

1

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

4 **and**

5 **Encoding Jet Substructure with a Deep Neural Network**

7 Dissertation

8 zur
9 Erlangung der naturwissenschaftlichen Doktorwürde
10 (Dr. sc. nat.)

11 vorgelegt der
12 Mathematisch-naturwissenschaftlichen Fakultät
13 der
14 Universität Zürich

15 von

16 Thea Klaeboe Arrestad
17 aus
18 Norwegen

19 Promotionskommission:

20 Prof. Dr. Ben Kilminster (Vorsitz)
21 Prof. Dr. Florencia Canelli
22 Prof. Dr. Jesse Thaler
23 Dr. Andreas Hinzmann

24 Zürich 2019

Abstract

28 In this doctoral thesis I will present three different searches for new heavy resonances with masses
29 above 1 TeV, decaying into pairs of vector bosons in the all-hadronic final state. The analyzed
30 data were collected by the CMS experiment at the Large Hadron Collider (LHC) during the first
31 three years of data-taking at a collision center-of-mass energy of 13 TeV, corresponding to an
32 integrated luminosity of 2.7(2015), 35.9(2016) and 77.3(2016+2017) fb^{-1} , and the searches were
33 the first of their kind to ever be performed at such a high collision energy. The diboson final
34 states under consideration are challenging to resolve due to the bosons being highly energetic
35 (“boosted”), resulting in the two quarks from the decay being collimated and merging into a single
36 jet. This leads to a dijet final state topology where each jet displays some energy substructure.
37 The first search I will present, was one of the two first CMS searches in boosted final states
38 with 13 TeV data to become published, and the first to take advantage of jet substructure at
39 the trigger level. It was a high profile analysis due to a previously observed 3.4 (1.3) σ excess
40 around 2 TeV in the 8 TeV dataset, as analyzed by ATLAS (CMS), and I brought the search to
41 a published result within six months after 13 TeV data taking began (and within ten months
42 after embarking on my PhD). Following this, in my second analysis I optimized, validated and
43 commissioned the novel PUPPI softdrop jet grooming algorithm for vector-boson tagging and,
44 in addition, developed dedicated mass corrections for the softdrop jet mass. The algorithm and
45 corresponding mass corrections are now the default for vector boson tagging in CMS and used
46 by several analyses. It was the first published result taking advantage of PUPPI softdrop jet
47 grooming. The third and final search I will present introduces a novel multidimensional search
48 framework, which can be used to search for resonances peaking anywhere in the 3D spectrum
49 of the dijet and groomed jet mass spectra. In this analysis, a simultaneous fit to the W and
50 Z jet mass peaks coming from the Standard Model $W(\bar{q}q)+\text{jets}$ and $Z(\bar{q}q)+\text{jets}$ background
51 has been performed, which has allowed for the extraction of the jet mass scale, resolution and
52 W-tagging efficiency in a way which has never been done before. Validated through a search with
53 hadronically decaying vector bosons in the final state, which I will present here, the framework
54 can be used to incorporate all resonance searches with hadronically decaying vector boson or
55 hadronically decaying Higgs boson final states, as well as for generic searches for any boosted
56 object peaking in jet mass. Finally, I will present a deep neural network for vector-boson tagging
57 which, through custom layers, encodes jet clustering- and substructure-like variables into the
58 deep layers themselves. This significantly improves the analysis sensitivity, and can also be used
59 as a stepping stone in the development of a generic anti-QCD tagger. The latter will be of
60 great importance when attempting to use the multidimensional framework for signal-independent
61 searches.

Zusammenfassung

In dieser Doktorarbeit werde ich drei verschiedene Suchen nach neuen starken Resonanzen vorstellen, die im all-hadronischen Endzustand in Paare von Vektorbosonen zerfallen. Die analysierten Daten wurden durch das CMS-Experiment am Large Hadron Collider (LHC) während der ersten drei Jahre der Datenaufnahme bei einer Kollisionsschwerpunktsenergie von 13 TeV gesammelt, und die Suchen waren die ersten ihrer Art jemals bei einer so hohen Kollisionsenergie durchgeführt werden. Die Dibosonendzustände, die in Betracht gezogen werden, sind schwierig zu lösen, da die Bosonen sehr energiereich (“boosted”) sind, was dazu führt, dass die beiden Quarks aus dem Zerfall sehr stark aufeinander abgestimmt sind und zu einem einzigen Jet zusammengeführt werden. Dies führt zu einer Dijet-Endzustands-Topologie, bei der jeder Jet eine Energieunterstruktur aufweist. Die erste Suche, die ich vorstellen werde, war eine der ersten CMS-Suchen in angehobenen Endzuständen mit 13 TeV Daten, die veröffentlicht wurden, und die erste, die die Jet-Substruktur auf Trigger-Ebene nutzte. Es war eine hochkarätige Analyse aufgrund eines beobachteten $3.4 (1.3) \sigma$ Überschusses um 2 TeV im 8 TeV Datensatz, der von ATLAS (CMS) analysiert wurde, und ich brachte die Suche innerhalb von sechs Monaten zu einem veröffentlichten Ergebnis nach 13 TeV wurde mit der Datenaufnahme begonnen (und innerhalb von zehn Monaten nach Beginn meiner Promotion). Durch die zweite Analyse wurde der neuartige PUPPI Softdrop Jet Grooming Algorithmus für das Vektor Boson Tagging optimiert, validiert und in Betrieb genommen. Außerdem wurden spezielle Massenkorrekturen für die Softdrop Jet Masse entwickelt. Der Algorithmus und die entsprechenden Massenkorrekturen sind jetzt der Standard für das Vektor Boson Tagging in CMS und werden von mehreren Analysen verwendet. Es war das erste veröffentlichte Ergebnis, bei dem die PUPPI Softdrop Jet grooming genutzt wurde. Die dritte und letzte Suche, die ich darlegen werde, führt ein neues multidimensionales Such-Framework ein, mit dem nach Resonanzen gesucht werden kann, die irgendwo im Dijet- und im Jet Groomed Massenspektrum ansteigen. Validiert durch eine Suche mit hadronisch zerfallenden Vektorbosonen im Endzustand, die ich hier vorstellen möchte, kann der Rahmen verwendet werden, um alle Resonanzsuchen mit hadronisch zerfallendem Vektorboson oder Higgs-Boson-Endzustand sowie generische Suchen nach geboosteten zu integrieren Objekt in Jet Masse.

Contents

95	1 The Standard Model and Beyond	15
96	1.1 The Standard Model	16
97	1.1.1 Fundamental particles: Quarks and leptons	16
98	1.1.2 The Standard Model Lagrangian	18
99	1.1.3 The electroweak sector	22
100	1.1.4 Spontaneous symmetry breaking and The Higgs Sector	25
101	1.2 Beyond Standard Model Physics	30
102	1.2.1 Theories of New Physics	31
103	2 Experimental setup	37
104	2.1 The Large Hadron Collider	38
105	2.2 The CMS detector	41
106	2.2.1 Coordinate system	42
107	2.2.2 Tracking detectors	43
108	2.2.3 Electromagnetic calorimeter	45
109	2.2.4 Hadronic calorimeter	47
110	2.2.5 Muon chambers	50
111	2.3 Trigger system	51
112	3 Event reconstruction	53
113	3.1 Track and primary vertex reconstruction	54
114	3.2 The Particle Flow Algorithm	55
115	3.2.1 Reconstruction of the Particle Flow inputs	55
116	3.2.2 Particle Flow identification	57
117	3.3 Pile-up removal	62
118	3.3.1 Charged Hadron Subtraction	62
119	3.3.2 Pile up per particle identification (PUPPI)	62

120	3.4	Jet reconstruction	64
121	3.4.1	Jet clustering	65
122	3.4.2	PF jets in CMS	67
123	3.4.3	Jet energy corrections	68
124	3.5	Jet substructure reconstruction	70
125	3.5.1	Grooming	71
126	3.5.2	N-subjettiness	74
127	3.5.3	Vector boson tagging	75
128	3.6	Monte Carlo Event Generators	78
129	4	Diboson resonance searches in CMS	81
130	4.1	Search I: First search for diboson resonances at 13 TeV	82
131	4.1.1	A small bump	83
132	4.1.2	Analysis strategy	84
133	4.1.3	Data and simulated samples	86
134	4.1.4	Event selection	87
135	4.1.5	Background modeling	103
136	4.1.6	Signal modeling	106
137	4.1.7	V-tagging scale factors	107
138	4.1.8	Systematic uncertainties	118
139	4.1.9	Results	119
140	4.1.10	Limits: All-hadronic analysis	119
141	4.1.11	Limits: Semi-leptonic and all-hadronic combination	120
142	4.2	Search II: A new pileup resistant, perturbative safe tagger	126
143	4.2.1	Towards robust boosted jet tagging	127
144	4.2.2	Analysis strategy	128
145	4.2.3	Data and simulated samples	129
146	4.2.4	Event selection	130
147	4.2.5	Developing a new W-tagger	134
148	4.2.6	W-tagging mistagging rate measurement	144
149	4.2.7	Mass and purity categorization	147
150	4.2.8	Background modeling: F-test	148
151	4.2.9	Signal modeling	148
152	4.2.10	Systematic uncertainties	149
153	4.2.11	Results	151

154	4.3	Search III: A novel framework for multi-dimensional searches	156
155	4.3.1	Small bumps and tri-bosons	157
156	4.3.2	Analysis strategy	158
157	4.3.3	Data and simulated samples	159
158	4.3.4	Event selection	159
159	4.3.5	Triggering	160
160	4.3.6	A mass and p_T decorrelated tagger	165
161	4.3.7	The multidimensional fit	174
162	4.3.8	Modeling of the signal	174
163	4.3.9	Modeling of the non-resonant background	176
164	4.3.10	Modeling of the resonant background	183
165	4.3.11	Systematic uncertainties	188
166	4.4	Background model validation	191
167	4.4.1	Variations of QCD multijet predictions	191
168	4.4.2	Fit to data control region	192
169	4.4.3	Bias test in pseudodata	194
170	4.4.4	Results	199
171	4.5	Summary and outlook	206
172	5	Encoding jet substructure in a deep neural network	209
173	5.1	Infusing deep neural networks with physics	210
174	5.2	LoLa	211
175	5.2.1	Architecture	211
176	5.2.2	Input	212
177	5.2.3	The Combination Layer	214
178	5.2.4	The Lorentz Layer	215
179	5.3	Performance	217
180	5.4	p_T and mass dependence	218
181	5.4.1	p_T decorrelation	219
182	5.4.2	Mass sculpting	221
183	5.5	Validation on an independent sample	223
184	5.6	Summary and outlook	226
185	6	Conclusion and outlook	227

186	A Search I: Limits per mass category	231
187	A.1 Limits per mass category	231
188	B Search II	241
189	B.1 W-tagging scale factor measurement: Additional plots	242
190	B.2 Efficiency scale factors for 2.5 fb^{-1}	242
191	B.3 Background fit checks	242
192	B.3.1 Background fit checks in data sideband	245
193	B.3.2 Background fit checks in QCD MC	250
194	B.4 F-test in signal region	251
195	C Search III	261
196	C.1 Signal fits	262
197	C.2 2016 kernels	262
198	C.3 Resonant background	262

Introduction

200 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, 201 and it has allowed us to incorporate all of particle physics into one single equation that explains 202 what we can see in the world around us. Almost all. Its greatest shortcoming is its failure to 203 successfully incorporate gravity, leaving large scale phenomena unexplained. This, together with 204 a few other shortcomings, has lead scientists to search for extensions to the Standard Model, 205 commonly referred to as *Beyond Standard Model physics (BSM)*. These models are usually 206 accompanied by predicted observables not included in the Standard Model, where the observation 207 of these, or the lack thereof, is a way of falsifying or supporting the model. In this thesis, I 208 attempt to do exactly that by searching for new massive particles predicted by SM extensions. 209 These particles have the property that they can decay into vector bosons, W^\pm and Z^0 , and 210 usually have a very small interaction probability. The vector bosons are heavy and unstable 211 and will quickly decay into leptons or quarks, and in order to make up for the small interaction 212 probability, I look for two vector bosons decaying hadronically, which occurs $\sim 70\%$ of the time, 213 significantly more frequently than leptonic decays. This final state is complicated by the presence 214 of an overwhelming QCD multijet background, and the fact that, due to the high mass of the 215 resonance, the vector bosons are highly energetic and their quark decay products are so collimated 216 they get merged into a single jet. Fortunately, the latter offers an opportunity to separate 217 vector boson jets from quark and gluon jets through the jets mass and how many prongs the jet 218 appears to have. Methods to do this separation are referred to as *jet substructure methods* and 219 is something I will heavily focus on in this thesis, due to my own personal contributions to the field. 220

221

222 Three analyses will be described, where each has benefitted from novel findings in the pre- 223 vious: The first search was the first analysis of its kind to ever be performed at a center-of-mass 224 energy of $\sqrt{s} = 13 \text{ TeV}$ and the first published result to take advantage of jet substructure at 225 trigger level. The second led to the development of a novel pileup robust and perturbative 226 safe vector boson tagging algorithm, which afterwards became the default tagging algorithm in

227 CMS. Finally, the third search introduces a completely new way of doing diboson searches in a
228 multi-dimensional space, allowing for the incorporation of all VV, VH and HH searches (where
229 $V = W, Z$ and $H = \text{Higgs boson}$) into one common framework as well as any generic search for
230 resonances peaking in jet mass and dijet invariant mass. I have performed every aspect of the
231 three analyses, making original contributions to all, despite the latter where we were a small
232 team of three analysts dividing the workload.

233

234 In the final chapter of this thesis, I will introduce an ongoing work on a deep neural net-
235 work for vector boson tagging intended to improve the analysis sensitivity for future searches and
236 which in addition could be used to develop a generic anti-QCD tagger due to its deep encoding of
237 jet substructure. Such a tagger would, in combination with the multidimensional fit framework,
238 lead to a completely new way of doing model independent searches.

239

240 This thesis is organized in the following way. In Chapter 1, I will go through the theoreti-
241 cal motivations behind the searches presented here through introducing the Standard Model,
242 its known shortcomings and possible alternatives. This is followed by a description of the
243 experimental setup used to collect the analyzed data, in Chapter 2, as well as the different
244 algorithms used in order to reconstruct each event in Chapter 3. The remaining part of the thesis
245 is dedicated to my own personal contributions: The three searches are presented chronologically
246 in Chapter 4, each with a personal introduction motivating the search in question. Following this,
247 and ending my thesis work, is a description of the deep neural network based vector boson tagger
248 for future analyses, in Chapter 5. Both chapters end with their own concluding summary and
249 outlook, Section sec:searchIII:outlook and Section sec:lola:outlook. To bring everything together,
250 Chapter 6, the final chapter of this thesis, provides a final conclusion of the work presented.

251

CHAPTER 1

252

The Standard Model and Beyond

²⁵³ **1.1 The Standard Model**

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

²⁶⁸ **1.1.1 Fundamental particles: Quarks and leptons**

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

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

276 The charged leptons can be positively or negatively charged, defined in units of electron charge
 277 e . By convention the leptons of matter are negatively charged, e^- , μ^- , and τ^- , whereas the
 278 positively charged leptons, e^+ , μ^+ , and τ^+ are considered their anti-particles. A summary of
 279 the lepton properties is listed in Table 1.1. Leptons interact with one another through the
 280 *electromagnetic and weak force*, which will be explained in more detail in Section 1.1.3.
 281 The other six fundamental particles of matter are the *quarks*. They are distinguished from the

Lepton	Mass	Charge
e^-	0.5 MeV	e
μ^-	106 MeV	e
τ^-	1777 MeV	e
ν_e	< 3 eV	0
ν_μ	< 0.19 MeV	0
ν_τ	< 18.2 MeV	0

Table 1.1: Lepton Properties [1].

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

leptons in that they interact with one another through the *strong force*, described in Section 1.1.2. 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/meson bound states. 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. As mentioned above, quarks can interact with one another through the strong force. However, they also interact through the weak and electromagnetic forces. 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 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 photons. All of the fermions, including the electrically neutral neutrinos, feel the weak force and undergo weak interactions. The weak force is mediated through vector bosons (PW^+ , PW^- 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 electrons and neutrinos. The strong force keeps us from observing quarks as free particles, and keeps them in bound states referred to as *hadrons*. 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.

1.1.2 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 , which 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

328 multiplet, but never to a field outside of that group. The symmetry group of the Standard Model
329 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 $\mathcal{G}_{SM} \ni x = (C, L)_{(Y)}$ (here for the 1st generation only) and are:

$$\begin{aligned} Q &= (3, 2)_{(1/3)} = \begin{pmatrix} u_r & u_g & u_b \\ d_r & d_g & d_b \end{pmatrix} && \sim \text{quark multiplet} \\ L &= (1, 2)_{(-1)} = \begin{pmatrix} \nu_e & e \end{pmatrix} && \sim \text{leptonic doublet} \\ u^c &= (\bar{3}, 1)_{(-4/3)} = \begin{pmatrix} u_{\bar{r}}^c & u_{\bar{g}}^c & u_{\bar{b}}^c \end{pmatrix} && \sim \text{anti-up quarks} \\ d^c &= (\bar{3}, 1)_{(2/3)} = \begin{pmatrix} d_{\bar{r}}^c & d_{\bar{g}}^c & d_{\bar{b}}^c \end{pmatrix} && \sim \text{anti-down quarks} \\ e^c &= (1, 1)_{(2)} && \sim \text{positron} \end{aligned}$$

330 The anti-neutrino is not included here because it is a SM *singlet* and does not transform under
331 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,
332 is included as its hypercharge is non-zero and it therefore undergoes $U(1)_Y$ interactions. These
333 five multiplets exist for each of the tree generations.

334 The numbers representing each multiplet correspond to which representation it belongs to. For
335 instance, we see that the quark multiplet transforms as a triplet under $SU(3)_C$, a doublet under
336 $SU(2)_L$ and has a non-zero hypercharge, corresponding to a non-trivial representation under
337 $U(1)_Y$. That corresponds to saying that quarks interact through all of the three fundamental
338 interactions. From that notation, it is also clear that leptons do not carry color charge and will
339 only interact via the electroweak interactions.

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

³⁴⁵ $U(1)_Y$.

³⁴⁶ These interactions; the strong, weak and electromagnetic, will be explained in more detail in the
³⁴⁷ following sections.

³⁴⁸ The Quantum Chromodynamics sector

³⁴⁹ The group $SU(3)_C$ describe the strong interaction mediated by gluons, and is described by
³⁵⁰ the quantum gauge theory Quantum Chromodynamics (QCD). The group is generated by
³⁵¹ 8 linearly independent matrices $T^a = \frac{\lambda^a}{2}$, where λ^a are the Gell-Mann matrices [2]. The
³⁵² generator matrices do not commute with one another, but rather satisfy the commutation relation
³⁵³ $[\lambda^i/2, \lambda^j/2] = if^{ijk}\lambda_k/2$. This property makes $SU(3)_C$ group *non-Abelian*, which consequently
³⁵⁴ results in the gluons themselves being charged and display self-interactions. Gluons are charged
³⁵⁵ with one unit of color and one unit of anti-color. Quarks, the only fundamental particles interacting
³⁵⁶ with the strong force, are arranged in the simplest representation of $SU(3)$ and come with one
³⁵⁷ unit of color or anti-color.

³⁵⁸ The generators are collectively referred to as the *group representation* of $SU(3)_C$, because any
³⁵⁹ group element can be written as $e^{-i\theta^a g_a}$, where a runs from 1 to 8, θ^a are real numbers
³⁶⁰ and g_a represent one of the eight linearly independent $\lambda/2$ matrices. That means that, given one
³⁶¹ representation, one can always find another one through any local gauge transformation that
³⁶² leaves the commutator unchanged. In this case, the group elements are the quark fields, and a
³⁶³ local gauge transformation of fields become

$$\Psi(x^\mu) \rightarrow e^{-ig_s\theta^a(x^\mu)T^a} \Psi(x^\mu) \quad (1.5)$$

³⁶⁴ where g_s is the strong coupling, $\theta^a(x^\mu)$ some arbitrary function and a runs over the eight generators
³⁶⁵ of the group. In order to keep the QCD Lagrangian invariant under such a transformation, an
³⁶⁶ additional term must be added, replacing the partial derivative ∂_μ with the *covariant derivative*

$$D_\mu = \partial_\mu + ig_s A_\mu^a T^a, \quad (1.6)$$

³⁶⁷ where we now have introduced a tensor A_μ^a that represents the 8 gluon fields. The QCD Lagrangian
³⁶⁸ then becomes

$$\mathcal{L}_{QCD} = -\frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^a + \bar{\Psi}(x^\mu)(i\lambda^\mu D_\mu - m)\Psi(x^\mu), \quad (1.7)$$

³⁶⁹ where $F_{\mu\nu}^a$ is the gauge field of the group, the gluon field tensor

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f^{abc} A_\mu^b A_\nu^c. \quad (1.8)$$

The first term in Equation 1.7 represents the quark-gluon interaction, leading to vertices like the one on the right in Figure 1.1. The second, the gluon field kinetic term, pick 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

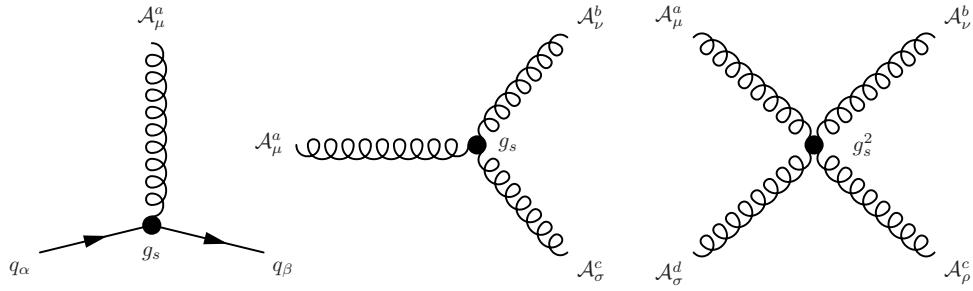


Figure 1.1: The QCD interaction vertices: Quark interaction with the gluon field (left), and three- (middle) and four-gluon (right) self-interaction vertices.

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

$$V(r) = -\frac{4\alpha_s}{3r} + kr, \quad (1.9)$$

where r is the distance between the quarks and α_s is the coupling strength of QCD, describing how the observed charge between two quarks increases depending on the distance between them. 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 $\Lambda_{QCD} = 100 - 500$ MeV (around the mass of the lightest hadrons). The effective

³⁹³ charge between the quarks, α_S , changes as a function of energy as

$$\alpha_S(Q) = -\frac{6\pi}{33 - 2n_f} \ln(Q/\Lambda_{QCD}) \quad (1.10)$$

³⁹⁴ where Q is the energy of the probe used to measure the charge and n_f is the number of quark
³⁹⁵ flavors (u, d, c, s, b, t) at that energy. α_S is around 0.1 for energies between 100-1000 GeV.

³⁹⁶ 1.1.3 The electroweak sector

³⁹⁷ The electromagnetic and weak interactions arise from the breaking of $SU(2)_L \otimes U(1)_Y$ symmetry.
³⁹⁸ While the unification of the electromagnetic and weak force is obtained under the $SU(2)_L \otimes U(1)_Y$
³⁹⁹ group, the predicted gauge bosons of such a group are not observed in nature (three charged
⁴⁰⁰ massless vector bosons, and one neutral massless boson). Rather, the W^\pm , Z^0 and the photon
⁴⁰¹ arise from the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y$ to $U(1)_{EM}$. This happens
⁴⁰² due to the *Higgs mechanism*, and exactly how this occurs will be the topic of Section 1.1.4.

⁴⁰³ The symmetry under $SU(2)_L$ is called weak left-handed isospin L , and the symmetry under
⁴⁰⁴ $U(1)_Y$ is the weak hypercharge Y . The name “left-handed” arise from the fact that *parity* is
⁴⁰⁵ violated in the electroweak interactions. All the fundamental fermions have a *chirality*, defined as
⁴⁰⁶ the projection of the particles spin along its direction of motion. From observations, the weak
⁴⁰⁷ interactions is observed to only interact with fermions of left-handed chirality (vector minus axial
⁴⁰⁸ coupling, V-A). The left-handed fermion fields are therefore in the simplest doublet representation
⁴⁰⁹ of $SU(2)$ with weak isospin $I = 1/2$, while the fermions of right-handed chirality are in the singlet
⁴¹⁰ representation with weak isospin $I = 0$, meaning they do not interact with the gauge bosons of
⁴¹¹ $SU(2)_L$. To obtain the correct chiral components Ψ_L and Ψ_R , The chirality of any fermion Ψ can
⁴¹² be defined through the operator γ^5 , the product of the four Dirac matrices [3] $\gamma^5 = i\gamma^1\gamma^2\gamma^3\gamma^4$,
⁴¹³ which has eigenvalues ± 1 . Any Dirac field can be projected into its chiral components Ψ_L or Ψ_R
⁴¹⁴ through the projection operation

$$\Psi_L = \frac{1 - \gamma^5}{2} \quad \text{and} \quad \Psi_R = \frac{1 + \gamma^5}{2}. \quad (1.11)$$

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 [4]. 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 [5]. This in turn implies self-interactions between the gauge bosons of the group, as

was the case for $SU(3)_C$. The left-handed fermion fields are doublets under $SU(2)_L$, while the right-handed components transform as singlets and hence do not interact with the gauge field tensor.

The latter group, $U(1)_Y$ of weak hypercharge Y , is abelian and hence display no self-interaction. The gauge field tensor $B_{\mu\nu}^a$ interacts with both left- and right-handed Dirac 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

$$D_\mu \Psi_L = (\partial_\mu + ig_2 T_a W_\mu^a + ig_1 \frac{Y}{2} B_\mu^a) \Psi_L \quad (1.12)$$

$$D_\mu \Psi_R = (\partial_\mu + ig_1 \frac{Y}{2} B_\mu^a) \Psi_R. \quad (1.13)$$

⁴¹⁵ The electroweak Lagrangian can be written as a sum of four terms

$$\mathcal{L}_{EW} = \mathcal{L}_{gauge} + \mathcal{L}_f + \mathcal{L}_{Yukawa} + \mathcal{L}_\phi. \quad (1.14)$$

⁴¹⁶ The first term, \mathcal{L}_{gauge} , represent the kinetic field tensor and is

$$\mathcal{L}_{gauge} = -\frac{1}{4} W_a^{\mu\nu} W_{\mu\nu}^a - \frac{1}{4} \mathcal{B}^{\mu\nu} \mathcal{B}_{\mu\nu} \quad (1.15)$$

where the gauge field tensors are

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g_2 \epsilon^{ijk} W_\mu^j W_\nu^k \quad , \text{ with } i = 1 - 3 \quad (1.16)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (1.17)$$

⁴¹⁷ The non-abelian nature of $SU(2)_L$ leads to trilinear and quadrilinear couplings between the
⁴¹⁸ photon and vector bosons as illustrated in Figure 1.2.

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

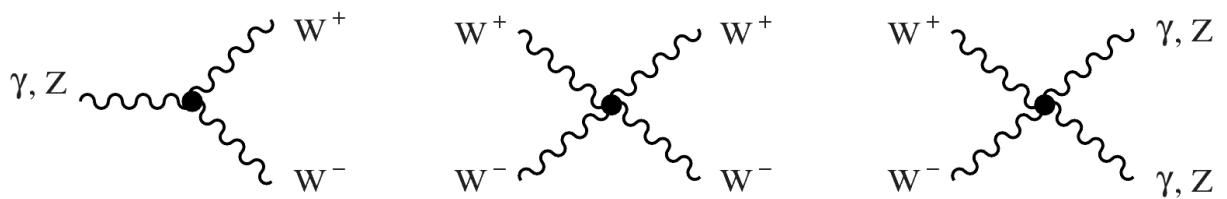


Figure 1.2: The electroweak self-interaction vertices: trilinear (left), and quadri-linear (middle and right) vertices.

419

$$\mathcal{L}_f = \bar{Q}_i i\lambda^\mu D_\mu Q_i + \bar{u}_i^c i\lambda^\mu D_\mu u_i^c + \bar{d}_i^c i\lambda^\mu D_\mu d_i^c + \bar{L}_i i\lambda^\mu D_\mu L_i + \bar{e}_i^c i\lambda^\mu D_\mu e_i^c, \quad (1.18)$$

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

423 Up until now we have considered the Lagrangian before spontaneous symmetry breaking, where
 424 we have three charged massless bosons and one massless neutral boson, a constellation we know
 425 to be wrong from observations. When I have referred to interaction vertices, I have loosely
 426 referred to W , Z and γ vertices without explicitly defining them (and we have only been working
 427 with the covariant derivatives as defined in Equation 1.12). We will show in Section 1.1.4 that,
 428 due to spontaneous symmetry breaking, the charged gauge boson fields W^\pm in reality are linear
 429 combinations of W_μ^1 and W_μ^2 :

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \quad (1.19)$$

430 These are responsible for *charged current interactions*, which turn up-type fermions into their
 431 corresponding down-type fermions within the same generation, and vice-versa. The electrically
 432 neutral Z boson and the photon fields are expressed in terms of W_μ^3 and B_μ through the weak
 433 mixing angle [6] as

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix} \quad (1.20)$$

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

$$\cos \theta_W = \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \quad \text{and} \quad \sin \theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}. \quad (1.21)$$

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

$$Q = I_3 + \frac{Y}{2}. \quad (1.22)$$

436 A few of the fermion-boson vertices are shown in Figure 1.3. The last two terms, \mathcal{L}_{Yukawa} and \mathcal{L}_ϕ ,
 437 are related to the Higgs boson and represent couplings to the Higgs field: its interactions with
 438 itself and the gauge bosons, and the Yukawa interaction with the fermions that generate their
 439 masses due to the non-zero Higgs vacuum expectation value. Introduced as an extension to the
 440 original SM, the Higgs sector is one of the great accomplishments of particle physics and one of
 441 the reasons why the LHC was built. How it arises is what we will turn to next.

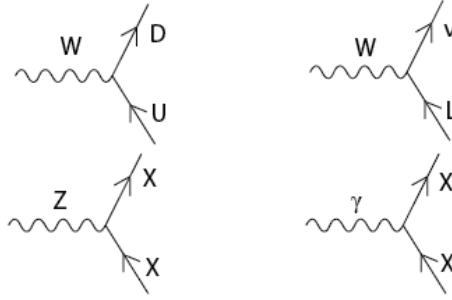


Figure 1.3: The electroweak fermion interaction vertices. Top: Charged current interaction vertex connecting a charged vector boson W^\pm to quarks (left) and leptons (right). Bottom: Neutral current interactions between the neutral Z^0 boson and any fermion (left), and between a γ and electrically charged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

442 1.1.4 Spontaneous symmetry breaking and The Higgs Sector

443 The problem of having massless gauge bosons under $SU(2)_L \times U(1)_Y$, while observing massive W
 444 and Z bosons, was independently solved by three different groups and has become known as the
 445 Englert–Brout–Higgs–Guralnik–Hagen–Kibble mechanism of spontaneous symmetry breaking [7–
 446 9], an accomplishment for which Peter Higgs and Francois Englert shared the 2013 Nobel Prize.
 447 It began with the realization that the breaking of a local gauge symmetry could give rise to a final
 448 mass for the gauge boson involved. This was first discovered in association with superconductivity,
 449 where it was found that when a normal metal becomes superconducting the photon field inside
 450 the superconductor would acquire a finite mass [10].

451 Spontaneous breaking of a global gauge symmetry

452 In order to achieve spontaneous symmetry breaking, Jeffrey Goldstone [11] suggested introducing
 453 a massive complex scalar field ϕ with quantum numbers of the vacuum and then give the field a
 454 vacuum expectation value. The field

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2) \quad (1.23)$$

455 would have a Lagrangian of the form

$$\mathcal{L} = \partial^\mu \bar{\phi} \partial_\mu \phi - \mu_0^2 \bar{\phi} \phi - \frac{\lambda_0}{6} \bar{\phi} \phi^2, \quad (1.24)$$

456 where λ_0 is the coupling constant and μ_0 is the mass. The Lagrangian is invariant under $U(1)$,
 457 though in this case under a global symmetry and not a local one. If one takes μ_0^2 to be negative,

458 the potential will get a minima along a circle of radius v such that

$$\phi_1^2 + \phi_2^2 = v^2 \quad \text{and} \quad v^2 = \frac{\mu_0^2}{\lambda_0} \quad (1.25)$$

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

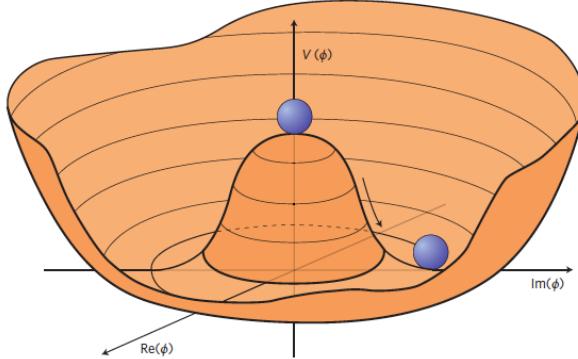


Figure 1.4: The potential $V(\phi)$ for a complex scalar field with $\mu_0^2 < 0$ [12].

459

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

$$\phi(x) = \frac{1}{\sqrt{2}} * (v + \eta(x) + i\epsilon(x)) \quad (1.26)$$

465 The Lagrangian then becomes

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \epsilon)^2 + \frac{1}{2}(\partial_\mu \eta)^2 + \mu_0^2 \eta^2. \quad (1.27)$$

466 The third term has the form of a mass term for the scalar η field. However, the ϵ field has no
467 mass term meaning that the theory contains a massless scalar, referred to as a *Goldstone boson*.
468 This is expressed through *Goldstones theorem*, which states that whenever a continuous symmetry
469 of a physical system is spontaneously broken, massless scalars will occur. So rather than solving
470 the problem of massless vector bosons, the theory acquired a massless scalar not observed in
471 Nature and the gauge theory of weak interactions had to look elsewhere for solutions.

⁴⁷² **The Higgs mechanism**

⁴⁷³ The solution to the problem came a few years later, in 1964, when spontaneous symmetry breaking
⁴⁷⁴ of a *local* gauge symmetry was studied rather than a global. For a $U(1)$ symmetry, this requires
⁴⁷⁵ the Lagrangian to be invariant under $\phi \rightarrow e^{i\theta(x)}\phi$ with the usual derivative replacement

$$D_\mu = \partial_\mu - ieA_\mu. \quad (1.28)$$

⁴⁷⁶ After translating the field ϕ to its true ground state and writing out the Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\epsilon)^2 + \frac{1}{2}(\partial_\mu\eta)^2 - v^2\lambda\eta^2 + \frac{1}{2}v^2e^2A_\mu A^\mu - evA_\mu\partial^\mu\epsilon - \frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a \quad (1.29)$$

⁴⁷⁷ we see that the particle spectrum now contains a massless Goldstone boson ϵ , a massive scalar
⁴⁷⁸ η and, finally, a massive vector A_μ with $m_A = ev$. After having succeeded with dynamically
⁴⁷⁹ creating the mass for the gauge field, one had to tackle the problem of the massless scalar. The
⁴⁸⁰ solution was found through the realization that one of the fields were unphysical; by giving mass
⁴⁸¹ to the vector A_μ , the polarization degrees of freedom had increased from 2 to 3 through adding
⁴⁸² a longitudinal polarization. However, this should not be possible when simply translating field
⁴⁸³ variables. It was found that through a simple gauge transformation with a different set of fields

$$A_\mu \rightarrow A_\mu + \frac{1}{ev}\partial_\mu\theta, \quad (1.30)$$

⁴⁸⁴ the Goldstone boson would disappear and turn into the longitudinal polarization of the massive
⁴⁸⁵ gauge boson and the theory was left with one massive vector gauge boson A_μ and another massive
⁴⁸⁶ scalar h . This is what is referred to as the *Higgs mechanism*.

⁴⁸⁷ **The Weinberg-Salam Model**

The final step is to formulate the Higgs mechanism such that the vector bosons W^\pm and Z^0 becomes 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:

$$\mathcal{L}_{Higgs} = \left| (i\partial_\mu - g_2 T_a W_\mu^a - vg_1 \frac{Y}{2} B_\mu^a) \phi \right|^2 - V(\phi) \quad (1.31)$$

- ⁴⁸⁸ The ϕ_i has to belong to $SU(2) \times U(1)$ multiplets, and the simplest choice is four fields in an
⁴⁸⁹ isospin doublet of weak hypercharge $Y = 1$:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \frac{\phi_1 + i\phi_2}{\sqrt{2}} \\ \frac{\phi_3 + i\phi_4}{\sqrt{2}} \end{pmatrix}. \quad (1.32)$$

- ⁴⁹⁰ To generate the masses, the Higgs potential from the section above is used, with a vacuum
⁴⁹¹ expectation value of

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.33)$$

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 term for the gauge bosons then finally become

$$M_W = \frac{1}{2} v g_1 \quad (1.34)$$

$$M_Z = \frac{1}{2} v \sqrt{g_1^2 + g_2^2} \quad (1.35)$$

$$\frac{M_W}{M_Z} = \cos \theta_W. \quad (1.36)$$

- ⁴⁹² As mentioned in Section 1.1.3, the fermions also get their mass through interaction with the
⁴⁹³ Higgs field, \mathcal{L}_{Yukawa} . This is done the same way as for the boson: an additional $SU(2) \times U(1)$
⁴⁹⁴ invariant term for each generation is added, for instance for the electron

$$-G_e \left[\begin{pmatrix} \bar{\nu}_e & \bar{e} \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} e^R + \bar{e}_R \begin{pmatrix} \phi^- & \bar{\phi}^0 \end{pmatrix} \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \right], \quad (1.37)$$

- ⁴⁹⁵ where G_E is the electron coupling. We then spontaneously break the symmetry with

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad (1.38)$$

- ⁴⁹⁶ where the neutral Higgs field $h(x)$ is the only remnant of the Higgs doublet after spontaneous
⁴⁹⁷ symmetry breaking. After substitution, the final Lagrangian for the electron mass becomes

$$\mathcal{L}_{Yukawa}^e = -\frac{G_e}{\sqrt{2}} v (\bar{e}_L e_R + \bar{e}_R e_L) (1 + \frac{h}{v}). \quad (1.39)$$

⁴⁹⁸ We can choose G_e such that

$$m_e = -\frac{G_e v}{\sqrt{2}} \quad (1.40)$$

⁴⁹⁹ and generate the electron mass as

$$\mathcal{L}_{Yukawa}^e = -m_e \bar{e} e - \frac{m_e}{v} \bar{e} e h. \quad (1.41)$$

⁵⁰⁰ In summary, all the fermion masses are generated through the interaction of the fermion fields
⁵⁰¹ with the Higgs field. From the equation above, we see that the fermion masses are not predicted
⁵⁰² from the theory as the coupling G is arbitrary. The Standard Model therefore cannot provide an
⁵⁰³ explanation the difference in hierarchy between the couplings. We also see that the Lagrangian
⁵⁰⁴ contains an interaction term coupling the Higgs scalar to the fermions and that this term depends
⁵⁰⁵ on the mass of the fermion. The Higgs boson therefore couples more strongly to heavy fermions
⁵⁰⁶ than to lighter ones.

507 1.2 Beyond Standard Model Physics

508 Despite being an extremely successful and predictive theory, the Standard Model has its short-
509 comings. The most obvious one is its failure to successfully incorporate the gravitational force.
510 Gravity is beautifully described in General Relativity as a classical theory: a force caused by
511 the curvature of space-time in the presence of matter and energy. The theory does not utilize
512 quantum fields and energy is not quantized. The scales between the Standard model, a quantum
513 field theory, and General Relativity are completely different: space-time is curved on astronomical
514 scales, where the force of gravity is measurable, while quantum field theories deal with things on
515 the smallest possible scales, where variations in space-time are essentially invisible. Hence, to the
516 Standard Model, space-time is approximately flat and gravity does not exist. In order to have an
517 elegant unified theory of all the forces, attempts has been made to have a quantum field theory of
518 the gravitational force by extending the Standard Model particle family to incorporate a particle
519 to mediate the gravitational force called the *graviton*, a massless gauge boson of spin-2. The
520 problem is that gravity is universally attractive, meaning nothing “cancels” it. That leads to loop
521 divergences that cannot be reabsorbed through renormalization and every effort of integrating
522 gravity in the SM has thus far failed. However, it has been shown that General Relativity is
523 an inevitable consequence of the quantum mechanics of interacting gravitons, which has led to
524 several proposals for extending the SM models to incorporate them.

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

540 Of course, it is possible that the Higgs boson mass just happens to be 125 GeV due to some

541 fine-tuned, large cancellations that keep the mass from approaching the Planck mass, as is
 542 currently held by the Standard Model. However, this is neither very elegant nor very probable
 543 without a well-motivated reason why such a cancellation should occur. Rather, in order to resolve
 544 the problem of scales, theories Beyond the Standard Model (BSM) have been introduced. The
 545 theories that I will probe in this thesis are amongst those.

546 1.2.1 Theories of New Physics

547 Two Beyond Standard Model theories will be considered in this thesis: Compositeness and
 548 extra dimensional theories. Compositeness attacks the hierarchy problem by assuming that the
 549 Standard Model breaks down at an energy between the weak and Planck scales and that, around
 550 the TeV scale, the Higgs boson no longer appears to be a single scalar particle but a composite
 551 state of two fermions. In the following, I will present the study of composite models in the context
 552 of the *Heavy Vector Triplet formalism*, described in Section 1.2.1. Warped extra dimensional
 553 theories attempt to solve the hierarchy problem by concentrating gravity on another “brane” and
 554 letting its strength fall off exponentially through an extra dimension. Both of these, compositeness
 555 and extra dimensional theories, will be described in the following.

556 Compositeness

557 In composite Higgs models, the Higgs boson is assumed to be a bound state of fundamental
 558 constituents held together by some new strong force [13, 14]. This removes the hierarchy problem
 559 since we no longer have an elementary scalar in the Standard Model, hence no loop corrections
 560 going to infinity. The compositeness of the Higgs boson becomes apparent at the energy scale Λ ,
 561 where Λ has to be at least 10 TeV, since anything below that is ruled out by electroweak precision
 562 measurements. The Higgs boson is assumed to be a pseudo-Goldstone boson of some approximate
 563 symmetry, where pseudo-Goldstone bosons are bosons with a tiny mass that approach zero in
 564 the limit of the symmetry becoming exact. The approximate symmetry is broken at the scale
 565 f , where $\Lambda = 4\pi f$. Being a pseudo-Goldstone boson, the Higgs boson mass is protected from
 566 divergent quantum loop corrections up to the scale of compositeness and, above that scale, is no
 567 longer an elementary scalar. The theory is based on the breaking of two parallel, large global
 568 symmetries $[SU(2)_1 \times U(1)_1] \times [SU(2)_2 \times U(1)_2]$, with Goldstone bosons becoming the longitudinal
 569 components of the three predicted gauge bosons of the symmetry group W'^{\pm} and Z' . These have
 570 masses of the order of the compositeness scale

$$M(W'^{\pm}) \simeq M(Z') = \frac{g}{\sin 2\theta} f, \quad (1.42)$$

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

$$\begin{aligned}\Gamma(W'^{\pm} \rightarrow \ell\nu, Z' \rightarrow \ell\ell) &= \frac{g^2 \cot^2 \theta}{96\pi} M \\ \Gamma(W'^{\pm} \rightarrow q\bar{q}', Z' \rightarrow q\bar{q}) &= \frac{g^2 \cot^2 \theta}{32\pi} M \\ \Gamma(W'^{\pm} \rightarrow WZ, Z' \rightarrow WW) &= \frac{g^2 \cot^2 2\theta}{192\pi} M\end{aligned}\quad (1.43)$$

573 Decays into fermions therefore dominate at $\cot \theta \geq 1/2$, whereas decays into bosons are enhanced
 574 for very low $\cot \theta$.

575 These generic composite models can be studied with the Heavy Vector Triplet formalism.

576 Heavy Vector Triplet formalism

577 There are many BSM theories that predict the presence of spin-1 particles with masses at the
 578 TeV scale, each with their own list of model parameters. In most cases, however, when looking for
 579 such new particles, experiments are not sensitive to the specifics of the model but only the masses
 580 and couplings of the resonances. We can therefore start from a *simplified model* that describes
 581 the dynamics of the new spin-1 vector through a simple phenomenological Lagrangian that only
 582 retains couplings and mass. In the Heavy Vector Triplet formalism [15], a real vector V_μ^a , where
 583 r runs from 1 to 3, is introduced in the adjoint representation of $SU(2)L$ and represents one
 584 charged and one neutral heavy spin-one particle with charge eigenstates

$$V_\mu^\pm = \frac{V_\mu^1 \mp iV_\mu^2}{\sqrt{2}} \quad \text{and} \quad V_\mu^0 = V_\mu^3. \quad (1.44)$$

585 The simplified Lagrangian governing the dynamics is given as

$$\begin{aligned}\mathcal{L}_V = -\frac{1}{4} \mathcal{D}_{[\mu} V_{\nu]}^a \mathcal{D}^{[\mu} V^{\nu]}_a + \frac{m_V^2}{2} V_\mu^a V^{\mu a} \\ + ig_V c_H V_\mu^a H^\dagger \tau^a \overleftrightarrow{\mathcal{D}}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a J_F^{\mu a} \\ + \text{additional terms.}\end{aligned}\quad (1.45)$$

586 The first line describes the kinematic and mass terms of the vector V , and the second line, which
 587 is of most interest to us, describes the coupling to the Higgs boson current and the left-handed
 588 fermionic currents. In the coupling to the Higgs current, the coefficient c_H leads to vertices
 589 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 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 [15]. 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 [15]. 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 1.5. For obvious reasons, model B is of most interest for the searches presented here.

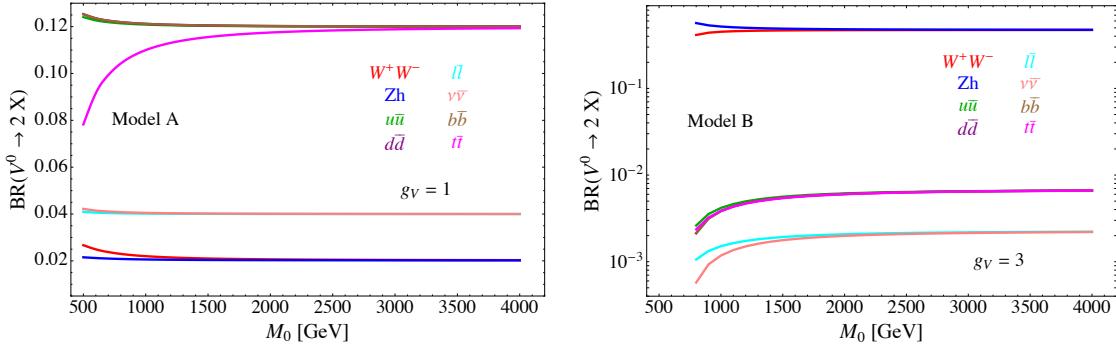


Figure 1.5: Predicted branching fractions of a Z' for two explicit HVT models: Model A $_{g_V=1}$ (left) and model B $_{g_V=3}$ (right) [15].

609 **Warped extra dimensions**

610 Extra dimensional theories also offer solutions to the hierarchy problem. This thesis focuses on
611 Randall-Sundrum (RS) warped extra dimensional scenarios [16]. In RS models, a new curved
612 spatial dimension y is proposed, leading to a 5-dimensional space-time bounded by two (3+1)-
613 dimensional planes, or *branes*: the UV/Planck and the IR/TeV brane. The new metric now
614 depends on the radius r and the curvature factor k of the new extra dimension

$$ds^2 = e^{-2ky} \eta_{\mu\nu} dx^\mu dx^\nu + dy^2; 0 < y < \pi R \quad (1.46)$$

615 Gravity is concentrated and relatively strong at the Planck brane at $y = 0$, which is separated
616 from us by the fifth dimension. Our observed four-dimensional reality and the Standard Model
617 particles reside at the TeV brane, at $y = \pi R$. Only gravity, transmitted through gravitons, is
618 allowed to propagate through the warped 5D space-time (the “bulk”) and is not confined to
either brane. Figure 1.6 illustrates how the branes and the bulk are connected. Due to the

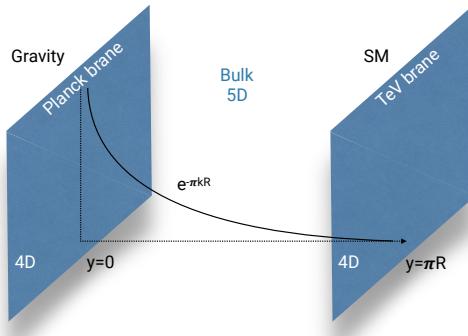


Figure 1.6: The RS model predicts an extra dimension where a 5D space-time stretches between two 4D branes: the Planck brane where gravity is concentrate, and the TeV brane where the SM particles are confined.

619
620 warping, the Planck mass on the Planck brane gets reduced by a factor of $e^{-\pi k R}$ at the TeV
621 brane, thereby solving the hierarchy problem. The Planck mass on the TeV brane, which depends
622 on the geometry of the extra dimension, becomes

$$\bar{M}_{Pl}^2 = V_1 M_*^3, \quad (1.47)$$

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 [17, 18]. 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 of the new resonances.

The branching ratios of the Bulk Graviton is shown in Figure 1.7.

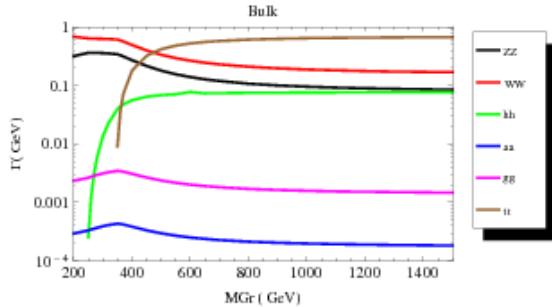


Figure 1.7: Predicted branching fractions for a Bulk Graviton.

641

CHAPTER 2

642

Experimental setup

643 **2.1 The Large Hadron Collider**

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

662 The Large Hadron Collider first circulated protons in September 2008 and, while having
663 the same 27-kilometer radius as the LEP collider, is capable of accelerating protons up to a
664 center-of-mass energy of around 14 TeV, 70 times that of LEP. The accelerator consists of two
665 oppositely circulating proton beams, isolated from each other and under ultrahigh vacuum. The
666 protons are accelerated up to speeds close to the speed of light through radio frequency (RF)
667 cavities, before being focused to collide at four different interaction points along the ring. These
668 four collision points correspond to the location of the four LHC particle detectors; ATLAS, CMS,
669 LHCb and ALICE. While ATLAS and CMS are general-purpose detectors, built in order to study
670 a large range of different physics processes, LHCb and ALICE are built for dedicated purposes;
671 LHCb for b-physics processes and ALICE for heavy ion collisions. The journey of a proton from
672 a gas to one of the LHC collision points is as follows. First, hydrogen nuclei are extracted from a
673 small tank of compressed hydrogen gas and stripped of their electrons. The remaining protons are
674 then injected into the LINAC2, a linear accelerator responsible for increasing the proton energy
675 to about 50 MeV through RF cavities that push charged particles forward by switching between
676 positive and negative electric fields. The LINAC2 additionally divides the constant stream of

677 particles into equally spaced “bunches” by careful tuning of the frequency of the field switch. The
 678 accelerated protons are then injected into the Proton Synchrotron Booster (PSB), where their
 679 energy is increased by thirty fold, to an energy of roughly 1.4 GeV. The two final acceleration
 680 stages before the protons reach the LHC ring are the Proton Synchrotron and Super Proton
 681 Synchrotron, eventually producing protons with a total energy of 450 GeV. The protons are now
 682 ready for the final stage of their travel and are injected into the two beam pipes of the LHC in
 683 opposite directions. They are injected in trains of 144 bunches each (on order of 10^{11} protons per
 684 bunch), where each bunch is roughly 7.5 meters apart (or 25 ns). There are some larger beam
 685 gaps present in each beam in order to give special magnets sufficient time to switch on in order
 686 to inject or dump the beam. The largest beam abort gap is roughly 3 ms or 900 m long. The
 687 ring is filled with proton bunches until these are equally distributed throughout the two rings, a
 688 process taking roughly 4 minutes. This is called a “fill”. Here, the protons are accelerated to
 689 their current maximum energy of 6.5 TeV, a process taking roughly 20 minutes, through eight RF
 690 cavities. These RF cavities are also responsible for keeping the proton bunches tightly bunched.
 691 A complete sketch of the CERN accelerator complex is shown in Figure 2.1. After the beams
 692 have reached their maximum energy and are stably circulating in the LHC ring, the bunches are
 693 brought to collide. The goal of such a collision, which occurs every 25 nanoseconds, is that some
 694 of the protons in each bunch will inelastically collide, allowing the quark and gluon constituents of
 695 each proton to interact with one another and produce new and interesting particles. The number
 696 of times such an interaction will take place inside a detector per area and time is quantified
 697 through the instantaneous luminosity \mathcal{L} , which is the proportionality factor between the number
 698 of observable events per second and the cross section σ of the process you are interested in,

$$\frac{dN_{events}}{dt} = \mathcal{L}\sigma. \quad (2.1)$$

699 The cross section is the probability that an event (like one which would produce new and interesting
 700 particles) will occur and is measured in barns, where 1 barn = 10^{-28} m². This luminosity should
 701 therefore be as high as possible. It depends only on parameters of the accelerator beams and can,
 702 in the case of the LHC, be defined through accelerator quantities as

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n \beta_*} F, \quad (2.2)$$

703 where N_b is the number of particles per bunch, n_b is the number of bunches, f_{rev} is their revolution
 704 frequency, γ_r is the relativistic gamma factor, ϵ_n is the normalized transverse beam emittance
 705 (how confined the particles are in space and momentum), β_* is the beta function at the collision
 706 point (how narrow, or “squeezed”, the beam is) and F is a reduction factor to account for

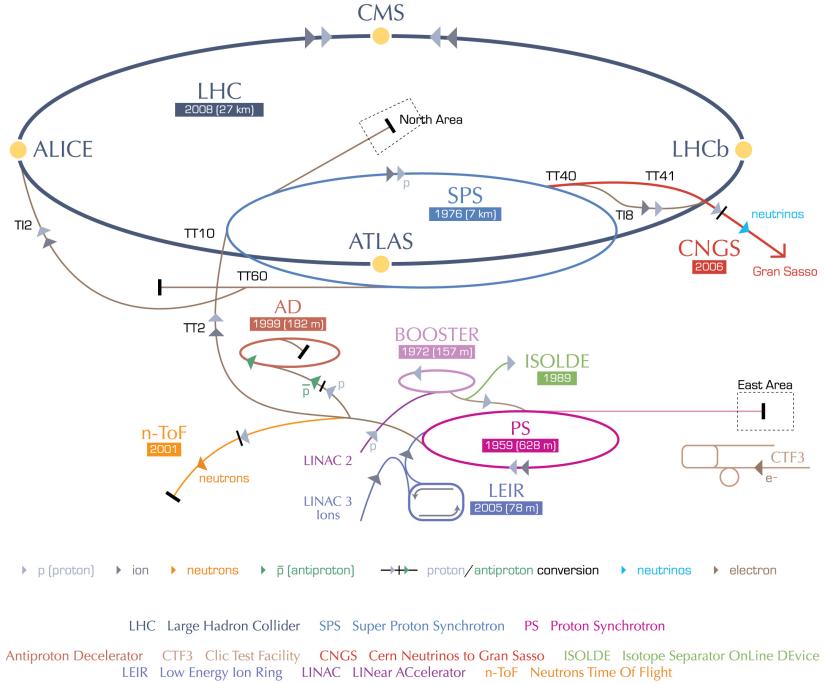


Figure 2.1: The Large Hadron Collider accelerator complex. The four collision points along the ring correspond to the location of the LHC particle detectors CMS, LHCb, ATLAS and ALICE [19].

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), 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

$$\mathcal{L}_{int} = \int \mathcal{L} dt. \quad (2.3)$$

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

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

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

725 Run 2 started in June 2015, and provides the dataset used in thesis. With the accelerator
 726 now running at 90% of its nominal energy, and with a peak luminosity between 1-2 times the
 727 design luminosity, the LHC managed to collect an impressive $\sim 160 \text{ fb}^{-1}$ with center-of-mass
 728 energy of 13 TeV up until its planned shutdown at the end of 2018. Some key LHC accelerator
 729 parameters that were in use for the datasets analyzed in this thesis are quoted in Table 2.1.

730 2.2 The CMS detector

731 The Compact Muon Solenoid (CMS) detector is true to its name. With a diameter of 15 meters and
 732 a weight of 14000 tons, it is 60% smaller but twice as heavy as its counterpart, the ATLAS detector.
 733 Its large weight is due to its solenoid: A superconducting niobium titanium coil circulating 18500
 734 Amps and capable of generating a magnetic field of 3.8 Tesla, making it the worlds largest and
 735 most powerful magnet. Together with its corresponding iron return yoke, responsible for returning
 736 the escaping magnetic flux, it accounts for 90% of the total detector weight. The CMS detector
 737 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 2.2. In the following, the different sub-detectors will be described in detail.

2.2.1 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 2.3.

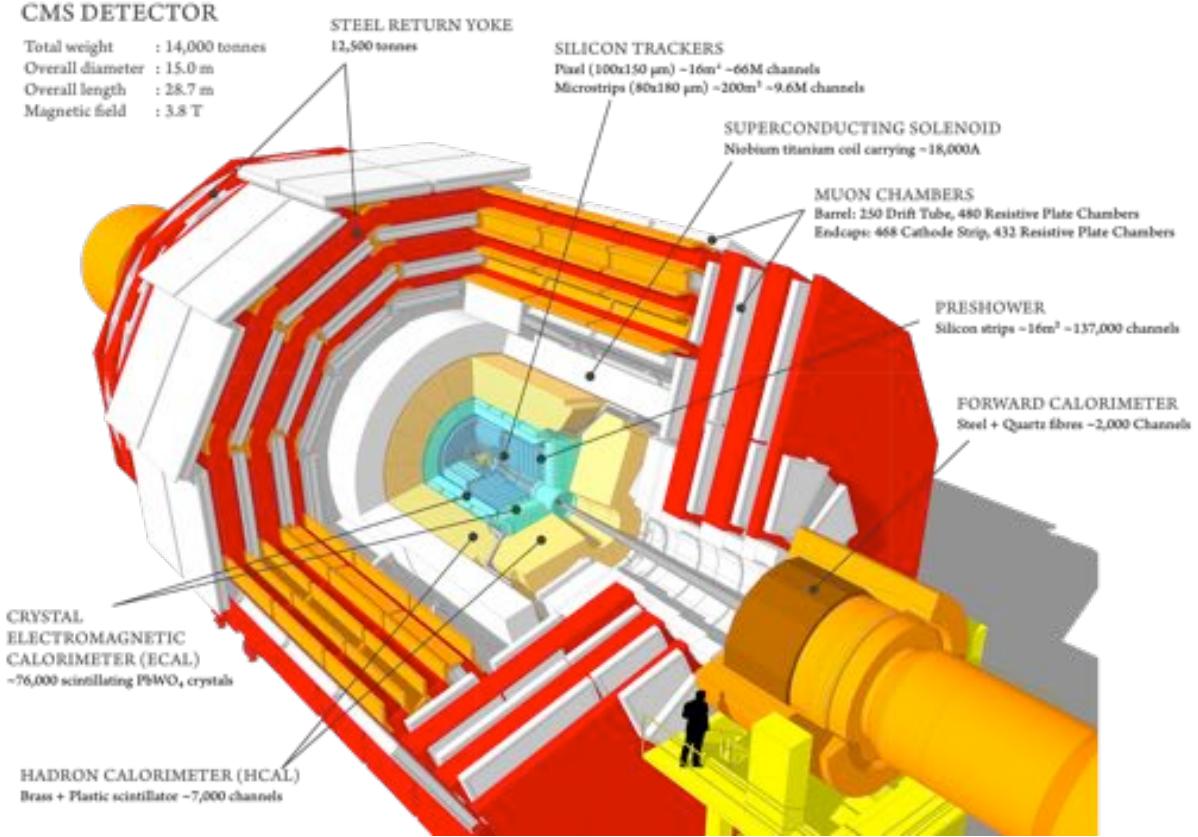


Figure 2.2: The CMS detector and its subsystems: the silicon tracker, electromagnetic and hadron calorimeters, superconducting solenoid and the muon chambers inter-layered with the steel return yoke [20].

2.2.2 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 2.4. The sensors located closest to the beam pipe are subject to hit intensities of $\mathcal{O}(\text{MHz}/\text{mm}^2)$ 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

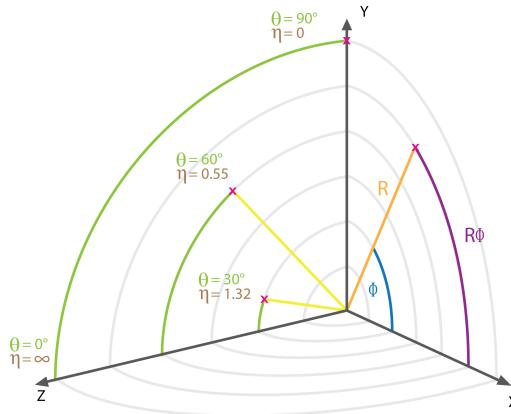


Figure 2.3: The CMS coordinate system [21], shown with some values of θ and η .

of 285 μm , 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/cm², while for the outer layers, readout chips with a rate capability of up to 200 MHz/cm² are sufficient. Since the hit intensity

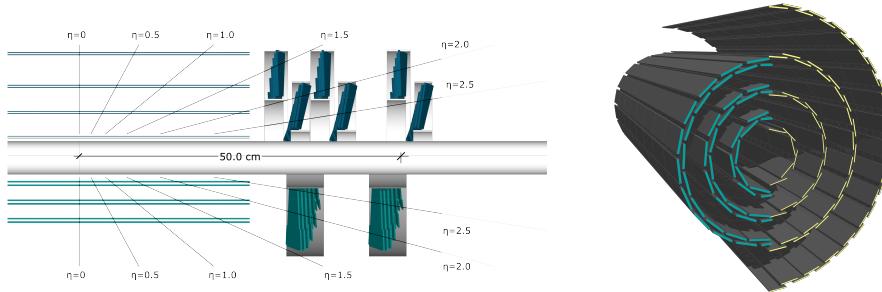


Figure 2.4: Left: The pixel detector layout before (bottom) and after (top) the Phase-1 upgrade. Right: The barrel pixel detector before (left) and after (right) the Phase-1 upgrade [22].

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 2.5. The strips in the TIB and TID are 10 cm long, with a width of 80

791 μm and a thickness of $320 \mu\text{m}$. The TOB and TEC sections consist of slightly larger strips of 25
792 $\text{cm} \times 180 \mu\text{m}$ and a thickness of $500 \mu\text{m}$. The strip tracker has a total of 10 million detector
793 strips and covers an area of $\sim 200 \text{ m}^2$. To prolong the silicon detector lifetime, the entire tracker
794 (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 $\sigma/p_T \approx 1.5 \times 10^{-5} p_T + 0.005$.

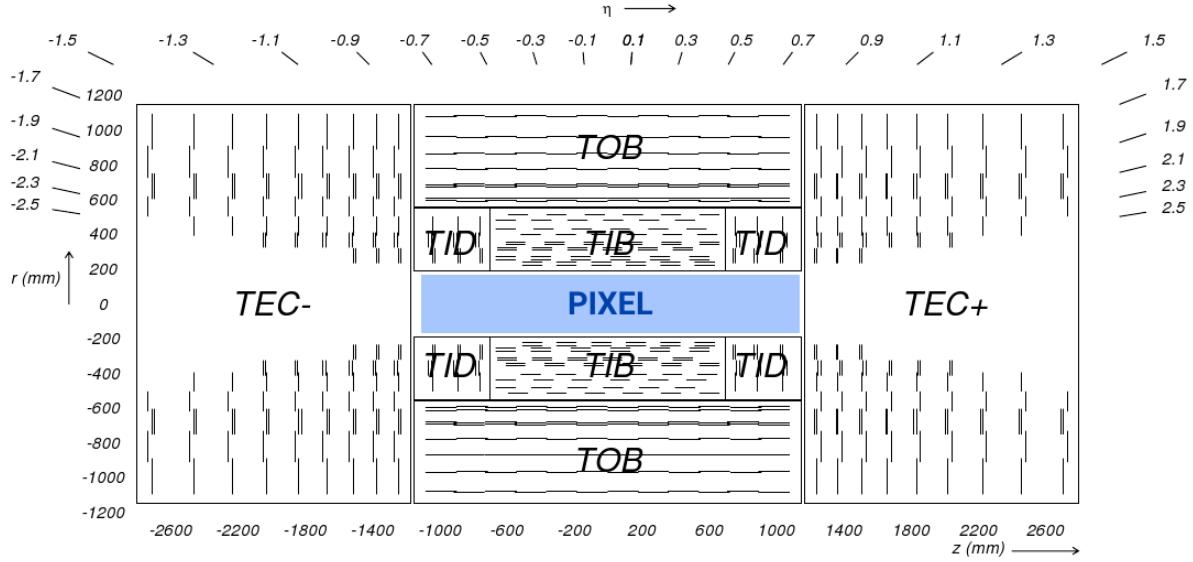


Figure 2.5: Schematic of the CMS silicon strip tracker and its four subsections: the inner barrel (TIB), inner endcaps (TID), the outer barrel (TOB), and the two endcaps (TEC) [23].

795

796 2.2.3 Electromagnetic calorimeter

797 Surrounding the tracking detectors is the electromagnetic crystal calorimeter (ECAL). Consisting
798 of $75\,848$ laterally segmented scintillating lead tungstate (PbWO_4) crystals, it was designed to
799 have the best possible photon energy and position resolution in order to resolve a Higgs boson
800 decaying into two photons, one of the cleanest discovery channels of the Higgs boson. With a
801 design energy resolution of 0.5% above 100 GeV for photons and electrons, the choice of detector
802 material for the ECAL has been its most crucial design feature. In order to withstand the high
803 doses of radiation and the high magnetic field present within the detector, while at the same time
804 generating well-defined signal responses within the 25 nanoseconds between particle collisions,
805 an extremely dense and transparent material capable of producing fast and clean photon bursts
806 when hit, is required. Metal-heavy lead tungstate crystals, each taking roughly two days to
807 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 Moli  re 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 PbWO₄ 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 2.6.

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 [24]. It is given as

$$\frac{\sigma E}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{0.128 \text{ GeV}}{E} \oplus 0.3\%,$$

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 [25], 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 2.7.

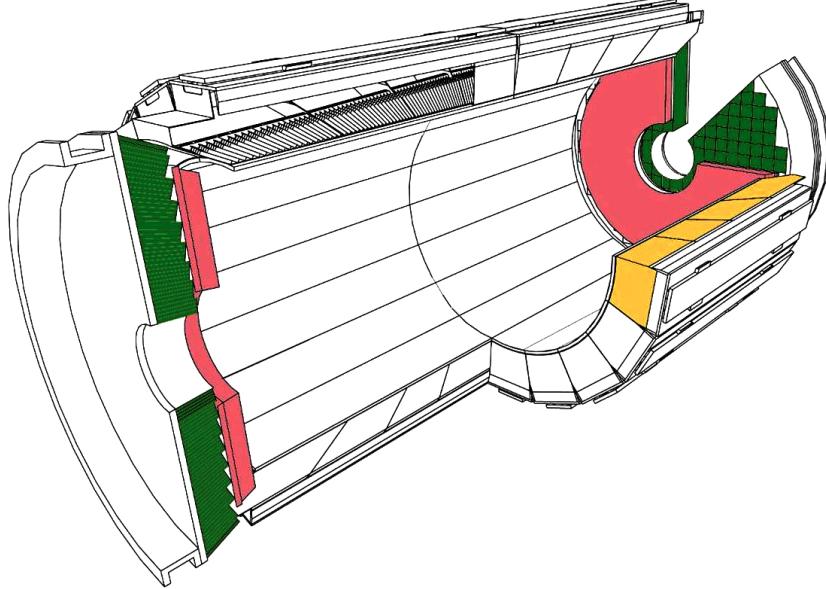


Figure 2.6: A schematic of the CMS electromagnetic calorimeter showing the barrel supermodules (yellow), the individual barrel crystals (black, top left), the endcap modules (green), and the pre-shower detectors (pink) [23].

2.2.4 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 2.8. 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 2.9. The wedges are made up of flat brass absorber plates oriented parallel to the beam axis. These plates consist of a 4-cm thick front steel plate followed by eight 5-cm

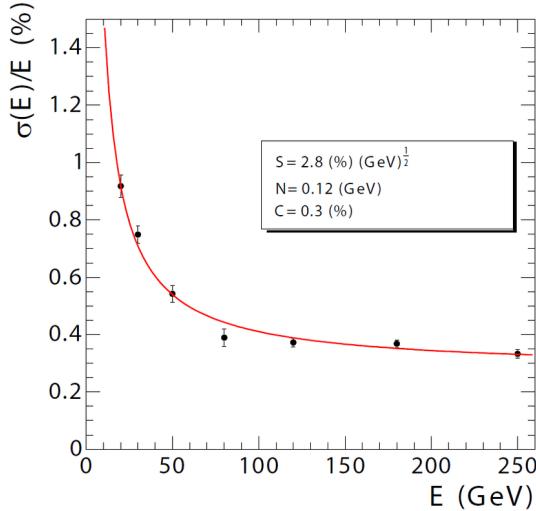


Figure 2.7: The ECAL energy resolution as a function of electron energy as measured in an electron test beam. [24]

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



Figure 2.8: The four regions of the CMS hadron calorimeter: the inner (HB) and outer (HO) barrel, the endcap (HE) and the forward region (HF) [23]

plates in the endcaps are mounted in a staggered geometry rather than on top of each other as is 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 2.10. 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 2.9: The installation of the barrel HCAL wedges, consisting of alternating layers of brass absorber plates and plastic scintillator, each weighing roughly 26 tonnes [26].

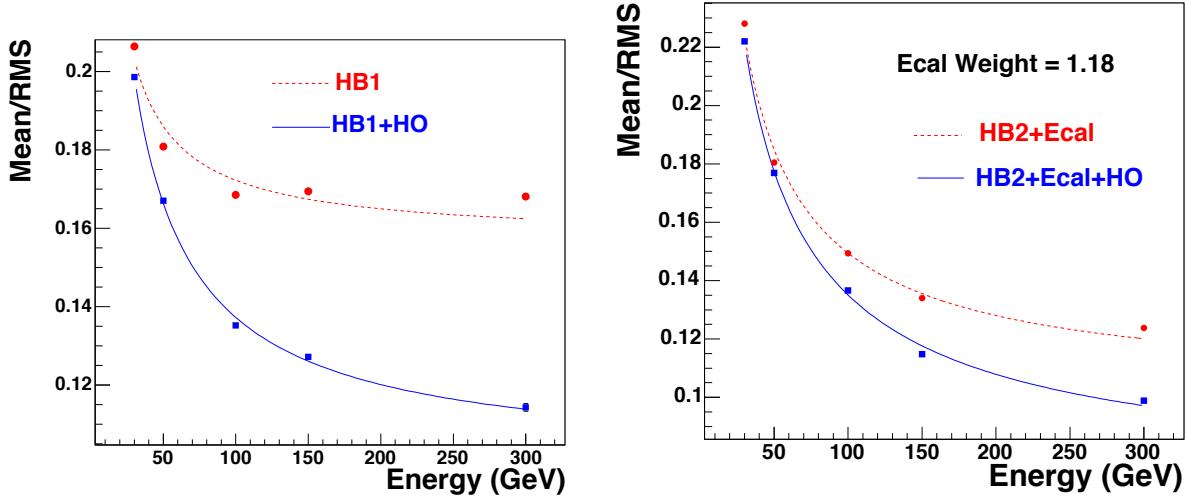


Figure 2.10: The calorimeter energy resolution as a function of pion energy using the HB only or HB+HO (left), and when adding ECAL and HCAL measurements (right) [27].

841 2.2.5 Muon chambers

842 The outer part of the CMS detector is dedicated to performing muon identification, momentum
 843 measurements, and triggering. The muon system is made up of three types of gaseous particle
 844 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 2.11.

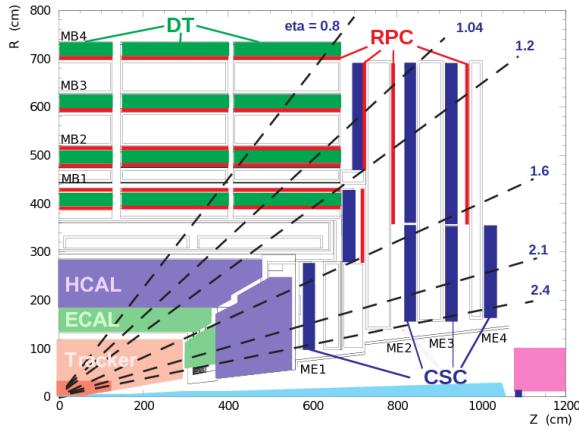


Figure 2.11: A schematic overview of the muon chambers: the DT chambers in the barrel, the CSCs in the endcaps, and the redundant RPC system stretching out to $|\eta| < 1.6$, which are used for triggering purposes [28].

854

855 2.3 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 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.

885

CHAPTER 3

886

Event reconstruction

887 3.1 Track and primary vertex reconstruction

888 The CMS tracker gets traversed by \mathcal{O} 1000 charged particles at each bunch crossing, produced by
 889 an average of roughly 34 proton-proton interactions happening simultaneously. This makes track
 890 reconstructions extremely challenging, and is the reason why a high granularity of the tracker is
 891 vital. The average number of vertices per event for the whole Run 2 is shown in Figure 3.1, with a
 combined average of 34 number of interactions per bunch crossing. Track reconstruction describes

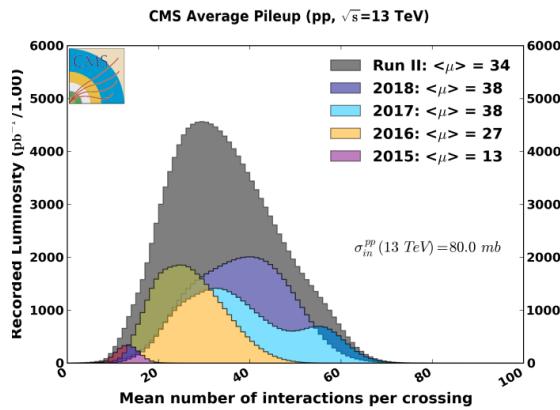


Figure 3.1: The average number of vertices per event in CMS during the Run 2 datataking [29].

892 the process of taking hits from the pixel and strip detectors, combining them and estimating the
 893 momentum and flight direction of the charged particle responsible for producing the hits. It is an
 894 extremely computationally heavy process and is based on what is called a combinatorial Kalman
 895 filter [30]. A Kalman filter is an algorithm that uses time-dependent observations in order to
 896 estimate unknown variables, by proceeding progressively from one measurement to the next,
 897 improving the knowledge of the trajectory with each new measurement. The track reconstruction
 898 software in CMS (called the Combinatorial Track Finder (CTF)) constructs its collection of tracks
 899 by iteratively looping over the hits and reconstructing tracks, then removing those which are
 900 already used as inputs for a previous track. It starts from a seed in the inner most tracker layers,
 901 usually two or three hits, and then extrapolates the seed trajectories searching for additional hits
 902 to associate to that candidate. It then disregards tracks that fail certain criteria based on a χ^2
 903 calculation taking both hit and trajectory uncertainties into account, as well as the number of
 904 missing hits. The track reconstruction algorithm is effective over the full tracker coverage range
 905 up to $|\eta| < 2.5$ and can reconstruct particles with momenta as low as 0.1 GeV or particles which
 906 are produced up to 60 cm from the beam line. In the central region, particles with a momentum
 907 of 100 GeV have a p_T -resolution of roughly 2.8 %, a transverse impact parameter resolution of 10

909 μm and a longitudinal impact parameter of $30 \mu\text{m}$.

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

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

923 3.2 The Particle Flow Algorithm

924 After track reconstruction, what remains is an incoherent collection of tracks, calorimeter clusters
925 and hits in the muon chambers. In order to connect these, CMS uses an algorithm called Particle
926 Flow (PF) [31] to combine the information obtained from all sub-detectors in order to infer which
927 particles were actually produced in the event. The reconstructed physics object in the order of
928 which they are reconstructed are

- 929 • Muons, through hits in the tracker and in the muon chambers
- 930 • Charged hadrons, through hits in the tracker and energy deposits in the calorimeters
- 931 • Neutral hadrons, through energy deposits in the calorimeters but no hits in the tracker
- 932 • Photons, through energy deposits in the ECAL but not in the HCAL and no hits in the
933 tracker
- 934 • Electrons, through hits in the tracker and energy deposits in the ECAL

935 How these different particles propagate through the CMS detector is illustrated in Figure 3.2.

936

937 3.2.1 Reconstruction of the Particle Flow inputs

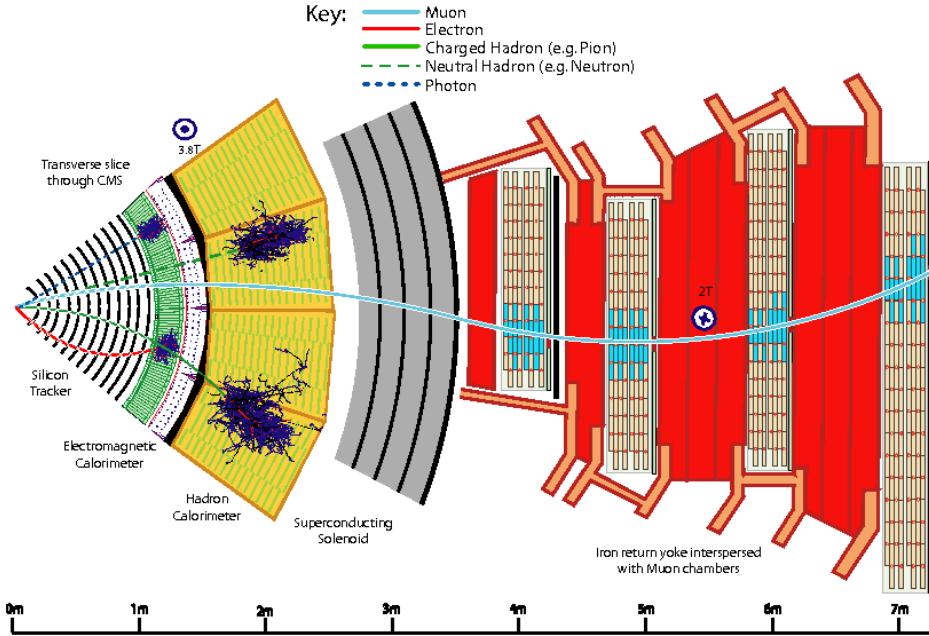


Figure 3.2: Particle interactions in the different subdetectors for a transverse slice through the CMS detector [31].

938 Electron tracking

939 Electron seeding is done in two different ways: ECAL-based or tracker-based electron seeding. In
 940 the ECAL-based method, electrons are seeded from ECAL clusters with $E_T > 4 \text{ GeV}$, where the
 941 position of the cluster is used to infer which hits in the inner tracker belongs to a given electron or
 942 positron. As a large fraction of the electron/positron energy is emitted through bremsstrahlung
 943 before even reaching the ECAL, ECAL superclusters covering a small window in η and a larger
 944 window in ϕ are defined in order to fully contain the electron as well as its bremsstrahlung
 945 photons. As these superclusters are prone to contamination, tight isolation requirements need to
 946 be applied, leading to reconstruction inefficiencies. Therefore, an additional tracker-based seeding
 947 approach has been developed. All tracks with $p_T > 2 \text{ GeV}$ are used as potential electron seeds.
 948 These tracks are then extrapolated to the ECAL and matched to the closest ECAL cluster. The
 949 ratio of the cluster energy to the track momentum is required to be 1. The electron candidates
 950 are then fit with a Gaussian-sum filter (GSF) [32] and required to pass certain criteria based on
 951 the score of a boosted-decision-tree (BDT) which combines the number of tracker hits, the χ^2 of
 952 the GSF track, the energy loss along the track, and the distance between the extrapolated track
 953 to the closest ECAL cluster.

954 Muon tracking

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

- 959 • Standalone muon: Muon tracks based on hits in the DT or CSC only
- 960 • Global muon: A standalone muon track matched to a track in the tracker if the track
961 parameters of the two are compatible
- 962 • Tracker muon: An inner track with $p_T > 0.5$ GeV, a total momentum greater than 2.5 GeV
963 and at least one muon segment matching the extrapolated inner track

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

967 Calorimeter clusters

968 The calorimeter clustering is performed separately for each calorimeter subdetector (ECAL barrel
969 and endcaps, HCAL barrel and endcaps and the preshower layers). The first step is to define
970 cluster seeds from cells with an energy exceeding some predefined threshold and in addition is
971 larger than the energy in its neighboring cells. Topological clusters are then formed by adding
972 cells to the seed which has at least one corner in common with a cell already in the cluster, and
973 that has an energy which is at least twice the noise level of the detector. In Figure 3.3, an example
974 of calorimeter clustering for a five-particle jet is shown for the HCAL (left) and ECAL (right). In
975 the HCAL (left), two seeds have been identified (gray filled areas) inside a topological cluster
976 consisting of 9 cells. These are then defined as two HCAL clusters, with a position as indicated
977 by the red circles. The green solid lines correspond to charged tracks reconstructed in the tracker,
978 both pointing to the center of the HCAL cluster seeds. The observed deposits left by the same
979 particles are shown on the right in Figure 3.3, where the K_L^0 , π^- and the two photons from the
980 decay of a π^0 leave distinct clusters in the ECAL. The π^+ leaves no energy deposit in this case.

981 3.2.2 Particle Flow identification

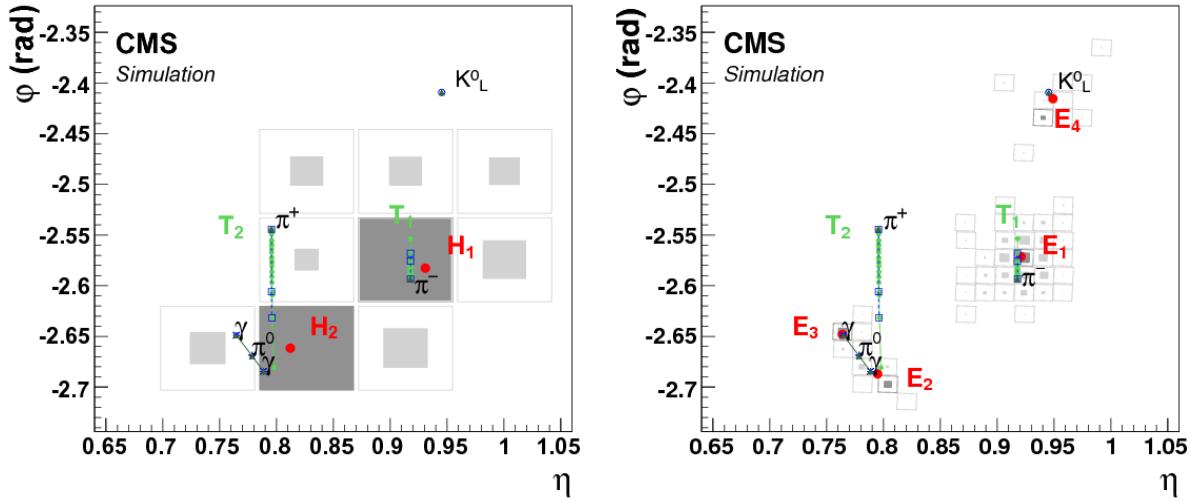


Figure 3.3: The $\eta - \phi$ vies of calorimeter clusters generated by a five-particle jet in the HCAL (left) and in the ECAL (right). The squares correspond to calorimeter cells, where the inner area is proportional to the logarithm of the cell energy. Cluster seeds are depicted in dark gray. The dotted blue lines correspond to the simulated particle trajectory, while the green lines correspond to charged tracks reconstructed in the tracker [31].

982 The link algorithm

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

- 988 • **Inner track - calorimeter cluster link:** The track is interpolated from its last hit,
 989 through the preshower layers, the ECAL and ending in the HCAL at an interaction length
 990 depth of 1. A link is made if the track is within the cluster area, where the areas is enlarged
 991 by up to a cell in each direction to account for detector gaps. In case several ECAL/HCAL
 992 clusters are linked to the same track, only the one with the smallest distance in $\eta - \phi$ is
 993 kept.
- 994 • **Calorimeter cluster - cluster link:** A link between ECAL and HCAL clusters as well
 995 as between ECAL and preshower clusters is made when the cluster position of the more
 996 granular calorimeter is within the cluster envelope in the less granular calorimeter. Also
 997 here, if there is link overlap, only the link with the smallest distance is kept

- 998 • **Inner tracker -muon chamber link:** As described in Section 3.2.1

999 For each PF block, the reconstruction proceeds in the following order. First, muons are
1000 reconstructed and their corresponding PF elements removed from the PF block. Then the
1001 electrons are reconstructed, with the hopes of removing their corresponding bremsstrahlung
1002 photons from the list of PF elements. Energetic photons are reconstructed in the same step.
1003 Finally, neutral and charged hadrons are reconstructed.

1004 **Muons**

1005 First, isolated global muons are selected by requiring the sum of track p_T and calorimeter energy
1006 deposits within a cone of $\Delta R = 0.3$ not belonging to the muon track, to be smaller than 10
1007 % of the muon p_T . If the muons are non-isolated, they are required to pass the tight muon
1008 requirement [33] and have at least three matching track segments in the muon detector or have
1009 matched calorimeter deposits compatible with being a minimum ionizing particle. Muons failing
1010 both the requirements above are kept if the standalone muon track is of high quality and have a
1011 lot of hits in the muon detectors, otherwise they are discarded. The muon momentum is defined
1012 from the inner tracker measurement if the muon p_T is less than 200 GeV. Otherwise, its chosen
1013 according to the track fit with the smallest χ^2 probability.

1014
1015 Muons used in this thesis are required to pass an isolation requirement in order to suppress
1016 the background from QCD multijet events where jet constituents are identified as muons. For this,
1017 a cone of radius $\Delta R = 0.3$ is constructed around the muon direction. The isolation parameter is
1018 defined as the scalar sum of the transverse momenta of all additional reconstructed tracks within
1019 the cone, divided by the muon p_T . Muon candidates with an isolation parameter less than 0.1
1020 are considered isolated and are used for further analysis.

1021 Further, the following selection criteria are applied:

- 1022 • The χ^2 of the global muon track fit must be less than 10
1023 • At least one muon-chamber hit is included in the global-muon track fit. The global muon
1024 track fit must include at least one muon chamber hit
1025 • Muon segments in at least two of the muon stations must be matched to the muon tracker
1026 track
1027 • The inner tracker track must be no more than 2 millimeters from the primary vertex in the
1028 xy plane and no more than 5 millimeters in the longitudinal direction
1029 • At least one hit in the pixel detector.

- 1030 • At least six layers of the inner tracker must contain hits.
 1031 • At least three matching track segments must be found in the muon detectors

1032 **Electrons**

1033 The electrons are seeded from a GSF track, as described in Section 3.2.1. To differentiate electrons
 1034 from charged hadrons, the energy deposit in the HCAL within a distance of 0.15 in the $\eta - \phi$ plane
 1035 of the supercluster , is required to be less than 10 % of that of the supercluster. The electron
 1036 candidate must further pass a requirement on the output of a dedicated electron-identification
 1037 BDT, using inputs such as track-cluster distance, track χ^2 and number of hits as input. In this
 1038 step, isolated photons are also reconstructed, seeded from ECAL superclusters with $|E_T > 10 \text{ GeV}$
 1039 and no link to a GSF track. All the tracks and calorimeter deposits used to reconstruct electrons
 1040 and isolated photons are further removed from the list of PF blocks.
 1041 Only electrons passing certain quality requirements, corresponding to the CMS electron HEEP
 1042 ID, are used in this thesis. These requirements are listed in Table 3.1, with the following variable
 1043 definitions:

- 1044 • E_T : The supercluster energy $x \sin(\theta_{track})$ where θ_{track} is the electron track polar angle
 1045 measured in the inner tracker layer and extrapolated to the interaction vertex.
- 1046 • η^{sc} : η of the electron supercluster.
- 1047 • **isEcalDriven**: Electron is found through ECAL requirements rather than through Particle
 1049 Flow and the tracker.
- 1050 • $\Delta\eta_{in}^{seed}$: η difference between the track position measured in the inner layer, extrapolated
 1052 to the interaction vertex and to the calorimeter, and the η of the supercluster.
- 1053 • $\Delta\phi_{in}$: ϕ difference between the track position measured in the inner layer, extrapolated to
 1055 the interaction vertex and to the calorimeter, and the ϕ of the supercluster.
- 1056 • **H/E**: Ratio of hadronic energy in the calorimeter towers within a cone of radius 0.15 cen-
 1058 tered at the electrons calorimeter position, to the electromagnetic energy of the supercluster.
- 1059 • $\sigma_{inj\eta}$: Measure of the energy spread in η in units of crystals of electron energy in a 5×5
 1061 block centered on the seed crystal.

- **ECAL Isolation:** The transverse electromagnetic energy of all reconstructed hits(with $E > 0.08 \text{ GeV}$) in a cone of radius 0.3 centered at the electron 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.

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 $\sigma_{inj\eta}$	n/a	< 0.03
full 5x5 E^{2x5}/E^{5x5}	> 0.94 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 3.1: Summary of the electron requirements allied to all electrons used in this analysis.

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 deposits as 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

1084 compared. If the calorimeter energy is compatible with the sum of track momenta, a particle for
 1085 each track is inferred, with its corresponding energy taken from the track momentum. If the
 1086 calorimeter energy is larger than the sum of track momenta, a photon or a neutral hadron is
 1087 added, together with one charged hadron for each track within the cluster area.

1088 Missing transverse energy

1089 Neutrinos (and other predicted non-SM weakly interacting particles) do not interact in the
 1090 detector and are instead inferred from the presence of a momentum imbalance in the detectors
 1091 transverse plane. The missing transverse momentum is defined as the negative p_T vector sum of
 1092 all reconstructed PF candidates in the event

$$\vec{p}_T^{\text{miss}} = - \sum_i^N \vec{p}_{T,i}. \quad (3.1)$$

1093 and its magnitude, $|\vec{p}_T^{\text{miss}}|$, is denoted the missing transverse energy E_T^{miss} (which is used as a
 1094 proxy for the neutrino p_T).

1095 3.3 Pile-up removal

1096 Particles originating from proton-proton interactions not associated with the hardest primary
 1097 vertex, are denoted pileup events. These distorts observables of interest from the hard scattering
 1098 event and must be mitigated through dedicated pileup removal techniques

1099 3.3.1 Charged Hadron Subtraction

1100 As mentioned previously, primary vertices are reconstructed using tracks from charged hadrons.
 1101 If a primary vertex does not correspond to the hard scattering vertex of the event, the charged
 1102 hadrons (as reconstructed through Particle Flow) associated to this vertex (called pileup vertex)
 1103 are removed from the event collection of particles and will not participate in any further object
 1104 reconstruction. This method is denoted charged hadron subtraction (CHS).

1105 3.3.2 Pile up per particle identification (PUPPI)

1106 CHS was the default pileup removal algorithm in CMS until very recently. In 2014, a new pileup
 1107 removal algorithm with improved performance was proposed; the pileup per particle identification
 1108 (PUPPI) [34] algorithm. PUPPI uses a combination of local shape information, event pileup

properties and tracking information to compute a weight describing the degree of “pileup-likeness” of a given particle. 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

$$\alpha_i = \log \sum_{\substack{j \in \text{Ch,PV} \\ j \neq i}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij}), \quad (3.2)$$

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

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{RMS_{PU}^2}, \quad (3.3)$$

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 3.4.

The top row shows the absolute mass resolution (left) as well as the mass resolution as a function 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 compared to CHS.

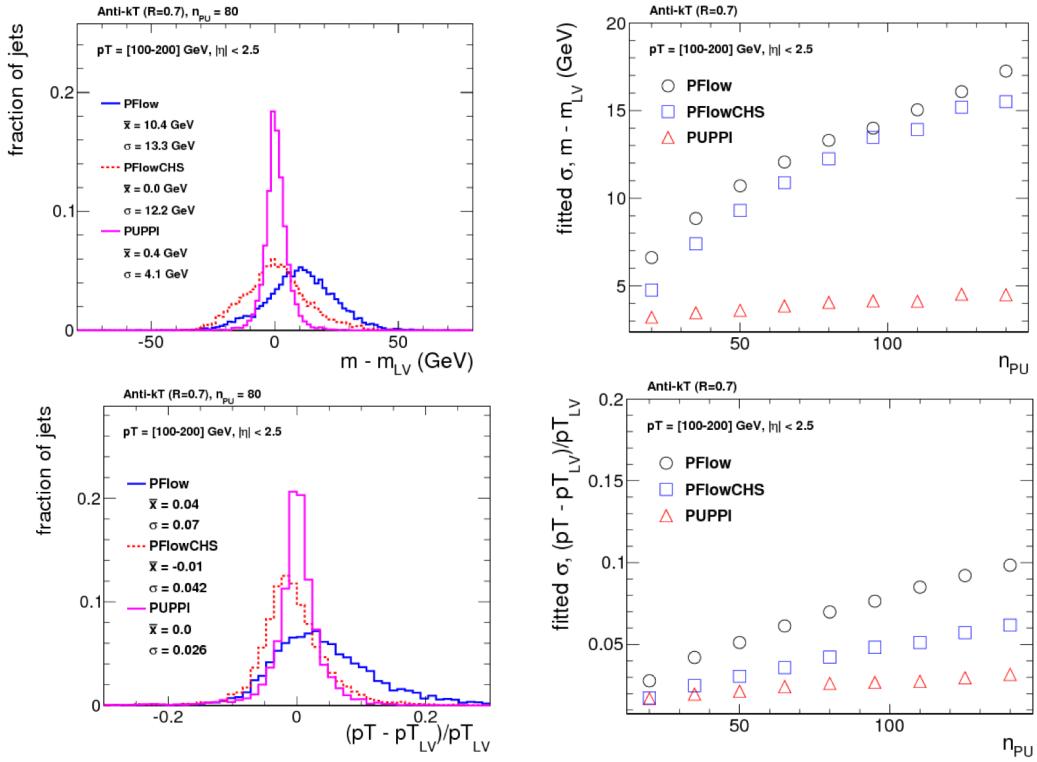


Figure 3.4: The mass (top) and p_T (bottom) resolution comparing PF only (blue), PF+CHS (red) and PUPPI (pink) jets. The absolute resolution (left) as well as the resolution as a function of the number of reconstructed primary vertices in the event (right) is shown [34].

1135 3.4 Jet reconstruction

As explained in Section 1.1.2, quarks and gluons are never themselves visible in a detector. Within 10^{-23} seconds, the timescale of the 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 designed to do, called 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.

¹¹⁴⁶ 3.4.1 Jet clustering

¹¹⁴⁷ The most common jet clustering algorithms used in hadron colliders are the Cambridge/Aachen
¹¹⁴⁸ algorithm [35], the k_T algorithm [36] and the anti- k_T algorithm [37]. These are all sequential
¹¹⁴⁹ recombination algorithms, meaning they systematically go through each particle pair in the event
¹¹⁵⁰ and recombines them into one particle if the combination satisfies certain criteria. The rules,
¹¹⁵¹ shared by all three algorithms, are as follows:

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

$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \text{ with } \Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \quad (3.4)$$

$$d_{iB} = p_{ti}^{2p}, \quad (3.5)$$

¹¹⁵² where d_{ij} is a measure of the relative transverse momenta between the particles, ΔR_{ij}^2 is
¹¹⁵³ the distance between them in the $\eta - \phi$ plane (which can be roughly translated into a
¹¹⁵⁴ jet radius), ΔR^2 corresponds to a distance parameter which controls the extension of the
¹¹⁵⁵ jet and d_{iB} is the distance between the particle and the beam. The parameter p is what
¹¹⁵⁶ separates the three algorithms from one another and controls the relative power of energy
¹¹⁵⁷ versus geometrical scales. For the anti- k_T algorithm, it is defined as $p = -1$, for the k_T
¹¹⁵⁸ algorithm $p = 1$ and in the case of the C/A algorithm, $p = 0$. The consequences of these
¹¹⁵⁹ 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 is defined to be a final state jet, and is removed from the list of
¹¹⁶² particles. The algorithm proceeds back to step 1.
¹¹⁶³
5. Repeat until no particles remain.

¹¹⁶⁵ Infrared and collinear safety

¹¹⁶⁶ There are two requirements that are extremely important when defining jet algorithms: They
¹¹⁶⁷ must be 1) *infrared* (IR) and 2) *collinear* (C) safe. *Infrared* safety corresponds to the requirement
¹¹⁶⁸ that if the final state particles are modified by the presence of a soft emission, and there are
¹¹⁶⁹ always soft emission in QCD events (both perturbative and non-perturbative), then the set of
¹¹⁷⁰ hard jets should remain unchanged. This is illustrated by the two left figures in Figure 3.5.

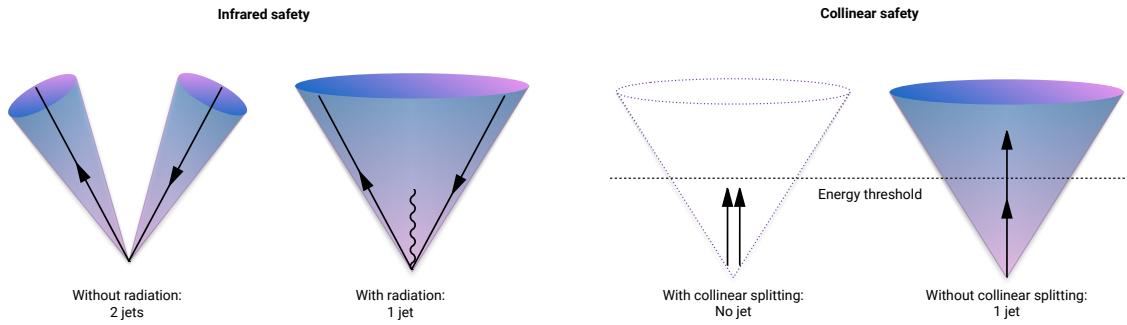


Figure 3.5: An illustration of what would happen for an infrared (left) and collinear (right) unsafe jet algorithm. If an algorithm is infrared unsafe, the presence of a soft emission changes the jet configuration. If an algorithm is collinear unsafe, then if a parton undergoes a collinear splitting this will change the configuration of the jet

1171 Here, the algorithm is infrared unsafe: the presence of an additional soft gluon changes the
 1172 jet configuration from 2 to 1 jets. If an algorithm is *collinear* unsafe, it means that the jet
 1173 configuration would change if the hard parton undergoes collinear splitting (which a hard parton
 1174 often does as part of the fragmentation process and which are also part of non-perturbative
 1175 dynamics, like the decay of highly energetic hadrons). This is shown in the two left figures of
 1176 Figure 3.5, where a hard parton undergoing collinear splitting fails to be reconstructed due to its
 1177 daughters being below the energy threshold of the algorithm.

1178 All sequential recombination algorithms are trivially infrared safe.

1179 The k_T algorithm

1180 The k_T algorithm is the oldest of the sequential recombination algorithms and, due to its $p = 1$
 1181 definition in the distance measures, follows the QCD branching structure in both p_T and in angle
 1182 (in reverse). Soft particles are clustered together first, and the final step is the clustering of
 1183 the two hardest particles. A consequence of this definition is that there is nothing that keeps
 1184 arbitrarily soft particles from being defined as jets, and a minimum cut on the jet p_T should be
 1185 introduced. Despite several favorable qualities, the k_T algorithm is not the algorithm of choice in
 1186 most hadron collider experiments due to the irregular jets it produces, a consequence of clustering
 1187 soft particles first.

1188 The Cambridge/Aachen algorithm

1189 The Cambridge/Aachen algorithm, with $p = 0$ in the distance measures, follows the QCD
 1190 branching structure only in angle as the clustering order is based solely on spatial separation.

1191 The simplest of the algorithms, it recombines all pairs close in ΔR until $\Delta R_{ij} > R$. The benefits
1192 of this is that the clustering history contains information about the presence of any geometrical
1193 substructure within a jet, a feature that will become important in Section 3.5.

1194 **The anti- k_T algorithm**

1195 The default jet clustering algorithm in CMS is the anti- k_T algorithm [37], which follows the
1196 rules and distance measures above with $p = -1$. The algorithm favors the clustering of high- p_T -
1197 high- p_T and high- p_T – low- p_T particles first, disfavoring clustering between soft particles. That
1198 means the algorithm grows around a hard core, yielding jets with a well-defined cone shaped area.
1199 Together with being IRC-safe and insensitive to the underlying event (any event not arising the
1200 primary hard scattering process) and pileup, makes it the main jet algorithm in CMS.

1201 A comparison of the resulting jet area in the $\phi - \eta$ plane after clustering with either k_T , C/A
1202 and anti- k_T , is shown in Figure 3.6. The z-axis correspond to the parton p_T . One can clearly see
1203 that when clustering with the anti- k_T algorithm, the produced jets are circular, with a radius set
1204 by R , around the hardest parton.

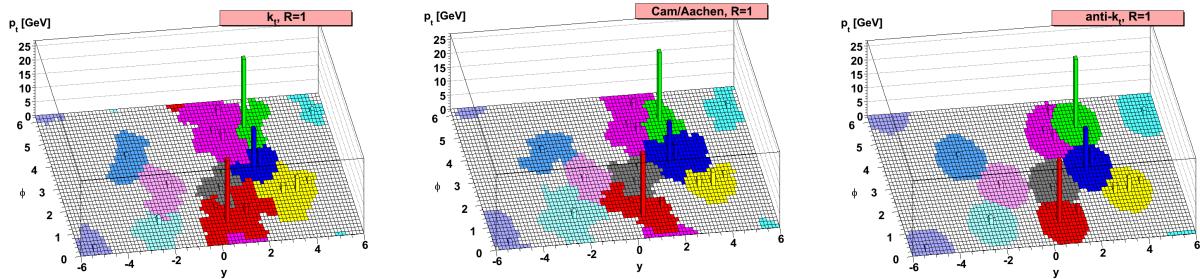


Figure 3.6: A comparison of the resulting jet cone area in the $\phi - \eta - p_T$ plane after clustering the same event with three different jet algorithms: k_T , C/A and anti- k_T . [37]

1205 **3.4.2 PF jets in CMS**

1206 Jet algorithms in CMS mainly use PF candidate four-vectors as input and a pileup removal
1207 algorithm is usually applied before clustering occurs. If using CHS (Section 3.3.1), charged
1208 hadrons not associated to the primary vertex are discarded before clustering. If PUPPI is
1209 used (Section 3.3.2), all the PF candidates are reweighted based on how likely they are to have
1210 originated from pileup. For the anti- k_T algorithm, CMS by default uses two jet cone sizes: $R=0.4$
1211 and $R=0.8$. Jets with $R=0.4$, called PFAK4, are used for single-prong jets while the larger $R=0.8$

1212 jets, PFAK8, are more often used when looking for jets containing multiple hard quarks/gluons
1213 in order to contain all the hadronization products.

1214 These jets are further required to pass certain jet identification requirements provided by the
1215 JetMET POG [38], in order to distinguish them from fake jets. All jets used in this analysis are
1216 required to pass the *tight ID* requirements which are as follows:

- 1217 • The jet must contain at least two PF constituents
- 1218 • At least one of these constituents must be a charged hadron
- 1219 • The fraction of jet energy coming from neutral hadrons must be < 0.90
- 1220 • The fraction of jet energy coming from neutral electromagnetic energy must be < 0.90
- 1221 • The fraction of jet energy coming from charged electromagnetic energy must be < 0.99

1222 3.4.3 Jet energy corrections

1223 All jets are further corrected for nonlinearities in p_T and rapidity using standard CMS jet energy
1224 corrections (JEC), as described in Ref. [39]. These are intended to bring the measured jet energy
1225 closer to the true jet energy by correcting the jet energy scale (JES) and jet energy resolution
1226 (JER). The energy corrections are derived in three steps:

- 1227 • L1: Energy offset corrections intended to remove pileup and electronic noise, both for data
1228 and simulation
- 1229 • L2L3: A relative (L2) and absolute (L3) correction to particle level jet response for
1230 simulation only
- 1231 • Residual: A correction for data only meant to correct for residual differences between data
1232 and simulation

1233 These are illustrated in Figure 3.7.

1234 L1 offset correction

1235 The largest correction is the L1 pileup offset correction, which are meant to subtract the additional
1236 energy in a jet due to pileup. This is done on an event-by-event basis through the *jet area method*
1237 which uses the jet effective area multiplied by the average event energy density to calculate the
1238 size of the offset energy to be subtracted from each jet. An additional p_T - and η -dependent term

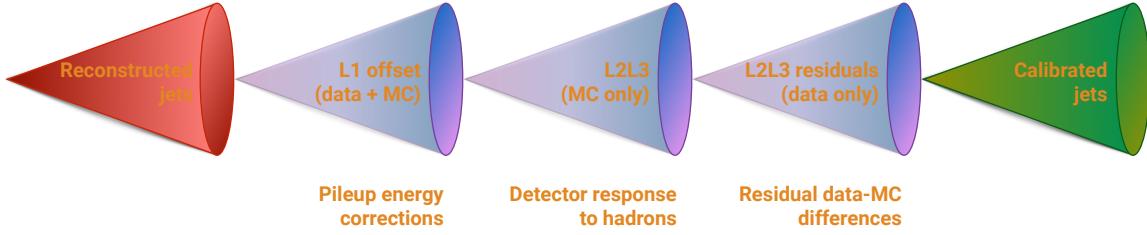


Figure 3.7: The CMS jet energy corrections are derived in three steps: A correction due to offset energy coming from pileup, applied to data and MC, a correction due to the particle level jet response, also applied to data and MC and finally a correction to account for residual differences between data and MC.

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 scalefactor 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

1244 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

$$R_{\text{particle}} = \frac{p_{\text{T, reco}}}{p_{\text{T, particle}}} \quad (3.6)$$

These are derived in bins of particle level p_{T} and reconstructed η : The L2 relative corrections are intended to uniform the detector response 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.

1252 Residual data corrections

After L1 and L2L3 corrections are applied, two additional corrections are derived for data only 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 dijet sample, 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 $W/\gamma + \text{jet}$ 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.

1255

1256

1257 The above description of jet energy corrections in CMS is meant as a rough, instructive
1258 summary only. A full description of the measurement techniques used in CMS, can be found
1259 in [39].

1260 3.5 Jet substructure reconstruction

1261 In analyses looking for highly energetic (“boosted”) vector bosons like, a main theme of this
1262 thesis, the opening angle between the vector boson quark decay products become so small that
1263 the highly boosted boson appears as a single large jet instead of two well-separated smaller jets.
1264 The distance between the two quarks, in the case of an hadronic decay, depends on the mass of
1265 the vector boson and its p_T and goes as

$$\Delta R = \frac{2M_V}{p_{T,V}}. \quad (3.7)$$

1266 Above a W boson p_T of 200 GeV, the two quarks are therefore merged into a single large cone jet
1267 of size $R = 0.8$. A sketch of the two different situation is shown in Figure 3.8. If the W p_T is well
1268 below 200 GeV, its decay products are well-defined jets in their own right (left). However, once
1269 the W transverse momenta starts exceeding 200 GeV, both the quarks are completely contained
1270 within a single jet (right). In order to distinguish hadronically decaying vector boson from QCD
1271 quark/gluon jets, the jet mass would in principle be a good discriminant as we know the W has a

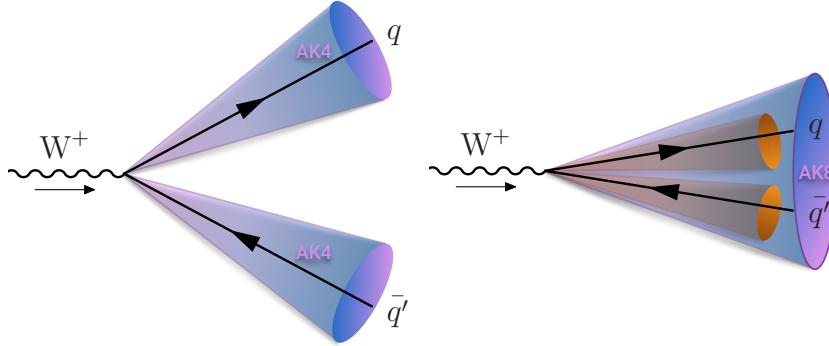


Figure 3.8: If the mass of the resonance is low enough, the quark decay products of each vector boson are well separated and clustered into distinguishable AK4 jets (left). If the transverse momentum of the vector boson is greater than 200 GeV, the vector boson decay products are merged into one single large cone AK8 jet.

mass of around 80 GeV while the quark/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 get rid of UE and pileup, algorithms like PUPPI and CHS can be used. Then, to improve the mass resolution further, dedicated grooming algorithms must be applied.

3.5.1 Grooming

Grooming was introduced as a tool to improve the signal, most often $W/Z/\gamma$, mass resolution 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 which then are removed.

Trimming

The trimming algorithm [40] is a grooming algorithm mostly used at trigger level in CMS (also where it is used in this thesis) due to it being less aggressive than other grooming algorithms. 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 jets and QCD quark/gluon jets for different values of r_{sub} and $p_{T,frac}$ is shown in Figure 3.9. 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 and background discrimination 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 $p_{T,frac} = 0.03$).

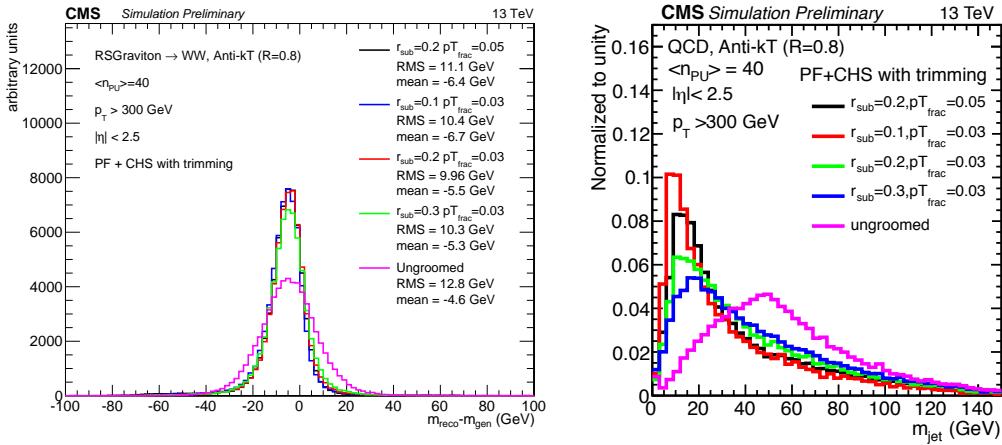


Figure 3.9: The effect of trimming on a signal jet (left) and a background jet (right) for different values of the tuned parameters r_{sub} and $p_{T,frac}$ [41].

1292

1293 Pruning

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

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

- The first requirement is a requirement on the hardness of the combination. $p_{T,i}$ 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

wider than D_{cut} relative to the axis of the recombination of the two protojets. In CMS, the tuned parameters are set to $r_{cut} = 0.5$ and where $z_{cut} = 0.1$. Figure 3.10 shows the ungroomed as well as the pruned jet mass distribution for signal (left) and background jets. The highest amount of signal and background separation in CMS, is achieved with $r_{cut} = 0.5$ and $z_{cut} = 0.1$.

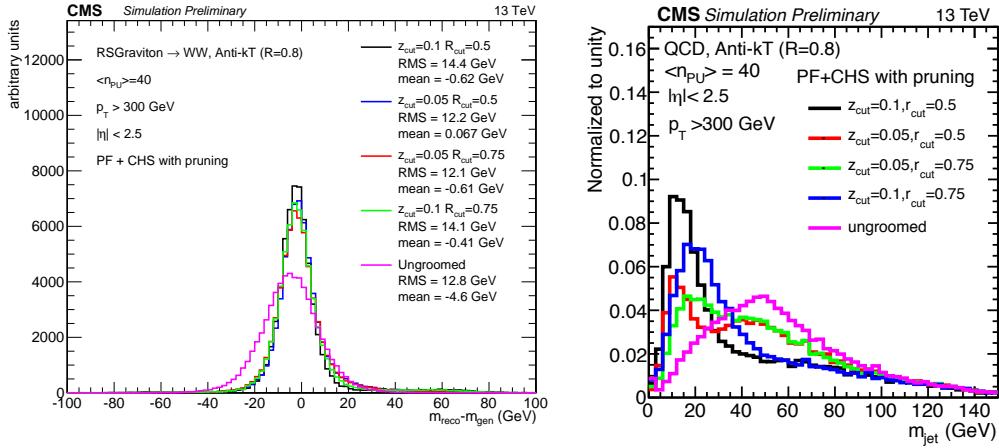


Figure 3.10: The effect of pruning on a signal jet (left) and a background jet (right) for different values of the tuned parameters z_{cut} and r_{cut} [41].

1301

1302 Modified Mass Drop Tagger and Soft Drop

The modified mass drop tagger (mMDT) [42] (a modified version of the originally suggested mass drop tagger [43]) is based on the idea that a W/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 mMDT tagger therefore starts from an already clustered jet, reclusters it with the C/A algorithm and then declusters it again defining subjets s_1 and s_2 . It then looks for a significant mass drop going from total jet mass to the mass of each subjet, and checks that the splitting is not too asymmetric. The modified mass drop condition is generalized through the soft drop declustering method [44], simply called Soft Drop, which allows for different types of angular requirements to enter the condition. The Soft Drop condition is the following

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

If no significant mass drop occurred and the splitting is not to 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 from the jet. 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 3.11. 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.

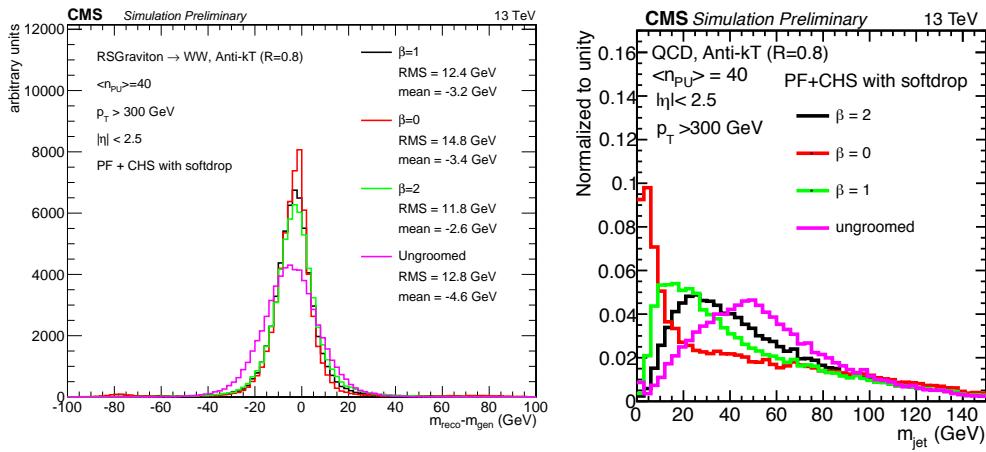


Figure 3.11: The effect of softdrop on a signal jet (left) and a background jet (right) for different values of the tuned parameters β . $\beta = 0$ corresponds to the Modified Mass Drop Tagger, which is the default Softdrop setting in CMS [41].

3.5.2 N-subjettiness

After hopefully having resolved the particle mass with one of the grooming algorithms above, there is still discriminating information to be gathered from the jet structure itself. A W or Z jet consists of two well-defined high- p_T subjets. A quark/gluon jet on the other hand, made from a single parton, consists of several large angle, asymmetric splittings, as illustrated in Figure 3.12. The N-subjettiness algorithm [45] takes advantage of this fact by attempting to count the number of hard subelements within a jet. This is quantified through the n-subjettiness variable, τ_N ,

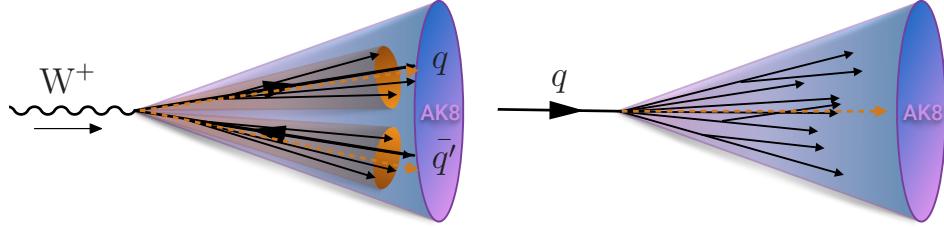


Figure 3.12: A jet stemming from the decay of a W will usually have two well-separated high- p_T subjets, while a jet with a single-prong origin consists of several large angle splittings.

1321 defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}) \quad (3.8)$$

1322 where k runs over all the jet constituents, $p_{T,k}$ is the constituent transverse momentum and
1323 $\Delta R_{i,k}$ is the distance between the constituent and candidate subjet axes. These subjet axes are
1324 obtained through a one-pass optimization procedure which minimizes N-subjettiness [46]. The
1325 normalization factor in front is given as

$$d_0 = \sum_k p_{T,k} R_0 \quad (3.9)$$

1326 where R_0 corresponds to the cone size of the initial jet. With this definition, jets with $\tau_N = 0$
1327 have most of its constituents aligned along the subjet axes. However, if $\tau_N \gg 0$, a large fraction
1328 of the energy is radiated away from the subjet directions and are more likely to have more
1329 than N subjets. In CMS, and as recommended by the authors in [45], the ratio τ_2/τ_1 is used to
1330 discriminate W jets from QCD jets. The reason for this is that, while signal jets are expected
1331 to have a large τ_1 , quark/gluon can similarly have large τ_1 due to the diffuse radiation present.
1332 However, QCD jets with a large τ_1 tend to have an equally large τ_2 , while signal jets do not, hence
1333 the ratio of the two provides greater separation power. In CMS, the n-subjettiness algorithm is
1334 by default applied to ungroomed jets. The distribution of τ_{21} for signal and background jets with
1335 different pileup subtraction algorithms applied are shown in Figure 3.13, where τ_{21} in combination
1336 with PF+PUPPI (green) yields a distribution most similar to the generated one (black).

1337 3.5.3 Vector boson tagging

1338 In order to discriminate W and Z bosons from quark/gluon jets a combination of a groomer and
1339 shape-tagger (like n-subjettiness) is usually used. Typical values for a W -tagger are jet groomed

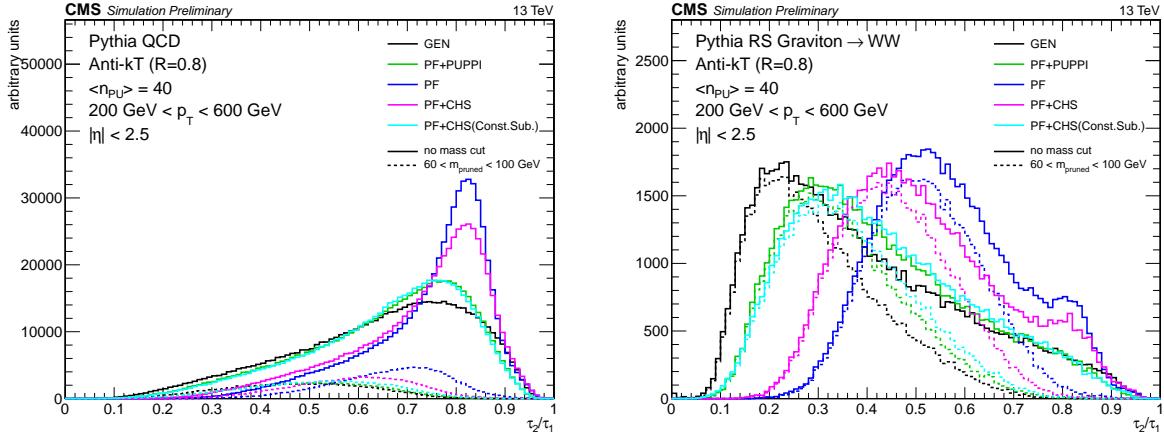


Figure 3.13: The distribution of the n-subjettiness ratio τ_{21} for signal jets (left) and background jets (right) with different combinations of pileup subtraction algorithms applied. The solid lines corresponds to the τ_{21} distribution with no mass cut applied, while the dotted lines are within a mass window of 60–100 GeV [41].

1340 mass between 60 and 100 GeV and $\tau_{21} < 0.5$. Which combination and which cuts to use is
 1341 analysis dependent, and have been optimized for each search presented in this thesis. The details
 1342 are thoroughly explained in each section.

1343 Polarization effects

1344 The vector boson polarization has a significant effect on the W-tagging efficiency. The helicity
 1345 angle θ , defined as the angle between the outgoing quark daughters of the W in the W bosons rest
 1346 frame relative to its direction of motion [47], is very different for longitudinally polarized vector
 1347 bosons, W_L , and transversely polarized vector bosons W_T [48]. Figure 3.14 shows the $\cos \theta^*$
 1348 distribution for the outgoing quarks from a $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decay, and it
 1349 can be observed that transversely polarized W bosons decay with the quarks emitted closer to the
 1350 vector boson direction of motion. The consequence of this, is that there is a higher asymmetry
 1351 in the transverse momenta of the two quarks from a W_T decay. This in turn makes grooming
 1352 algorithms, designed to remove soft constituents of a jet, tend to reject particles coming from the
 1353 softer quark, resulting in a lower jet mass and a drop in tagging efficiency. Figure 3.15 shows the
 1354 W-jet tagging efficiency versus q/g jet mistagging rate for a selection on the jet pruned mass of
 1355 $60 \text{ GeV} < m_{\text{pruned}} < 100 \text{ GeV}$, scanning τ_{21} cuts (here for CA R=0.8 jets).

1356 The tagging efficiency for transversely polarized W bosons (green) is significantly lower than
 1357 the tagging efficiency for longitudinally polarized bosons (red). This can be explained by looking

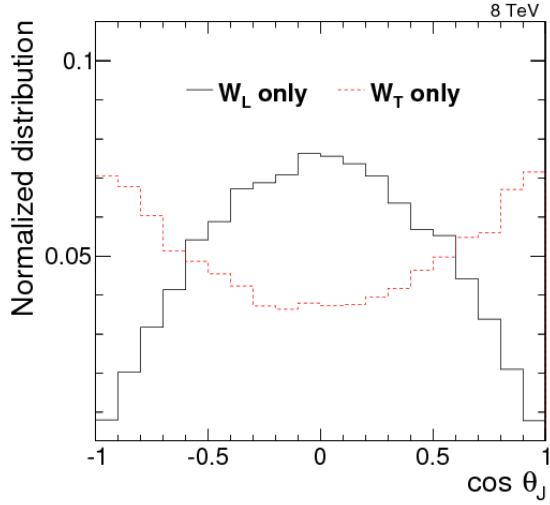


Figure 3.14: The helicity angle for generated quarks from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decays [48].

at the $\cos \theta^*$ distribution on reconstructed level, using the C/A subjets, with a cut on the jet pruned mass of $60 \text{ GeV} < m_{\text{pruned}} < 100 \text{ GeV}$, shown in Figure 3.16. When comparing to the distribution at generator level with no groomed mass window applied, Figure 3.14, one can see that the W_T jets with $\cos \theta^* \approx 1$ 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 $R=0.8$.

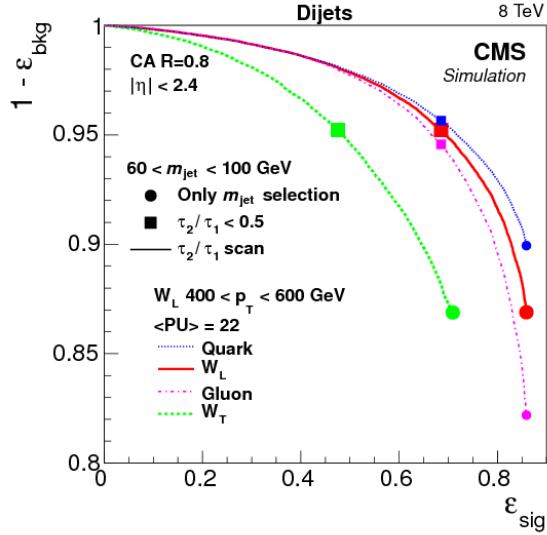


Figure 3.15: The helicity angle for generated quarks from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decays [48].

1365 3.6 Monte Carlo Event Generators

1366 Monte Carlo event generators offer a realistic estimate of high-energy collisions on an event-
 1367 by-event basis, allowing us to estimate signal and background processes accurately. Simulated
 1368 events are usually produced in three steps, describing a process from very short timescales up
 1369 until hadronization and decay. First, a matrix element generator simulates the hard scattering
 1370 process and subsequent decays. Secondly, the showering and hadronization of unstable particles
 1371 is performed and, lastly, the final state particles are passed through a full detector simulation in
 1372 order to reproduce a range of experimental effects.

1373 General-purpose Monte Carlo (GPMC) generators, like HERWIG ++ [49] and PYTHIA 8 [50],
 1374 deal with both perturbative as well as hadronization phenomena, simulating an event all the
 1375 way up until detector simulation. In HERWIG ++ and PYTHIA 8, the hardest processes are only
 1376 simulated at the lowest order of perturbative expansion, meaning $2 \rightarrow 2$ or $2 \rightarrow 3$ scatterings. In
 1377 order to have tree-level matrix elements with an arbitrary final-state multiplicity, they can be
 1378 combined with programs used to generate parton-level events at higher accuracy, which are then
 1379 processed through showering and hadronization with the GPMC generators. One popular program
 1380 for generating matrix elements is MADGRAPH [51]. This, however, still correspond to a tree-level
 1381 (leading order) approach. To go to next-to-leading-order (NLO), meaning the inclusion of virtual
 1382 corrections, two methods exist: MC@NLO [52, 53] and POWHEG [54]. These combine the full

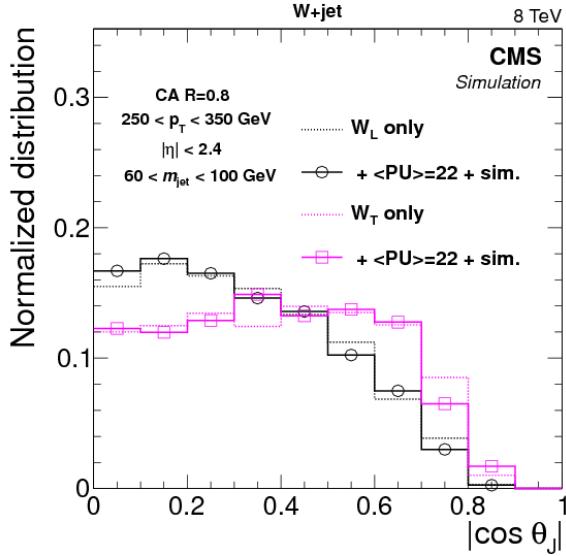


Figure 3.16: The helicity angle for subjets from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (pink) decays. [48].

1383 next-to-leading-order prediction for inclusive processes with the subsequent parton showers, either
 1384 by a subtraction method regularizing the real contributions, or by a matrix-element correction
 1385 of the parton shower branching probability. After hadronization, all final state particles are
 1386 passed through a full simulation of the CMS detector. This is done with GEANT4 [55], where
 1387 experimental effects like object reconstruction and detector resolution are accounted for.

1388 For the work presented in this thesis, simulated samples of the Standard Model background
 1389 processes are used to optimize the analysis and in some cases provide flexible background templates.
 1390 QCD multijet production is simulated with four generator configurations: 1. PYTHIA standalone,
 1391 2. the LO mode of MADGRAPH matched with PYTHIA, 3. POWHEG matched with PYTHIA and 4.
 1392 HERWIG++ 2.7.1 with tune CUETHS1 [56]. Top quark pair production is modeled with POWHEG
 1393 and showered with PYTHIA unless otherwise stated. $W+jets$ and $Z+jets$ production are simulated
 1394 with the leading-order (LO) mode of MADGRAPH matched with PYTHIA. Signal samples are
 1395 generated with standalone PYTHIA.

1396 All samples are processed through a GEANT4-based simulation of the CMS detector. To
 1397 simulate the effect of additional proton-proton collisions within the same or adjacent bunch
 1398 crossings (pileup), additional inelastic events are generated using PYTHIA and superimposed
 1399 on the hard-scattering events. The MC simulated events are finally weighted to reproduce the
 1400 distribution of the number of pileup interactions observed in data.

1401

CHAPTER 4

1402

Diboson resonance searches in CMS

¹⁴⁰³ 4.1 Search I: First search for diboson resonances at 13 TeV

¹⁴⁰⁴ When the LHC started its Run II data taking period in summer 2015, it would be the first time
¹⁴⁰⁵ ever for a particle collider to produce collisions with center-of-mass energies of 13 TeV. The Higgs
¹⁴⁰⁶ boson for which the LHC was designed to find had been discovered at the end of the previous data
¹⁴⁰⁷ taking era, leaving us with a Standard Model that we know is, in the best case, in need of extensions
¹⁴⁰⁸ and, in the worst case, an effective theory valid only in a certain energy domain. The Run II
¹⁴⁰⁹ search program would therefore be oriented around two main efforts: Precision measurements of
¹⁴¹⁰ the newly discovered Higgs boson and searches for Beyond Standard Model physics.

¹⁴¹¹

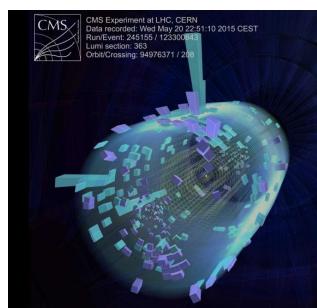
¹⁴¹² I started my PhD four months before the first 13 TeV collisions took place and had to consider the following: What was the most interesting search that could be done on a short time scale
¹⁴¹³ (to be presented 6 months after first collisions, at the CERN end-of-year “Jamboree”), would be
¹⁴¹⁴ manageable for a student with no previous analysis experience and would be robust enough incase
¹⁴¹⁵ there were issues with the never-before-validated 13 TeV Monte Carlo?

¹⁴¹⁷

¹⁴¹⁸ The attention of the high-energy physics community has in the past years been focused on certain
¹⁴¹⁹ “hot topics”: In 2018, this was most certainly leptoquarks (driven by a dimuon excess around 30
¹⁴²⁰ GeV), in 2016 and 2017 is was diphoton resonances (with $> 3\sigma$ excesses observed both in ATLAS
¹⁴²¹ and in CMS). And in 2015 during the 13 TeV LHC start-up, attention was centered on diboson
¹⁴²² resonances in the all-hadronic final state. The choice was therefore clear: My first analysis would
¹⁴²³ be a search for diboson resonances in the boosted dijet final state. With a background model
¹⁴²⁴ based on a smooth fit to data in the signal region, eliminating the need for accurate QCD MC
¹⁴²⁵ predictions, this was a simple one-background only (QCD) analysis, feasible for a first-year PhD
¹⁴²⁶ student to finalize within a year. Despite its straightforwardness, due to observed 8 TeV excesses,
¹⁴²⁷ it was in addition considered a high-profile analysis.

¹⁴²⁸

¹⁴²⁹ This search became one of the first “boosted” searches published with 13 TeV data as well
¹⁴³⁰ as the first search to take advantage of dedicated “grooming” triggers. Published with 2.7 fb^{-1} of
¹⁴³¹ 2015 data.



¹⁴³² 4.1.1 A small bump

¹⁴³³ On June 2nd 2015, the day before CMS recorded its first ever 13 TeV event, a pre-print appeared
¹⁴³⁴ on the arXiv “Search for high-mass diboson resonances with boson-tagged jets in proton-proton
¹⁴³⁵ collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector” [51]. It was an analysis of the full ATLAS
¹⁴³⁶ Run 1 dataset, corresponding to 20.3 fb^{-1} , searching for heavy resonances decaying to vector
¹⁴³⁷ bosons in the all-hadronic state. The analysis documented a 3.4σ excess for a heavy resonance
¹⁴³⁸ decaying to W Z around 2 TeV. The corresponding CMS analysis, published the previous year,
¹⁴³⁹ had a 1.3σ excess at roughly the same resonance mass, but mostly compatible with a W W
¹⁴⁴⁰ final state hypothesis [57]. Figure 4.1 shows the corresponding dijet invariant mass spectrum as
¹⁴⁴¹ seen by ATLAS (left) and the upper limit on the production times the cross section for a G_{Bulk}
¹⁴⁴² decaying to W W (right) as documented by CMS.

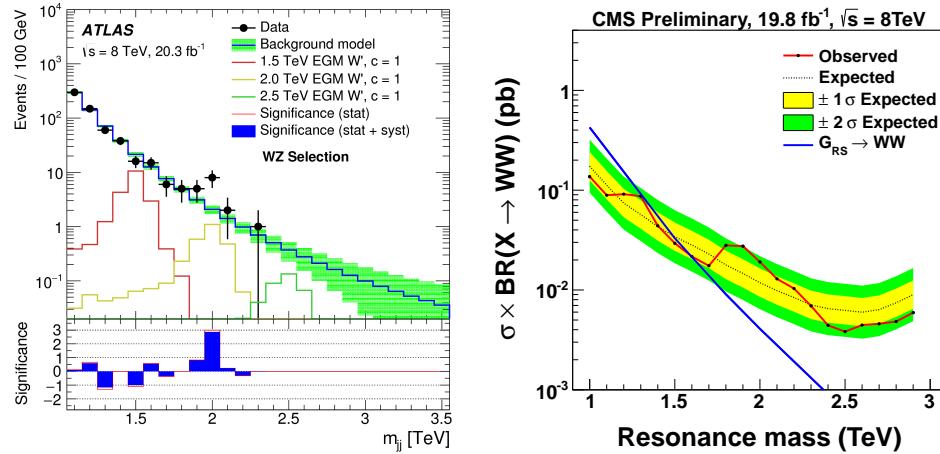


Figure 4.1: A “bump” corresponding to 3.4σ in the dijet invariant mass spectrum around 2 TeV (left) observed by ATLAS when analyzing the full 8 TeV dataset [51], together with a similar excess (1.3σ) observed in the corresponding CMS analysis [57].

¹⁴⁴³ The two measurements were found to be compatible, favoring a heavy resonance with a
¹⁴⁴⁴ production cross section of around 5 fb^{-1} and a mass between 1.9 and 2.0 TeV decaying to either
¹⁴⁴⁵ W W, W Z or Z Z [58]. Figure 4.2 show the obtained p-value of the ATLAS (red) and CMS
¹⁴⁴⁶ (blue) search as well as their combination (black).

¹⁴⁴⁷ The combination of the two excesses and the timing of the ATLAS paper, naturally lead to
¹⁴⁴⁸ some excitement and in the coming weeks, the arXiv was flooded with theory papers seeking an
¹⁴⁴⁹ explanation for the measurements.

¹⁴⁵⁰ In addition, one of the main benefits of increasing the LHC center-of-mass energy from 8 to 13
¹⁴⁵¹ TeV, was that the partonic luminosity would increase. That meant that you could expect them

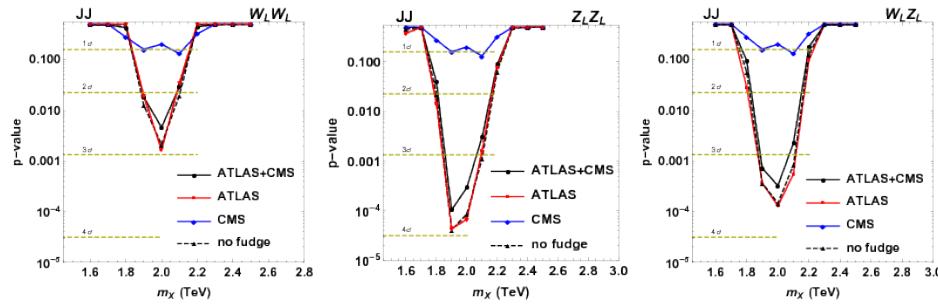


Figure 4.2: p-values as a function of resonance mass obtained with an emulation of the ATLAS (red) and CMS (blue) searches as well as the combination of the two (black). Here for a $W W$ (left), $W Z$ (middle) and $Z Z$ (right) hypothesis [58].

same number of signal events as you would expect for the full 8 TeV dataset ($\sim 20 \text{ fb}^{-1}$), for a considerably smaller 13 TeV dataset. Figure 4.3 shows the system mass that can be probed with 3 fb^{-1} of 13 TeV data (y-axis), the expected 2015 integrated luminosity, as a function of the probe-able system mass with 20 fb^{-1} of 8 TeV data (x-axis) for different partonic channels. The probable 13 TeV mass is defined by finding the system mass which results in the same number of expected events at 8 TeV, if assuming cross sections scale with partonic luminosity and $1/m^2$. Three different partonic scattering channels are considered: qq, qg and gg. We see that, for instance for a resonance with a mass of 2000 GeV, the reach at 13 TeV is 2241(gg), 2091(qg), 1851(qq, one type) and 2046(qq, all types) GeV.

What this meant was that, if we saw hints of a 2 TeV G_{bulk} (mainly produced through gluon fusion, gg) with the 8 TeV dataset, we should be able to confirm it with nothing but the expected 3 fb^{-1} of data expected in 2015. The pressure on seeing early results with 13 TeV data in the VV all-hadronic final state was therefore extremely high, and it was agreed with CMS Physics Coordination that a preliminary analysis would be ready in December that same year, only 6 months after the first 13 TeV collision.

4.1.2 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). The boson is co-called “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 very collimated, as described in Section 3.5. This results in a final state with two large high- p_T jets, where an AK R=0.8 jet is expected to fully contain the two quarks coming from the vector boson decay. This is illustrated in Figure 4.4.

The two jets are both expected to have a mass around the W or Z mass, and some intrinsic

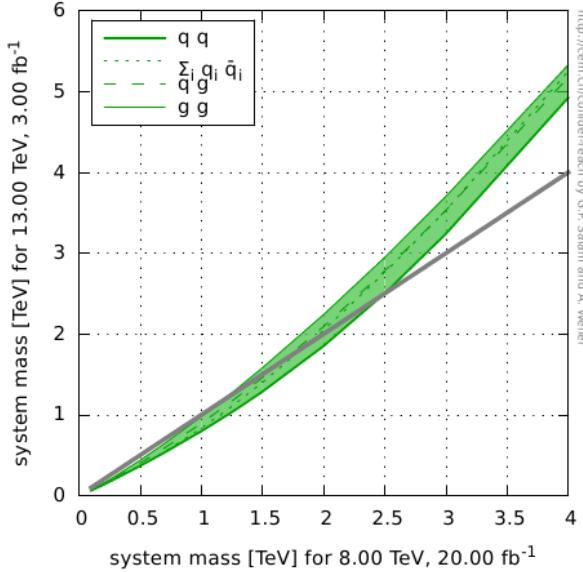


Figure 4.3: The system mass that can be probed with 3 fb^{-1} of 13 TeV data (y-axis) as a function of the probe-able system mass with 20 fb^{-1} of 8 TeV data (x-axis) for different partonic channels (Generated with [59]).

1475 substructure stemming from their two-prong origin. The invariant mass of the dijet system, m_{jj} ,
 1476 should be roughly equal to the resonance mass M_X . This dijet system is the final state under
 1477 scrutiny and the dijet invariant mass is the parameter of interest. Both WW and ZZ, as well as
 1478 WZ final states are of interest.

1479 The main background for such an analysis, is QCD multijet events. As mentioned in Section 3.5,
 1480 quark/gluon jets can obtain a high mass due to diffuse radiation and QCD processes have such a
 1481 large cross section that the number of QCD jets with a mass compatible with the W mass can be
 1482 large. In order to discriminate between the two, we take advantage of three properties: 1. The
 1483 groomed mass of signal and background jets should be very different, 2. signal jets should appear
 1484 two-prong like, quark/gluon jets not, and 3. the dijet invariant mass for a signal process should
 1485 peak around the resonance mass while the QCD spectrum is predicted to be smoothly falling
 1486 (we will get back to why this assumption is justified in Section 4.1.5). The strategy therefore
 1487 consists of performing a smoothness test on m_{jj} of the observed data, a so-called “bump-hunt”,
 1488 by assuming that the signal will appear as a bump on top of a smooth distribution. This is
 1489 illustrated in Figure 4.5.

1490 The benefit of such a method is that there is no need for any background simulation and
 1491 the strategy is simple and robust. The disadvantage is that the analysis is intrinsically limited

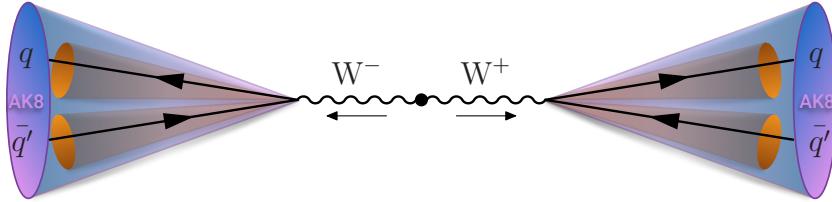


Figure 4.4: If a heavy ($> 1 \text{ TeV}$) resonance decays into vector bosons, the transverse momentum of each boson will be large and its decay products are merged into one single large cone AK8 jet.

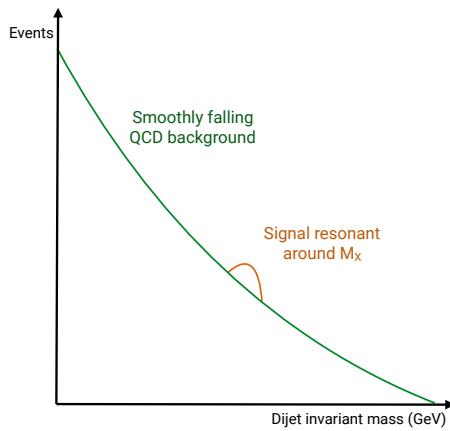


Figure 4.5: The search strategy consists of looking for signal “bumps” in the dijet invariant mass on top of a smoothly falling QCD multijet background.

1492 to regions where the dijet invariant mass spectrum is smooth, hence must avoid regions with
 1493 continuities due to trigger turn-ons or kinematic selections.

1494 4.1.3 Data and simulated samples

1495 The data analyzed in this search correspond to a total integrated luminosity of 2.7 fb^{-1} collected at
 1496 a center-of mass energy of 13 TeV between June and December 2015. The instantaneous luminosity
 1497 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
 1498 an average number of primary vertices per event of $\langle \mu \rangle = 13$.

1499 The bulk graviton model (see Section 1.2.1) and the HVT model (W' and Z' , see Section 1.2.1)
 1500 are used as benchmark signal processes. In these models, the vector gauge bosons are produced
 1501 with a longitudinal polarization in more than 99% of the cases, which leads to a 24% higher
 1502 acceptance per boson for reasons explained in Section 3.5.3. For the HVT model, a scenario
 1503 (model B) with $g_V = 3$, $c_H = -0.976243$, and $c_F = 1.02433$ is chosen, where the heavy resonance

1504 predominantly couple to bosons and the coupling to fermions is suppressed. The bulk graviton
1505 samples were generated with $\tilde{k} = 0.5$. The resonance masses considered lie in the range 1.2 to
1506 4 TeV and has a width of 0.1% of the resonance mass. The narrow width allows us to neglect
1507 detector effect as the natural width of the resonance is smaller than the detector resolution,
1508 making the modeling of detector effects on the signal shape independent of the model. All signal
1509 samples are generated at leading order with MADGRAPH5_AMC@NLO v2.2.2 [60]

1510 Simulated samples of the production of QCD multijet events are generated to leading order
1511 using PYTHIA version 8.205 [61] with the CUETP8M1 tune [56].

1512 4.1.4 Event selection

1513 Triggering

1514 The first selection to be confronted in any analysis, is the trigger selection. Due to an overwhelming
1515 QCD background in all-hadronic final states, the threshold for fully-hadronic triggers is very
1516 large in order to keep the trigger rate low (preferably around 10-30 Hertz). In this analysis, we
1517 therefore decided to take advantage of triggers that place requirements on the jet groomed mass
1518 in addition to the “standard” triggers based on the scalar sum of jet transverse energy H_T . These
1519 “boosted” triggers were never before tested in data, and this analysis was the first published
1520 result taking advantage of grooming at the trigger level in CMS. The following H_T -based triggers
1521 (called inclusive triggers in the following) are used

- 1522 • HLT_PFHT650_WideJetMJJ900DEtaJJ1p5
- 1523 • HLT_PFHT650_WideJetMJJ950DEtaJJ1p5,
- 1524 • HLT_PFHT800

1525 where the two first triggers apply an additional cut on the $|\Delta\eta|$ between the two jets for reasons
1526 that will be explained below. In addition, two grooming triggers cutting on the jet trimmed mass
1527 (see Section 3.5.1) of 30 and 50 GeV are used

- 1528 • HLT_AK8PFJet360_TrimMass30
- 1529 • HLT_AK8PFHT700_TrimR0p1PT0p03Mass50.

1530 The tuneable parameters for the trimming algorithm at HLT are $r_{sub} = 0.2$ and $p_{T,frac} = 0.03$.
1531 The HLT_AK8PFJet360_TrimMass30 trigger is seeded by single-object Level 1 triggers with jet
1532 p_T thresholds of 176 or 200 GeV (L1_SingleJet176 or L1_SingleJet200), and the remaining
1533 triggers requires an online $H_T > 150$ or 175 GeV (L1_HTT150 or L1_HTT175).

1534 In order to avoid any kinks in the dijet invariant mass spectrum due to the presence of a
 1535 trigger turn-on, we need to define for which dijet invariant mass the analysis triggers are fully
 1536 efficient ($> 99\%$), then cut away everything below that point.

In order to estimate the trigger efficiency, we use a lower threshold prescaled H_T trigger HLT_PFT650 as reference trigger. This trigger has a prescale of 40, meaning it only stores information for every 40 events that trigger it, and is seeded by L1 triggers L1_HTT150 or L1_HTT175. We then define the efficiency as

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

1537 where $N_{trigger+ref}$ corresponds to events passing the trigger under study as well as the reference
 1538 trigger and N_{ref} corresponds to all events passing the reference trigger. Figure 4.6 shows the
 1539 trigger turn-on curves as a function of dijet invariant mass for jets where one of the jets is required
 1540 to have a pruned mass larger than 65 GeV (in other words, compatible with a W jet). A sharp
 1541 turn-on for the inclusive triggers (top left) is observed, reaching the 100% efficiency plateau for
 1542 dijet masses of around 1.0–1.1 TeV. The grooming triggers, however, turn on more slowly and
 1543 are not fully efficient before dijet invariant masses of around 1.2 TeV (top right). The real power
 1544 of the grooming triggers become clear when adding them in OR with the H_T -based triggers. The
 1545 bottom plot in Figure 4.6 compares the trigger turn-on curves as a function of dijet invariant
 1546 mass for jets passing one of the three inclusive triggers only, one of the grooming triggers only
 1547 and when combining all of them. Here, one can see that the 99% efficiency threshold is lowered
 1548 by 75 GeV when including the substructure triggers, once substructure is required at analysis
 1549 level. This allowed for the analysis to start at a dijet invariant mass of 1 TeV.

1550 As a measure of the performance of the grooming triggers, we have in addition looked at
 1551 the trigger efficiencies as a function of the offline groomed mass (pruned and softdrop, see
 1552 Sections 3.5.1 and 3.5.1), for the grooming trigger with the lowest mass threshold (30 GeV).
 1553 This is shown in Figure 4.7, where an additional cut on the jet transverse momentum of one of
 1554 the jets of 600 GeV is required and no other mass cut is applied. The trigger plateau is reached
 1555 for offline groomed-jet masses around 50 GeV, an impressively sharp turn-on for a trigger paths
 1556 first test i data (as reference trigger for this study, the prescaled trigger HLT_PFT650 was used).

1557 Preselection

1558 After trigger selections and the corresponding requirement of a dijet invariant mass above 1 TeV
 1559 to ensure a smooth falling background, the process of maximizing the signal significance while
 1560 keeping the background low can begin. This is done through a set of jet requirements. The jets

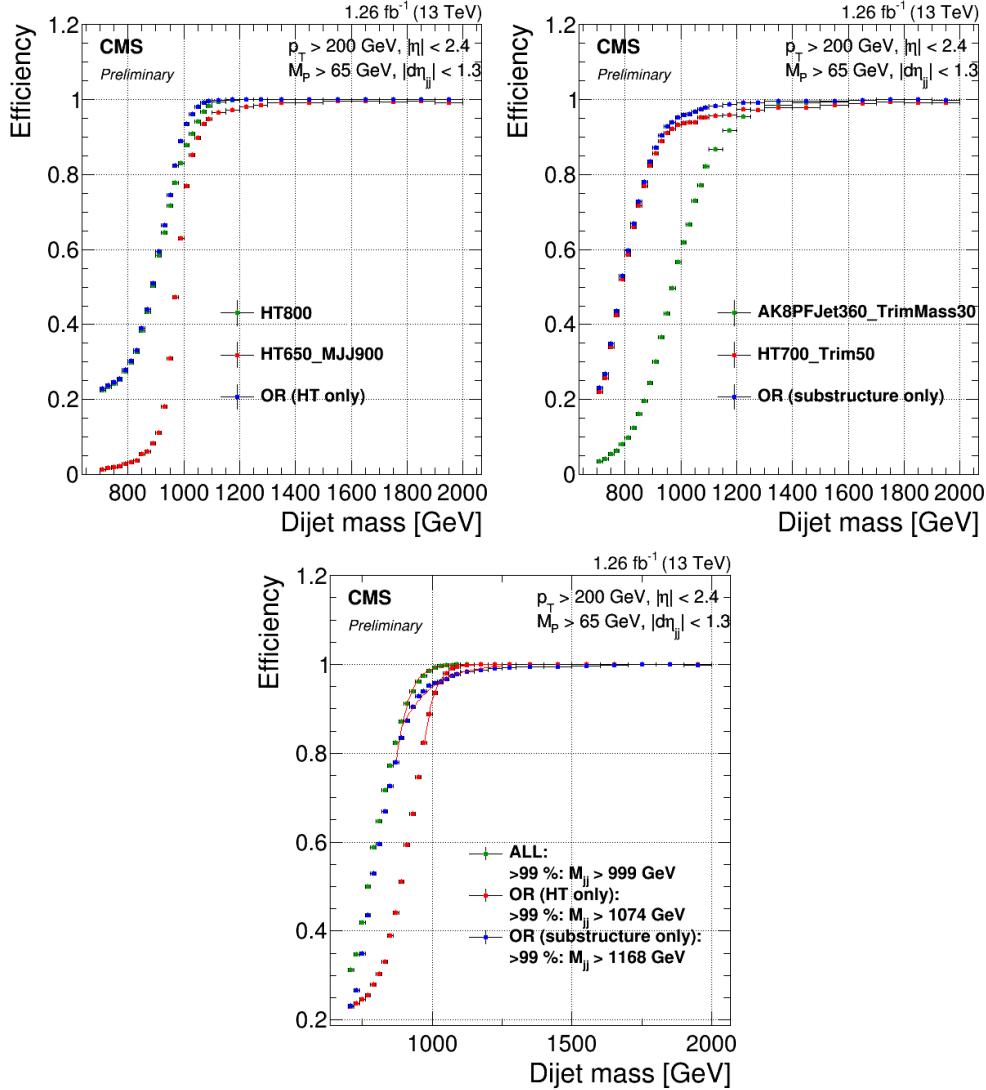


Figure 4.6: Top: Efficiency for the inclusive triggers (top left) and the grooming triggers (top right) as a function of dijet invariant mass for jet pairs where one jet has a pruned mass larger than 65 GeV. Bottom: Comparison of trigger efficiencies for jets passing one of the HT-triggers only (red), 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.

used in this analysis are clustered with the anti- k_T jet clustering algorithm with a clustering parameter of $R = 0.8$ (see Section 3.4) to allow containment of the full vector boson decay products. As we know that a minimum transverse of 200 GeV is required for the decay products

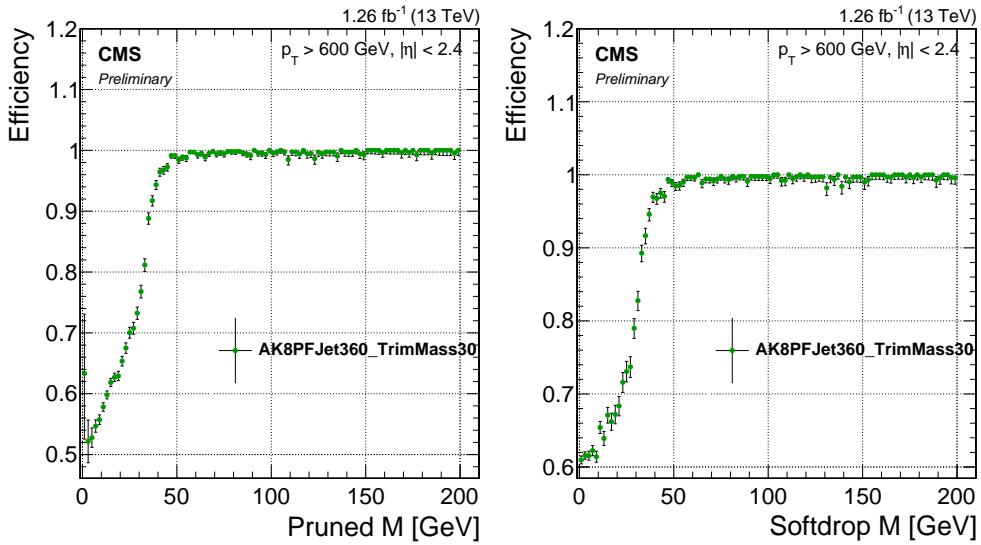


Figure 4.7: Efficiency for the lowest threshold grooming trigger as a function of pruned-jet (left) and softdrop-jet (right) mass for jets with $p_T > 600 \text{ GeV}$.

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 $p_T > 200 \text{ GeV}$. 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 forwardly produced, 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 4.8

A cut of $|\Delta\eta|_{jj} < 1.3$ makes sure to remove the t-channel pole at $\cos \theta^* = 1$ and is in addition found to yield the best separation between signal and the QCD background.

In addition the these requirements on the jets themselves, an overlap veto with leptons in the event is applied. 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 3.2.2 and 3.2.2, have a transverse momentum larger than 35

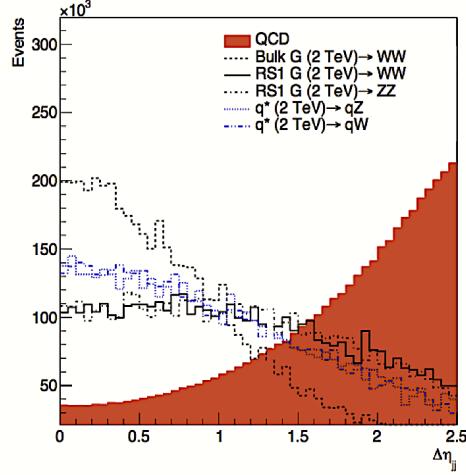


Figure 4.8: $|\Delta\eta|$ between the two highest- p_T jets for QCD jets and jets stemming from different signal scenarios.

1583 (30) GeV and a pseudorapidity smaller than 2.5 (2.4) in case of electrons (muons).

1584 The p_T , η , dijet invariant mass and $|\Delta\eta|_{jj}$ distribution for the two leading jets in the event
1585 after the above preselections have been applied is shown in Figure 4.67.

1586 Vector boson tagging

1587 After preselections, we take advantage of the jet substructure algorithms described in Section 3.5 to
1588 further separate boosted W/Z jets from the QCD multijets background. In the 8 TeV analysis [57]
1589 published the previous year, the pruning algorithm was the groomer of choice. However, recent
1590 progress had been made in the development of alternative groomers which had favorable properties
1591 from a theoretical point of view (see Sections 3.5.1 and 4.2.5). We therefore studied two different
1592 grooming algorithms: pruning and softdrop (with $\beta = 0$ and $z_{cut} = 0.1$). A comparison of
1593 the softdrop (dotted lines) and pruned (solid lines) jet mass for W, Z and H jets is shown in
1594 Figure 4.10.

1595 One of the first observations we made comparing the two groomers, was that there appeared
1596 to be a strong dependence of softdrop mass on the jet p_T . Figure 4.11 shows the pruned (left)
1597 and softdrop (right) mass distributions for W jets coming from the decay of a G_{bulk} with a
1598 resonance mass of $0.8 \text{ TeV} < M_X < 4 \text{ TeV}$. While the pruned jet mass mean appeared stable as
1599 the jet transverse momenta of the jet increased ($p_T \sim M_X/2$), the softdrop jet mass mean shifted
1600 towards lower values as jet p_T increased.

1601 In order to investigate whether this was a reconstruction effect or an algorithmic effect, we

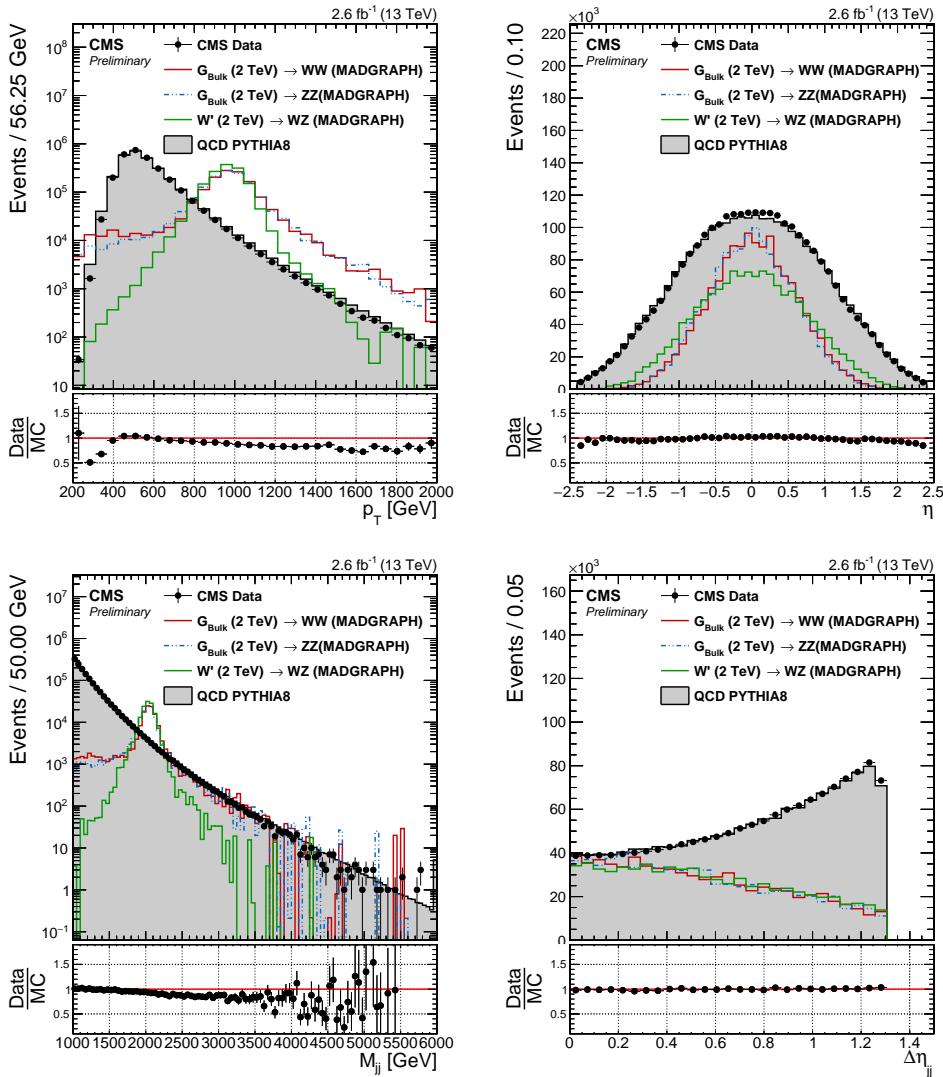


Figure 4.9: Jet p_T (top left), η (top right), dijet invariant mass (bottom left) and $|\Delta\eta|_{jj}$ (bottom right) distribution for the two leading jets in the event after loose preselections are applied. The signal is scaled by an arbitrary number.

1602 additionally looked at the pruned and softdrop mass for generator level jets (jets clustered with
 1603 generator level particles before they are passed through the detector simulation). Figure 4.12
 1604 shows the reconstructed (solid line) and generator level (dotted line) jet mass distributions after
 1605 pruning (left) or softdrop (right) have been applied. Again, the distributions are compared for jets
 1606 with very different p_T profiles, here for W jets coming from a $G_{\text{bulk}} \rightarrow WW$ of mass $M_X = 0.8 \text{ TeV}$
 1607 (red), roughly $p_T \sim 400 \text{ GeV}$, and $M_X = 2.0 \text{ TeV}$ (blue), $p_T \sim 1 \text{ TeV}$. Interestingly, we observe a

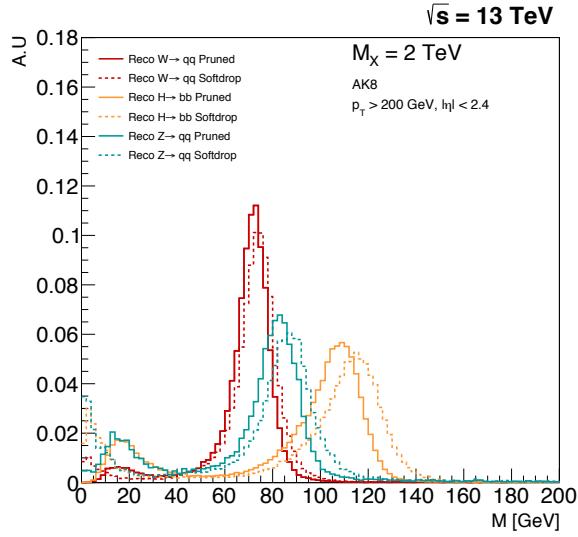


Figure 4.10: The softdrop (dotted lines) and the pruned (solid lines) jet mass for W, Z and H jets.

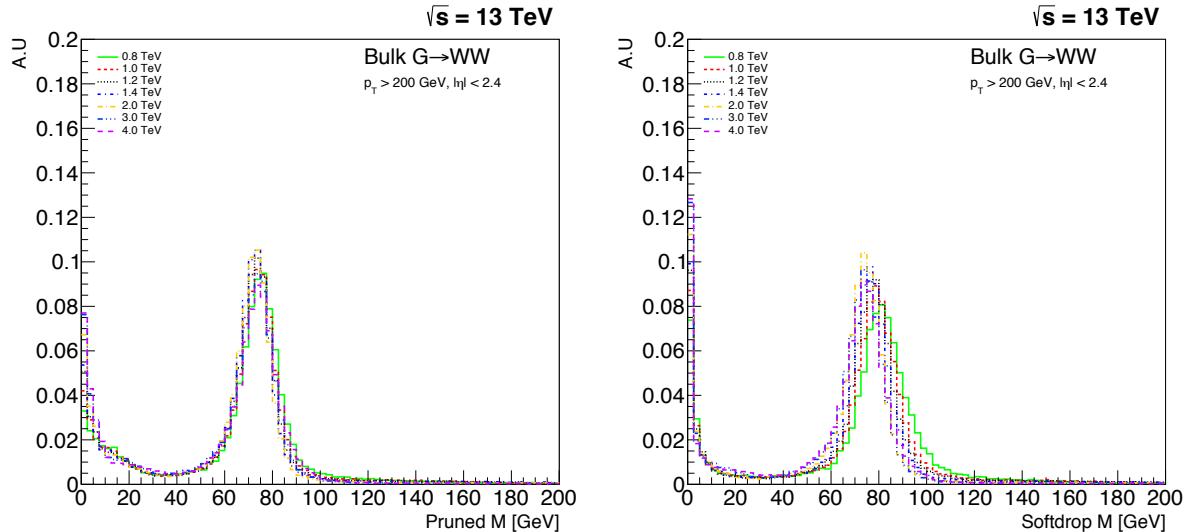


Figure 4.11: The jet mass distribution for W jets coming from a G_{bulk} of masses in the range $0.8 \text{ TeV} < M_X < 4 \text{ TeV}$ decaying to WW, here with pruning applied (left) and softdrop (right). A strong shift in the jet mass mean as a function of p_T ($\sim M_X/2$), is observed for jets groomed with the softdrop algorithm. Charge hadron subtraction is applied to all jets before clustering.

¹⁶⁰⁸ p_T -dependent mass shift already for generator level softdrop jets (comparing the dotted lines in

1609 the right plot); an effect further enhanced at reconstruction level. This effect is not present for
 1610 pruned jets, neither at generator level nor reconstruction level.

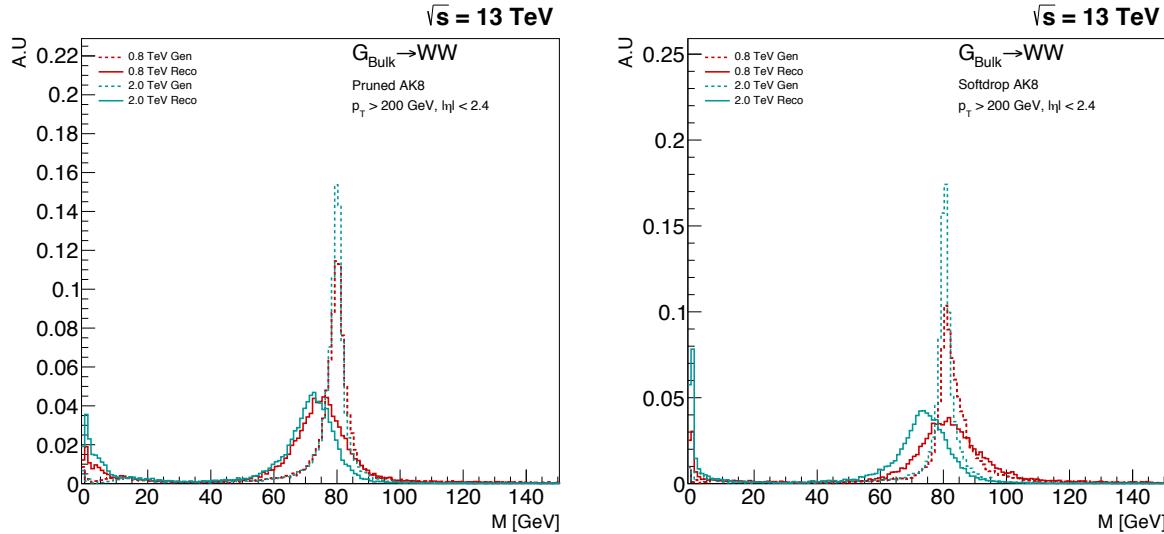


Figure 4.12: The reconstructed (solid line) and generator level (dotted line) jet mass distribution for W jets coming from a $G_{\text{bulk}} \rightarrow WW$ of mass $M_X = 0.8 \text{ TeV}$ (red), roughly $p_T \sim 400 \text{ GeV}$, and $M_X = 2.0 \text{ TeV}$ (blue), $p_T \sim 1 \text{ TeV}$. Here for the pruned (left) and softdrop (right) jet mass.

1611 The observed softdrop mass p_T -dependence was problematic, due to the fact that it would
 1612 require a p_T dependent mass window. This would again require several different measurements
 1613 of data to simulation tagging efficiency scale factors, for the respective mass windows, or a
 1614 significantly higher uncertainty on the signal yield. Due to these observations, the grooming
 1615 algorithm of choice for this analysis is pruning, with $65 \text{ GeV} < m_{\text{pruned}} < 105 \text{ GeV}$. However, this
 1616 would be a study we would return to in Search II (Section 4.2).

1617 The shape tagger we chose for this analysis was the n-subjettiness ratio τ_{21} . τ_{21} is strongly
 1618 correlated to the pruned jet mass, and the discriminating power of the variable is reduced when
 1619 applying a pruned mass cut. The τ_{21} distribution for the QCD background and W jets from a
 1620 signal decay before (left) and after (right) a pruned mass cut of $65 \text{ GeV} < m_{\text{pruned}} < 105 \text{ GeV}$
 1621 have been applied, is shown in Figure 4.13.

We therefore perform a cut optimization on τ_{21} after all analysis selections, including a pruned mass window of $65 \text{ GeV} < m_{\text{pruned}} < 105 \text{ GeV}$, have been applied. This is done by scanning the

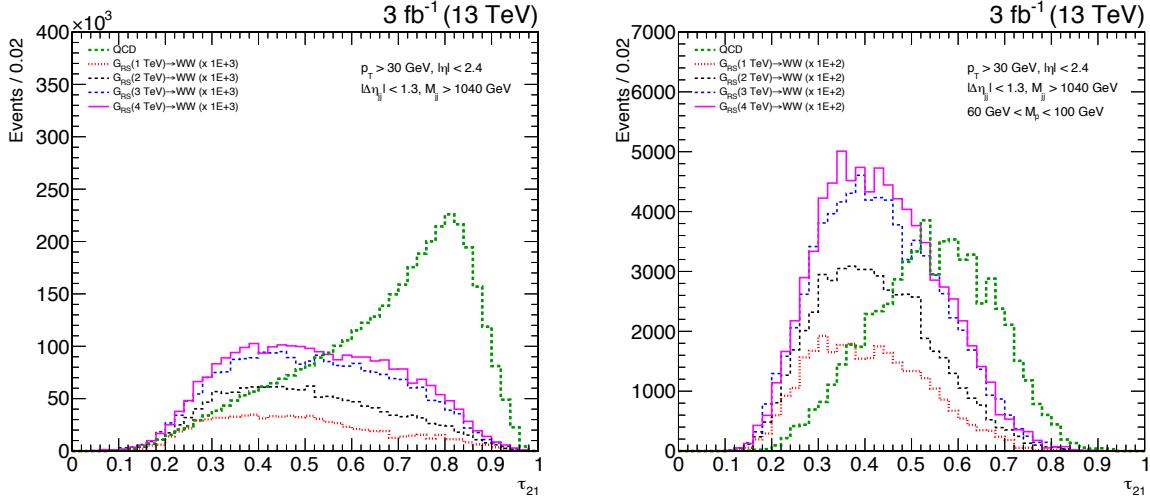


Figure 4.13: The τ_{21} distribution for QCD background and signal jets before (left) and after (right) a pruned mass window is applied. The discriminating power of τ_{21} is strongly reduced after grooming.

τ_{21} cut, and for each cut computing the Punzi significance [62] defined as

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

where ϵ_S is the signal efficiency and B is the total background. The cut with the highest significance is defined as the optimal cut value. The signals under consideration are W jets coming from the decay of a G_{bulk} with $M_X = 1 - 4 \text{ TeV}$, 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 4.14.

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 4.67. 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 mass points, we look at the ratio of a given τ_{21} cut over the significance of the best cut at that mass points. This is shown in the right plot of Figure 4.14. The cut $\tau_{21} < 0.45$ has the most stable performance out of the investigated cut values and is due to that, and due to the desire of keeping the background as low as possible at low m_{jj} , chosen as the nominal cut. In order to account for the fact that background is lower at

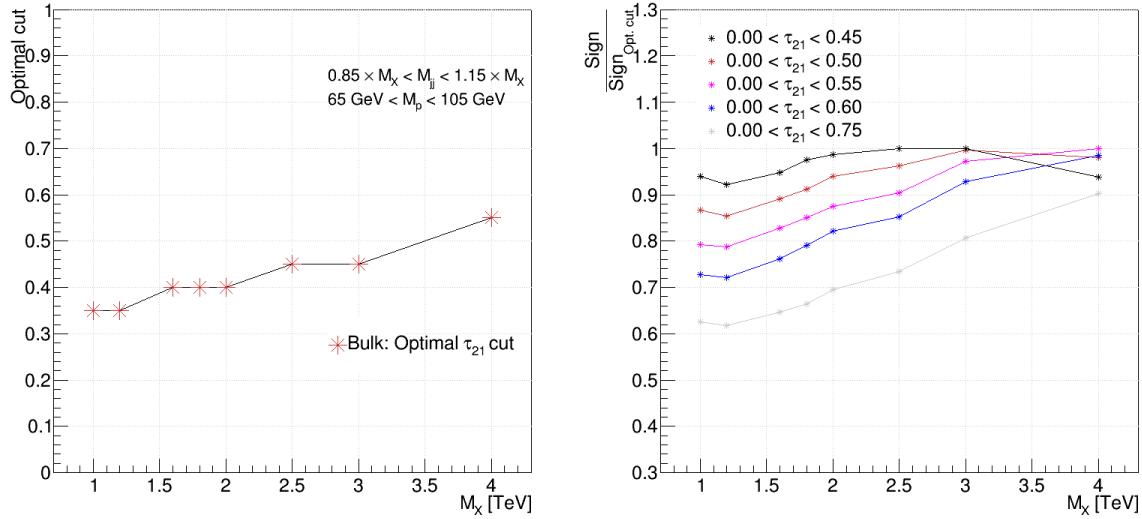


Figure 4.14: Optimal upper cut on τ_{21} as a function of G_{bulk} mass (left). The optimal τ_{21} selection for W' (HTV model) resembles the Bulk graviton selection.

1638 high- m_{jj} , we add an additional analysis category, $0.45 < \tau_{21} < 0.75$, which contains $> 95\%$ of the
 1639 signal and enhances the analysis sensitivity where the background is low. These categories are
 1640 hereafter referred to as the “high purity” (HP) category, for jets with $0 < \tau_{21} \leq 0.45$, and the low
 1641 purity (LP) category, for jets with $0.45 < \tau_{21} \leq 0.75$.

1642
 1643 The W -tagging efficiency and QCD light-flavored jet mistagging rate for a W -tagger consisting of
 1644 $0 < \tau_{21} \leq 0.45$ and $65 \text{ GeV} < m_{\text{pruned}} < 105 \text{ GeV}$ is shown in Figure 4.15, both as a function of
 1645 jet p_T and as a function of number of primary vertices in the event.

1646 The signal efficiency for a pruned jet mass cut only, is around 80 % with a mistag rate of
 1647 $\sim 15\%$. After adding a τ_{21} cut, the signal efficiency drops to around 55% and the mistagging
 1648 rate to $\sim 2\%$. Another interesting feature is the dependence of τ_{21} on p_T on pileup, compared
 1649 to the resilience of the groomed mass as a function of the same variables. This will be another
 1650 feature we explore in Search II (Section 4.2). Figure 4.1.7 shows the pruned-jet mass (left) and
 1651 the n-subjettiness τ_{21} distribution (right) for signal and background Monte Carlo, as well as the
 1652 distributions measured in data.

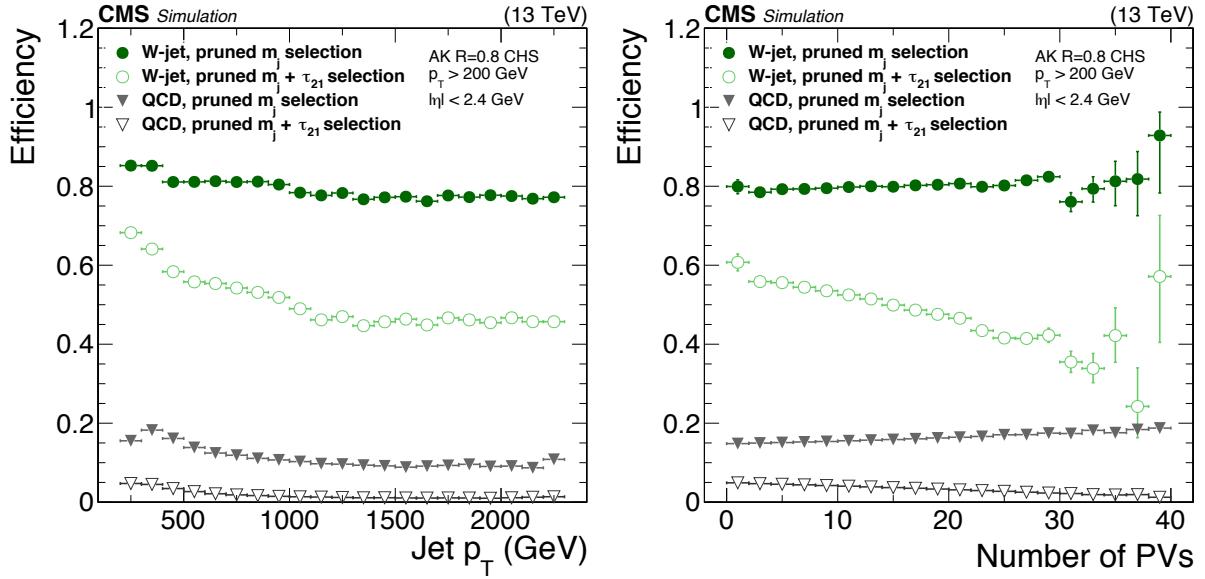


Figure 4.15: The W-tagging efficiency (green) and light jet mistag rate (grey) for a pruned jet mass cut only and pruned jet mass + τ_{21} cut as a function of p_T (left) and number of primary vertices (right).

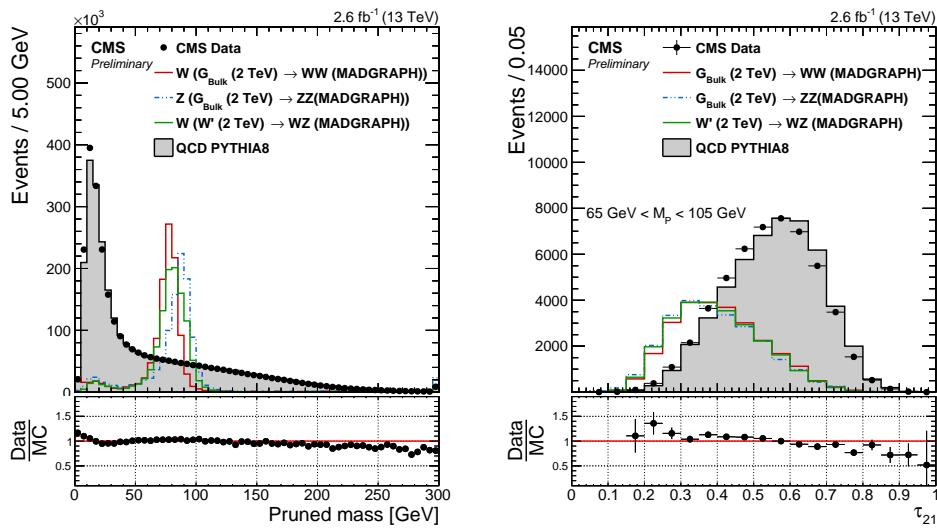


Figure 4.16: Pruned jet mass distribution (left) and n-subjettiness τ_{21} (right) distribution for data and simulated samples. Simulated samples are scaled to match the distribution in data. The τ_{21} distribution is shown for jets after a cut of $65 \text{ GeV} < M_p < 105 \text{ GeV}$ has been applied.

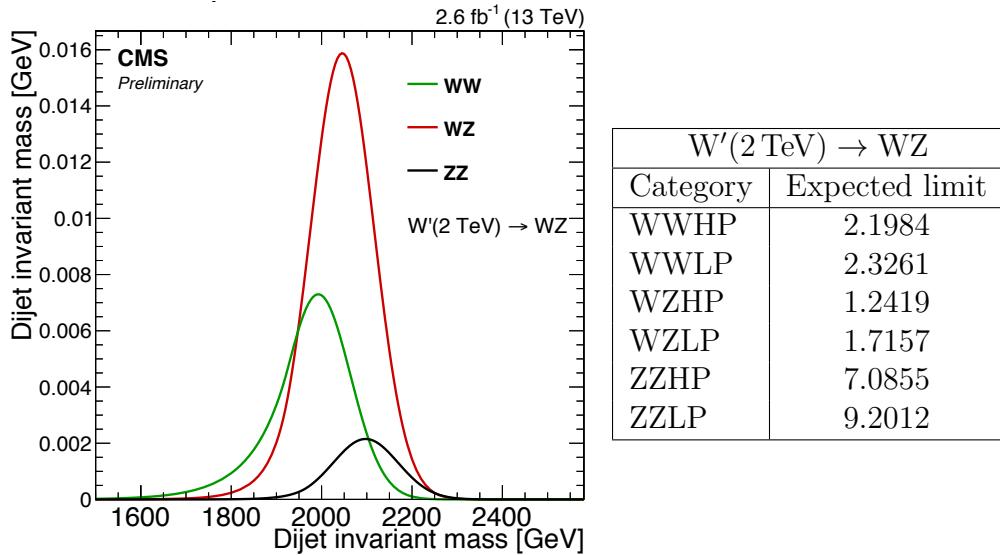


Figure 4.17: The expected signal yield per mass category for W' (2 TeV) decaying to a W and Z (left) together with the expected limit per mass category for the same signal (right).

1653 Analysis categorization

1654 As the analysis requires two W/Z-tags, we always require one HP tagged jet and then divide into
 1655 LP and HP categories depending on whether the other jet is of high or low purity. In addition, in
 1656 order to further enhance the analysis sensitivity, we further split the pruned jet mass window
 1657 into a W and a Z boson window where the W window is defined as $65 \text{ GeV} < m_{\text{pruned}} < 85 \text{ GeV}$
 1658 and the Z boson window as $85 \text{ GeV} < m_{\text{pruned}} < 105 \text{ GeV}$. This has the added benefit of allowing
 1659 us to make a discrimination between a G_{bulk} decaying to WW or ZZ, and a W' decaying into
 1660 WZ through counting events in each category. We, for instance, expect a higher signal yield
 1661 in WZ category for a W' decaying to a W and Z boson than for a G_{bulk} decaying to WW or
 1662 ZZ. Figure 4.17 shows the relative expected signal yield (left) and expected limits (left) in the
 1663 different mass categories for a 2 TeV W' .

1664 All categories are combined in the end, leading to the same or better sensitivity than when
 1665 using the whole pruned mass window. Figure 4.18 shows the expected 95% CL upper limits on
 1666 the production cross section of a W' decaying to WZ (left) and a G_{bulk} decaying to WW (right)
 1667 as function of the resonance mass in the HP category. The blue line corresponds to the expected
 1668 limits obtained when not splitting into mass categories and the red line corresponds to the limit
 1669 using the combination of two categories. The dotted and solid black lines are the limits in the W
 1670 and Z categories, respectively. The combination of two mass categories leads to a slightly better
 1671 (10%) or to the same sensitivity as when using one large mass window.

1672 The real benefit of splitting into mass categories becomes obvious when defining a test
1673 statistics based on the likelihood ratios of each signal hypothesis, $q = -2 \ln(L_{G_{\text{bulk}}}/L_{W'})$, shown
1674 in Figure 4.19. For a signal with a signal strength corresponding to a $3\text{-}4 \sigma$ excess, the test
1675 statistics for each signal hypothesis are well separated ($\sim 3.5\sigma$), allowing us to make a statement
1676 of how G_{bulk} or W' like a possible signal is. With the high-purity and low-purity categories as
1677 defined above for each mass window combination, this leaves us with six different signal categories.
1678 They are as follows:

- 1679 • High-purity, 3 mass categories: WW, ZZ and WZ
1680 • Low-purity , 3 mass categories: WW, ZZ and WZ

1681 In parallel to the mass-category based analysis, we perform an analysis without categorization
1682 in mass (similar to the 8 TeV analysis) as a cross-check, and found the sensitivity with mass
1683 categories to be better. The final tagging efficiency for different signal hypothesis (top) together
1684 with the QCD mistag rate (bottom) in the different signal categories is shown in Figure 4.20.
1685 The solid lines represent the tagging efficiency in the full mass window ($65 \text{ GeV} < M_p < 105 \text{ GeV}$)
1686 before splitting into mass categories. A lower signal efficiency the ZZ mass category is observed in
1687 all cases. This can be explained from the pruned jet mass distribution on the left in Figure 4.1.7,
1688 where a cut at 85 GeV leaves a large fraction of the Z peak in the W mass window. As the main
1689 benchmark models under consideration preferably decays to W bosons (in the Bulk Graviton
1690 model the branching ratio $\text{BR}(G_{\text{Bulk}} \rightarrow W W) = 2 * \text{BR}(G_{\text{Bulk}} \rightarrow ZZ)$, and in the HVT model
1691 $W'/Z' \rightarrow WZ/WW$ (but not ZZ)), a high tagging efficiency for the W boson is preferred. In
1692 the limit-setting procedure all the categories are combined and the overall signal efficiency is
1693 conserved. For the combined mass-categories (solid line) the signal efficiency is between 16 and
1694 23 % in the double-tag categories, and between 20 and 34 % in the single-V tag categories. The
1695 mistagging rate in the double-V tag categories is below 1 % in the high-purity category. The full
1696 analysis selections and final categories are listen in Table 4.1.

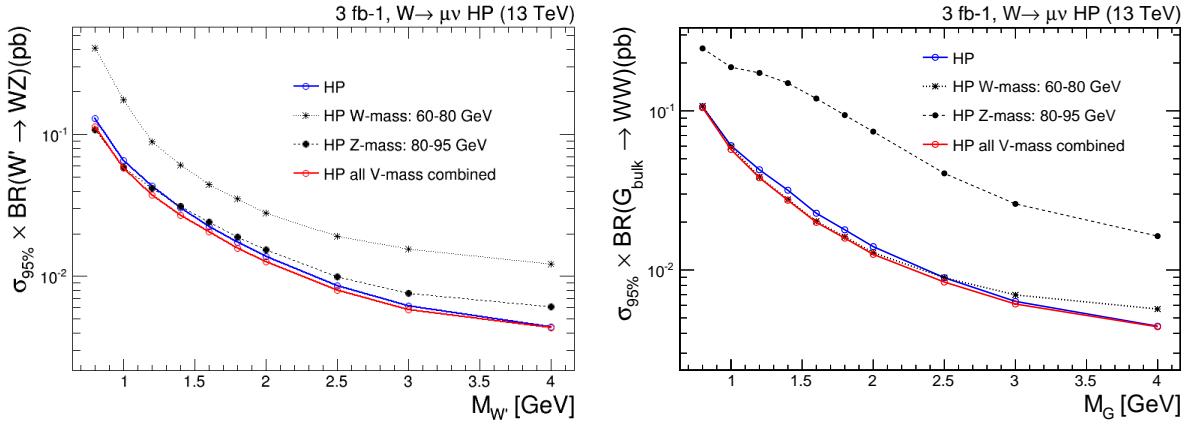


Figure 4.18: Expected 95% CL upper limits on the production cross section of a W' (left) and G_{bulk} (right) signal as function of the resonance mass for the different mass categories for events passing the high-purity τ_{21} selections.

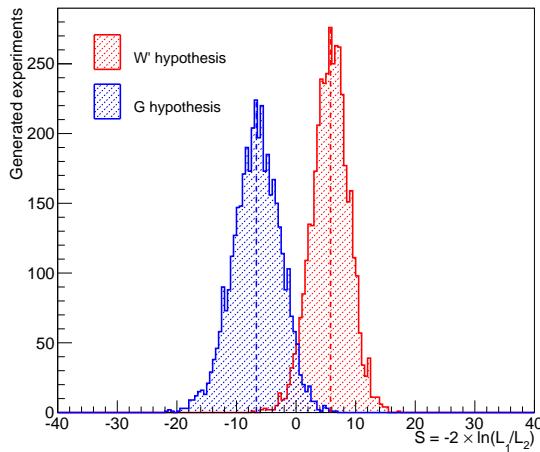


Figure 4.19: Distribution of the test statistic $q = -2 \ln(L_{G_{\text{bulk}}} / L_{W'})$ for a G_{bulk} (blue) and W' signal hypothesis.

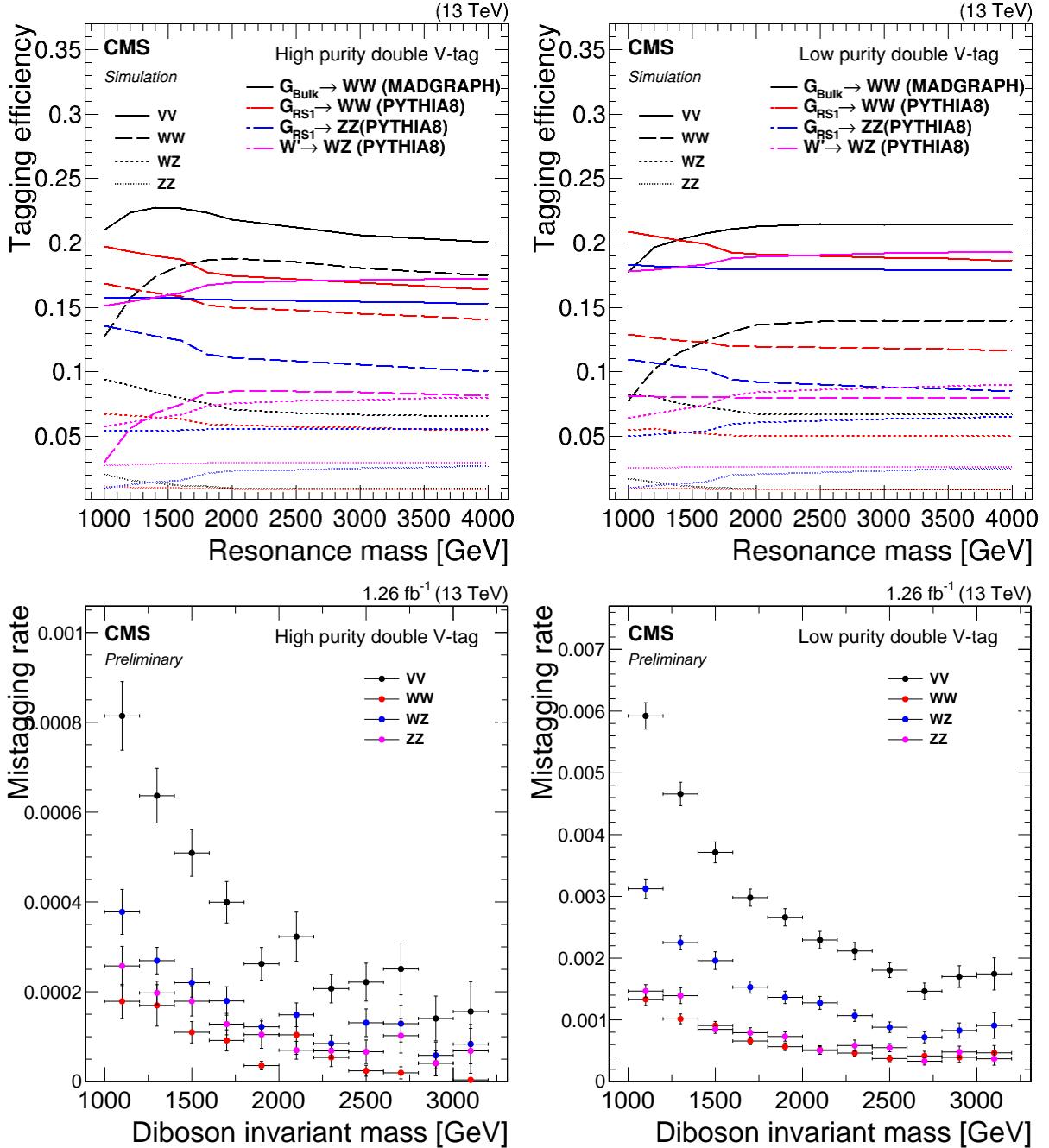


Figure 4.20: Tagging efficiency (top) and mistagging rate (bottom) in the different pruned mass categories in the high-purity category (left) and in the low-purity category (right)

Selection	Value
<hr/>	
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	$\tau_{21} < 0.75$
<hr/>	
m_{jet} categories	
WW	$65 < m_{\text{jet}_1} < 85 \text{ GeV}, 65 < m_{\text{jet}_2} < 85 \text{ GeV}$
WZ	$65 < m_{\text{jet}_1} < 85 \text{ GeV}, 85 < m_{\text{jet}_2} < 105 \text{ GeV}$
ZZ	$85 < m_{\text{jet}_1} < 105 \text{ GeV}, 85 < m_{\text{jet}_2} < 105 \text{ GeV}$
<hr/>	
τ_{21} categories	
High-purity	$\tau_{21,\text{jet}1} < 0.45, \tau_{21,\text{jet}2} < 0.45$
Low-purity	$\tau_{21,\text{jet}1} < 0.45, 0.45 < \tau_{21,\text{jet}2} < 0.75$

Table 4.1: The full analysis selections, mass and τ_{21} categories.

¹⁶⁹⁷ 4.1.5 Background modeling

¹⁶⁹⁸ The background modeling in this analysis is based on a smoothness test performed directly on
¹⁶⁹⁹ unblinded data, similar to what is done in previous CMS analyses looking for bumps in the dijet
¹⁷⁰⁰ invariant mass spectrum [63, 64]. We assume that the QCD multijet background in the different
¹⁷⁰¹ analysis categories can be described by smooth, monotonically decreasing functions of 2 or 3
¹⁷⁰² parameters

$$\frac{dN}{dm_{jj}} = \frac{P_0}{(m_{jj}/\sqrt{s})^{P_2}} \quad \text{and} \quad \frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}}, \quad (4.1)$$

where m is the dijet invariant mass, \sqrt{s} the centre of mass energy and 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 Fishers F-test [65]. In this test, we start from the 2 parameter function and compare the goodness of fit (χ^2 divided by degrees of freedom) when fitting the data signal region with a 2, 3, 4 and 5 parameter function. We then check at 10% confidence level (CL) if additional parameters are needed to model the background distribution. The 4 and 5 parameter functions are

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2+P_3 \times \log(m_{jj}/\sqrt{s})}} \quad (4.2)$$

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2+P_3 \times \log(m_{jj}/\sqrt{s})+P_4 \times \log(m_{jj}/\sqrt{s})^2}} \quad (4.3)$$

¹⁷⁰³ where P_3 and P_4 are additional free parameters. As an additional cross check, an alternative fit
¹⁷⁰⁴ function is also tested:

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s} + P_3(m_{jj}/\sqrt{s})^2)^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}}. \quad (4.4)$$

¹⁷⁰⁵ The fit range is chosen such that it start where the trigger efficiency has reached its plateau
¹⁷⁰⁶ to avoid bias from trigger inefficiency, and extends to the bin after the highest m_{VV} mass point.
¹⁷⁰⁷ The binning chosen for the fit follows the detector resolution as in [63, 64]. Before unblinding the
¹⁷⁰⁸ signal region, we check that the QCD dijet invariant mass spectrum is expected to be smooth from
¹⁷⁰⁹ the distribution in QCD MC as well as exercise the F-test in QCD MC and in a data sideband.

¹⁷¹⁰ The fits to data in the signal region using the different fit functions, are shown in Figure 4.21,
¹⁷¹¹ and the corresponding F-test output are given in Table 4.2 through Table 4.4. The findings can
¹⁷¹² be summarized as follows: for the WW enriched category a 2 parameter fit is sufficient to describe
¹⁷¹³ the data in both the high- and low-purity categories. In the WZ category, a two parameter fit

1714 is sufficient in the high-purity category, while three parameters are needed for the low-purity
1715 category. For the ZZ category, a 3 parameter fit is needed for both purity categories. The 2
1716 and 3 parameters fit functions as defined in Equation B.2 will therefore be used to model the
1717 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
Fishers23	-0.053	CL	1.0	Fishers23	-1.723	CL	1.0
Fishers34	-1.456	CL	1.0	Fishers34	-1.191	CL	1.0

Table 4.2: Residuals, χ^2 , and degrees of freedom for the WW enriched HP and LP categories. A 2 parameter fit is needed to describe the data in both categories.

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
Fishers23	-2.598	CL	1.0	Fishers23	55.258	CL	0.0
Fishers34	-0.491	CL	1.0	Fishers34	0.078	CL	0.783

Table 4.3: Residuals, χ^2 , and degrees of freedom 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
Fishers23	6.302	CL	0.029	Fishers23	40.438	CL	0.0
Fishers34	1.246	CL	0.290	Fishers34	0.356	CL	0.56

Table 4.4: Residuals, χ^2 , and degrees of freedom for the ZZ enriched LP and HP categories. A 3 parameter fit is sufficient to describe the data in both categories.

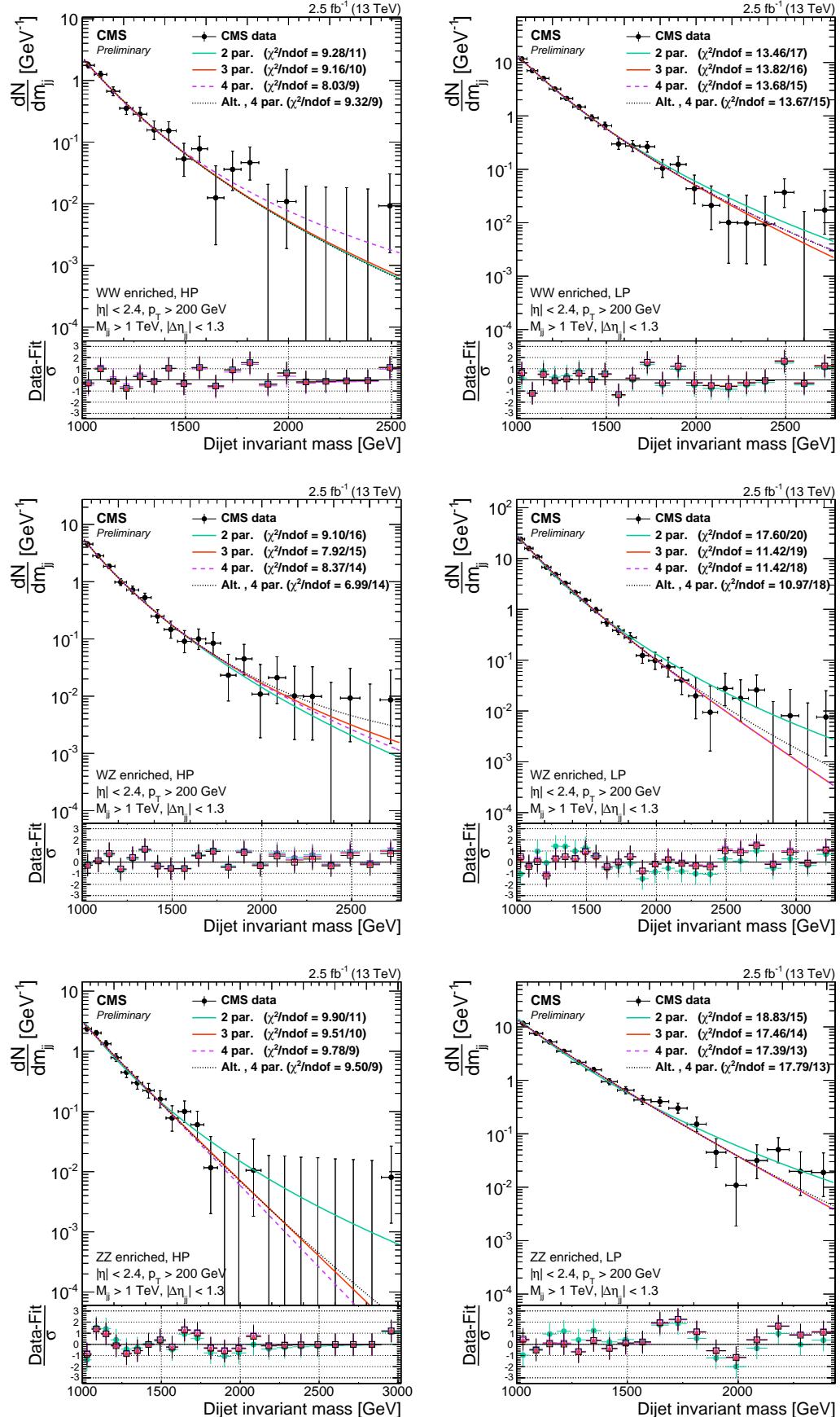


Figure 4.21: Fitted dijet mass spectrum in the different mass and purity categories in data for

₁₇₁₈ **4.1.6 Signal modeling**

₁₇₁₉ The signal shape is extracted from signal MC with masses in the range from 1 to 4 TeV. A
₁₇₂₀ linear interpolation provides shapes for the mass points in between in steps of 100 GeV. From
₁₇₂₁ these shapes, pdf models are constructed as composite models with a Gaussian core due to
₁₇₂₂ detector resolution and an exponential tail to account for parton distribution function effects.
₁₇₂₃ Parametric shape uncertainties due to jet energy scale and resolution uncertainties are inserted by
₁₇₂₄ variations of the Gaussian peak position and width. The dijet invariant mass shape for different
₁₇₂₅ benchmark model signals are shown in Figure 4.22. The signal and background components are
then simultaneously fitted to the data points.

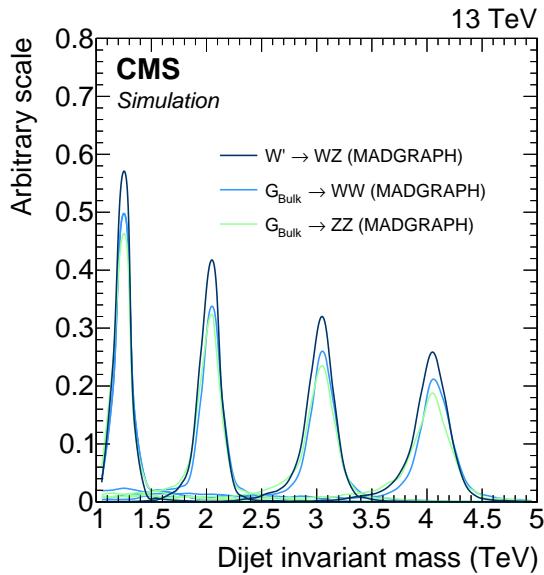


Figure 4.22: Dijet invariant mass from signal MC used to extract the signal shape. Here for 1.2, 2, 3 and 4 TeV resonances.

¹⁷²⁷ 4.1.7 V-tagging scale factors

¹⁷²⁸ As seen in Figure , some discrepancy is observed in the τ_{21} distribution between data and MC.
¹⁷²⁹ This can lead to a bias in the signal efficiency estimation and we therefore measure the real data
¹⁷³⁰ signal efficiency in an orthogonal data sample. The W-tagging efficiency is measured using real
¹⁷³¹ boosted W-jets in a semi-leptonic $t\bar{t}$ enriched data sample. This region is mainly quark-enriched,
¹⁷³² as opposed to the QCD gluon-enriched region we saw previously, and substructure variables are
¹⁷³³ better described here. The sample is obtained through requiring a final state compatible with
¹⁷³⁴ two b-jets and two W bosons, where one of the bosons decay leptonically and the other one
¹⁷³⁵ hadronically. There are several good reasons to use this channel: Top quark pair production
¹⁷³⁶ events are plentifully produced at the LHC, we can ensure a high purity of the sample through
¹⁷³⁷ high-energy lepton, b-tag and missing energy requirements and lastly we can ensure that the
¹⁷³⁸ W jets are boosted by requiring the leptonic leg, together with the hadronic W candidate, to
¹⁷³⁹ have high transverse momentum. The final state is illustrated in Figure 4.23, with the object of
¹⁷⁴⁰ interest being the AK R=0.8 jet containing the two quark daughters of the hadronically decaying
W.

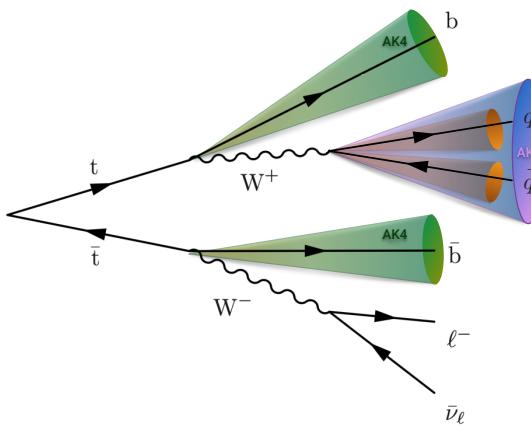


Figure 4.23: A top quark pair decaying into two b quarks and two W bosons, one of which decays leptonically and one on which decays hadronically

¹⁷⁴¹

¹⁷⁴² Event selection

The W can decay either to an electron or a muon, both final states (“channels”) are used in the analysis. We select events through triggering and selections on the leptonic leg. First, we require a high-energy lepton at trigger level, with an online p_T above 45 GeV for the muon and

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 3.2.2 and Section 3.2.2, and events containing additional leptons (passing the same ID requirements, but looser cuts as defined in Table 4.5) are vetoed. Offline, we further require a high missing energy of 40(80) GeV in the muon(electron) channel. To insure a high signal (boosted hadronic W) purity, the leptonic W four-vector is reconstructed such that we can put tight momentum requirements on the leptonic leg (ensuring that both tops, and therefore vector bosons, have a high momentum). The leptonic W 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 mass

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

1743 This results in a completely defined neutrino four-vector, which is then added to the lepton
 1744 four-vector. The sum of the two defines the leptonic W and its momentum is required to be
 1745 greater than 200 GeV.

1746 Further, we require at least one AK R=0.4 jet to be b-tagged with the Combined Secondary
 1747 Vertex (CSV) algorithm [66, 67]. This algorithm exploits the relatively long lifetime of b quarks
 1748 leading to the presence of a displaced vertex, in order to distinguish between jets originating from
 1749 b quarks to those originating from light flavor quarks. More information on the CSV algorithm
 1750 can be found in [66, 67]. The reason for requiring only one b-tagged jet is to ensure a high
 1751 selection efficiency.

1752 Finally, we require at least one AK R=0.8 jet in the event with a momentum greater than 200
 1753 GeV which will be the hadronic W candidate. It's pruned jet mass is required to be between 40
 1754 GeV and 150 GeV. After reconstructing and selecting all our objects, a set of angular selections
 1755 are applied to ensure a diboson like topology. These are the following:

- 1756 • $\Delta R(l, W_{AK8}) > \pi/2$
- 1757 • $\Delta\phi(W_{AK8}, E_T^{\text{miss}}) > 2$
- 1758 • $\Delta\phi(W_{AK8}, W_{lep}) > 2$

1759 With these requirements, we have a nearly pure sample of $t\bar{t}$ events, with a small contamination
 1760 from single top, W+jets and VV events. A summary of the final selection criteria is presented in
 1761 Table 4.5. The pruned jet mass and τ_{21} variables in data and in MC are shown in Figure 4.24.

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	$p_T > 200 \text{ GeV}$	For hadronic
Jet η	$ \eta < 2.4$	W reconstruction
AK4 jet selections		
Jet p_T	$p_T > 30 \text{ GeV}$	Used for b-tag
Jet η	$ \eta < 2.4$	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 4.5: Summary of the final semi-leptonic $t\bar{t}$ selections.

1762 Fitting procedure

1763 For this measurement, what we are interested in is to extract and compare the W-tagging efficiency
 1764 of the combined jet mass and τ_{21} selection in data and in MC. We are additionally interested in
 1765 the difference in jet mass scale (mean of the W jet mass peak) and jet mass resolution (width of
 1766 W jet mass peak), as this also affects the signal jet mass shape and therefore efficiency. In order
 1767 to study these variables, we look at the pruned jet mass spectrum between 40 and 150 GeV in
 1768 two regions:

- 1769 • Pass region: $0 < \tau_{21} \leq 0.45 \sim \text{high purity}$
- 1770 • Fail region: $0.45 < \tau_{21} \leq 0.75 \sim \text{low purity}$

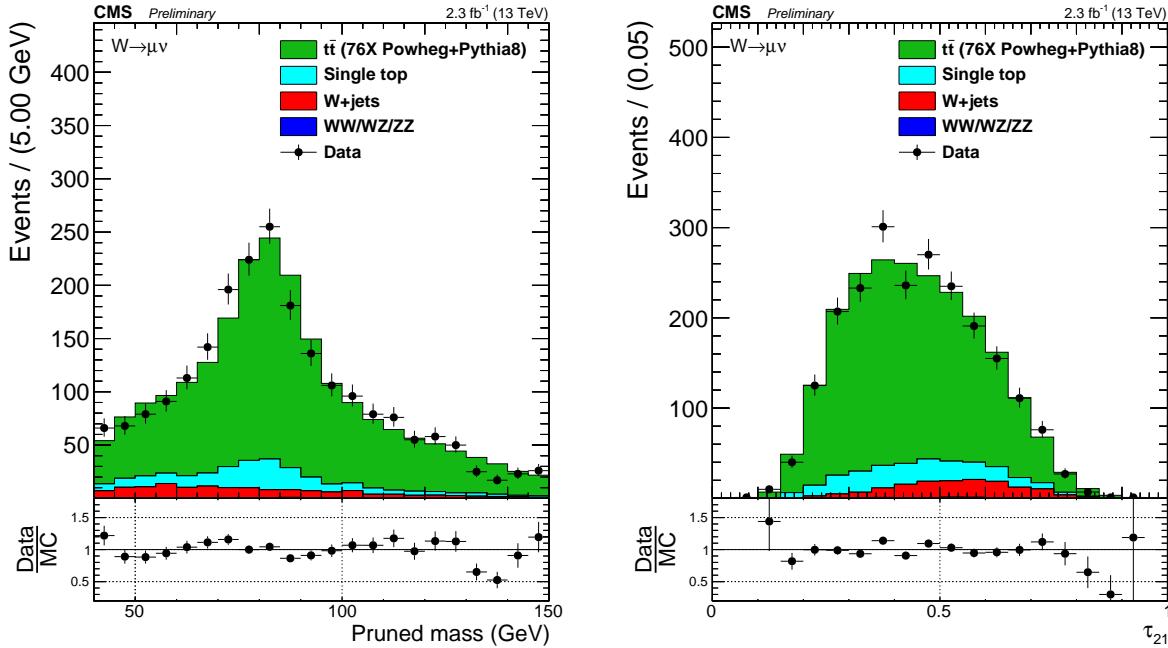


Figure 4.24: Distribution of pruned jet mass (left) and n-subjettiness (right) in the $t\bar{t}$ control sample.

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 jets falling outside of this region is negligible). The strategy is the following: We first derive probability density functions (PDFs) which describe the distribution of fully merged W jets and non-W 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. As simultaneous fit of pass and fail is then performed (using the two composite W +non-W PDFs), where the fraction of real W jets in both pass and fail is constrained such that, if the signal efficiency in pass is ϵ_S , the signal efficiency in fail 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 yield in pass and fail combined multiplied by some fraction, ϵ_S . The normalization of the PDF describing real W jets in the fail category is then the total real W yield multiplied by $(1 - \epsilon_S)$.

To understand which part of the $t\bar{t}$ jet mass distribution contains “real” merged Ws and which are only pure combinatorial background, non-PWs, we start from $t\bar{t}$ MC. By matching the AK8 jet with quarks coming from the hadronic W at generator level, in a cone of $\Delta R < 0.8$, we can

access the real merged W and non-merged W 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:

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

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

1771 Figure 4.25 shows the fitted PUPPI softdrop mass spectrum for $t\bar{t}$ real W (top) and non-W
1772 (bottom) distributions for jets that passed (left) and failed (right column) the N-subjettiness
1773 selection PUPPI $\tau_{21} < 0.4$. The corresponding plots for the jet pruned mass can be found in
1774 Figure B.1.

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 have 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, hereby 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 [68]), 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 tests 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 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 estimating the cut

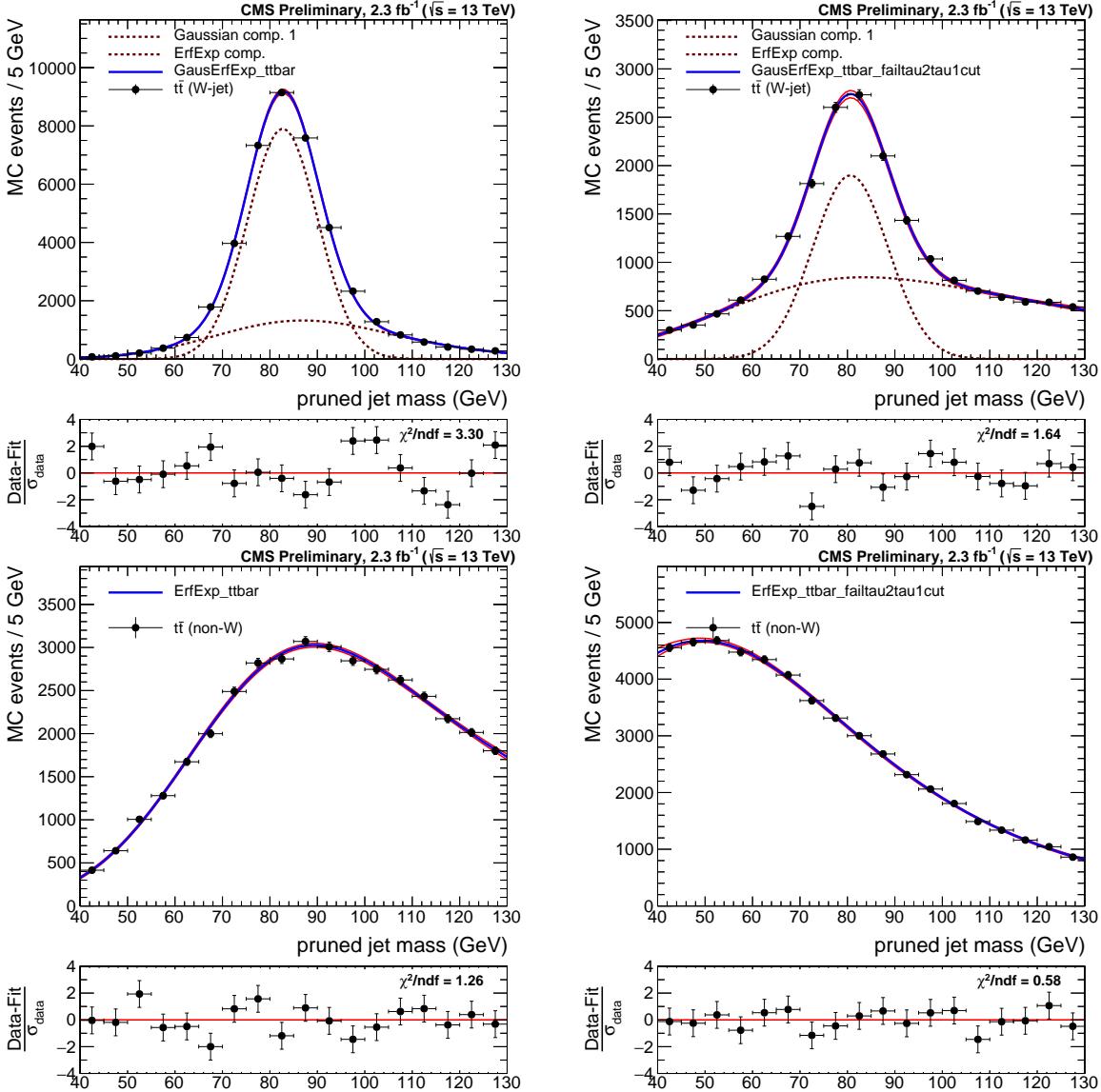


Figure 4.25: Fit to the real W (top) and non-W (bottom) pruned jet mass distribution for jets that pass (left) and fail (right) the cut on $\tau_{21} < 0.45$.

efficiency (ϵ_{HP}) on both data and simulated samples fitting, simultaneously, pass and fail samples:

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

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

where N_W is the number of real W jets, N_2 and N_3 are the number of combinatorial background events passing and failing the τ_{21} cut respectively. 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

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

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

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

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

$$f_{\text{pass}}^{\text{wjet}} = F_{\text{ErfExp}}(x) = e^{c_0x} \cdot \frac{1 + \text{Erf}((x - a)/b)}{2}$$

$$f_{\text{fail}}^{\text{wjet}} = F_{\text{ErfExp}}(x) = e^{c_0x} \cdot \frac{1 + \text{Erf}((x - a)/b)}{2}$$

with the corresponding distributions shown in Figure 4.26. 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 are then taken as the W-tagging scale factor:

$$SF_{HP} = \frac{\epsilon_{HP}(\text{data})}{\epsilon_{HP}(\text{sim})} \quad (4.5)$$

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

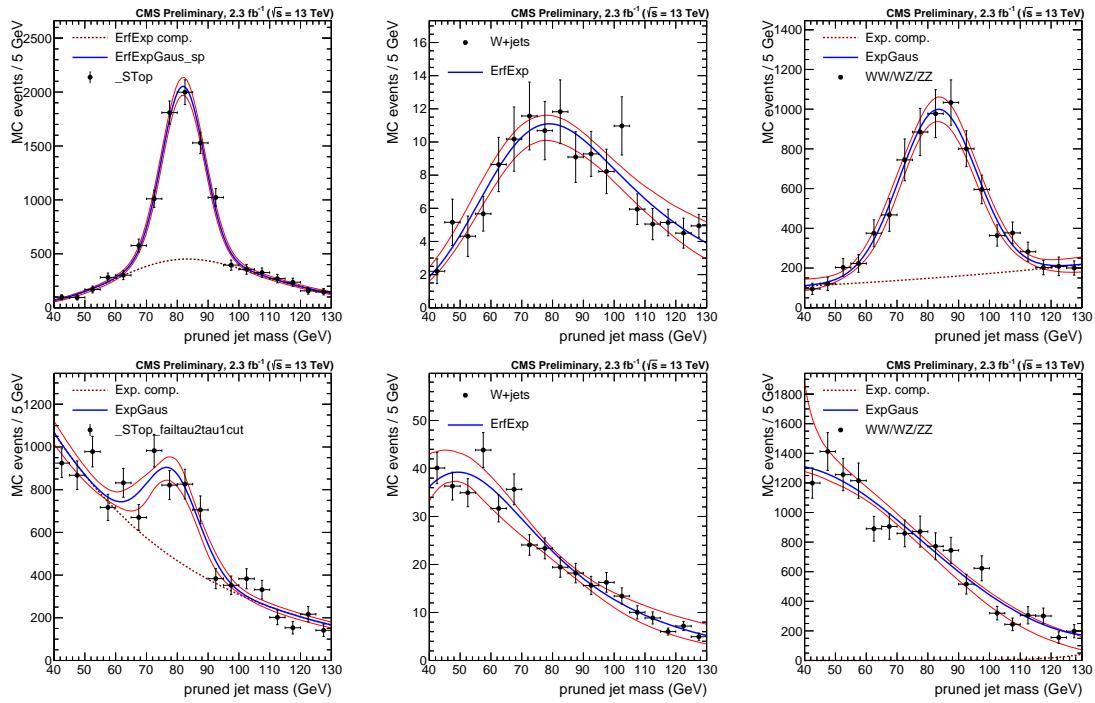


Figure 4.26: Fits to the pruned jet mass spectrum for the non-dominant backgrounds (Single top, W+jets and VV respectively) in the pass (top) and fail (bottom) regions.

1781 the definition of the low purity scale factor simplifies to

$$SF_{LP} = \frac{1 - \epsilon_{HP}(\text{data})}{1 - \epsilon_{HP}(\text{sim})} \quad (4.6)$$

1782 Systematic uncertainties

1783 As systematic uncertainties, we consider effects due to differences in $t\bar{t}$ simulation as well as effects
 1784 due to choice of fit method. The former is evaluated by comparing the extracted scale factor
 1785 when using $t\bar{t}$ MC samples produced with different matrix element (ME) and shower generators:
 1786 POWHEG (NLO) interfaced with PYTHIA8 , MADGRAPH (LO) QCD interfaced with HERWIG++
 1787 and POWHEG interfaced with HERWIG++. The uncertainty due to different ME generators
 1788 (POWHEG versus MADGRAPH) correspond to 3(13)% and are listed in Table 4.6 as the first quoted
 1789 systematic uncertainty. The uncertainty due to parton showering (PYTHIA8 versus HERWIG++)
 1790 is 8.6%, but are not relevant for analyses where no HERWIG++ based simulation is used, as is
 1791 the case for the search presented in this chapter. For the systematic uncertainty accounting for
 1792 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 4.27 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 being corrected for the fraction of W jets in the tails, agree within 0.3(0.8)% and are listed as systematic uncertainty in Table 4.6. 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 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 scalefactor changes as a function of p_T . This is estimated by comparing the difference in tagging between $G_{\text{bulk}} \rightarrow WW$ signal MC showered by PYTHIA8 and HERWIG++ as a function of 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.

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 B.4, with the corresponding extracted efficiencies and scale factors summarized in Table 4.6.

Category	Working point	Eff. data	Eff. simulation	Scale factor
HP	$\tau_2/\tau_1 < 0.45$	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	$0.45 < \tau_2/\tau_1 < 0.75$	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 4.6: Efficiencies in data and in MC together with the corresponding W-tagging scale factors for the high purity and low purity categories.

We additionally extract the jet mass scale and jet mass resolution, used to scale 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 4.7. Both the jet mass scale as well as the jet mass resolution is larger in

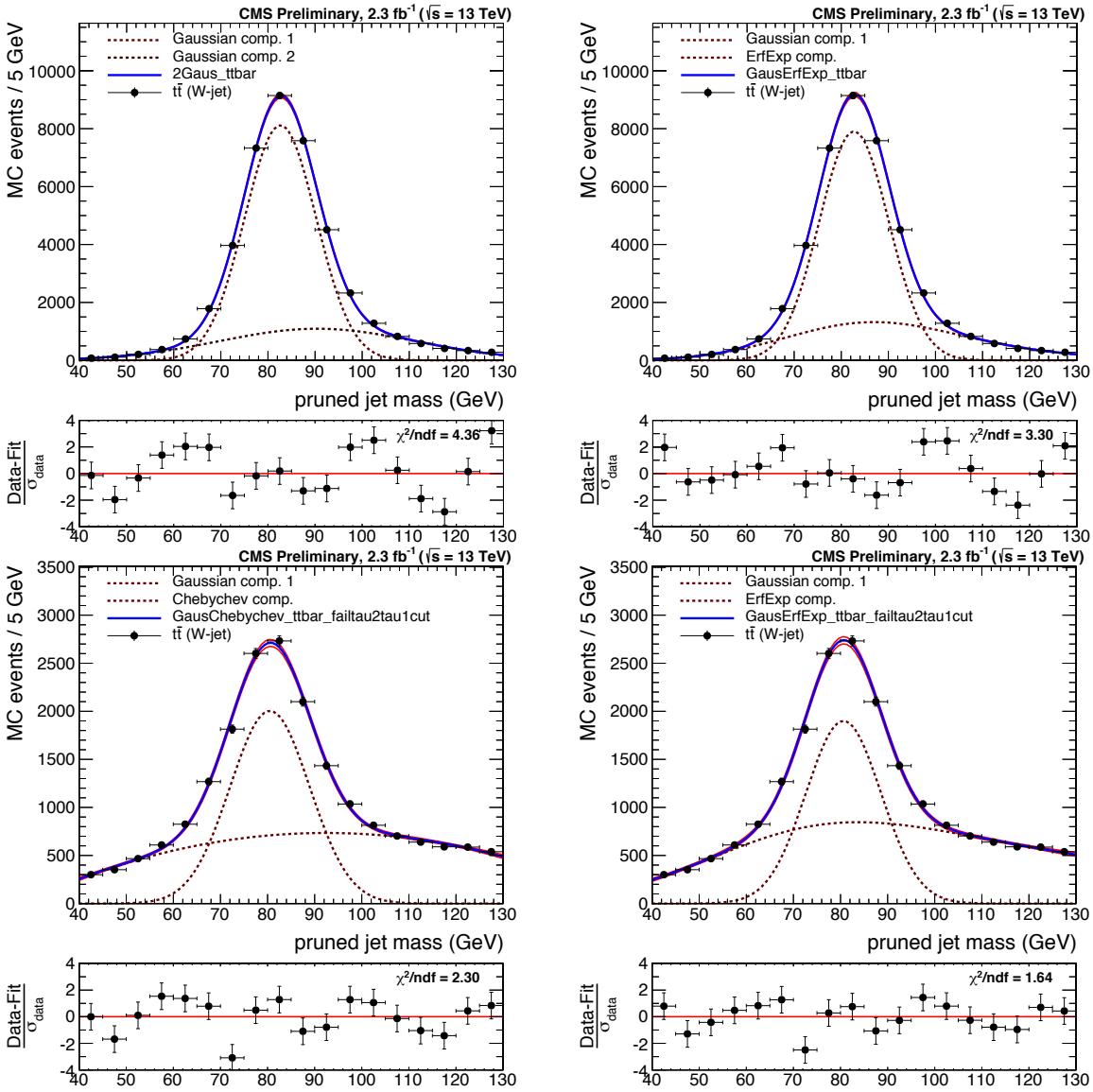


Figure 4.27: Fits obtained in the pass (top) and fail (bottom) regions using two different models: An alternative model with tails (top and bottom, left) where the tail component is contributing to the total W-tagging efficiency. When using the default model (top and bottom, right), only the Gaussian component of the fit contributes to the W-tagging efficiency.

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).

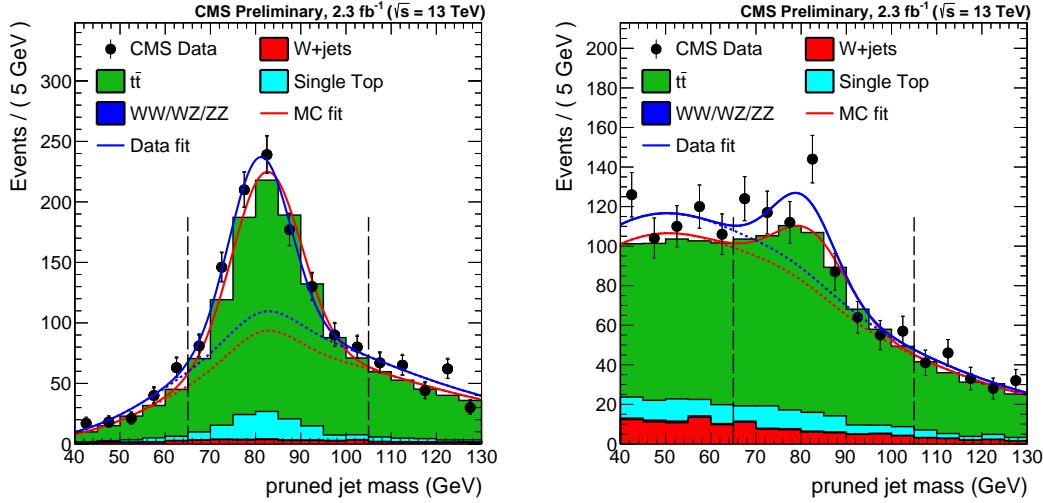


Figure 4.28: Pruned jet mass distribution that pass (left) and fail (right) the $\tau_2/\tau_1 < 0.45$ selection. Results of both the fit to data (blue) and simulation (red) are shown. The background components of the fit are shown as short-dashed lines.

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

1824 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 HPHP or HPLP, 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

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

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

1825 The uncertainties on the scale factors are considered as anti-correlated between the HP and
 1826 the LP categories. The jet mass scale and resolution are used to scale and smear the signal
 1827 Monte Carlo. An uncertainty on the signal yield based on the uncertainty on jet mass scale and
 1828 resolution is also considered by scaling and smearing the jet mass up and down within the quoted

1829 uncertainties and then recomputing the signal efficiency. The results are listed in Table 4.8.

1830 4.1.8 Systematic uncertainties

1831 The uncertainty on the background parametrization is statistical only and is taken as the
1832 covariance matrix of the dijet fit function. As demonstrated in the F-test, we study different
1833 background parameterizations and we have found these to be within the fit uncertainty of the
1834 nominal fit. The remaining uncertainties concern the signal shape and yield and are listed in
1835 Table 4.8. Jet reconstruction uncertainties affect both the signal yield and the signal shape.
1836 These are evaluated by rescaling the jet four-momenta according to uncertainties on the jet
1837 energy scale and resolution and recomputing the signal efficiency. The difference in efficiency
1838 with and without smearing/scaling is taken as systematic uncertainties, as described above. The
1839 jet mass/energy scale and resolution also affect the signal shape, and are added as uncertainties
1840 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 4.8: Summary of systematic uncertainties and the quantities they affect. Migration uncertainties result in events switching between the purity/mass categories and changes the efficiency in each category, but do not affect the total signal efficiency.

¹⁸⁴¹ 4.1.9 Results

¹⁸⁴² The background fits for each analysis category in the data signal region are shown in Figure
¹⁸⁴³ 4.29. Here a background only fit is performed while, as described above, a simultaneous fit is
¹⁸⁴⁴ used for the limit setting procedure. The filled area correspond to the 1 sigma error band of the
¹⁸⁴⁵ background fit, obtained using linear error propagation.

¹⁸⁴⁶ We proceed by setting limits on the cross section of the process $X \rightarrow VV$, using the asymptotic
¹⁸⁴⁷ CL_S method as described in Section ???. The binned likelihood is defined as

$$L = \prod_i \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!} \quad (4.7)$$

¹⁸⁴⁸ with

$$\mu_i = \sigma \cdot N_i(S) + N_i(B) \quad (4.8)$$

¹⁸⁴⁹ Here σ is the signal strength scaling the expected number of signal events in the i -th dijet invariant
¹⁸⁵⁰ mass bin $N_i(S)$, $N_i(B)$ is the expected number of background events in dijet invariant mass bin i
¹⁸⁵¹ and n_i is the observed number of events in the i th dijet invariant mass bin. The background per
¹⁸⁵² bin $N_i(B)$ is estimated from the background component of the best signal+background fit to
¹⁸⁵³ the data points with the signal cross section set to zero. The number of signal events in the i -th
¹⁸⁵⁴ dijet invariant mass bin, $N_i(S)$, is then estimated from the signal templates, where only a dijet
¹⁸⁵⁵ invariant mass in a 20% window around the resonance mass is considered, containing most of the
¹⁸⁵⁶ signal contribution while making sure to keep a good description of the core.

¹⁸⁵⁷ 4.1.10 Limits: All-hadronic analysis

¹⁸⁵⁸ As mentioned in Section 4.1.3, we set limits on three different signal scenarios: $G_{\text{bulk}} \rightarrow WW$,
¹⁸⁵⁹ $G_{\text{bulk}} \rightarrow ZZ$ and $W' \rightarrow WZ$, with a $\tilde{k} = 0.5$ for the G_{bulk} . Figure 4.30 shows the asymptotic
¹⁸⁶⁰ limits at 95 % confidence level on signal cross section as a function of the resonance mass obtained
¹⁸⁶¹ with 2.7 fb^{-1} of 13 TeV CMS data after combining all mass and purity categories (top). The
¹⁸⁶² corresponding p-values are shown in the bottom panel.

¹⁸⁶³ The statistics are too low to exclude the excess around 2 TeV observed in the corresponding
¹⁸⁶⁴ Run 1 analysis and in addition an under-fluctuation in data is present in this region. The largest
¹⁸⁶⁵ excess is observed for a $G_{\text{bulk}} \rightarrow ZZ$ hypothesis at a resonance mass of 2.8-3 TeV, around 2.3σ .
¹⁸⁶⁶ This is driven by the ZZ high-purity category, the category with the lowest statistics, where one
¹⁸⁶⁷ event at 3 TeV yields a local significance of 2.8σ . A 3 parameter fit is the default background fit
¹⁸⁶⁸ function for this category, however, a 2 parameter fit could also be used to describe these data. In

1869 Figure 4.31 we compare the limits and p-values obtained using a 2 parameter and a 3 parameter
 1870 fit to describe the background in this category. The significance at 3 TeV is reduced from 2.8
 1871 to 1.5σ with a 2 parameter fit, reflecting the fact that the fit is poorly constrained in the high
 1872 mass tail due to low statistics. The fit to data using both a 2 and 3 parameter fit in the ZZHP
 1873 category is shown in Figure 4.32 and we in addition see that the 2 parameter fit lies within the
 1874 fit uncertainties of the nominal fit.

1875

1876

1877 The lack of constraint on the fit in the dijet invariant mass tail when statistics are very low,
 1878 is a drawback of a method relying fully on a parametric fit and reduces the analysis sensitivity
 1879 in the high- m_{jj} region. In Search II (Section 4.2) we will keep taking advantage of the dijet fit,
 1880 however, the integrated luminosity is ~ 15 times higher, resulting in more datapoints in the
 1881 m_{jj} tail which further constrains the fit. In Search III (Section 4.3), we will explore alternate
 1882 methods which allow more control over the background shape across the full mass spectrum.

1883 4.1.11 Limits: Semi-leptonic and all-hadronic combination

1884 To maximize the search sensitivity, we combine the results obtained above with those of the
 1885 corresponding semi-leptonic analysis. We assume the uncertainties on luminosity, V-tagging
 1886 efficiency, jet mass scale and resolution to be fully correlated.

1887 The obtained exclusion limits are shown in Figure 4.33 shows the resulting expected and
 1888 observed exclusion limits. As before, we consider a scenario where only either a W' or Z' resonance
 1889 is expected, called the singlet hypothesis (upper two plots). In addition, we set limits on the
 1890 triplet hypothesis, assuming the W' and Z' bosons to be degenerate in mass (bottom left plot).
 1891 Due to larger branching fraction, the all-hadronic analysis sets stronger upper limits than the
 1892 semi-leptonic analysis above 1.7 TeV for Z' and > 1.3 TeV for W' ($\mathcal{B}(WW \rightarrow q\bar{q}q\bar{q}) = 44\%$,
 1893 $\mathcal{B}(WW \rightarrow \ell\nu q\bar{q}) = 31\%$, $\mathcal{B}(WZ \rightarrow q\bar{q}q\bar{q}) = 46\%$, and $\mathcal{B}(WZ \rightarrow \ell\nu q\bar{q}) = 16\%$). The analysis
 1894 sensitivity for G_{bulk} is too weak to set limits, but cross sections between 3–1200 fb are excluded.
 1895 For the HVT model A and B, W' is excluded below < 2.0 and 2.2 TeV, respectively. Z' resonances
 1896 are excluded below < 1.6 (1.7) TeV for HVT model B(A). If assuming a HVT Model A(B) triplet
 1897 hypothesis, resonances below $< 2.3 (< 2.4)$ TeV are excluded.

1898 The combined results would therefore just exclude a W' with a mass around 2 TeV, the
 1899 favored candidate to explain the 8 TeV diboson excess. However, Bulk Graviton signals were
 1900 still far from excluded and, with the expected ten times increase in luminosity in 2016, we were
 1901 excited to keep on searching.

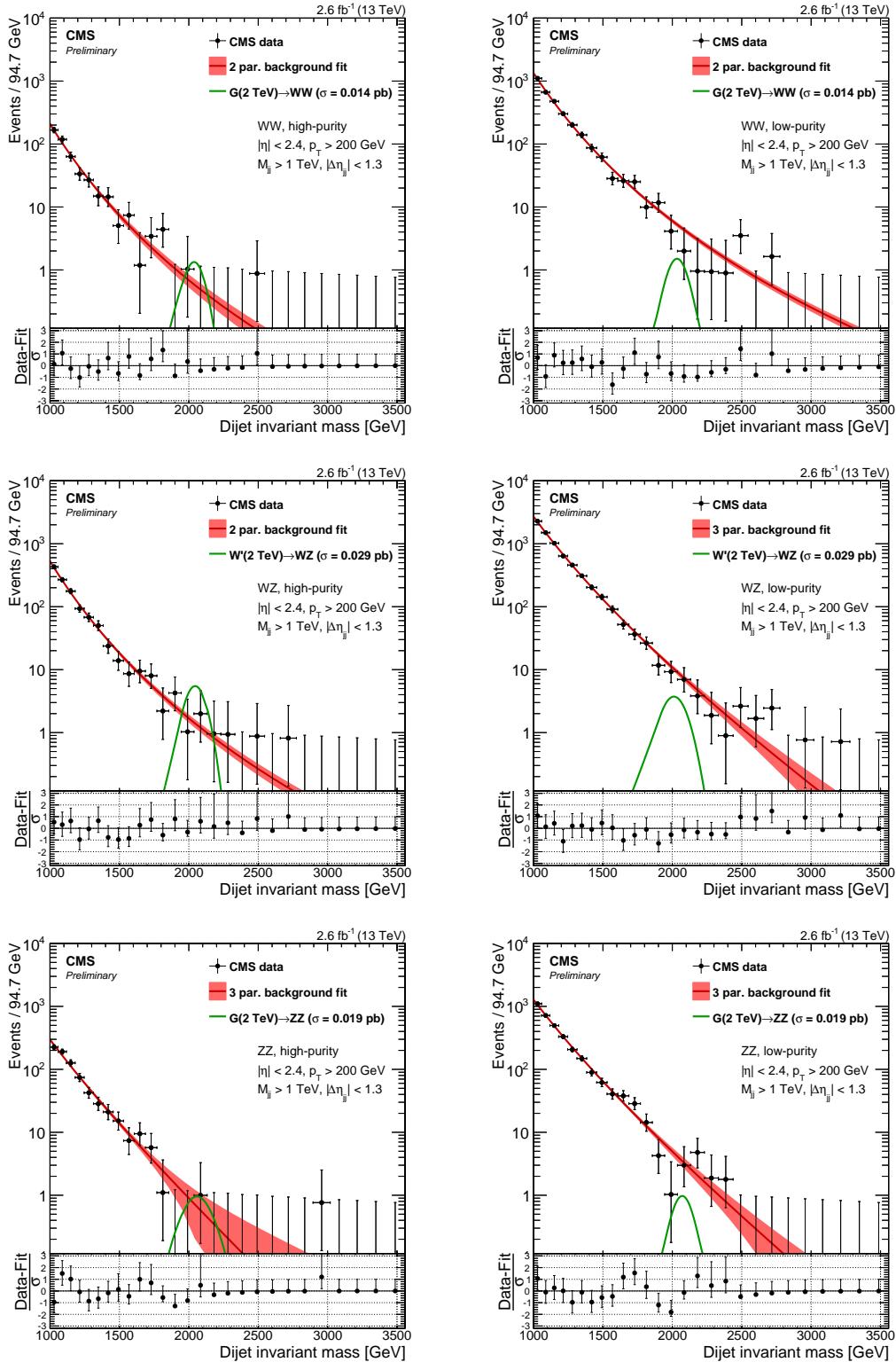


Figure 4.29: Fit to data in the signal region using the background fit only for the different mass and purity categories. The filled red area correspond to the 1 sigma statistical error of the fit.

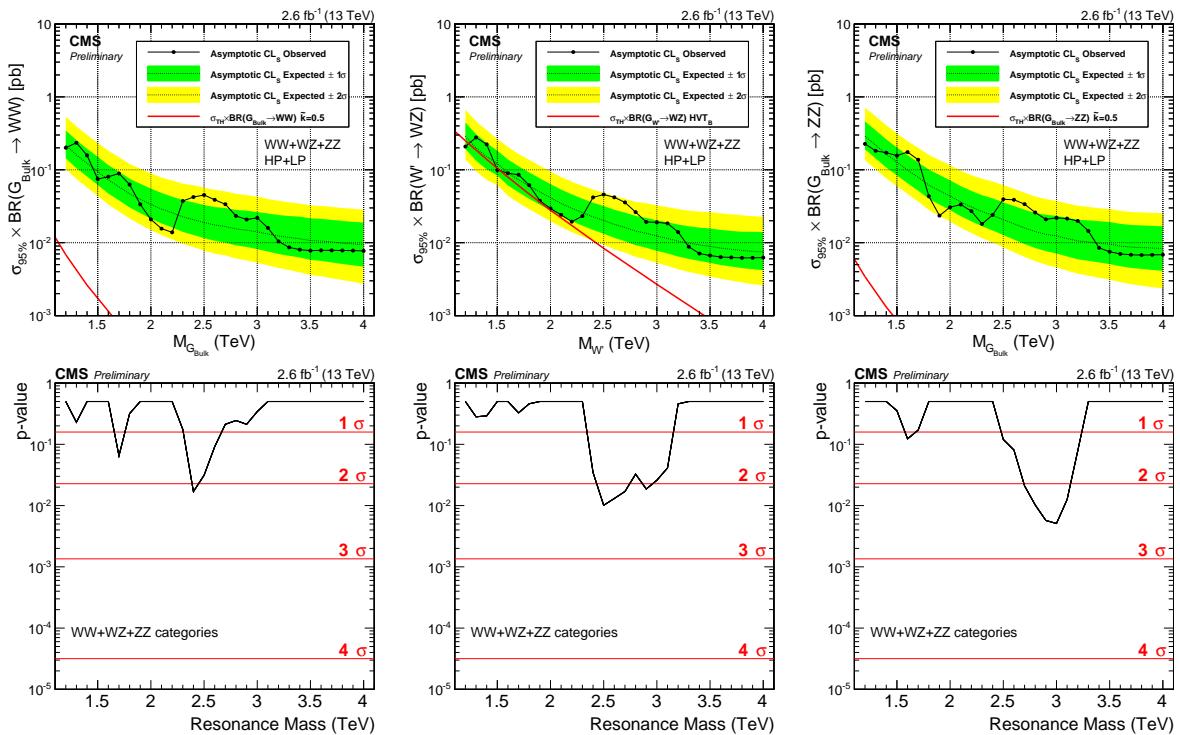


Figure 4.30: Expected and observed limits with corresponding p-values obtained using 2.6 fb^{-1} of CMS data after combining all mass and purity categories. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle) and $G \rightarrow ZZ$ (right) signal.

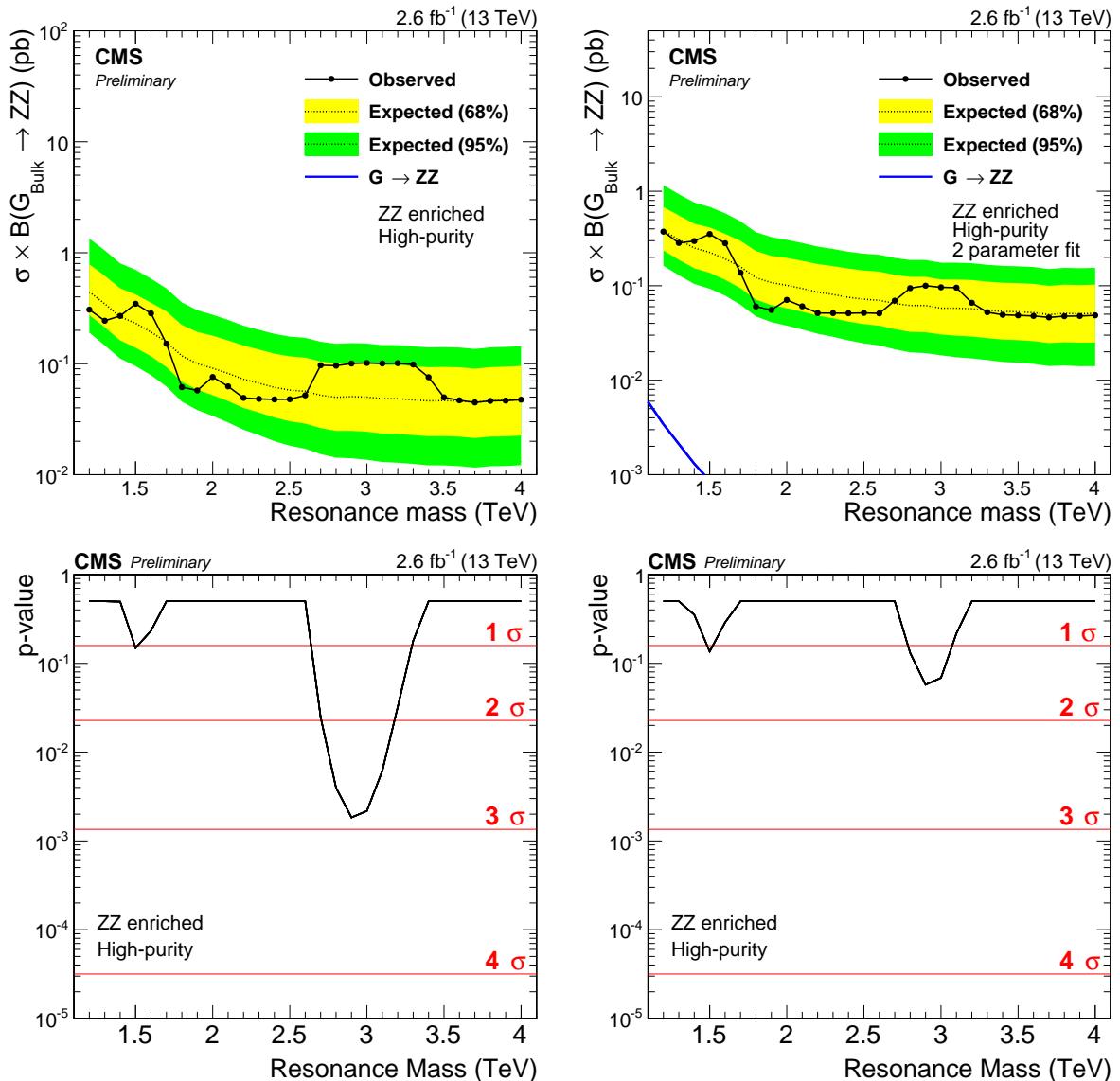


Figure 4.31: Expected/observed limits and corresponding p-values obtained in the ZZHP category using a 3 (left) and two (right) parameter fit to describe the background. The significance at 3 TeV is reduced from 2.8σ to 1.5σ .

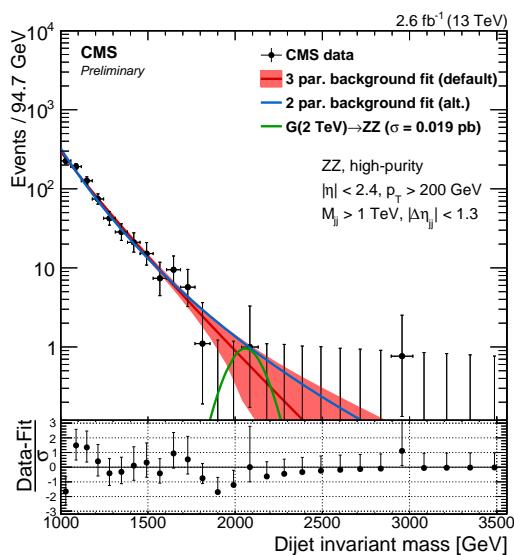


Figure 4.32: Background fit to data in the ZZHP category using the default 3 (red) and an alternate 2 (blue) parameter fit to describe the background.

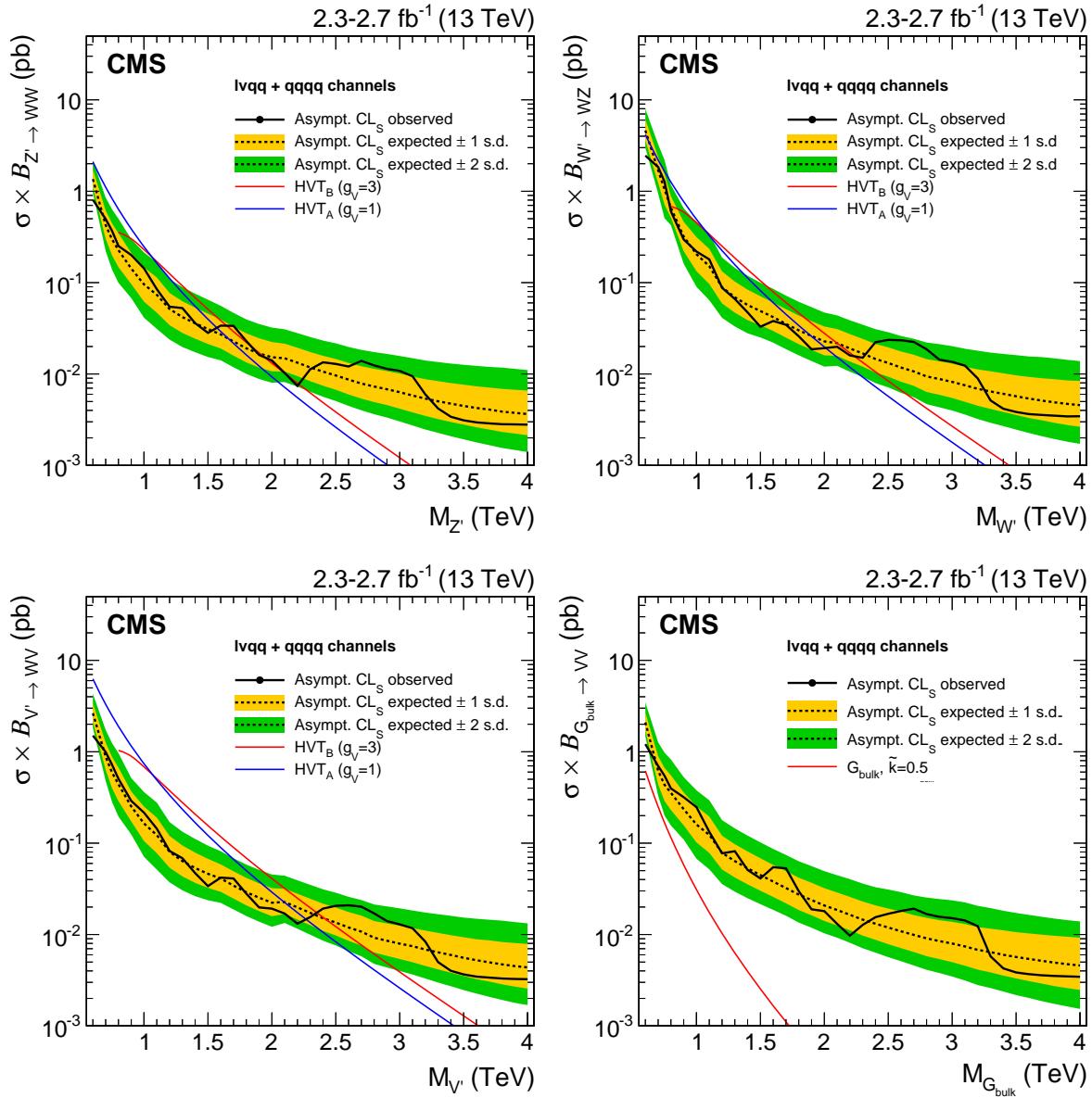


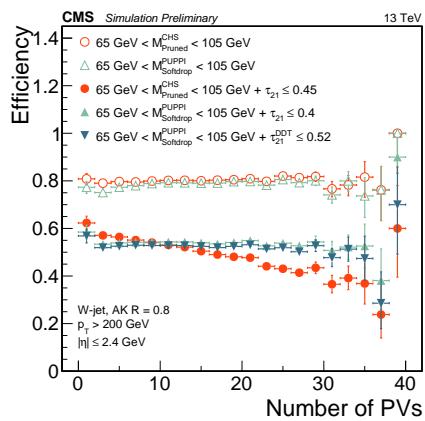
Figure 4.33: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of a narrow-width resonance decaying to a pair of vector bosons for different signal hypotheses. In the upper plots, limits are set in the context of a spin-1 neutral Z' (left) and charged W' (right) resonances, 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'). 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.

1902 4.2 Search II: A new pileup resistant, perturbative safe tagger

1903 With the first 13 TeV diboson resonance search published, we could conclude that more data
 1904 would be needed in order to fully exclude the observed Run 1 excess. Luckily, 2016 was right
 1905 around the corner and, with the LHC planning to reduce β^* from 80 cm to 40, the machine
 1906 was expected to deliver an instantaneous luminosity three times that of the 2015 peak luminosity.
 1907 Higher instantaneous luminosity, however, meant double the pileup.

1908 We knew that a novel pileup subtraction algorithm had been developed, which provided far better
 1909 pileup and underlying event rejection than the current default (CHS). We also knew that there
 1910 had been made progress on the theory side in the development of a groomer which was insensitive
 1911 to the soft divergences of QCD and allowed to accomplish jet grooming in a theoretically calculable
 1912 way, SoftDrop (mMDT). With more time at hand than in 2015, I therefore decided to pursue
 1913 a novel W-tagger for this second search. This included work like optimization, development of
 1914 dedicated jet mass corrections (in use today and recommended by the jet physics object group) as
 1915 well as validation of the new tagger. The tagger, together with the mass corrections, afterwards
 1916 became the default W-tagging algorithm in CMS.

1918 Search II became the first published analysis to use the novel PUPPI+softdrop algorithm, now
 1919 default for W-tagging in CMS. Through this search, the tagger was optimized, commissioned and
 1920 validated, making it available for several analysis to come. In addition, the search was extended
 1921 to setting limits on three additional signal hypothesis. Two of these were in a final state never
 1922 before explored at 13 TeV, the $q^* \rightarrow qV$ single V-tag analyses. Published with $35.9(12.9)$ fb^{-1} of
 1923 2016 data.



1925 4.2.1 Towards robust boosted jet tagging

1926 When we first studied W-tagging at 13 TeV in context with the analysis of the 2015 dataset,
1927 Section 4.2.5, two interesting correlations were observed:

- 1928 1) A strong dependence of the AK8 CHS softdrop ($\beta = 0$) jet mass on jet p_T and
- 1929 2) a strong dependence of the AK8 CHS τ_{21} cut efficiency on pileup.

1930 The reason we studied the softdrop algorithm as an alternative to pruning in 2015 was, besides
1931 the possibility it would result in a higher signal efficiency, that we knew it had certain favorable
1932 qualities compared to other groomers: Softdrop removes all sensitivity to the soft divergences of
1933 QCD, by removing all soft emission, more specifically the non-global logarithmic terms (NGLs)
1934 in the jet mass [42]. These arise from constellations where, for instance, a soft gluon is radiated
 into the jet cone, as illustrated in Figure 4.34. The consequence of this is that you can calculate

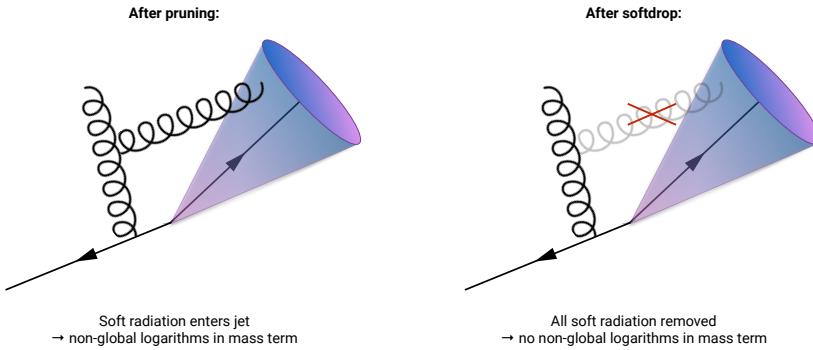


Figure 4.34: The pruning algorithm does not remove all soft emission and therefore has non-global logarithmic terms in the jet mass. Softdrop ($\beta = 0$) completely removes soft emissions and is therefore free of non-global logarithms.

1935

- 1936 the softdrop jet mass to way higher precision than what is possible for other grooming algorithms
- 1937 or for the plain jet mass (NGLs are the main reason a full resummation of the plain jet mass
- 1938 beyond NLL (considering e.g multiple-emission effects) accuracy does not exist). Despite this not
- 1939 being a precision measurement analysis, we had theoretically well-motivated reasons for wanting
- 1940 the baseline CMS V-tagger to be softdrop-based. However, despite being less sensitive to soft
- 1941 radiation for QCD jets, signal jets groomed with softdrop were found to be far more sensitive to
- 1942 the underlying event than pruned jets [69]. Figure 4.35 shows the signal efficiency for pruning
- 1943 (left) and softdrop (right) as a function of jet transverse momenta when including FSR only,
- 1944 FSR+ISR, hadronization and hadronization + underlying event. On parton level, as well as after
- 1945 hadronization, the two algorithms perform very similar as a function of p_T . However, once UE

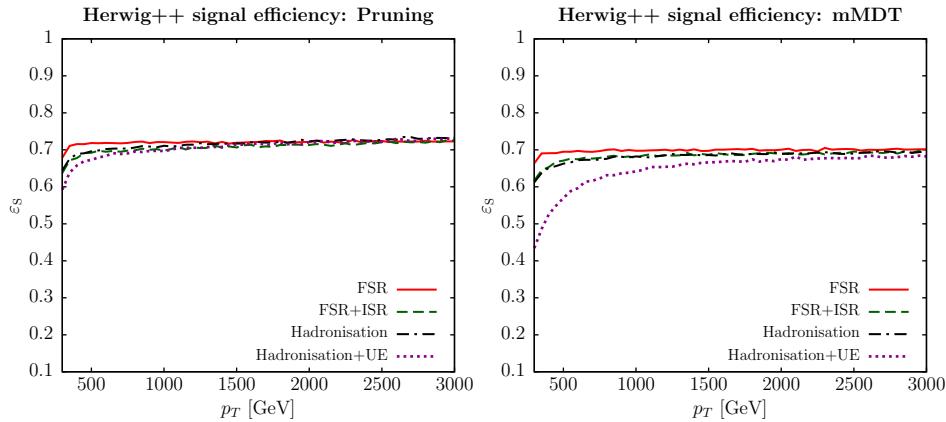


Figure 4.35: The signal efficiency for pruning (left) and softdrop (right) as a function of jet p_T when adding FSR, ISR, hadronization and UE. The UE has a severe impact on the softdrop efficiency for signal jets [69].

contamination is added, the softdrop tagging efficiency is severely affected. This can be explained by the larger effective radius considered by the softdrop algorithm ($\propto m_V/p_T \sqrt{z_{cut}(1-z_{cut})}$) in comparison to pruning ($\propto m_V/p_T$). This observation corresponds very well with the shift in jet mass we observed for softdrop as a function of p_T in Section 4.2.5: As the jet p_T decreases the softdrop effective radius increases and the jet mass mean shifts to higher values, due to absorbing more background radiation. If softdrop would be our new default tagger, a better rejection of pileup and UE contamination would be needed. In parallel to the ongoing theoretical work on groomers, a novel pileup removal algorithm had been proposed: Pileup per particle identification (PUPPI) [34]. Described in detail in Section 3.3.2, PUPPI considers not only charged pileup but rather reweights each particle in the jet with its probability of arising from pileup. PUPPI had proven it self far superior to the current CHS algorithm in terms of jet observables for large radius jets, and therefore seemed like the obvious choice to address both issues listed above: The sensitivity of softdrop regarding UE contamination 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 changes and inclusions in 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.

4.2.2 Analysis strategy

The analysis strategy for this search is conceptually the same as for Search I. In addition, we'll take advantage of the n-subjettiness categorization and do an additional analysis in parallel: A

1965 search for excited quark resonances q^* [70, 71] decaying to qW or qZ . We call this the single V-tag
 1966 analysis, and the analysis selection only differs in that one jet is not required to pass the V-tag
 1967 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 4.36. In addition,

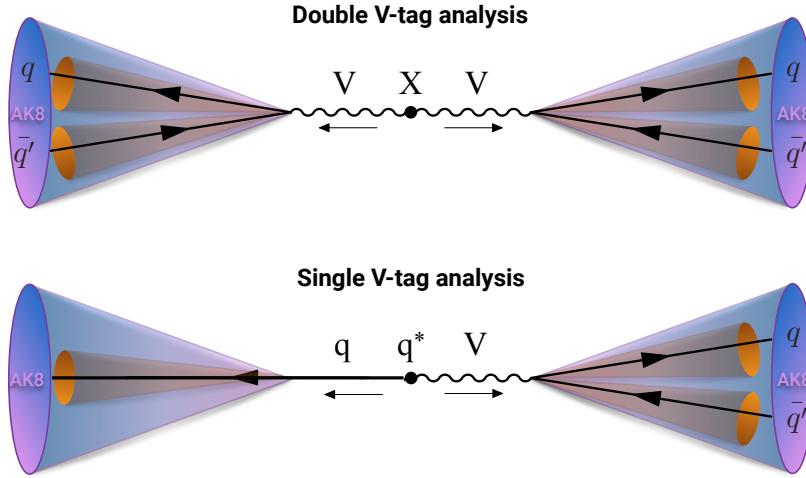


Figure 4.36: The double (top) and single (bottom) W/Z-tag analysis.

1968 limits are set on a $Z' \rightarrow WW$ signal hypothesis in the double V-tag analysis, another 13 TeV first.
 1969 This analysis was published in two steps: An early Physics Analysis Summary (PAS) based
 1970 on 12.9 fb^{-1} of 2016 data [68], describing the new PUPPI+softdrop based V-tagger as well as
 1971 the single V-tag analysis, and a second analysis topping up with the full 2016 data [72]. The
 1972 commissioning of the new W Z-tagger has also been documented in a jet performance Physics
 1973 Analysis Summary [73]. As the new V-tagger was developed and commissioned in the context
 1974 of the early analysis, which was also where the single V-tag analysis was first published with 13
 1975 TeV data, the main emphasis will be on the work presented in CMS-PAS-B2G-16-021 [68]. The
 1976 second part of the results chapter, Section 4.2.11, includes the results obtained using the full 2016
 1977 dataset of 35.9 fb^{-1} .

1979 **4.2.3 Data and simulated samples**

1980 As mentioned above, the analysis of the 2016 dataset was done in two steps: One analysis based
 1981 on 12.9 fb^{-1} of early 2016 data, describing the new W-tagger and single V-tag category, and a
 1982 second paper topping up with the full 2016 dataset, corresponding to 35.9 fb^{-1} .

1983 The G_{bulk} and HVT signal samples are modeled in precisely the same way as in 2015. For the
 1984 single V-tag q^* samples, we simulate unpolarized boson with a compositeness scale Λ set equal to

1985 the resonance mass. These are generated to leading order using PYTHIA version 8.212 [61].

1986 The background Standard Model processes; QCD, W+jets and Z+jets are all simulated to
 1987 leading order. V+jets is simulated with MADGRAPH5_AMC@NLO [60, 74], while three different
 1988 combinations of matrix element and shower generators is used for QCD as these predictions are
 1989 known to differ: PYTHIA only, the leading order mode of MADGRAPH5_AMC@NLO matched with
 1990 PYTHIA, and HERWIG++ 2.7.1 [49] with tune CUETHS1 [56].

1991 4.2.4 Event selection

1992 Triggering

1993 The triggers used in this analysis are the same ones as in 2015 (see Section 4.1.4), however, due
 1994 to the new single V-tag analysis, the trigger turn-ons have this time been re-evaluated separately
 1995 requiring either one or two jets to have an offline softdrop jet mass above 65 GeV.

1996 Figure 4.37 shows the trigger turn-on curves as a function of dijet invariant mass for jets
 1997 passing one of the three inclusive triggers only, one of the grooming triggers only and when
 1998 combining all of them. The turn-on curves are shown for all jet pairs passing loose selections as
 1999 described in Section 4.1.4. Zero, one or two of the two jets is further required to have a softdrop
 2000 mass larger than 65 GeV.

2001 Including grooming triggers lowers the 99% trigger efficiency threshold by around 50(80) GeV
 2002 in the single (double) tag category once substructure is requested on the analysis level. Using the
 2003 or of all triggers, we are safely on the trigger plateau for dijet invariant masses above 955(986)
 2004 GeV in the double (single) tag category, setting the analysis threshold at a dijet invariant mass
 2005 of 955 GeV for the double tag analysis and 990 GeV for the single tag analysis. For controlplots,
 2006 where no groomed mass window is applied, a trigger threshold of 1020 GeV is used.

2007 Trigger efficiencies as a function of the offline softdrop-jet mass for the
 2008 HLT_AK8PFJet360_TrimMass30 trigger are shown in Figure 4.38. Here the jet transverse momen-
 2009 tum of one of the jets is required to be at least 600 GeV and no other mass cut is applied. This
 2010 trigger requires one jet to have a trimmed mass above 30 GeV at HLT level and reaches the
 2011 trigger plateau for groomed-jet masses around 50 GeV. As reference trigger, the prescaled trigger
 2012 HLT_PFJet320 is used.

2013 Preselection

2014 The same preselections as in Search I, described in , have been applied: We require two AK R=0.8
 2015 jets with CHS applied pre-clustering, required to pass the tight jet ID requirement, $p_T > 200$ GeV

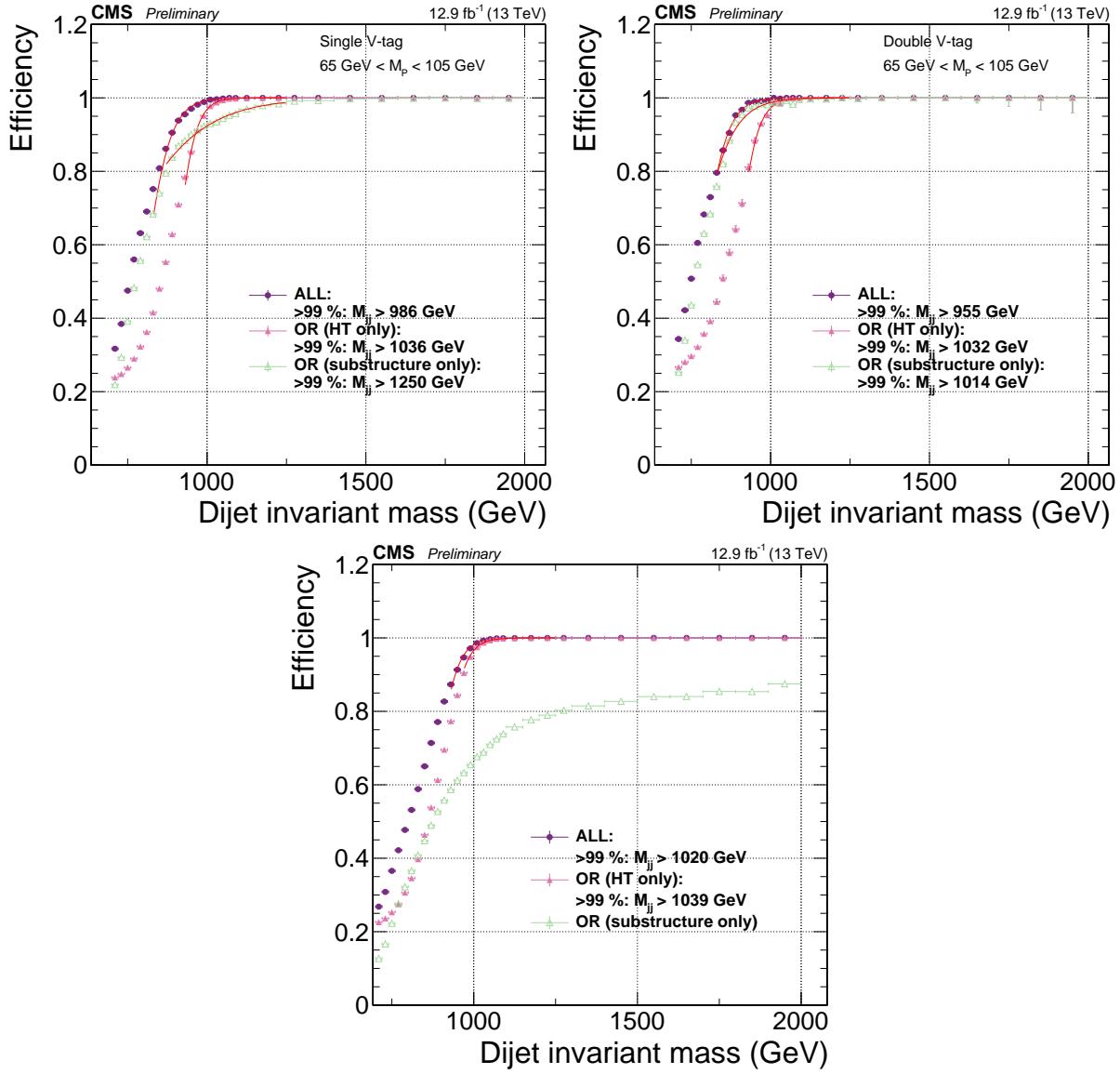


Figure 4.37: Comparison of trigger efficiencies for jets passing one of the HT-triggers only (pink), for jets passing one of the grooming-triggers only (green) and for jets passing one of the HT-triggers or one of the grooming triggers (purple). Here as a function of dijet invariant mass for all jet pairs passing loose selections and where one jet has a softdrop mass larger than 65 GeV (top left), both jets have a softdrop mass larger than 65 GeV (top right) and where no mass cut is applied (bottom).

²⁰¹⁶ and $|\eta| < 2.5$. The same QCD t-channel suppressing cut of $|\Delta\eta| < 1.3$ is required together with
²⁰¹⁷ the following trigger thresholds on the dijet invariant mass: $m_{jj} > 955 \text{ GeV}$ for the double V-tag
²⁰¹⁸ and 990 GeV for the single V-tag analysis. The jet p_T (top left), η (top right), $\Delta\eta_{jj}$ and dijet

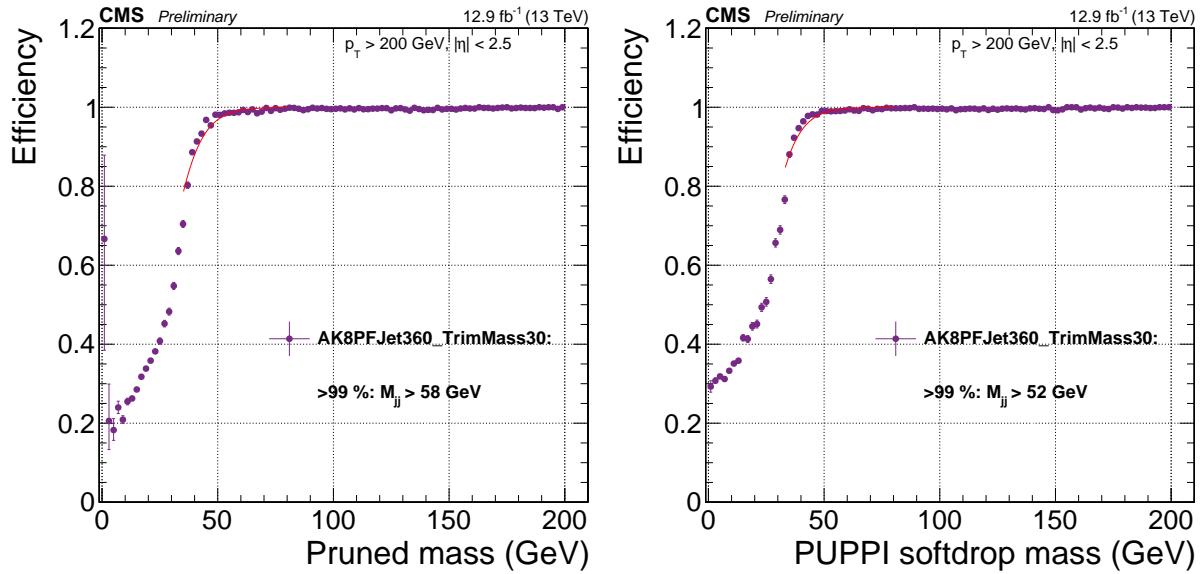


Figure 4.38: Efficiency for the HLT_AK8PFJet360_TrimMass30 trigger as a function of pruned-jet (left) and softdrop-jet (right) mass for jets with $p_T > 600 \text{ GeV}$.

2019 invariant mass (bottom left) for the two leading jets in the event after loose preselections are
2020 applied is shown in Figure 4.39. A large difference in slope in the jet p_T and dijet invariant mass
2021 spectrum depending on the QCD matrix element or shower generator is observed. Pure PYTHIA
2022 QCD MC describes the data best, while HERWIG++ and MADGRAPH5_AMC@NLO+PYTHIA tend
2023 to under- or over-estimate the number of high p_T/m_{jj} jets, respectively. Pure PYTHIA QCD MC
2024 is therefore used for all background checks in this analysis.

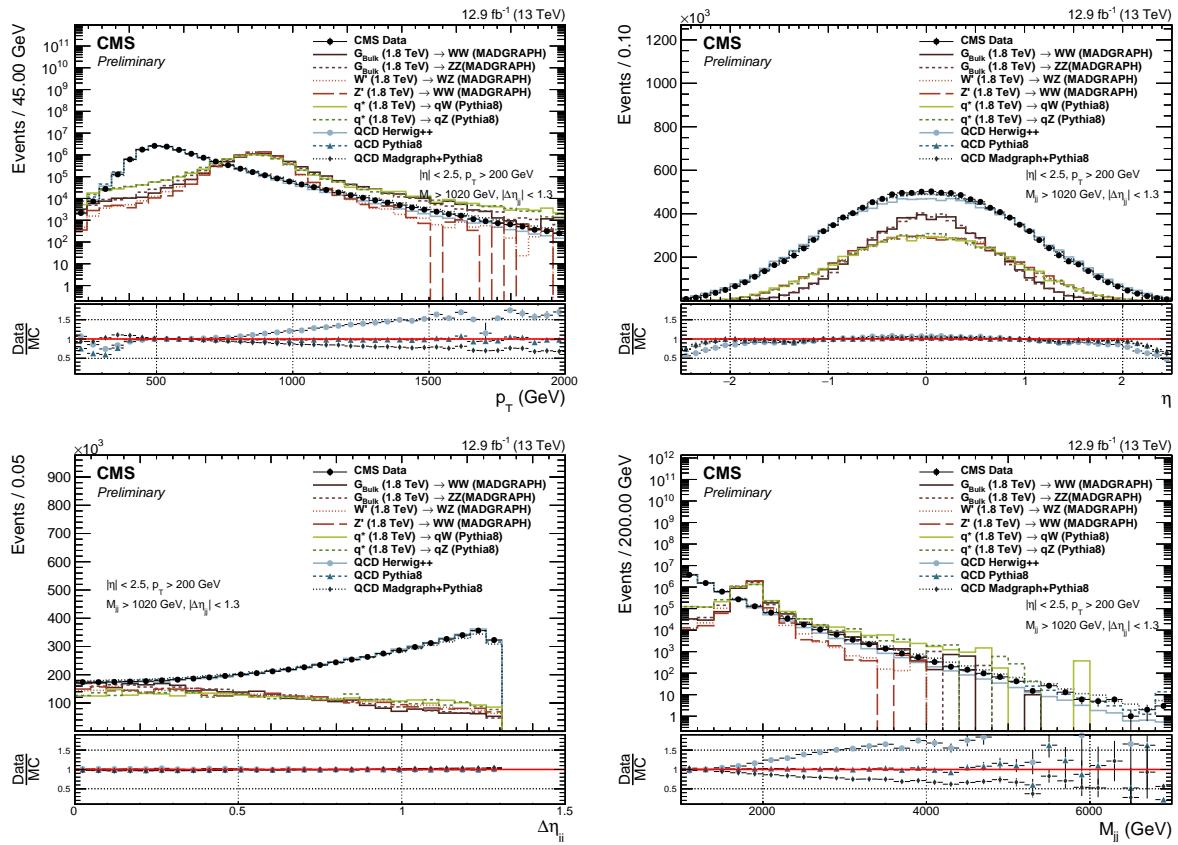


Figure 4.39: Jet p_T (top left), η (top right), $\Delta\eta_{jj}$ and dijet invariant mass (bottom left) for the two leading jets in the event after loose preselections are applied. The signal is scaled by an arbitrary number.

4.2.5 Developing a new W-tagger

As mentioned in the introduction to this chapter, early 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 CHS+Softdrop, as hoped. Figure 4.40 shows the PUPPI softdrop mass for W-jets from a 1 TeV ($p_T \sim 500$ GeV) and 4 TeV ($p_T \sim 2$ TeV) resonance, exhibiting the desired reduced p_T dependence in jet mass scale. However, when apply-

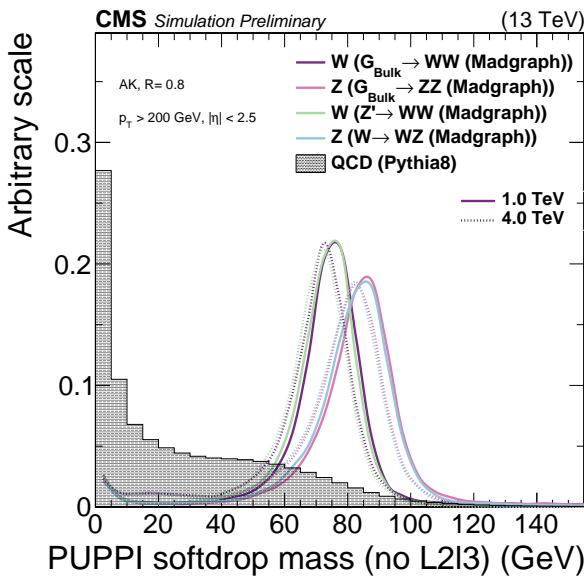


Figure 4.40: The PUPPI softdrop jet mass distribution with no jet energy corrections applied

ing centrally provided L2 and L3 jet energy corrections (see Section 3.4.3) to the jet groomed mass, as is recommended, a strong p_T dependence is re-introduced. This effect is not present for the pruned jet mass. Figure 4.41 show the softdrop (top left) and pruned (top right) jet mass distribution with recommended L2L3 corrections applied. Here, the PUPPI+softdrop jet mass shift is significantly increased with respect to what was observed for the uncorrected mass, while CHS+pruned jet mass is stable. This points to the PUPPI jet energy corrections not being optimal for scalar jet mass variables, while they may be good for correcting jet 4-vectors. The jet energy corrections derived for CHS and PUPPI jets as a function of jet p_T is shown in the bottom plot in Figure 4.41 . A significant slope in JEC as a function of p_T is measured for PUPPI, while

not present for CHS.

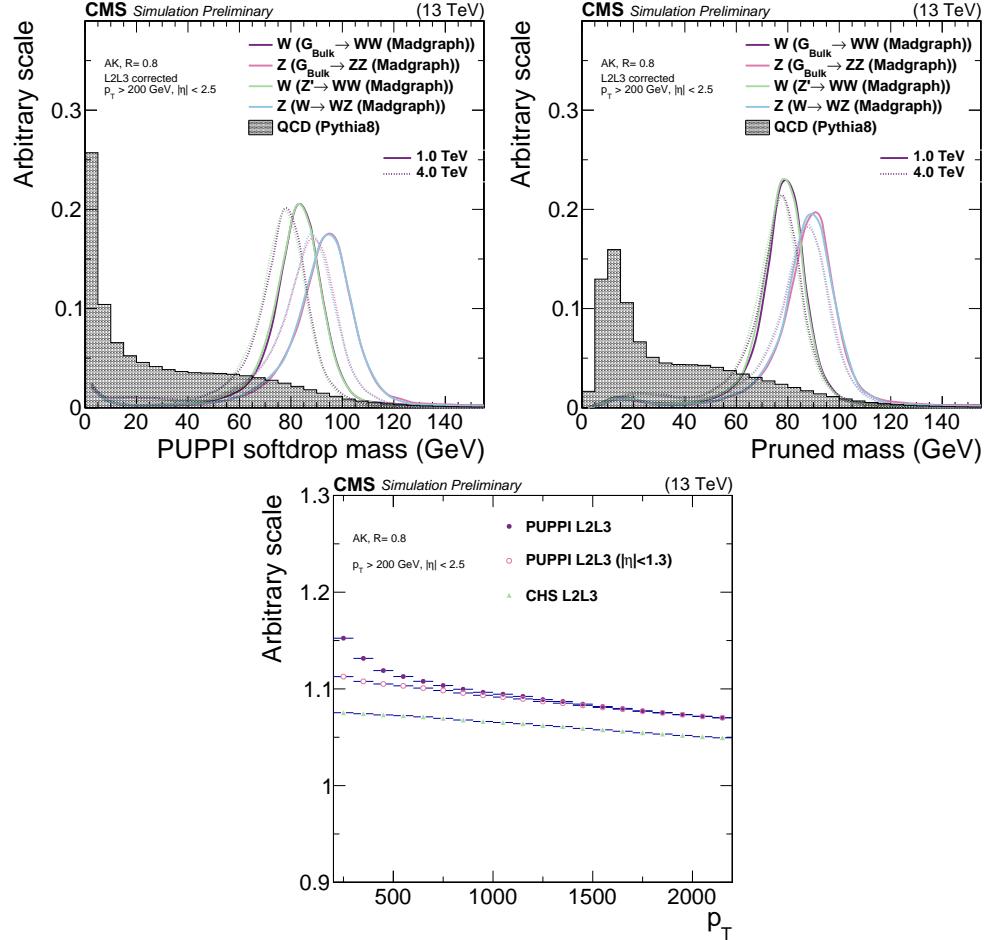


Figure 4.41: Top: PUPPI softdrop mass distribution (top left) and pruned jet mass distribution (top right) with L2 and L3 corrections applied. Bottom: The projection of CHS and PUPPI jet energy corrections versus jet p_T .

2044

2045 Dedicated PUPPI softdrop mass corrections

In order to minimize p_T dependence in the PUPPI softdrop jet mass, all jet energy corrections to the softdrop jet mass are removed. However, this still leaves a residual p_T dependence and, in addition, the uncorrected mass does not peak at the correct W-mass of 80.4 GeV. Figure 4.42 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 significantly lower than the W-mass. In order to use PUPPI+softdrop for W-tagging, we

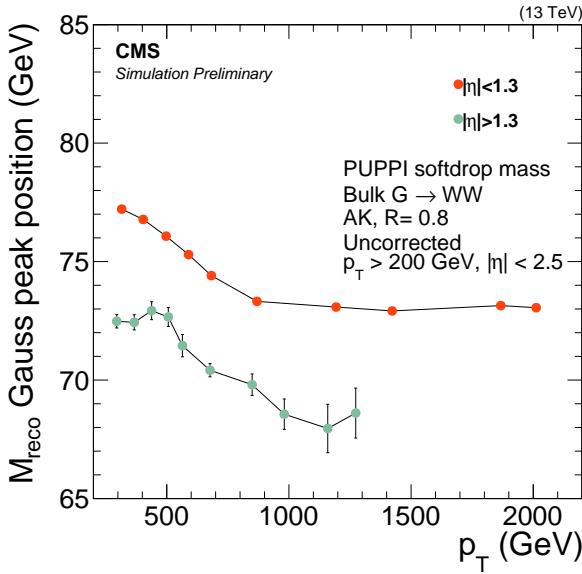


Figure 4.42: The mean of a Gaussian fit to the W-jet PUPPI softdrop mass peak as a function of jet p_T in two different η bins (smaller or greater than $|\eta| = 1.3$). No corrections have been applied to the softdrop mass.

therefore derive dedicated jet mass corrections to compensate for two factors: A generator level p_T -dependence, as first observed in , and a reconstruction level p_T - and η -dependence, most likely caused by UE effects and the growing effective softdrop radius at low jet p_T . Figure 4.43 shows the mean of the generated softdrop mass (left) and the normalized difference in reconstructed and generated softdrop 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 $M_{\text{PDG}}/M_{\text{GEN}}$ 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 as shown in the left plot in Figure 4.43. 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 4.43 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

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

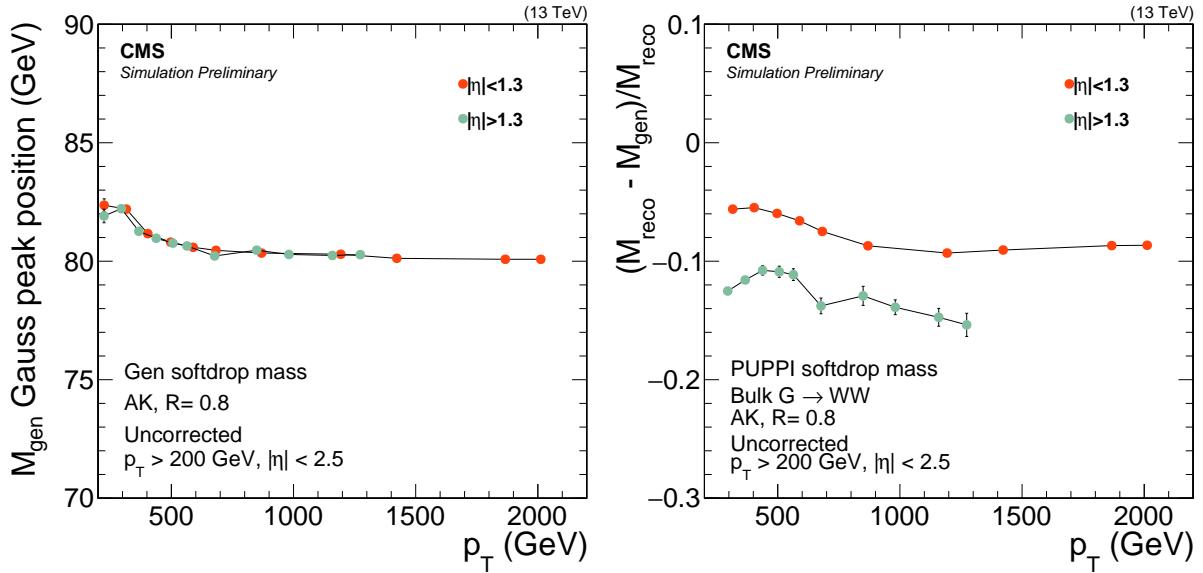


Figure 4.43: The mean of the fitted generator level W-jet softdrop mass distribution as a function of jet p_T (left) and the normalized difference in reconstructed and generated softdrop mass (right).

2046 The distribution and corresponding fits for the two weights is shown in Figure 4.44 for the “gen correction” (left) and “reco correction” (right). The two corrections are then applied to the

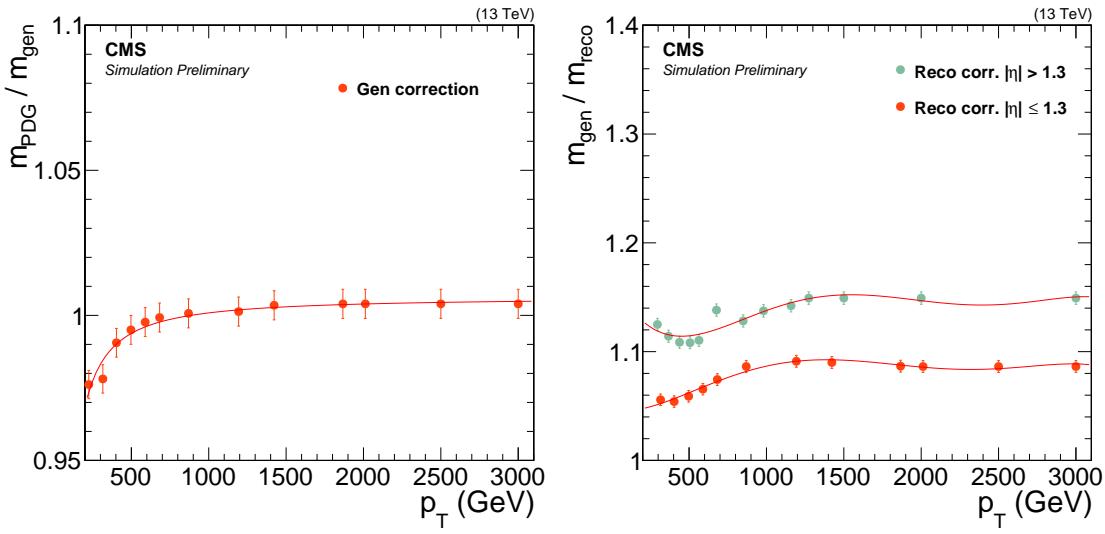


Figure 4.44: Fit to $m_{\text{PDG}}/m_{\text{GEN}}$ as a function of jet p_T (left), where $m_{\text{PDG}} = 80.4 \text{ GeV}$ and m_{GEN} is the fitted mean of the generator level mass and a fit to $(M_{\text{RECO}} - M_{\text{GEN}})/M_{\text{RECO}}$ (right), where M_{RECO} is the reconstructed softdrop mass, as a function of jet p_T in two η bins.

2048 uncorrected PUPPI softdrop mass both in data and in MC as

$$M_{SD} = M_{SD, \text{uncorr}} \times w_{\text{GEN}} \times w_{\text{RECO}} \quad (4.9)$$

2049 where w_{GEN} and w_{RECO} correspond to the gen and reco corrections respectively and $M_{SD, \text{uncorr}}$
2050 is the uncorrected PUPPI softdrop mass.

2051 Finally, a closure test is performed in order to check that the corrected PUPPI+softdrop
2052 W-jet mass peaks at 80.4 GeV and is stable with p_T and η . The fitted mean of the corrected
2053 PUPPI softdrop mass peak as a function of jet p_T in two different η bins is shown in Figure 4.45.
2054 Good closure is observed, with the corrected mass peaking around 80 GeV independent of the jet
2055 p_T and η . The PUPPI softdrop jet mass peak for W/Z-jets from different signal samples after jet
2056 mass corrections have been applied is shown in Figure 4.46, for resonances with a mass of 1 and 4
TeV. The corrections applied to Z-jets yield a mass stable with p_T , peaking around the Z mass.

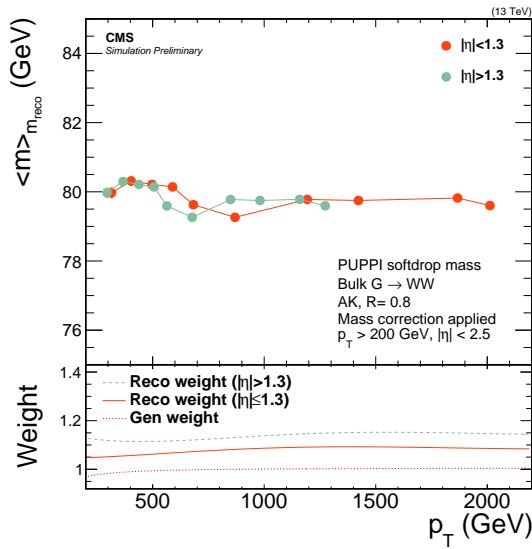


Figure 4.45: The mean of the fitted W-jet corrected PUPPI softdrop mass peak as a function of jet p_T in two different η bins.

2057

2058 W-tagging performance

2059 The new PUPPI+softdrop based W/Z-tagger uses a mass window of $65 \text{ GeV} < m_{SD} < 105 \text{ GeV}$
2060 in combination with a cut of PUPPI $\tau_{21} < 0.4$. We compare its performance to that of the
2061 CHS+pruning based tagger used in Search I as well as to that of a “DDT-transformed” τ_{21} based

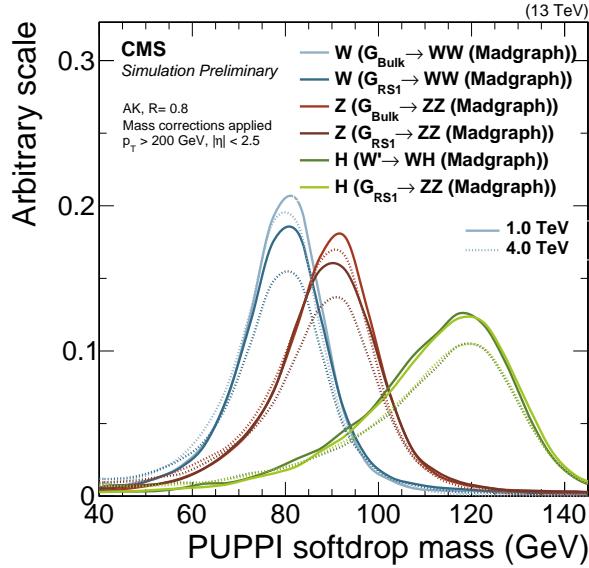


Figure 4.46: The W/Z/H-jet corrected PUPPI softdrop mass peak for jets from different signal samples with masses of 1 and 4 TeV.

2062 tagger [75]. The τ_{21}^{DDT} variable is a linear transformation of τ_{21} given as

$$\tau_{21}^{DDT} = \tau_{21} + M \times \log \left(\frac{m^2}{p_T \times 1 \text{ GeV}} \right) \quad (4.10)$$

2063 where $M = -0.063$ is obtained from a fit of τ_{21} against the variable $\rho' = \log(m^2/p_T/\mu)$, where
 2064 $\mu = 1 \text{ GeV}$. The purpose of this is to de-correlate τ_{21} from the softdrop mass and p_T , yielding
 2065 a mass and dijet invariant mass spectrum minimally sculpted by a cut on the τ_{21}^{DDT} tagging
 2066 variable. This is tagger that will be further explored and explained in detail in the context of
 2067 Search III, Section 4.3.6.

2068 The background rejection efficiency for QCD light flavor jets as a function of W-jet signal
 2069 efficiency is shown in Figure 4.47. The efficiency is measured requiring a fixed jet mass window of
 2070 65–105 GeV, while scanning the cut on τ_2/τ_1 . The general performance of each tagger is very
 2071 similar, with the PUPPI+softdrop based taggers displaying a slightly higher signal efficiency for a
 2072 given mistag rate at high p_T and CHS+pruning slightly better at low p_T . To better understand
 2073 the difference between each tagger, we look at the tagging performance as a function of jet p_T as
 2074 well as pileup, shown in Figure 4.48 and 4.49.

2075 Starting with the tagger p_T -dependence in Figure 4.48, we observe that the signal efficiency of a
 2076 PUPPI+softdrop of CHS+pruned jet mass cut is flat as a function of p_T , at around 80%. The QCD
 2077 mistagging rate drops for both groomers, with a 1–3% lower mistag rate using PUPPI+softdrop

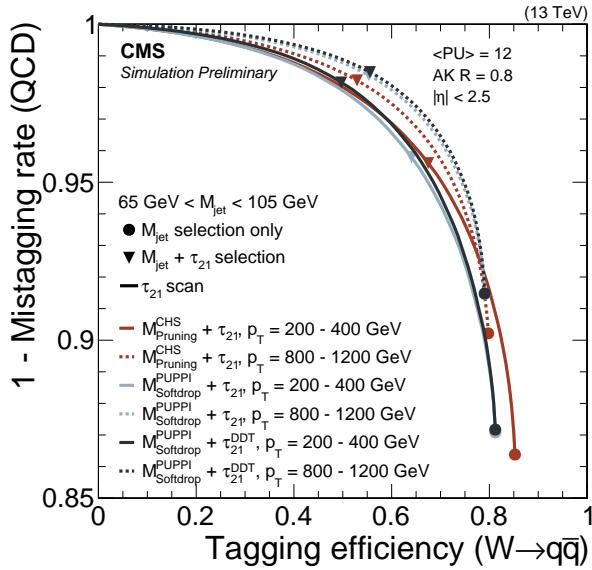


Figure 4.47: The background rejection efficiency for QCD light flavor jets as a function of W-jet signal efficiency. A cut on CHS pruned or PUPPI softdrop jet mass of $65 < m_{\text{jet}} < 105$ GeV is applied while scanning the cut on τ_{21} . The cuts corresponding to $\tau_2/\tau_1 < 0.45$ for CHS+pruning, PUPPI $\tau_2/\tau_1 < 0.4$ for PUPPI+softdrop or $\tau_{21}^{\text{DDT}} < 0.52$ are indicated with triangles, while the solid circles represent the efficiency and mistag rate for a mass cut only.

that CHS+pruning. Once applying an n-subjettiness cut, the signal efficiency as well as the mistag rate for the PUPPI τ_{21} and CHS τ_{21} taggers drops as a function of 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 mistag rate is flat as a function of p_T , as is the purpose of decorrelated taggers, the signal efficiency improves as the p_T increases, outperforming the other taggers above 1 TeV.

Turning to the tagger pileup dependence, shown in Figure 4.49, the expected benefit from using the PUPPI algorithm is observed: The tagging efficiency for the CHS+pruning (red solid circles) based tagger falls off steeply versus the number of primary vertices in the event, while the PUPPI+softdrop based taggers (light and dark blue solid circles) are more or less insensitive to pileup. Based on general performance, tagging stability versus pileup and due to theoretical considerations, PUPPI softdrop mass with dedicated mass corrections applied together with PUPPI τ_{21} is chosen as this analysis W-tagger. The per-jet efficiency is around 50-55% for a 1-2% mistag rate.

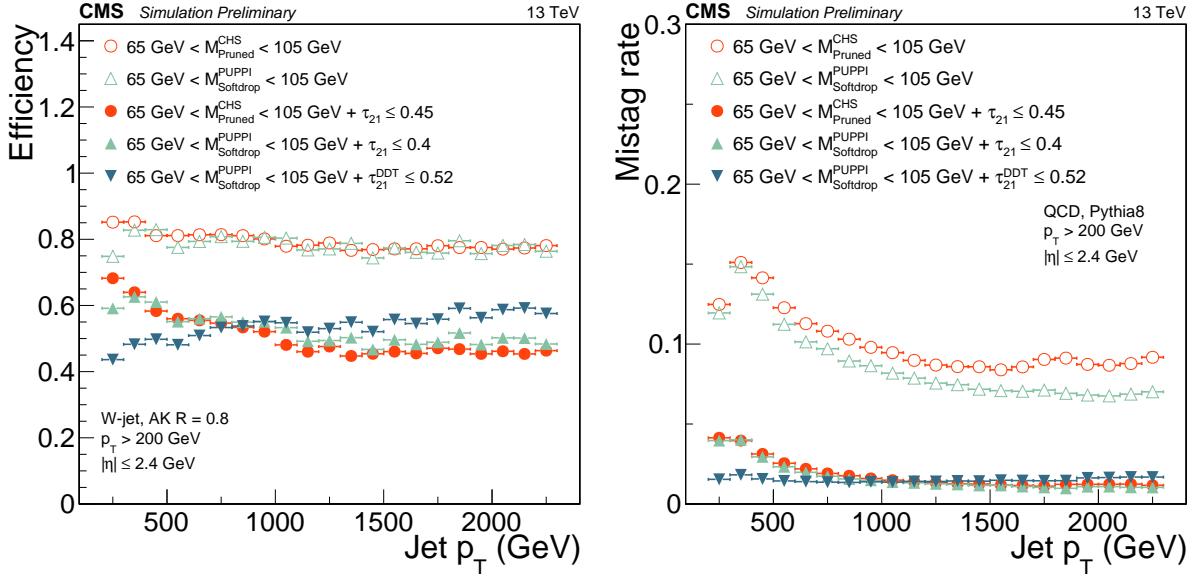


Figure 4.48: W-jet efficiency (left) and QCD light jet mistag rate (right) for a PUPPI+softdrop or CHS+pruned jet mass selection only (hollow circles) and the combined $m_{\text{jet}} + (\text{PUPPI}) \tau_2/\tau_1$ (DDT) selection (solid circles) as a function of jet p_T .

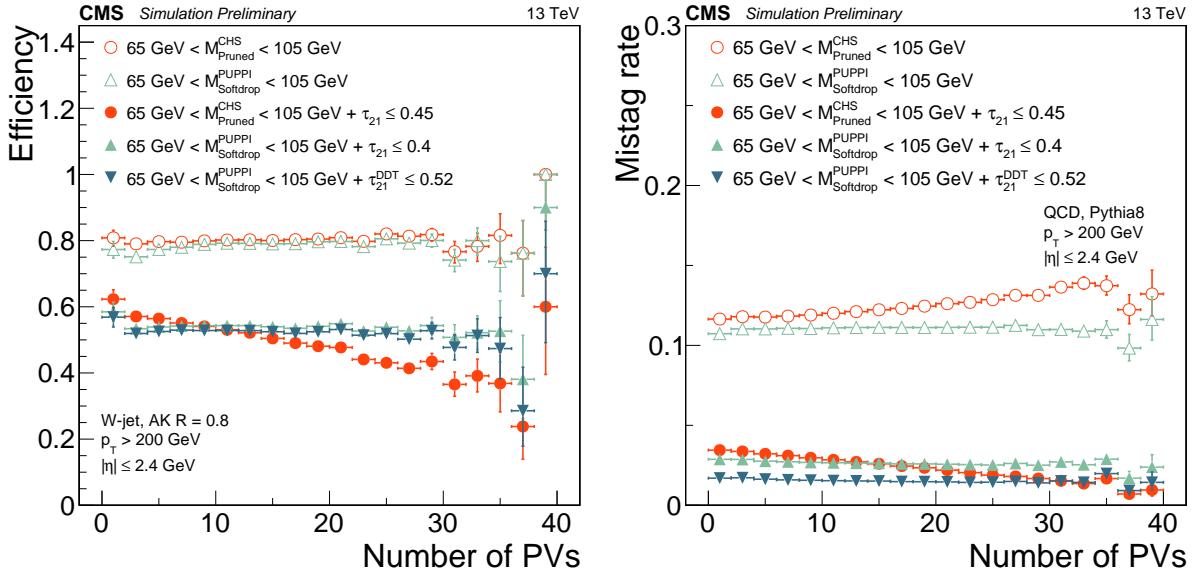


Figure 4.49: W-jet efficiency (left) and QCD light jet mistag rate (right) for a PUPPI+softdrop or CHS+pruned jet mass selection only (hollow circles) and the combined $m_{\text{jet}} + (\text{PUPPI}) \tau_2/\tau_1$ (DDT) selection (solid circles) as a function of jet pileup.

2092 Efficiency scale factors and mass scale/resolution measurement

2093 subsubsectionEfficiency scale factors and mass scale/resolution measurement In order to measure
2094 the W-tagging efficiency, jet mass scale and resolution for the new PUPPI softdrop based tagger,

we use the same procedure as outlined in Section 4.1.7. We first did an early measurement of
 the efficiency using 2.3 fb^{-1} of data collected in 2015, which was published in a jet algorithms
 performance note [73] and served as the first commissioning of the new tagger. We then redid the
 measurement with 12.9 and 35.9 fb^{-1} of 2016 data, respectively, for the two analyses presented in
 this chapter (the latter measurement performed by a separate analysis team). The results shown
 in the following will be those obtained when the tagger was first presented in analysis context,
 corresponding to the 12.9 fb^{-1} of early 2016 data. In order to better understand the differences
 between the CHS+pruning and PUPPI+softdrop based taggers, the first efficiency measurement
 was done in parallel for both algorithms, requiring either a softdrop or a pruned jet mass between
 40 GeV and 150 GeV . The softdrop mass is computed after PUPPI and the jet mass corrections
 as described in Section 4.2.5 are applied, while the pruned mass is corrected with L2L3 jet energy
 corrections. The method is the same as the one outlined in detail in Section 4.1.7 and 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 B.1.
 The PUPPI softdrop jet mass and PUPPI τ_{21} variables in data and in MC are shown in Figure B.3
 and can be compared to the corresponding plots for the CHS pruned jet mass and CHS τ_{21}
 distributions in Figure B.3. The data to MC agreement as well as the observed spectra, is very
 similar between CHS pruning and PUPPI softdrop in this region.

Following what was done in Section 4.1.7, we extract and compare the W-tagging efficiency,
 jet mass scale and 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: $0 < \tau_{21} \leq 0.40 \sim \text{high purity}$
- Fail region: $0.40 < \tau_{21} \leq 0.75 \sim \text{low purity}$

The corresponding fits are shown in Figure B.4, with the corresponding extracted efficiencies
 from the Gaussian component of the total fit and scale factors summarized in Table B.1. The
 quoted systematic uncertainties are evaluated the same was as described in Section 4.1.7 and
 correspond to the uncertainty due to ME generator and due to choice of fit method.

	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 4.9: 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} .

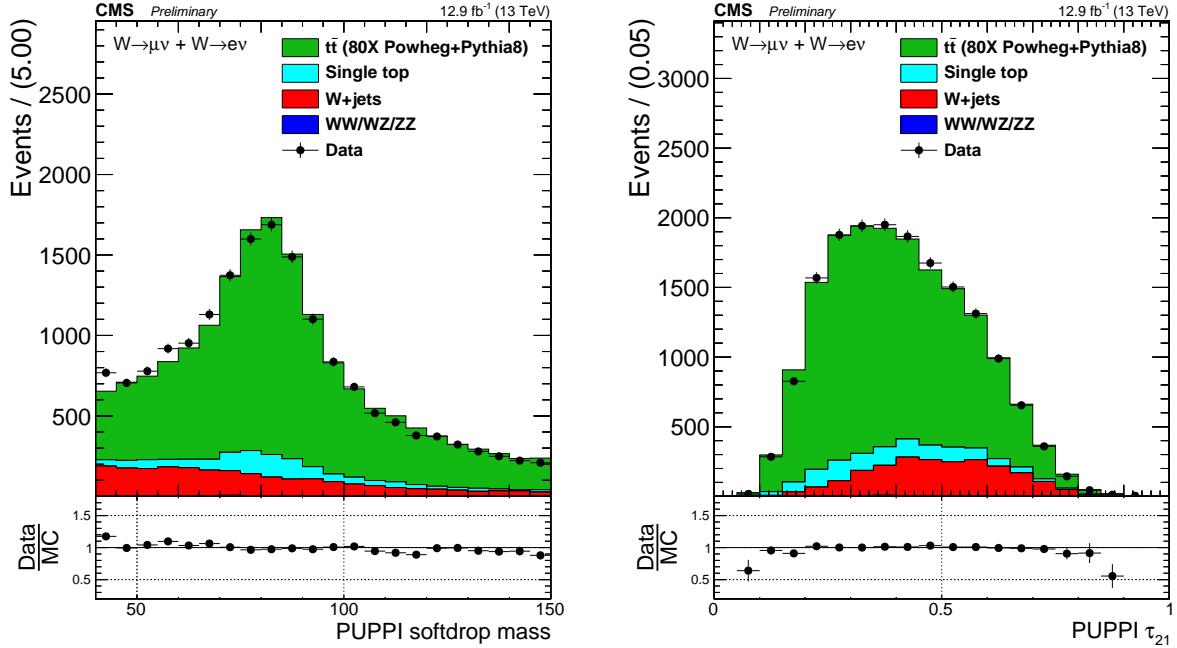


Figure 4.50: Distribution of the PUPPI softdrop mass (left) and PUPPI n-subjettiness (right) distribution in the $t\bar{t}$ control sample.

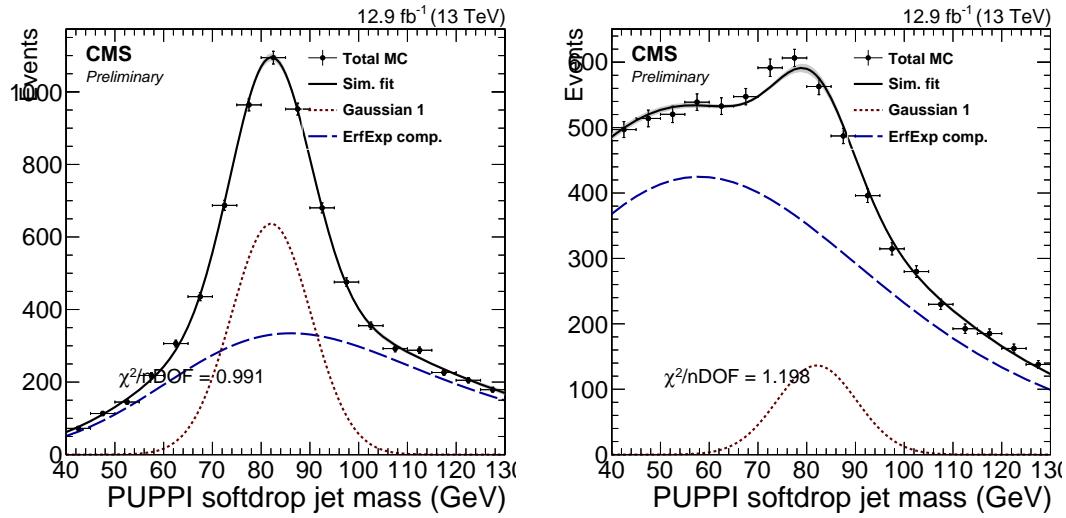


Figure 4.51: PUPPI softdrop jet mass distribution that pass (left) and fail (right) the PUPPI $\tau_2/\tau_1 < 0.40$ selection. Results of both the fit to data (blue) and simulation (red) are shown and the background components of the fit are shown as short-dashed lines.

Both scale factors are comparable to unity, within uncertainties. We additionally extract the jet mass scale and jet mass resolution from the mean and width, respectively, of the Gaussian component of the total fit in the pass region. These are summarized in Table B.2. As for pruning (Table 4.7), we find that the W jet mass scale is larger in simulation than in data, of roughly 2%. The jet mass resolution, on the other hand, is larger in data, of roughly 8%, whereas for pruning the resolution is larger in simulation (11%). However, both are statistically insignificant and comparable with unity within uncertainties.

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

Table 4.10: Summary of the fitted W-mass peak fit parameters.

The W-tagging efficiency scale factors, jet mass scale and resolution affects the signal yield and are included as described in Section 4.1.7: as a scale of the total signal yield and an uncertainty on the signal efficiency due to a shift and broadening of the W-jet mass peak.

4.2.6 W-tagging mistagging rate measurement

We additionally measure the W-tagging fake rate in data in a QCD dijet enriched region and compare this to the prediction from QCD MC using the three different combination of generators: HERWIG++, PYTHIA and MADGRAPH +PYTHIA. Figure 4.52 shows the mistag rate as a function of p_T for three different taggers: CHS pruning + τ_{21} , PUPPI softdrop + PUPPI τ_{21} and PUPPI softdrop + 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 tagging p_T dependence well. We additionally study the difference in the total quark/gluon-content for the two PUPPI softdrop based taggers: τ_{21} and τ_{21}^{DDT} . Figure 4.53 shows the stacked relative q/g content in a Pythia 8 QCD dijet sample for a cut on PUPPI τ_{21} and τ_{21}^{DDT} . We see that the quark content increases as a function of jet p_T when cutting on the DDT, while it decreases when cutting on τ_{21} . This can be attributed to the fact that the m/p_T distribution for quark and gluon jets are very different from one another, and this difference increase as the jet p_T increases. Figure 4.54 shows the m/p_T for jets originating from a quark (blue) and a gluon (red) for a jet p_T of 200 GeV (left) and 1600 GeV. We see that the mass over p_T for gluon jets is significantly higher for gluon jets than for quark jets. With the

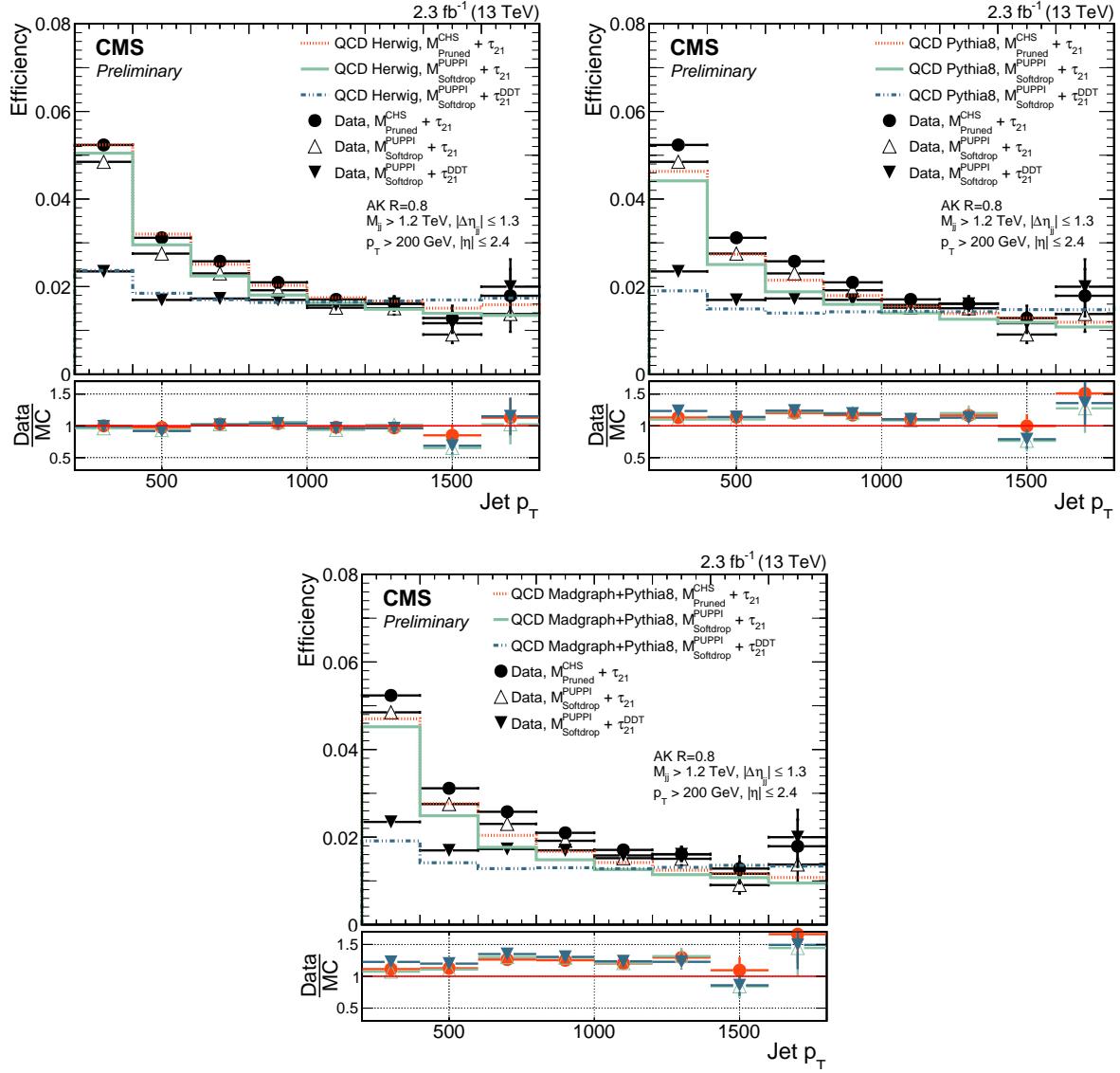


Figure 4.52: The fraction of jets that pass the $m_{jet} + \tau_2/\tau_1$ selections in a dijet enriched sample for data and for simulation as a function of jet p_T . Here comparing HERWIG++ (left), PYTHIA8 (right) and PYTHIA8 with MADGRAPH as matrix-element generator (left).

²¹⁵⁰ τ_{21}^{DDT} tagger being defined as in Equation 4.10, the DDT will therefore act more aggressive on
²¹⁵¹ jets with a high m/p_T , effectively removing more gluon jets.

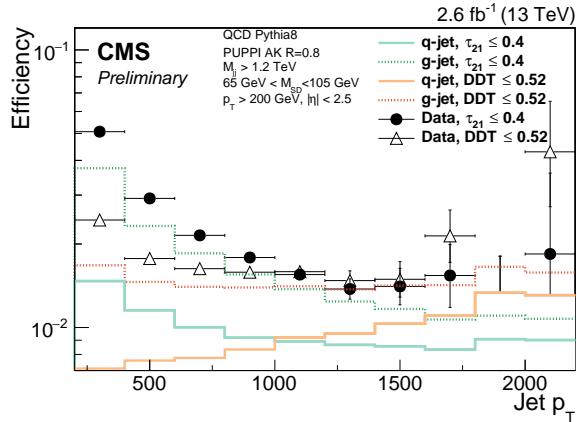


Figure 4.53: The fraction of jets that pass the PUPPI softdrop m_{jet} with τ_2/τ_1 (turquoise) or τ_{21}^{DDT} (orange) selections in a dijet enriched sample. The jets from QCD MC are split into two contributions: jets originating from gluons (dotted line) and jets originating from quarks (solid line).

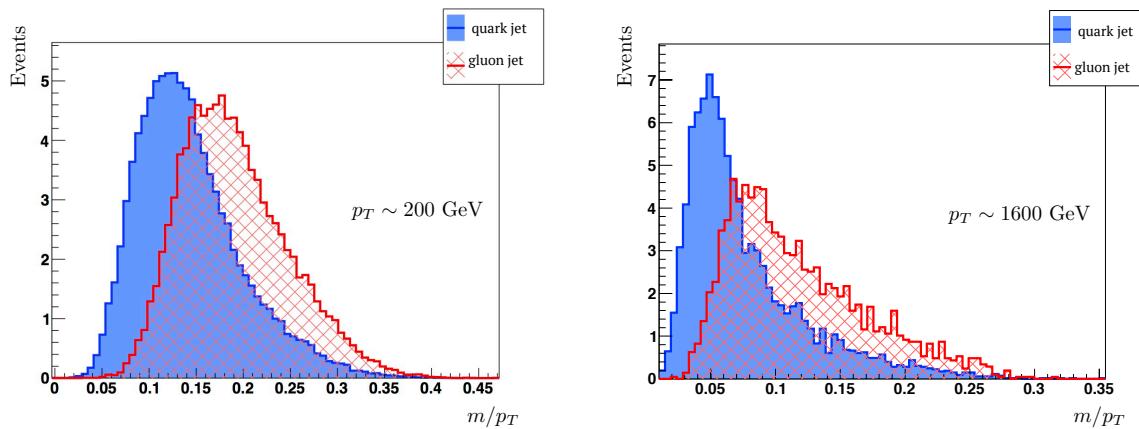


Figure 4.54: The jet mass divided by the jet p_T for quark (blue) and gluon (red) jets for a jet p_T of 200 (left) and 1600 GeV (right). Created with [76].

4.2.7 Mass and purity categorization

The PUPPI softdrop jet mass and PUPPI τ_{21} distribution after loose analysis preselections, as outlined in Section 4.2.3, are shown in Figure 4.55. We see some disagreement between data and MC, especially in the high-purity region ($\text{PUPPI } \tau_{21}^{DDT} < 0.4$), confirming what we observed in Section 4.2.6. As this analysis is sensitive to both heavy resonances decaying into two vector

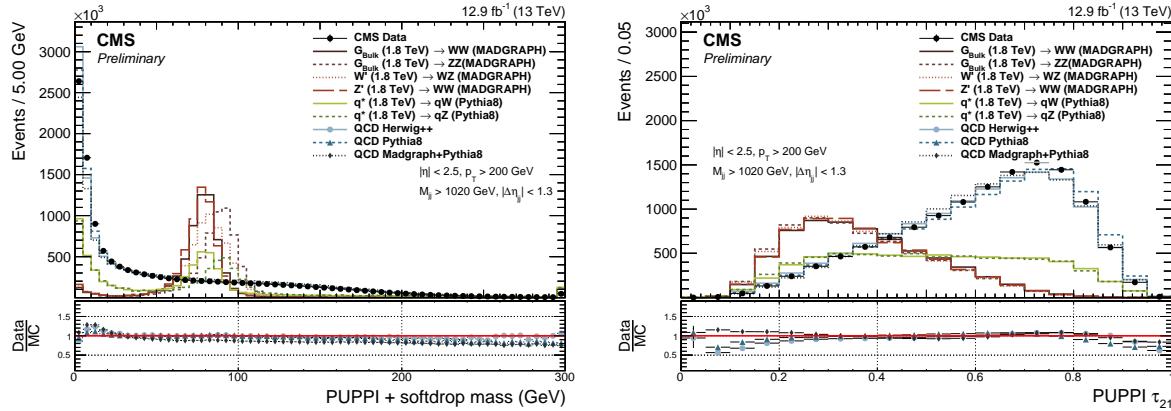


Figure 4.55: PUPPI softdrop jet mass distribution (left) and PUPPI n-subjettiness τ_{21} (right) distribution for data and simulated samples. Simulated samples are scaled to match the distribution in data.

bosons and excited quark resonances q^* decaying to qW and qZ, we look for events with both a single W/Z-tag and events with two W/Z-tags. Vector boson candidates are selected with a PUPPI softdrop jet mass of $65 < m_{\text{jet}} < 105$ GeV. Further, and similar to what was done in Search I, we select “high purity” (HP) W/Z jets by requiring PUPPI $0 < \tau_{21} \leq 0.40$ and “low purity” (LP) jets with $0.40 < \tau_{21} \leq 0.75$. The events with one W/Z-tag are classified in HP and LP events according to the two categories described previously. 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 into two mass categories in order to enhance the analysis sensitivity, with the W window defined as $65 \text{ GeV} < m_{\text{pruned}} < 85 \text{ GeV}$ and the Z boson window as $85 \text{ GeV} < m_{\text{pruned}} < 105 \text{ GeV}$. This results in ten different signal categories. They are as follows:

- High-purity double W/Z-tag, 3 mass categories: WW, ZZ and WZ
- Low-purity double W/Z-tag, 3 mass categories: WW, ZZ and WZ
- High-purity single W/Z-tag, 2 mass categories: qW and qZ

- 2171 • Low-purity single W/Z-tag, 2 mass categories: qW and qZ

2172 4.2.8 Background modeling: F-test

2173 With the full analysis selections and categorization defined, we move to the determination of
 2174 background fit function. Following the same strategy as in Section 4.1.5, we determine the number
 2175 of necessary parameters in order to describe the background through a Fishers F-test, comparing
 2176 the same fit functions as in Section 4.1.5. This test is first exercised in QCD MC and then in a
 2177 data sideband before the final determination in the data signal region. As the F-test method
 2178 was presented in detail in the context of Search I, only a brief summary and the fits in the new
 2179 single-tag categories will be presented here, while all fits and F-test results can be found in
 2180 Appendix B.3.

2181 A two or three parameter fit is sufficient to describe the background for all the double tag
 2182 categories: a two parameter fit is sufficient for the “high-purity” WZ and ZZ categories. as
 2183 well as the “low-purity” WW category, while the remaining analysis categories require a three
 2184 parameter background fit. From the fits to the single tag categories, shown in Figure 4.56, a three
 2185 parameter fit is sufficient for all categories except the “high-purity” qW category. In this category
 2186 the improvement in fit quality when increasing the number of parameters is so large adding that
 2187 adding an additional fit parameter is justified, and we continue by using a 5 parameter fit for
 2188 this category. A summary of what fit functions are used for each analysis category is listen in
 Table 4.11.

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

Table 4.11: Fit parameters used in each analysis category

2189

2190 4.2.9 Signal modeling

2191 The signal is modeled from signal MC in the same way as was done in Section 4.1.6, assuming a
 2192 Gaussian core and an exponential tail. The interpolated signal shapes for $q^* \rightarrow qW$ and $q^* \rightarrow qZ$

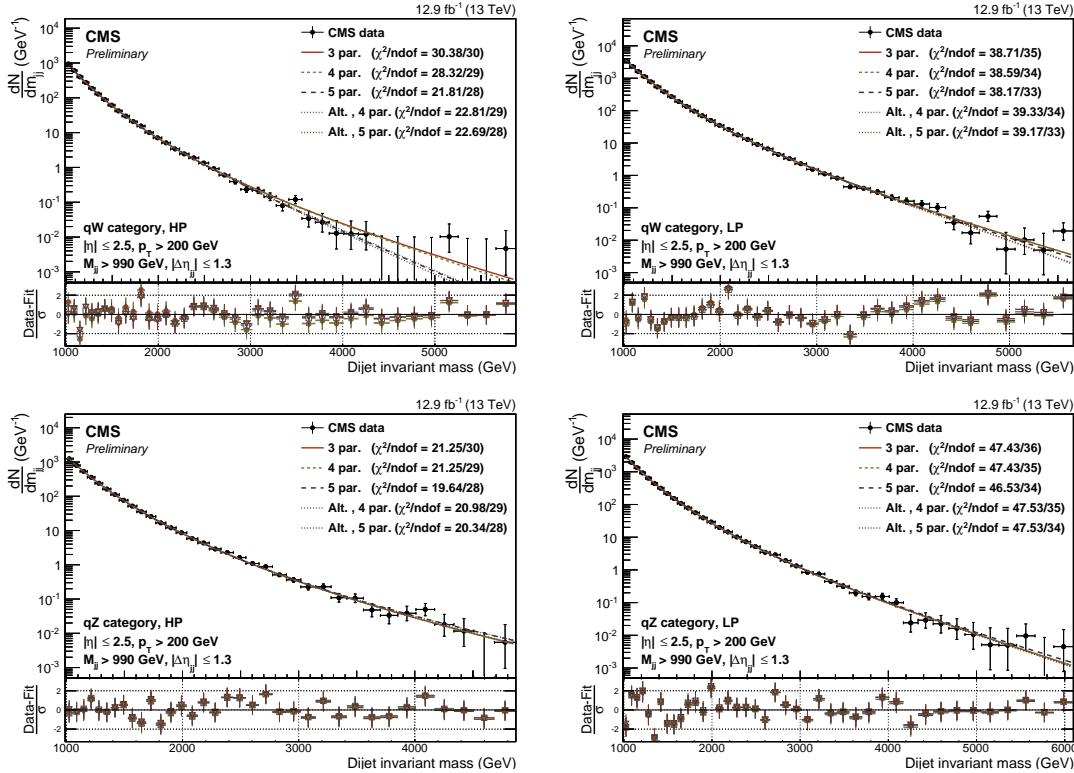


Figure 4.56: Background fit for the M_{jj} distribution in the data signal region for the single-tag analysis. Here for the high- (left) and low-purity (right) single W/Z-tag categories qW (top) and qZ (bottom).

in their most sensitive analysis categories (qW and qZ, respectively) are shown in Figure 4.57.
The signal shapes for the double-tag category can be compared to those in Figure 4.22.

4.2.10 Systematic uncertainties

The largest sources of systematic uncertainty for this analysis is, as for Search I, related to the signal modeling and are due to the uncertainty in the tagging efficiency of the W/Z-tagger, the jet energy/mass scale, the jet energy/mass resolution and integrated luminosity. The W/Z tagging uncertainty is estimated in $t\bar{t}$ events, as described in Section 4.1.7, 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 were approximated by a constant 2% and 10% uncertainty in Search I (Section 4.1.7) and are not expected to change for the 2016 analysis. The jet energy response and resolution uncertainty are taken into account as shape uncertainty by shifting and widening the signal resonance model, while all other signal uncertainties only

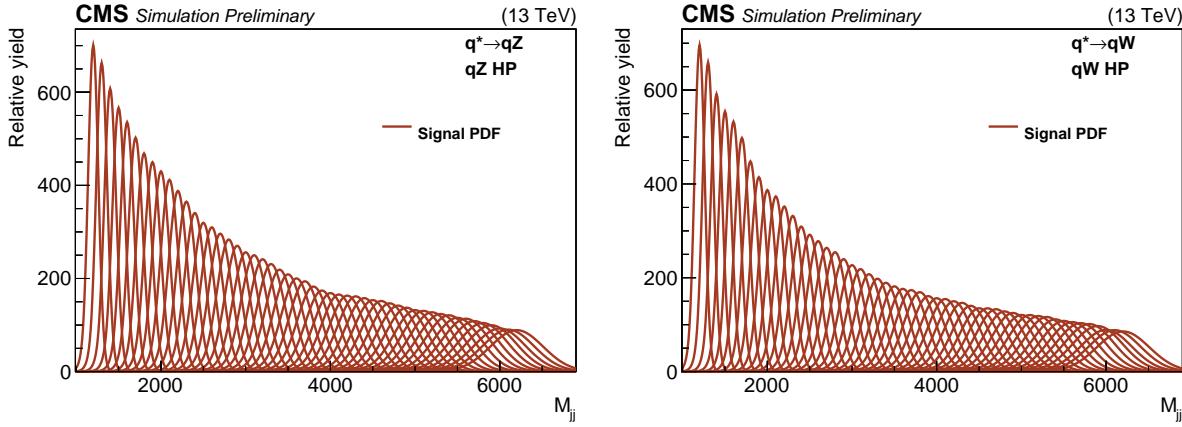


Figure 4.57: Interpolated signal shapes for a $q^* \rightarrow qZ$ (left) and $q^* \rightarrow qW$ (right) signal.

affect the yield. The list of most relevant systematic uncertainties are listed in Table 4.12.

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

4.2.11 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 early 2016 data, demonstrating the new PUPPI softdrop based tagger and single-tag analysis categories, and one topping up with the full 35.9 fb^{-1} dataset. The results from both will be presented in the following.

Early analysis

Exclusion limits are set in the context of the bulk graviton model, the HVT model B scenario and excited quark resonances, assuming the resonances to have a natural width negligible with respect to the experimental resolution (as in Search I).

Figure 4.58 shows the 95% confidence level (CL) expected and observed exclusion limits on the signal cross section as a function of the resonance mass for the different signal hypotheses in the double-tag analysis. The limits are compared with the cross section times the branching fraction to WW and ZZ for a bulk graviton with $\tilde{k} = 0.5$, and with the cross section times the branching fraction to WZ and WW for spin-1 particles predicted by the HVT model B for both the singlet (W' or Z') and triplet (W' and Z') hypothesis. 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 ?? shows the corresponding exclusion limits for excited quarks decaying into qW and qZ. We exclude excited quark resonances 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.

Full 2016 dataset

The results obtained with the full $\sim 36 \text{ fb}^{-1}$ of 2016 data are as follows: 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 resonance masses above 3.6 TeV. 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. Figure 4.60 and 4.61 show the resulting 95% confidence level expected and observed exclusion limits on the signal cross section as a function of the resonance mass for

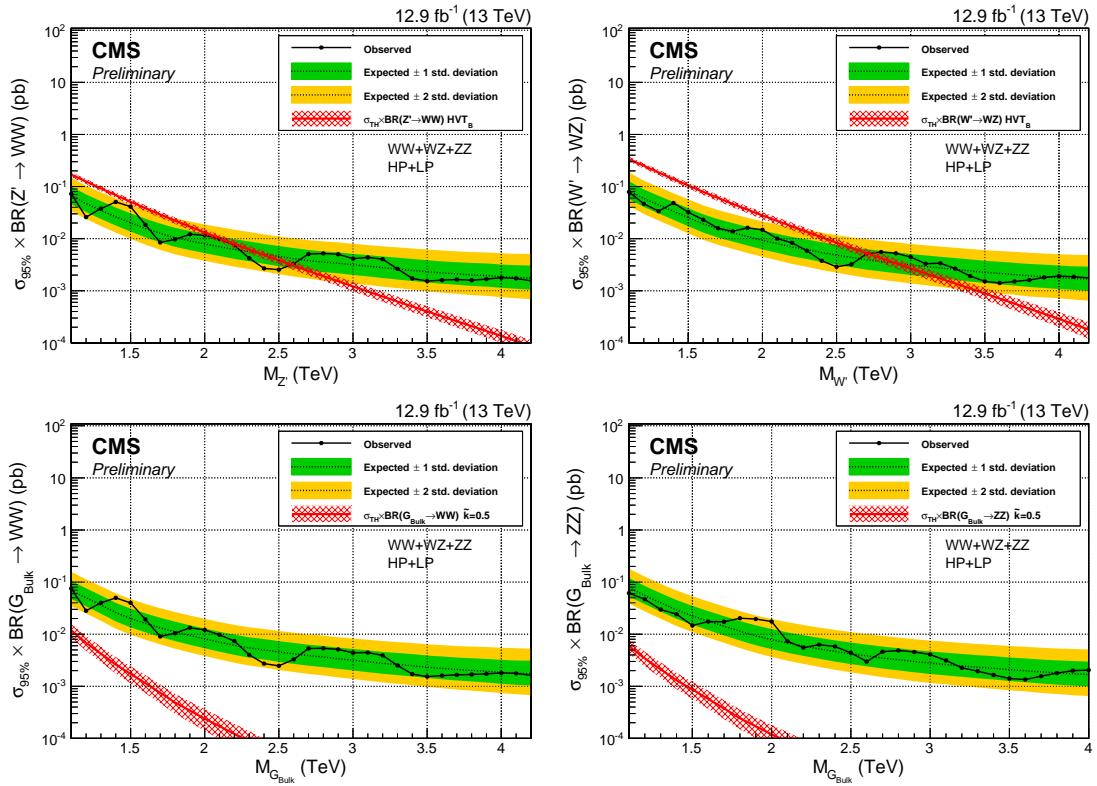


Figure 4.58: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of a narrow-width resonance decaying to a pair of vector bosons for different signal hypotheses. Limits are set in the context of a spin-1 neutral Z' (left) and charged W' (right) resonances resonance, 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.

2237 VV and QV resonances, respectively.

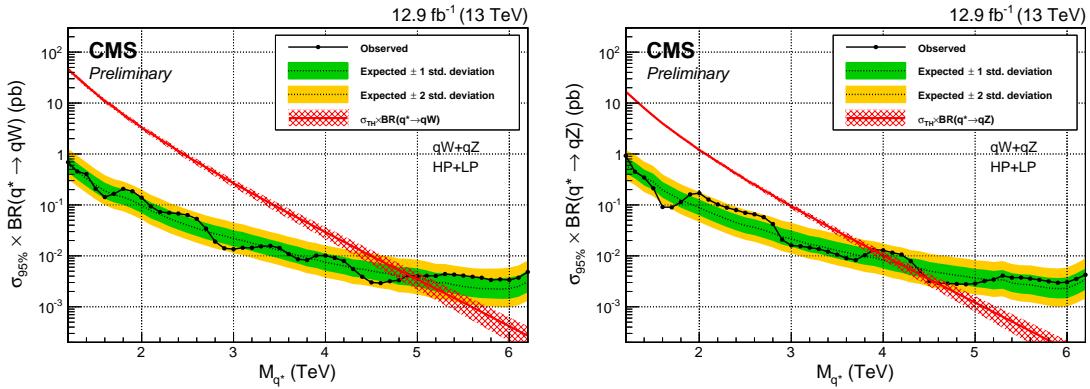


Figure 4.59: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of an excited quark resonance decaying into qW (left) or qZ (right). Signal cross section uncertainties are displayed as a red checked band.

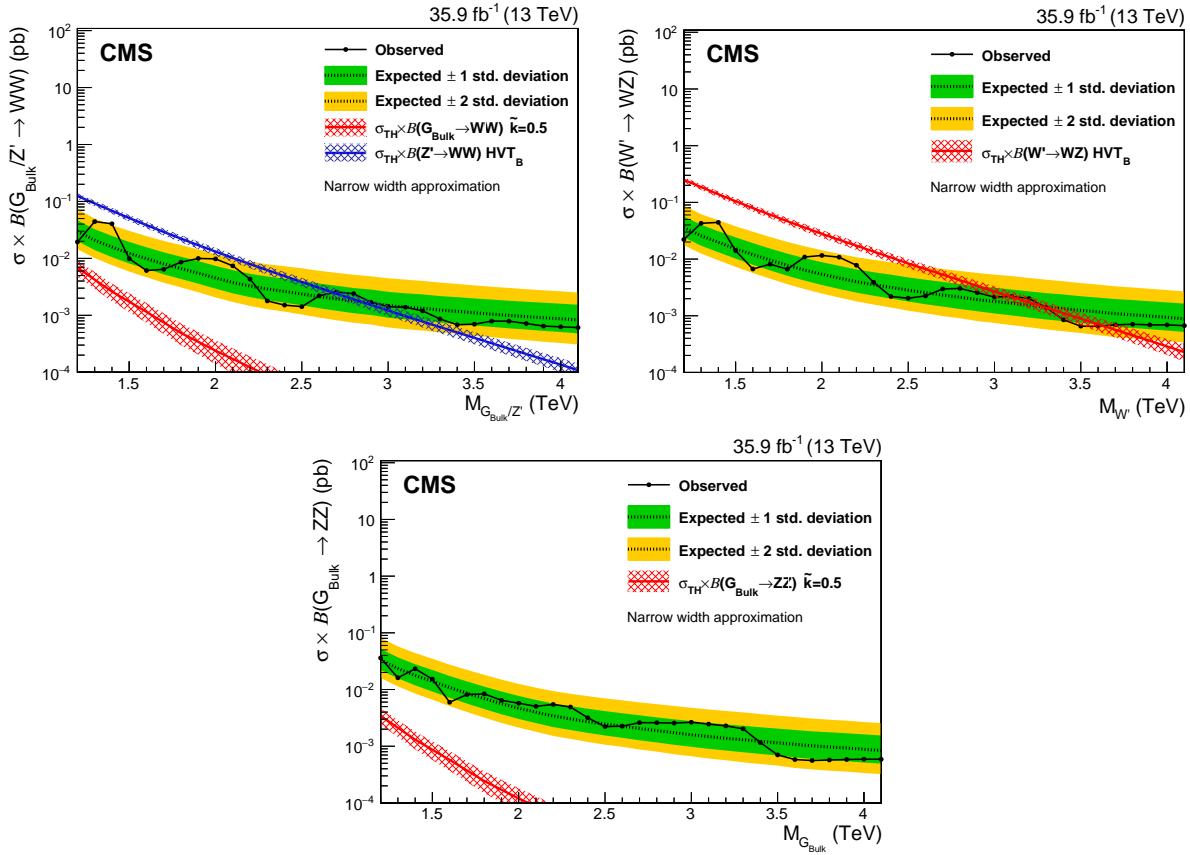


Figure 4.60: Observed (solid line) and expected (dashed line) 95% CL upper limits on the production cross section of a narrow resonance decaying into two vector bosons for different signal hypotheses: A Z' or G_{bulk} resonance decaying into WW (top left), a Z' decaying into WZ (top right) and a bulk graviton decaying into ZZ (bottom).

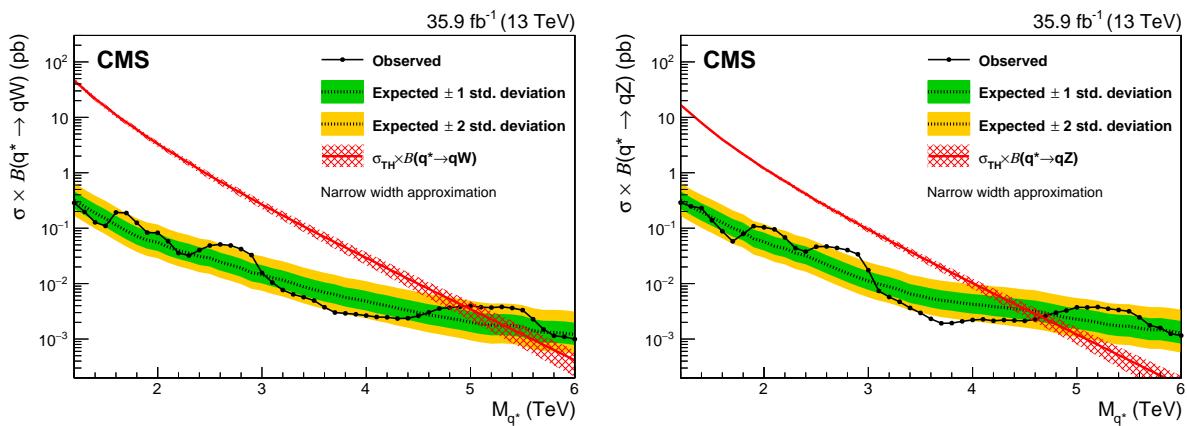
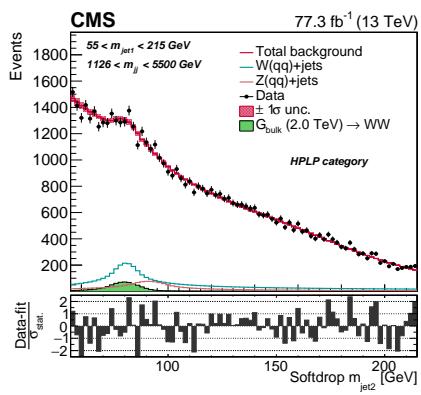


Figure 4.61: Observed (solid line) and expected (dashed line) 95% CL upper limits on the production of an excited quark resonance decaying into qW (left) or qZ (right).

2238 4.3 Search III: A novel framework for multi-dimensional searches

2239 After two successful analyses of 13 TeV data, no excess had confirmed the 8 TeV bump and the
 2240 available phase space for New Physics to hide out was shrinking. However, this fact wouldn't
 2241 disarm everybody. On the BSM theory front, ideas were simmering about whether it was possible
 2242 that the small bumps we were observing here and there in the dijet invariant mass spectrum were
 2243 due to us catching the tail of another type of boson with a mass slightly different from that of a
 2244 W or a Z boson? And that perhaps these jets were not 2-prong objects, but in reality 4-prong?
 2245 With Run 2 coming to an end, marking the beginning of a two year long shut-down, it was time
 2246 to think about how we could probe alternative BSM models as effectively as possible. Our idea
 2247 was therefore the following: We would build a novel framework capable of easily scanning the
 2248 full softdrop jet mass and N-prong spectrum, and which, in addition, would lead to a gain in
 2249 sensitivity for the standard VV all-hadronic search. We would do this by taking advantage of
 2250 the fact that we were looking for bumps in a three-dimensional plane: the softdrop mass of the
 2251 two jets as well as their invariant mass. As a validation of the method, the method would be
 2252 demonstrated in the context of the VV all-hadronic search, replacing the dijet fit method. We
 2253 would then extend this to simultaneously search for resonances decaying to $W(qq)$, $Z(qq)$ and
 2254 $H(qq)$ and, finally, take full advantage of the framework and look for generic resonances peaking
 2255 anywhere in the jet mass and dijet invariant mass spectrum.

2256
 2257 Search III introduces a novel three-dimensional search method, allowing to simultaneously search
 2258 for $W/Z/H$ signals, and eventually non-SM bosons, in the softdrop jet mass spectrum. It is
 2259 the first analysis to measure the $V+jets$ cross section and the jet mass scale/resolution from a
 2260 $W(qq)+jets$ and $Z(qq)+jets$ mass peak. Published with full 2016+2017 dataset, $\sim 80 \text{ fb}^{-1}$.



2261 4.3.1 Small bumps and tri-bosons

2262 Ever since the ATLAS observation of a 3.4σ excess in the search for VV resonances in the
 2263 all-hadronic final state [51], several little bumps kept re-appearing near 2 TeV. These were not
 2264 statistically insignificant, as we've already seen in Search I and Search II, but rather small elusive
 enhancements, illustrated by the collection of ATLAS/CMS observations in Figure 4.62. Due to

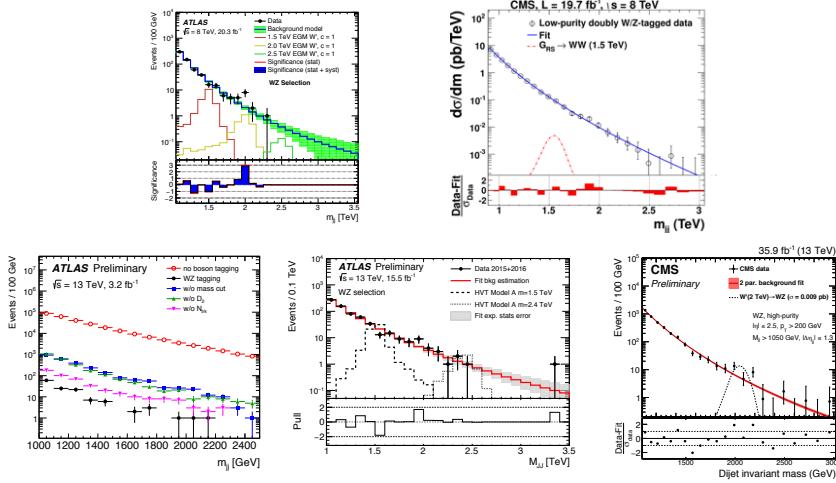


Figure 4.62: Several small bumps observed in VV resonance searches in the all-hadronic final state, both in ATLAS and in CMS [77].

2265
 2266 their small size and the way the excesses seemed to slightly shift around, these were obviously
 2267 not diboson resonances. However, could they be caused by us catching the tails of some non-SM
 2268 boson with a mass slightly different from that of a SM vector boson, as illustrated in Figure 4.63?
 Further, these could be 4-pronged objects rather than 2-prong, which would cause the excess to

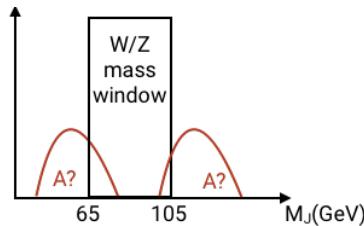


Figure 4.63: Slight excesses in diboson analyses could be caused by catching the tail of a non-SM object peaking at slightly higher/lower jet mass than at the W/Z mass.

2269
 2270 vary in size depending on the 4-prong efficiency of the analysis specific W-tagger used.
 2271 An explanation for the observed excesses was proposed in [78]. This paper pointed out that, if

2272 particles like W' and Z' exist, an extended scalar sector is needed in order to give mass to the
 2273 vector bosons. These heavy scalars will decay to lighter bosons, if kinematically allowed, leading
 2274 to multiboson signals from cascade decays. Some example signatures are illustrated in Figure 4.64.
 Signatures like these would peak in the groomed jet mass spectrum and, depending on what the

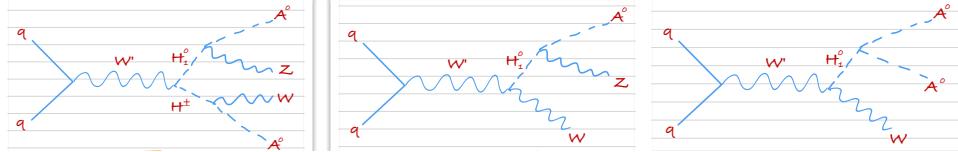


Figure 4.64: A W' decaying to a neutral H^0 and a charged H^\pm scalar particle leading to a quadriboson final state (left), and a W' decaying to a neutral scalar particle H^0 and a W leading to a triboson final state (middle and right) [78].

2275
 2276 final bosons decay into, have very different substructure profiles (4- and 4-prong, 2- and 4-prong
 2277 etc.).

2278 In order to effectively search for such types of signals, or any signal peaking in the softdrop
 2279 jet mass spectrum, we therefore wanted to build a generic new framework allowing to look for
 2280 peaks anywhere in the groomed mass - dijet invariant mass spectrum. Rather than selecting
 2281 jets with a groomed mass between 65 and 105 GeV and look for resonances peaking in the dijet
 2282 invariant mass, we'd look for resonances peaking anywhere in the three dimensional plane formed
 2283 by the groomed mass of each jet and their invariant mass. The benefits of this procedure, was
 2284 that it would allow us to scan the full groomed mass spectrum in one analysis. We would first
 2285 demonstrate the method through the VV all-hadronic analysis, which is the paper introduced
 2286 here.

2287 4.3.2 Analysis strategy

2288 The background estimation used in Search I and Search II rely on a one dimensional fit of the
 2289 dijet invariant mass signal region after a tight jet mass cut (65-105 GeV) has been applied. We
 2290 now take advantage of the fact that the signal peaks in three dimensions; dijet invariant mass
 2291 (M_{VV}) and the jet groomed mass of jet 1 and jet 2 (M_{jet1} and M_{jet2}), and attempt to extract
 2292 the signal from the three dimensional M_{VV} - M_{jet1} - M_{jet2} plane. The benefits of doing so is that
 2293 we now can perform different searches in the WW , WZ , ZZ , WH or XX final states encoded in
 2294 the same analysis. Additionally, tight jet mass cuts are no longer needed as we fit the full jet
 2295 mass line-shape to extract the signal. This effectively increases our signal statistics as a large
 2296 fraction of the W and Z signal fall outside the above mass window. Fitting the jet groomed mass
 2297 and resonance mass together also allow us to add nuisance parameters that simultaneously affect

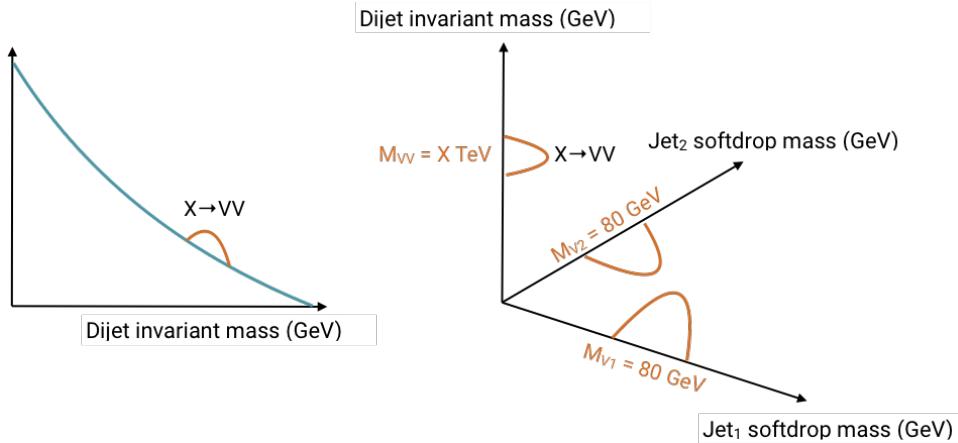


Figure 4.65: The one dimensional VV analysis versus the three dimensional fit.

the jet groomed mass and the resonance mass, fully accounting for the correlation between the variables. We would model the background starting from simulation, rather than the dijet fit to data, which would allow us to model peaky distributions like a trigger turn-on. This could allow the search to go even lower in M_{VV} , something we will discuss further in Section 4.3.5. This chapter, and its corresponding publication, is an analysis of the 2016 and 2017 dataset, corresponding to $\sim 80 \text{ fb}^{-1}$, and serves as the first documentation and demonstration of the novel three-dimensional fit method. In Section 4.5, I'll discuss how we plan to take this framework further in searches for VH and HH as well as for generic resonances peaking in jet softdrop mass.

4.3.3 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 4.2.3, with specific detector conditions to match the 2016 and 2017 dataset.

4.3.4 Event selection

Events are selected following the same criteria as in Search I and Search II (see Section 4.1.4) and can be summarized as follows:

- PF jet Tight ID applied

- Jet $\eta < 2.5$

- 2316 • Jet $p_T > 200$ GeV

- 2317 • $|\Delta\eta|_{jj} < 1.3$

2318 The two jets with the highest groomed jet mass in the event are selected as potential vector boson
 2319 candidates. In addition, the dijet invariant mass is required to be > 1126 GeV in order to be on
 2320 the trigger plateau. As already mentioned in the introduction, the background modeling used
 2321 in this analysis is capable of modeling turn-ons and is something we explored. However, while
 2322 the background modeling was reliable, we found it difficult to extract a signal peaking on top of
 2323 a turn-on and had to abandon the trigger modeling for this first demonstration of the method.
 2324 More details will be given in Section 4.3.5.

2325

2326 The dijet invariant mass and $|\Delta\eta|_{jj}$ distribution for the two leading jets in the event after the
 2327 above preselections have been applied is shown in Figure 4.66. The jet p_T and η distributions for
 2328 signal and for background is shown in Figure 4.67.

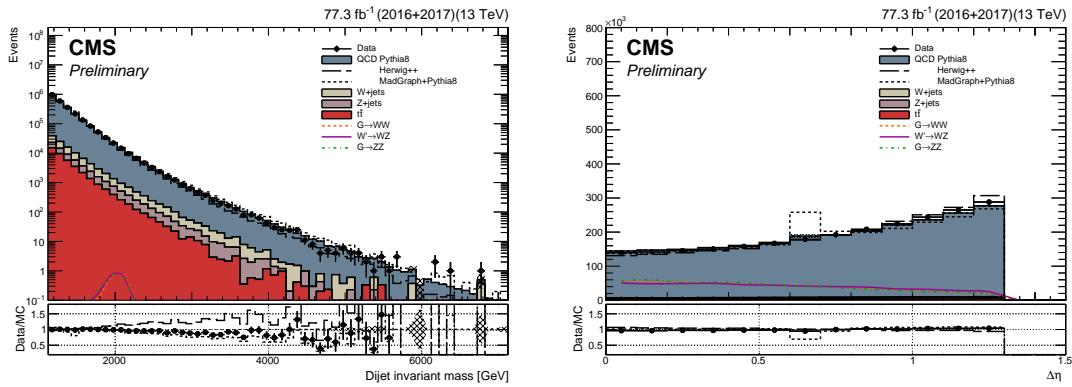


Figure 4.66: The dijet invariant mass (left) and $|\Delta\eta|_{jj}$ (right) for the two leading jets after preselections are applied. The signal is scaled by an arbitrary number.

2329 4.3.5 Triggering

2330 The triggers used for 2016 data are the same as in Section 4.1.4, while the thresholds in 2017
 2331 have increased in order to push the trigger rate to a level acceptable for the increased luminosity.
 2332 The triggers used for 2017 data are

- 2333 • HLT_PFHT1050

- 2334 • HLT_AK8PFJet500

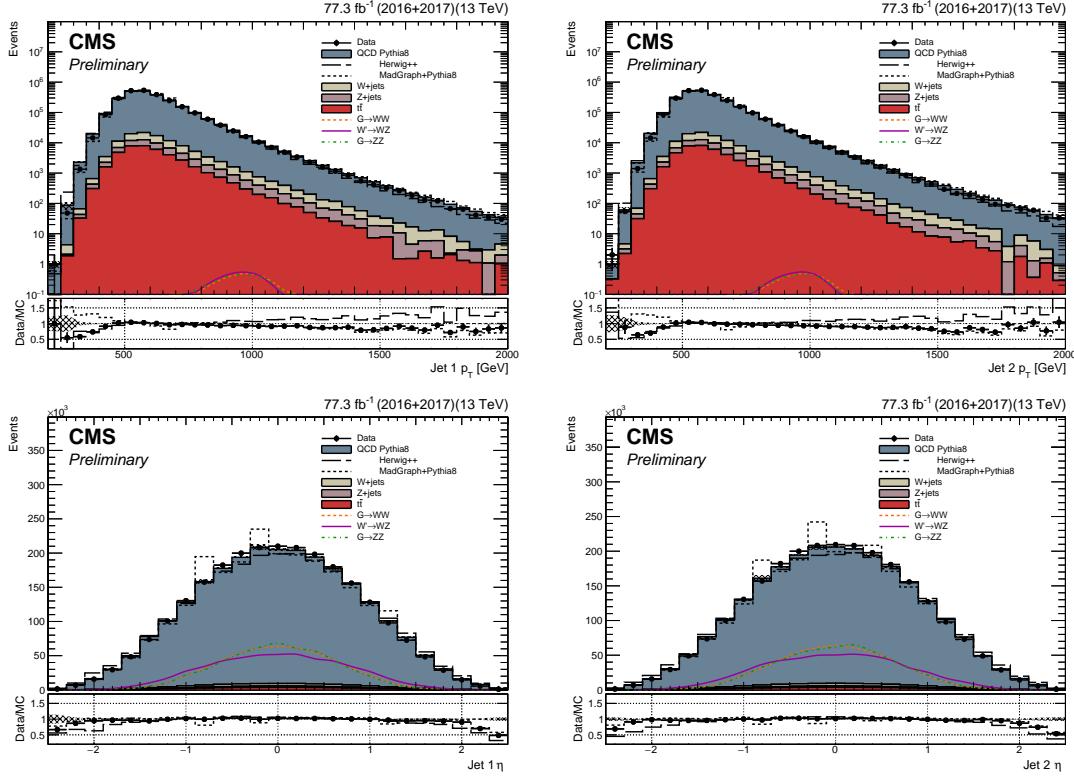


Figure 4.67: Jet p_T (top) and η (bottom) of the leading (left) and second leading (right) jet in the event. The signal is scaled by an arbitrary number.

- 2335 • HLT_AK8PFJet360/380/400/420_TrimMass30

- 2336 • HLT_AK8PFHT750/800/850/900_TrimMass50

2337 . For the results presented here, the analysis threshold is set by the trigger turn-on point (where
2338 the combination of all triggers are > 99 percent efficient). The trigger turn-on is evaluated in the
2339 Single Muon dataset, using the HLT_Mu50 and HLT_IsoMu27 triggers as reference triggers. The
2340 trigger turn-on curves as a function of dijet invariant mass and jet soft drop mass are shown in
2341 Figure 4.68. The combination of all triggers are $> 99\%$ efficient above a dijet invariant mass of
2342 1126 GeV and this sets the analysis threshold.

2343 Trigger turn-on modeling

The analysis threshold for searches depending on the dijet fit is set by the trigger turn-on point as the analysis relies on a background fit of the dijet invariant mass spectrum with a smoothly falling function. As the background modeling for this analysis does not depend on a smoothly

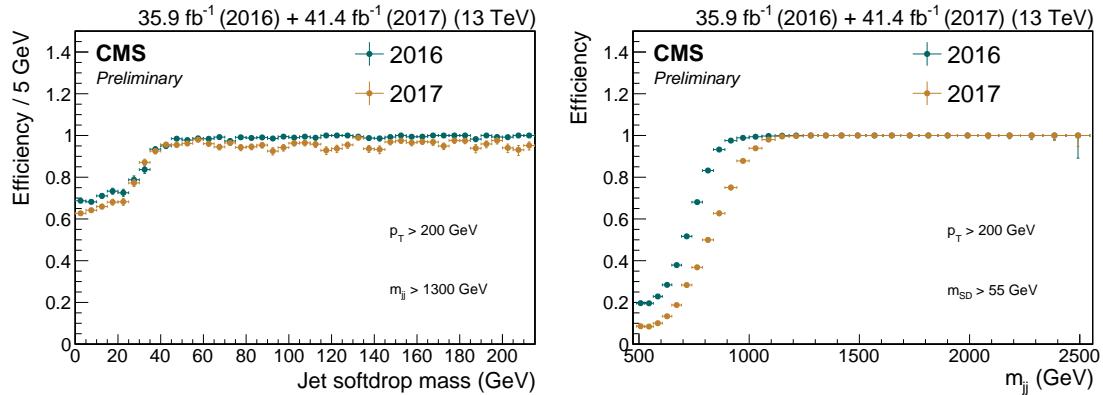


Figure 4.68: Trigger turn-on curves in the 2016 and 2017 datasets for the H_T based (left) and groomed mass based (right) trigger paths.

falling spectrum (as will be described in detail in Section ??), and in order to compensate for a loss in acceptance due to increased trigger thresholds, we attempted to model the trigger turn-on directly from data and apply a trigger weight in simulation. To do so, we first study the one dimensional trigger turn-ons versus dijet invariant mass and softdrop jet mass to understand where in M_{VV} and $M_{\text{jet}1}/M_{\text{jet}2}$ the triggers are fully efficient. We then derive a three dimensional histogram of the trigger efficiency versus dijet invariant mass (M_{VV}) and the jet groomed mass of jet 1 and jet 2 ($M_{\text{jet}1}$ and $M_{\text{jet}2}$), where each bin corresponds to the trigger efficiency for a given value of $M_{VV}-M_{\text{jet}1}-M_{\text{jet}2}$. The procedure is as follows: From the one dimensional histograms, the points of full efficiency versus M_{VV} , $M_{\text{jet}1}$ and $M_{\text{jet}2}$ are defined. For every bin above this threshold, the trigger efficiency is fixed to one (100 percent efficiency). For all bins below this trigger threshold, we fit slices of M_{VV} with a sigmoid function, evaluate the trigger weights from this function, and set the bin content of the three dimensional weight histogram accordingly. For every bin below this point, the trigger weight is extracted by fitting slices of M_{VV} in each bin of $M_{\text{jet}1}/M_{\text{jet}2}$. As the trigger efficiency falls below 50 percent around a dijet invariant mass of 800 GeV (Figure 4.68), searching for resonances with masses below this point is non-feasible. In addition, the full signal shape needs to be contained within the dijet invariant mass spectrum, excluding resonance masses of 0.8 and 0.9 TeV. The lowest mass point signal sample to set limits on for this analysis is therefore at 1 TeV and the analysis threshold is fixed at $M_{VV}=900$ GeV in order to fully contain the signal shape. Starting from $M_{VV}=893$ GeV (dijet bin closest to 900 GeV) and $M_{\text{jet}1}/M_{\text{jet}2}=40$ GeV, a coarsely binned three-dimensional histogram (10 GeV $M_{\text{jet}1}/M_{\text{jet}2}$ binning and “dijet” binning in M_{VV}) is filled with the fraction of events that pass

one of the signal triggers,

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

The resulting coarse histogram is then expanded in M_{VV} to 10 GeV dijet invariant mass bins, interpolating between the points using sigmoid fit for each $M_{\text{jet}1} - M_{\text{jet}2}$ bin (the fine binning in dijet invariant mass is sufficient enough to yield a smooth reweighted distribution so no expansion is done in $M_{\text{jet}1}/M_{\text{jet}2}$). From this histogram, each slice in M_{VV} is fitted with a sigmoid function,

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

- ²³⁴⁴ the trigger weight of each bin $M_{VV}-M_{\text{jet}1}-M_{\text{jet}2}$ is extracted from the fit and set accordingly. Figure 4.69 shows the total projections on each axis for the full trigger weight histogram. The

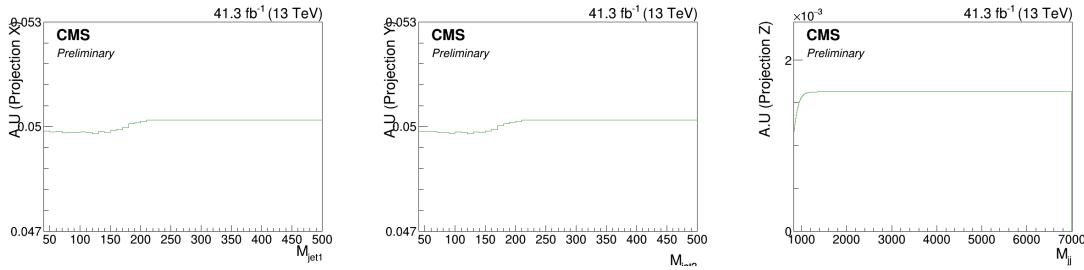


Figure 4.69: One-dimensional projections of the trigger weight histogram for $M_{\text{jet}1}$, $M_{\text{jet}2}$ and M_{VV} respectively.

- ²³⁴⁵
- ²³⁴⁶ M_{jet} and M_{VV} spectra for the lowest mass-point signal sample and for the QCD background
²³⁴⁷ before and after trigger weights are applied, are shown in Figure 4.70 and are compared to data.
²³⁴⁸ > 95% of the signal efficiency is retained, and reweighted QCD simulation agrees well with data.
²³⁴⁹ The modeling of the trigger turn-on was successful and implemented in the background fit method
²³⁵⁰ in the 3D analysis. We found that the method could model the turn-on well. However, when
²³⁵¹ studying the bias on the extracted signal rate for a possible signal in this turn-on region, we
²³⁵² found that this was large due to us attempting to fit a peak on top of a peaky background. As
²³⁵³ we wanted this analysis describing the 3D fit method to become available as soon as possible, we
²³⁵⁴ therefore abandoned modeling of the trigger turn-on for this paper. However, we still wish to
²³⁵⁵ pursue this strategy in the future.

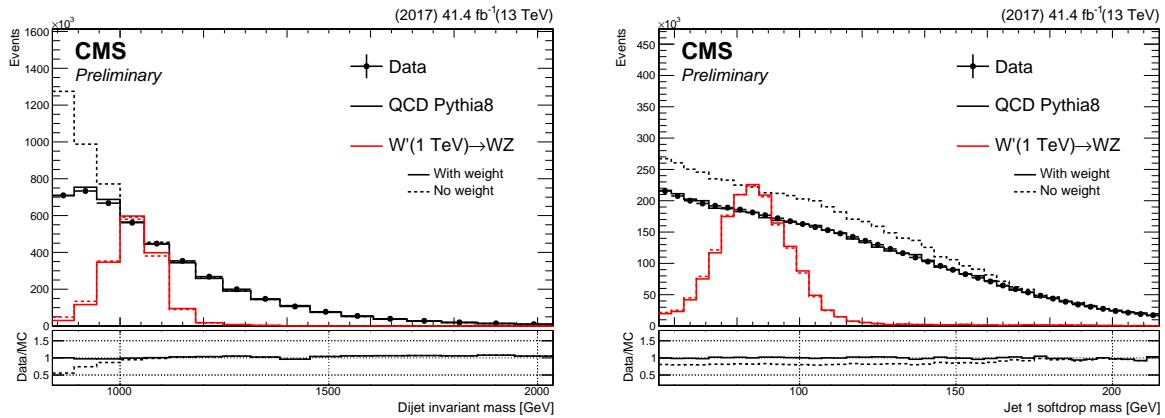


Figure 4.70: The M_{VV} (left) and M_{jet} (right) spectra for signal and background before and after trigger weights are applied.

²³⁵⁶ **4.3.6 A mass and p_T decorrelated tagger**

²³⁵⁷ In order to identify W and Z jet candidates, two algorithms are run on the AK8 PUPPI jet:
²³⁵⁸ softdrop [?] and the N-subjettiness ratio τ_{21} [?]. The softdrop jet mass is used to improve the
²³⁵⁹ mass resolution of the jet, while N-subjettiness serves as a discriminant by yielding a probability
²³⁶⁰ of how compatible the jet is with having N axes. For this search, we require the softdrop-jet mass
²³⁶¹ to be in a window around the W/Z/H/top mass, between 55 and 215 GeV, something which we
²³⁶² plan to extend in the future. In order to improve the statistical power of the jet substructure
²³⁶³ variable τ_{21} and ensure a minimal sculpting of the jet mass as a function of jet p_T , we decorrelate
²³⁶⁴ the variable from the jet softdrop mass and the jet p_T -scale dependence following what was
²³⁶⁵ done in Ref. [?]. This decorrelation is performed by flattening the τ_{21} profile dependence on
²³⁶⁶ $\rho' = \log(m^2/p_T/\mu)$, where $\mu = 1$ GeV. Figure 4.71 shows the profile distribution of τ_{21} as a
²³⁶⁷ function of $\rho' = \log(m^2/p_T/\mu)$, applying the preselections as listed above together with a softdrop
mass cut of $55 \text{ GeV} < M_{\text{jet}} < 215 \text{ GeV}$ (right). A linear transformation is then defined as

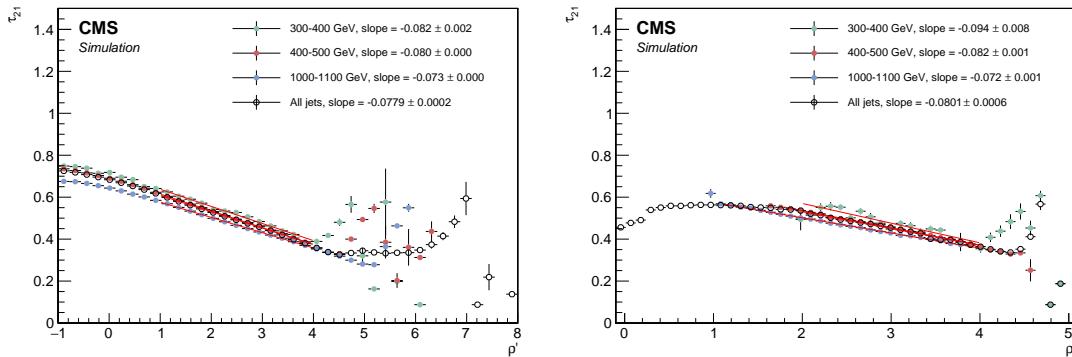


Figure 4.71: Profile distributions of τ_{21} as a function of $\rho' = \log(m^2/p_T/\mu)$, where $\mu = 1$ GeV (bottom), before applying a softdrop mass cut (left) and after applying a softdrop mass cut of $55 \text{ GeV} < M_{\text{jet}} < 215 \text{ GeV}$ (right).

²³⁶⁸

$$\tau_{21}^{DDT} = \tau_{21} - M \times \rho', \quad (4.11)$$

²³⁶⁹ where the slope M is fitted from the linear part of the spectra of the τ_{21} profile versus ρ' with full
²³⁷⁰ selections (bottom left plot). For our purposes, the slope is extracted from fitting the inclusive
²³⁷¹ p_T -spectrum (“All jets”) with a mass window applied as it most closely corresponds to our full
²³⁷² analysis selections. The resulting slope is $M = -0.080$, slightly steeper than than the 2016 value
²³⁷³ of $M = -0.063$ [?]. The profile of the retuned τ_{21}^{DDT} versus ρ' is shown in Figure 4.72, exhibiting
²³⁷⁴ the desired flattened spectra. Working points for τ_{21}^{DDT} are chosen in the following way: First,

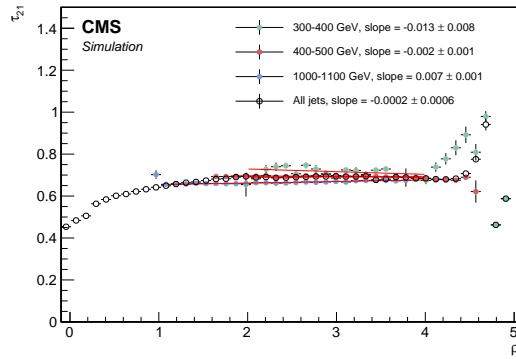


Figure 4.72: Profile distributions of τ_{21}^{DDT} as a function of $\rho' = \log(m^2/p_T/\mu)$, where $\mu = 1$ GeV.

2375 we check which τ_{21}^{DDT} cut corresponds to the highest Punzi significance as a function of the
 2376 resonance mass for different signal samples, shown in Figure 4.73. All other analysis cuts cuts
 have been applied. The cut maximizing the Punzi significance at low resonance mass, where

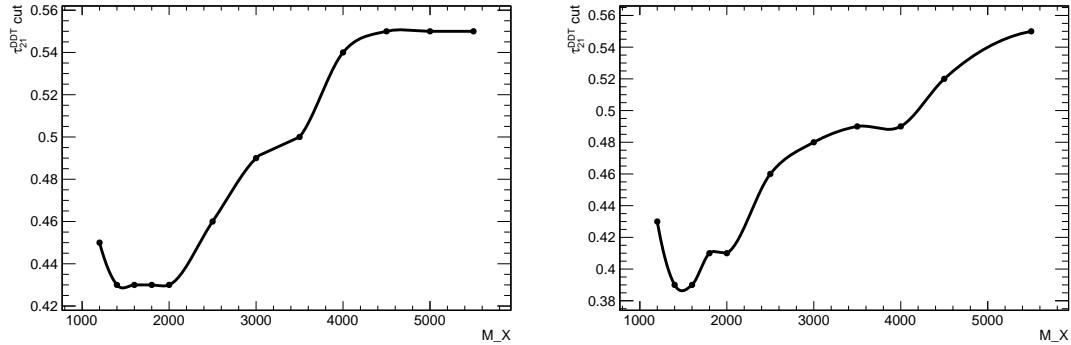


Figure 4.73: The τ_{21}^{DDT} cut corresponding to the highest Punzi significance for a given signal resonance mass, here for Bulk $G \rightarrow WW$ (left) and Bulk $G \rightarrow ZZ$.

2377
 2378 the background is highest, is chosen as the “high purity” (HP) working point. This corresponds
 2379 to $\tau_{21}^{DDT} \leq 0.43$. Second, we find the cut which, together with events falling in the HP region,
 2380 contains at least 95 percent of the signal as well as optimizes the Punzi significance. This is
 2381 found to be $0.43 < \tau_{21}^{DDT} \leq 0.79$, and is classified as the low purity (LP) category. The purpose
 2382 of this category is to enhance the overall sensitivity, especially where the background is low. We
 2383 observe a significant gain in signal efficiency at a fixed mistag rate with the retuned DDT tagger.
 2384 The signal efficiency versus mistagging rate for all three taggers is shown in Figure 4.74, and we
 2385 see the retuned τ_{21}^{DDT} performing better than τ_{21} and the version of τ_{21}^{DDT} using the old tune. In

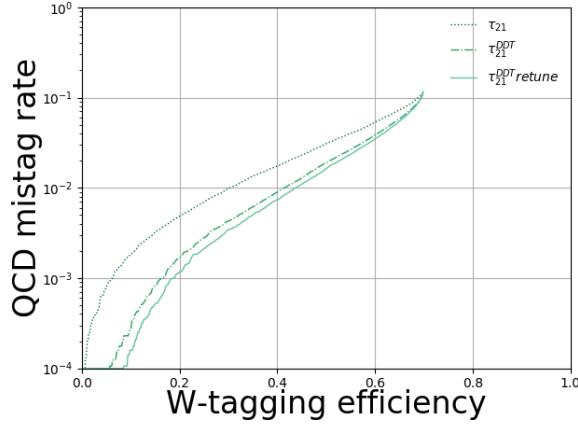


Figure 4.74: Performance of τ_{21} and τ_{21}^{DDT} (2016 and 2017 tune) in the background-signal efficiency plane.

addition to cutting on τ_{21}^{DDT} a loose cut of $\rho = \log(m^2/p_T^2) < -1.8$ is applied. The reason for this is that, while the distribution of ρ is flat as a function of jet transverse momentum for QCD jets, this only holds in the region where perturbative contributions dominate and breaks down at around $\rho = \log(m^2/p_T^2) < -2.0$ due to the AK8 cone size being too small to contain the full jet at high masses. This has a negligible effect on the signal, which mainly peaks around 80 GeV and has a relatively high jet transverse momenta. Figure 4.75 shows the signal and background distribution for the PUPPI softdrop jet mass and τ_{21}^{DDT} . The signal softdrop mass distribution peaks nicely around the W mass, while the multijets background spectrum is peaked at lower softdrop masses. Also, in addition to having a higher signal efficiency for a given mistag rate, τ_{21}^{DDT} has the added benefit of being better modeled in MC than τ_{21} .

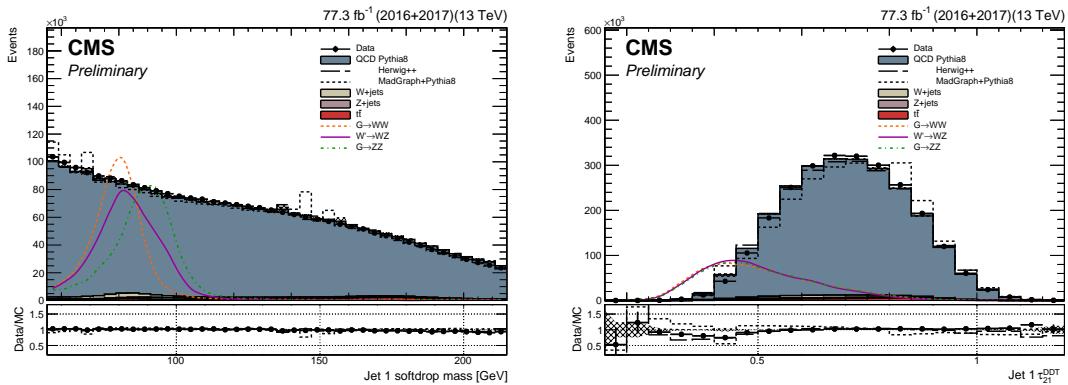


Figure 4.75: PUPPI softdrop jet mass distribution (left) and PUPPI N-subjettiness τ_{21}^{DDT} (right). Signal is scaled with an arbitrary number.

2395 **Data to simulation scale factors**

2396 Following what was done in Section 4.1.7 and 4.2.5, we derive W-tagging scale factors for the
2397 efficiency of the selection on τ_{21}^{DDT} by estimating the ratio of the selection efficiencies on data and
2398 simulation. The PUPPI softdrop mass range is extended to 55 to 215 GeV, and the two purity
2399 categories are

- 2400 • Pass region: $0 < \tau_{21}^{DDT} \leq 0.43 \sim$ high purity
- 2401 • Fail region: $0.43 < \tau_{21} \leq 0.79 \sim$ low purity

2402 The obtained scale factors are listed in Tables 4.13 and 4.14 for 2016 and 2017 data, respectively,
2403 with the corresponding simultaneous fits shown in Fig. 4.76. The jet mass scale and resolution
2404 together with their are estimated in the same fits and also listed in Table 4.14. Two additional
2405 uncertainties are added: one due to generator differences and one due to NNLO corrections.
2406 These are evaluated by comparing the extracted efficiency with and without top p_T reweighting (a
2407 weight derived from data in order to better describe the observed p_T distribution. Calculated for
2408 each top jet as $w = \exp^{0.0615 - 0.0005 * p_{T,top}}$) and when using $t\bar{t}$ simulation produced with different
2409 generators. The scale factors, jet mass scale and jet mass resolution with their total uncertainty
2410 after adding systematics, are listed in Table 4.15. As before, the scale factor is added as a scale
2411 of the signal yield and the jet mass scale and resolution are used to smear MC, and are
2412 additionally inserted as systematic uncertainties in the final fit (scale up/down).

	m [GeV]	σ [GeV]	W-tag efficiency
$\tau_{21}^{DDT} < 0.43$			
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 4.13: Jet mass scale, jet mass resolution and τ_{21}^{DDT} scale factors as evaluated in the full 2016 Single Muon dataset.

2413

2413 **p_T dependence**

2414 As the τ_{21}^{DDT} working point used for this search is so tight, the statistics when evaluating data to
2415 simulation scalefactors are very low in the pass category. A p_T -binned measurement has therefore

$\tau_{21}^{DDT} < 0.43$	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 4.14: Jet mass scale, jet mass resolution and τ_{21}^{DDT} scalefactors as evaluated in the full 2017 Single Muon dataset.

	SF $\pm \sqrt{\text{Stat.} + \text{Sys}_{\text{Generator}} + \text{Sys}_{\text{NNLO}}}$	SF \pm 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 4.15: Final jet mass scale, jet mass resolution and τ_{21}^{DDT} scalefactors.

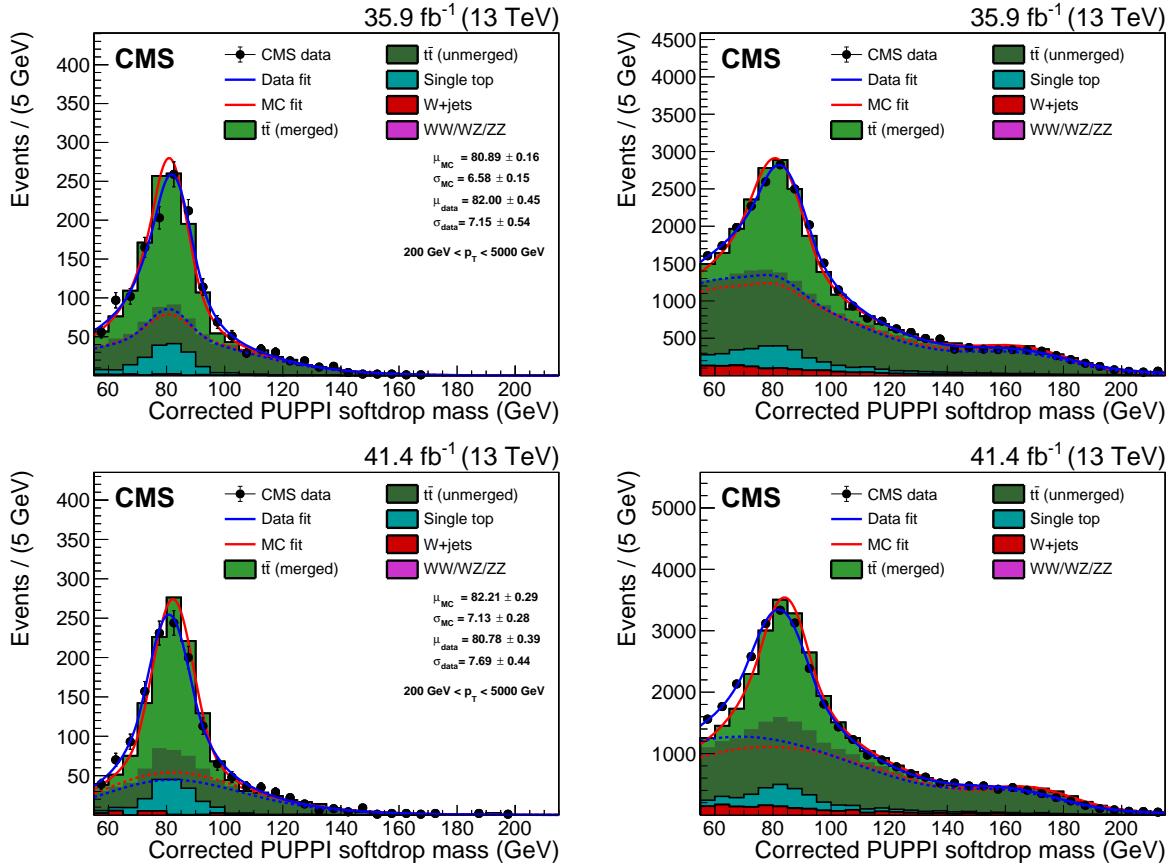


Figure 4.76: PUPPI softdrop jet mass distribution that pass (left) and fail (right) the $\tau_{21}^{DDT} 0.43$ selection in the $t\bar{t}$ control sample. The result of the fit to data and simulation are shown by the solid blue and solid red line, respectively. The background components of the fit are 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.

not been possible. In order to get a feeling for how W-tagging efficiency, PUPPI softdrop jet mass scale and mass resolution scale factors change with jet transverse momentum, we do a measurement in 4 different p_T bins using the looser PUPPI softdrop + τ_{21} tagger. The systematics are evaluated the same way as above (one due to top p_T reweighting and one comparing different $t\bar{t}$ samples). Figure 4.77 shows the extracted W-tagging efficiency for data (black markers) and for simulation (red markers) using a PUPPI softdrop + $\tau_{21} < 0.4$ based tagger 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 4.16. With the observation

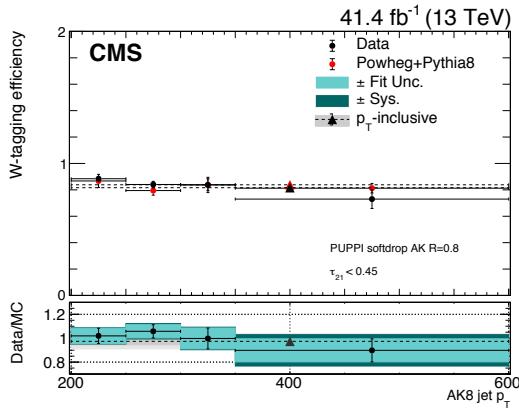


Figure 4.77: The W-tagging efficiency in data (black circles) and in simulation (red circles) as a function of jet p_T . The p_T -inclusive measurement is marked with triangles. The lower panel shows the efficiency in data divided by the efficiency in simulation, corresponding to the W-tagging uncertainty. The blue bands mark the fit and systematic uncertainties.

Bin	SF $\pm \sqrt{\text{Stat.} + \text{Sys}_{\text{Generator}} + \text{Sys}_{NNLO}}$	SF \pm 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 4.16: The data to simulation scalefactor scale factor for the PUPPI softdrop + τ_{21} based tagger in bins of jet p_T . All scalefactors are compatible with unity.

of this clear trend of an uncertainty increase as a function of p_T , we evaluate a p_T -dependent W-tagging scalefactor uncertainty in the following way: Using signal Monte Carlo generated with two different shower generators, PYTHIA8 and HERWIG++, we compute the difference in tagging efficiency between the two 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

$$\sigma_{p_T, \text{Bin}=i} = \frac{\left(\frac{\epsilon_{\text{PYTHIA}}}{\epsilon_{\text{HERWIG}}} p_{T,\text{Bin}=i}\right)}{\left(\frac{\epsilon_{\text{PYTHIA}}}{\epsilon_{\text{HERWIG}}}\right)_{500 \text{ GeV}}} \quad (4.12)$$

This parametrization is then applied as a growing uncertainty on the signal yield as a function of resonance mass, where $p_{T,\text{Bin}=i} = M_{X,i}/2$, due to an uncertainty on the W-tagging efficiency. In contrast with what was found for the τ_{21} based tagger (Section 4.1.7), where this uncertainty grew logarithmically with p_T , we find that the corresponding double ratio stabilizes around 1 TeV

²⁴³⁶ when using τ_{21}^{DDT} and is better described by a sigmoid function. The resulting parametrization is shown in Figure 4.78.

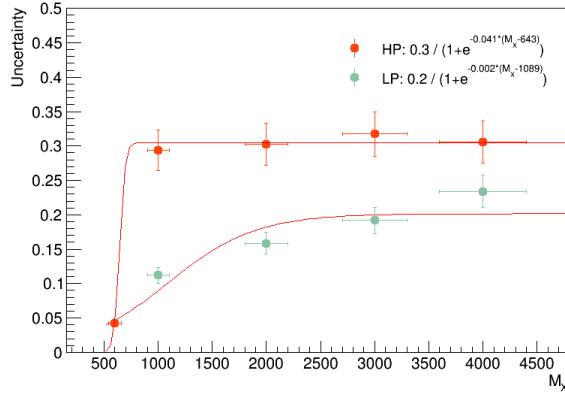


Figure 4.78: The parametrized uncertainty on W-tagging efficiency as a function of resonance mass ($2 \times p_T$) extracted using the difference in tagging efficiency between PYTHIA and HERWIG++ Monte Carlo relative to the difference at 500 GeV

²⁴³⁷

²⁴³⁸ In addition to measuring the tagging p_T dependence we also extract the change in PUPPI
²⁴³⁹ softdrop jet mass scale and resolution as a function of p_T , as this should be roughly the same
²⁴⁴⁰ independent of whether τ_{21} or τ_{21}^{DDT} is used. We find that the jet mass scale ranges between 0.5
²⁴⁴¹ and 2.5%, and the jet mass resolution between 4 and 10%, the latter measurement not being
²⁴⁴² statistically significant as the uncertainties are large, around $\sim 10\%$). We therefore use a fixed
²⁴⁴³ uncertainty of 2 and 10% for the PUPPI softdrop jet mass scale and resolution, respectively,
²⁴⁴⁴ which should be sufficient to cover a broadening and a shift at high p_T .

2445 4.3.7 The multidimensional fit

2446 As mentioned in the introduction to this chapter, the three-dimensional fit method takes advantage
 2447 of the fact that the signal peaks in three dimensions; dijet invariant mass (M_{VV}) and the jet
 2448 groomed mass of jet 1 and jet 2 (M_{jet1} and M_{jet2}) and attempt to extract the signal from the
 2449 three dimensional $M_{jet1}(x)$ - $M_{jet2}(y)$ - $M_{VV}(z)$ plane. In order to do so, four different types of PDFs
 need to be created in order to have a complete model:

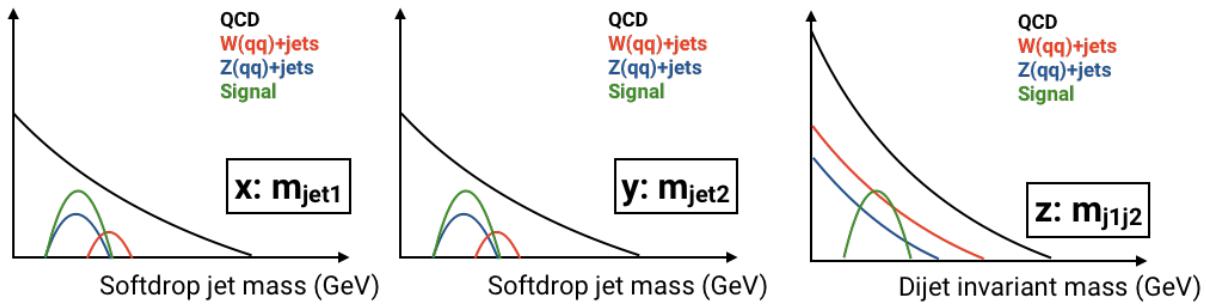


Figure 4.79: An illustration of the shape of the signal and the relative background contributions in the three relevant dimensions $M_{jet1}(x)$, $M_{jet2}(y)$ and $M_{VV}(z)$.

2450

- 2451 • **Signal 3D PDF:** Resonant in x, y and z. Parametrized as function of the resonance mass
 2452 M_X
- 2453 • **Non-resonant background:** Non-resonant in x, y and z and dominant background.
 2454 Created through a forward folding kernel approach
- 2455 • **Resonant background:** Mainly $W/Z+jets$ (some $t\bar{t}$). Resonant in x and y, smoothly
 2456 falling in z.
- 2457 • **Alternative shapes:** 5 additional shape uncertainties implemented through vertical
 2458 morphing

2459 These are illustrated in Figure 4.79 and will be described in detail in the following.

2460 4.3.8 Modeling of the signal

2461 The signal shape in three dimensions is defined as a product of the shape of the resonance mass
 2462 and the jet masses:

$$P_{sig}(M_{VV}, M_{jet1}, M_{jet2} | \theta(M_X)) = P_{VV}(M_{VV} | \theta_1(M_X)) \times P_{j1}(M_{jet1} | \theta_2(M_X)) \times P_{j2}(M_{jet2} | \theta_2(M_X)). \quad (4.13)$$

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 signal is parametrized by fitting the resonance mass and jet mass line shapes for each mass point, 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 4.80 shows the derived parameters and interpolation as a function of resonance mass. The final M_{VV} shapes as extracted from the parametrization are shown in Figure 4.81.

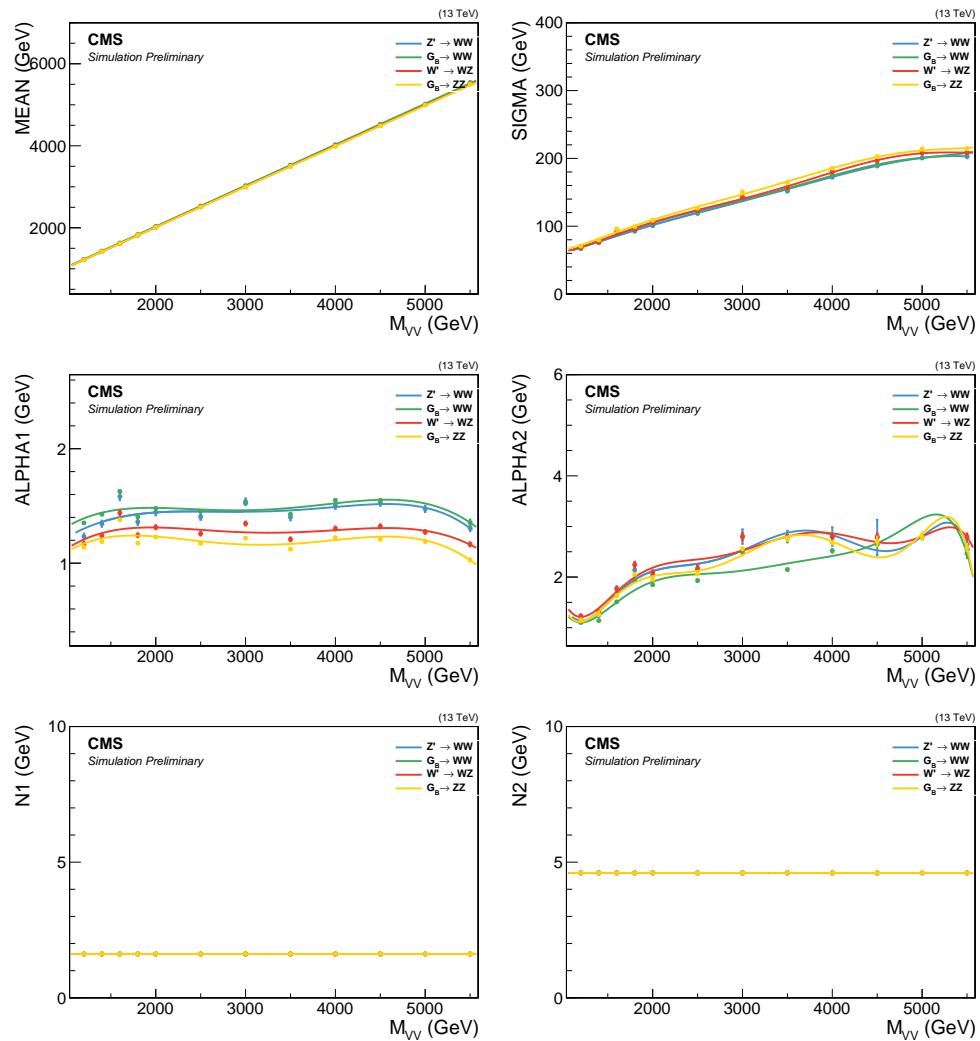


Figure 4.80: The interpolated Crystal-ball parameters for the dijet invariant mass as a function of M_X . The small variations for ALPHA2 have been shown to have no effect on the overall modeling.

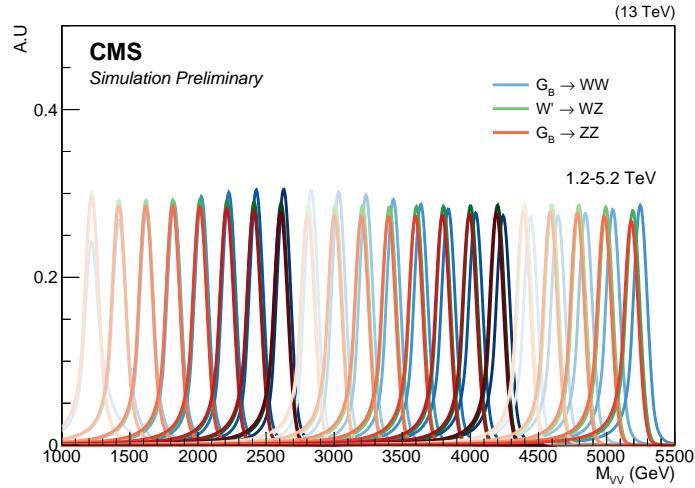


Figure 4.81: Final M_{VV} signal shapes extracted from the parametrization. Here for a G_{bulk} decaying to WW (blue) and ZZ (red) and for a W' decaying to WZ (green).

2471 The same procedure is used to model the jet mass: The $M_{\text{jet}1}$ spectrum for each resonance
 2472 mass hypothesis is fitted using a double Crystal-ball function, the fitted parameters are extracted
 2473 and interpolated as functions of the resonance mass. This is done separately for $M_{\text{jet}1}$ and $M_{\text{jet}2}$.
 2474 The fitted parameters and interpolations are shown in Figure 4.82 for $M_{\text{jet}1}$, the corresponding
 2475 distributions for $M_{\text{jet}2}$ are in Appendix C.1. The final $M_{\text{jet}1}$ shapes as extracted from the
 2476 parametrization are shown in Figures 4.83. Finally, the signal yield is parametrized as a function
 2477 of the resonance mass. For each mass point M_X and each purity category, the signal yield per
 2478 picobarn of cross section is calculated as the integral of the Monte Carlo histogram. The yields
 2479 are then interpolated as a function of M_X . The signal efficiency as a function of resonance mass
 2480 is shown in Figure 4.84.

2481 4.3.9 Modeling of the non-resonant background

2482 In order to model the QCD multijets background in the three-dimensional M_{VV} - $M_{\text{jet}1}$ - $M_{\text{jet}2}$
 2483 plane, we use the following conditional product:

$$P(M_{VV}, M_{\text{jet}1}, M_{\text{jet}2}) = P_{VV}(M_{VV}|\theta_1) \times P_{cond,1}(M_{\text{jet}1}|M_{VV}, \theta_2) \times P_{cond,2}(M_{\text{jet}2}|M_{VV}, \theta_2). \quad (4.14)$$

2484 This probability density requires a computation of the conditional two-dimensional shapes of
 2485 $M_{\text{jet}1}/M_{\text{jet}2}$ given M_{VV} , as well as a one dimensional shape of the M_{VV} distribution.

2486 The following fit range and binning is used for the three axes: $M_{\text{jet}1}/M_{\text{jet}2}$ is fitted from 55
 2487 to 215 GeV using 2 GeV bins. M_{VV} is fitted from 1126 to 5500 GeV. The lower bound is chosen

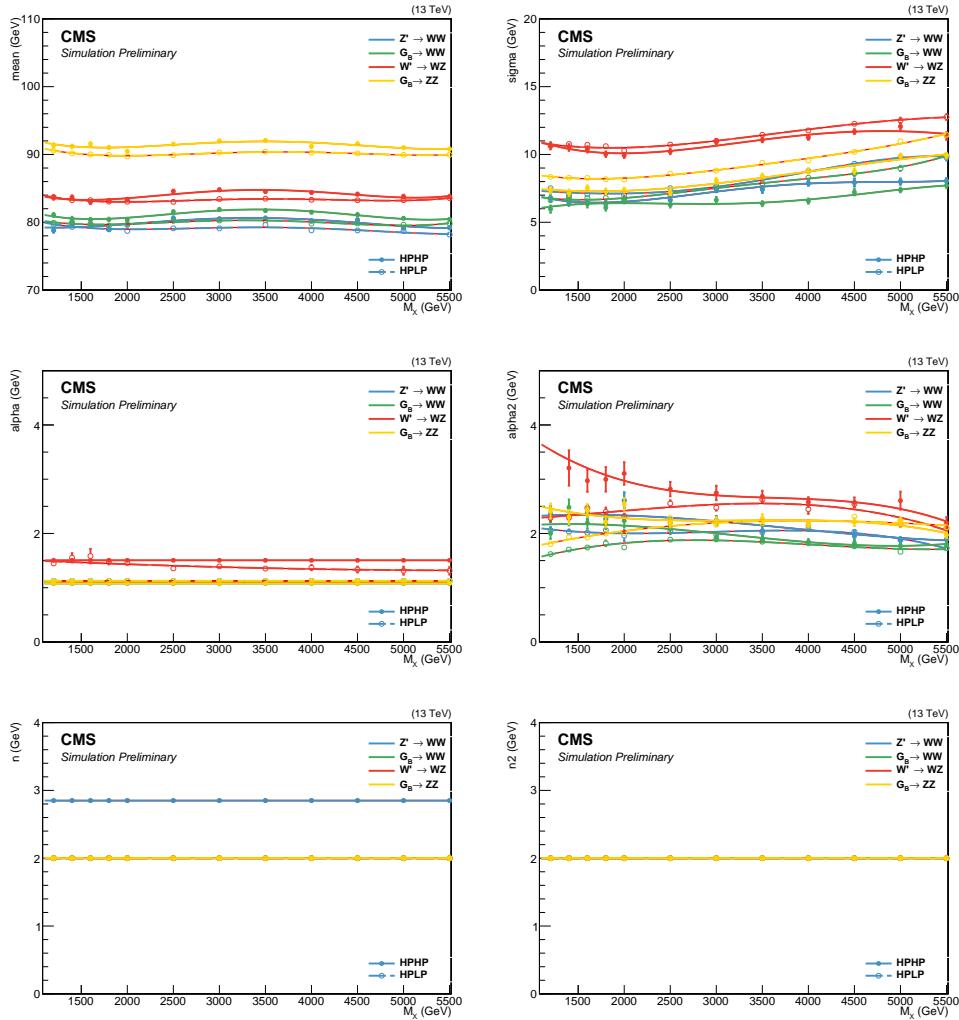


Figure 4.82: The interpolated double Crystal-ball parameters for the softdrop jet mass as a function of M_X . To improve the stability of the fit some parameters are set constant. Here for jet 1.

such that to avoid complications in the fitting procedure due to trigger turn-on effects, while the upper bound is chosen considering the highest dijet invariant mass found in data as well as avoiding mis-reconstruction effects at very large m_{jj} and low jet masses. For M_{VV} , the “dijet binning” is used. This binning corresponds to the actual dijet mass resolution and is, in units of GeV:

2492

2493 Dijet binning = 1126, 1181, 1246, 1313, 1383, 1455, 1530, 1607, 1687, 1770, 1856, 1945,
 2494 2037, 2132, 2231, 2332, 2438, 2546, 2659, 2775, 2895, 3019, 3147, 3279, 3416, 3558, 3704,

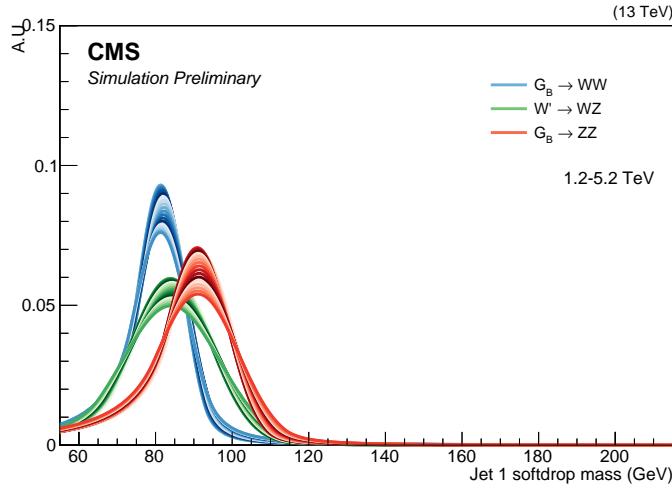


Figure 4.83: Final M_{jet} signal shapes extracted from the parametrization for a G_{bulk} decaying to ZZ, a G_{bulk} decaying to WW and for a W' decaying to WZ.

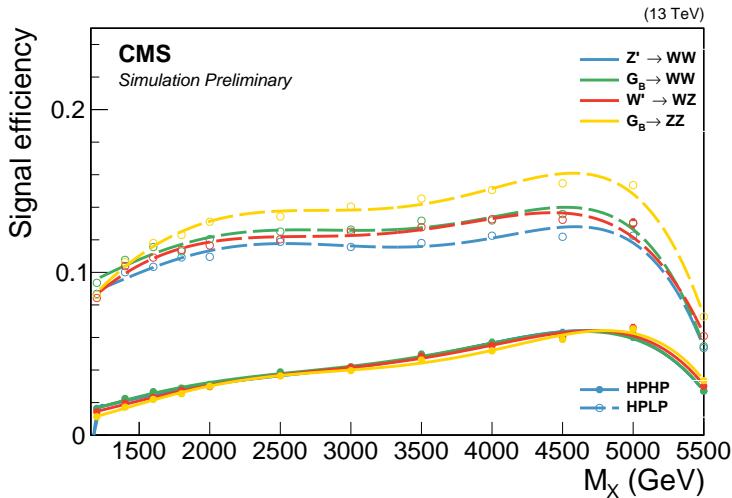


Figure 4.84: Signal efficiency as a function of resonance mass.

2495 3854, 4010, 4171, 4337, 4509, 4686, 4869, 5058, 5253, 5500

2496

2497 The background model is built starting from simulation and we encode sufficient nuisance
 2498 parameters into the fit, allowing the shape to adapt itself to data. For this we use a “forward-
 2499 folding” approach. For each MC event in the two(one)-dimensional $M_{\text{vv}}-M_{\text{jet}}$ (M_{jet}) space, a
 2500 2D(1D) Gaussian kernel is built starting from generator level quantities. Each of these Gaussians

then contribute to the total probability density of the final two(one)-dimensional probability density functions

First, the resonance mass and softdrop jet mass scale and resolution are derived. For this we use the anti- k_T generated jet collection matched to jets identified as V-jets on reconstruction level. We then derived the M_{jet} and M_{VV} scale and resolution from a Gaussian fit to $M_i(\text{reco})/M_i(\text{gen})$ ($i = M_{\text{jet}}$ or $i = M_{\text{VV}}$) in bins of generator jet p_T . Figure 4.85 shows the fit to $M_{\text{jet}}(\text{reco})/M_{\text{jet}}(\text{gen})$ (left) and $M_{jj}(\text{reco})/M_{jj}(\text{gen})$ (right) for an arbitrary bin. The Gaussian mean yields the mass scale and the Gaussian width the mass resolution. The resonance mass and softdrop jet mass

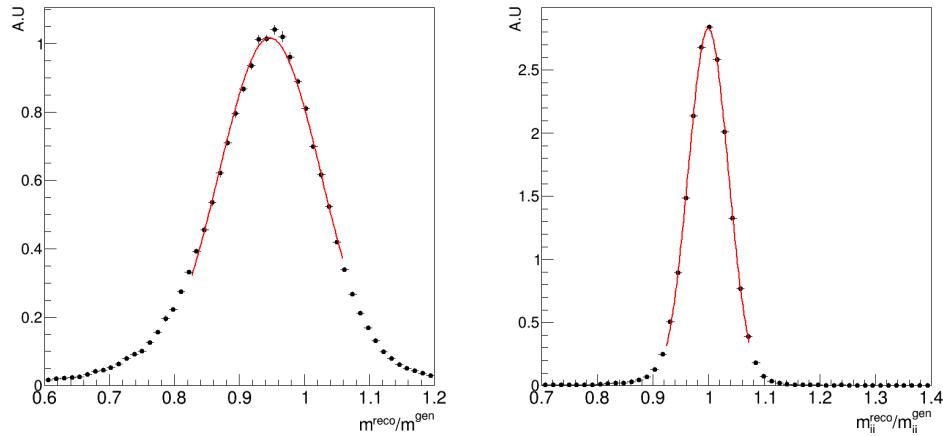


Figure 4.85: Fit to $M_{\text{jet}}(\text{reco})/M_{\text{jet}}(\text{gen})$ (left) and $M_{jj}(\text{reco})/M_{jj}(\text{gen})$. The mass resolution is taken as the width of the fitted Gaussian, while the Gaussian mean yields the mass scale.

scale and resolution as a function of generator jet p_T is shown in Figure 4.86. The projection of these resolution functions are shown in Figure 4.87. The 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

$$k(M_{\text{jet}}, M_{\text{VV}}) = \frac{w_i}{\sqrt{2\pi}r_{M_{\text{VV}},i} \cdot r_{M_{\text{jet}},i}} \exp \left(-\frac{1}{2} \left(\frac{M_{\text{VV}} - s_{M_{\text{VV}},i}}{r_{M_{\text{VV}},i}} \right)^2 - \frac{1}{2} \left(\frac{M_{\text{jet}} - s_{M_{\text{jet}},i}}{r_{M_{\text{jet}},i}} \right)^2 \right), \quad (4.15)$$

where s_i, r_i are the scale and the resolution derived in the previous step and w_i is the event weight product (e.g PU, cross sections etc.). The resulting kernel values are 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 one-dimensional Gaussian kernel only depending on M_{VV} . The templates are then added together to form a three-dimensional PDF. Finally, we fit this 3D PDF to QCD

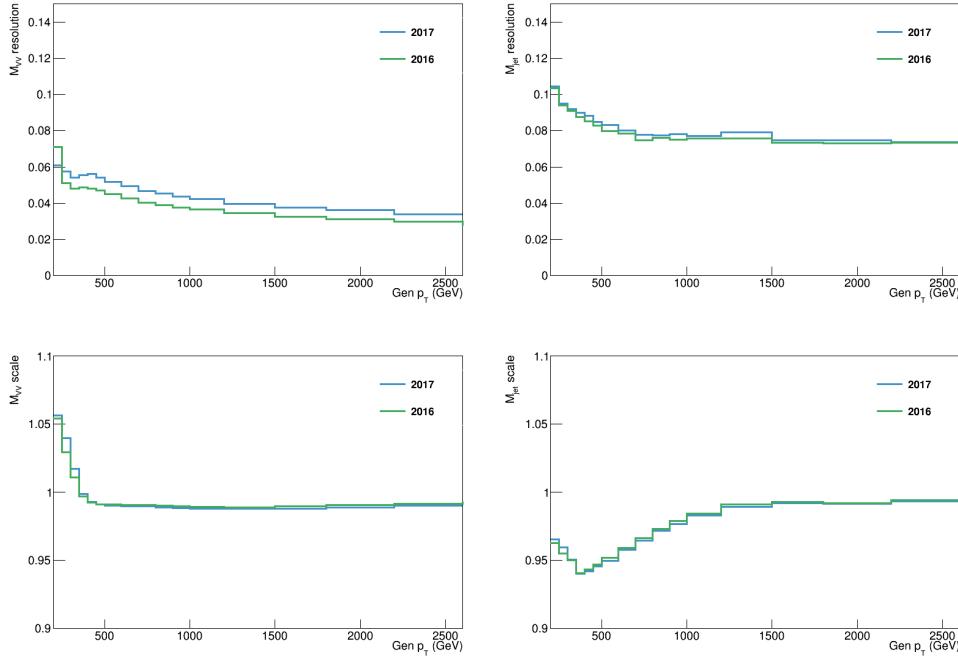


Figure 4.86: Resolution (top) and scale (bottom) for M_{VV} (left) and the M_{jet} (right) as a function of generator jet p_T .

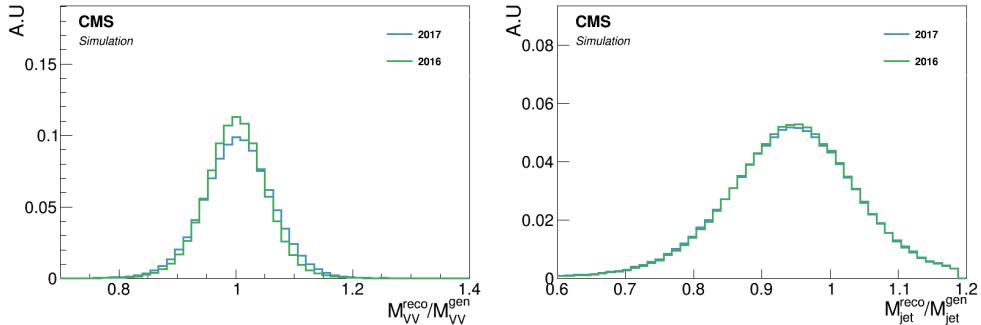


Figure 4.87: Projections of the resolution functions for all generator jet p_T bins for M_{VV} (left) and M_{jet} (right).

2519 MC in order to remove any residual bias (mainly at the extreme ends of the spectra). The result
 2520 is a full, smooth shape replacing the prediction from simulation.

2521 As the high purity category is limited by statistics, we rather build this template starting
 2522 from the low purity templates. This is done by fitting the low purity 3D template to QCD MC in
 2523 the high purity category. Figure 4.88 show the final templates (solid lines) together with the QCD

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

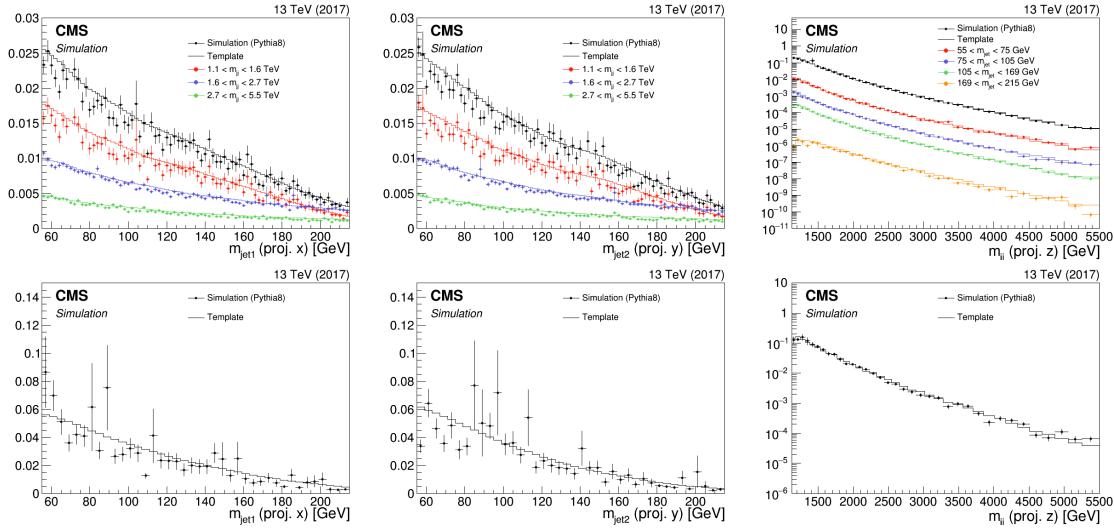


Figure 4.88: Comparison between QCD MC simulation (markers) and kernels derived from generator level quantities (lines) for the HPHP (top) and HPLP (bottom) categories using 2017 MC. The kernels are shown for $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{vv} (right).

In order to validate the kernel transfer method, we check that we can fit a higher-statistics 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, without being camouflaged by large error bars. The resulting kernel versus MC spectra are shown in Figure 4.89. Good closure is observed in all three dimensions, demonstrating that the HPLP kernels adapt well to the HPHP MC data points even when statistics are sufficient, and we consider the method sound.

2535

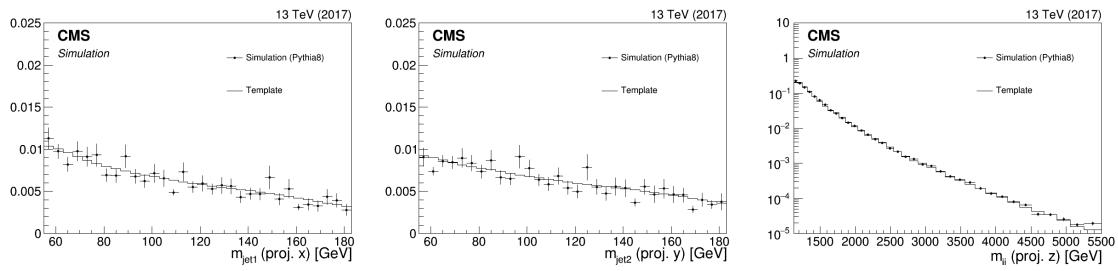


Figure 4.89: Comparison between QCD MC simulation (markers) and kernels derived from generator level quantities (lines) in the HPHP category, using a looser cut on τ_{21}^{DDT} .

²⁵³⁶ **4.3.10 Modeling of the resonant background**

In addition to the QCD multijet background, there are a few sub-dominant processes to consider which contain one real vector boson and at least one QCD-jet. These are W+jets, Z+jets and events from $t\bar{t}$ processes. They are resonant around the W/Z mass in the two softdrop jet mass dimensions and must therefore be treated differently than the non-resonant QCD background. Figure 4.90 shows the projections on M_{jet1} (left), M_{jet2} (middle) and M_{VV} (right). As the jets are

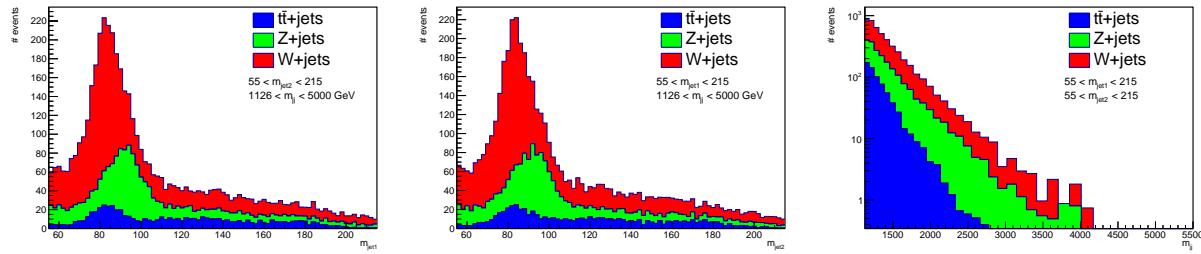


Figure 4.90: Projections of the sub-dominant backgrounds on the jet mass axes M_{jet1} (left) and M_{jet2} (middle), as well as on the dijet invariant mass M_{VV} (right). Here in the low purity category.

randomly sorted, each jet mass dimension contains two contributions: A resonant part consisting of real vector boson jets, peaking around the V 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 as we know that the non-resonant part of the jet mass spectrum is correlated with the dijet invariant mass (like QCD), while the resonant part is not. We additionally want to encode the fact that we know these backgrounds in reality only peaks in one jet mass axis (there is only one real vector boson) by requiring the PDFs to consist of a resonant part on one axis, and a non-resonant part on the other axis. A three dimensional PDF for the sub-dominant backgrounds is built as a product of three one dimensional pdf's as follows:

$$P_{res}(M_{jet1}, M_{jet2}, M_{VV}) = P_{VV}(M_{VV}) \times P_{jet1}(M_{jet1}, M_{jet2}) \times P_{jet2}(M_{jet2}, M_{jet1}) \quad (4.16)$$

where

$$P_{jet1}(M_{jet1}, M_{jet2}) = f \times R(M_{jet2}) \times P_{res}(M_{jet1}) + (1 - f)P_{non-res}(M_{jet1}) \quad (4.17)$$

$$P_{jet2}(M_{jet2}, M_{jet1}) = (1 - f) \times R(M_{jet1}) \times P_{res}(M_{jet2}) + fP_{non-res}(M_{jet2}) \quad (4.18)$$

2537 Here, $R(M_{\text{jet}})$ parametrizes the correlation between $M_{\text{jet}1}$ and $M_{\text{jet}2}$ and f is a fit parameter that
 2538 is used to adjust the fraction of real vector boson jets in $M_{\text{jet}1}$ compared to $M_{\text{jet}2}$. Its value can
 2539 vary 10% around $f = 0.5$, the expected median when using a random jet sorting.
 2540 As the contribution of the $t\bar{t}$ background is much smaller than the one coming from V+jets (less
 2541 than 2%), it is modeled together with the W+jets contribution. Therefore only two PDFs are
 2542 built for the resonant background: One for the W+jets and $t\bar{t}$, and one for Z+jets.
 2543 The available MC statistics in the high purity category is very low and therefore the parametrization
 2544 of the resonant background is done for the low purity category only, and the resulting shapes are
 2545 then used for both purity categories. The uncertainties for the different purity categories are,
 2546 however, included as separate nuisance parameters in the fit.

2547 The non-resonant M_{VV} PDF is constructed using the same kernel approach as is used to
 2548 model the QCD multijet background with one minor difference: Due to the low statistics in the
 2549 high- M_{VV} tail, an additional smoothing of the jet mass tail using a leveled exponential

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/s)^{P_1}}{(m_{jj}/s)^{P_2}} \quad (4.19)$$

2550 is performed. Here, s is the center of mass energy. The function is fitted to the spectrum starting
 2551 from a dijet mass of 1.1 (2.1) TeV for high purity (low purity). Two uncertainties on this shape is
 2552 added in order to accommodate for MC mis-modeling due to higher order QCD and electroweak
 2553 corrections: One proportional to M_{VV} and one proportional to $1/M_{\text{VV}}$. The resulting M_{VV}
 2554 kernels (solid lines) for the W+jets background are shown in Figure 4.91 and are compared to
 2555 MC simulation (markers). The blue line corresponds to the nominal shape, while the red and
 2556 green lines correspond to the uncertainties proportional to M_{VV} and $1/M_{\text{VV}}$, respectively. The
 2557 corresponding distributions for the Z+jets background are shown in Appendix C.3.

As mentioned above, the modeling of the M_{jet} spectrum is split into two different PDFs: One describing the resonant and one describing the non-resonant part. We distinguish between the two by requiring the resonant contribution to consist of jets matched to a generated boson within a cone of $\Delta R = 0.8$. A double-sided crystal-ball function is then fitted to the resonant spectrum for W+jets and Z+jets separately. This allows us to fully correlate uncertainties on the mean and width of the M_{jet} distribution for signal with the corresponding values for W/Z+jets, as these uncertainties affect all jets generated from real vector bosons in the same way. This effectively gives us a way of constraining these parameters directly from data. The parametrization of the resonant part of the jet mass spectrum for $M_{\text{jet}1}$ is shown in Figure 4.92 for W+jets and $t\bar{t}$ (left) and Z+jets. The small enhancement around 170 GeV is caused by fully merged top jets and is so small (< 2% in the low purity and < 0.5% in the high purity category) that we do not take

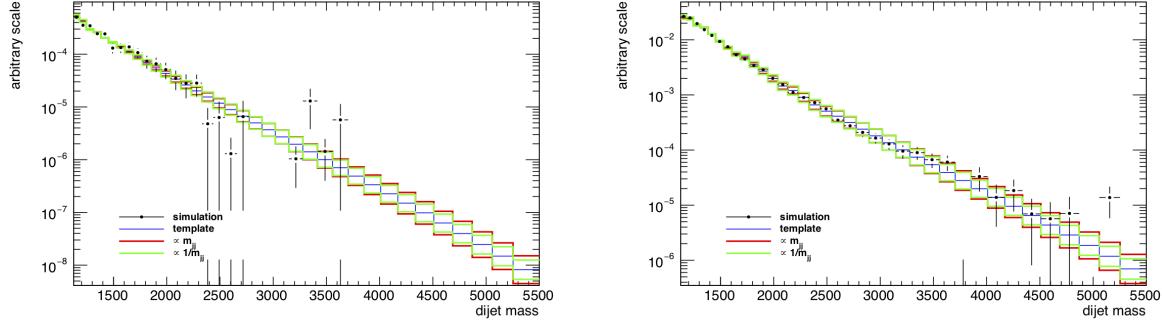


Figure 4.91: One dimensional M_{VV} kernels (solid line) compared to MC (markers) for the $W+jets$ background in the HPHP (left) and HPLP (right) categories. The blue line corresponds to the nominal shape, while the red and green lines correspond to uncertainties proportional to M_{VV} and $1/M_{VV}$, respectively.

it into account in the final PDF. The non-resonant part of the $V+jets$ background is modeled

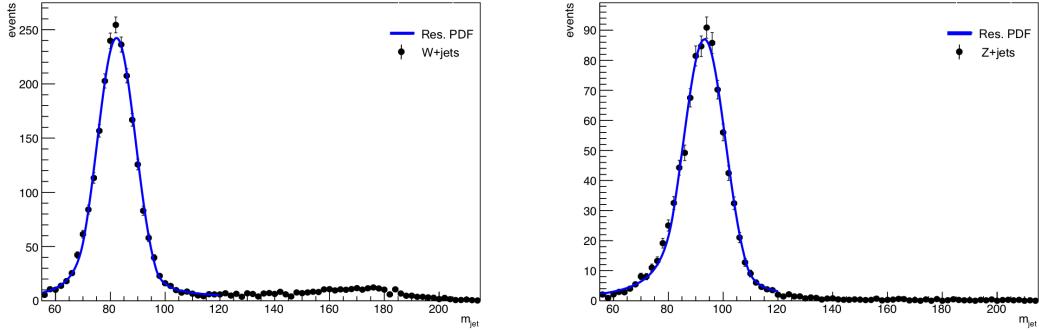


Figure 4.92: Fit to the resonant part of the $V+jets$ spectrum for $W+jets$ and $t\bar{t}$ (left) and $Z+jets$ (right).

using a simple Gaussian fit to the spectrum of jets not matched to a real vector bosons, as shown in Figure 4.93. For the resonant modeling, correlations between the jet mass M_{jet} and dijet invariant mass M_{VV} have been found small enough to be neglected (the jet mass spectrum does not, within statistical uncertainties, depend on the jet p_T). However, there is a strong correlation between the softdrop jet mass of each jet due to the fact that when one of the two jets originates from a real boson peaking around the V mass, the other is bound to be a quark jet. Therefore, the fraction of real boson jets contained in M_{jet1} affects the fraction of real V jets in M_{jet2} . To account for this, the fraction of real V jets versus quark/gluon jets jets, is parametrized as a

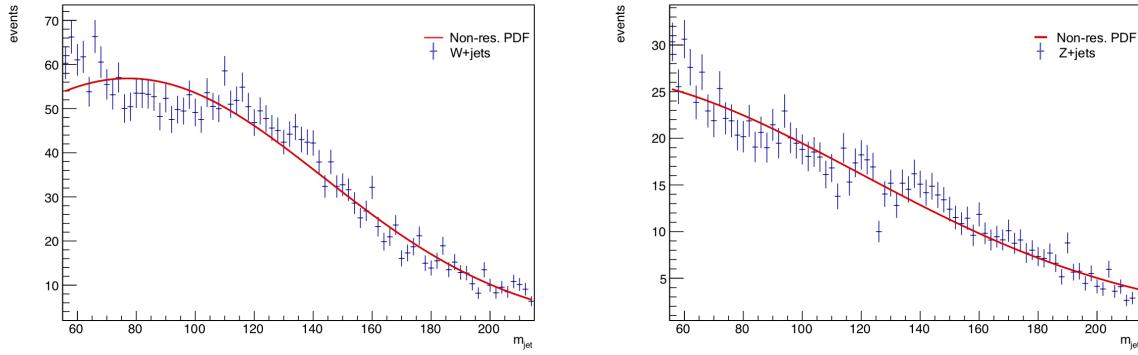


Figure 4.93: Fit to the non-resonant part of the $V+jets$ and $t\bar{t}$ spectrum for $W+jets$ and $t\bar{t}$ (left) and $Z+jets$ (right).

function of the jet mass. As the jets are randomly sorted, we define the parametrization as

$$R(M_{jet}) = \frac{N_{res,jet1}(M_{jet2}) + N_{res,jet2}(M_{jet1})}{N_{non-res,jet1}(M_{jet2}) + N_{non-res,jet2}(M_{jet1})}. \quad (4.20)$$

Where $N(M_{jet})$ denotes the number of events within a given M_{jet} window. The resulting ratio is then fitted with a polynomial function, as shown in Figure 4.94. As a closure test, the

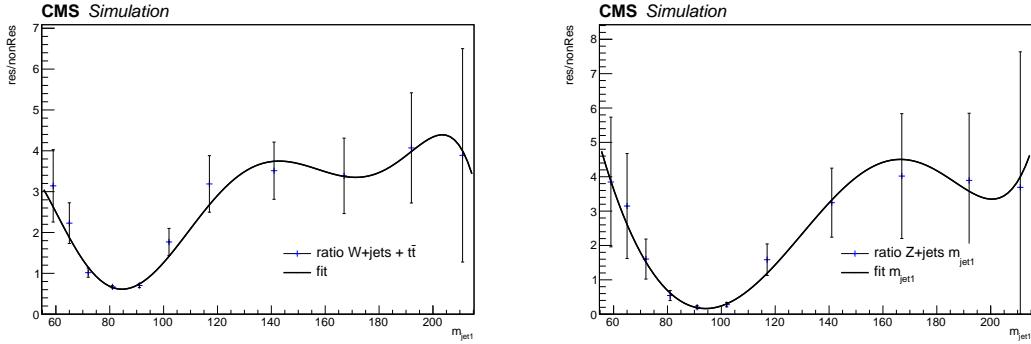


Figure 4.94: Ratio of resonant to non-resonant events in $W+jets$ (left) and $Z+jets$ (right) as a function of jet mass.

three dimensional kernel as defined in Equation 4.16 is fitted to the $V+jets$ and $t\bar{t}$ simulation. Figure 4.95 shows the fitted kernel (red) together with the MC data points (black markers) in the low (top) and high (bottom) purity 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 dominated with very little statistics and is completely swallowed up by the

QCD multijets background which has the same shape.

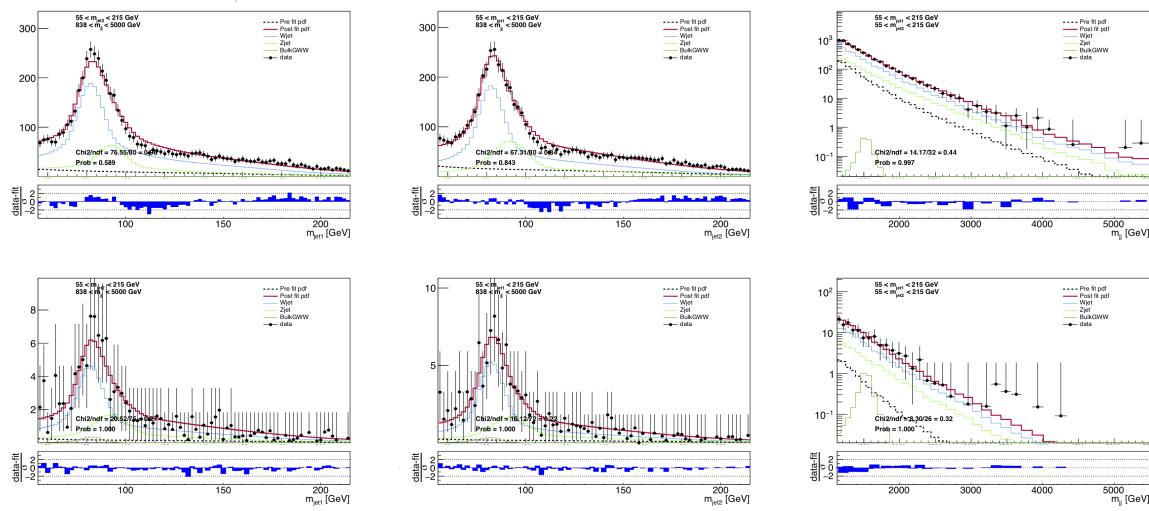


Figure 4.95: Final fits of the complete background model (red line) to the MC simulation of the V+jets backgrounds (black markers) for the high purity category.

2566 4.3.11 Systematic uncertainties

2567 Systematic uncertainties are inserted as nuisance parameters in the fit and affect the normalization
 2568 and shape of both signal and background.

2569 Signal normalization uncertainties

2570 As all background contributions are data driven, the largest systematic uncertainties affect only
 2571 the signal.

- 2572 • **luminosity** 2.6%(2.3%) for the 2016(2017) dataset
- 2573 • τ_{21}^{DDT} **efficiency** 1-12% systematic uncertainty on the signal efficiency scale factor for the
 2574 τ_{21}^{DDT} selection (listed in Table 4.15). Anti-correlated between the HP and LP categories.
- 2575 • τ_{21}^{DDT} **p_T extrapolation** An additional uncertainty arising from the extrapolation of
 2576 the V-tagging efficiency scale factors towards higher transverse momenta. Treated as
 2577 correlated between the τ_{21}^{DDT} categories and is given as $3.9\% \times \ln(p_T/200 \text{ GeV})$ and
 2578 $8.5\% \times \ln(p_T/200 \text{ GeV})$ for the LP and HP regions, respectively.
- 2579 • **PDF and factorization/renormalization scale** 2% uncertainty on the signal accep-
 2580 tance due to choice of PDFs and factorization/renormalization scale.

2581 Background normalization uncertainties

2582 The QCD multijet background has a poorly known cross section and is therefore allowed to
 2583 float within 50% of the yield expected by simulation. For the resonant V+jets background, the
 2584 following uncertainties on the normalization is added: One due to cross section ($\sim 5\%$), one due
 2585 to NLO QCD corrections ($\sim 10\%$) and one due NLO electroweak corrections ($\sim 15 - 35\%$).

2586 Signal and resonant background shape uncertainties

2587 There are two shape uncertainties correlated between the signal and V+jets background: System-
 2588 atics due to jet mass scale (JMS) and jet mass resolution (JMR). These are evaluated in a $t\bar{t}$
 2589 control region, listed in Table 4.15, and affect the mean and width of the jet mass PDFs. Three
 2590 additional shape uncertainties are added for the signal: Uncertainties due to jet energy scale
 2591 (JES), jet energy resolution (JER) and PDF. The uncertainty due to JES/JER is evaluated by
 2592 reweighting the transverse momentum of each MC event in signal MC, then fitting a Gaussian
 2593 to the dijet invariant mass spectrum and calculating the change in the mean and variance with

respect to no reweighting. These are then used as uncertainties on the signal M_{VV} shape (2% for the dijet mass scale, 5% for the resolution).

Uncertainties due to the PDF are evaluated by reweighting each event according to ≈ 100 PDF variations. The impact on the signal shapes are again evaluated by fitting a Gaussian to the m_{jet} and m_{jj} distributions before and after changing PDF weights. The obtained variation is < 1% for the 2017 signal MC, but slightly higher in 2016. An overall uncertainty on the acceptance of 3% is therefore applied.

Non-resonant background shape uncertainties

Alternate shapes for the non-resonant background components are added to the fit through vertical template morphing. This creates nuisance parameters for each additional shape, allowing the derived nominal 3D PDF to vary within the respective nuisances to match the data. Shape uncertainties that simultaneously affect all three dimensions are used, totaling to 5 to take all possible effects into account. The first effect accounts for variation of the underlying transverse momentum spectrum and is obtained through an alternate shape produced by varying the jet masses M_{jet} and dijet mass M_{VV} by a quantity proportional to M_{VV} and M_{jet} . The second effect considered is a variation of the scale, where the corresponding alternate shape is obtained by simultaneously varying M_{jet} and M_{VV} up and down by a quantity proportional to $1/M_{VV}$ and $1/M_{jet}$. Finally, as we do not know a priori whether Nature behaves more like PYTHIA8, HERWIG++ or MADGRAPH+PYTHIA8, we account for differences due to MC generation and parton shower modeling by adding three additional shapes corresponding to the PDFs obtained using different QCD MC generators.

Those five shape uncertainties are assigned very large pre-fit values (allowed to float within 33%), effectively allowing the simulation to take any value to describe the data. The alternate shapes described above are shown in Figure 4.96, here for the 2017 analysis.

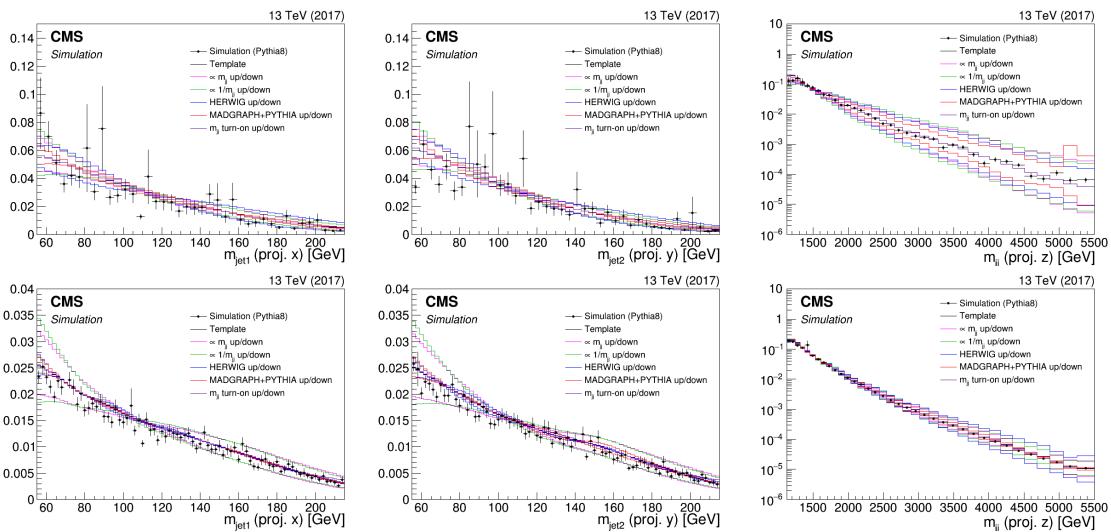


Figure 4.96: The nominal MC data (markers) and smooth nominal kernel obtained from PYTHIA8 (black line), together with the five alternate shapes added to the fit as shape nuisance parameters. Here for the high (top) and low (bottom) categories.

2618 **4.4 Background model validation**

2619 In order to convince ourselves that the fit is robust, several checks are performed in simulation
2620 and in a data control region before unblinding the data signal region.

2621 **4.4.1 Variations of QCD multijet predictions**

2622 First, we check that the main template (generated starting from QCD PYTHIA8 MC) can fit
2623 predictions from alternate QCD multijets samples in order to demonstrate that systematic
2624 uncertainties in the modeling of parton showers are covered by the relevant nuisance parameters.
2625 Fits to three different QCD samples are compared: PYTHIA8 (nominal), HERWIG++ (alternate
2626 shape 1) and MADGRAPH+PYTHIA8 (alternate shape 2). To ensure a smooth data distribution
2627 not affected by low statistics present in the MC spectra, we generate a toy dataset under the
2628 templates rather than using MC events directly. The postfit distributions after fitting each toy, is
 shown in Figures 4.97-4.99 for the HPHP (top) and HPLP category (bottom).

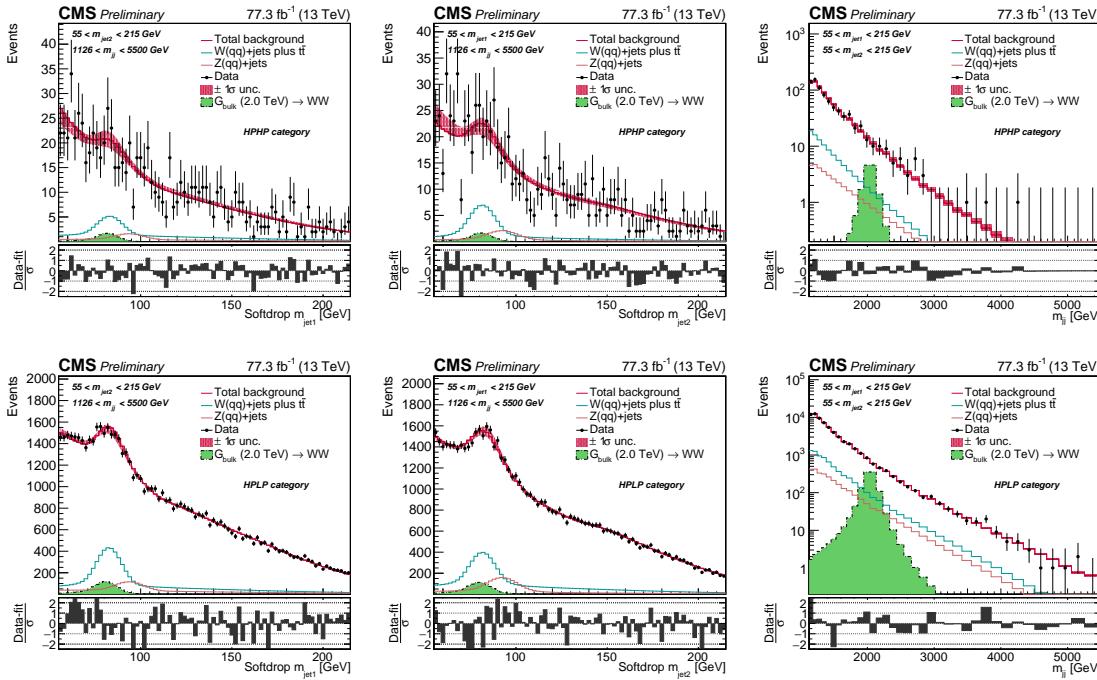


Figure 4.97: Postfit distributions after a combined fit to a toy datasets generated under the QCD PYTHIA8 template. The projections of $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right) are shown for the HPHP (top) and HPLP (bottom) category.

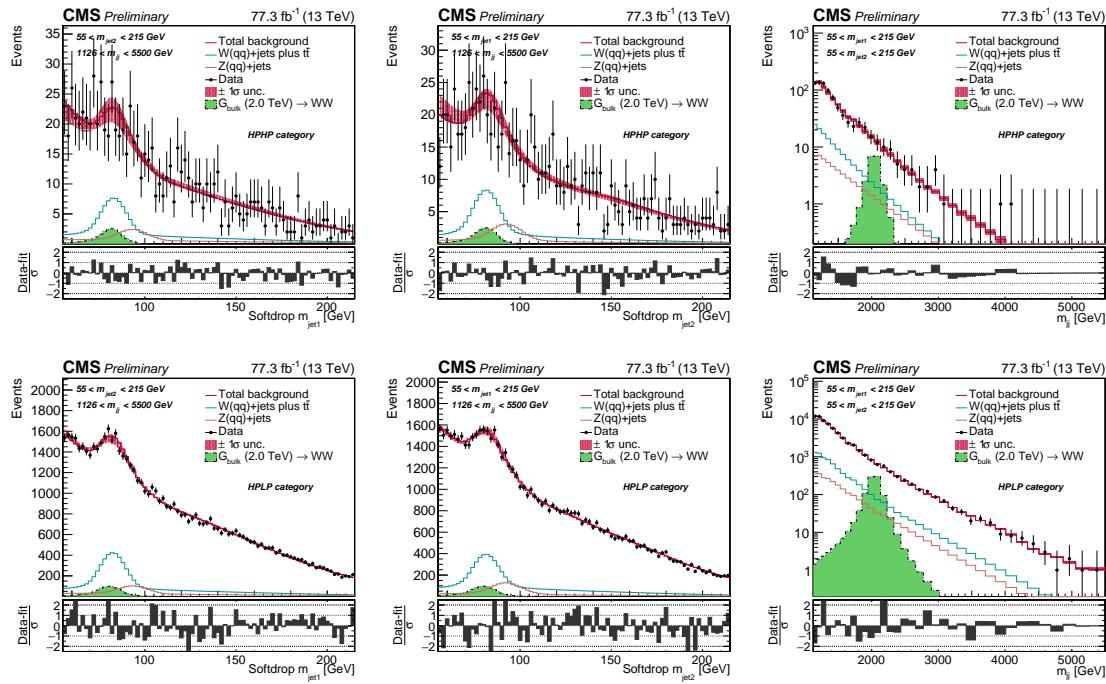


Figure 4.98: Postfit distributions after a combined fit to a toy datasets generated under the QCD HERWIG++ template. The projections of $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right) are shown for the HPHP (top) and HPLP (bottom) category.

To quantify the fit quality of these, a goodness-of-fit test is performed using a saturated model (which is valid when data are non-Gaussian), shown in Figure 4.100. 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.

4.4.2 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 4.3.9. 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. The post-fit distributions in the data control region are shown in Figures 4.101–4.103 for different jet mass and dijet invariant mass bins. A goodness-of-fit check is performed, shown on the left plot in Figure 4.104, and we again find that the toys are Gaussian distributed and the LPLP data is in good agreement with the background-only hypothesis. The post-fit value and uncertainty of each nuisance parameter involved in the fit is also studied, where

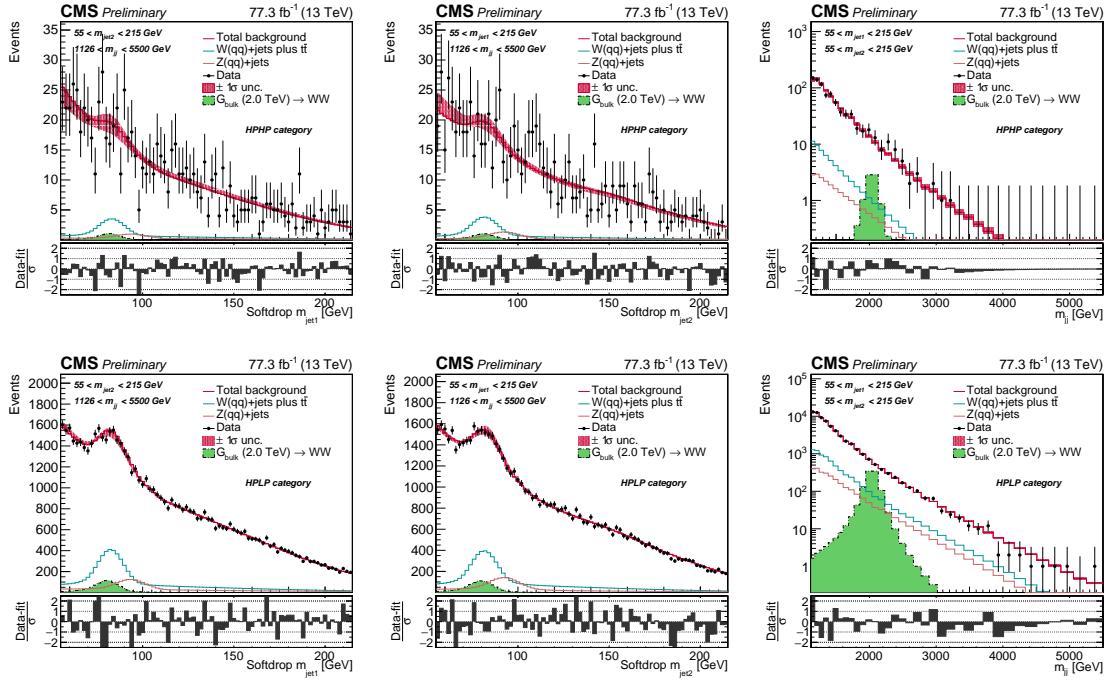


Figure 4.99: Postfit distributions after a combined fit to a toy datasets generated under the QCD QCD MADGRAPH+PYTHIA8 template. The projections of $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right) are shown for the HPHP (top) and HPLP (bottom) category.

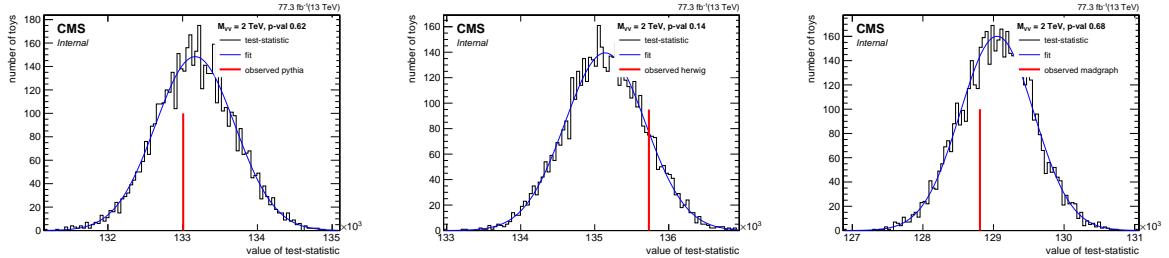


Figure 4.100: The likelihood for toys generated around the background-only hypothesis compared to the likelihood value of a toy dataset generated under the PYTHIA (left), HERWIG (middle) and MADGRAPH+PYTHIA8 template (right).

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 the right plot in Figure 4.104, 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

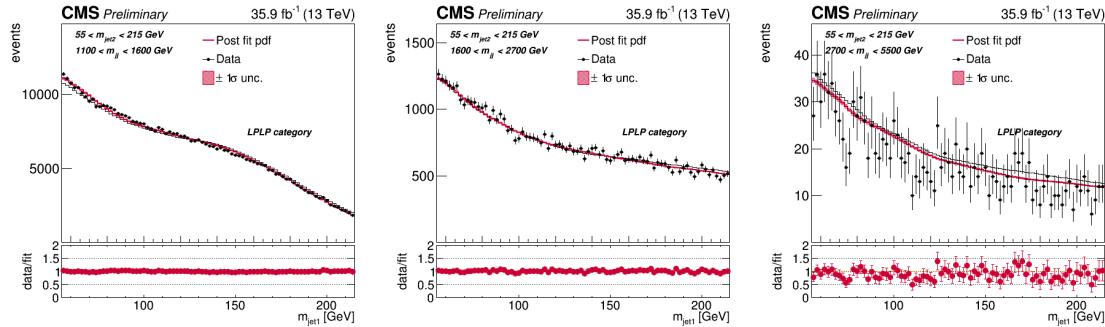


Figure 4.101: Distributions obtained from the fit to 2016 LPLP data. Here the projections of $M_{\text{jet}1}$ are shown for the several ranges of M_{VV} and for the full $M_{\text{jet}2}$ range.

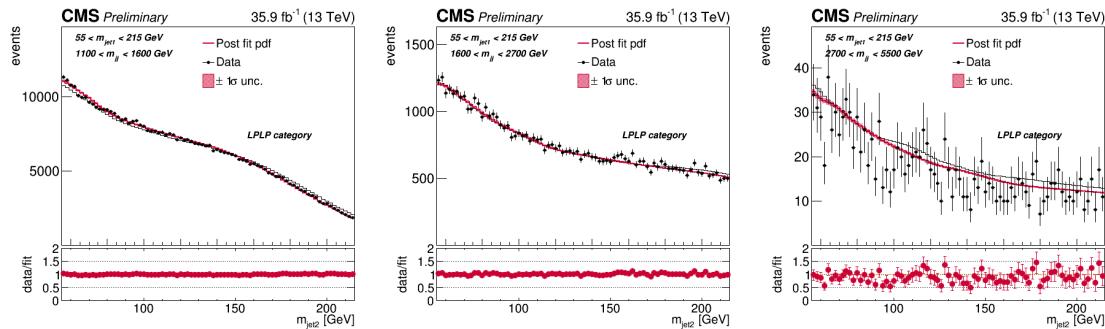


Figure 4.102: Distributions obtained from the fit to 2016 LPLP data. Here the projections of $M_{\text{jet}2}$ are shown for the several ranges of M_{VV} and for the full $M_{\text{jet}1}$ range.

2649 are strongly constrained by data, as expected given the large pre-fit uncertainty assigned to let
 2650 the shapes adjust to the real data.

2651 4.4.3 Bias test in pseudodata

2652 Finally, we study whether there is any bias on the extracted signal rate due to the background
 2653 model. A $G_{\text{bulk}} \rightarrow WW$ signal, with different mass and signal strength μ_{in} , is injected on top of the
 2654 full background. The signal strength is chosen such that it corresponds to a significance of 4-4.5
 2655 standard deviations for each signal mass, resulting in jet mass and dijet invariant mass spectra as
 2656 shown in Figure 4.105. For each tested signal mass point, a signal plus background fit is performed.
 2657 The signal normalization is free to float in the fit, which determines the signal strength μ and its
 2658 uncertainty σ_μ . For each toy, the pull of the signal strength $p_\mu = (\mu - \mu_{\text{in}})/\sigma_\mu$ is calculated. The
 2659 procedure is repeated ~ 1000 times for each category, and the cumulative distribution of the pulls

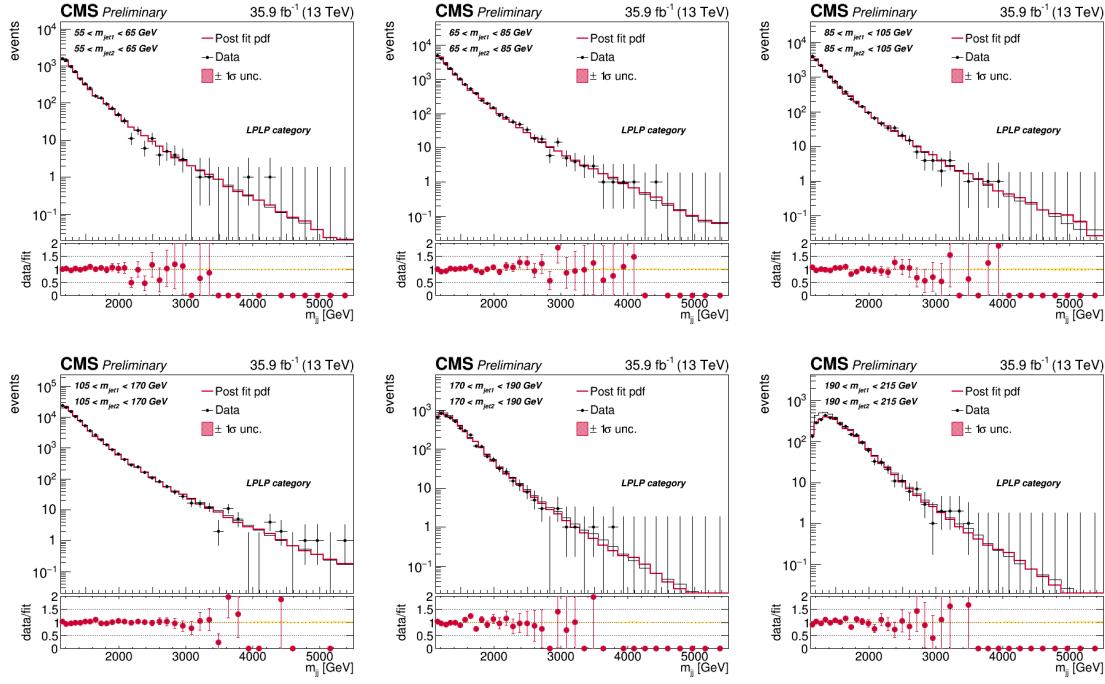


Figure 4.103: Distributions obtained from the fit to 2016 LPLP data. Here the projections of M_{VV} are shown for several equal ranges of M_{jet1} and M_{jet2} .

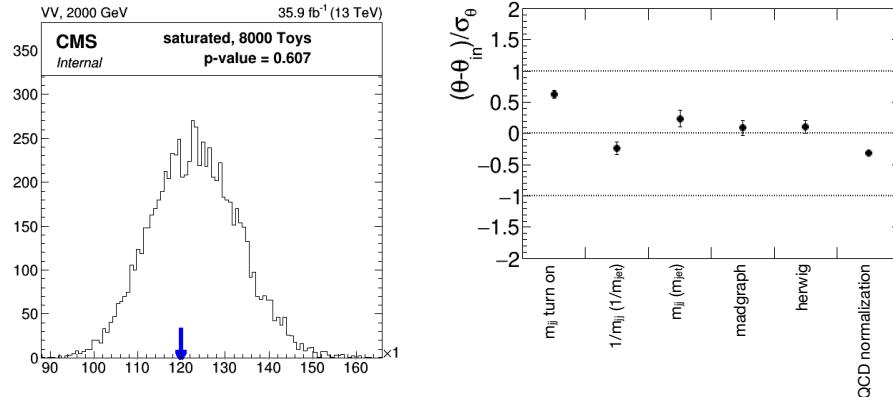


Figure 4.104: Left: the likelihood for toys generated around the background-only fit to LPLP data, compared to the likelihood on data. Data is in good agreement with the background-only hypothesis. Right: pulls of the nuisance parameters for a background-only fit.

2660 is fitted with a Gaussian function to determine the mean and its uncertainty, which represent
 2661 the bias. This is shown in Figure 4.106. Ideally, the distribution should be Gaussian distributed
 2662 with a mean around 0 and a width of 1. The bias, as shown in Figure 4.107, is consistency below

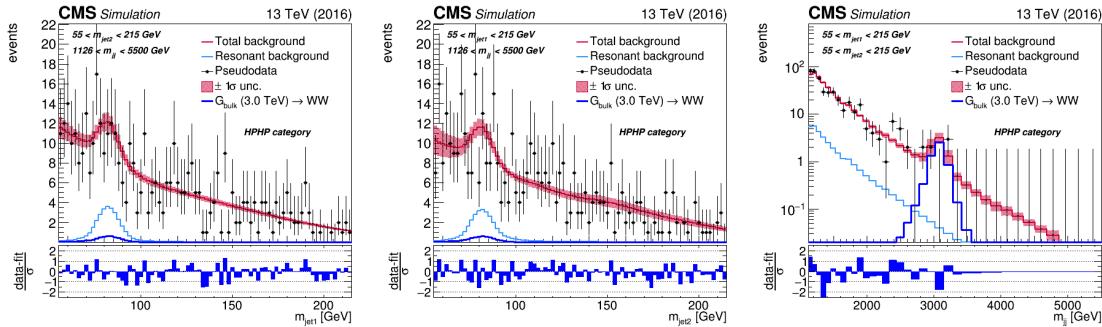


Figure 4.105: The post-fit distribution after injecting a signal with a mass of 3 TeV on top of the background in the HPHP (top) and HPLP (bottom) category. are shown separately for the two categories.

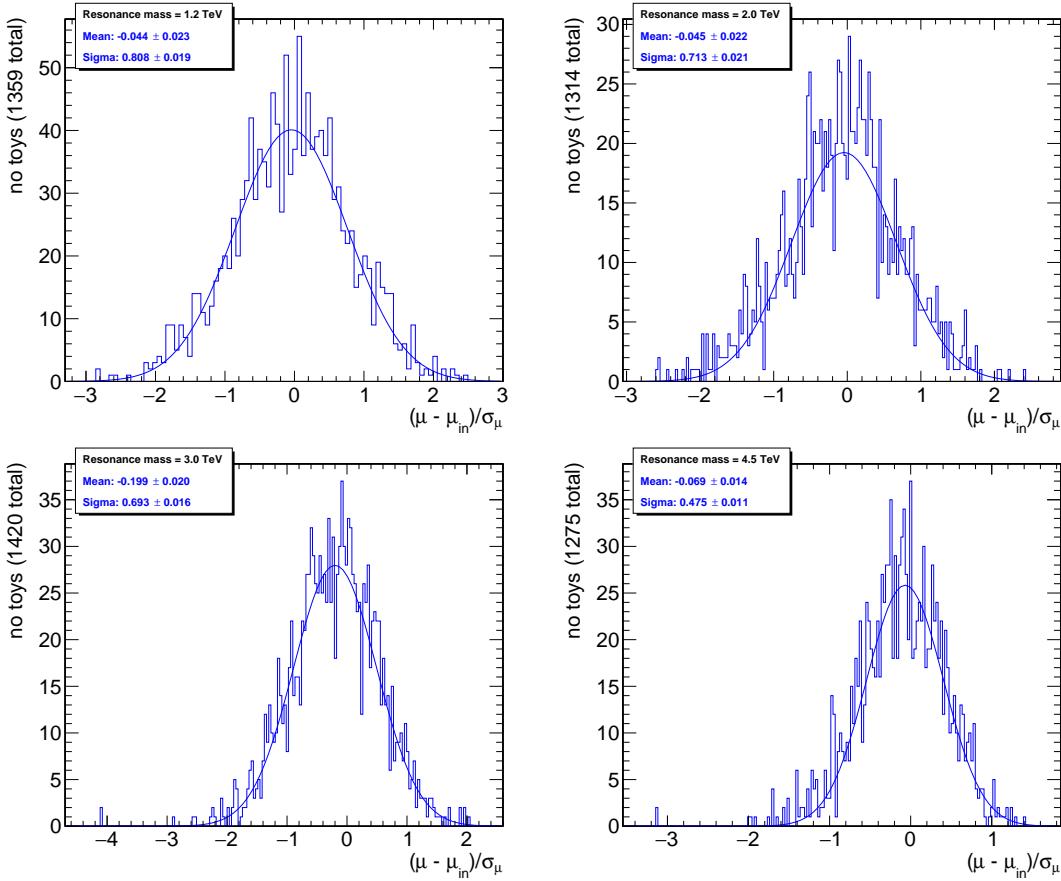


Figure 4.106: Cumulative distributions of the pulls for the signal strength for 4 different signal mass-points.

50% and therefore no additional correction/bias term is introduced. In addition, we calculate

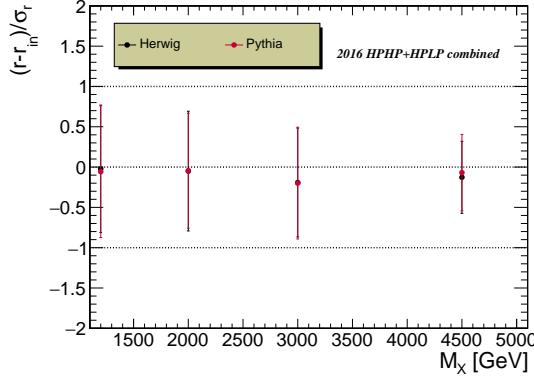


Figure 4.107: Estimated bias as a function of the resonance mass.

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 4.108. 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.

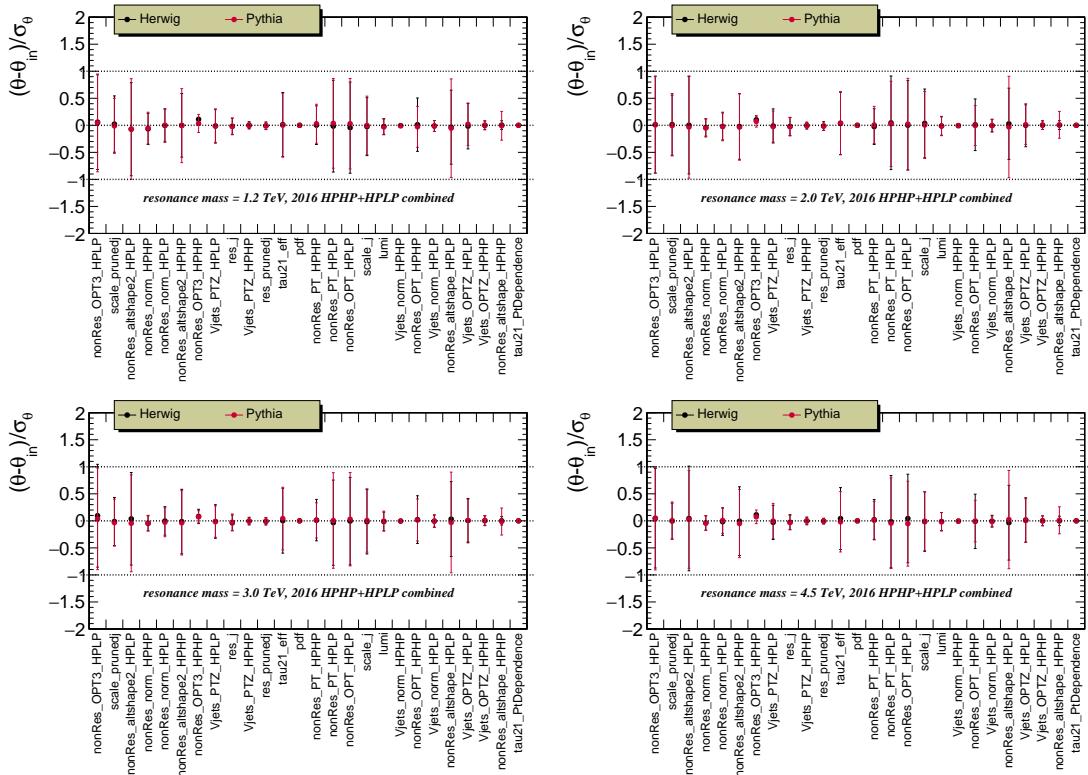


Figure 4.108: Pulls of the nuisance parameters estimated for the signal masses and strengths under test. Results are shown of the signal+background combined fits of the two HPHP and HPLP categories to toy datasets generated from the nominal PYTHIA8 templates corresponding to the 2016 integrated luminosity and signal+background pdf.

2670 **4.4.4 Results**

2671 The distributions obtained from a combined fit to the observed data in 2016 and 2017 are shown
2672 in Figure 4.109 and 4.110, with the corresponding predicted and observed number of background
2673 events in the signal region summarized in Table 4.17. We observe a beautiful double peak from
2674 the $W(q\bar{q})$ and $Z(q\bar{q})+jets$ background, especially clear in the low purity category. This allows us
2675 to, for the very first time extract the softdrop jet mass scale and resolution from a $V(q\bar{q})+jets$
2676 double peak, which we'll discuss in Section 4.4.4. It also gives us the opportunity to measure the
 $V(q\bar{q})+jets$ cross section, a measurement we are currently planning on how to best extract.

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

Table 4.17: Expected and observed yields and their total uncertainty (stat.+sys.) in the two purity categories.

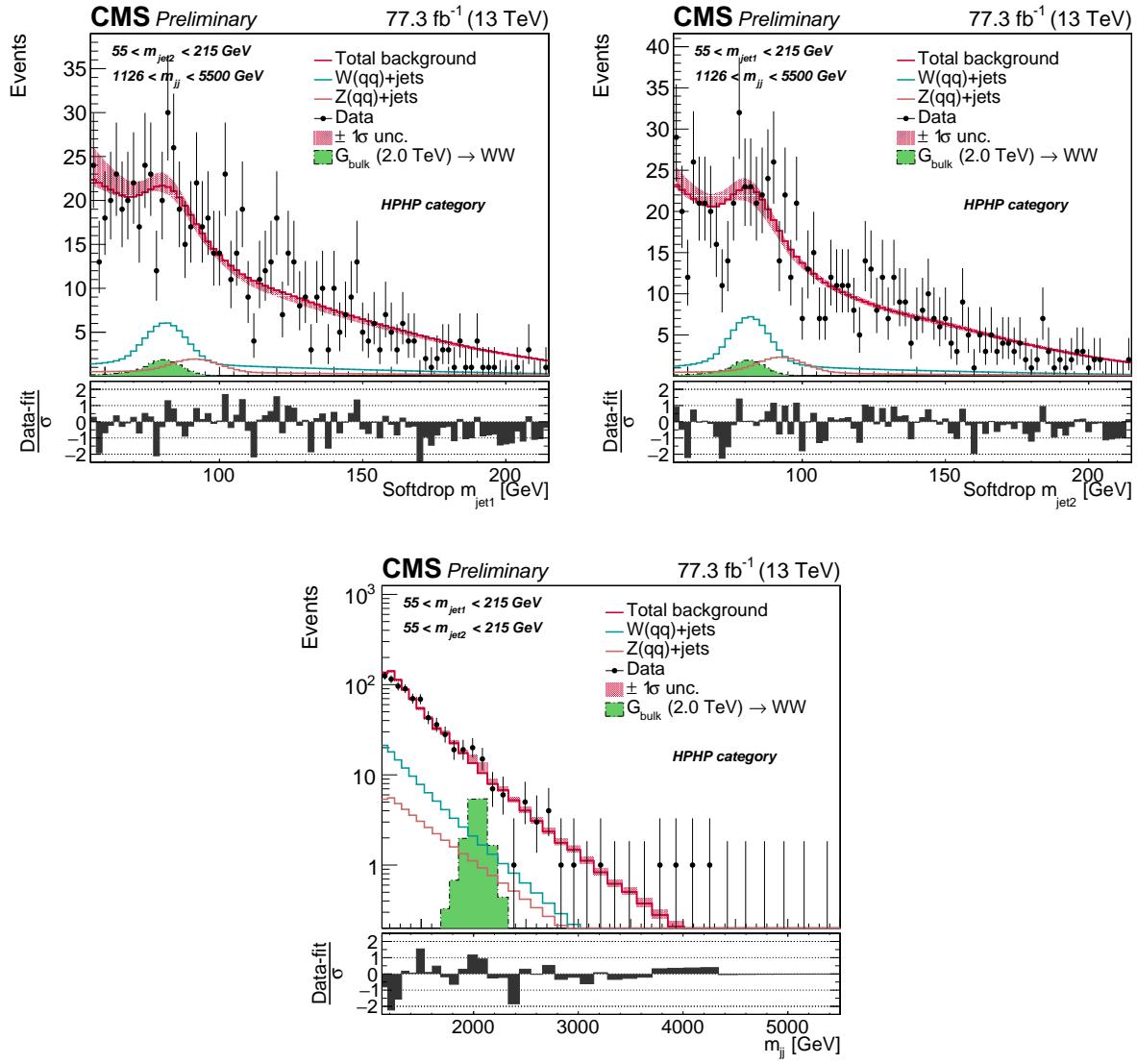


Figure 4.109: Postfit distribution after a fit to 2016 and 2017 data projected onto the $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle), and M_{ll} (right) axis. Here for the high purity category. The background shape uncertainty is shown as a red shaded band, and the statistical uncertainties of the data are shown as vertical bars. The overlaid signal distribution is arbitrarily normalized.

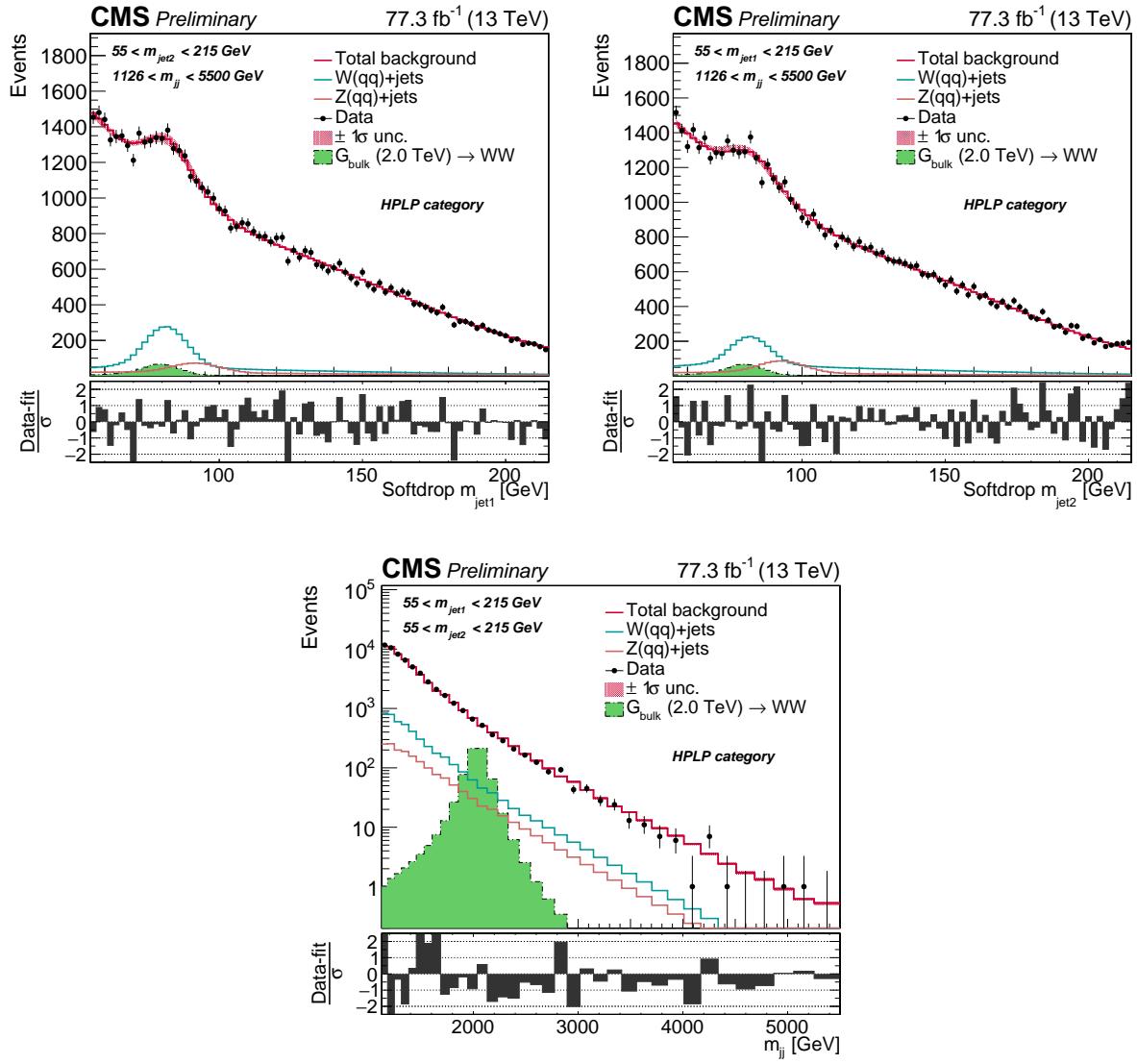


Figure 4.110: Postfit distribution after a fit to 2016 and 2017 data projected onto the M_{jet1} (left), M_{jet2} (middle), and M_{VV} (right) axis. Here for the HPLP (bottom) category. The background shape uncertainty is shown as a red shaded band, and the statistical uncertainties of the data are shown as vertical bars. The overlaid signal distribution is arbitrarily normalized.

2678 **Limits**

2679 As for Search I and Search II, exclusion limits on the cross section of the process $X \rightarrow VV$ are
2680 set in the context of the Bulk Graviton model and the HVT model B scenario (again obtained
2681 using the asymptotic CL_S method). Figure 4.111 show the resulting expected and observed 95%
2682 confidence level exclusion limits on the signal cross section as a function of the resonance mass for
2683 all signal models. The obtained limits are compared with the resonance production cross section
times the branching fraction to WW, ZZ and WZ. To settle the question of whether the 3D

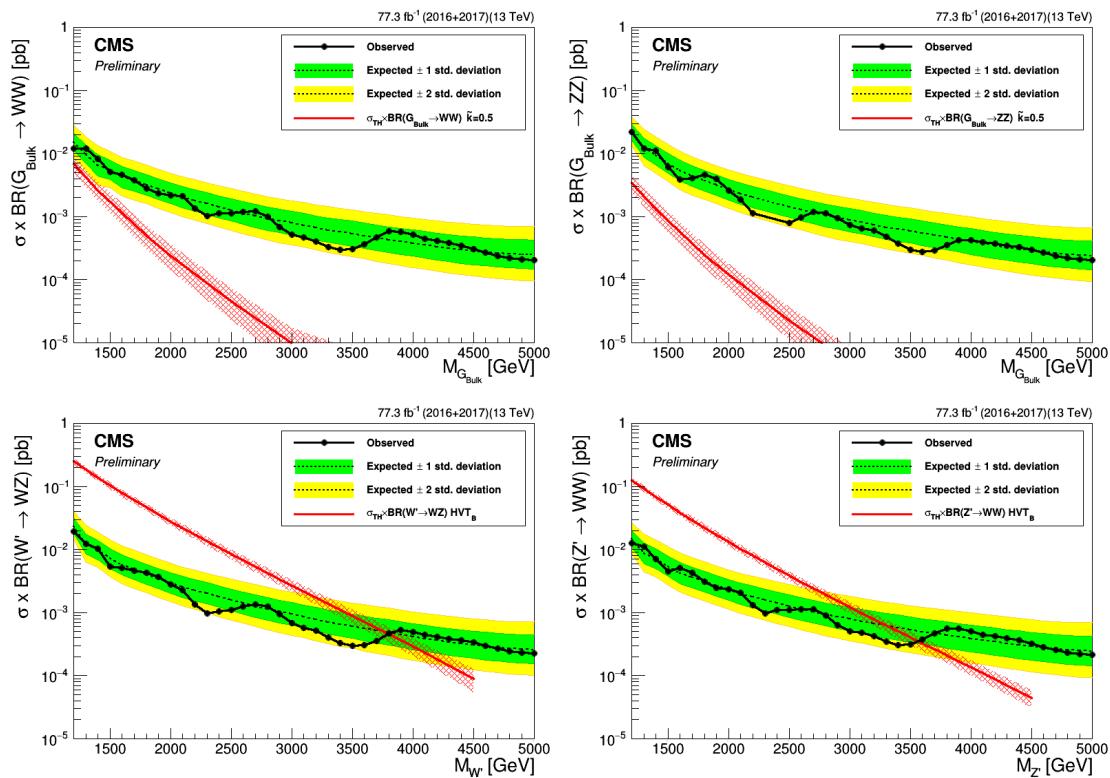


Figure 4.111: Expected limits obtained combining 35.9 fb^{-1} and 41.4 fb^{-1} of data of data after combining all purity categories for Bulk $G \rightarrow WW$ (top left), Bulk $G \rightarrow ZZ$ (top right), $W' \rightarrow WZ$ (bottom left) and $Z' \rightarrow WW$ (bottom right) signals.

2684 fit method yields an improvement upon the 1D search method, we compare the limits obtained
2685 above to those obtained using the 1D fit method (the results from SearchII). In Figure 4.112
2686 we see a 20–30% improvement for all signal hypothesis with the new method (comparing the
2687 1D and 3D 2016 limits only), and a total gain of about 35–40% when combining the 2016 and
2688 2017 datasets. In Figure 4.113, we additionally compare the 3D limits to those obtained by the
2689 ATLAS collaboration in a similar search [79] and find this search to be up to 35% more sensitive

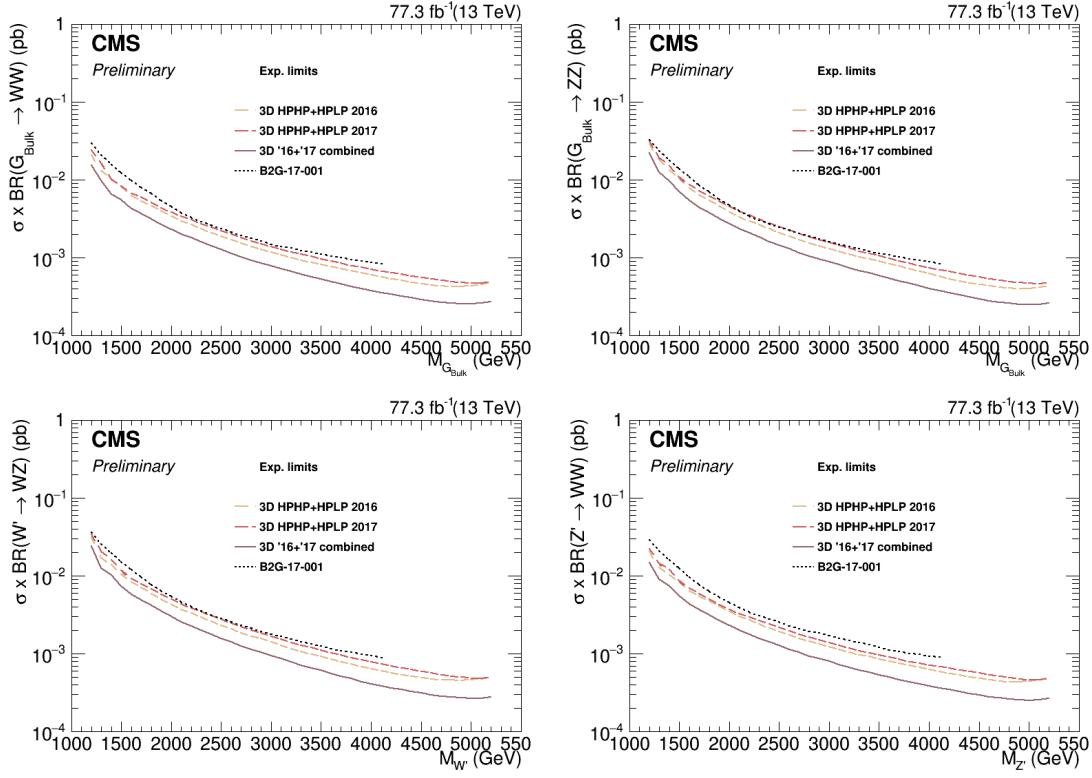


Figure 4.112: CA comparison of the 3D expected limits split into dataset (2016 and 2017), to that obtained with the 1D fit using 2016 data. Here for $G_{\text{bulk}} \rightarrow WW$ (top left), $G_{\text{bulk}} \rightarrow ZZ$ (top right), $W' \rightarrow WZ$ (bottom left) and $Z' \rightarrow WW$ (bottom right) signals.

for the two signal scenarios considered (the G_{bulk} limits can not be compared due to different values of \tilde{k}). Finally, in Figure 4.114 we show a breakdown of the limits per purity category. As expected, the HPHP category dominates at low mass where background is high and the HPLP category dominates at high mass due to low background and high signal efficiency.

2695 Pulses of nuisance parameters

As summarized in Section 4.3.11, we add a list of systematic uncertainties to the fit as nuisance parameters. To quantify how much the nuisances we insert differs from the ones preferred by the fit, we compute the pull

$$p_\theta = (\theta - \theta_{in})/\sigma_\theta \quad (4.21)$$

where θ_{in} is the nuisance value pre-fit, θ the corresponding parameter post-fit and σ_θ its error. and its error bar calculated as the ratio between post- and pre-fit uncertainty. Figure 4.115

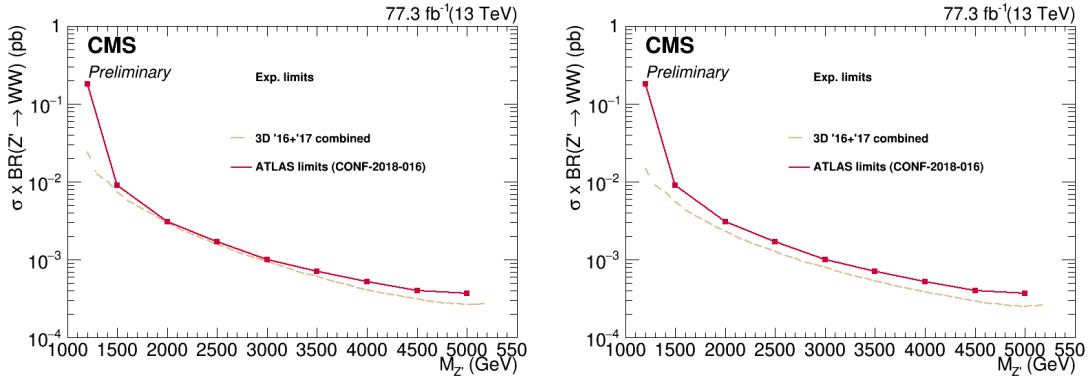


Figure 4.113: A comparison of the limits obtained above, to those by the ATLAS collaboration in a similar search [79]. Here for W' (left) and Z' (right) signal hypotheses.

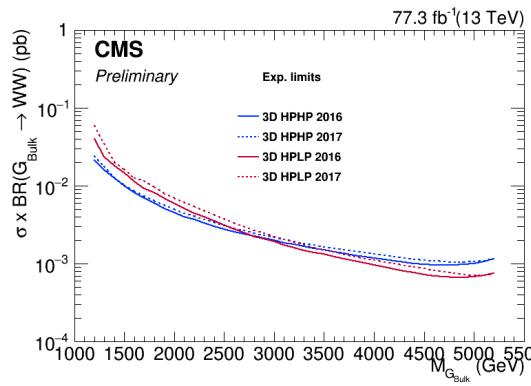


Figure 4.114: Limits split into purity categories (HPHP and HPLP) for a Bulk $G \rightarrow WW$ signal hypothesis.

2701 shows the pulls for a signal+background fit to the combined (2016+2017) dataset (left) and
 2702 when fitting the two separately (right), here using a signal hypothesis corresponding to a 2
 2703 TeV G_{bulk} . We observe that the W -tagging efficiency (“`CMS_VV_JJ_tau21_eff`”), the softdrop
 2704 jet mass scale (“`CMS_scale_prunedj`”) and resolution (“`CMS_res_prunedj`”) gets pulled and
 2705 constrained by the $W(\text{qq})$ and $Z(\text{qq})+\text{jets}$ mass peaks. In addition, the QCD shape parameters
 2706 (“`CMS_VV_JJ_nonRes_*`”) are significantly pulled and constrained by data because of their large
 2707 pre-fit uncertainty and unknown a-priori pre-fit value (again, we do not know if Nature is PYTHIA8,
 2708 HERWIG++ or MADGRAPH+PYTHIA8. Though from this measurement, HERWIG++ seems to
 2709 take the prize (“`altshape2`”)).

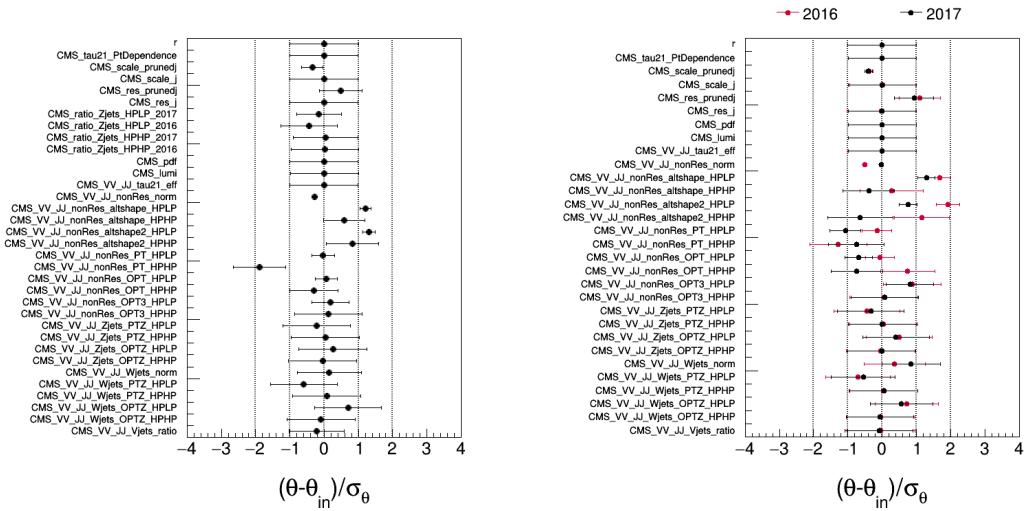


Figure 4.115: Pulls of each nuisance parameter for a combined signal+background fit to the combined 2016+2017 dataset (left) and when fitting the two separately.

2710 4.5 Summary and outlook

2711 In this chapter, we have followed the search for VV resonances in the all-hadronic final state
 2712 through three stages and corresponding publications: From being one of the first ever analysis in
 2713 the “boosted” final state at 13 TeV and the very first to take advantage of substructure at trigger
 2714 level, through leading the development of a new W-tagging algorithm and mass corrections now
 2715 default in CMS and finally ending with the development of a multi-dimensional fit for generic
 2716 searches in jet groomed mass and dijet invariant mass.

2717 Each analysis has built on significant improvements that came with the analysis before it:
 2718 The substructure triggers and mass corrected softdrop jet mass are both used for the 3D fit, the
 2719 early discovery of the softdrop signal efficiency dependence on p_T led us to derive corrections for
 2720 it. Now the question which remains is: *What comes next?*.

2721 A few ideas were already mentioned in the introduction to Search III, Section 4.3. The natural
 2722 next step for this search is an incorporation of the VH(bb) and H(bb)H(bb) final states into the
 2723 three-dimensional fit. Orthogonality between the three is guaranteed through b-tagging categories,
 as illustrated in Figure 4.116. This process is already underway, aiming for a publication of the

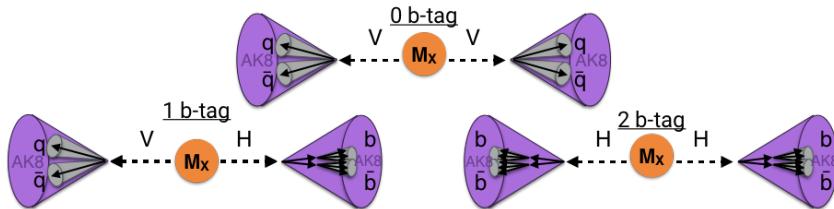


Figure 4.116: The VV, VH(bb) and H(bb)H(bb) analyses can all be incorporated into the multidimensional framework. Orthogonality between the analyses is ensured through b-tagging categories.

2724
 2725 full Run 2 dataset (data collected in 2016, 2017 and 2018) in one common framework.

2726 Secondly, after 14 TeV there will not be another increase in the collision center-of-mass energy
 2727 at the LHC. That means that we want the best possible sensitivity when analyzing the dataset
 2728 which is to come. One way of doing so is through changing the search method, as we did with
 2729 the three dimensional fit, another is to work towards a better W-tagger.

2730 Beyond that, and perhaps more interestingly, is the search for generic resonances peaking
 2731 anywhere in the softdrop jet mass and dijet invariant mass spectrum, where the jets themselves
 2732 could have other compositions than two subjets (for instance a scalar decaying to two vector
 2733 bosons, who’s decay products are merged into one jet; a 4-prong object). One caveat of the
 2734 current setup is that it constrains the signal to be a 2-pronged object through its n-subjettiness

2735 cut. In order for the multi-dimensional framework to be truly generic, the tagger needs to be
2736 replaced by a generic anti-QCD tagger. Such a tagger works as an anomaly detector by encoding
2737 the probability density function for quark/gluon jets as a function of certain variables, variables
2738 for which signal jets are assumed to have a different probability density. In our case, good
2739 variables would for instance be groomed mass or substructure, as any generic signal is assumed
2740 to be peaking in softdrop mass and have some (unknown) substructure. The tagger itself is given
2741 no information about what a potential signal looks like and will only return the probability of
2742 any jet being a QCD jet. Such taggers are usually deep neural network (DNN) based where the
2743 quark/gluon jet PDF is obtained through training of the network, as demonstrated in [80, 81].
2744 In order for such an encoding to work, the deep neural network needs access to the features
2745 distinguishing q/g jets from signal jets without these features being biased towards any signal in
2746 particular. The network must learn how to encode “non-substructure”.
2747 As a side project in parallel to working on the multi-dimensional fit, I spent the last half year of
2748 my PhD working on a deep neural network capable of discriminating q/g jets from W jets in
2749 order to improve W-tagging performance and improve the search sensitivity of VV analyses to
2750 come. Based solely on jet constituent four-vectors, the idea is to let the neural network itself
2751 compute grooming and substructure like variables, without feeding it any high-level features
2752 (like softdrop mass and τ_{21}). This type of architecture is, in addition to improving W-tagging
2753 performance, ideal for the purpose described above: Encoding QCD in terms of substructure like
2754 features. The final chapter of this thesis is therefore dedicated to the two last points: How to
2755 improve W-tagging in CMS for future analyses and how to design a neural network capable of
2756 learning jet substructure in an unbiased manner.

2757

CHAPTER 5

2758

Encoding jet substructure in a deep neural network

2759

2760 5.1 Infusing deep neural networks with physics

2761 The previous chapter ended by mentioning two ingredients that will become important for future
 2762 searches with the multi-dimensional fit: A better vector boson tagger, and a generic anti-QCD
 2763 tagger for signal independent searches. As a side project during my final PhD semester, I worked
 2764 on a solution for the first, which has the added benefit of being a stepping stone towards the latter.
 2765 This is what I will cover in the final chapter of this thesis.

2766

2767 When applying machine learning to particle physics problems, the input has historically con-
 2768 sisted of pre-computed high-level features (quantities based on lower-level variables and certain
 2769 theoretical assumptions). With the rise of deep learning however, computational graphs have
 2770 achieved an increased capability to find even the smallest correlations in datasets, allowing them
 2771 to construct complex features on their own. The deep neural network (DNN) I will present in
 2772 the following is based on the assumption that, given sufficient instructions about the laws of
 2773 Nature, a neural network should be capable of reconstructing its own high-level features based on
 2774 lower-level variables only. In addition, if smartly designed, the network should be capable of finding
 2775 novel correlations and physical features, *a-priori* unknown, by allocating a physical meaning to
 2776 the training weights deep within the network. The deep neural network I will present here, is
 2777 trained to discriminate quark/gluon jets from W -jets. However, as I will discuss in the final
 2778 section of this chapter, it is also the perfect starting point for developing a generic anti-QCD tagger.

2779

2780 The work presented in the following has not been published and still qualifies as work in progress.
 2781 However, I believe developing taggers such as these is of great importance for future versions of
 2782 the searches presented here, and is something I hope to continue working on in the future..

The notes are handwritten and organized into several sections:

- COLA:** A diagram showing a 2x2 matrix with columns E_1, E_2 and rows $P_x^1, P_x^2, P_y^1, P_y^2$. Below it is the equation $k_{\mu\nu} = k_{\mu i} \cdot L_{i\nu}$.
- Linear Combinations:** A diagram showing a 2x2 matrix with columns E_1, E_2 and rows $P_x^1, P_x^2, P_y^1, P_y^2$. Below it is the equation $L_{i\nu} = w_{1i} E_1 + w_{2i} E_2$.
- Sum of Energies:** A diagram showing a 2x2 matrix with columns E_1, E_2 and rows $P_x^1, P_x^2, P_y^1, P_y^2$. Below it is the equation $\sum E = E_1 + E_2$.
- Sum of Momentum:** A diagram showing a 2x2 matrix with columns E_1, E_2 and rows $P_x^1, P_x^2, P_y^1, P_y^2$. Below it is the equation $\sum p = P_x^1 + P_x^2 + P_y^1 + P_y^2$.
- Work in Progress:** A diagram showing a 2x2 matrix with columns E_1, E_2 and rows $P_x^1, P_x^2, P_y^1, P_y^2$. Below it is the equation $\sum p = w_{1i} P_x^1 + w_{2i} P_x^2 + w_{3i} P_y^1 + w_{4i} P_y^2$.
- Final Result:** A diagram showing a 2x2 matrix with columns E_1, E_2 and rows $P_x^1, P_x^2, P_y^1, P_y^2$. Below it is the equation $\sum p = w_{1i} P_x^1 + w_{2i} P_x^2 + w_{3i} P_y^1 + w_{4i} P_y^2$.
- Conclusion:** A bracketed summary at the bottom right states "linear combinations of momenta".

"What can we teach the machine?" → "What can we learn from the machine?".

Work in progress

2783 **5.2 LoLa**

2784 LoLa is a deep neural network architecture which was first introduced for top tagging [82]. It is
2785 based on the idea that, given enough information about the laws of Nature, a neural network
2786 should be capable of calculating jet substructure observables on its own given only low-level
2787 information. The network is designed to discriminate between AK R=0.8 jets originating from
2788 W bosons from those originating from quarks or gluons, solely based on the jet constituent
2789 four-vectors (variables with little discriminating power on their own) as illustrated in Figure 5.1.
 Rather than being fed high-level features, the neural network is given tools to perform calculations

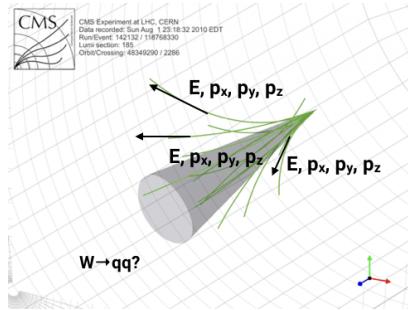


Figure 5.1: LoLa uses only jet constituent four-vectors as input to discriminate W from q/g jets.

2790
2791 on Lorentz vectors using the Minkowski metric. Through two novel layers, linear combinations
2792 similar to jet clustering and jet substructure algorithms are performed, allowing the algorithm
2793 to create its own substructure variables. Additionally, training weights deep within the network
2794 correspond to physical quantities reconstructed by the algorithm; distance between particles,
2795 masses and energies, linear combinations of particle four-vectors etc. Besides the end goal of
2796 discriminating Ws from quarks and gluons, one could therefore hope to learn of new correlations
2797 separating QCD from vector boson jets.

2798 **5.2.1 Architecture**

2799 The LoLa architecture is designed as a four layer deep, feed-forward sequential network doing
2800 supervised learning on fixed size input vectors. Two novel layers are introduced, the Combination
2801 Layer (CoLa) and the Lorentz Layer (LoLa), which perform basic jet clustering and substructure
2802 calculations as well as implements the Minkowski metric. These two layers are then followed by
2803 two fully connected layers, consisting of 100 and 50 nodes respectively, before the final output
2804 is computed using a Softmax activation function, yielding two output probabilities between 0

and 1. The loss function to be minimized is “categorical crossentropy” (or log loss) where the two categories in use are W versus non-W probabilities. Only the W jet probability is stored. The optimizer used in the training is the, now standard, ADAM optimizer [83], which adapts the learning rate of the model parameters during training. The code itself is written using the Keras [84] interface with a TensorFlow [85] backend. The full architecture with input and output dimension per layer is shown in Figure 5.2. The three first boxes are matrices, while the final four boxes correspond to vectors of different length. In the following, each layer will be explained in detail.

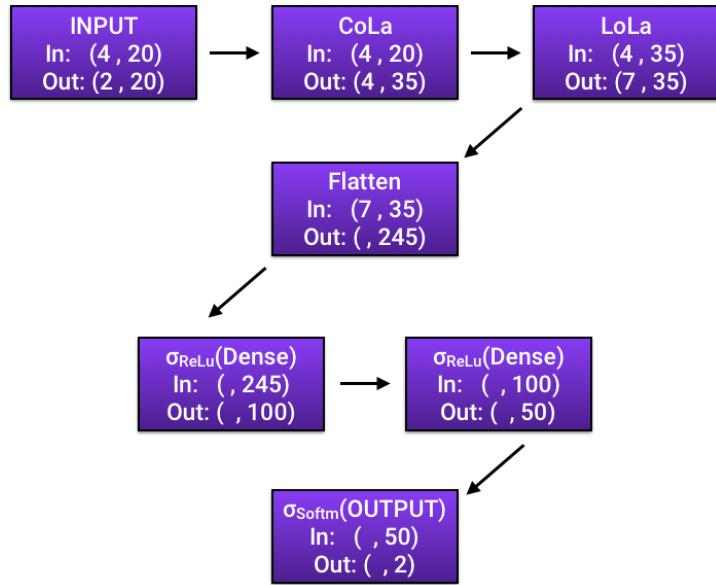


Figure 5.2: The full LoLa architecture. “In” denotes the dimension of the input tensor to the given layer, “Out” is the output tensors dimensions.

2812

2813 5.2.2 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/gluon jets from a QCD sample generated with PYTHIA8Pythia 8. 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 bosons, 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 5.3. From these signal and background jets, only the jet constituent four

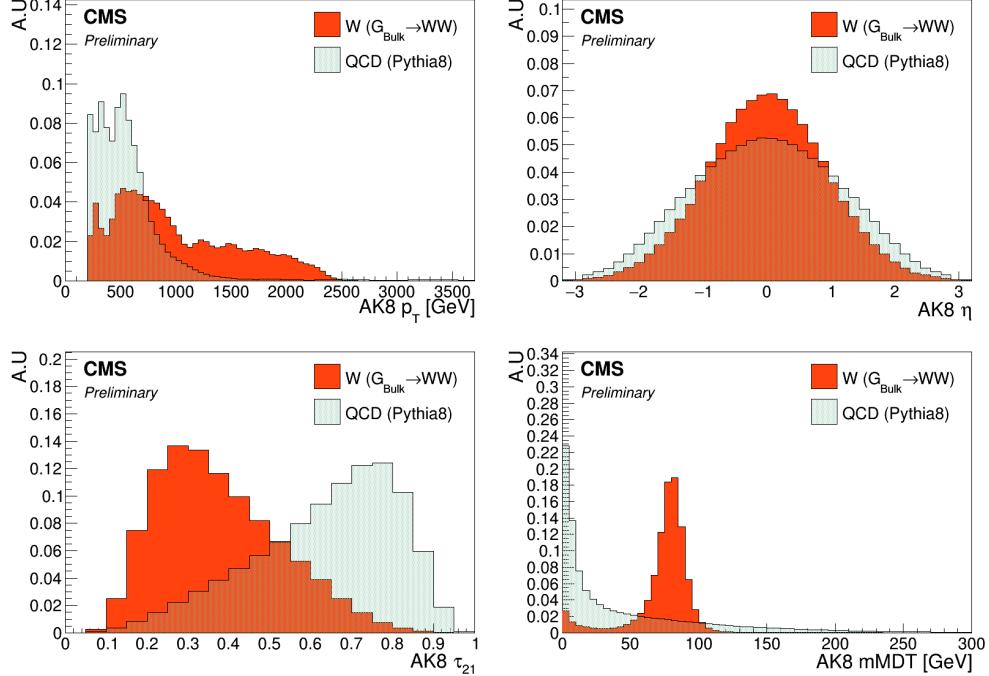


Figure 5.3: Jet p_T (top left), η (top right), τ_{21} (bottom left) and softdrop jet mass (bottom right) for signal and background jets.

2822

2823 vectors of the 20 highest- p_T particles are used as input to the deep neural network: E , p_x , p_y and
 2824 p_z . I use 20 constituents as any larger number has a negligible affect on the performance, while
 2825 performance tends to drop once going below 15. The input is therefore a $4 \times N = 20$ matrix for
 2826 each signal and background jet, one four-vector for each of the 20 jet constituents:

$$x_{\mu,i} = \begin{pmatrix} E^1 & E^2 & \dots & E^N \\ p_x^1 & p_x^2 & \dots & p_x^N \\ p_y^1 & p_y^2 & \dots & p_y^N \\ p_z^1 & p_z^2 & \dots & p_z^N \end{pmatrix} \quad (5.1)$$

2827 The total number of jet constituents is shown in Figure 5.4, and the input variables (here for all
 2828 constituents) is shown in Figure 5.5.

2829 It is clear that the input variables provide little discriminating power on their own. Therefore,
 2830 the network must learn how to derive other physical quantities where the signal and background
 2831 PDFs differ to a larger extent. This is achieved through the two custom layers described in the

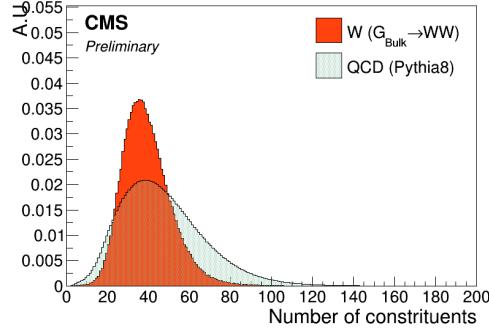


Figure 5.4: The number of jet constituents for signal (red) and background (blue). Only the 20 highest- p_T constituents are used during training.

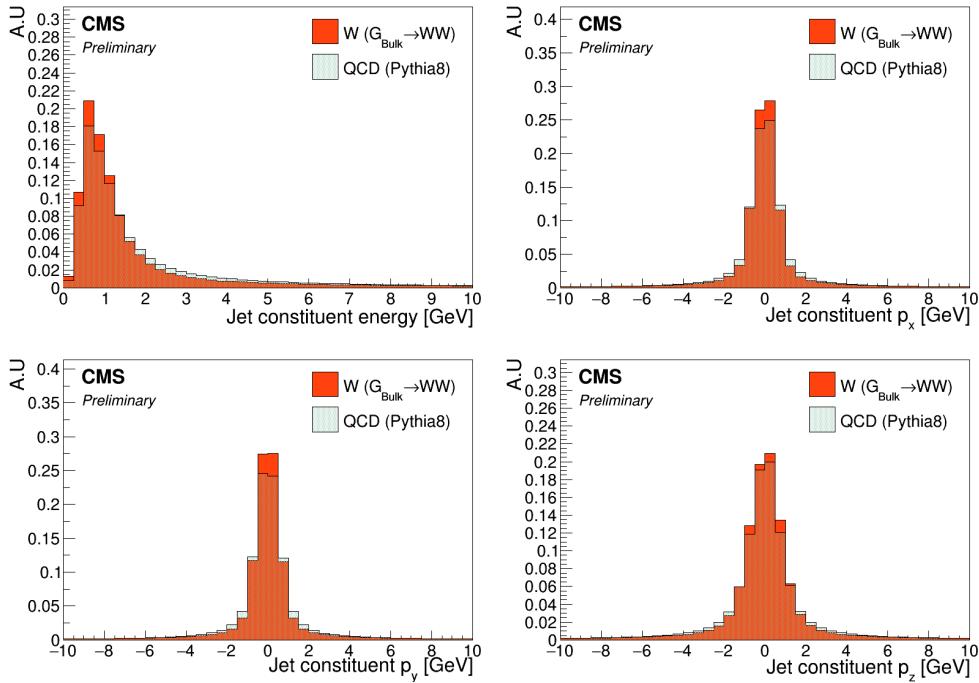


Figure 5.5: Energy (top left), p_x (top right), p_y (bottom left) and p_z (bottom right) for all jet constituents. These values are used as input to the neural network training.

2832 following.

2833 5.2.3 The Combination Layer

2834 The Combination Layer (CoLa) consists of a matrix which, when taking the scalar product with
2835 the input matrix, computes linear combinations of the jet constituents, similar to what is done in

2836 recombination jet algorithms. The main goal here is to create additional four-vectors as input for
2837 the next layer. The CoLa matrix is a concatenation of the following: A vector of 1's of length N ,
2838 the $N \times N$ identity matrix ($N = 20$) and a matrix of $N \times M$ trainable weights.

$$C_{i,j} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & w_{1,N+2} & w_{1,N+3} & \dots & w_{1,(N+2)+M} \\ 1 & 0 & 1 & \dots & 0 & w_{2,N+2} & w_{2,N+3} & \dots & w_{2,(N+2)+M} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & 0 & 1 & w_{N,N+2} & w_{N,N+3} & \dots & w_{N,(N+2)+M} \end{pmatrix} \quad (5.2)$$

2839 When performing the following multiplication

$$x_{\mu,i}^C = x_{\mu,i} C_{i,j} \quad (5.3)$$

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:

$$\begin{pmatrix} E^1 & E^2 \\ p_x^1 & p_x^2 \\ p_y^1 & p_y^1 \\ p_z^1 & p_z^2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & w_{1,4} & w_{1,5} \\ 1 & 0 & 1 & w_{2,4} & w_{2,5} \end{pmatrix} = \begin{pmatrix} E^1 + E^2 & E^1 & E^2 & w_{1,4}E^1 + w_{2,4}E^2 & w_{1,5}E^1 + w_{2,5}E^2 \\ p_x^1 + p_x^2 & p_x^1 & p_x^2 & w_{1,4}p_x^1 + w_{2,4}p_x^2 & w_{1,5}p_x^1 + w_{2,5}p_x^2 \\ p_y^1 + p_y^1 & p_y^1 & p_y^1 & w_{1,4}p_y^1 + w_{2,4}p_y^1 & w_{1,5}p_y^1 + w_{2,5}p_y^1 \\ p_z^1 + p_z^2 & p_z^1 & p_z^2 & w_{1,4}p_z^1 + w_{2,4}p_z^2 & w_{1,5}p_z^1 + w_{2,5}p_z^2 \end{pmatrix}$$

2840 In the two last columns, the neural network makes two “subjet” four-vectors by weighting the
2841 relative contribution of each particle as it sees fit. This is similar to jet grooming (Section 3.5.1)
2842 or PUPPI pileup subtraction (Section 3.3.2), and should allow the network to learn which
2843 constituents are part of the hard scatter and which are not. The $x_{\mu,i}^C$ matrix is finally passed on
2844 to the next layer, the Lorentz Layer.

5.2.4 The Lorentz Layer

2845 The Lorentz Layer (LoLa) is responsible for encoding how particles move in space-time through
2846 a simple set of rules. Each column (four-vector) of $x_{\mu,i}^C$, is used to compute, and afterwards is

2848 replaced by, the following $k = 7$ features:

$$x_{k,i}^L = \begin{pmatrix} m^2(x_{\mu,i}^C) \\ p_T(x_{\mu,i}^C) \\ w_{ij}^E E(x_{\mu,j}^C) \\ w_{ij}^{s1} \sum d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{s2} \sum d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{m1} \min d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{m2} \min d^2(x_{\mu,i}^C, x_{\mu,j}^C) \end{pmatrix} \quad (5.4)$$

2849 Going through from top to bottom, these are:

- 2850 • The invariant mass and p_T of each four-vector
- 2851 • A linear combination of all four-vector energies where each is scaled by a trainable weight
- 2852 • The sum of distances between the four-vector under consideration and every other column
- 2853 reweighted with a trainable weight
- 2854 • The minimum distance between the four-vector under consideration and every other column
- 2855 where each distance again is reweighted by a trainable weight.

2856 The Minkowski metric enters explicitly in the first and in the last four calculations, where the
2857 neural network is told to abide by the rules

$$m^2(x_{\mu,i}^C) = g^{\mu\nu} x_{\mu,i}^C x_{\nu,i}^C \quad (5.5)$$

2858 and

$$d^2(x_{\mu,i}^C, x_{\mu,j}^C) = (x_{\mu,i}^C - x_{\mu,j}^C)_\mu g^{\mu\nu} (x_{\mu,i}^C - x_{\mu,j}^C)_\nu \quad (5.6)$$

2859 with $g^{\mu,\nu} = [-1, 1, 1, 1]$, when calculating the invariant mass and distance between particles/subjets.
2860 This tells the neural network to use a space-time geometry in all its calculations to respect Lorentz
2861 Invariance. The four final rows of LoLa are the most interesting: Here the network computes
2862 quantities similar to n-subjettiness by summing up the distances between all constituents, the
2863 jet axis and the subjets produced by CoLa. If, for instance, the network has been capable of
2864 reconstructing two hard subjets in the final columns of CoLa, which do linear combinations of
2865 particles, it can create its own “ τ_2 ” variable by taking the distance between those subjets and
2866 all the jet constituents (and weighing down the column corresponding to the full jet four-vector,

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) and compute “ τ_2 ”.

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?”.

5.3 Performance

The deep neural network is trained on 320k 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 60k W jets and 60k quark/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 5.6. The point where the blue curves end, represent the signal efficiency for a mass cut of $65 \text{ GeV} < \text{Softdrop jet mass} < 105 \text{ GeV}$, here roughly 70%. We clearly see that LoLa performs significantly better than the current baseline W-taggers

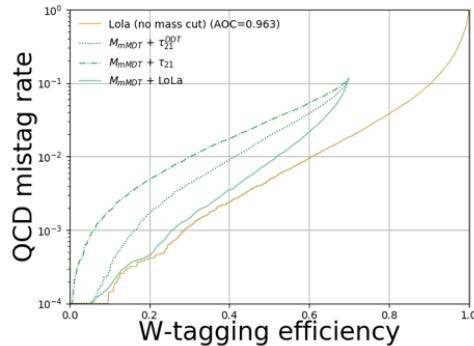


Figure 5.6: Performance of LoLa compared to other W-tagging discriminants in the background-signal efficiency plane: PUPPI softdrop + τ_{21} (dashed blue), PUPPI softdrop + τ_{21}^{DDT} (dotted blue), LoLa with a softdrop mass window applied (solid blue) and the nominal LoLa tagger with no mass cut applied.

2885

2886 based on τ_{21} and τ_{21}^{DDT} , with a roughly 20% higher signal efficiency at a given mistagging rate.

2887 LoLa also has a higher signal acceptance, as it can be used without a mass window applied. If
 2888 LoLa were to replace the tagger used in Search II (a better comparison than Search III as the
 2889 latter uses a rather unconventional mass window), which has a signal efficiency of $\sim 42\%$ at a 2%
 2890 mistagging rate for a single jet, the signal efficiency for the same mistagging rate would be 65%,
 2891 a 55% increase. For an analysis requiring two tagged jets, that would imply going from an 18 to
 2892 a 43% total signal efficiency, a significant gain.

2893 5.4 p_T and mass dependence

2894 Despite being a key feature, absolute performance is not all that quantifies how good one tagger
 2895 is compared to another. One a tagger is planned to be used in physics analysis, there are three
 2896 key questions one needs to consider:

- 2897 • Is the absolute performance better (compared to common methods)?
- 2898 • Is the tagger p_T -dependent?
- 2899 • Does the tagger sculpt the mass spectrum?

2900 These three measures are equally important in quantifying performance and, in the following,
 2901 I will attempt to explain why this is the case and which approaches are used here in order to
 2902 tackle them.

2903 Any deep neural network trained to distinguish W jets from q/g jets, will naturally learn that
 2904 p_T and mass are discriminating features unless it is penalized for it. Figure 5.7 shows the LoLa
 discriminant as a function of jet p_T and softdrop jet mass. A strong correlation is observed both

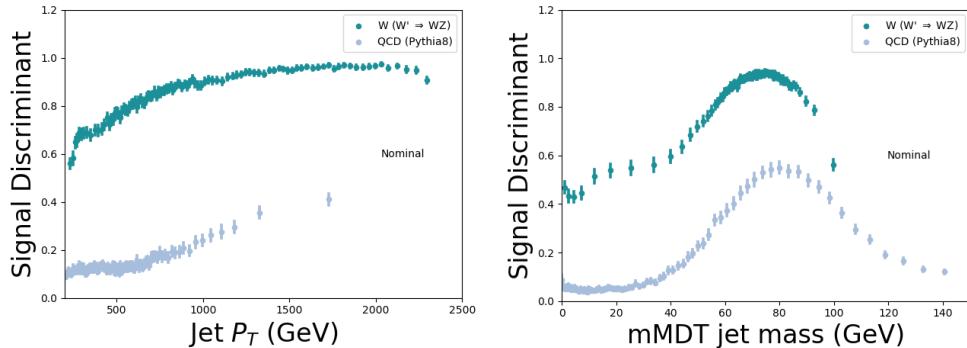


Figure 5.7: The LoLa discriminant as a function of jet p_T (left) and softdrop jet mass (right). A strong correlations with both variables is observed

2905
 2906 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 more signal like) and a bump around the W mass for both signal and for background.

5.4.1 p_T decorrelation

A tagger which is p_T dependent is a problem for 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 performance is 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 in addition can introduce sculpting of the dijet invariant mass spectrum.

From the top left plot in Figure 5.3, 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 each jets contribution to the total loss (making high mass QCD jets and low mass signal jets count more). Figure 5.8 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

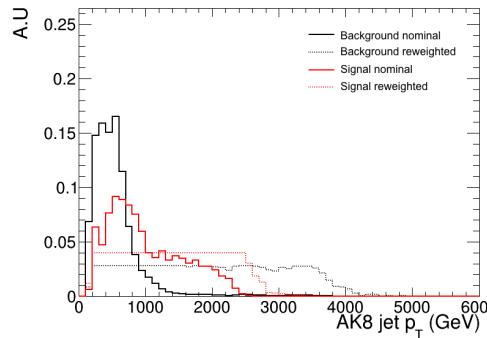


Figure 5.8: Jet p_T distribution before (solid lines) and after (dashed line) applying a weight intended to flatten the jet p_T spectrum.

with each jet, and the final discriminant compared to the nominal training. Figure 5.9 shows the performance of the same taggers as above but with one additional line, 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 5.10,

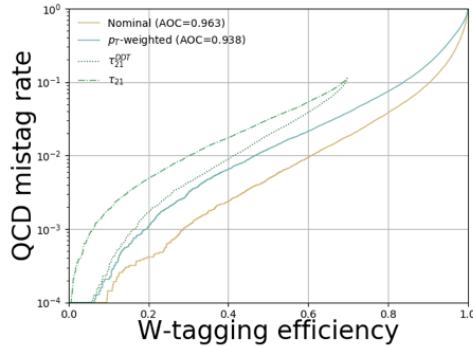


Figure 5.9: Performance of the p_T -reweighted LoLa tagger (solid blue) and the nominal LoLa tagger (solid yellow).

2929 the correlation we observed before has vanished and we are left with a tagger not depending on the jet p_T . For completeness, Figure 5.11 shows the τ_{21} and τ_{21}^{DDT} discriminants versus jet

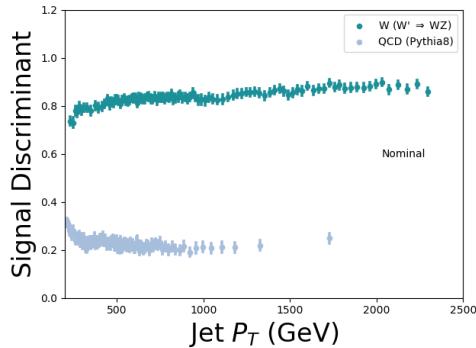


Figure 5.10: The LoLa discriminant as a function of jet p_T after training with a weight intended to flatten the sample p_T spectrum.

2930
 2931 p_T . Whereas the nominal LoLa discriminant had a much larger correlation with jet p_T than
 2932 the τ_{21} -based taggers, the p_T -reweighted version is as decorrelated from p_T as the τ_{21} variables
 2933 while still exhibiting a better absolute performance than the baseline taggers. In summary,
 2934 reweighting strategies as the one described above yield a loss in overall performance, as expected
 2935 when removing information from the training. However, the p_T dependence of the tagger is
 2936 strongly reduced, meaning that it might perform better overall in physics analysis when systematic
 2937 uncertainties are taken into account. There is therefore no clear answer as to which method is
 2938 better before running a full analysis including systematics for p_T -dependent tagging.

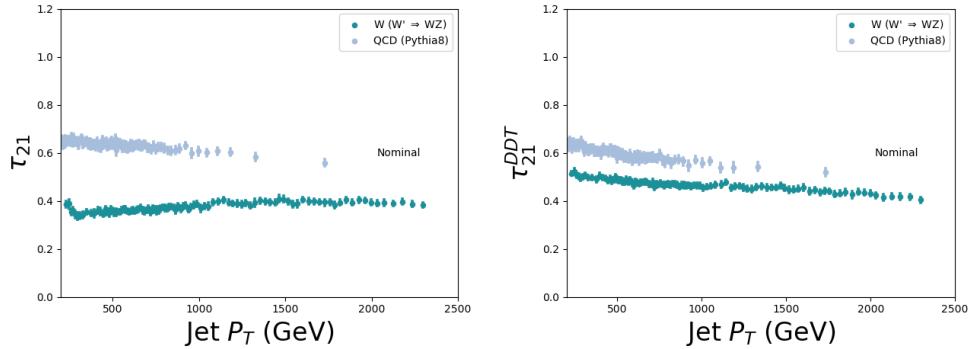


Figure 5.11: The τ_{21} (left) and τ_{21}^{DDT} (right) discriminant as a function of jet p_T .

5.4.2 Mass sculpting

Any smart deep neural network intended to separate Ws from quarks and gluons, will inevitably learn the W mass as it clearly is very different from the q/g mass. Unfortunately, as these taggers are meant to be used in physics analysis 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 in Figure 5.12 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 5.12, where all histograms are normalized to unit area, we see that the bulk of the remaining 1% QCD jets is right below the W 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 analysis, 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 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 5.13 shown the same softdrop jet mass spectrum before and after a cut

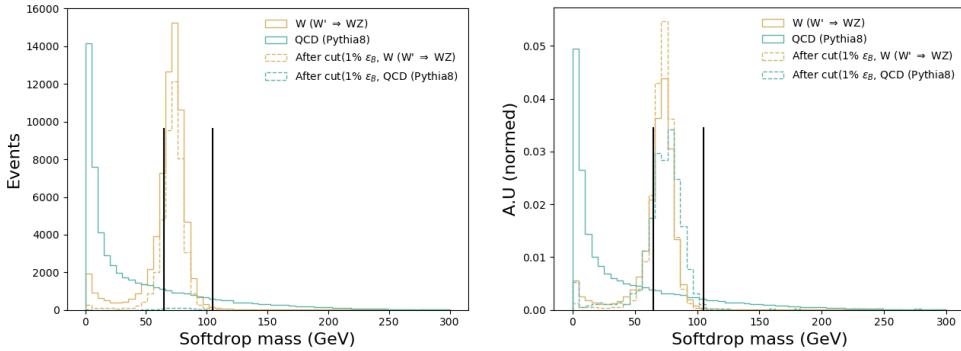


Figure 5.12: The softdrop jet mass distribution before (solid lines) and after (dotted lines) a cut on the LoLa discriminant corresponding to a 1% mistagging rate has been applied. The left plot shows the real number of events left after the cut, the right is normalized to area.

2963 corresponding to a 1% mistagging rate on τ_{21} (left) and τ_{21}^{DDT} (right). Here τ_{21} clearly exhibits
 2964 mass sculpting, but not as peaky as was the case for LoLa. τ_{21}^{DDT} exhibits the least amount of
 2965 sculpting, but is also the tagger with the worst absolute performance.

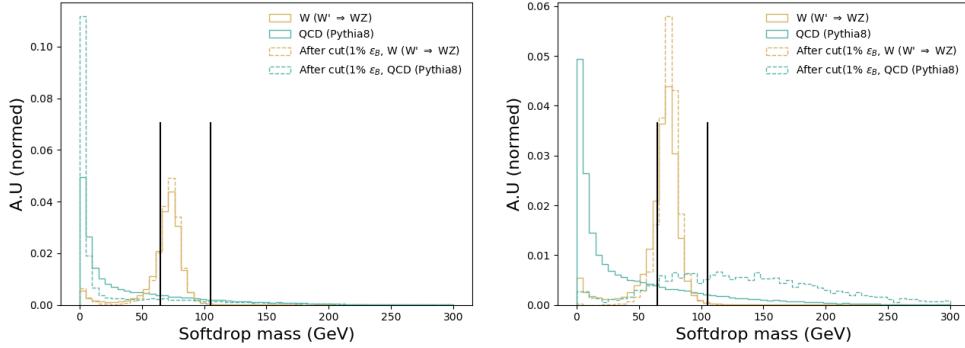


Figure 5.13: The softdrop jet mass distribution before (solid lines) and after (dotted lines) a cut on τ_{21} (left) and τ_{21}^{DDT} (right). All spectra are normalized to unit area.

2966 I have not yet had the chance to implement a mass decorrelation strategy for LoLa, but I
 2967 see two ways going forward: The first is, following the example of what was done to decorrelate
 2968 LoLa from jet p_T , to pass a mass dependent sample weight to the training. LoLa would then be
 2969 trained with a weight derived to flatten the two dimensional jet mass - jet p_T plane. Another
 2970 option would be to train LoLa together with an adversarial, a dedicated deep neural network
 2971 running in parallel to LoLa and attempting to learn the jet mass from the LoLa output. The
 2972 total loss function would then be a sum of the two, where the better the adversarial is in learning
 2973 the mass, the worse the total loss function gets. Both these options are something I'd like to

2974 explore in the future.

2975 In summary, mass- and p_T -dependence are in their own right not a problem for a tagger.
2976 The problem occurs when using these taggers in actual physics analyses where background rate
2977 uncertainties and tagging p_T dependence uncertainties has a large impact on the final sensitivity.
2978 There is a trade-off between signal efficiency and (analysis-dependent) systematics. For LoLa,
2979 rather than choosing, I'd like to provide two different taggers: A nominal tagger, where no mass/ p_T -
2980 decorrelation is attempted, and a decorrelated version. Then both can be tested in a full analysis
2981 chain before deciding on which tagger to use when looking at data.

2982 5.5 Validation on an independent sample

2983 LoLa is additionally validated on independent samples as an unbiased measure of performance
2984 allowing to compare different CMS W-tagging algorithms to one another: A $Z' \rightarrow WW$ sample
2985 with $M_{Z'} = 3\text{ TeV}$ produced with MadGraph and a QCD PYTHIA 8 sample in a p_T bin of 1000
2986 to 1400 GeV. Here, only jets with $1000\text{ GeV} < p_T < 1400\text{ GeV}$ and $|\eta| < 1.5$ are used. The
2987 signal efficiency versus mistagging rate for LoLa compared to the baseline PUPPI Softdrop +
2988 τ_{21} tagger, is shown in Figure ???. As was pointed out in Section ???, a mass cut is not necessary
2989 when using LoLa, but has been added to this plot for completeness. A significant improvement in
2990 tagging efficiency is observed for LoLa compared to the default tagger, also when being validated
2991 on a sample completely independent from the training sample. The cut corresponding to a 30
2992 % signal efficiency working point are used as reference working points when we will look at
2993 the tagging performance as a function of jet p_T and pileup in the following, and is marked by
2994 triangles in the plot. The signal efficiency and mistagging rate as a function of jet p_T , is shown in
2995 Figure 5.15. Again we observe the strong correlation between LoLa tagging efficiency and jet
2996 transverse momenta. There is, however, no point in the spectra where the τ_{21} tagger has a higher
2997 signal over background ratio than LoLa. LoLa performs its worst at very high jet p_T , but in this
2998 region the background is very small (dijet invariant masses around 2.5-3 TeV) so the absolute
2999 performance here matters less than at lower p_T . Figure 5.16 shows the tagging efficiency and
3000 mistagging rate as a function of pileup. Both taggers under study are more or less decorrelated
3001 from pileup, with a flat efficiency up to 50 reconstructed primary vertices. In Run 3, this number
3002 is of course expected to be significantly higher, around 140-200, and the study should be redone
3003 up to a higher number of reconstructed primary vertices.

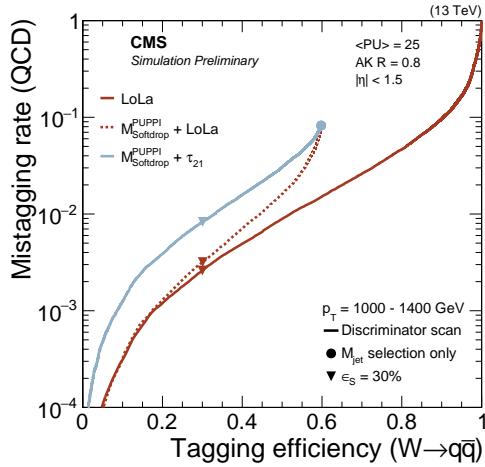


Figure 5.14: Performance of LoLa and PUPPI Softdrop + τ_{21} in the background-signal efficiency plane. The PUPPI softdrop jet mass selection of $65 < M_{SD} < 105 \text{ GeV}$, and the 30 percent efficiency points are indicated with symbols.

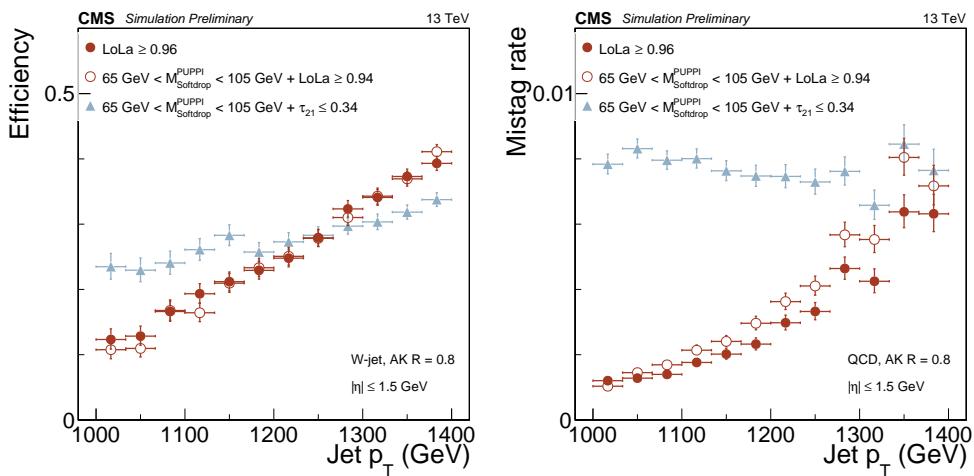


Figure 5.15: Efficiency (left) and mistag rate (right) of the LoLa selection corresponding to a 30 percent signal efficiency as a function of jet p_T .

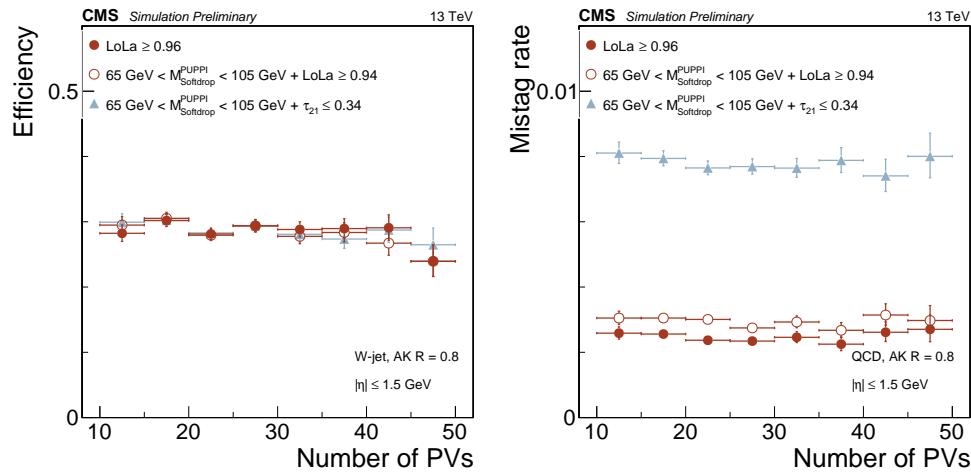


Figure 5.16: Efficiency (left) and mistagging rate (right) of the LoLa selection corresponding to a 30 percent signal efficiency as a function of the number of reconstructed vertices.

3004 5.6 Summary and outlook

3005 In this chapter, we have seen a promising new W-tagging algorithm for future VV searches. Its
3006 absolute performance is better than that of the baseline PUPPI softdrop + τ_{21} tagger up to a jet
3007 p_T of at least 1400 GeV, roughly corresponding to a dijet invariant mass of 2.5-3 TeV, and could
3008 lead to an increase in total signal efficiency from 18 to 43 % for the searches presented here. With
3009 a p_T decorrelation method already in place, it could already now be used for the one dimensional
3010 VV search presented in Search I and Search II. However, if to be used in the multidimensional
3011 search framework, a mass decorrelation method needs to be established. I have already outlined
3012 two possibilities of how to achieve this in Section 5.4.2, where one of these has already been
3013 shown to work in the context of p_T decorrelation. This is, as of this writing, left to future studies.

3014
3015 When discussing the future of the multidimensional search, I mentioned how a deep neural
3016 network such as the one presented here could be used to encode jet substructure in a way useful in
3017 order to develop a generic anti-QCD tagger. This has already been achieved by a parallel analysis
3018 team through the use of auto-encoders, published ten days before this writing and documented
3019 in [80], and has shown very promising results. However, this strategy is, to my knowledge after
3020 discussing with the authors, no longer pursued after observing that the auto-encoder version of
3021 LoLa was very difficult to decorrelate from the jet mass. It is my belief that this can be overcome
3022 by changing some of the features calculated in the Lorentz Layer (in [80], only the invariant
3023 mass is calculated and the other features listed in Equation 5.4 are stripped away) and this is
3024 something I would also like to pursue in future studies in order to achieve the truly generic search
3025 for boosted dijet resonances in the $M_{\text{jet}1}$ - $M_{\text{jet}2}$ - M_{VV} plane.

3026

CHAPTER 6

3027

Conclusion and outlook

3028 In this doctoral thesis I have presented three searches for new heavy ($> 1\text{TeV}$) resonances decaying
3029 to two vector bosons in the all-hadronic final state, using datasets collected by the CMS experiment
3030 corresponding to an integrated luminosity of 2.7(2015), 35.9(2016) and 77.3(2016+2017) fb^{-1} .
3031 Due to the high energy (“boost”) of the vector bosons, their decay products are so collimated
3032 that they get merged into one single jet, leading to a dijet final state topology. Dedicated jet
3033 grooming and jet substructure techniques are therefore explored in order to discriminate vector
3034 bosons from the overwhelming QCD multijet background.

3035

3036 Each of the analyses presented has provided original contributions to the field: The first search
3037 was the first of its kind to ever be performed at $\sqrt{s} = 13\text{ TeV}$, following an observed excess of
3038 3.4 (1.3) σ by ATLAS (CMS) in the 8 TeV dataset, and the first time CMS demonstrated the
3039 efficiency of using jet grooming techniques at trigger level. It, at the time, set the most stringent
3040 limits to date for the signal scenarios under consideration.

3041 The second search introduces a novel pileup resistant and perturbative safe vector boson tagging
3042 algorithm based on PUPPI softdrop jet mass, ensuring a high and stable signal efficiency up to
3043 pileup of at least 50 interactions per event. The optimization, validation and full commissioning
3044 of the tagger was performed in the context of this search. Dedicated jet mass corrections in order
3045 to account for an observed p_T and η dependence in PUPPI softdrop jet mass, due to the nature
3046 of the softdrop algorithm, were also developed. The PUPPI softdrop based tagger, together with
3047 the jet mass corrections, became, and still is, the recommended algorithm for W-tagging in CMS.
3048 The final analysis introduces a band new way of doing diboson resonance searches through a tree
3049 dimensional fit of the dijet invariant mass and the softdrop mass of the two jets. By optimization
3050 of the W-tagging algorithm used in this search and the nature of the search method, we have,
3051 for the first time, been able to extract the jet mass scale and resolution for a merged $W(\bar{q}q)$
3052 and $Z(\bar{q}q)$ peak from the V+jets Standard Model process. The method itself leads to a 20-30 %
3053 higher sensitivity than the default search method.

3054

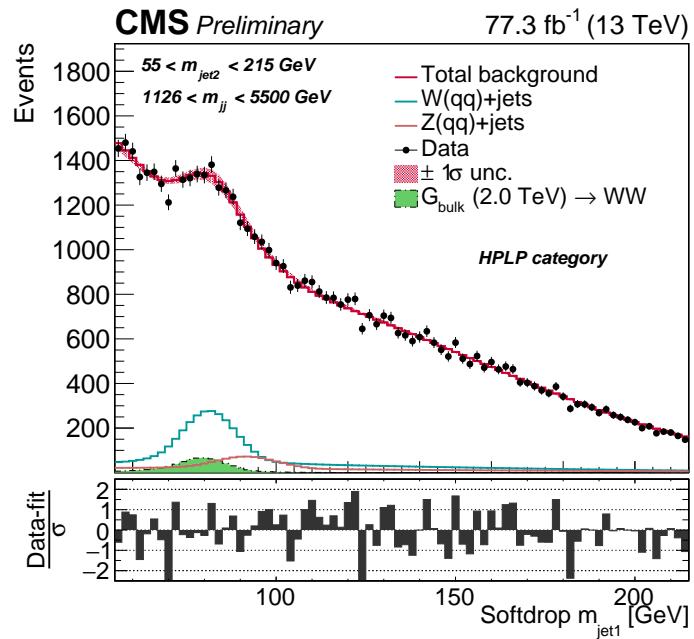
3055 The benefit of using a three-dimensional fit based on dijet invariant mass and the softdrop
3056 jet masses of the two jets, is that one can look for resonances peaking anywhere in the jet
3057 mass and dijet invariant mass spectrum. The natural next step for this search is therefore the
3058 incorporation of $VH(bb)$ and $H(bb)H(bb)$ final states into the three-dimensional fit framework,
3059 where orthogonality is guaranteed through b-tagging categories. This process is ongoing and
3060 planned for the full Run 2 dataset (including the data collected in 2018).

3061 Going further, one can incorporate searches for generic resonances peaking anywhere in the
3062 softdrop jet mass and dijet invariant mass spectrum in the multi-dimensional fit, where the

3063 jets themselves can have other substructure compositions than two subjets. This type of model
 3064 independent search requires a generic anti-QCD tagger in order to be truly model independent.
 3065 In the final chapter of this thesis I presented ongoing work on a deep neural network based
 3066 W-tagging algorithm for future searches, capable of more than doubling the analysis signal
 3067 efficiency by incorporating jet substructure algorithms within the deep layers. As there will be no
 3068 center-of-mass energy increase after the LHC reaches 14 TeV, achieving the best possible analysis
 3069 sensitivity for the dataset to come will be of key importance. I also showed how such a deep
 3070 neural network model is the ideal starting point for building a signal independent anti-QCD tagger.

3071

3072 With $\sim 80 \text{ fb}^{-1}$ of 13 TeV data analyzed and no excess observed, the future of this search
 3073 therefore lies in: Increasing the analysis sensitivity through novel taggers, as the one I have
 3074 presented here, and, making the search strategy as generic as possible through multi-dimensional
 3075 scans and generic anti-QCD taggers, both of which have been built a foundation for in this
 3076 doctoral work..



APPENDIX A

Search I: Limits per mass category

A.1 Limits per mass category

The asymptotic limits obtained with 2.6 fb^{-1} of 13 TeV CMS data per mass and purity category.

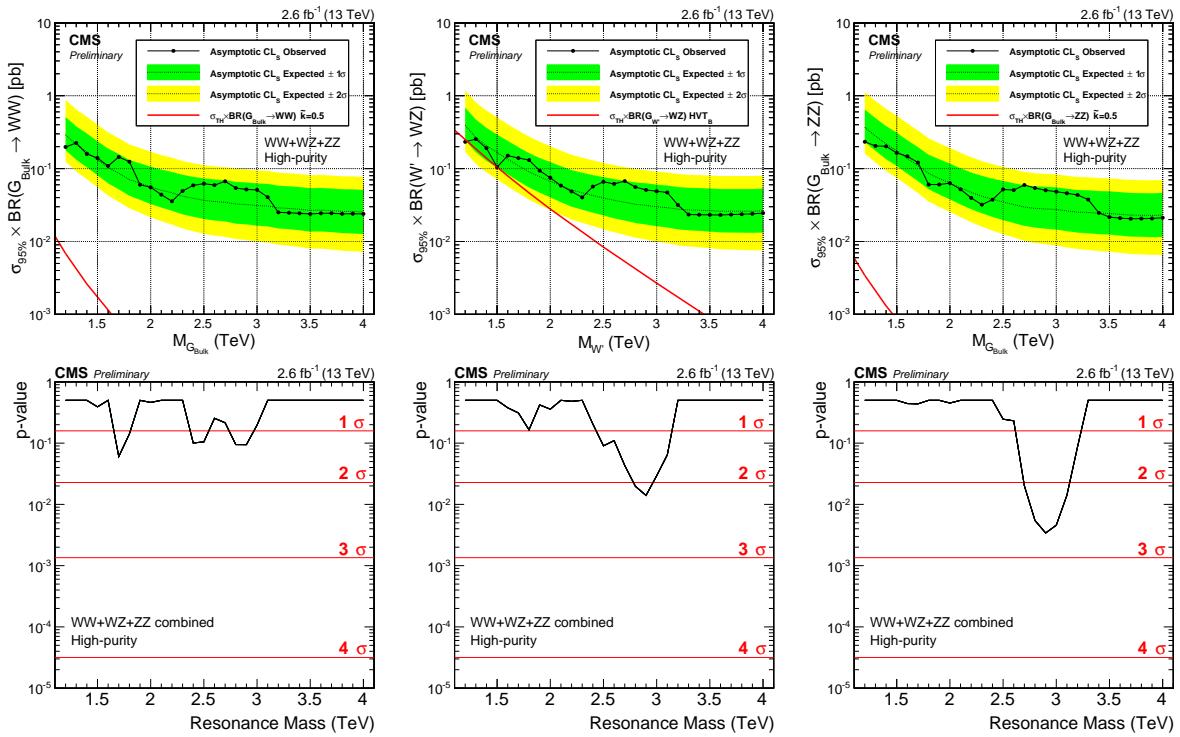


Figure A.1: Expected/observed limits and corresponding p-values obtained in the high purity category using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle) and $G \rightarrow ZZ$ (right) signal.

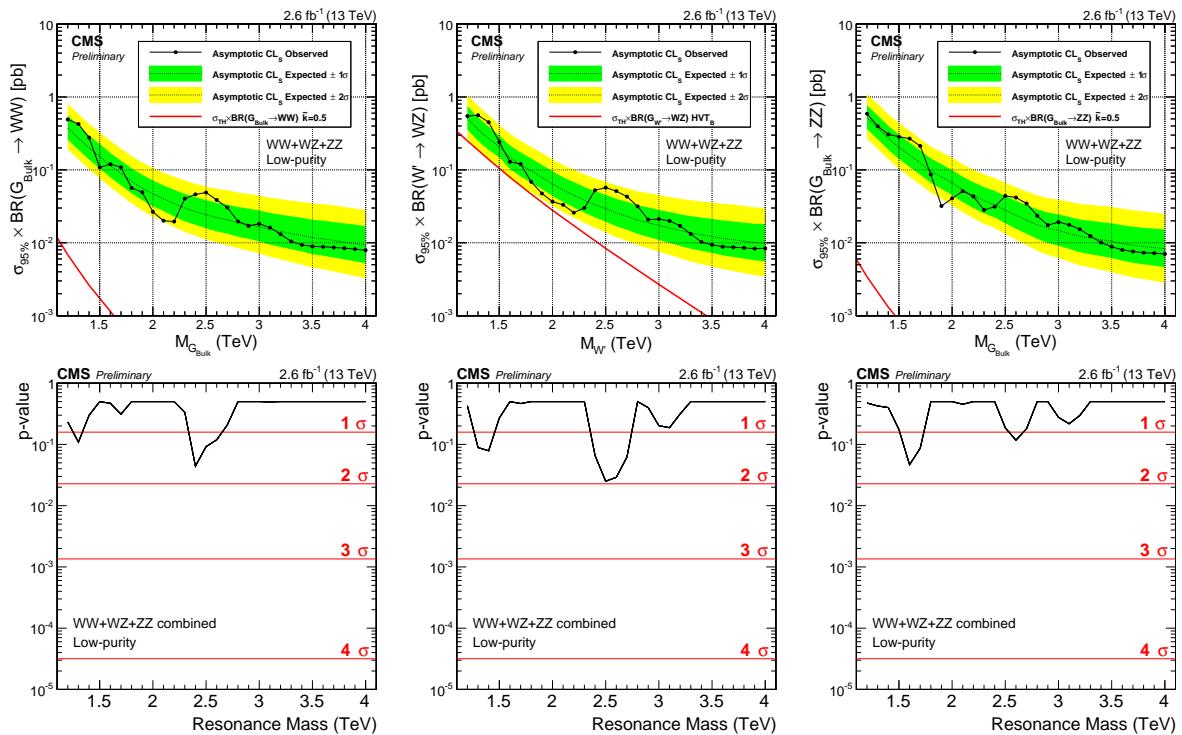


Figure A.2: Expected/observed limits and corresponding p-values obtained in the low purity category using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle) and $G \rightarrow ZZ$ (right) signal.

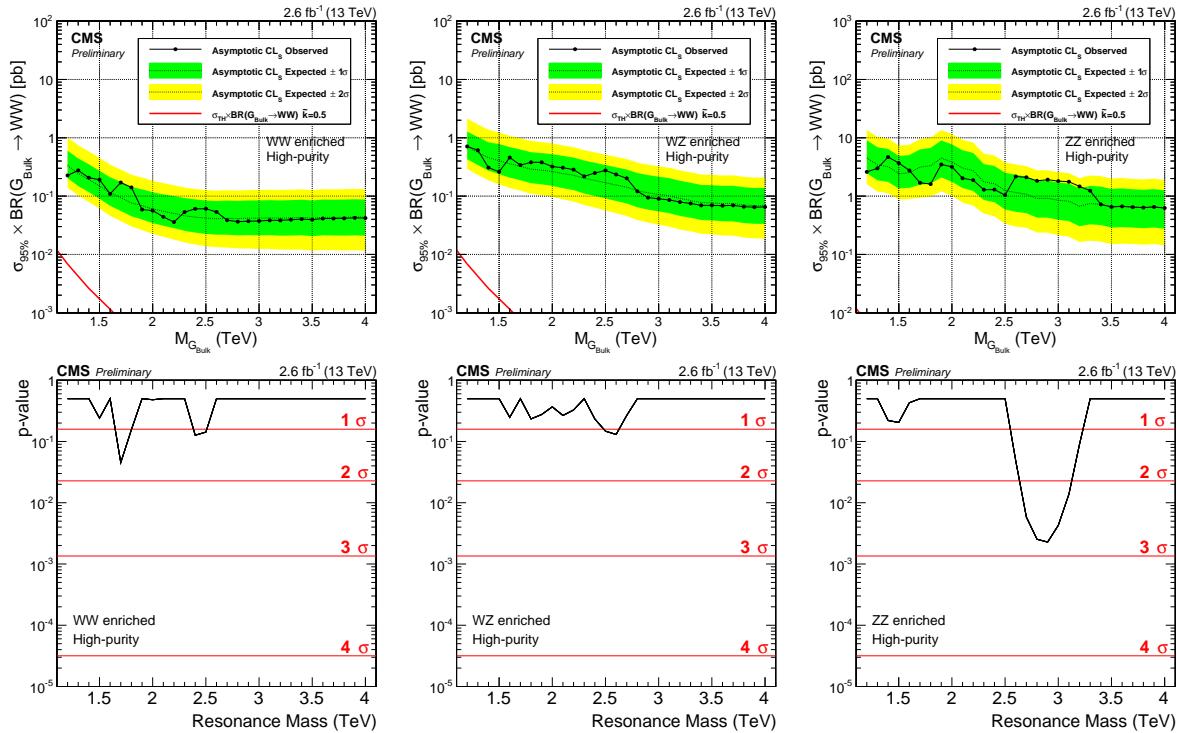


Figure A.3: Expected/observed limits and corresponding p-values obtained for the different mass categories using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ signal in the HP category

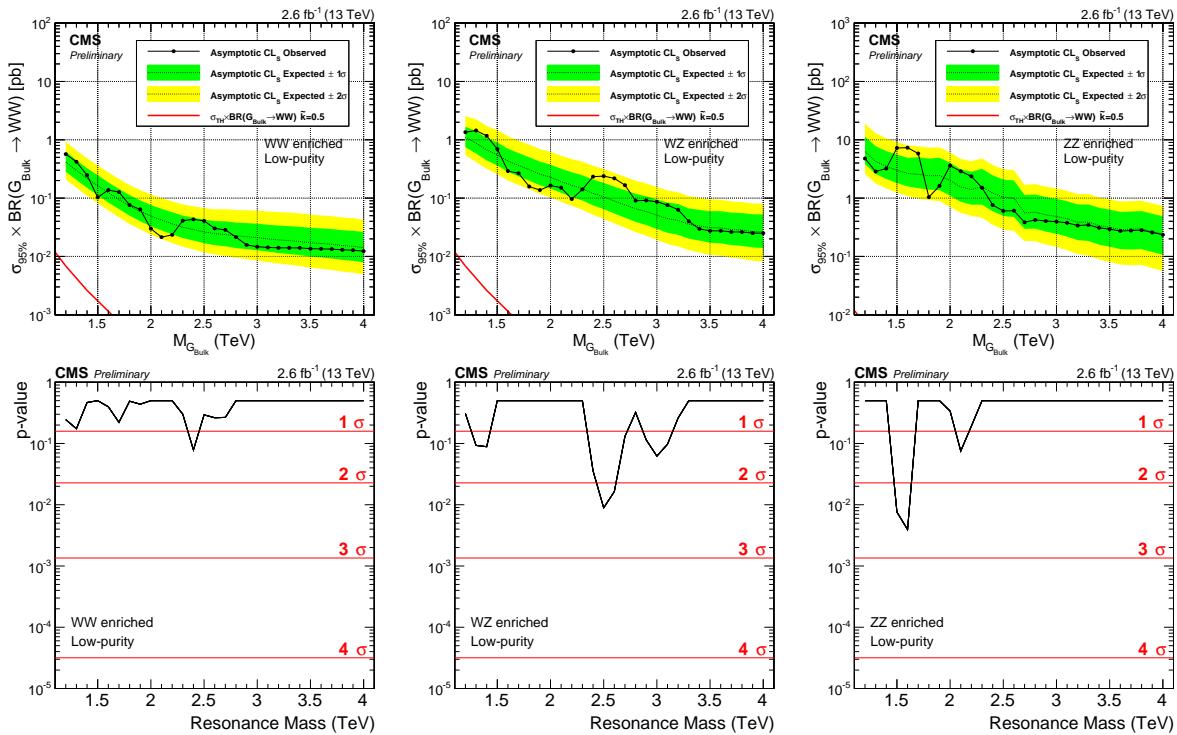


Figure A.4: Expected/observed limits and corresponding p-values obtained in the different mass categories using 2.6 fb⁻¹ of CMS data. Here for a Bulk $G \rightarrow WW$ signal in the LP category.

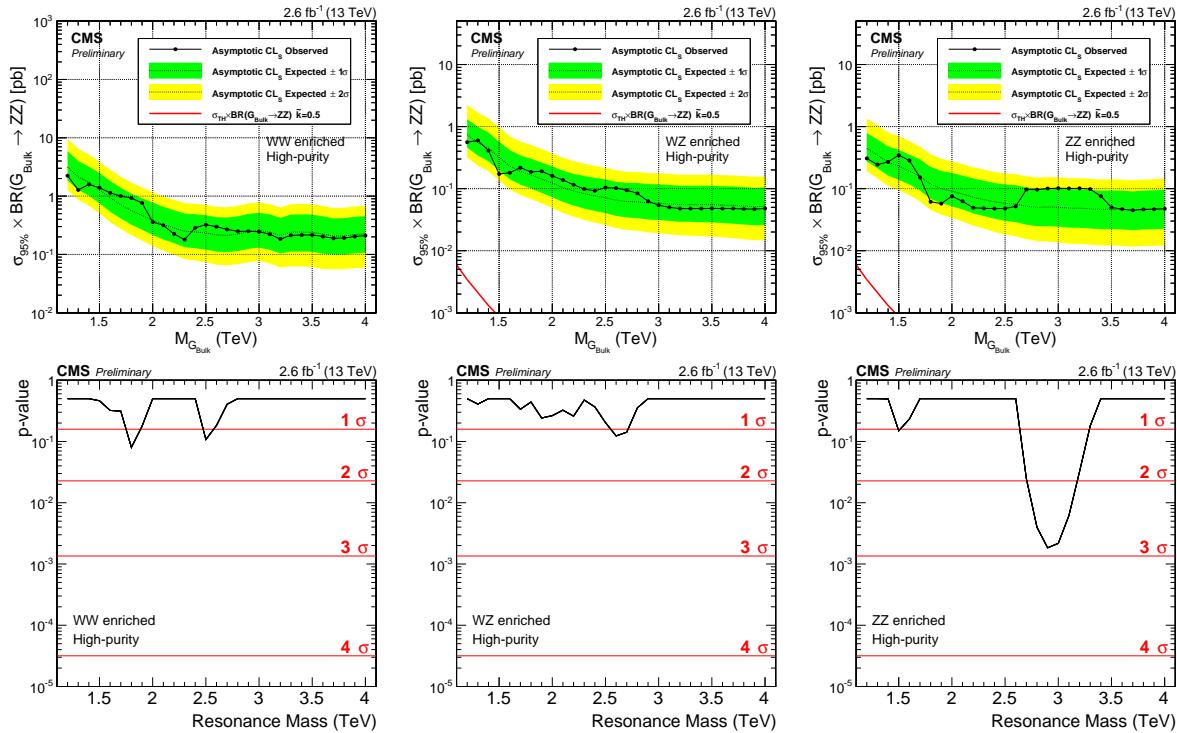


Figure A.5: Expected/observed limits and corresponding p-values obtained in the different mass categories. Here for a $G \rightarrow ZZ$ signal in the HP category.

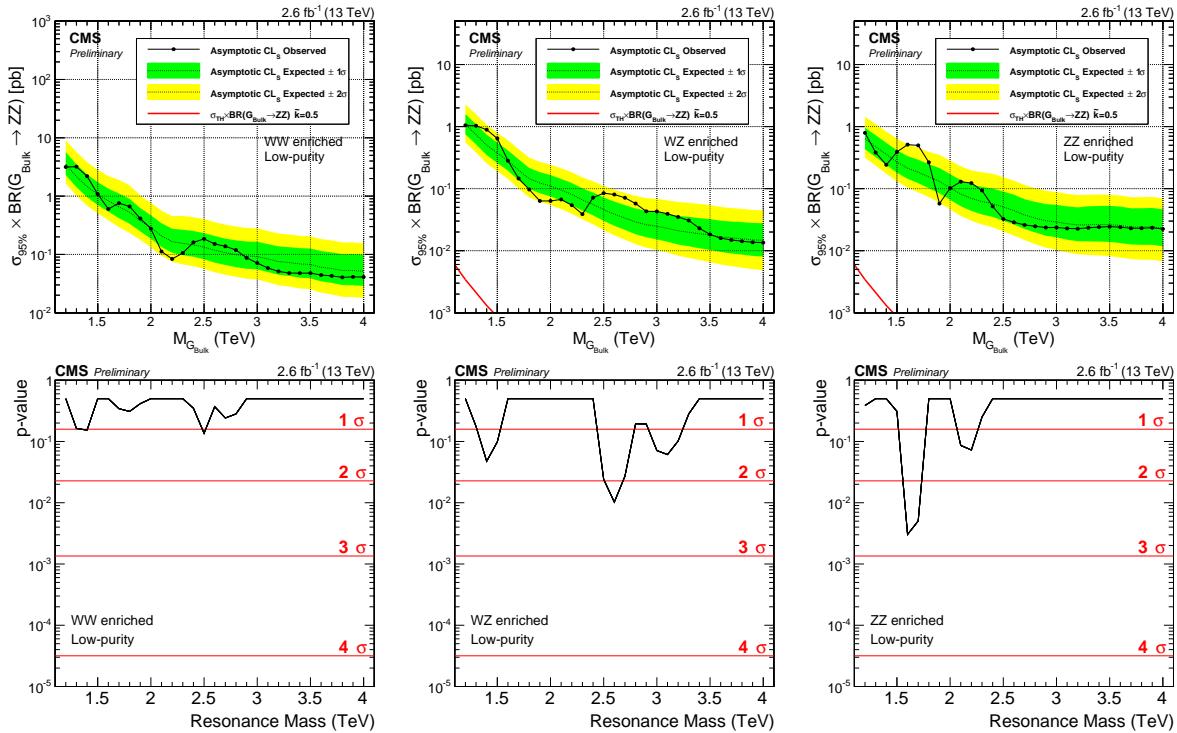


Figure A.6: Expected/observed limits and corresponding p-values obtained in the different mass categories. Here for a $G \rightarrow ZZ$ signal in the LP category.

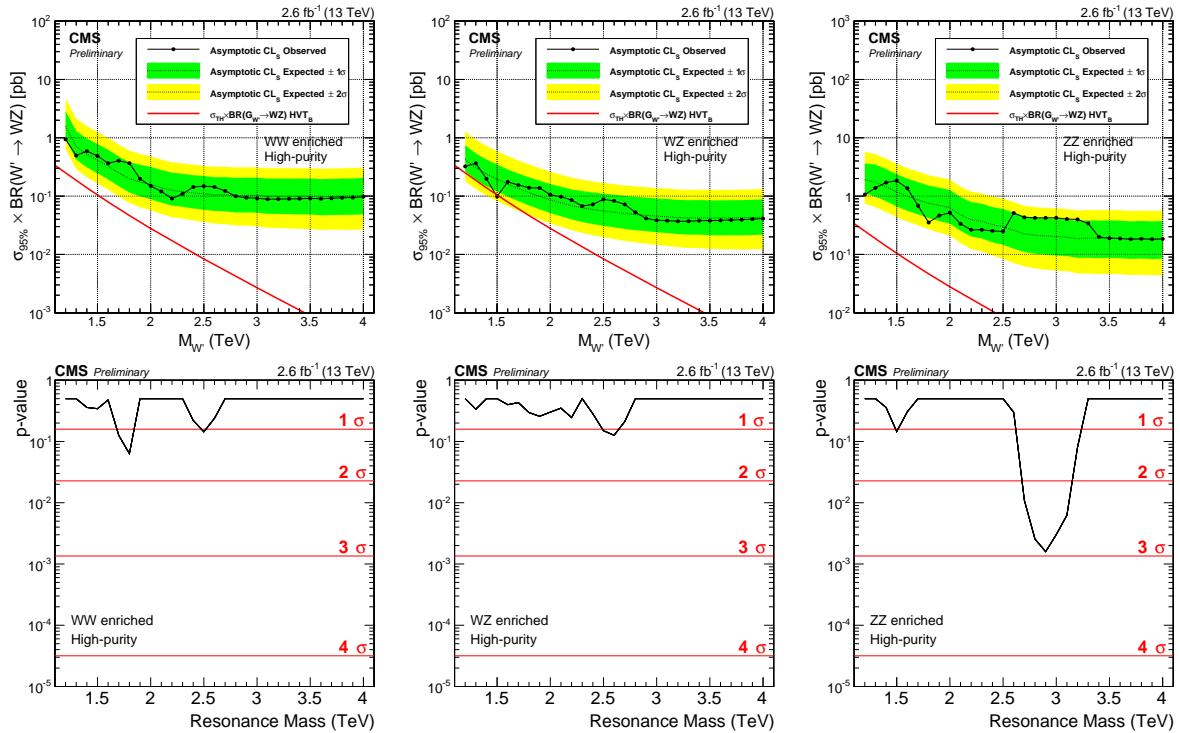


Figure A.7: Expected/observed limits and corresponding p-values obtained in the different mass categories. Here for a $W' \rightarrow WZ$ signal in the high-purity category.

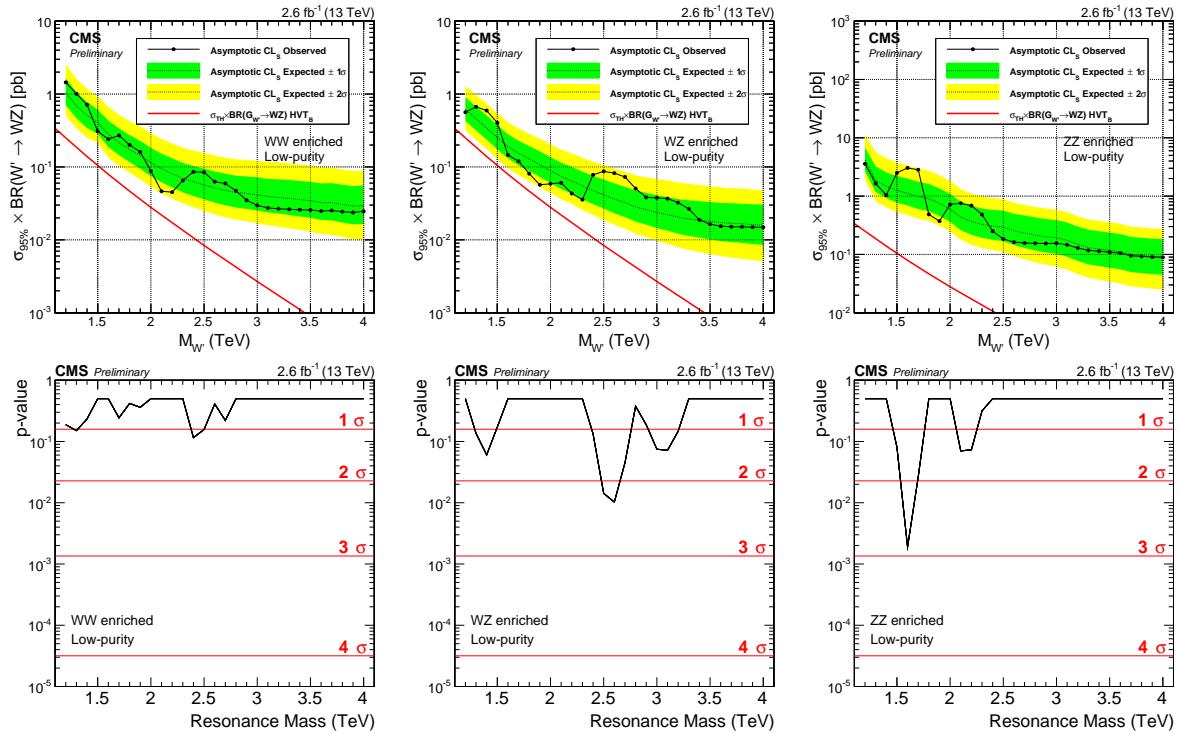


Figure A.8: Expected/observed limits and corresponding p-values obtained in the different mass categories. Here for a $W' \rightarrow WZ$ signal in the low purity category.

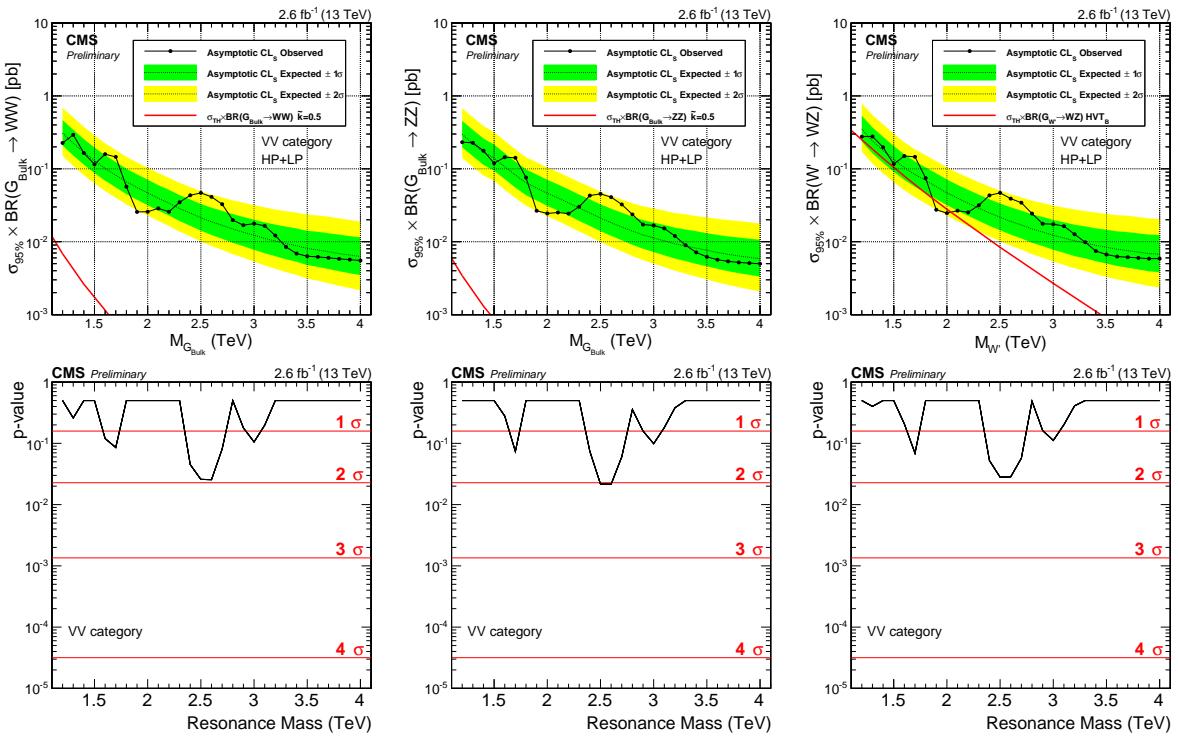


Figure A.9: Expected/observed limits and corresponding p-values obtained without splitting into mass categories. This analysis is performed as a cross check analysis and directly compares with the method used in the corresponding Run 1 analysis [?]. Here for a Bulk $G \rightarrow WW$ (left), $G \rightarrow ZZ$ (middle) and $W' \rightarrow WZ$ signal (right).

3082

APPENDIX B

3083

Search II

3084 B.1 W-tagging scale factor measurement: Additional plots

3085 Figures B.1 show the $t\bar{t}$ real W (top) and non-W (bottom) PUPPI softdrop jet mass distributions
 3086 for jets that passed (left) and failed (right column) the N-subjettiness selections $\text{PUPPI } \tau_{21} < 0.40$.
 Figures B.2 shows the fitted PUPPI softdrop jet mass distributions for the non-dominant

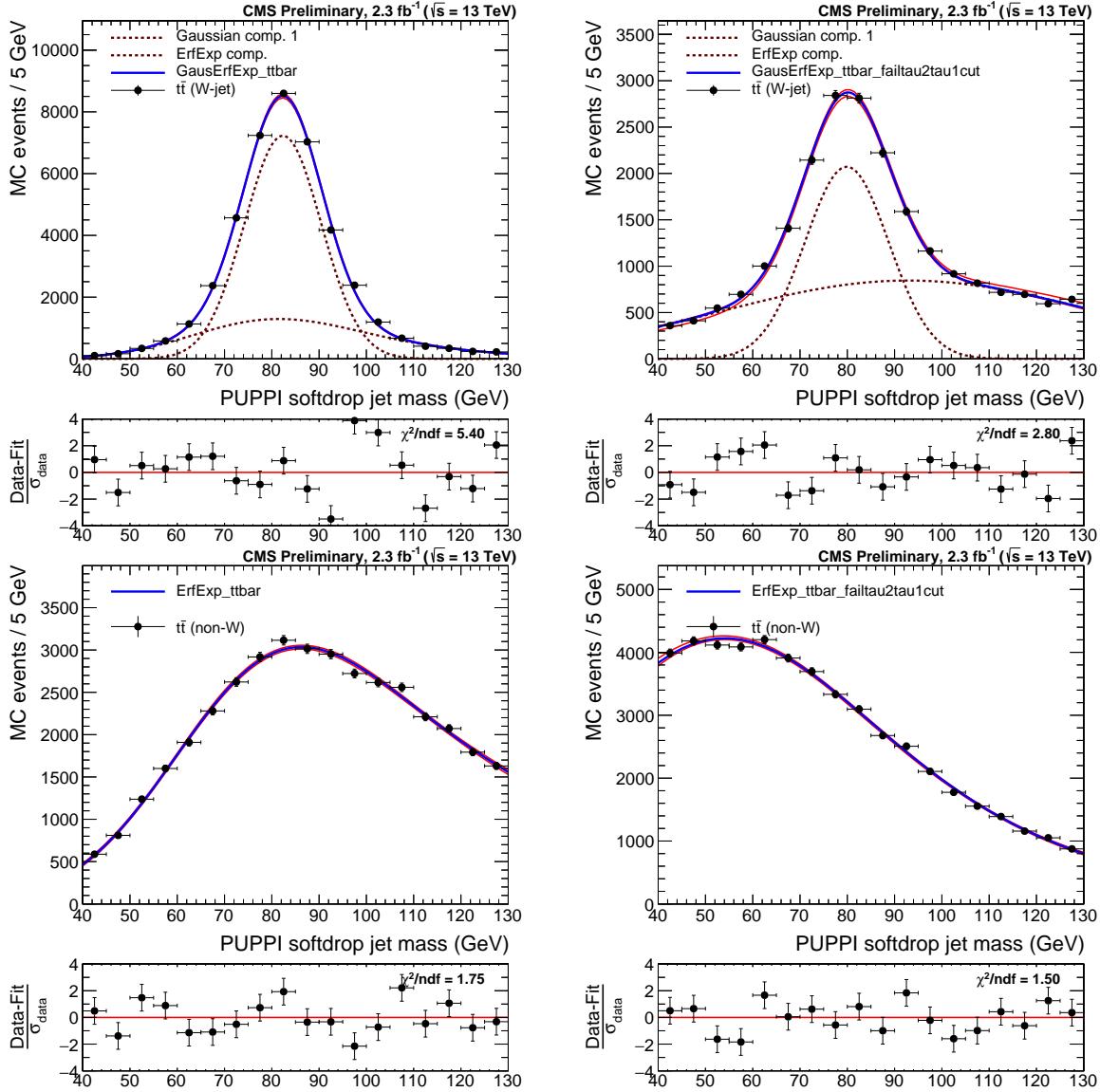


Figure B.1: Fit to the real W (top) and non-W (bottom) softdrop jet mass distribution for jets that pass (left) and fail (right) the cut on $\text{PUPPI } \tau_{21} < 0.4$.

3087

3088 backgrounds in the evaluation of the W-tagging scale factors. Here for jets that pass (top) and

failed (bottom) the N-subjettiness selections PUPPI $\tau_{21} < 0.40$.

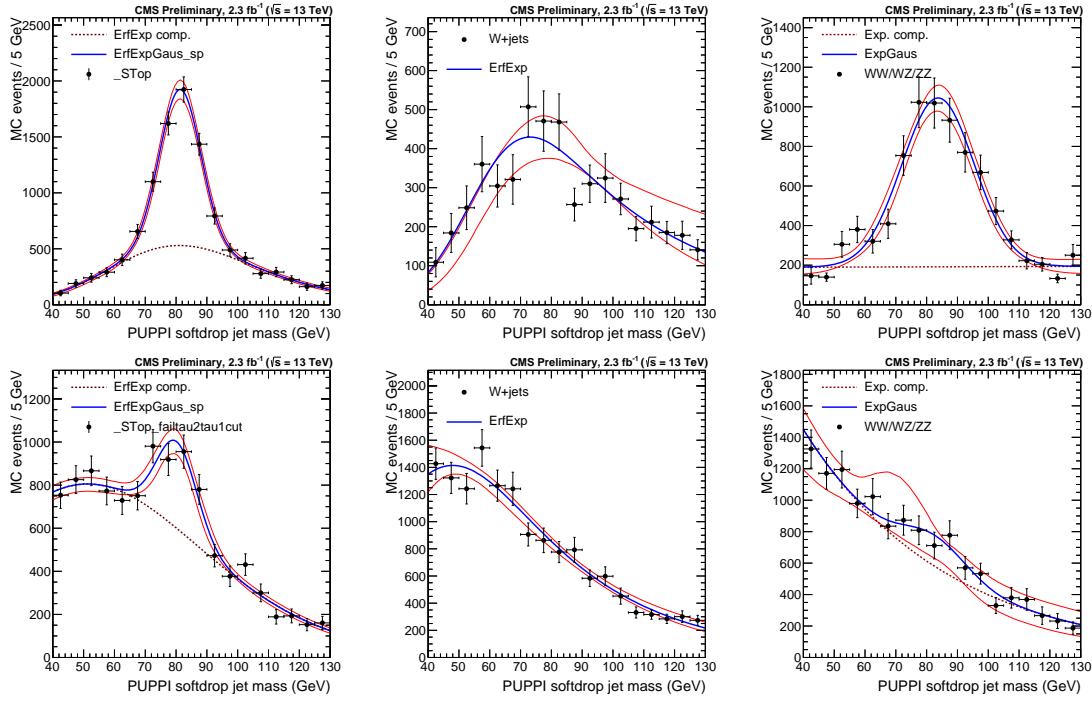


Figure B.2: Fits to the PUPPI softdrop jet mass spectrum for the non-dominant backgrounds (Single top, W+jets and VV respectively) in the pass (top) and fail (bottom) regions.

3089

3090 B.2 Efficiency scale factors for 2.5 fb^{-1}

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	$0.45 < \tau_2/\tau_1 < 0.75$	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 B.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} .

3091 B.3 Background fit checks

The background from QCD multijet events is modelled by a smoothly falling distribution in each analysis category. The method consists of a smoothness test of the observed data where the

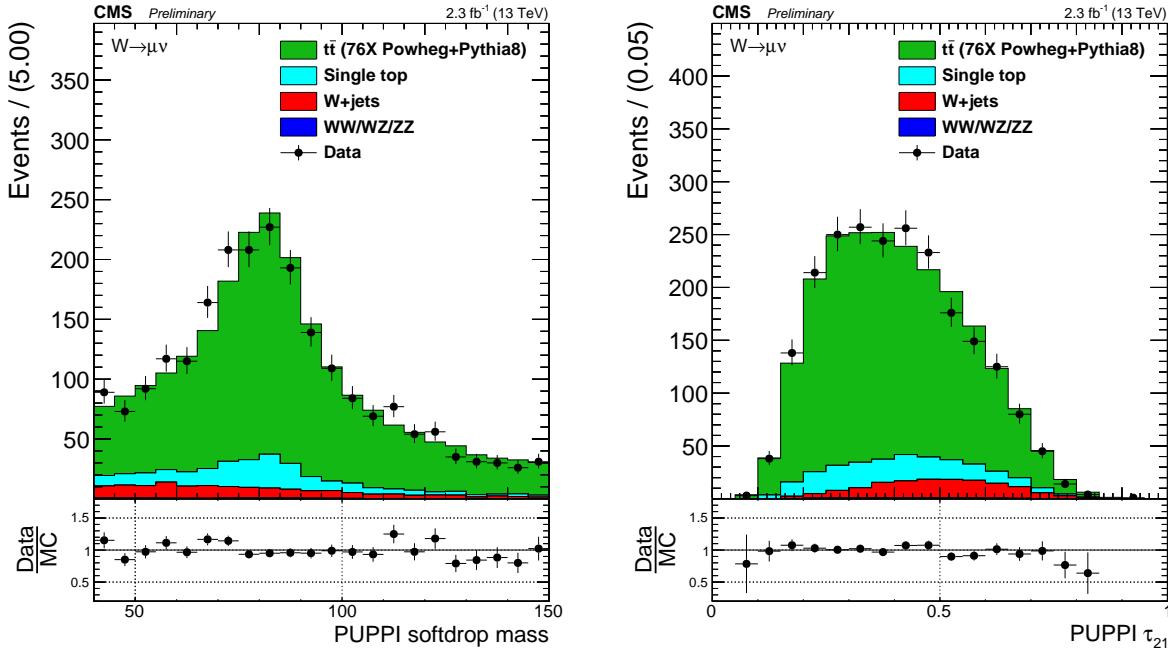


Figure B.3: Distribution of the PUPPI softdrop mass (left) and PUPPI n-subjettiness (right) distribution in the $t\bar{t}$ control sample.

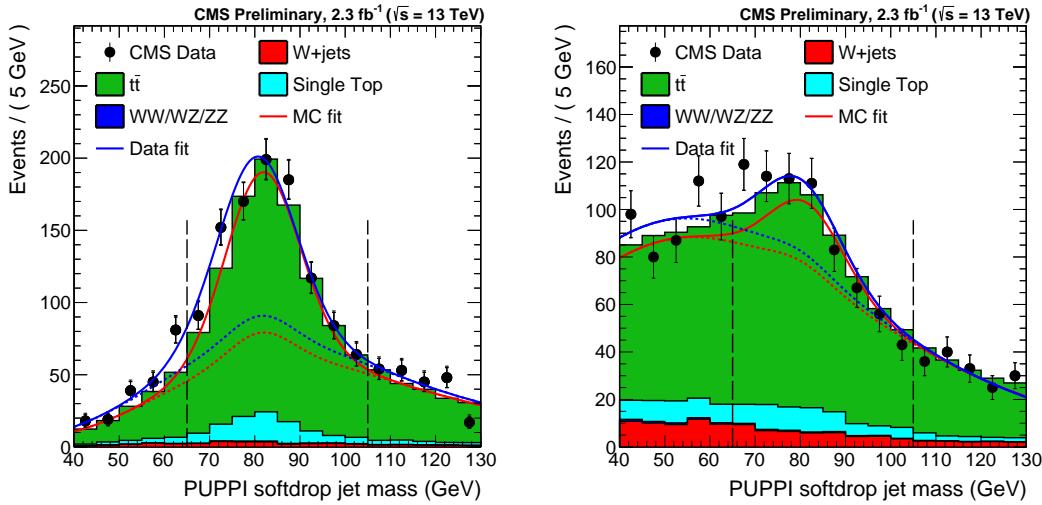


Figure B.4: PUPPI softdrop jet mass distribution that pass (left) and fail (right) the PUPPI $\tau_2/\tau_1 < 0.40$ selection. Results of both the fit to data (blue) and simulation (red) are shown and the background components of the fit are shown as short-dashed lines.

Parameter	Data	Simulation	Data/Simulation
PUPPI softdrop $\langle m \rangle$	80.3 ± 0.8 GeV	81.9 ± 0.01 GeV	0.98 ± 0.01
PUPPI softdrop σ	9.0 ± 0.9 GeV	8.5 ± 0.4 GeV	1.07 ± 0.12

Table B.2: Summary of the fitted W-mass peak fit parameters.

background is assumed to be described by the following empirical probability density function:

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2}} \quad (\text{B.1})$$

where m is the dijet invariant mass, \sqrt{s} the centre of mass energy, P_0 is a normalisation 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 systematics 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 $20 \text{ GeV} < M_{\text{Softdrop}} < 65 \text{ GeV}$. In order to quantify how many parameters are necessary, a Fishers F-test [65] 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:

$$\frac{dN}{dm} = \frac{P_0}{(m/\sqrt{s})^{P_2}} \quad (\text{B.2})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2 + P_3 \times \log(m/\sqrt{s})}} \quad (\text{B.3})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2 + P_3 \times \log(m/\sqrt{s}) + P_4 \times \log(m/\sqrt{s})^2}} \quad (\text{B.4})$$

Additionally, fits with an alternative fit function has also been performed (for the single-tag categories we try both 4 and 5 parameter versions):

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s} + P_3(m/\sqrt{s})^2)^{P_1}}{(m/\sqrt{s})^{P_2}} \quad (\text{B.5})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s} + P_3(m/\sqrt{s})^2)^{P_1}}{(m/\sqrt{s})^{P_2 + P_4 \times \log(m/\sqrt{s})}} \quad (\text{B.6})$$

3107 B.3.1 Background fit checks in data sideband

3108 We perform a test in a data sideband to make sure the fit functions work on real data and
 3109 to exercise the estimation of number of necessary fit parameters via an F-test. The sideband
 3110 is constructed by requiring one of the two jets two have a mass in the low softdrop jet mass
 3111 sideband, between $20 \text{ GeV} < M_{\text{Softdrop}} < 65 \text{ GeV}$, while the full W/Z-tag selections are applied to
 3112 the other jet. The low-mass jet is also required to pass the τ_{21} cut corresponding to the given
 3113 category. For the single-tag category, the sideband is constructed by requiring one of the two jets
 3114 to have a mass in the low softdrop jet mass sideband, between $20 \text{ GeV} < M_{\text{Softdrop}} < 65 \text{ GeV}$ and
 3115 the other in a high-mass sideband, between $105 \text{ GeV} < M_{\text{Softdrop}} < 200 \text{ GeV}$. One of the jets is
 3116 also required to pass the τ_{21} cut corresponding to the given category. We first check whether
 3117 the sideband can be used to exercise the F test by checking whether or not there are features
 3118 introduced in the dijet mass spectrum that may be hard to cover with the fit using QCD MC. To
 3119 do so we look at the dijet invariant mass spectrum in the signal region divided by the distribution
 3120 in the sideband. The obtained distributions are shown in Figure B.5 for three different generators,
 3121 where the pure Pythia8 QCD samples (top left) has the highest statistics. The distributions are
 3122 mostly smooth, but we do see features introduced in the “WW HP” and “ZZ HP” categories
 3123 which might prove difficult to fit, as well as in the tail of the single-tag categories. These kinks
 3124 shift around depending on what MC generator is used and do not seem to be a systematic feature
 3125 caused by the cuts that have been applied, but rather due to limited statistics. We proceed with
 3126 exercising the F-test in a data sideband, with the caveat that there might be features introduced
 3127 in the spectrum where statistics are low.

3128 Figure B.6 shows the fit to data in the data sideband for the WW and ZZ mass categories,
 3129 both in the HP and in the LP n-subjettiness categories. The corresponding residuals, χ^2 and
 3130 F-test results are shown i Table B.7 through B.10. For the double-tag categories, a two or three
 3131 parameter function is sufficient to describe the data and we conclude that the function as defined
 3132 in B.1 is sufficient for all mass categories. For the single-tag category a five parameter fit seems
 3133 to be required in order to describe the data and the fit quality is not optimal. To ensure that the
 3134 fit functions with sufficient number of parameters is able to describe the shape in the single-tag
 3135 categories, we have additionally looked at the fit quality in QCD MC (see below). Here, the
 3136 default dijet function seems sufficient to describe the distributions. The sideband in the single-tag
 3137 categories in QCD MC do not show the same features as the data sideband as shown in Figure
 3138 B.7.

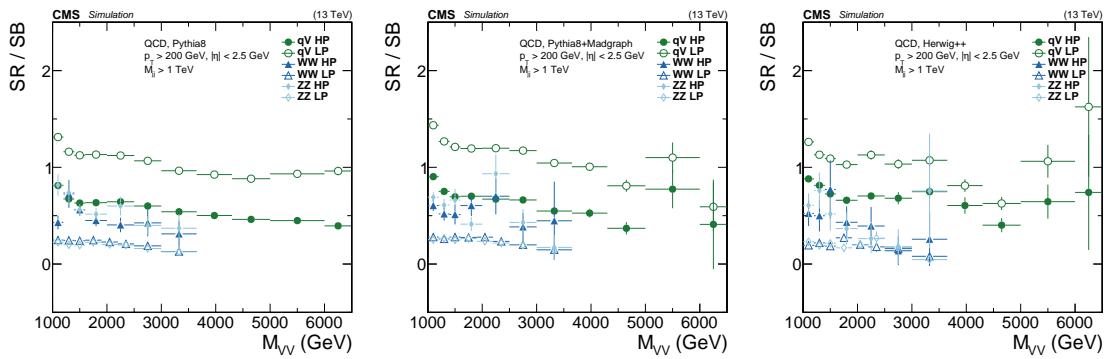


Figure B.5: Dijet mass spectrum in the signal region divided by the dijet mass spectrum in the sidebands using QCD Pythia8 (left), QCD Pythia8+Madgraph (middle) and QCD Herwig++ (right) simulated samples. Here for the double W/Z-tag and the single V-tagged HP and LP categories. Some jumps are observed 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.

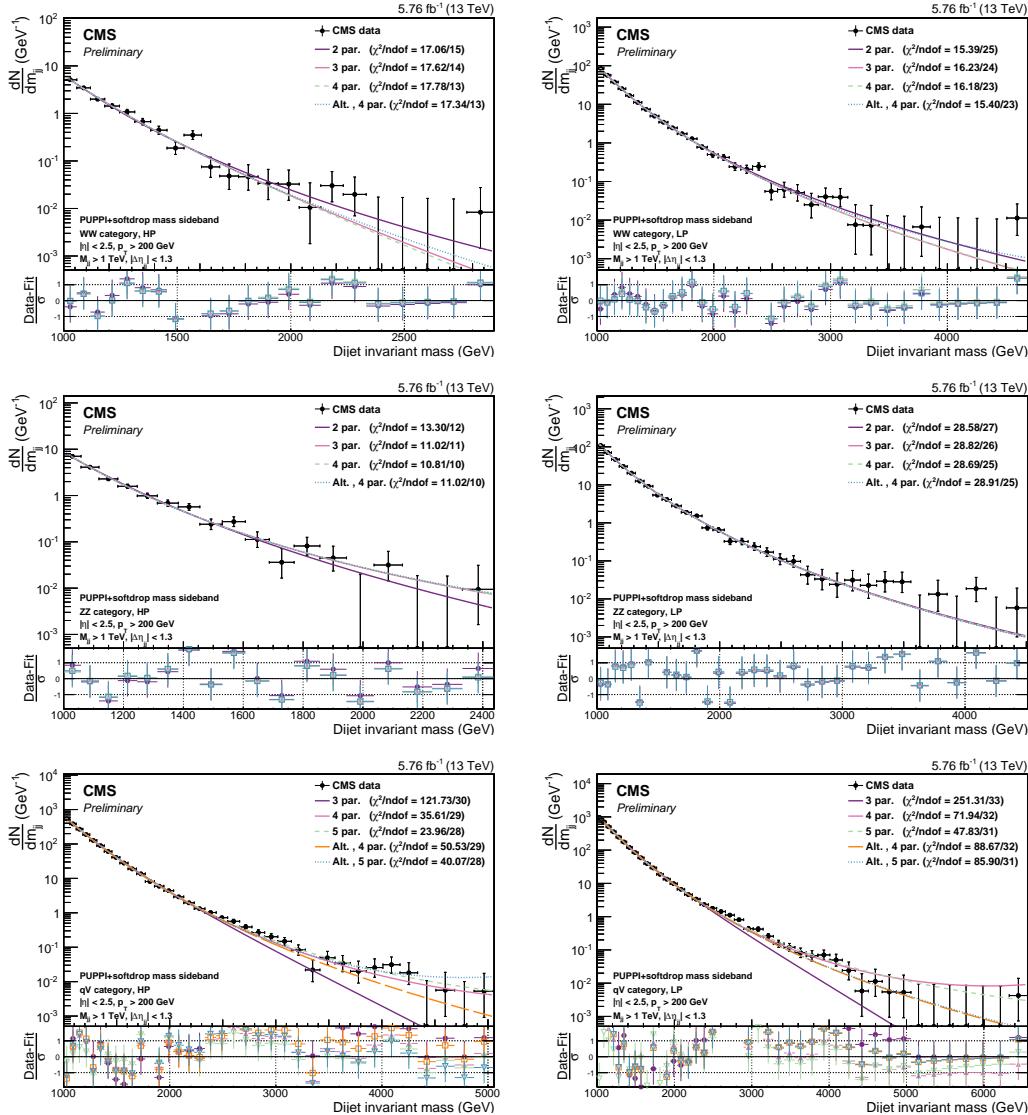


Figure B.6: Fitted dijet mass spectrum in the different mass and purity categories in a data sideband: WW high-purity (top left) and low-purity (top right), ZZ high-purity (middle left) and low-purity (middle right), qV high-purity (bottom left) and low-purity (bottom right).

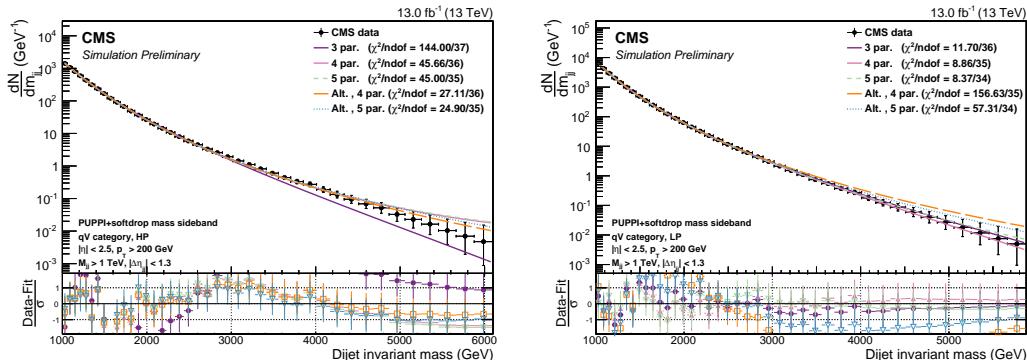


Figure B.7: Fitted dijet mass spectrum in the QCD MC sideband: qV high-purity (left) and low-purity (right).

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 B.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 B.4: Residuals, χ^2 , and degrees of freedom for the WW category, LP category. A 3 parameter fit is needed to describe these data.

3139 B.3.2 Background fit checks in QCD MC

3140 As an additional check, we look at the fit functions in the different signal categories using QCD
3141 MC. This is shown in Figure B.8 for the double and Figure B.9 for the single tag categories. Here
3142 all fit functions and their pull distributions are plotted. The fits are performed in a mass range
3143 corresponding to the expected distribution in the different categories for 13 fb^{-1} of data. We
3144 have adapted the error bars to correspond to the maximum of the expected Poisson error for 13
3145 fb^{-1} of data and the pure simulation error (accounting for the different weights assigned to the
3146 p_T -binned QCD MC sample). The reason for this choice is to get an estimate of whether the
3147 set of fit functions we plan to use to fit the background distribution in data, and plan to use in
3148 order to understand the systematic uncertainty due to our choice of fit function, are appropriate
3149 and do not produce fake bumps/kinks. As this distribution is the pure MC simulation curve,
3150 whose variation at high masses is much smaller than the expected poisson error for 13 fb^{-1} of
3151 data, we expect the χ^2/ndof to be lower than one. In order to protect against the fact that the
3152 MC simulation at lower dijet masses does not have more statistics than the expected data for

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 B.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 B.6: Residuals, χ^2 , and degrees of freedom for the ZZ category, LP category. A 2 parameter fit is needed to describe these data.

3153 13 fb^{-1} of data, we use the largest of the Poisson and the MC error. The resulting errors are
 3154 therefore a mixture of Poisson and the MC error and the χ^2/ndof for the QCD MC fits should
 3155 not be considered. Fit quality in the form of χ^2/ndof should only be estimated from the fits to
 3156 data sideband where pure Poisson errors are used Overall the fits to QCD MC in the different
 3157 categories describe the data well, with a two or three parameter function sufficient to describe the
 3158 distributions. However, due to an under fluctuation of the first bin in the “ZZHP” category, the
 3159 higher parameter fits are steered by the first bin leading to discrepancies in the tail. This is the
 3160 lowest statistics category and the danger for underfluctuations does exist. We have investigated
 3161 the dijet invariant mass distribution down to a dijet invariant mass threshold of 800 GeV to
 3162 make sure we are not seeing a turn-on effect. This is shown in Figure B.10. The fits in the data
 3163 sideband for the same category (Figure B.6, middle left) do not show the same trend.

B.4 F-test in signal region

3164 The final F-test performed in order to define the number of fit parameters to be used to fit the
 3165 background in each analysis category, is performed in the signal region. The resulting fits and

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 B.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 B.8: Residuals, χ^2 , and degrees of freedom for the qV category, LP category. A 5 parameter fit is needed to describe these data.

3167 F-test values for the double tag categories are shown in Figure B.11 and Tables B.9 to B.14.

3168 The F-test results for the single-tag category are listed in Tables B.15 to B.18. Here, a three
 3169 parameter fit is sufficient for all categories except the “high-purity” qW category where a 5
 3170 parameter fit is preferred.

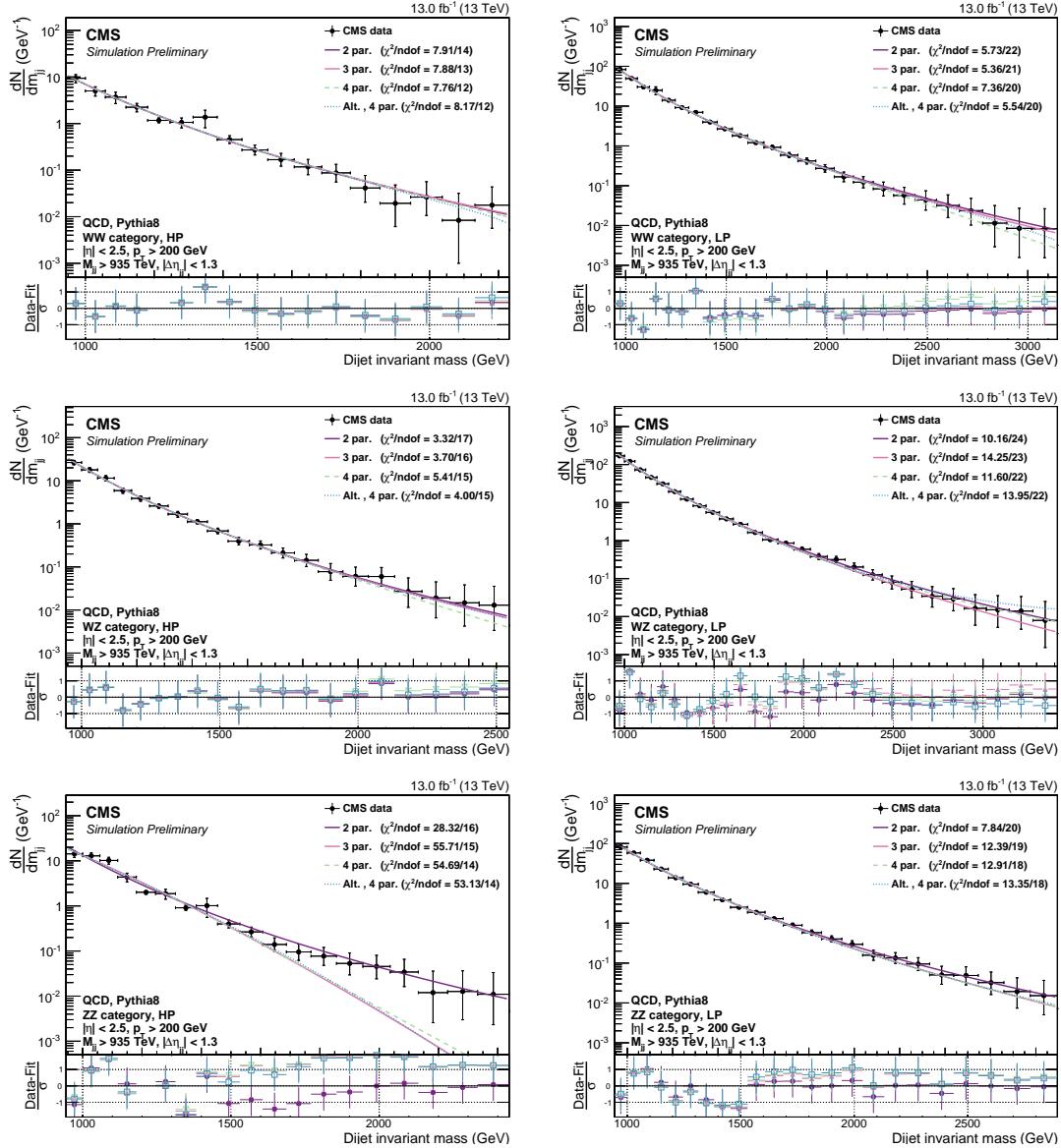


Figure B.8: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Here for the high- and low-purity double W/Z-tag category for the three different mass categories: WW category (top), WZ category (middle) and ZZ category (bottom).

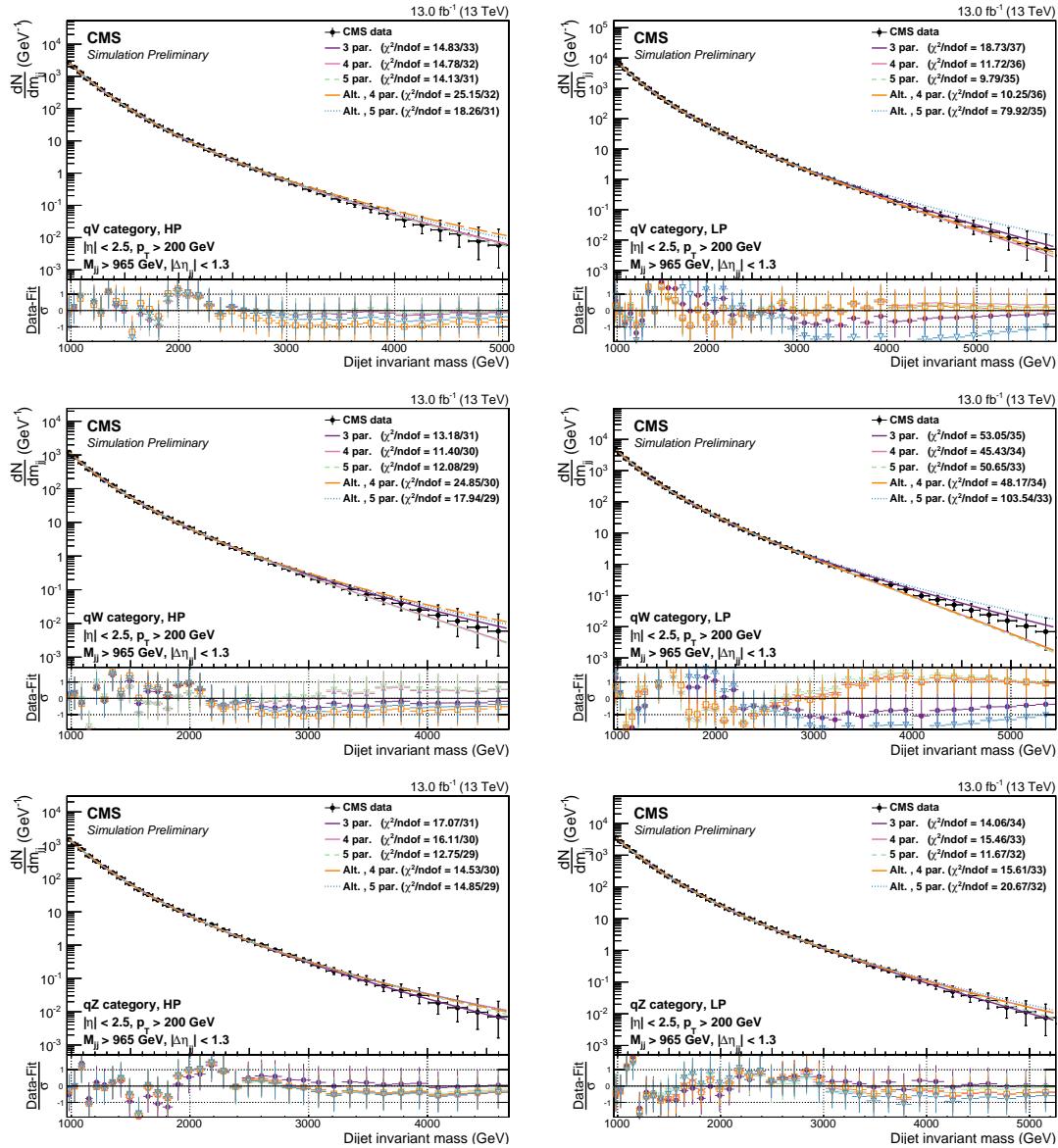


Figure B.9: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Here for the high- and low-purity single W/Z-tag category for the two different mass categories: wW category (top) and qZ category (bottom).

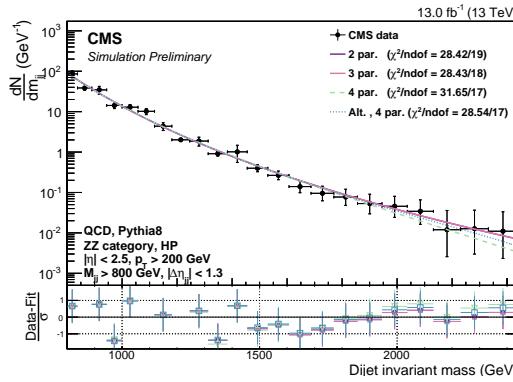


Figure B.10: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Here for the high-purity double Z-tag category using a dijet invariant mass threshold of 800 GeV. No turn-on effect at low invariant masses is observed.

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 B.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 B.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 B.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 B.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 B.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 B.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 B.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 B.16: Residuals, χ^2 , and degrees of freedom for the qW category, LP category. A 3 parameter fit is needed to describe these data.

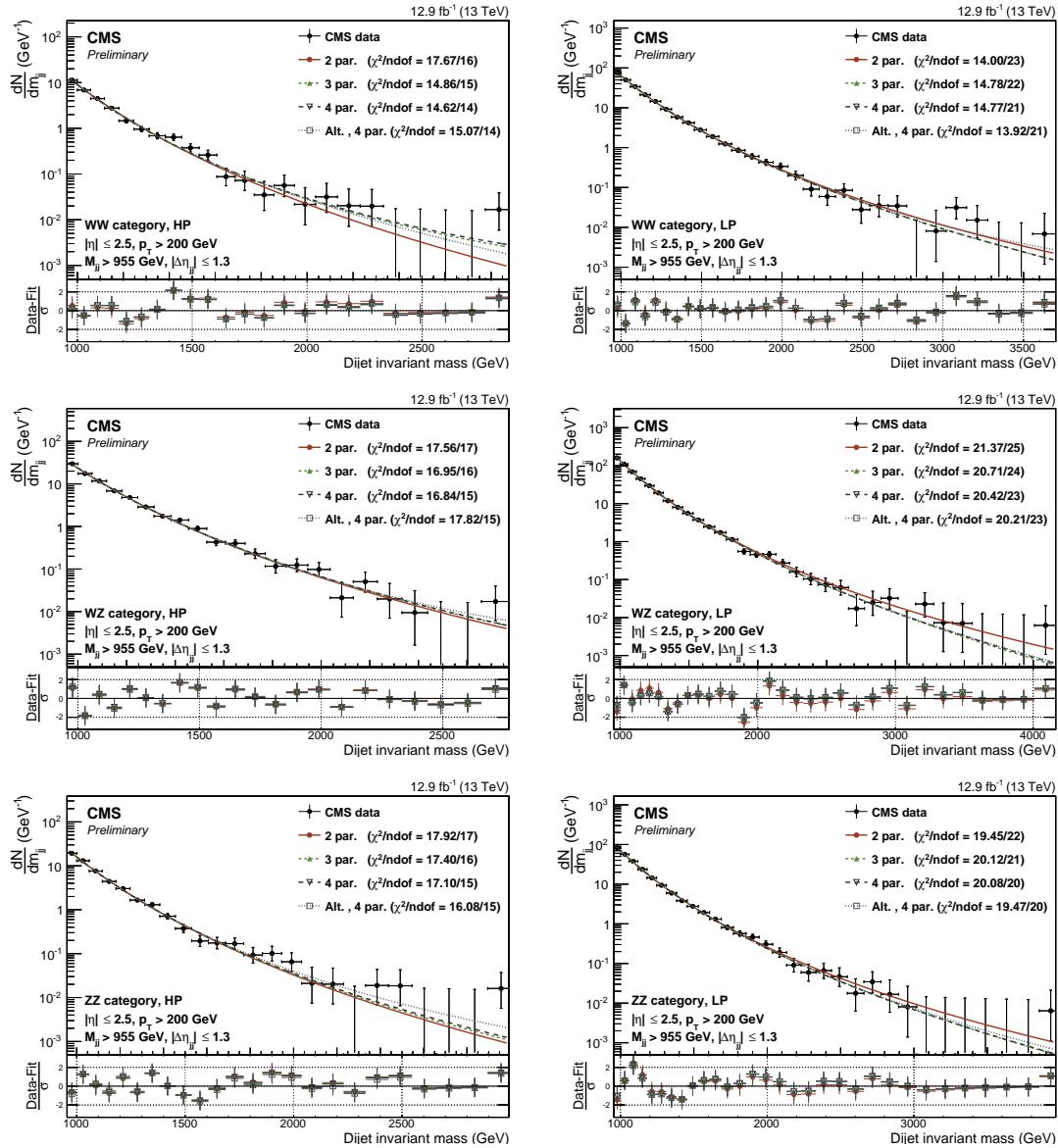


Figure B.11: Background fit for the M_{jj} distribution in the data signal region. Here for the high- (left) and low-purity (right) double W/Z-tag categories WW (top), WZ (middle) and ZZ (bottom).

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 B.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 B.18: Residuals, χ^2 , and degrees of freedom for the qZ category, LP category. A 3 parameter fit is needed to describe these data.

3171

APPENDIX C

3172

Search III

3173 C.1 Signal fits

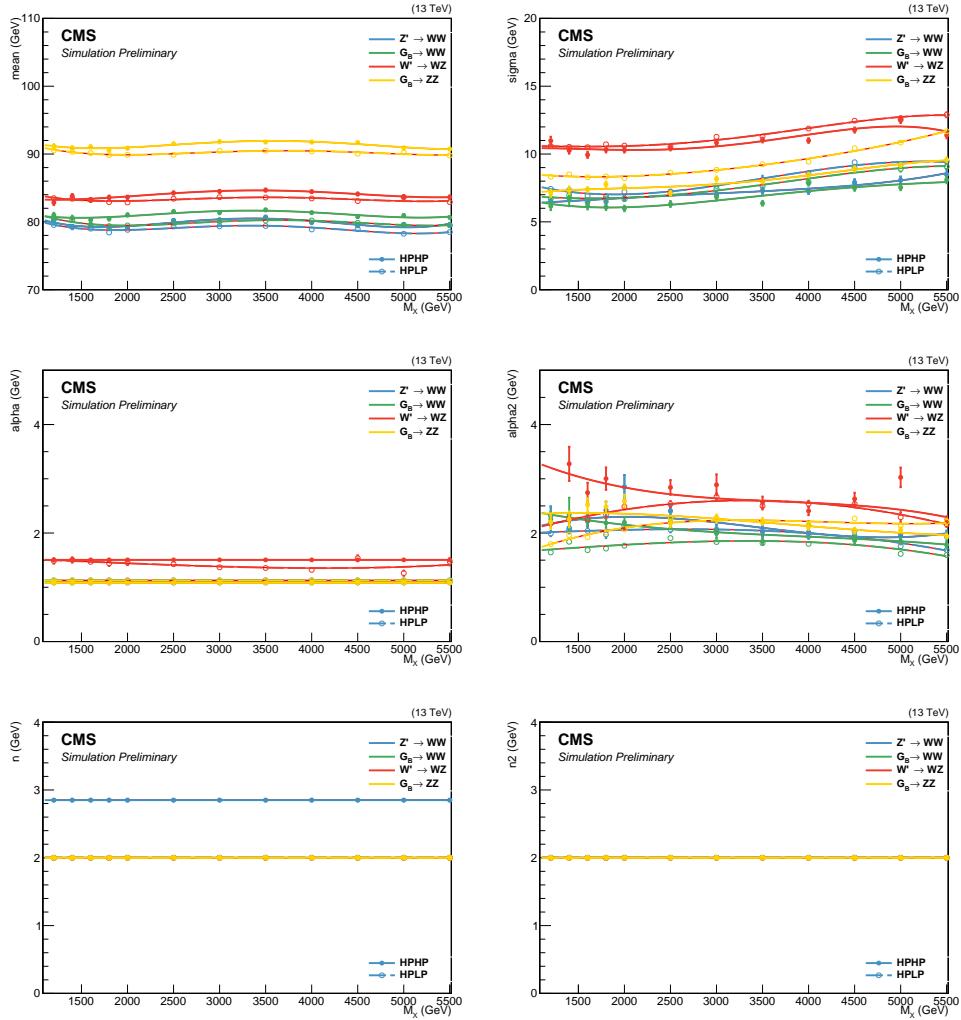


Figure C.1: The interpolated double Crystal-ball parameters for the softdrop jet mass as a function of M_X . To improve the stability of the fit some parameters are set constant. Here for jet 2.

3174 C.2 2016 kernels

3175 C.3 Resonant background

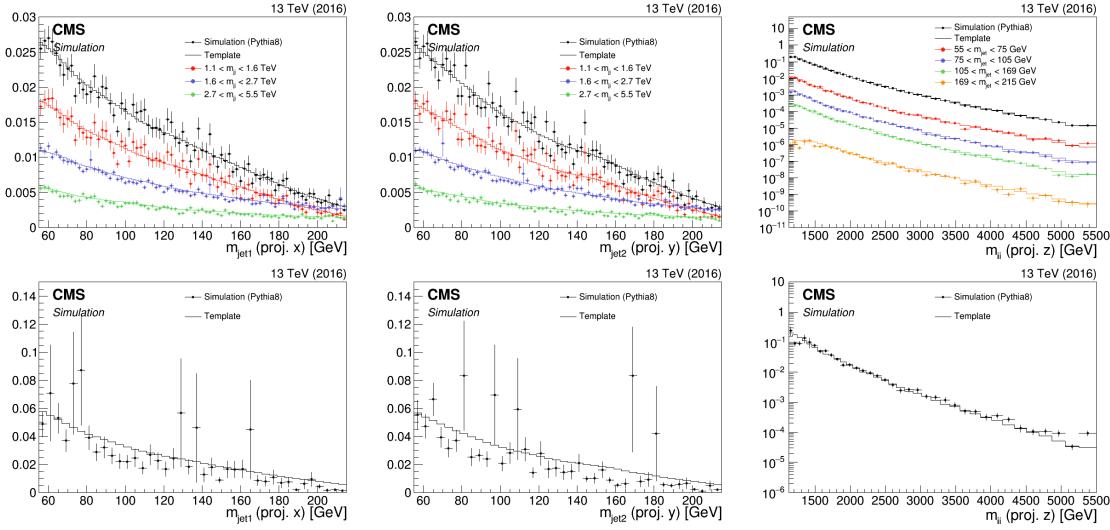


Figure C.2: Comparison between QCD MC simulation (markers) and kernels derived from generator level quantities (lines) for the HPHP category (top) and the HPLP category (bottom). The kernels are shown for $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right).

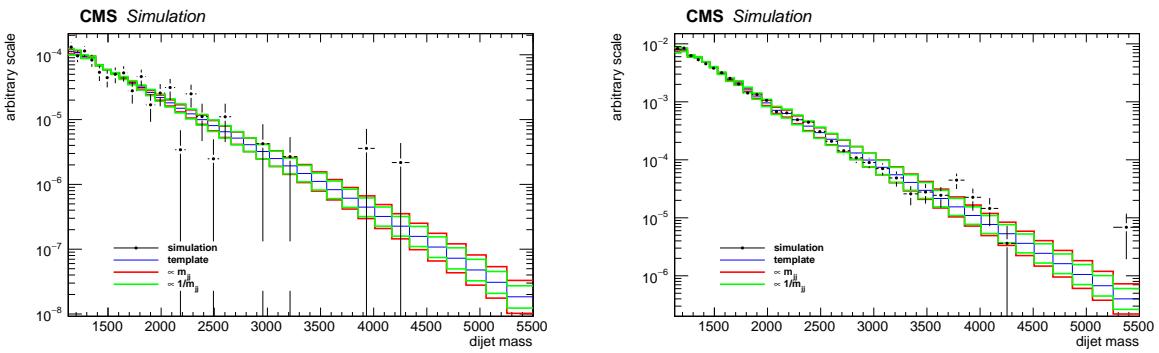


Figure C.3: One dimensional M_{VV} kernels (solid line) compared to MC (markers) for the $Z + \text{jets}$ background in the HPHP (left) and HPLP (right) categories. The nominal shape derived from the smoothing procedure can be seen as blue line, alternative shapes derived from varying the slope of the M_{VV} spectrum are shown in green and red.

3176

Bibliography

- 3177 [1] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys.* **C40** (2016),
3178 no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.
- 3179 [2] M. Gell-Mann, “Symmetries of Baryons and Mesons”, *Phys. Rev.* **125** (Feb, 1962)
3180 1067–1084, doi:10.1103/PhysRev.125.1067.
- 3181 [3] W. Pauli, “Mathematical contributions to the theory of Dirac’s matrices”, *Ann. Inst. H.
3182 Poincaré Phys. Theor.* **6** (1936) 109–136.
- 3183 [4] R. Tumulka, “Pauli Spin Matrices”, pp. 470–472. 07, 2009.
3184 doi:10.1007/978-3-540-70626-7_142.
- 3185 [5] S. Duplij and F. Klinkhamer, “Levi-Civita Symbol”, pp. 227–227. Springer Netherlands,
3186 Dordrecht, 2004. doi:10.1007/1-4020-4522-0_298.
- 3187 [6] S. Weinberg, “Conceptual Foundations of the Unified Theory of Weak and Electromagnetic
3188 Interactions”, *Rev. Mod. Phys.* **52** (1980) 515–523, doi:10.1103/RevModPhys.52.515.
3189 [,543(1979)].
- 3190 [7] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys.
3191 Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- 3192 [8] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13**
3193 (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- 3194 [9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless
3195 Particles”, *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.
- 3196 [10] P. W. Anderson, “Coherent Excited States in the Theory of Superconductivity: Gauge
3197 Invariance and the Meissner Effect”, *Phys. Rev.* **110** (May, 1958) 827–835,
3198 doi:10.1103/PhysRev.110.827.

- 3199 [11] J. Goldstone, “Field Theories with Superconductor Solutions”, *Nuovo Cim.* **19** (1961)
 3200 154–164, doi:10.1007/BF02812722.
- 3201 [12] J. Ellis, “Higgs Physics”, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School
 3202 of High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report.
- 3203 [13] B. Bellazzini, C. Csáki, and J. Serra, “Composite Higgses”, *Eur. Phys. J.* **C74** (2014),
 3204 no. 5, 2766, doi:10.1140/epjc/s10052-014-2766-x, arXiv:1401.2457.
- 3205 [14] R. Contino, D. Marzocca, D. Pappadopulo, and R. Rattazzi, “On the effect of resonances in
 3206 composite Higgs phenomenology”, *JHEP* **10** (2011) 081, doi:10.1007/JHEP10(2011)081,
 3207 arXiv:1109.1570.
- 3208 [15] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, “Heavy Vector Triplets: Bridging
 3209 Theory and Data”, *JHEP* **09** (2014) 060, doi:10.1007/JHEP09(2014)060,
 3210 arXiv:1402.4431.
- 3211 [16] L. Randall and R. Sundrum, “Large Mass Hierarchy from a Small Extra Dimension”, *Phys.*
 3212 *Rev. Lett.* **83** (Oct, 1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370.
- 3213 [17] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, “Warped gravitons at the CERN LHC
 3214 and beyond”, *Phys. Rev. D* **76** (Aug, 2007) 036006, doi:10.1103/PhysRevD.76.036006.
- 3215 [18] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, “Searching for the Kaluza-Klein
 3216 Graviton in Bulk RS Models”, *JHEP* **09** (2007) 013,
 3217 doi:10.1088/1126-6708/2007/09/013, arXiv:hep-ph/0701150.
- 3218 [19] CERN, “Accelerators and Schedules”, 2018 (accessed November 12, 2018),
 3219 <https://beams.web.cern.ch/content/accelerators-schedules>.
- 3220 [20] CMS, “CMS Detector Design”, 2018 (accessed November 21, 2018),
 3221 <https://cms.cern/news/cms-detector-design>.
- 3222 [21] T. Lenzi, “Development and Study of Different Muon Track Reconstruction Algorithms for
 3223 the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors”, Master’s thesis, U.
 3224 Brussels (main), 2013.
- 3225 [22] A. Dominguez et al., “CMS Technical Design Report for the Pixel Detector Upgrade”,
 3226 Technical Report CERN-LHCC-2012-016. CMS-TDR-11, CERN, Sep, 2012.
- 3227 [23] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004,
 3228 doi:10.1088/1748-0221/3/08/S08004.

- [24] 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.
- [25] 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.
- [26] L. V. . EP, “Assembly of the 7th wedge. Assemblage du 7eme module”, (Aug, 2000). CMS Collection.
- [27] S. Sharma, “Understanding the performance of CMS calorimeter”, *Pramana* **69** (Dec, 2007) 1069–1074, doi:10.1007/s12043-007-0229-8.
- [28] on behalf of the CMS Collaboration 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.
- [29] CMS, “CMS Luminosity - public results”, 2018 (accessed December 11, 2018),
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [30] 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).
- [31] A. S. et. Al., “Particle-flow reconstruction and global event description with the CMS detector”, *Journal of Instrumentation* **12** (2017), no. 10, P10003.
- [32] 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.
- [33] T. C. collaboration, “Performance of CMS muon reconstruction in pp collision events at s = 7 TeV”, *Journal of Instrumentation* **7** (2012), no. 10, P10002.
- [34] 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.
- [35] 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](https://arxiv.org/abs/hep-ph/9707323).

- 3258 [36] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”,
 3259 *Phys. Rev.* **D48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160,
 3260 arXiv:hep-ph/9305266.
- 3261 [37] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04**
 3262 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- 3263 [38] J. POG, “Jet Identification: Recommendations for 13TeV data analysis”, 2018 (accessed
 3264 December 18, 2018), <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID>.
- 3265 [39] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum
 3266 Resolution in CMS”, *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002,
 3267 arXiv:1107.4277.
- 3268 [40] D. Krohn, J. Thaler, and L.-T. Wang, “Jet Trimming”, *JHEP* **02** (2010) 084,
 3269 doi:10.1007/JHEP02(2010)084, arXiv:0912.1342.
- 3270 [41] CMS Collaboration Collaboration, “Pileup Removal Algorithms”, Technical Report
 3271 CMS-PAS-JME-14-001, CERN, Geneva, 2014.
- 3272 [42] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet
 3273 substructure”, *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.
- 3274 [43] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet substructure as a new
 3275 Higgs search channel at the LHC”, *Phys. Rev. Lett.* **100** (2008) 242001,
 3276 doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.
- 3277 [44] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05** (2014) 146,
 3278 doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.
- 3279 [45] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP*
 3280 **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.
- 3281 [46] J. Thaler and K. Van Tilburg, “Maximizing boosted top identification by minimizing
 3282 N-subjettiness”, *Journal of High Energy Physics* **2012** (Feb, 2012) 93,
 3283 doi:10.1007/JHEP02(2012)093.
- 3284 [47] S. Bolognesi et al., “Spin and parity of a single-produced resonance at the LHC”, *Phys.
 3285 Rev. D* **86** (Nov, 2012) 095031, doi:10.1103/PhysRevD.86.095031.
- 3286 [48] CMS Collaboration, “Identification techniques for highly boosted W bosons that decay into
 3287 hadrons”, *JHEP* **12** (2014) 017, doi:10.1007/JHEP12(2014)017, arXiv:1410.4227.

- 3288 [49] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J.* **C58** (2008) 639–707,
3289 [doi:10.1140/epjc/s10052-008-0798-9](https://doi.org/10.1140/epjc/s10052-008-0798-9), [arXiv:0803.0883](https://arxiv.org/abs/0803.0883).
- 3290 [50] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191**
3291 (2015) 159–177, [doi:10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024), [arXiv:1410.3012](https://arxiv.org/abs/1410.3012).
- 3292 [51] The ATLAS collaboration, “Search for high-mass diboson resonances with boson-tagged jets
3293 in proton-proton collisions at $s=8$ TeV with the ATLAS detector”, *Journal of High Energy*
3294 *Physics* **2015** (Dec, 2015) 1–39, [doi:10.1007/JHEP12\(2015\)055](https://doi.org/10.1007/JHEP12(2015)055).
- 3295 [52] S. Frixione and B. R. Webber, “Matching NLO QCD computations and parton shower
3296 simulations”, *JHEP* **06** (2002) 029, [doi:10.1088/1126-6708/2002/06/029](https://doi.org/10.1088/1126-6708/2002/06/029),
3297 [arXiv:hep-ph/0204244](https://arxiv.org/abs/hep-ph/0204244).
- 3298 [53] S. Frixione, P. Nason, and B. R. Webber, “Matching NLO QCD and parton showers in
3299 heavy flavor production”, *JHEP* **08** (2003) 007, [doi:10.1088/1126-6708/2003/08/007](https://doi.org/10.1088/1126-6708/2003/08/007),
3300 [arXiv:hep-ph/0305252](https://arxiv.org/abs/hep-ph/0305252).
- 3301 [54] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton
3302 Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070,
3303 [doi:10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070), [arXiv:0709.2092](https://arxiv.org/abs/0709.2092).
- 3304 [55] S. A. et. al., “Geant4—a simulation toolkit”, *Nuclear Instruments and Methods in Physics*
3305 *Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506**
3306 (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\)](https://doi.org/10.1016/S0168-9002(03)01368-8).
- 3307 [56] CMS Collaboration, “Event generator tunes obtained from underlying event and
3308 multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155,
3309 [doi:10.1140/epjc/s10052-016-3988-x](https://doi.org/10.1140/epjc/s10052-016-3988-x), [arXiv:1512.00815](https://arxiv.org/abs/1512.00815).
- 3310 [57] CMS Collaboration Collaboration, “Search for massive resonances in dijet systems
3311 containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s}= 8$ TeV”, *JHEP* **08**
3312 (May, 2014) 173. 37 p. Comments: Replaced with published version. Added journal reference
3313 and DOI.
- 3314 [58] F. Dias et al., “Combination of Run-1 Exotic Searches in Diboson Final States at the LHC”,
3315 *JHEP* **04** (2016) 155, [doi:10.1007/JHEP04\(2016\)155](https://doi.org/10.1007/JHEP04(2016)155), [arXiv:1512.03371](https://arxiv.org/abs/1512.03371).
- 3316 [59] CERN, “Collider Reach”, 2019 (accessed January 12, 2019),
3317 <http://collider-reach.web.cern.ch/>.

- 3318 [60] J. Alwall et al., “The automated computation of tree-level and next-to-leading order
 3319 differential cross sections, and their matching to parton shower simulations”, *JHEP* **07**
 3320 (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- 3321 [61] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “A brief introduction to PYTHIA 8.1”, *Comput.
 3322 Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.
- 3323 [62] G. Punzi, “Sensitivity of searches for new signals and its optimization”, *eConf* **C030908**
 3324 (2003) MODT002, arXiv:physics/0308063. [,79(2003)].
- 3325 [63] CMS Collaboration, “Search for heavy resonances in the W/Z-tagged dijet mass spectrum
 3326 in pp collisions at 7 TeV”, *Phys. Lett.* **B723** (2013) 280–301,
 3327 doi:10.1016/j.physletb.2013.05.040, arXiv:1212.1910.
- 3328 [64] CMS Collaboration Collaboration, “Search for Narrow Resonances using the Dijet Mass
 3329 Spectrum with 19.6fb-1 of pp Collisions at sqrt(s)=8 TeV”, Technical Report
 3330 CMS-PAS-EXO-12-059, CERN, Geneva, 2013.
- 3331 [65] D. J. Hand, “Statistical Concepts: A Second Course, Fourth Edition by Richard G. Lomax,
 3332 Debbie L. Hahs-Vaughn”, *International Statistical Review* **80** (2012), no. 3, 491–491.
- 3333 [66] T. C. collaboration, “Identification of b-quark jets with the CMS experiment”, *Journal of
 3334 Instrumentation* **8** (2013), no. 04, P04013.
- 3335 [67] T. C. collaboration, “Identification of heavy-flavour jets with the CMS detector in pp
 3336 collisions at 13 TeV”, *Journal of Instrumentation* **13** (2018), no. 05, P05011.
- 3337 [68] CMS Collaboration Collaboration, “Search for massive resonances decaying into WW, WZ,
 3338 ZZ, qW and qZ in the dijet final state at $\sqrt{s} = 13$ TeV using 2016 data”, Technical Report
 3339 CMS-PAS-B2G-16-021, CERN, Geneva, 2016.
- 3340 [69] M. Dasgupta, A. Powling, and A. Siódmiak, “On jet substructure methods for signal jets”,
 3341 *JHEP* **08** (2015) 079, doi:10.1007/JHEP08(2015)079, arXiv:1503.01088.
- 3342 [70] U. Baur, I. Hinchliffe, and D. Zeppenfeld, “Excited quark production at hadron colliders”,
 3343 *Int. J. Mod. Phys. A* **2** (1987) 1285, doi:10.1142/S0217751X87000661.
- 3344 [71] U. Baur, M. Spira, and P. M. Zerwas, “Excited-quark and -lepton production at hadron
 3345 colliders”, *Phys. Rev. D* **42** (1990) 815, doi:10.1103/PhysRevD.42.815.

- 3346 [72] CMS Collaboration Collaboration, “Search for massive resonances decaying into WW , WZ ,
3347 ZZ , qW , and qZ with dijet final states at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **97** (Apr, 2018)
3348 072006, doi:10.1103/PhysRevD.97.072006.
- 3349 [73] CMS Collaboration Collaboration, “Jet algorithms performance in 13 TeV data”, Technical
3350 Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.
- 3351 [74] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers
3352 and matrix elements in hadronic collisions”, *Eur. Phys. J.* **C53** (2008) 473–500,
3353 doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- 3354 [75] J. Dolen et al., “Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet
3355 substructure”, arXiv:1603.00027.
- 3356 [76] J. Gallicchio and M. D. Schwartz, “Quark and Gluon Tagging at the LHC”, *Phys. Rev.*
3357 *Lett.* **107** (2011) 172001, doi:10.1103/PhysRevLett.107.172001, arXiv:1106.3076.
- 3358 [77] J. A. Aguilar-Saavedra, “Stealth bosons and where to find them”, 2018,
3359 <https://indico.cern.ch/event/649482/contributions/2993323/attachments/1688557/2716149/SBAWTFD.pdf>.
- 3361 [78] J. A. Aguilar-Saavedra, “Running bumps from stealth bosons”, *Eur. Phys. J.* **C78** (2018),
3362 no. 3, 206, doi:10.1140/epjc/s10052-018-5717-0, arXiv:1801.08129.
- 3363 [79] ATLAS Collaboration Collaboration, “Search for diboson resonances in hadronic final states
3364 in 79.8 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, Technical Report
3365 ATLAS-CONF-2018-016, CERN, Geneva, Jun, 2018.
- 3366 [80] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, “QCD or What?”,
3367 arXiv:1808.08979.
- 3368 [81] J. A. Aguilar-Saavedra, J. H. Collins, and R. K. Mishra, “A generic anti-QCD jet tagger”,
3369 *JHEP* **11** (2017) 163, doi:10.1007/JHEP11(2017)163, arXiv:1709.01087.
- 3370 [82] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, “Deep-learned Top Tagging with a
3371 Lorentz Layer”, *SciPost Phys.* **5** (2018), no. 3, 028,
3372 doi:10.21468/SciPostPhys.5.3.028, arXiv:1707.08966.
- 3373 [83] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, *CoRR*
3374 abs/1412.6980 (2014) arXiv:1412.6980.

- 3375 [84] F. Chollet et al., “Keras”. <https://keras.io>, 2015.
- 3376 [85] M. A. et al, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”, 2015.
3377 Software available from tensorflow.org.