

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

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

Zusammenfassung

In dieser Doktorarbeit werde ich drei verschiedene Analysen über neue schwere Resonanzen vorstellen, die in Paare von Vektorbosonen zerfallen und welche wiederum in Quarks zerfallen (hadronic). Die analysierten Daten wurden vom CMS-Experiment am LHC während der ersten drei Jahre der Datenaufnahme bei einer Kollisionsschwerpunktsenergie von 13 TeV gesammelt, was einer integrierten Luminosität von 2.7 (2015), 35.9 (2016) und 77.3 (2016 + 2017) fb^{-1} entspricht. Diese Analysen waren die ersten ihrer Art, die jemals bei einer so hohen Kollisionsenergie durchgeführt wurden. Die Endzustände des Dibosons sind schwierig zu lösen, da die Bosonen hochenergetisch (“boosted”) sind, was dazu führt, dass die beiden Quarks aus dem Zerfall kollimiert werden und zu einem einzigen Jet verschmelzen. Dies führt zu einer Dijet-Endzustands-Topologie, bei der jeder Jet eine Energieunterstruktur aufweist. Die erste Analyse, die ich vorstellen werde, war eine der ersten veröffentlichten CMS-Analysen in “boosted” Endzuständen mit 13 TeV-Daten, und die erste, welche die Jet-Substruktur auf Trigger-Ebene nutzte. Es war eine hoch gehandelte Analyse aufgrund eines zuvor beobachteten $3,4\ (1,3)\ \sigma$ -Überschusses um 2 TeV im 8 TeV-Datensatz, der von ATLAS (CMS) analysiert wurde. Ich habe die Analyse innerhalb von sechs Monaten nachdem die 13 TeV-Datenaufnahme begonnen hatte zu einem veröffentlichten Ergebnis gebracht. Im Anschluss daran habe ich in meiner zweiten Analyse den neuartigen PUPPI-Softdrop-Algorithmus für das Vektor-Boson-Tagging optimiert, validiert und in Betrieb genommen und zusätzlich spezielle Massenkorrekturen für die Softdrop-Jet-Masse entwickelt. Der Algorithmus und die entsprechenden Massenkorrekturen sind jetzt der Standard für das Vektor-Boson-Tagging in CMS und werden von mehreren Analysen verwendet. Es war das erste veröffentlichte Ergebnis, bei dem PUPPI-Softdrop genutzt wurde. Die dritte und letzte Analyse führt ein neuartiges multidimensionales Analyse-Framework ein, mit dem nach Resonanzen irgendwo im 3D-Spektrum der Dijet- und Groomed Jet-Massenspektren gesucht werden kann. Validiert durch die VV-All-Hadronic Analyse, kann dieses Framework für alle Resonanzanalysen über “hadronic” zerfallende Vektorbosonen oder “hadronic” zerfallende Higgs-Boson-Endzustände sowie generische Analysen über geboostete Objekt, die in der Jet-Masse ihren Höhepunkt erreichen, verwendet werden. Schließlich werde ich ein Deep Neural Network für das Vektor-Boson-Tagging vorstellen, welches Jet-Clustering- und substrukturähnliche Variablen in die neuronalen Netzwerkschichten encodiert. Dieser Algorithmus verbessert die Analyseempfindlichkeit erheblich und kann auch als Grundlage für die Entwicklung eines generischen Anti-QCD-Taggers verwendet werden. Letzteres ist von großer Bedeutung, wenn versucht wird, dieses mehrdimensionale Framework für modellunabhängige Analysen zu verwenden.

Contents

⁹⁷	I Theoretical background	15
⁹⁸	1 The Standard Model	17
⁹⁹	1.1 Fundamental particles: quarks and leptons	17
¹⁰⁰	1.2 The Standard Model Lagrangian	19
¹⁰¹	1.2.1 The quantum chromodynamics sector	21
¹⁰²	1.2.2 The electroweak sector	23
¹⁰³	1.2.3 The Higgs sector	26
¹⁰⁴	2 Beyond Standard Model Physics	31
¹⁰⁵	2.1 Shortcomings of the Standard Model	31
¹⁰⁶	2.2 Compositeness	33
¹⁰⁷	2.2.1 Heavy Vector Triplet formalism	33
¹⁰⁸	2.3 Warped extra dimensions	36
¹⁰⁹	II Experimental setup	39
¹¹⁰	3 The Large Hadron Collider	41
¹¹¹	4 The CMS detector	45
¹¹²	4.1 Coordinate system	46
¹¹³	4.2 Tracking detectors	47
¹¹⁴	4.3 Electromagnetic calorimeter	48
¹¹⁵	4.4 Hadronic calorimeter	51
¹¹⁶	4.5 Muon chambers	54

¹¹⁷	5 The CMS trigger system	57
¹¹⁸	6 Event reconstruction	59
¹¹⁹	6.1 Track and primary vertex reconstruction	59
¹²⁰	6.2 The Particle Flow Algorithm	61
¹²¹	6.2.1 Reconstruction of the Particle Flow inputs	62
¹²²	6.2.2 Particle Flow identification	64
¹²³	6.3 Pile-up removal	68
¹²⁴	6.3.1 Charged Hadron Subtraction	68
¹²⁵	6.3.2 Pile up per particle identification (PUPPI)	68
¹²⁶	6.4 Jet reconstruction	70
¹²⁷	6.4.1 Jet clustering	70
¹²⁸	6.4.2 PF jets in CMS	72
¹²⁹	6.4.3 Jet energy corrections	73
¹³⁰	6.5 Jet substructure reconstruction	75
¹³¹	6.5.1 Grooming	76
¹³²	6.5.2 N-subjettiness	80
¹³³	6.5.3 Vector boson tagging	82
¹³⁴	6.6 Monte Carlo Event Generators	85
¹³⁵	III Diboson resonance searches in CMS	87
¹³⁶	7 Search I: First search for diboson resonances at 13 TeV	89
¹³⁷	7.1 A small bump	91
¹³⁸	7.2 Analysis strategy	92
¹³⁹	7.3 Data and simulated samples	94
¹⁴⁰	7.4 Event selection	95
¹⁴¹	7.4.1 Triggering	95
¹⁴²	7.4.2 Preselection	97
¹⁴³	7.4.3 Vector boson tagging	99
¹⁴⁴	7.4.4 Analysis categorization	106
¹⁴⁵	7.5 Background modeling	111
¹⁴⁶	7.6 Signal modeling	114
¹⁴⁷	7.7 V-tagging scale factors	115
¹⁴⁸	7.7.1 Event selection	115

149	7.7.2	Fitting procedure	119
150	7.7.3	Systematic uncertainties	122
151	7.7.4	Fit results	124
152	7.7.5	Impact on search variables	125
153	7.8	Systematic uncertainties	126
154	7.9	Results	127
155	7.10	Limits	127
156	7.10.1	All-hadronic analysis only	127
157	7.10.2	Semi-leptonic and all-hadronic combination	128
158	8	Search II: A new pileup resistant and perturbative safe tagger	135
159	8.1	Towards robust boosted jet tagging	137
160	8.2	Analysis strategy	138
161	8.3	Data and simulated samples	139
162	8.4	Event selection	139
163	8.4.1	Triggering	140
164	8.4.2	Preselection	140
165	8.5	Developing a new W-tagger	144
166	8.5.1	Dedicated PUPPI softdrop jet mass corrections	145
167	8.5.2	W-tagging performance	149
168	8.5.3	Efficiency scale factors, jet mass scale and resolution	151
169	8.5.4	W-tagging mistag rate measurement	154
170	8.6	Mass and purity categorization	157
171	8.7	Background modeling	158
172	8.8	Signal modeling	160
173	8.9	Systematic uncertainties	161
174	8.10	Results	162
175	8.10.1	Early analysis	162
176	8.10.2	Full 2016 dataset	162
177	9	Search III: A novel framework for multi-dimensional searches	173
178	9.1	Small bumps and tri-bosons	175
179	9.2	Analysis strategy	177
180	9.3	Data and simulated samples	178
181	9.4	Event selection	178
182	9.5	Triggering	180

183	9.5.1 Trigger turn-on modeling	181
184	9.6 A mass- and p_T -decorrelated tagger	184
185	9.6.1 Data to simulation scale factors	189
186	9.6.2 p_T -dependence	192
187	9.7 The multidimensional fit	195
188	9.7.1 Modeling of the signal	196
189	9.7.2 Modeling of the non-resonant background	199
190	9.7.3 Modeling of the resonant background	203
191	9.8 Systematic uncertainties	208
192	9.9 Background model validation	211
193	9.9.1 Variations of QCD multijet predictions	211
194	9.9.2 Fit to data control region	211
195	9.9.3 Bias test in pseudodata	212
196	9.10 Results	224
197	9.10.1 Limits	227
198	9.10.2 Pulls of nuisance parameters	231
199	10 Summary	235
200	IV Encoding jet substructure in a deep neural network	239
201	11 A Lorentz invariance based Deep Neural Network for W-tagging	243
202	11.1 LoLa	243
203	11.1.1 Architecture	244
204	11.1.2 Input	244
205	11.1.3 The Combination Layer	247
206	11.1.4 The Lorentz Layer	249
207	11.2 Performance	251
208	11.3 Dependence on jet mass and p_T	252
209	11.3.1 p_T -decorrelation	253
210	11.3.2 Mass sculpting	255
211	11.4 Validation	257
212	11.5 What <i>does</i> LoLa learn?	260
213	12 Summary	263

214	V Final conclusions and outlook	265
215	A V-tagging	271
216	A.1 mMDT and underlying event	272
217	A.2 N-subjettiness on groomed jets	273
218	B Search I: Limits per mass category	277
219	B.1 Limits per mass category	278
220	C Search II	287
221	C.1 W-tagging scale factor: Additional plots	288
222	C.2 Efficiency scale factors for 2.5 fb^{-1}	290
223	C.3 Background fit checks	292
224	C.3.1 Background fit checks in data sideband	293
225	C.3.2 Background fit checks in QCD MC	300
226	C.4 F-test in signal region	304
227	D Search III	309
228	D.1 Signal fits	310
229	D.2 2016 kernels	311
230	D.3 Resonant background	311

Introduction

231

232 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, 233 and it has allowed us to incorporate all of particle physics into one single equation that explains 234 what we can see in the world around us. However, it has some shortcomings. One major problem 235 is that the Standard Model and gravity are incompatible at very high energies and that any 236 incorporation of gravity into a quantum field theory framework results in a non-renormalizable 237 theory. The SM also fails to explain why gravity is so much weaker than the electromagnetic 238 and nuclear forces. This, together with a few other problems, has lead scientists to search for 239 extensions to the Standard Model, commonly referred to as *Beyond Standard Model physics* 240 (BSM). These models are usually accompanied by predicted observables not included in the 241 Standard Model, where the observation of these, or the lack thereof, is a way of falsifying or 242 supporting the model.

244

245 In this thesis, I look for such observables by searching for new massive particles predicted 246 by SM extensions. These particles have the property that they can decay into vector bosons, W^\pm 247 and Z^0 , and usually have a very small interaction probability. The vector bosons are heavy and 248 unstable and will quickly decay into leptons or quarks. In order to counterbalance the small 249 interaction probability associated with generating such a new heavy particle, a final state with 250 two vector bosons decaying hadronically is required since the branching ratio for a vector boson 251 decaying to hadrons is significantly higher than that into leptons. This final state is complicated 252 by the presence of an overwhelming QCD multijet background and the fact that, due to the high 253 mass of the resonance, the vector bosons are highly energetic and their quark decay products get 254 merged into a single jet due to the small angular opening between them. The latter offers an 255 opportunity to distinguish between vector boson jets and jets coming from a quark or a gluon due 256 to the expected differences in jets mass and geometrical substructure within the jet. Algorithms 257 designed to improve the jet mass resolution and resolve jet substructure are commonly referred to

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

260

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

270

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

278

279 This thesis is organized in five parts. Part I provides the theoretical background and moti-
280 vation for the searches presented here. First, the Standard Model is introduced together with a
281 discussion of its known shortcomings, followed by a chapter presenting two possible extensions to
282 the Standard Model, both of which are probed in this thesis. Part II consists of a description of
283 the experimental setup used to collect the data which is analyzed here, as well as the different
284 algorithms used in order to reconstruct each event. Part I and Part II mainly consist of work
285 done by others which has been vital for the completion of this work. The remaining three parts
286 are dedicated to my own personal contributions. In Part III, the three searches described above
287 are presented in chronological order, each with a personal introduction motivating the analysis in
288 question. Following this, in Part IV, the deep neural network based vector boson tagger for future
289 analyses will be presented. Both parts end with their own concluding summary and outlook, in
290 Section 10 and Section 12. In Part V, a final summary of the obtained results and a discussion of
291 the future for the analysis is given.

292

Part I

293

Theoretical background

294

CHAPTER 1

295

The Standard Model

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

310

1.1 Fundamental particles: quarks and leptons

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

³¹⁷ three generations can be arranged in a doublet structure, and are as follows.

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

³¹⁸ The charged leptons can be positively or negatively charged, defined in units of electron charge ³¹⁹ e . By convention the leptons of matter are negatively charged, e^- , μ^- , and τ^- , whereas the ³²⁰ positively charged leptons, e^+ , μ^+ , and τ^+ are considered their anti-particles. A summary of the lepton properties is listed in Table 1.1. Leptons interact with one another through the

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

Table 1.1: Lepton Properties [1, 2].

³²¹

³²² *electromagnetic and weak forces*, which will be explained in more detail in Section 1.2.2. ³²³ The other six fundamental particles of matter are the *quarks*. They are distinguished from the ³²⁴ leptons in that they, in addition to interacting through the weak and electromagnetic forces, ³²⁵ interact with one another through the *strong force*, described in Section 1.2.1. This force binds ³²⁶ the quarks together to form baryons (like protons and neutrons) or mesons (like pions), and in ³²⁷ addition, keeps the quarks from being observed as free particles such that they are only visible ³²⁸ through their baryon or meson bound states (commonly referred to as *hadrons*). Also organized ³²⁹ in three generations, the six quarks are called *up*, *down*, *charm*, *strange*, *top* and *bottom*, and are ³³⁰ organized in flavor doublets as follow

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

³³¹ 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 ³³² charge. As with the leptons, there are also distinct particles of opposite charge, anti-quarks. ³³³ Some of the quark properties are listed in Table 1.2.

³³⁴ These 12 fermions, together with their corresponding anti-particles, represent the fundamental ³³⁵ particles of the universe and constitute all matter around us. There are four fundamental forces ³³⁶ that we know of: gravity, electromagnetism, the weak force and the strong force. Gravity is ³³⁷ extremely weak compared to the other forces and we currently lack a quantum field theory of

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

Table 1.2: Quark Properties

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

349 1.2 The Standard Model Lagrangian

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

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

354 where $\Psi(x^\mu)$ represents any spin-1/2 fermion field, also called *Dirac field*, as a function of space-
 355 time; $\bar{\Psi} = \Psi^\dagger \gamma_0$, where γ^0 is one of the gamma matrices γ^μ and is included in order to make
 356 $\bar{\Psi}\Psi$ invariant under Lorentz transformation; and m is the mass of the fermion in question. Any
 357 interaction between the fundamental particles due to the fundamental forces, can be described as
 358 variations in the Lagrangian of quantum fields and are represented as additional terms in the

³⁵⁹ equation above.

³⁶⁰ Being a gauge theory, the Standard Model has the property of gauge invariance, meaning
³⁶¹ that measurable quantities stay the same despite the fields themselves changing. If observables
³⁶² stay the same after a field transformation, there is a symmetry in the system. The symmetries
³⁶³ of the Standard Model arise due to the fact that fermions of a given type are indistinguishable
³⁶⁴ from one another. These symmetries result in the presence of *force mediators*, which arise
³⁶⁵ as a representation of infinitesimal generators of some symmetry group. The fermion fields
³⁶⁶ can be arranged as particle multiplets where one transforms into the other under a symmetry
³⁶⁷ transformation. A symmetry transformation produce rotations between the particles of a given
³⁶⁸ multiplet, but never to a field outside of that group. The symmetry group of the Standard Model
³⁶⁹ is the direct product

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

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

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

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

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

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

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

$$(1.10)$$

³⁷⁰ The right-handed neutrino (which in needed to generate a Dirac mass for the neutrinos) is not
³⁷¹ included here because it is a SM *singlet* and does not transform under the SM group \mathcal{G}_{SM} , but
³⁷² it could be written as $\nu^c = (1, 1)_{(0)}$. The positron e^c , on the other hand, is included as its
³⁷³ hypercharge is non-zero and it therefore undergoes $U(1)_Y$ interactions. These five multiplets exist
³⁷⁴ for each of the three generations.

³⁷⁵ The numbers representing each multiplet correspond to which representation it belongs to. For

376 instance, we see that the quark multiplet transforms as a triplet under $SU(3)_C$, a doublet under
 377 $SU(2)_L$ and has a non-zero hypercharge, corresponding to a non-trivial representation under
 378 $U(1)_Y$. That corresponds to saying that quarks interact through all of the three fundamental
 379 interactions. From that notation, it is also clear that leptons do not carry color charge, and will
 380 only interact via the electroweak interactions.

381 From the multiplets above, we see that u^c and d^c transform as singlets under $SU(2)_L$, meaning
 382 they do not interact. This, however, does not mean they do not feel the electroweak force. The
 383 electroweak gauge bosons W and Z are not directly part of $SU(2)_L$, rather, they are a linear
 384 combination of $SU(2)_L$ and $U(1)_Y$ and anything with a non-zero weak hypercharge Y will interact
 385 with them. This is also true for the photon, which also is a linear combination of $SU(2)_L$ and
 386 $U(1)_Y$. These interactions; the strong, weak and electromagnetic, will be explained in more detail
 387 in the following sections.

388 1.2.1 The quantum chromodynamics sector

389 The group $SU(3)_C$ describes the strong interaction mediated by gluons, and is described by
 390 the quantum gauge theory Quantum Chromodynamics (QCD). The group is generated by
 391 8 linearly independent matrices $T^a = \frac{\lambda^a}{2}$, where λ^a are the Gell-Mann matrices [3]. The
 392 generator matrices do not commute with one another, but rather satisfy the commutation relation
 393 $[\lambda^i/2, \lambda^j/2] = if^{ijk}\lambda_k/2$. This property makes the $SU(3)_C$ group *non-Abelian*, which consequently
 394 results in the gluons themselves being charged and displaying self-interactions. Gluons are charged
 395 with one unit of color and one unit of anti-color. Quarks, the only fundamental particles interacting
 396 with the strong force, are arranged in the simplest representation of $SU(3)$ and come with one
 397 unit of color or anti-color.

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

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

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

407 covariant derivative

$$D_\mu = \partial_\mu + ig_s A_\mu^a T^a, \quad (1.12)$$

408 introducing a tensor A_μ^a that represents the 8 gluon fields. The QCD Lagrangian then becomes

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

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

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

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

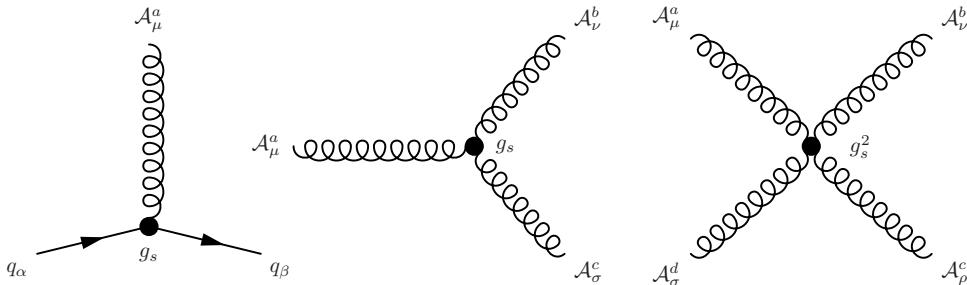


Figure 1.1: The QCD interaction vertices: quark interaction with the gluon field (left), and three-gluon (middle) and four-gluon (right) self-interaction vertices.

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

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

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

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

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

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

436 1.2.2 The electroweak sector

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

443 The symmetry under $SU(2)_L$ is called weak left-handed isospin L , and the symmetry under
 444 $U(1)_Y$ is the weak hypercharge Y . The name “left-handed” arise from the fact that *parity* is
 445 violated in the electroweak interactions. All the fundamental fermions have a *chirality*, defined
 446 as the projection of the particles spin along its direction of motion. Charged weak interactions
 447 are only observed for fermions of left-handed chirality (vector minus axial coupling, V-A). While
 448 the left-handed fermion fields are in the simplest doublet representation of $SU(2)$ with weak
 449 isospin $I = 1/2$, the fermions of right-handed chirality are therefore in the singlet representation
 450 with weak isospin $I = 0$, meaning they do not interact with the gauge bosons of $SU(2)_L$. The
 451 chirality of any fermion Ψ can be defined through the operator γ^5 , the product of the four Dirac
 452 matrices [4] $\gamma^5 = i\gamma^1\gamma^2\gamma^3\gamma^4$. Any fermion field can be projected into its chiral components Ψ_L or

⁴⁵³ Ψ_R through the projection operation

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

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

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

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

⁴⁵⁴ After the substitution, the electroweak Lagrangian can be written as a sum of four terms:

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

⁴⁵⁵ The first term, \mathcal{L}_{gauge} , represent the kinetic field tensor and is

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

where the gauge field tensors are

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

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

⁴⁵⁶ The non-abelian nature of $SU(2)_L$ leads to trilinear and quadrilinear couplings between the photon and vector bosons as illustrated in Figure 1.2.

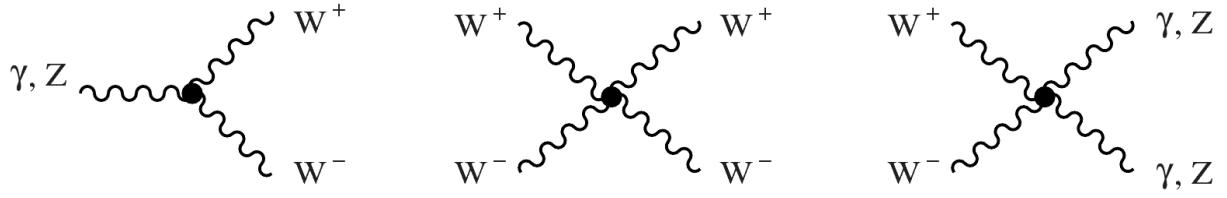


Figure 1.2: The electroweak self-interaction vertices: trilinear (left), and quadri-linear (middle and right) vertices.

458

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

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

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

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

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

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

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

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

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

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

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

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

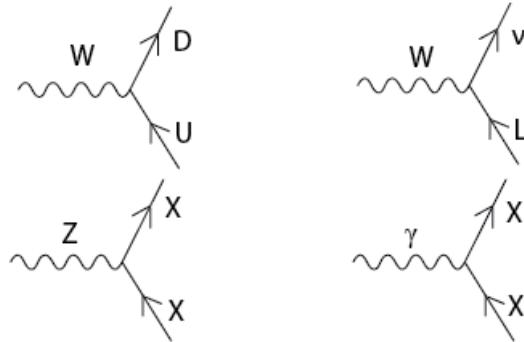


Figure 1.3: The electroweak fermion interaction vertices. Top: charged-current interaction vertex connecting a charged vector boson W^\pm to quarks (left) and leptons (right). Bottom: neutral-current interactions between the neutral Z^0 boson and any fermion (left), and between a γ and electrically charged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

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

481 1.2.3 The Higgs sector

482 The problem of having massless gauge bosons under $SU(2)_L \times U(1)_Y$, while observing massive W
483 and Z bosons, was independently solved by three different groups and has become known as the
484 Englert–Brout–Higgs–Guralnik–Hagen–Kibble mechanism of spontaneous symmetry breaking [8–
485 10], an accomplishment for which Peter Higgs and Francois Englert shared the 2013 Nobel Prize.
486 It began with the realization that the breaking of a local gauge symmetry could give rise to a final
487 mass for the gauge boson involved. This was first discovered in association with superconductivity,
488 where it was found that when a normal metal becomes superconducting the photon field inside
489 the superconductor would acquire a finite mass [11].

⁴⁹⁰ Spontaneous breaking of a global gauge symmetry

⁴⁹¹ In order to achieve spontaneous symmetry breaking, Jeffrey Goldstone [12] suggested introducing
⁴⁹² a massive complex scalar field ϕ with quantum numbers of the vacuum and then giving the field
⁴⁹³ a vacuum expectation value. The field

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

⁴⁹⁴ would have a Lagrangian of the form

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

⁴⁹⁵ where λ_0 is the coupling constant and μ_0 is the mass. The Lagrangian is invariant under $U(1)$,
⁴⁹⁶ though in this case under a global symmetry and not a local one. If one takes μ_0^2 to be negative,
⁴⁹⁷ the potential will get a minima along a circle of radius v such that

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

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

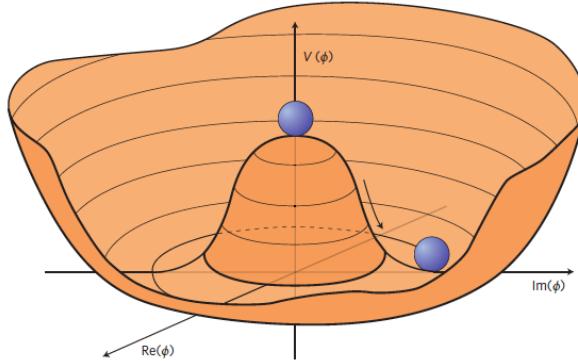


Figure 1.4: The potential $V(\phi)$ for a complex scalar field with $\mu_0^2 < 0$ [13].

⁴⁹⁸

⁴⁹⁹ vacuum expectation value (VEV). The lowest value of the Hamiltonian is now at $\phi = v$ rather
⁵⁰⁰ than at $\phi = 0$. Goldstone then translated the field to a minimum energy position $\phi' = \phi + v$ and
⁵⁰¹ gave the field a vacuum expectation value, effectively breaking the symmetry between the two
⁵⁰² field components, but keeping the Lagrangian invariant. This complex field can be expanded
⁵⁰³ around the ground state in terms of two real fields η and ϵ that represent deviations from the

504 minimum:

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

505 The Lagrangian then becomes

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

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

513 The Higgs mechanism

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

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

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

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

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

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

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

528 **The Weinberg-Salam Model**

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

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

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

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

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

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

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

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

537 with the Weinberg angle θ_W defined as

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

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

⁵⁴¹ each generation is added, for instance for the electron:

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

⁵⁴² where G_E is the electron coupling. We then spontaneously break the symmetry with

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

⁵⁴³ where the neutral Higgs field $h(x)$ is the only remnant of the Higgs doublet after spontaneous
⁵⁴⁴ symmetry breaking. After substitution, the final Lagrangian for the electron mass becomes

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

⁵⁴⁵ We can choose G_e such that

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

⁵⁴⁶ and generate the electron mass as

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

⁵⁴⁷ In summary, all the fermion masses are generated through the interaction of the fermion fields
⁵⁴⁸ with the Higgs field. From the equation above, we see that the fermion masses are not predicted
⁵⁴⁹ from the theory as the coupling G is arbitrary. The Standard Model therefore cannot provide an
⁵⁵⁰ explanation for the difference in hierarchy between the couplings. We also see that the Lagrangian
⁵⁵¹ contains an interaction term coupling the Higgs scalar to the fermions and that this term depends
⁵⁵² on the mass of the fermion. The Higgs boson therefore couples more strongly to heavy fermions
⁵⁵³ than to lighter ones.

CHAPTER 2

Beyond Standard Model Physics

556 2.1 Shortcomings of the Standard Model

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

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

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

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

609 2.2 Compositeness

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

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

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

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

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

629 2.2.1 Heavy Vector Triplet formalism

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

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

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

The simplified Lagrangian governing the dynamics is given as

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

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

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

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

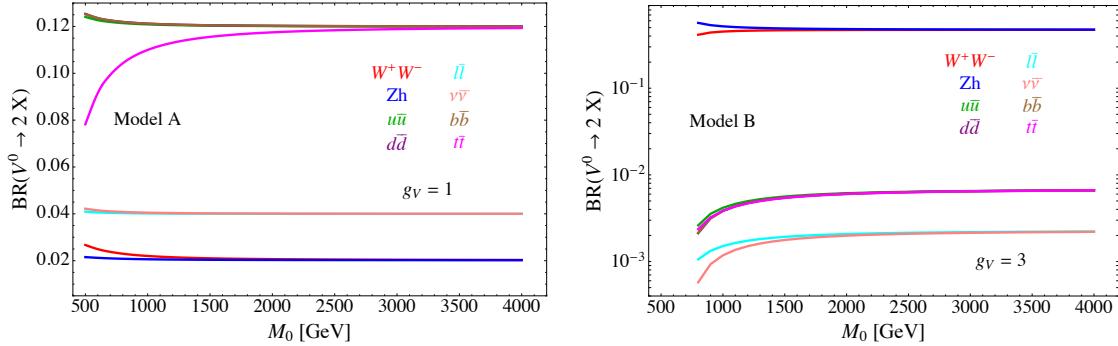


Figure 2.1: Predicted branching fractions of a Z' for two explicit HVT models: Model A $_{g_V=1}$ (left) and model B $_{g_V=3}$ (right) [16].

662 2.3 Warped extra dimensions

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

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

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

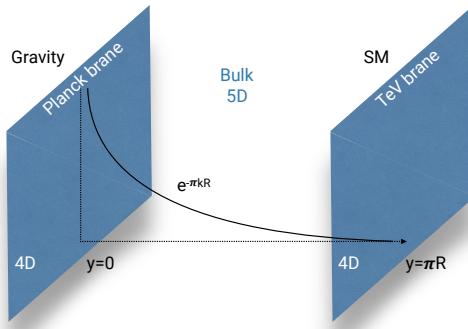


Figure 2.2: The RS model predicts an extra dimension where a 5D space-time stretches between two 4D branes: the Planck brane where gravity is concentrated, and the TeV brane where the SM particles are confined.

672 warping, the Planck mass on the Planck brane gets reduced by a factor of $e^{-\pi k R}$ at the TeV
 673 brane, thereby solving the hierarchy problem. The Planck mass on the TeV brane, which depends
 674 on the geometry of the extra dimension, becomes
 675

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

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

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

The branching ratios of the Bulk Graviton is shown in Figure 2.3.

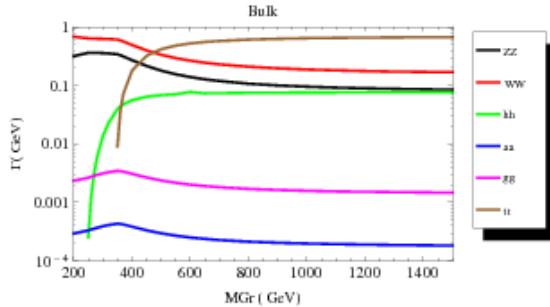


Figure 2.3: Predicted branching fractions for a Bulk Graviton as a function of its mass.

Part II

694

Experimental setup

695

696

CHAPTER 3

697

The Large Hadron Collider

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

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

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

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

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

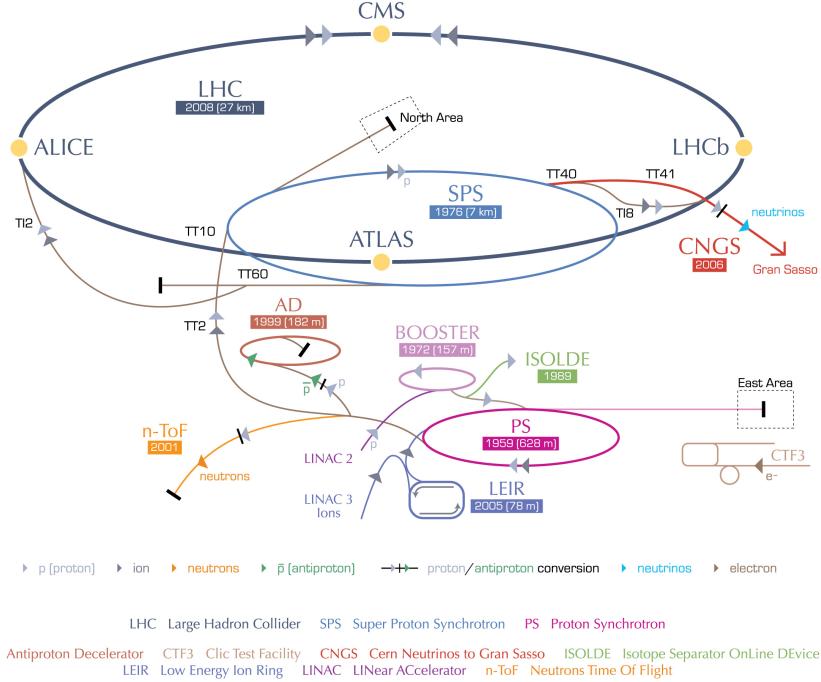


Figure 3.1: The Large Hadron Collider accelerator complex. The four collision points along the ring correspond to the location of the LHC particle detectors CMS, LHCb, ATLAS and ALICE [20].

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

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

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

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

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

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

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

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

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

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

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

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

784

CHAPTER 4

785

The CMS detector

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

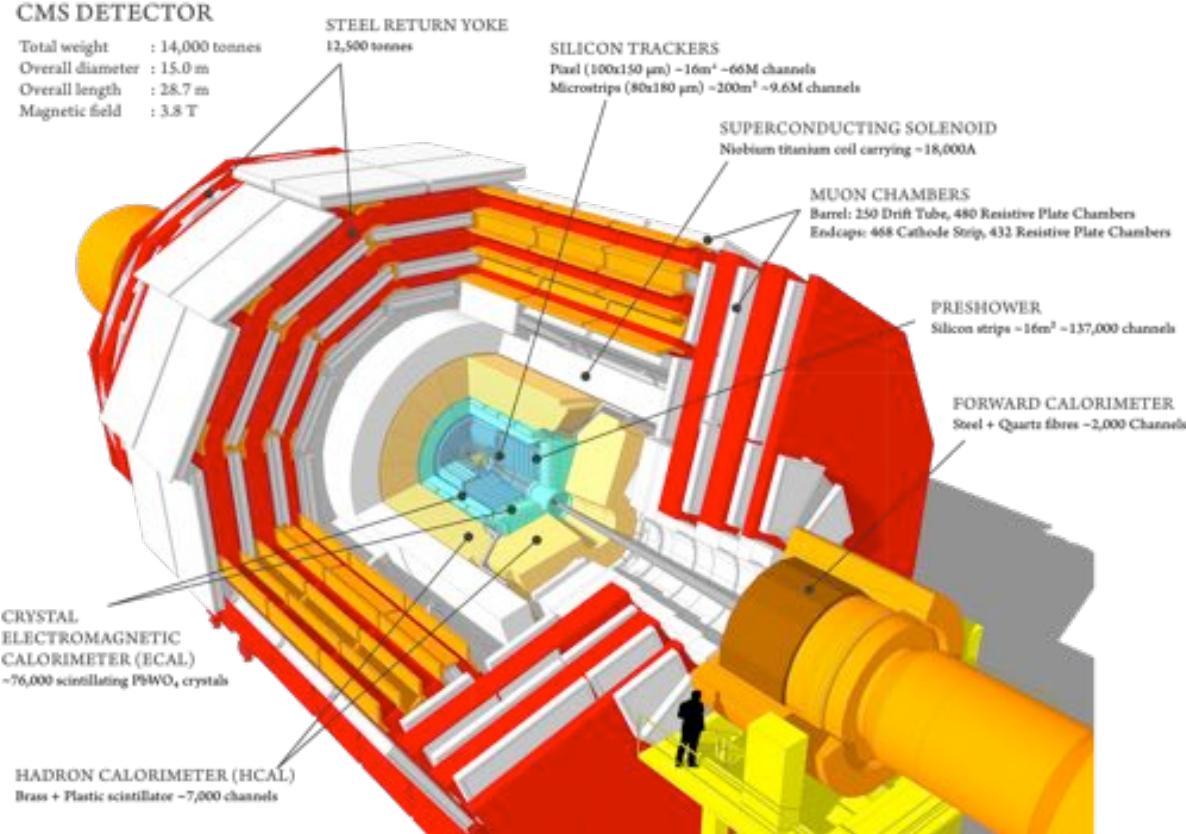


Figure 4.1: The CMS detector and its subsystems: the silicon tracker, electromagnetic and hadron calorimeters, superconducting solenoid and the muon chambers inter-layered with the steel return yoke [21].

4.1 Coordinate system

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

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

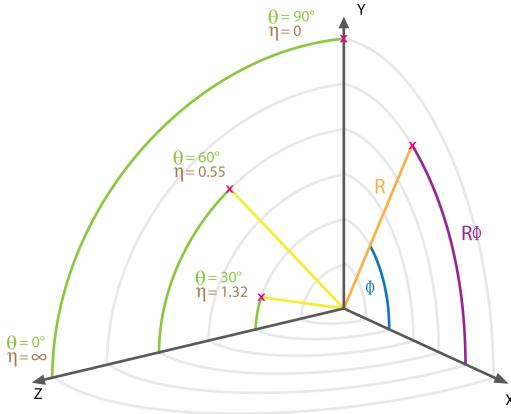


Figure 4.2: The CMS coordinate system [22], shown with some values of θ and η .

822

823 4.2 Tracking detectors

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

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

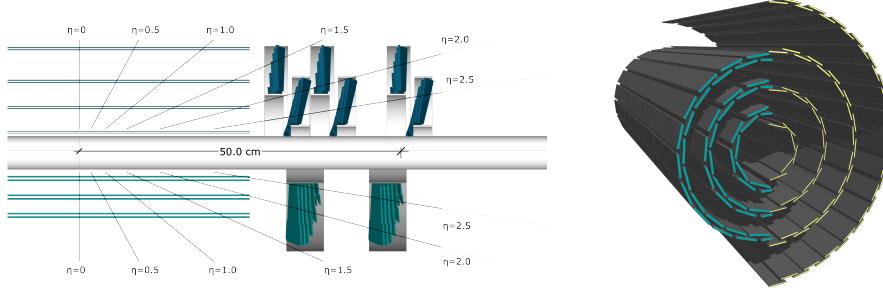


Figure 4.3: Left: The pixel detector layout before (bottom) and after (top) the Phase-1 upgrade. Right: The barrel pixel detector before (left) and after (right) the Phase-1 upgrade [23].

838
839 reduces as you go further away from the beam pipe, the pixel sensors are replaced by silicon
840 strip sensors of larger size, making up the second of the two tracker sub-systems, the silicon strip
841 tracker. There are ten strip layers in total, stretching out to a radius of roughly 130 cm. These
842 are divided into four sections: The inner barrel (TIB) with four strip layers, the two inner endcaps
843 (TID) consisting of three disks each, the outer barrel (TOB) consisting of 6 cylindrical layers,
844 and the two endcaps (TEC) with 9 strip layers each. A schematic overview of the strip tracker
845 layout is shown in Figure 4.4. The strips in the TIB and TID are 10 cm long, with a width of 80
846 μm and a thickness of 320 μm . The TOB and TEC sections consist of slightly larger strips of 25
847 cm \times 180 μm and a thickness of 500 μm . The strip tracker has a total of 10 million detector
848 strips and covers an area of $\sim 200 \text{ m}^2$. To prolong the silicon detector lifetime, the entire tracker
849 (pixel and strip) is kept at a temperature of -20°C with a dedicated cooling system. The tracker
850 has a coverage of up to $|\eta| < 2.6$ and a resolution of roughly $\sigma/p_T \approx 1.5 \times 10^{-5} p_T + 0.005$.

851 4.3 Electromagnetic calorimeter

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

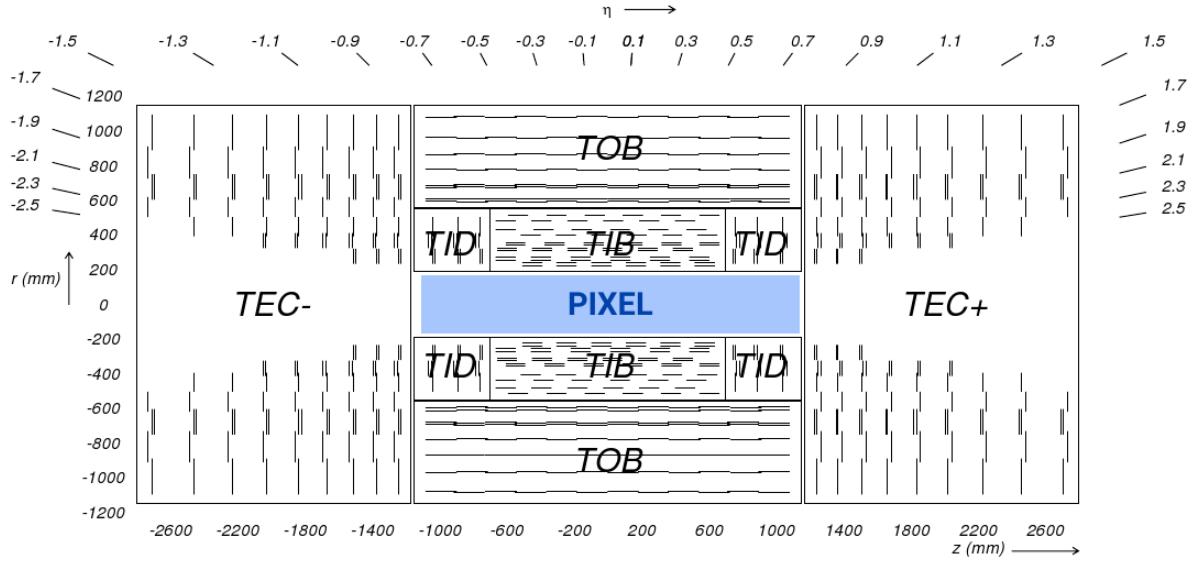


Figure 4.4: Schematic of the CMS silicon strip tracker and its four subsections: the inner barrel (TIB), inner endcaps (TID), the outer barrel (TOB), and the two endcaps (TEC) [24].

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

(ES) using lead absorbers and silicon sensors covers the forward region between $1.65 < |\eta| < 2.6$. The crystals in the barrel are organized into supermodules, each consisting of about 1700 crystals, while the endcap is divided into two half disks consisting of 3662 crystals each (so-called “Dees”). Each 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 4.5.

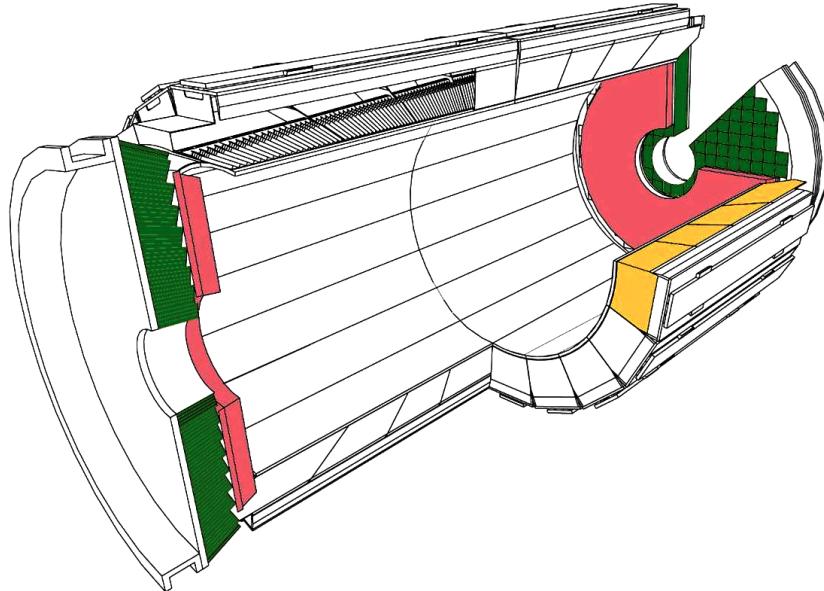


Figure 4.5: A schematic of the CMS electromagnetic calorimeter showing the barrel supermodules (yellow), the individual barrel crystals (black, top left), the endcap modules (green), and the pre-shower detectors (pink) [24].

Having no longitudinal segmentation, the ECAL relies on an accurate reconstruction of the event primary vertex, provided by the tracker, in order to reconstruct the photon angle correctly. The obtained energy resolution of the ECAL can be parametrized in three parts: a stochastic, a noise, and a constant term [25]. It is given as

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

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

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

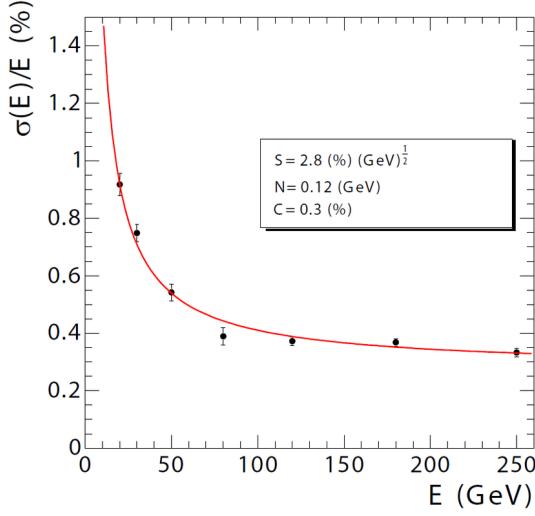


Figure 4.6: The ECAL energy resolution as a function of electron energy as measured in an electron test beam. [25]

894

895 4.4 Hadronic calorimeter

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



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

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



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

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

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

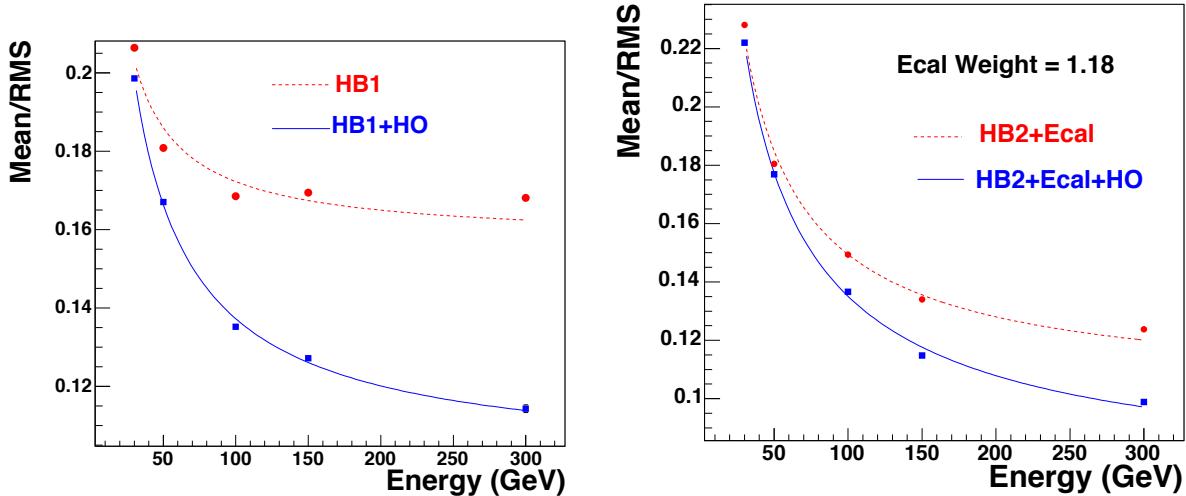


Figure 4.9: The calorimeter energy resolution as a function of pion energy using the HB only or HB+HO (left), and when adding ECAL and HCAL measurements (right) [28].

896 4.5 Muon chambers

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

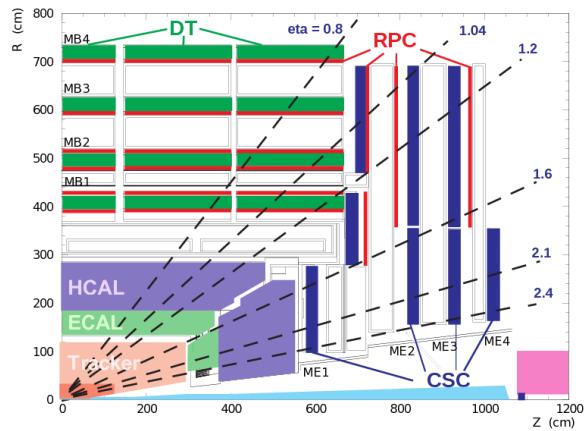


Figure 4.10: A schematic overview of the muon chambers: the DT chambers in the barrel, the CSCs in the endcaps, and the redundant RPC system stretching out to $|\eta| < 1.6$, which are used for triggering purposes [29].

910

CHAPTER 5

911

The CMS trigger system

912 With bunches in CMS colliding at a rate of 40 MHz, there are only 25 nanoseconds between
913 collisions available to process event data. For typical instantaneous luminosities, one billion
914 collisions take place every second, and with an event size of roughly 1 MB, it is impossible for
915 all of these events to be read out and stored to disk. The CMS triggering system is therefore
916 designed to make ultra fast high-quality decisions about which events are interesting and which
917 events are not. The first stage of triggering, called Level 1 (L1), is designed to reduce the event
918 rate to a maximum of 100 kHz through custom-designed hardware. It uses coarse data from
919 the muons system and calorimeters in order to make a decision on whether the event should be
920 recorded or not, a decision that needs to happen within $3.2\mu s$. During this time, the full event
921 data from each sub-detector is stored in detector front-end electronics, awaiting the L1 decision.
922 The information used by L1 is gathered in three steps. First, trigger primitives are created. For
923 the muon system, these consist of track segments from each of the three types of muon detectors.
924 For the calorimeter, trigger primitives are generated by calculating the transverse energy in a
925 trigger tower ($\Delta\eta - \Delta\phi$ of 0.087×0.087) and assigning it to the correct bunch crossing. Trigger
926 primitives from the calorimeter are then passed on to a regional trigger, which defines electrons,
927 muons and jet candidates. Some of this information is passed to the regional muon trigger in
928 order to provide information about whether the particle is minimum ionizing. The global muon
929 trigger then combines track information with calorimeter information and selects a maximum of
930 four muon candidates and calculates their momentum, position, charge and quality. The output
931 from the regional calorimeter trigger is also passed to a global calorimeter trigger which provides
932 information about the jets, total transverse energy and missing energy in the event. Combining
933 the information from the global muon trigger and the global calorimeter trigger, the L1 trigger
934 decides whether to keep the event or not by combining several decisions by simple logic operations
935 (AND/OR/NOT), forming up to 128 algorithms.

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

941

CHAPTER 6

942

Event reconstruction

943

6.1 Track and primary vertex reconstruction

The CMS tracker gets traversed by $\mathcal{O}(1000)$ charged particles at each bunch crossing, produced by several proton-proton interactions happening simultaneously. This makes track reconstructions extremely challenging, and is the reason why a high granularity of the tracker is vital. The average number of vertices per event for the whole Run 2 data taking period is shown in Figure 6.1, with an average of 34 interactions per bunch crossing. Track reconstruction encompasses the process of

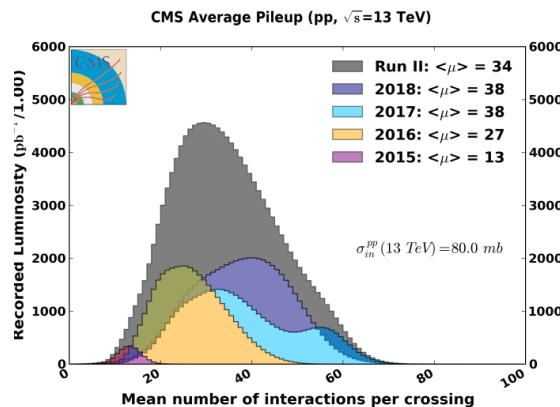


Figure 6.1: The average number of vertices per event in CMS for the whole Run 2 data taking period [30].

948

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

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

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

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

979 6.2 The Particle Flow Algorithm

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

- 985 • muons, through hits in the tracker and in the muon chambers;
- 986 • charged hadrons, through hits in the tracker and energy deposits in the calorimeters;
- 987 • neutral hadrons, through energy deposits in the calorimeters but no hits in the tracker;
- 988 • photons, through energy deposits in the ECAL but not in the HCAL, and no hits in the
 989 tracker; and
- 990 • electrons, through hits in the tracker and energy deposits in the ECAL.

How these different particles propagate through the CMS detector is illustrated in Figure 6.2.

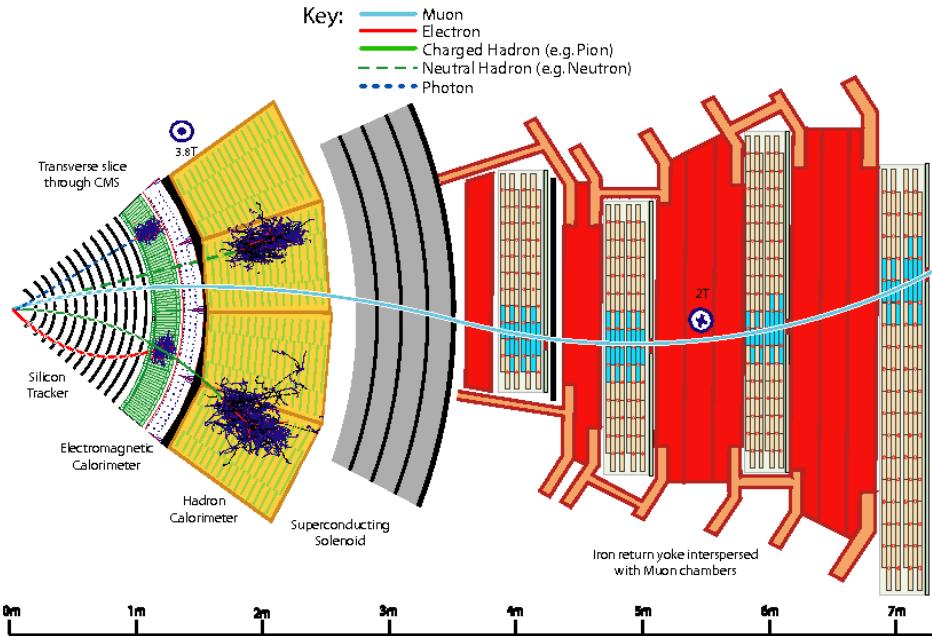


Figure 6.2: Particle interactions in the different subdetectors for a transverse slice through the CMS detector [32].

6.2.1 Reconstruction of the Particle Flow inputs

Electron tracking

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

Muon tracking

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

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

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

1022 Calorimeter clusters

1023 The calorimeter clustering is performed separately for each calorimeter subdetector (ECAL barrel
 1024 and endcaps, HCAL barrel and endcaps and the preshower layers). The first step is to define
 1025 cluster seeds from cells with an energy exceeding some predefined threshold and larger than the
 1026 energy in its neighboring cells. Topological clusters are then formed by adding cells to the seed
 1027 which has at least one corner in common with a cell already in the cluster, and that has an energy
 1028 which is at least twice the noise level of the detector. In Figure 6.3, an example of calorimeter
 1029 clustering for a five-particle jet is shown for the HCAL (left) and ECAL (right). In the HCAL
 1030 (left), two seeds have been identified (gray-filled areas) inside a topological cluster consisting of 9
 1031 cells. These are then defined as two HCAL clusters, with a position as indicated by the red circles.
 1032 The green solid lines correspond to charged tracks reconstructed in the tracker, both pointing to
 1033 the center of the HCAL cluster seeds. The observed deposits left by the same particles are shown
 1034 on the right in Figure 6.3, where the K_L^0 , π^- and the two photons from the decay of a π^0 leave
 distinct clusters in the ECAL. The π^+ leaves no energy deposit in this case.

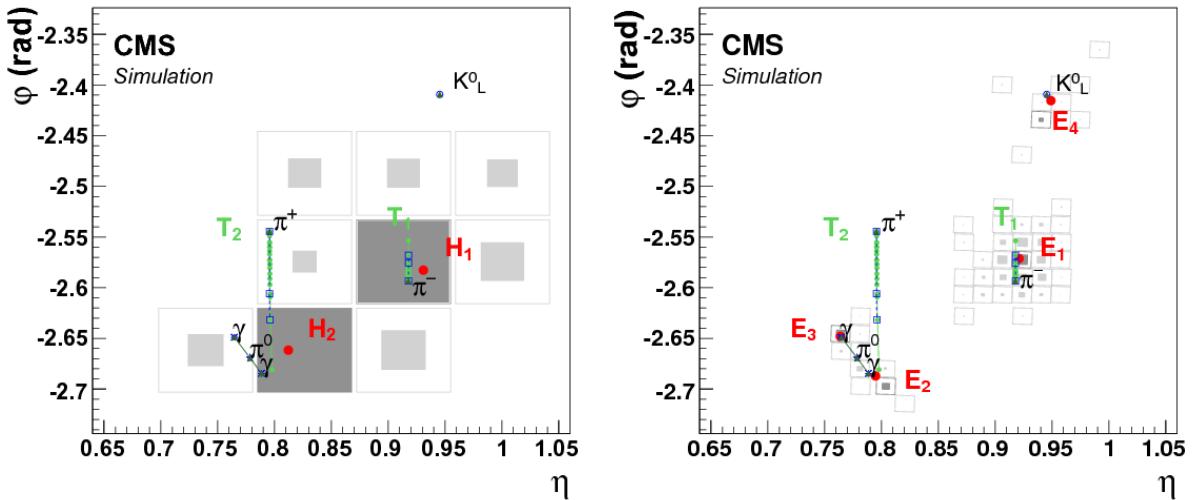


Figure 6.3: The $\eta - \phi$ views of calorimeter clusters generated by a five-particle jet in the HCAL (left) and in the ECAL (right). The squares correspond to calorimeter cells, where the inner area is proportional to the logarithm of the cell energy. Cluster seeds are depicted in dark gray. The dotted blue lines correspond to the simulated particle trajectories, while the green lines correspond to charged tracks reconstructed in the tracker [32].

1036 6.2.2 Particle Flow identification

1037 The link algorithm

1038 The link algorithm is the algorithm responsible for combining the particle flow elements from
 1039 different subdetectors. It can test any pair of elements in the event based on specific requirements
 1040 depending on the nature of the element, but is restricted to the nearest neighbors in the $\eta - \phi$
 1041 plane. The outputs of the link algorithm are so-called *PF blocks* of linked elements, either directly
 1042 linked or linked through having common elements.

1043 • **Inner track - calorimeter cluster link:** The track is interpolated from its last hit,
 1044 through the preshower layers, the ECAL, and ending in the HCAL at a nuclear interaction
 1045 length depth of 1. A link is made if the track is within the cluster areas, where the area
 1046 is enlarged by up to a cell in each direction to account for detector gaps. In case several
 1047 ECAL/HCAL clusters are linked to the same track, only the one with the smallest distance
 1048 in $\eta - \phi$ is kept.

1049 • **Calorimeter cluster - cluster link:** A link between ECAL and HCAL clusters, as well
 1050 as between ECAL and preshower clusters, is made when the cluster position of the more
 1051 granular calorimeter is within the cluster envelope of the less granular calorimeter. If there
 1052 are overlapping links, the one with the smallest distance is kept.

1053 • **Inner tracker - muon chamber link:** This procedure is described in Section 6.2.1.

1054 For each PF block, the reconstruction proceeds in the following order. First, muons are recon-
 1055 structed and their corresponding PF elements removed from the PF block. This is followed by
 1056 the reconstruction and subsequent removal of electrons and energetic photons. Finally, neutral
 1057 and charged hadrons are reconstructed.

1058 Muons

1059 First, isolated global muons are selected by requiring the sum of track p_T and calorimeter energy
 1060 deposits within a cone of $\Delta R = 0.3$ and not belonging to the muon track, to be smaller than
 1061 10 % of the muon p_T . If the muons are non-isolated, they are required to pass the tight muon
 1062 requirement [34] and have at least three matching track segments in the muon detector or have
 1063 matched calorimeter deposits compatible with being a minimum ionizing particle. Muons failing
 1064 both the requirements above are kept if their standalone muon track is of high quality and have a
 1065 lot of hits in the muon detectors, otherwise they are discarded. The muon momentum is defined

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

1068

1069 Muons used in this thesis are required to pass an isolation requirement in order to suppress
1070 the background from QCD multijet events where jet constituents are identified as muons. For this,
1071 a cone of radius $\Delta R = 0.3$ is constructed around the muon direction. The isolation parameter is
1072 defined as the scalar sum of the transverse momenta of all additional reconstructed tracks within
1073 the cone, divided by the muon p_T . Muon candidates with an isolation parameter less than 0.1
1074 are considered isolated and are used for further analysis.

1075 Further, the following selection criteria are applied:

- 1076 • the χ^2 of the global muon track fit must be less than 10;
- 1077 • at least one muon-chamber hit is included in the global-muon track fit and the global muon
1078 track fit must include at least one muon chamber hit;
- 1079 • muon segments in at least two of the muon stations must be matched to the muon tracker
1080 track;
- 1081 • the inner tracker track must be no more than 2 millimeters from the primary vertex in the
1082 xy plane and no more than 5 millimeters in the longitudinal direction;
- 1083 • at least one hit in the pixel detector;
- 1084 • at least six layers of the inner tracker must contain hits; and
- 1085 • at least three matching track segments must be found in the muon detectors

1086 **Electrons**

1087 The electrons are seeded from a GSF track, as described in Section 6.2.1. To differentiate electrons
1088 from charged hadrons, the energy deposit in the HCAL within a distance of 0.15 in the $\eta - \phi$
1089 plane of the supercluster is required to be less than 10 % of that of the supercluster. The electron
1090 candidate must further pass a requirement on the output of a dedicated electron-identification
1091 BDT, using inputs such as track-cluster distance, track χ^2 , and number of hits as input. In this
1092 step, isolated photons are also reconstructed, seeded from ECAL superclusters with $|E_T| > 10$ GeV
1093 and no link to a GSF track. All the tracks and calorimeter deposits used to reconstruct electrons
1094 and isolated photons are further removed from the list of PF blocks.

1095 Only electrons passing certain quality requirements as listed in Table 6.1 are used in this thesis,
1096 with the following variable definitions:

- 1097 • E_T : The supercluster energy ($x \sin(\theta_{track})$), where θ_{track} is the polar angle of the electron
1098 track as measured in the inner tracker layer and extrapolated to the interaction vertex.
- 1099 • η^{sc} : η of the electron supercluster.
- 1100 • **isEcalDriven**: The electron is found through ECAL requirements rather than through
1101 Particle Flow and the tracker.
- 1102 • $\Delta\eta_{in}^{seed}$: difference in η between the track position as measured in the inner layer and extrapolated
1103 to the interaction vertex and then to the calorimeter, and the η of the supercluster.
- 1104 • $\Delta\phi_{in}$: difference in ϕ between the track position as measured in the inner layer and extrapolated
1105 to the interaction vertex and then to the calorimeter, and the ϕ of the supercluster.
- 1106 • **H/E**: Ratio of hadronic energy in the calorimeter towers, within a cone of radius 0.15 centered
1107 at the electron's calorimeter position to the electromagnetic energy of the supercluster.
- 1108 • $\sigma_{in\eta\eta}$: Measure of the energy spread in η in units of crystals of electron energy in a 5×5
1109 block centered on the seed crystal.
- 1110 • **ECAL Isolation**: The transverse electromagnetic energy of all reconstructed hits (with
1111 $E > 0.08$ GeV) in a cone of radius 0.3 centered at the electron's calorimeter position, excluding
1112 those in an inner cone with a radius of 3 crystals and an η strip with a width of 3 crystals.
- 1113 • **Hadronic Depth Isolation**: Defined as the transverse depth of the hadronic energy in
1114 the HCAL inside a cone of 0.3 centered on the electron calorimeter position, excluding
1115 towers in a cone of 0.15 radius.
- 1116 • **Track p_T Isolation**: The sum p_T of the tracks in a ΔR cone of 0.04 to 0.3, excluding an
1117 η region of 0.015.
- 1118 • d_{xy} : Transverse distance between the electron track and the primary vertex.

1130 Hadrons

1131 Finally, after the removal of muons and electrons, the remaining hadrons and non-isolated photons
1132 are identified. HCAL clusters with no track link are defined as neutral hadrons, while ECAL
1133 clusters with no track link are defined as photons (photons are exclusively associated to the ECAL

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

Table 6.1: Summary of the electron requirements applied to all electrons used in this analysis.

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

Missing transverse energy

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

$$\vec{p}_T^{\text{miss}} = - \sum_i^N \vec{p}_{T,i}, \quad (6.1)$$

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

1148 6.3 Pile-up removal

1149 Particles originating from proton-proton interactions not associated with the hardest primary
1150 vertex, are denoted as pileup events. These distort observables of interest from the hard scattering
1151 event and must be mitigated through dedicated pileup removal techniques.

1152 6.3.1 Charged Hadron Subtraction

1153 As mentioned previously, primary vertices are reconstructed using tracks from charged hadrons.
1154 If a primary vertex does not correspond to the hard scattering vertex of the event, the charged
1155 hadrons (as reconstructed through Particle Flow) associated to this vertex (called a pileup vertex)
1156 are removed from the event collection of particles and will not participate in any further object
1157 reconstruction. This method is denoted charged hadron subtraction (CHS).

1158 6.3.2 Pile up per particle identification (PUPPI)

1159 CHS was the default pileup removal algorithm in CMS until very recently. In 2014, a new pileup
1160 removal algorithm with improved performance was proposed; the pileup per particle identification
1161 (PUPPI) [35] algorithm. PUPPI uses a combination of local shape information, event pileup
1162 properties, and tracking information to compute a weight describing the degree to which a given
1163 particle is likely to arise from pileup. First, a variable denoted α is computed based on the
1164 difference between soft radiation coming from pileup and the harder collinear QCD pattern. The
1165 shape of α for charged particles is then used as a proxy for all pileup particles and is used on an
1166 event-by-event basis to calculate a weight for each particle. This weight in turn describes the
1167 degree to which particles are pileup-like and are used to rescale the particle four-momenta. The
1168 shape variable for a given particle i is defined as

$$\alpha_i = \log \sum_{\substack{j \in \text{Ch,PV} \\ j \neq i}} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij}), \quad (6.2)$$

1169 where Θ is the step function and j refers to the neighboring charged particles from the primary
1170 vertex within a cone of radius $R_0 = 0.4$. Charged particles are defined as coming from the
1171 primary vertex if they are associated to the leading vertex of the event or are within a distance of
1172 $d_z < 0.3$ 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

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{RMS_{PU}^2}, \quad (6.3)$$

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 6.4. The top row shows the absolute mass resolution (left), as well as the mass resolution as a function

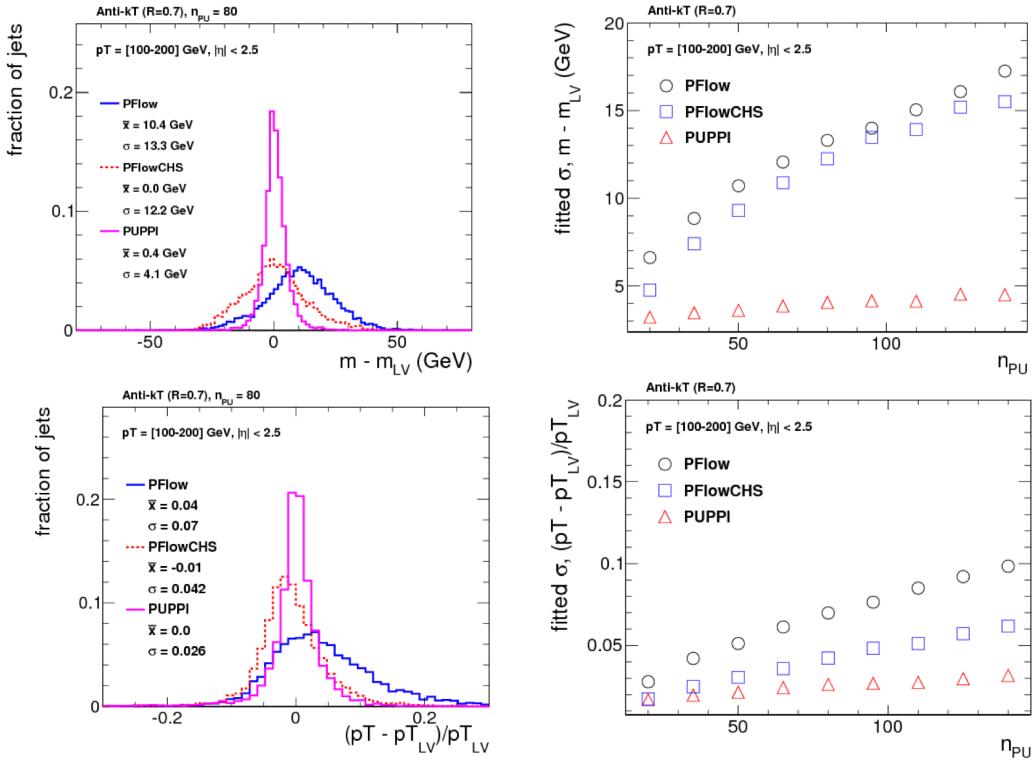


Figure 6.4: The mass (top) and p_T (bottom) resolution comparing PF-only (blue), PF+CHS (red) and PUPPI (pink) jets. The absolute resolution (left) as well as the resolution as a function of the number of reconstructed primary vertices in the event (right) is shown [35].

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

1184 can be achieved using PUPPI as compared to CHS.

1185 6.4 Jet reconstruction

1186 As explained in Section 1.2.1, quarks and gluons are never themselves visible in a detector. Within
1187 10^{-23} seconds, the timescale of strong interactions, they fragment and hadronize into a collimated
1188 spray of hadrons, a so-called jet. In order to infer the properties of the original parton generating
1189 the jet, the properties of the full particle spray needs to be evaluated. Combining these particles
1190 algorithmically is non-trivial, and several algorithms jet-clustering algorithms exist. These provide
1191 a set of rules for grouping particles together into jets and are usually based on certain distance
1192 requirements between particles as well as rules for how to recombine their momenta. Thanks to
1193 Particle Flow, objects like charged hadrons, neutral hadrons and photons, together with their
1194 estimated energy and direction, are already defined and jet clustering in CMS therefore consists
1195 of associating these particles to one common origin.

1196 6.4.1 Jet clustering

1197 The most common jet clustering algorithms used in hadron colliders are the Cambridge/Aachen
1198 algorithm [36], the k_T algorithm [37] and the anti- k_T algorithm [38]. These are all sequential
1199 recombination algorithms, meaning they systematically go through each particle pair in the event
1200 and recombines them into one particle if the combination satisfies certain criteria. The rules,
1201 shared by all three algorithms, are as follows:

- 1201 For each pair of particles i and j , compute the longitudinally invariant distances

$$d_{ij} = \min(p_{ti}^{2p}, p_{tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad , \text{with} \quad \Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2, \text{ and} \quad (6.4)$$

$$d_{iB} = p_{ti}^{2p}, \quad (6.5)$$

1202 where d_{ij} is a measure of the relative transverse momenta between the particles, ΔR_{ij}^2 is
1203 the distance between them in the $\eta - \phi$ plane (which can be roughly translated into a
1204 jet radius), ΔR^2 corresponds to a distance parameter which controls the extension of the
1205 jet and d_{iB} is the distance between the particle and the beam. The parameter p is what
1206 separates the three algorithms from one another and controls the relative power of energy
1207 versus geometrical scales. For the anti- k_T algorithm, it is defined as $p = -1$; for the k_T
1208 algorithm, $p = 1$; and in the case of the C/A algorithm, $p = 0$. The consequences of these
1209 choices are explained in detail below.

- 1210 2. Find the minimum distance of d_{ij} and d_{iB} .
 1211 3. If this is d_{ij} , recombine particles i and j and return to step 1.
 1212 4. If it is d_{iB} , the particle i is defined to be a final state jet, and is removed from the list of
 1213 particles. The algorithm proceeds back to step 1.
 1214 5. Repeat until no particles remain.

1215 **Infrared and collinear safety**

1216 There are two requirements that are extremely important when defining jet algorithms: They
 1217 must be 1) *infrared* (IR) and 2) *collinear* (C) safe. *Infrared* safety corresponds to the requirement
 1218 that if the final state particles are modified by the presence of a soft emission, and there are always
 1219 soft emission in QCD events (both perturbative and non-perturbative), then the set of hard jets
 should remain unchanged. This is illustrated by the two left figures in Figure 6.5. Here, the

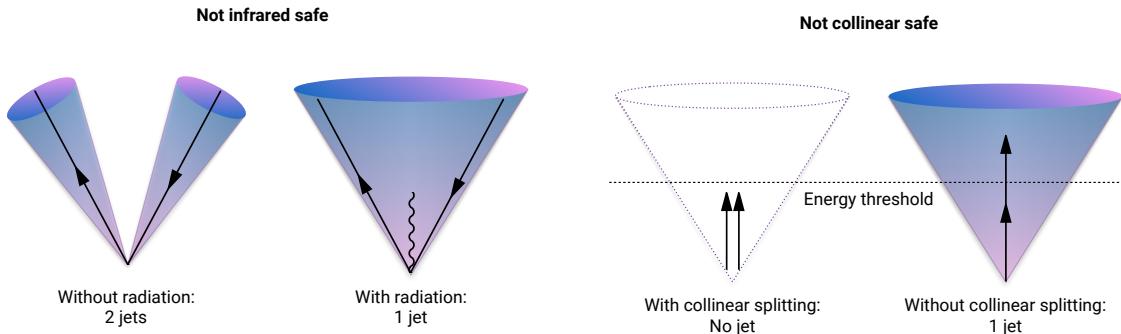


Figure 6.5: An illustration of what would happen for an infrared (left) and collinear (right) unsafe jet algorithm. If an algorithm is infrared unsafe, the presence of a soft emission changes the jet configuration. If an algorithm is collinear unsafe, then if a parton undergoes a collinear splitting this will change the configuration of the jet.

1220
 1221 algorithm is infrared unsafe: the presence of an additional soft gluon changes the jet configuration
 1222 from 2 to 1 jets. If an algorithm is *collinear* unsafe, it means that the jet configuration would
 1223 change if the hard parton undergoes collinear splitting (which a hard parton often does as part of
 1224 the fragmentation process and which are also part of non-perturbative dynamics, like the decay of
 1225 highly energetic hadrons). This is shown in the two left figures of Figure 6.5, where a hard parton
 1226 undergoing collinear splitting fails to be reconstructed due to its daughters being below the energy
 1227 threshold of the algorithm. All sequential recombination algorithms are trivially infrared safe.

1228 **The k_T algorithm**

1229 The k_T algorithm is the oldest of the sequential recombination algorithms and, due to its $p = 1$
1230 definition in Equations 6.4 and 6.5, follows the QCD branching structure in both p_T and in angle
1231 (in reverse). Soft particles are clustered together first, and the final step is the clustering of
1232 the two hardest particles. A consequence of this definition is that there is nothing that keeps
1233 arbitrarily soft particles from being defined as jets, and a minimum cut on the jet p_T should be
1234 introduced. Despite several favorable qualities, the k_T algorithm is not the algorithm of choice in
1235 most hadron collider experiments due to the irregular jets it produces, a consequence of clustering
1236 soft particles first.

1237 **The Cambridge/Aachen algorithm**

1238 The Cambridge/Aachen algorithm, with $p = 0$ in Equations 6.4 and 6.5, follows the QCD
1239 branching structure only in angle as the clustering order is based solely on spatial separation.
1240 The simplest of the algorithms, it recombines all pairs close in ΔR until $\Delta R_{ij} > R$. The benefits
1241 of this is that the clustering history contains information about the presence of any geometrical
1242 substructure within a jet, a feature that will become important in Section 6.5.

1243 **The anti- k_T algorithm**

1244 The default jet clustering algorithm in CMS is the anti- k_T algorithm [38], which follows the
1245 clustering rules in Equations 6.4 and 6.5 with $p = -1$. The algorithm favors the clustering
1246 between high p_T particles, and high and low p_T particles first, disfavoring clustering between soft
1247 particles. That means the algorithm grows around a hard core, yielding jets with a well-defined
1248 cone shaped area. Since the algorithm is infrared- and collinear-safe (IRC safe) and insensitive to
1249 the underlying event (parton remnants not from the hard process) and pileup, , it is chosen as
1250 the main jet algorithm in CMS. A comparison of the resulting jet area in the $\phi - \eta$ plane after
1251 clustering with either k_T , C/A and anti- k_T , is shown in Figure 6.6. The z-axis correspond to the
1252 parton p_T . One can clearly see that when clustering with the anti- k_T algorithm, the produced
1253 jets are circular, with a radius set by R , around the hardest parton.

1254 **6.4.2 PF jets in CMS**

1255 Jet algorithms in CMS mainly use the four-vectors of PF candidates as input and a pileup
1256 removal algorithm is usually applied before clustering occurs. If using CHS (Section 6.3.1),
1257 charged hadrons not associated to the primary vertex are discarded before clustering. If PUPPI

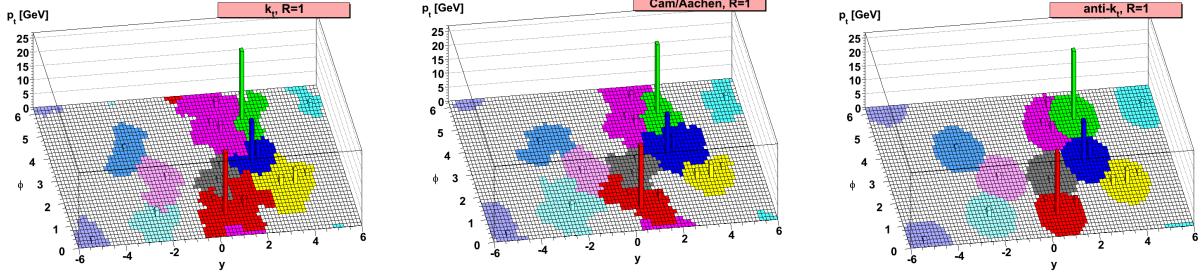


Figure 6.6: A comparison of the resulting jet cone area in the $\phi - \eta - p_T$ plane after clustering the same event with three different jet algorithms: k_T , C/A and $anti-k_T$. [38]

is used (Section 6.3.2), all the PF candidates are reweighted based on how likely they are to have originated from pileup. For the $anti-k_T$ algorithm, CMS by default uses two jet cone sizes: $R=0.4$ and $R=0.8$. Jets with $R=0.4$, called PFAK4, are used for single-prong jets while the larger $R=0.8$ jets, PFAK8, are more often used when looking for jets containing multiple hard quarks/gluons in order to contain all the hadronization products.

These jets are further required to pass certain jet identification requirements provided by the JetMET POG [39], in order to distinguish them from fake jets. All jets used in this analysis are required to pass the *tight ID* requirements which are as follows:

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

6.4.3 Jet energy corrections

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

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

- 1280 • Residual: A correction for data only meant to correct for residual differences between data
 1281 and simulation.

These are illustrated in Figure 6.7.

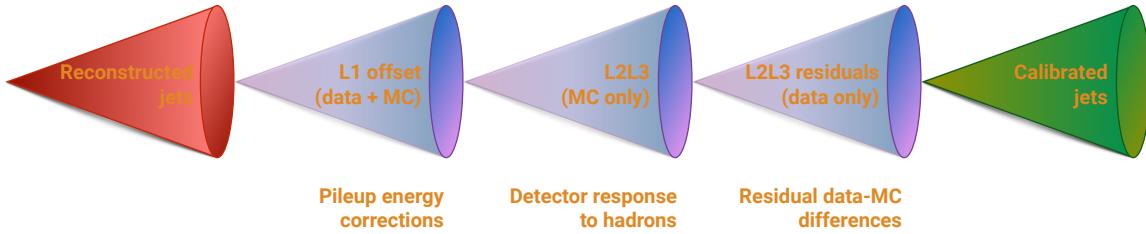


Figure 6.7: The CMS jet energy corrections are derived in three steps: A correction due to offset energy coming from pileup, applied to data and MC, a correction due to the particle-level jet response, also applied to data and MC and finally a correction to account for residual differences between data and MC.

1282

1283 L1 offset correction

1284 The largest correction is the L1 pileup offset correction, which is meant to subtract the additional
 1285 energy in a jet due to pileup. This is done on an event-by-event basis through the *jet area*
 1286 *method* which uses the effective area of the jet multiplied by the average event energy density
 1287 to calculate the size of the offset energy to be subtracted from each jet. An additional p_T - and
 1288 η -dependent term is added in order to account for different pileup densities in different parts
 1289 of the detector and for different jet energies. For data, an additional scalefactor to account for
 1290 data and simulation differences is computed. This is done by constructing a *Random Cone* (*RC*)
 1291 centered at a given η, ϕ and dividing the energy density within that cone in data, evaluated in a
 1292 dataset with no hard interactions (*Zero Bias*), by that of the true energy offset in simulation

1293 L2 relative and L3 absolute corrections

1294 After L1 corrections are applied, corrections to account for the detector response to hadrons are
 1295 derived based on the true detector response in QCD MC. The simulated particle response is
 1296 defined as the ratio

$$R_{\text{particle}} = \frac{p_{T,\text{reco}}}{p_{T,\text{particle}}}. \quad (6.6)$$

1297 These are derived in bins of particle-level p_T and reconstructed η : The L2 relative corrections are
 1298 intended to make the detector response uniform and are derived as a function of η , while the L3

1299 absolute corrections are derived as a function of jet p_T . These corrections are applied both to
1300 data and to MC.

1301 **Residual data corrections**

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

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

1302 is then evaluated in data and in simulation. The ratio of the two, $R_{\text{data}}/R_{\text{MC}}$, defines the residual
1303 corrections.

1304

1305 The above description of jet energy corrections in CMS is meant as a rough, instructive summary
1306 only. A full description of the measurement techniques used in CMS can be found in [40].

1307 **6.5 Jet substructure reconstruction**

1308 In analyses looking for highly energetic (“boosted”) vector bosons, the opening angle between the
1309 vector boson quark decay products becomes so small that the highly boosted boson appears as a
1310 single large jet instead of two well-separated smaller jets. The distance between the two quarks,
1311 in the case of a hadronic decay, depends on the mass of the vector boson and its p_T and goes as

$$\Delta R = \frac{2M_V}{p_{T,V}}. \quad (6.7)$$

Above a W boson p_T of 200 GeV, the two quarks are therefore merged into a single large cone jet of size $R = 0.8$. A sketch of the two different situations is shown in Figure 6.8. If the W p_T is well below 200 GeV, its decay products are two well-defined jets (left). However, once the W boson transverse momenta is approximately 200 GeV, both the quarks are completely contained within a single jet (right), referred to as a W jet. In order to distinguish jets from hadronically decaying

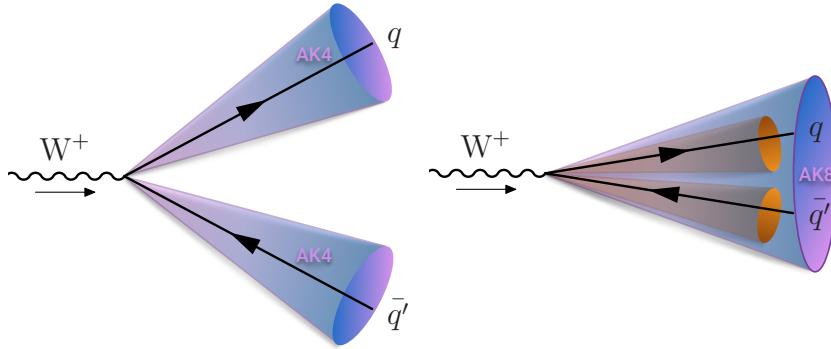


Figure 6.8: If the mass of the resonance is low enough, the quark decay products of each vector boson are well separated and clustered into distinguishable AK4 jets (left). If the transverse momentum of the vector boson is greater than 200 GeV, the vector boson decay products are merged into one single large cone AK8 jet.

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

6.5.1 Grooming

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

1332 **Trimming**

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

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

1333 If the subjet fails this requirement, it is removed. The remaining subjets are then assembled into
 1334 a new “trimmed” jet. The effect of trimming on real W boson jets and QCD quark or gluon jets
 1335 for different values of r_{sub} and $p_{T,frac}$ is shown in Figure 6.9. The best signal mass resolution is
 1336 obtained with $r_{sub} = 0.2$ and $p_{T,frac} = 0.03$, which is also the parameter setting that provides the
 1337 best signal discrimination from background by pushing the QCD jet mass closer to zero. These
 1338 are the default values of the tuned parameters of the trimming algorithm in CMS ($r_{sub} = 0.2$ and
 $p_{T,frac} = 0.03$).

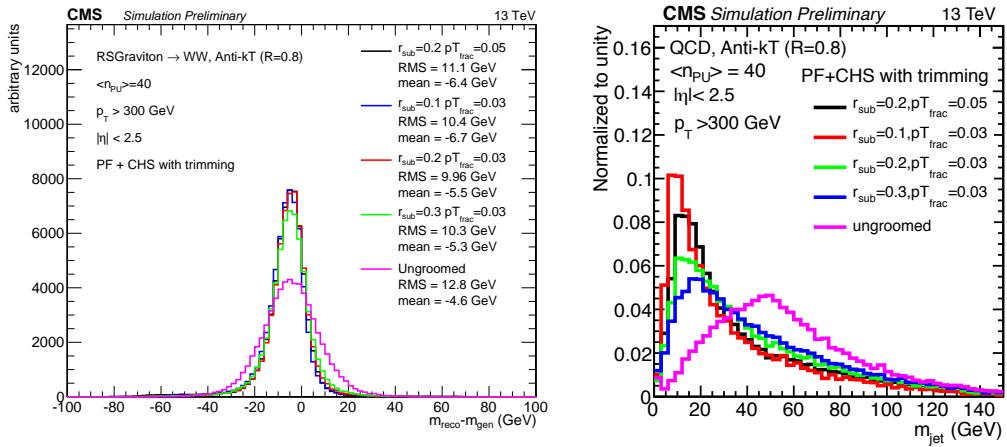


Figure 6.9: The effect of trimming on a signal jet (left) and a background jet (right) for different values of the tuned parameters r_{sub} and $p_{T,frac}$ [42].

1339

1340 **Pruning**

1341 The pruning algorithm, in addition to removing soft particles, has an additional requirement on
 1342 the distance between any recombination that is at wide angle. It proceeds by reclustering the jet

1343 with the C/A algorithm, requiring at each step that

$$\frac{\min(p_{T,i}, p_{T,j})}{p_{T,P}} > z_{cut} \quad \text{and} \quad \Delta R_{i,j} < D_{cut} = \frac{2r_{cut}m_{jet}}{p_T}. \quad (6.8)$$

1344 The first requirement is a requirement on the hardness of the combination. The variables $p_{T,i}$ and
 1345 $p_{T,j}$ correspond to the transverse momenta of each protojet (single particle or group of particles
 1346 already combined in a previous step) and $p_{T,P}$ is the combined p_T of the two. The protojet with
 1347 the lowest transverse momenta is removed if its hardness is below z_{cut} , or if it forms an angle
 1348 wider than D_{cut} relative to the axis of the recombination of the two protojets. In CMS, the tuned
 1349 parameters are set to $r_{cut} = 0.5$ and where $z_{cut} = 0.1$. Figure 6.10 shows the ungroomed as well
 1350 as the pruned jet mass distribution for signal (left) and background (right) jets. The highest
 amount of signal and background separation in CMS, is achieved with $r_{cut} = 0.5$ and $z_{cut} = 0.1$.

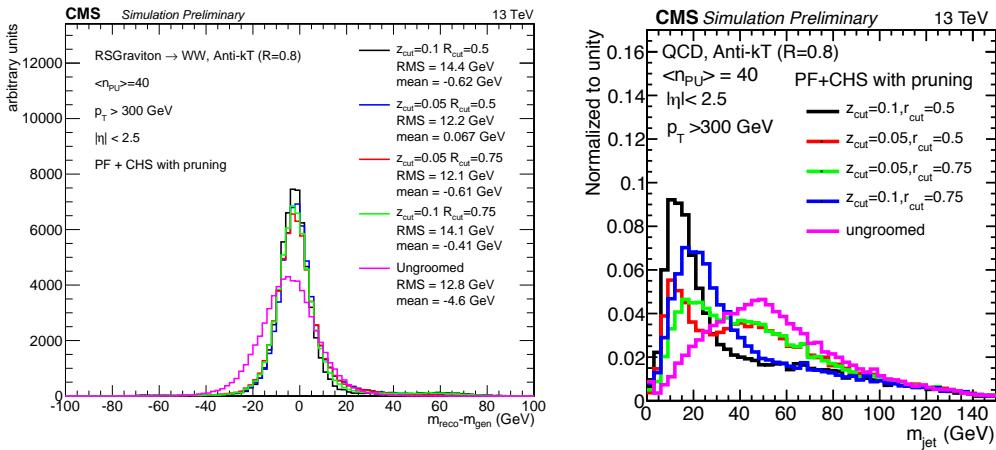


Figure 6.10: The effect of pruning on a signal jet (left) and a background jet (right) for different values of the tuned parameters z_{cut} and r_{cut} [42].

1351

1352 Modified Mass Drop Tagger and Soft Drop

1353 The mass drop tagger (MDT) [43] is based on the idea that a W or Z jet is formed by two quark
 1354 subjets and that, therefore, the mass of each subjet is much smaller than their combined mass
 1355 (and much smaller than the mass of the boson itself). A QCD jet is, on the other hand, formed
 1356 by continuous soft radiation, meaning that its heaviest subjet should be close to the mass of the
 1357 jet itself. The MDT tagger therefore starts from a jet j clustered with the C/A algorithm and
 1358 then declusters it again, defining subjets s_1 and s_2 , where $m_{s_1} > m_{s_2}$. If a significant mass drop
 1359 occurred during declustering, $m_{s_1} < \mu m_j$, where m_j is the mass of the jet j , and the splitting
 1360 is not too asymmetric, $\min(p_{T,s_1}^2, p_{T,s_2}^2) \Delta R(s_1, s_2) / m_j^2 > y_{cut}$, the jet j is selected as the tagged
 1361 jet. Otherwise j is set equal to s_1 and the procedure starts over. The modified mass drop
 1362 tagger (mMDT) [44] is a modification of MDT where the subjet with the largest transverse mass,
 1363 $m^2 + p_T^2$, instead of bare mass, m , is followed. This fixes a flaw in the MDT tagger so that, in
 1364 cases where the mass drop and asymmetry conditions are not satisfied, the more energetic rather
 1365 than the heavier branch is followed. The recommended default version of mMDT additionally
 1366 does not use the mass drop condition, $m_{s_1} < \mu m_j$, as it was shown to have a negligible impact
 1367 on the jet mass distribution for QCD jets [44]. In addition, the mMDT algorithm by default uses
 1368 a z_{cut} criteria, as for pruning, rather than a y_{cut} criteria. The modified mass drop condition is
 1369 generalized through the soft drop declustering method [45], simply called Soft Drop, which allows
 1370 for different types of angular requirements to enter the condition. The Soft Drop condition is the
 1371 following,

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut} \frac{\Delta R_{12}^\beta}{R_0}, \quad (6.9)$$

1372 where the asymmetry condition now is defined directly through the transverse momentum fractions
 1373 of the subjets, rather than through a k_T distance to the jet mass. If the splitting is not too
 1374 asymmetric, the condition is met and the full jet is deemed the softdrop jet. Otherwise only the
 1375 highest- p_T subjet is kept and the declustering continues. If the jet can not be declustered any
 1376 further, it can either be removed from consideration, so-called “tagging”-mode, or deemed the
 1377 final soft-dropped jet, “grooming”-mode. A $\beta = 0$ corresponds to the modified mass drop tagger
 1378 and removes all soft emission wider than the dominant two-prong substructure. For $\beta > 0$, soft
 1379 radiation is removed, but some fraction of soft-collinear radiation is kept. Lastly, with $\beta < 0$,
 1380 Soft Drop can remove soft as well as collinear radiation. The performance of Soft Drop on W jets
 1381 and QCD quark/gluon jets for different values of β is shown in Figure 6.11. The modified mass
 1382 drop tagger (Softdrop with $\beta=0$) with $z_{cut} = 0.1$ is the default Soft Drop settings in CMS, due
 1383 to it providing the best signal/background discrimination while maintaining an excellent signal
 1384 mass resolution.

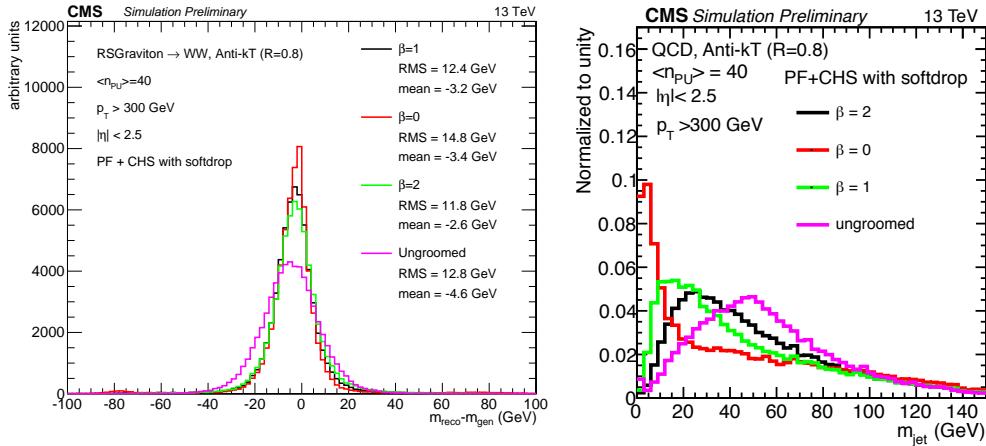


Figure 6.11: The effect of softdrop on a signal jet (left) and a background jet (right) for different values of the tuned parameters β . $\beta = 0$ corresponds to the Modified Mass Drop Tagger, which is the default Softdrop setting in CMS [42].

1385 6.5.2 N-subjettiness

1386 After using the algorithms above, there is still information in the jet structure itself that can
 1387 distinguish W/Z jets from quark/gluon jets. A W or Z jet consists of two well-defined high- p_T
 1388 subjets. A quark/gluon jet instead is made from a single parton, and consists of several large
 angle, asymmetric splittings, as illustrated in Figure 6.12. The N-subjettiness algorithm [46]

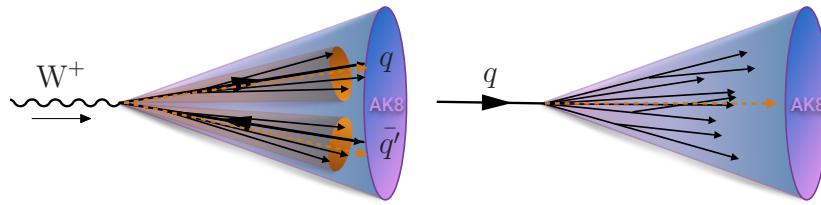


Figure 6.12: A jet stemming from the decay of a W will usually have two well-separated high- p_T subjets, while a jet with a single-prong origin consists of several large angel splittings.

1389 takes advantage of this fact by attempting to count the number of hard sub-elements within a jet.
 1390 This is quantified through the n-subjettiness variable, τ_N , defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min(\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}), \quad (6.10)$$

1392 where k runs over all the jet constituents, $p_{T,k}$ is the constituent transverse momentum, and
 1393 $\Delta R_{i,k}$ is the distance between the constituent and candidate subjet axes. These subjet axes are

1394 obtained through a one-pass optimization procedure which minimizes N-subjettiness [47]. The
1395 normalization factor in front is given as

$$d_0 = \sum_k p_{T,k} R_0, \quad (6.11)$$

1396 where R_0 corresponds to the cone size of the initial jet. With this definition, jets with $\tau_N = 0$
1397 have most of their constituents aligned along the subjet axes. However, if $\tau_N \gg 0$, a large
1398 fraction of the energy is radiated away from the subjet directions and the jet is more likely to
1399 have more than N subjets. In CMS, and as recommended by the authors in [46], the ratio τ_2/τ_1
1400 is used to discriminate W jets from QCD jets. The reason for this is that, while signal jets are
1401 expected to have a large τ_1 , quark/gluon can similarly have large τ_1 due to the diffuse radiation
1402 present. However, QCD jets with a large τ_1 tend to have an equally large τ_2 , while signal jets
1403 do not, hence the ratio of the two provides greater separation power. This can be seen from
Figure 6.13. In CMS, the n-subjettiness algorithm is by default applied to ungroomed jets. The

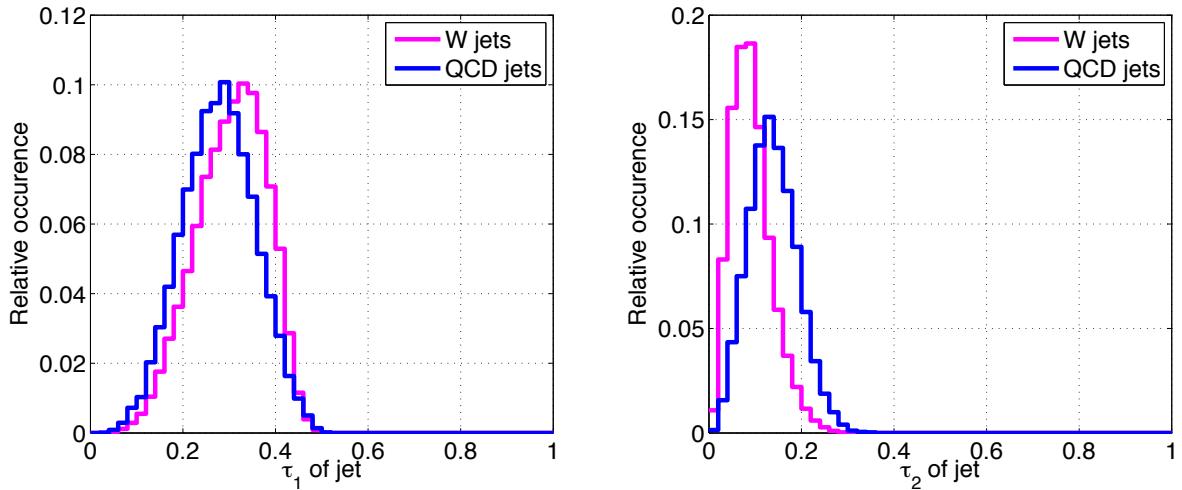


Figure 6.13: The distribution of the n-subjettiness variables τ_1 (left) and τ_2 (right) for signal jets and background jets. On their own, the τ_N variables provide little discriminating power [46].

1404
1405 distribution of τ_{21} for signal and background jets with different pileup subtraction algorithms
1406 applied is shown in Figure 6.14, where τ_{21} in combination with PF+PUPPI (green), yields a
1407 distribution most similar to the generated one (black).

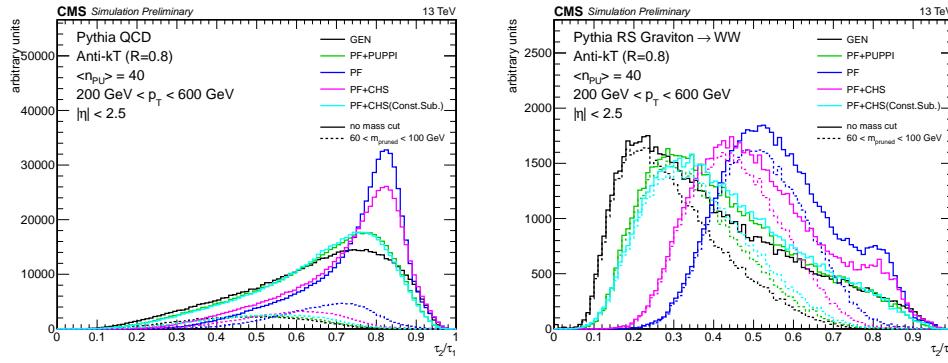


Figure 6.14: The distribution of the n-subjettiness ratio τ_{21} for signal jets (left) and background jets (right) with different combinations of pileup subtraction algorithms applied. The solid lines corresponds to the τ_{21} distribution with no mass cut applied, while the dotted lines are within a mass window of 60–100 GeV [42].

6.5.3 Vector boson tagging

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

Polarization effects

The vector boson polarization has a significant effect on the W-tagging efficiency. The helicity angle θ^* , defined as the angle between the outgoing quark daughters of the W boson in its rest frame relative to its direction of motion [48], is very different for longitudinally polarized vector bosons, W_L , and transversely polarized vector bosons W_T [49]. Figure 6.15 shows the $\cos \theta^*$ distribution for the outgoing quarks from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decays, and it can be observed that transversely polarized W bosons decay with the quarks emitted closer to the vector boson direction of motion. The consequence of this, is that there is a higher asymmetry in the transverse momenta of the two quarks from a W_T decay. This in turn causes grooming algorithms, designed to remove soft constituents of a jet, tend to reject particles coming from the softer quark, resulting in a lower jet mass and a drop in tagging efficiency. Figure 6.16 shows the W-jet tagging efficiency versus q/g jet mistagging rate for a selection on the jet pruned mass of $60 \text{ GeV} < m_{\text{pruned}} < 100 \text{ GeV}$, scanning τ_{21} cuts (here for CA R=0.8 jets).

The tagging efficiency for transversely polarized W bosons (green) is significantly lower than

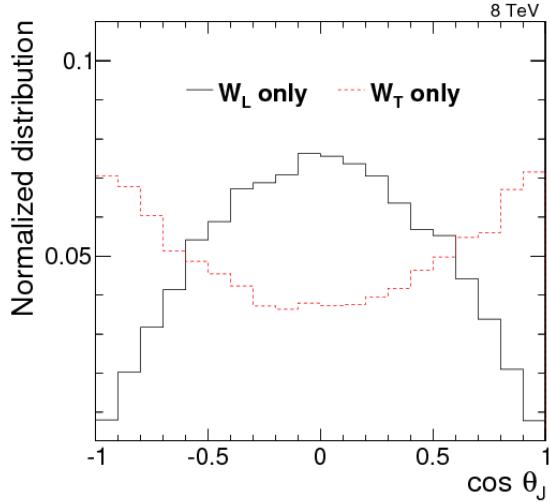


Figure 6.15: The helicity angle for generated quarks from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decays [49].

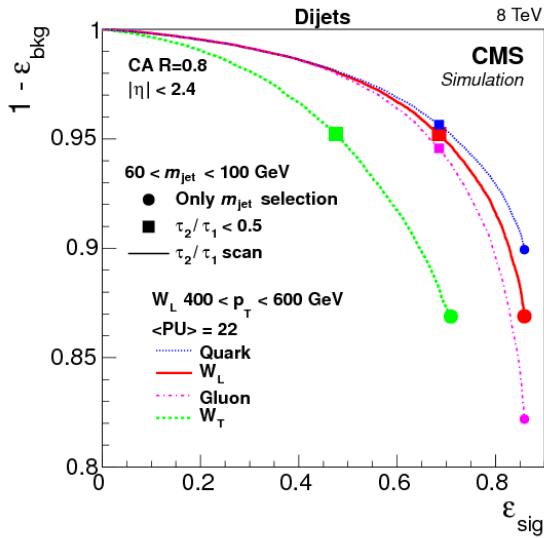


Figure 6.16: The helicity angle for generated quarks from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (red) decays [49].

the tagging efficiency for longitudinally polarized bosons (red). This can be explained by looking at the $\cos \theta^*$ distribution on reconstructed level, using the C/A subjets, with a cut on the jet pruned mass of $60 \text{ GeV} < m_{\text{pruned}} < 100 \text{ GeV}$, as shown in Figure 6.17. When comparing to

¹⁴³¹ the distribution at generator level with no groomed mass window applied, Figure 6.15, one can see that the W_T jets with $\cos \theta \sim 1$ are completely removed. This is due to two effects: the

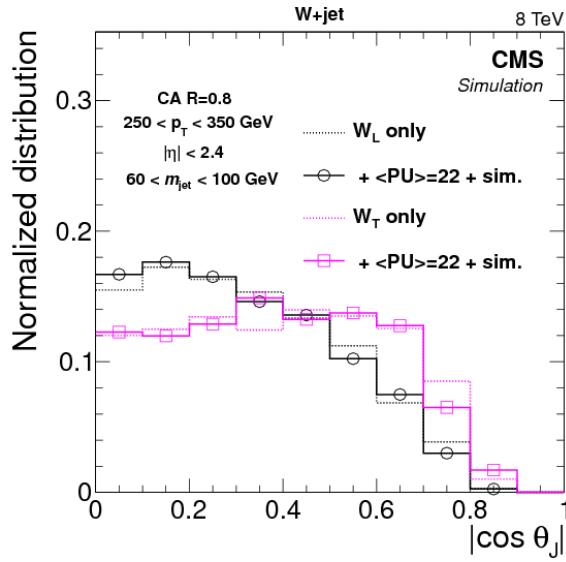


Figure 6.17: The helicity angle for subjets from $W_L \rightarrow q\bar{q}$ (black) and $W_T \rightarrow q\bar{q}$ (pink) decays [49].

¹⁴³²
¹⁴³³ p_T -asymmetry explained above and the fact that the ΔR distribution between the two quarks is
¹⁴³⁴ much smaller in the case of W_L , making them more likely to be fully contained within a jet cone
¹⁴³⁵ of $R=0.8$.

1436 6.6 Monte Carlo Event Generators

1437 Monte Carlo event generators offer a realistic estimate of high-energy collisions on an event-by-
1438 event basis, allowing us to estimate signal and background processes accurately. Simulated events
1439 are usually produced in three steps, beginning with the hard process through hadronization and
1440 decay. First, a matrix element generator simulates the hard scattering process and subsequent
1441 decays. Secondly, the showering and hadronization of unstable particles is performed and, lastly,
1442 the final-state particles are passed through a full detector simulation in order to reproduce a
1443 range of experimental effects.

1444 General-purpose Monte Carlo (GPMC) generators, like HERWIG ++ [50] and PYTHIA 8 [51],
1445 deal with both perturbative as well as hadronization phenomena, simulating an event all the
1446 way up until detector simulation. In HERWIG ++ and PYTHIA 8, the hardest processes are only
1447 simulated at the lowest order of perturbative expansion, meaning $2 \rightarrow 2$ or $2 \rightarrow 3$ scatterings.
1448 In order to have tree-level matrix elements with an arbitrary final-state multiplicity, they can
1449 be combined with programs used to generate parton-level events at higher accuracy, which are
1450 then processed through showering and hadronization with the GPMC generators. One popular
1451 program for generating matrix elements is MADGRAPH [52]. This, however, still correspond
1452 to a tree-level (leading order) approach. To go to next-to-leading-order (NLO), meaning the
1453 inclusion of virtual corrections, two methods exist: MC@NLO [53, 54] and POWHEG [55]. These
1454 combine the full next-to-leading-order prediction for inclusive processes with the subsequent
1455 parton showers, either by a subtraction method regularizing the real contributions, or by a
1456 matrix-element correction of the parton shower branching probability. After hadronization, all
1457 final state particles are passed through a full simulation of the CMS detector. This is done with
1458 GEANT4 [56], which models the interaction and showering of particles with materials, and outputs
1459 position-dependent energy deposits.

1460 For the work presented in this thesis, simulated samples of the Standard Model background
1461 processes are used to optimize the analysis and in some cases provide flexible background templates.
1462 QCD multijet production is simulated with four generator configurations: 1. PYTHIA standalone,
1463 2. the LO mode of MADGRAPH matched with PYTHIA, 3. POWHEG matched with PYTHIA and 4.
1464 HERWIG++ 2.7.1 with tune CUETHS1 [57]. Top-quark pair production is modeled with POWHEG
1465 and showered with PYTHIA unless otherwise stated. W+jets and Z+jets production are simulated
1466 with the leading-order (LO) mode of MADGRAPH matched with PYTHIA. Signal samples are
1467 generated with standalone PYTHIA. All samples are processed through a GEANT4-based simulation
1468 of the CMS detector. To simulate the effect of additional proton-proton collisions within the
1469 same or adjacent bunch crossings (pileup), additional inelastic events are generated using PYTHIA

¹⁴⁷⁰ and superimposed on the hard-scattering events. The simulated MC events are finally weighted to reproduce the distribution of the number of pileup interactions observed in data.

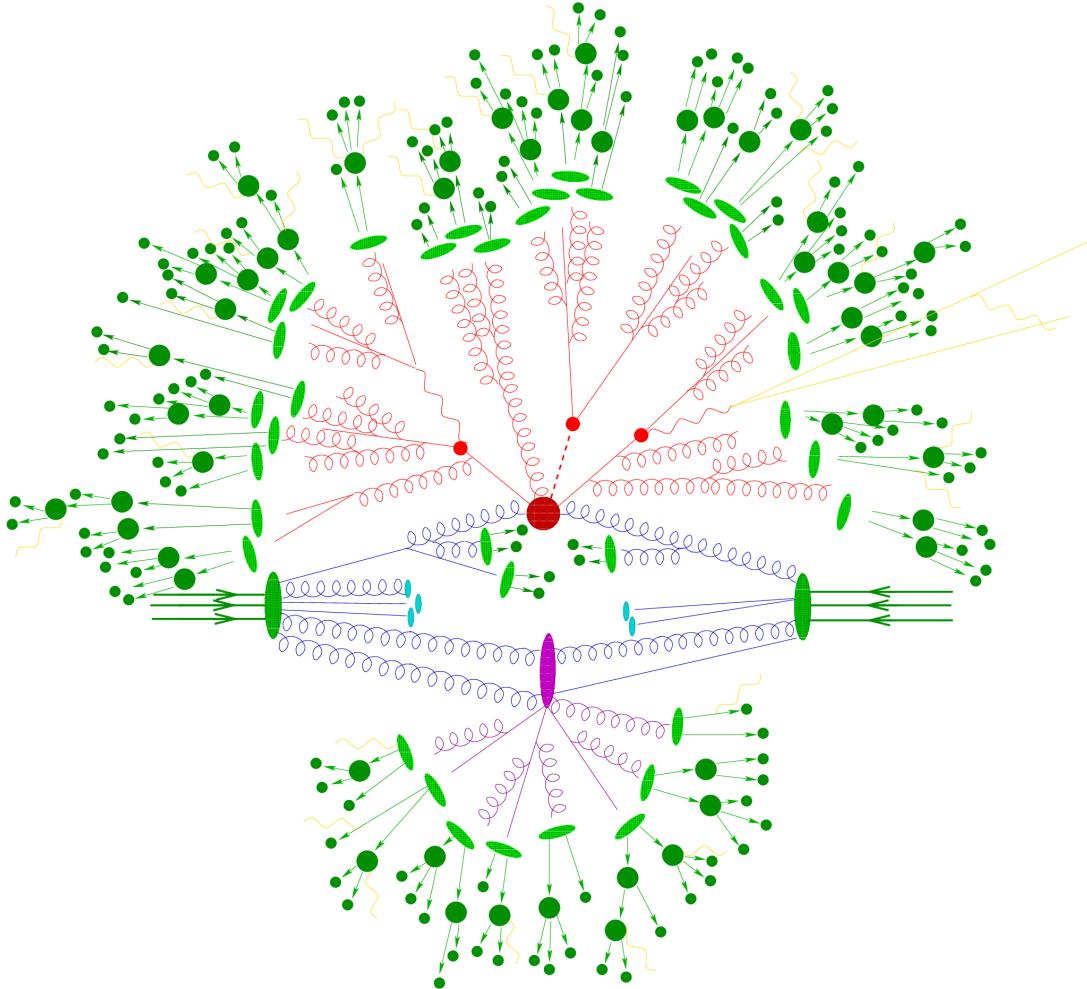


Figure 6.18: An illustration of a hadron-hadron collision as simulated by a MC generator. The hard collision center corresponds to the red circle and the branches represent parton showers. The purple circle is a second hard scattering event. The dark green circles represent the decay of hadrons and the yellow lines illustrate photons.

1472

Part III

1473

Diboson resonance searches in CMS

1474

CHAPTER 7

1475

Search I: First search for diboson resonances at 13 TeV

1476

1477 When the LHC started its Run II data taking period in summer 2015, it would be the first time
1478 ever for a particle collider to produce collisions with center-of-mass energies as high as 13 TeV.
1479 The Higgs boson, for which the LHC was designed to observe, had been discovered at the end of
1480 the previous data taking era, leaving us with a Standard Model that we know is either in need
1481 of extensions or only an effective theory valid in a certain energy domain. The Run II search
1482 program would therefore be oriented around two main efforts: Precision measurements of the
1483 newly discovered Higgs boson and searches for physics beyond the standard model.

1484

1485 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
1486 scale (to be presented 6 months after first collisions, which would be at the CERN end-of-year
1487 “Jamboree”), whose physics objects could be reconstructed and understood on a relatively short
1488 time scale, and would be robust enough in case there were issues with the never-before-validated
1489 13 TeV Monte Carlo?

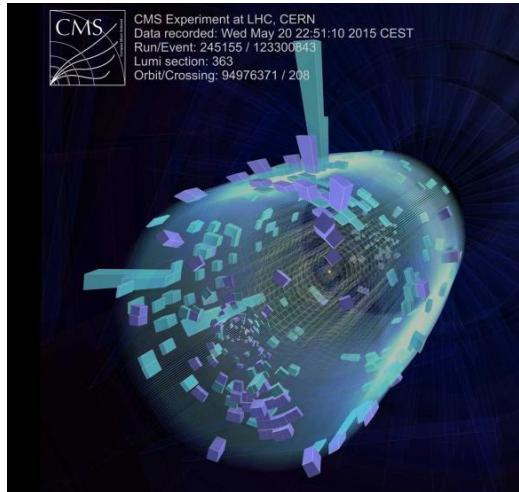
1491

1492 The attention of the high-energy physics community has in the past years been focused on certain
1493 “hot topics”: In 2018 and currently in 2019, the excitement is over leptoquarks, which could
1494 explain anomalies observed by LHCb and b-factories; in 2016 and 2017 it was diphoton resonances,
1495 with $> 3\sigma$ excesses observed at the same mass in both CMS and ATLAS. And in 2015 during the
1496 13 TeV LHC start-up, attention was centered on diboson resonances in the all-hadronic final state.
1497 The choice was therefore clear: My first analysis would be a search for diboson resonances in the
1498 boosted dijet final state. With a background model based on a smooth fit to data in the signal region,

1499 *eliminating the need for accurate QCD MC predictions, this was a simple one-background only*
 1500 *(QCD) analysis, feasible to finalize in one year, given dedication and sufficient effort. Despite its*
 1501 *straightforwardness, due to observed 8 TeV excesses, it was in addition considered a high-profile*
 1502 *analysis.*

1503

1504 *This search became one of the first “boosted” searches published with data collected with a 13*
 1505 *TeV center-of-mass energy, as well as the first search to take advantage of dedicated “grooming”*
triggers. It was published with 2.7 fb^{-1} of 2015 data.



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

1506

1507 7.1 A small bump

1508 On June 2nd, 2015, the day before CMS recorded its first ever 13 TeV event, a pre-print appeared
 1509 on the arXiv titled, “Search for high-mass diboson resonances with boson-tagged jets in proton-
 1510 proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector” [52]. It was an analysis of the full
 1511 ATLAS Run 1 dataset, corresponding to 20.3 fb^{-1} , searching for heavy resonances decaying to
 1512 vector bosons in the all-hadronic state. The analysis documented a 3.4σ excess for a heavy
 1513 resonance decaying to WZ with a mass of around 2 TeV. The corresponding CMS analysis,
 1514 published the previous year, had a 1.3σ excess at roughly the same resonance mass, but was
 1515 mostly compatible compatible with a WW final state hypothesis [58]. Figure 7.1 shows the
 1516 corresponding dijet invariant mass spectrum as seen by ATLAS (left) and the upper limit on the
 1517 production times the cross section for a G_{Bulk} decaying to WW (right) as documented by CMS.

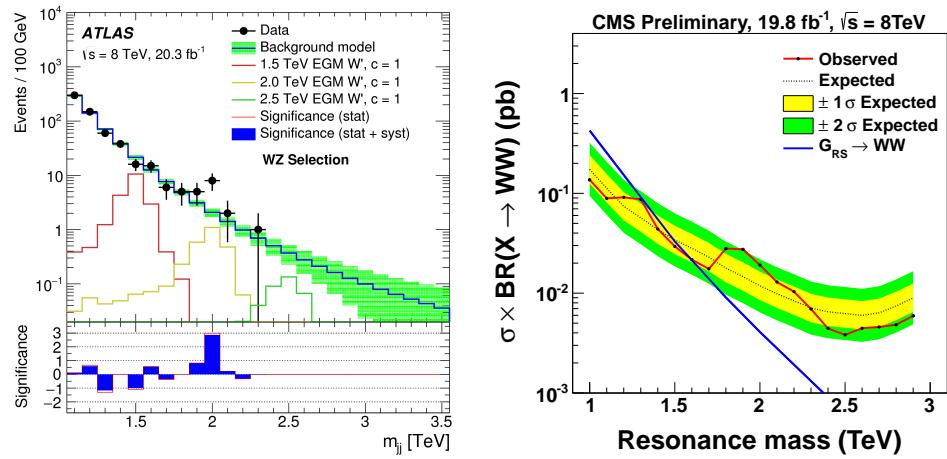


Figure 7.1: A “bump” corresponding to 3.4σ in the dijet invariant mass spectrum around 2 TeV (left) observed by ATLAS when analyzing the full 8 TeV dataset [52], together with a similar excess (1.3σ) observed in the corresponding CMS analysis [58].

1518 The two measurements were found to be compatible, favoring a heavy resonance with a
 1519 production cross section of around 5 femtobarn and a mass between 1.9 and 2.0 TeV decaying to
 1520 either WW, WZ or ZZ [59]. Figure 7.2 shows the obtained p-value from the ATLAS (red) and
 1521 CMS (blue) searches, as well as their combination (black). In addition to the observed excesses
 1522 in the vector boson final states, another 3σ excess for a resonance with a mass of 1.8 TeV had
 1523 been observed in the search for heavy resonances decaying to a W and a Higgs boson [60] at 1.8
 1524 TeV. The combination of the excesses and the timing of the ATLAS paper, naturally led to some
 1525 excitement, and in the coming weeks, the arXiv was flooded with theory papers attempting an

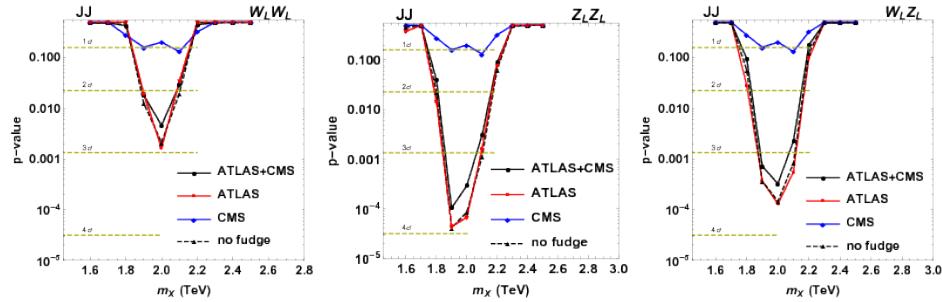


Figure 7.2: p-values as a function of resonance mass obtained with an emulation of the ATLAS (red) and CMS (blue) searches as well as the combination of the two (black). Here for a W_W (left), W_Z (middle) and Z_Z (right) hypothesis [59].

explanation of the deviations.

In addition, one of the main benefits of increasing the LHC center-of-mass energy from 8 to 13 TeV was that the partonic luminosity would increase. One could therefore expect the same number of signal events in the 20 fb^{-1} data set collected with a center-of-mass energy of 8 TeV, for a considerably smaller luminosity with a center-of-mass energy of 13 TeV. Figure 7.3 shows the system mass that can be probed with the expected 2015 integrated luminosity of 3 fb^{-1} collected with a center-of-mass energy of 13 TeV, as a function of the probeable mass with 20 fb^{-1} of 8 TeV data for different partonic channels of qq , qg , and gg . For example, a 2 TeV mass resonance would be observable in both datasets. We therefore expected that the small excess observed in the VV all-hadronic final state would be observable in the 2015 dataset if the signal was genuine.

7.2 Analysis strategy

When a resonance X with a mass above 1 TeV decays into a vector-boson pair, the bosons have a very high energy ($\tilde{p}_T = M_X/2 = 500 \text{ GeV}$, assuming X is produced at rest) is referred to as boosted. The decay products of a hadronically decaying boosted vector boson will therefore not appear as back-to-back in the lab frame but rather be collimated, as described in Section 6.5. This results in a final state with two high-pt, large-radius jets, such that the AK algorithm with an $R=0.8$ is expected to fully contain the two quarks coming from the vector boson decay. This is illustrated in Figure 7.4. The two jets are each expected to have a mass around the W or Z boson mass, and some intrinsic substructure stemming from their two-pronged decay. The invariant mass of the dijet system, m_{jj} , should be roughly equal to the resonance mass M_X . This dijet system is the final state under scrutiny and the dijet invariant mass is the parameter of interest.

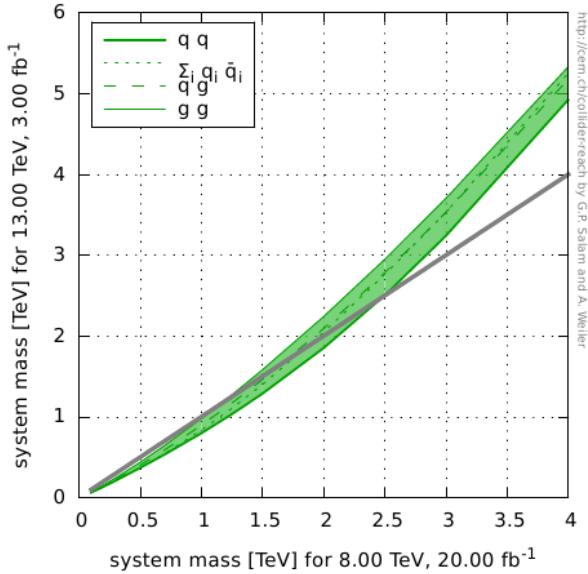


Figure 7.3: The system mass that can be probed with 3 fb^{-1} of 13 TeV data (y-axis) as a function of the probeable system mass with 20 fb^{-1} of 8 TeV data (x-axis) for different partonic channels (generated with [61]).

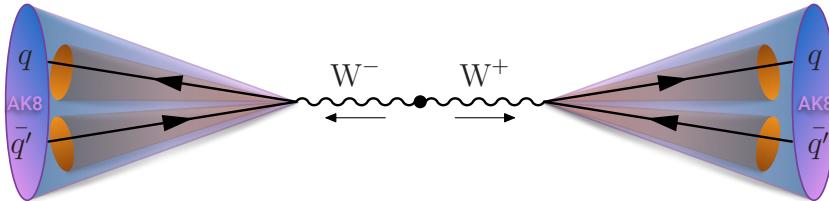


Figure 7.4: If a heavy ($> 1 \text{ TeV}$) resonance decays into vector bosons, the transverse momentum of each boson will be large and its decay products are merged into one single large cone AK8 jet.

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

The main background for such an analysis is QCD multijet events. As mentioned in Section 6.5, quark/gluon jets can obtain a high mass due to diffuse radiation and QCD processes have such a large cross section that the number of QCD jets with a mass compatible with the W mass can be large. In order to discriminate between the two, we take advantage of three properties. First, the groomed mass of signal and background jets should be very different. Second, signal jets should appear two-prong like, as opposed to quark/gluon jets, and third, the dijet invariant mass for the signal process should peak around the resonance mass while the QCD spectrum is predicted to be smoothly falling. Section 7.5 explains this assumption in more detail. The strategy therefore

1557 consists of performing a smoothness test on m_{jj} of the observed data, a so-called “bump-hunt”,
 1558 by assuming that the signal will appear as a bump on top of a smooth distribution. This is
 illustrated in Figure 7.5. The benefit of such a method is that there is no need for a simulation

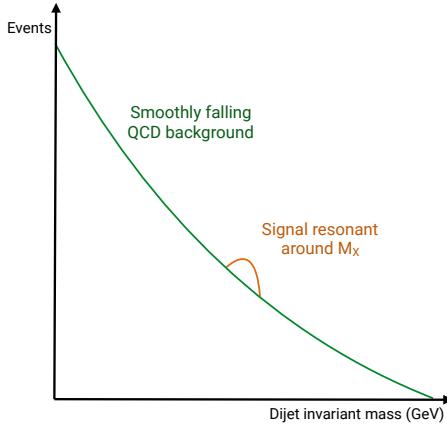


Figure 7.5: The search strategy consists of looking for signal “bumps” in the dijet invariant mass on top of a smoothly falling QCD multijet background.

1559
 1560 of the background and the strategy is simple and robust. The disadvantage is that the analysis
 1561 is intrinsically limited to regions where the dijet invariant mass spectrum is smooth and hence
 1562 regions with discontinuities due to trigger turn-ons or kinematic selections must be avoided.

1563 7.3 Data and simulated samples

1564 The data analyzed in this search correspond to a total integrated luminosity of 2.7 fb^{-1} collected at
 1565 a center-of mass energy of 13 TeV between June and December 2015. The instantaneous luminosity
 1566 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
 1567 an average number of primary vertices per event of $\langle \mu \rangle = 13$.

1568 The bulk graviton model (see Section 2.3) and the HVT model (W' and Z' , see Section 2.2.1)
 1569 are used as benchmark signal processes. In these models, the vector gauge bosons are produced
 1570 with a longitudinal polarization in more than 99% of the cases, which leads to a 24% higher
 1571 acceptance per boson for reasons explained in Section 6.5.3. For the HVT model, a scenario
 1572 (model B) with $g_V = 3$, $c_H = -0.976243$, and $c_F = 1.02433$ is chosen, where the heavy resonance
 1573 predominantly couple to bosons and the coupling to fermions is suppressed. The bulk graviton
 1574 samples were generated with $\tilde{k} = 0.5$. The resonance masses considered lie in the range 1.2 to
 1575 4 TeV and are generated under the assumption of a natural width negligible with respect to

1576 the experimental resolution (narrow-width approximation). All signal samples are generated at
 1577 leading order with MADGRAPH5_AMC@NLO v2.2.2 [62].

1578 Simulated samples of the production of QCD multijet events are generated to leading order
 1579 using PYTHIA version 8.205 [63] with the CUETP8M1 tune [57] and are used to validate the
 1580 analysis procedure.

1581 7.4 Event selection

1582 7.4.1 Triggering

1583 The first selection to be confronted in any analysis is the trigger selection. Due to an overwhelming
 1584 QCD background in all-hadronic final states, the threshold for fully-hadronic triggers is very
 1585 large in order to keep the trigger rate low (preferably around 10-30 Hz). In this analysis, we
 1586 therefore decided to take advantage of triggers that place requirements on the jet’s groomed mass
 1587 in addition to the “standard” jet triggers based on the scalar sum of jet transverse energy H_T .
 1588 These “boosted” triggers were never before tested in data, and this analysis was the first published
 1589 result taking advantage of grooming at the trigger level in CMS. The following H_T -based High
 1590 Level Triggers (HLT), referred to as inclusive triggers in the following, are used:

- 1591 • HLT_PFHT650_WideJetMJJ900DEtaJJ1p5,
- 1592 • HLT_PFHT650_WideJetMJJ950DEtaJJ1p5, and
- 1593 • HLT_PFHT800.

1594 Here, *PFHT650* refers to a total H_T of at least 650 GeV. *WideJet* means jets reconstructed
 1595 with the *wide jet algorithm* [64], an algorithm inspired by jet grooming intended to reduce
 1596 sensitivity to gluon radiation. The two AK R=0.4 jets with the largest p_T in the event are
 1597 used as seeds. Geometrically close jets are then combined into the closest jet seed if they are
 1598 within $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, and these two jets form a dijet system used for further selections.
 1599 *MJJ900* refers to a wide-jet dijet mass of at least 900 GeV, and *DataJJ1p5* means there is an
 1600 additional cut on the $|\Delta\eta|$ between the two wide jets for reasons that will be explained below. In
 1601 addition, two triggers based on jet grooming are used. These require a trimmed jet mass (see
 1602 Section 3.5.1) of 30 and 50 GeV, yielding the triggers:

- 1603 • HLT_AK8PFJet360_TrimMass30 and
- 1604 • HLT_AK8PFHT700_TrimR0p1PT0p03Mass50.

1605 The tuneable parameters for the trimming algorithm are set to $r_{sub} = 0.2$ and $p_{T,frac} = 0.03$.
 1606 The trigger requiring a trimmed jet mass of at least 30 GeV is seeded by single-object Level 1
 1607 triggers with jet p_T thresholds of 176 or 200 GeV, and the remaining triggers require an online
 1608 H_T of at least 150 or 175 GeV.

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

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

1609 where $N_{trigger+ref}$ corresponds to the number of events passing the trigger under study as well as
 1610 the reference trigger, and N_{ref} corresponds to the number of events passing the reference trigger.
 1611 Figure 7.6 shows the trigger turn-on curves as a function of dijet invariant mass for jets where
 1612 one of the jets is required to have a pruned mass larger than 65 GeV (in other words, compatible
 1613 with a W jet). A sharp turn-on for the inclusive triggers (top left) is observed, reaching the
 1614 100% efficiency plateau for dijet masses of around 1.0–1.1 TeV. The grooming triggers, however,
 1615 turn on more slowly and are not fully efficient until dijet invariant masses reach around 1.2 TeV
 1616 (top right). The real power of the grooming triggers become clear when considering them in
 1617 addition to the H_T -based triggers. The bottom plot in Figure 7.6 compares the trigger turn-on
 1618 curves as a function of dijet invariant mass for jets passing one of the three inclusive triggers
 1619 only, one of the grooming triggers only, and when combining all of them. Here, one can see that
 1620 the 99% efficiency threshold is lowered by 75 GeV when including the substructure triggers, once
 1621 substructure is required at analysis level. This combination of triggers allowed the analysis to
 1622 consider dijet invariant masses as low as 1 TeV. As a measure of the performance of the grooming
 1623 triggers, we have in addition looked at the trigger efficiencies as a function of the offline groomed
 1624 mass (using the pruned and softdrop algorithms described in Sections 6.5.1 and 6.5.1), for the
 1625 grooming trigger with the lowest mass threshold (30 GeV). This is shown in Figure 7.7, where an
 1626 additional cut on the jet transverse momentum of one of the jets of 600 GeV is required and no
 1627 other mass cut is applied. The trigger plateau is reached for offline groomed-jet masses around
 1628 50 GeV.

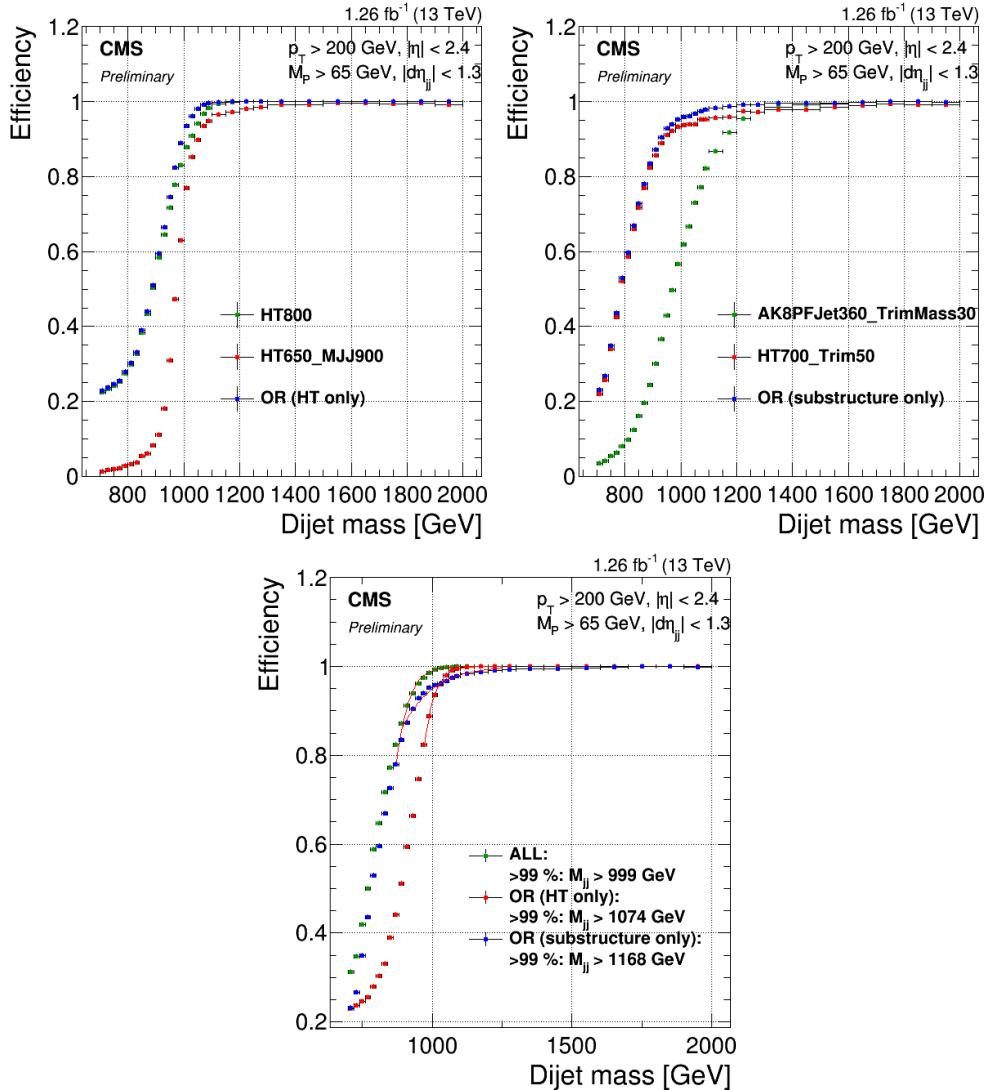


Figure 7.6: Top: Efficiency for the inclusive triggers (top left) and the grooming triggers (top right) as a function of dijet invariant mass for jet pairs where one jet has a pruned mass larger than 65 GeV. Bottom: Comparison of trigger efficiencies for jets passing one of the HT-triggers only (red), 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.

7.4.2 Preselection

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

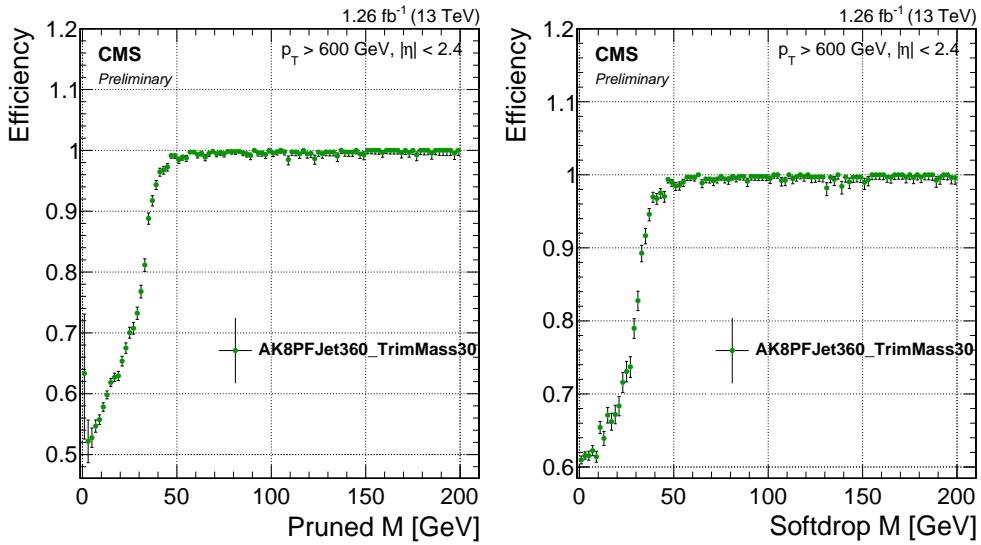


Figure 7.7: Efficiency for the lowest threshold grooming trigger as a function of pruned-jet (left) and softdrop-jet (right) mass for jets with $p_T > 600$ GeV.

1632 while keeping the background low. This is done by optimizing the selection requirements on the
 1633 jets. The jets used in this analysis are clustered with the anti- k_T jet clustering algorithm with a
 1634 clustering parameter of $R = 0.8$ (see Section 6.4) to allow containment of the full vector-boson
 1635 decay products. Since a minimum transverse momentum of 200 GeV is required for the decay
 1636 products of a W/Z to be fully contained within an $R=0.8$ jet, events are further selected by
 1637 requiring at least two jets with $p_T > 200$ GeV. These are in addition required to be central, with
 1638 an $|\eta| < 2.4$. The two highest p_T jets in the event passing these criteria are selected as potential
 1639 vector boson candidates. As our main background is QCD multijet events, we further take
 1640 advantage of the fact that the angular distribution between these, mainly t-channel, processes
 1641 are very different from the s-channel signal processes under study. The crosssection for QCD
 1642 t-channel processes as a function of the opening angle with respect to the beam axis (θ^*), exhibit
 1643 a pole around $\cos \theta^* = 1$, meaning QCD t-channel jets are mostly produced in the forward
 1644 direction, with an opening angle with respect to the beam axis close to zero. The signal jets on
 1645 the other hand, produced through an s-channel process, are concentrated in the barrel region.
 1646 We therefore require the jets to have a separation of $|\Delta\eta| < 1.3$ in order to reduce the QCD
 1647 multijets background. The distribution of $|\Delta\eta|$ between the two highest- p_T jets for QCD, as
 1648 well as for different signal scenarios, is shown in Figure 7.8. A cut of $|\Delta\eta|_{jj} < 1.3$ removes the
 1649 t-channel pole at $\cos \theta^* = 1$ and is in addition found to yield the highest signal sensitivity. In
 1650 addition to these requirements on the jets themselves, a veto on jets overlapping with leptons is

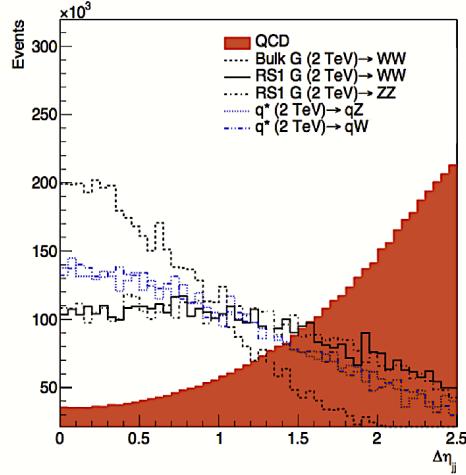


Figure 7.8: $|\Delta\eta|$ between the two highest- p_T jets for QCD jets and jets stemming from different signal scenarios.

1651 applied. Here the overlap $\Delta R(\text{jet}, \text{lepton})$ between the jet candidate and a lepton is required to
 1652 be larger than 0.8. Leptons used for this veto are required to pass the identification requirements
 1653 described in Section 6.2.2 and 6.2.2, have a transverse momentum larger than 35 (30) GeV, and a
 1654 pseudorapidity smaller than 2.5 (2.4) in the case of electrons (muons). The p_T , η , dijet invariant
 1655 mass, and $|\Delta\eta|_{jj}$ distribution for the two leading jets in the event after the above preselections
 1656 have been applied is shown in Figure 9.6.

1657 7.4.3 Vector boson tagging

1658 After preselections, we take advantage of the jet substructure algorithms described in Section 6.5 to
 1659 further separate boosted W/Z jets from the QCD multijet background. In the 8 TeV analysis [58]
 1660 published the previous year, the pruning algorithm was the chosen grooming algorithm of CMS.
 1661 However, recent progress had been made in the development of alternative grooming algorithms
 1662 that had favorable properties from a theoretical point of view (see Sections 6.5.1 and 8.5). We
 1663 therefore studied two different grooming algorithms: pruning and softdrop (with $\beta = 0$ and
 1664 $z_{cut} = 0.1$). A comparison of the jet mass for W, Z, and H jets after either the softdrop (dotted
 1665 lines) or pruning (solid lines) algorithms were applied is shown in Figure 7.10. One of the first
 1666 observations we made comparing the two grooming algorithms was that there appeared to be
 1667 a strong dependence of the softdrop mass on the jet p_T . Figure 7.11 shows the pruned (left)
 1668 and softdrop (right) mass distributions for W jets coming from the decay of a G_{bulk} with a
 1669 resonance mass of $0.8 \text{ TeV} < M_X < 4 \text{ TeV}$. While the pruned jet mass mean appeared stable

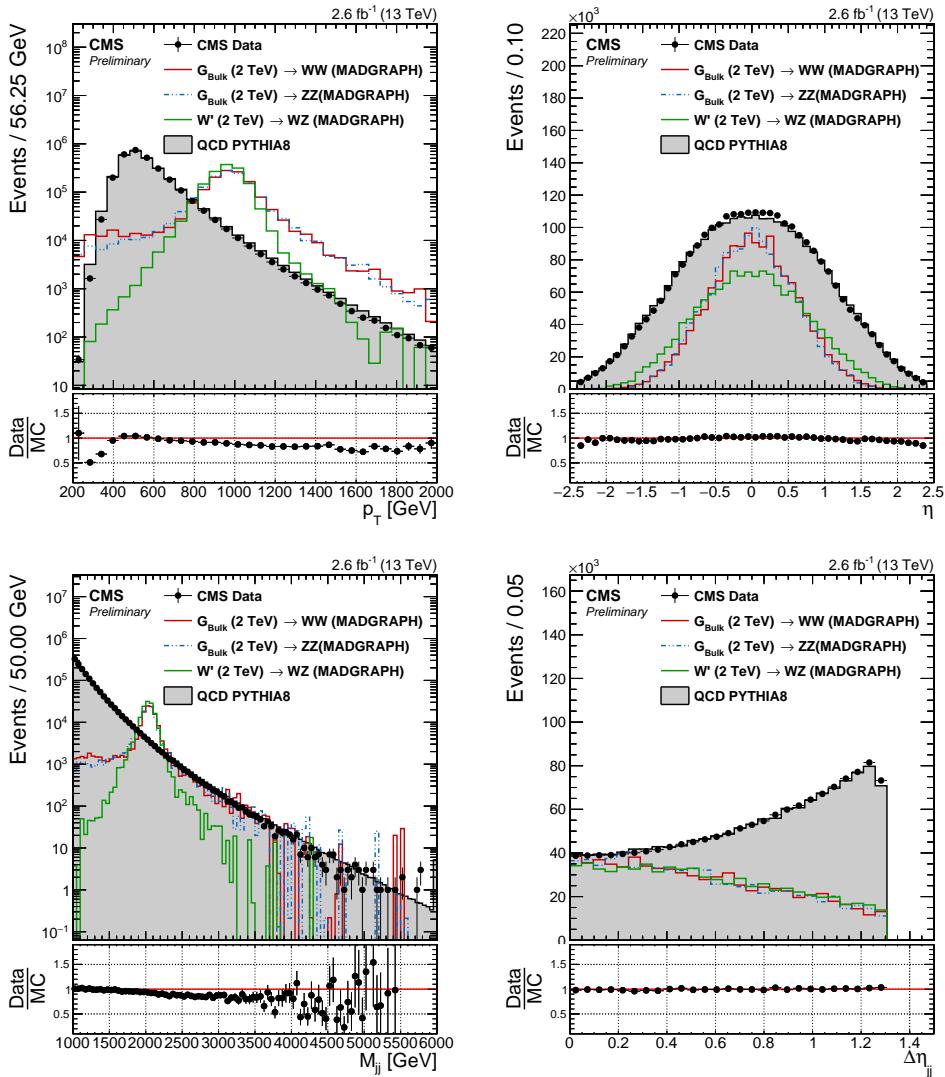


Figure 7.9: Jet p_T (top left), η (top right), dijet invariant mass (bottom left) and $|\Delta\eta|_{jj}$ (bottom right) distribution for the two leading jets in the event after loose preselections are applied. The signal is scaled by an arbitrary number.

as the jet transverse momenta of the jet increased ($p_T \sim M_X/2$), the mean of the softdrop jet mass shifted towards lower values as jet p_T increased. In order to investigate whether this was a reconstruction effect or an algorithmic effect, we additionally looked at the pruned and softdrop mass for generator-level jets (jets clustered with generator-level particles not passed through the detector simulation). Figure 7.12 shows the reconstructed (solid line) and generator-level (dotted line) jet mass distributions after pruning (left) or softdrop (right) have been applied. Again,

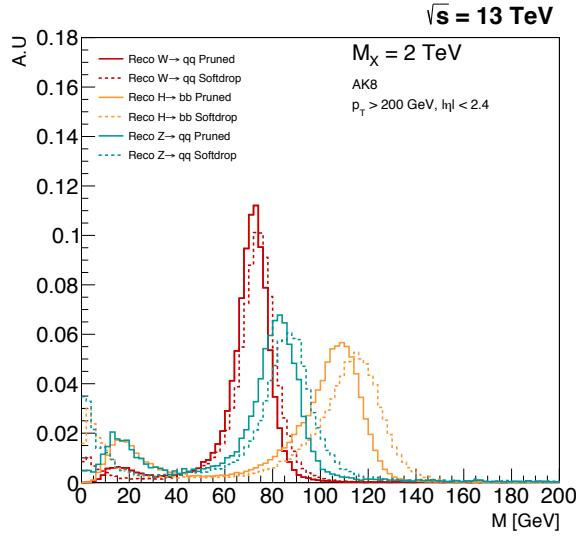


Figure 7.10: The softdrop (dotted lines) and the pruned (solid lines) jet mass for W, Z and H jets.

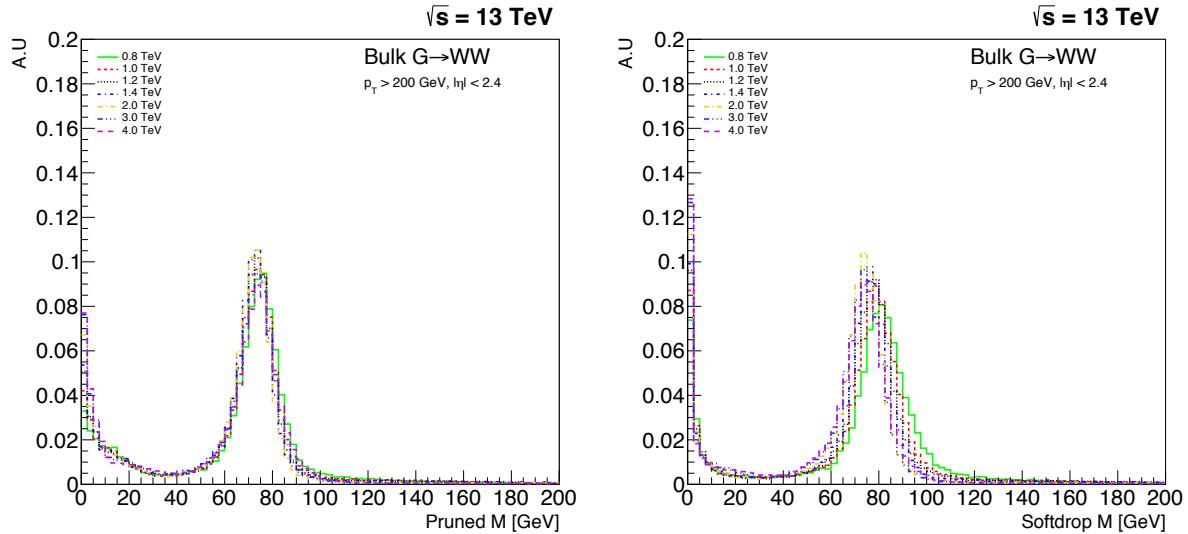


Figure 7.11: The jet mass distribution for W jets coming from a G_{bulk} of masses in the range $0.8 \text{ TeV} < M_X < 4 \text{ TeV}$ decaying to WW, here with pruning applied (left) and softdrop (right). A strong shift in the jet mass mean as a function of p_T ($\sim M_X/2$), is observed for jets groomed with the softdrop algorithm. Charge hadron subtraction is applied to all jets before clustering.

¹⁶⁷⁶ the distributions are compared for jets with very different p_T profiles, here for W jets coming

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

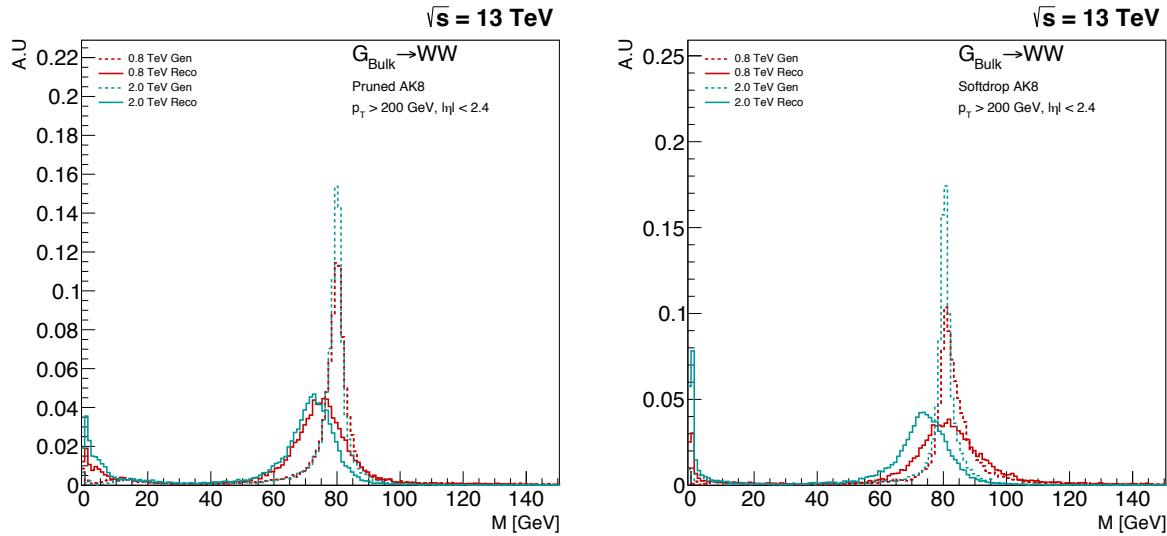


Figure 7.12: The reconstructed (solid line) and generator level (dotted line) jet mass distribution for W jets coming from a $G_{\text{bulk}} \rightarrow WW$ of mass $M_X = 0.8 \text{ TeV}$ (red), roughly $p_T \sim 400 \text{ GeV}$, and $M_X = 2.0 \text{ TeV}$ (blue), $p_T \sim 1 \text{ TeV}$. Here for the pruned (left) and softdrop (right) jet mass.

due to the fact that it would require a p_T -dependent mass window. This would again require several different measurements to produce an efficiency scale factor between simulation and data, for each mass window, or a significantly higher uncertainty on the signal yield. Due to these findings, the grooming algorithm of choice for this analysis is pruning, with the signal selection window defined as $65 \text{ GeV} < m_p < 105 \text{ GeV}$. The above findings will be important for subsequent analyses and are revisited in Section 8.

The variable used to determine the substructure of the V jets is the n-subjettiness ratio τ_{21} , as described in Section 6.5.2. The τ_{21} variable is correlated to the pruned jet mass, however, it still provides additional signal discrimination when applied after the pruned jet mass selection. Figure 7.13 shows the τ_{21} distribution for the QCD background and W jets from a signal decay before (left) and after (right) a pruned mass cut of $65 \text{ GeV} < m_p < 105 \text{ GeV}$ has been applied. We perform a cut optimization on τ_{21} after all analysis selections, including the pruned mass window of $65 \text{ GeV} < m_p < 105 \text{ GeV}$, have been applied. This is done by scanning over thresholds

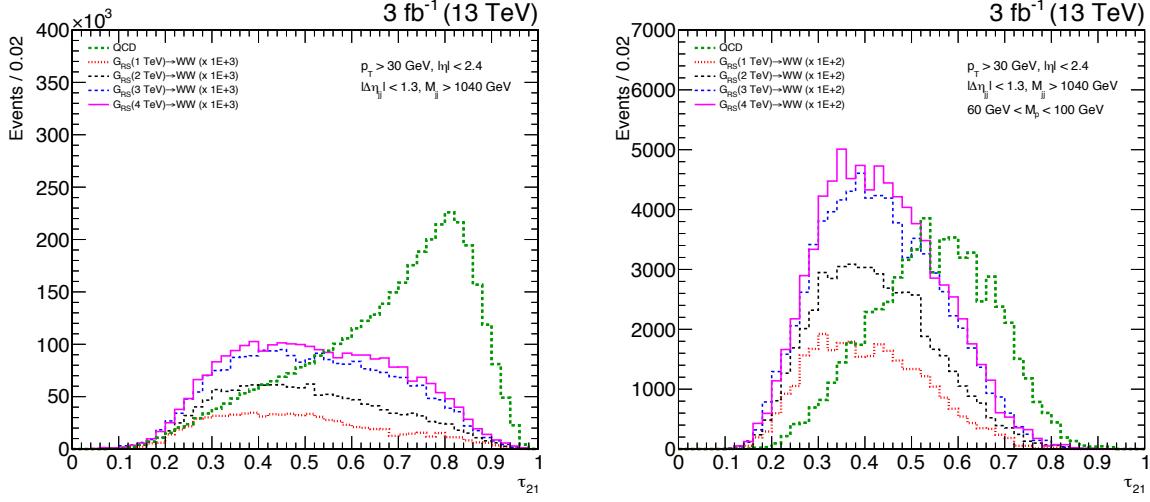


Figure 7.13: The τ_{21} distribution for QCD background and signal jets before (left) and after (right) a pruned mass window is applied. The discriminating power of τ_{21} is strongly reduced after grooming.

for the τ_{21} variable, and for each threshold, computing the Punzi significance [65] defined as

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

where ϵ_S is the signal efficiency and B is the total number of background events. The selection with the highest significance is defined as the optimal value. The signals under consideration are W jets coming from the decay of a G_{bulk} with $1 \text{ TeV} < m_X < 4 \text{ TeV}$, against a background of light-flavored QCD jets. Only jets with a dijet invariant mass in a 20% window around the resonance mass are considered. The Punzi significance as a function of the upper cut value on τ_{21} is shown on the left in Figure 7.14. 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 9.6. The number of QCD jets falls off exponentially with m_{jj} , meaning that the background at 4 TeV is considerably lower than at 1 TeV. This allows for a looser cut on τ_{21} as m_{jj} increases. In order to choose a single cut which works reasonably well for all values of resonance mass, we look at the ratio of a given τ_{21} cut over the significance of the best cut at that mass value. This is shown in the right plot of Figure 7.14. Choosing signal events with a $\tau_{21} < 0.45$ yields the most stable performance out of the investigated τ_{21} selection requirements and also maintains low background rates at low m_{jj} . This selection is therefore used for our main analysis category. In order to account for the fact that background rates are lower for

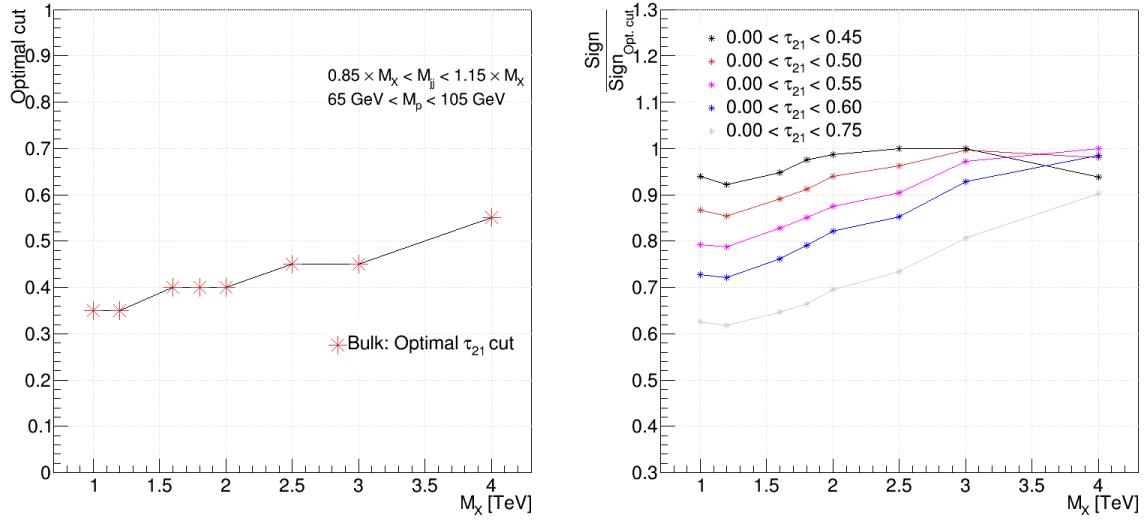


Figure 7.14: Left: Optimal upper value of τ_{21} for selecting the signal as a function of G_{bulk} mass. Right: The ratio of a given τ_{21} cut over the significance of the best cut at that mass value.

higher values of m_{jj} , we add an additional analysis category, $0.45 < \tau_{21} < 0.75$, which contains $> 95\%$ of the signal and enhances the analysis sensitivity in the case when the background is low. These categories are hereafter referred to as the *high-purity* (HP) category, for jets with $0 < \tau_{21} \leq 0.45$, and the *low-purity* (LP) category, for jets with $0.45 < \tau_{21} \leq 0.75$. QCD jets from light-flavor quarks and gluons (u,d,s,g) can be incorrectly identified as W-jets, and these events are referred to as “mistags” of the W-tagger. The W-tagging efficiency and QCD light-flavored jet mistagging rate for a W-tagger consisting of $0 < \tau_{21} \leq 0.45$ and $65 \text{ GeV} < m_p < 105 \text{ GeV}$ is shown in Figure 7.15, both as a function of jet p_T and as a function of number of primary vertices in the event. The signal efficiency when applying only the pruned jet mass selection is around 80% with a mistag rate of $\sim 15\%$. After applying the τ_{21} selection, the signal efficiency drops to around 55% and the mistagging rate to $\sim 2\%$. Another interesting feature is the dependence of τ_{21} on jet p_T and pileup, compared to the resilience of the groomed mass as a function of the same variables. This will be another feature we explore in the second analysis (Section 8). Figure 7.16 shows the pruned jet mass (left) and the τ_{21} distribution (right) for signal and background Monte Carlo, as well as the distributions measured in data.

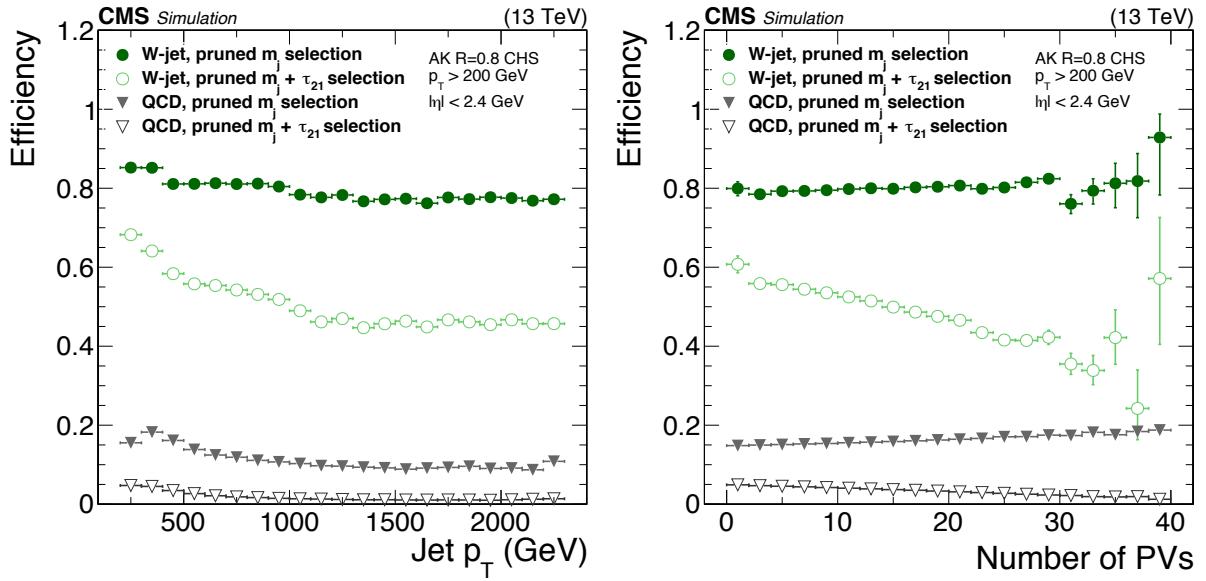


Figure 7.15: The W-tagging efficiency (green) and light jet mistag rate (grey) for a selection based on either the pruned jet mass or the pruned jet mass and τ_{21} as a function of p_T (left) and the number of primary vertices (right).

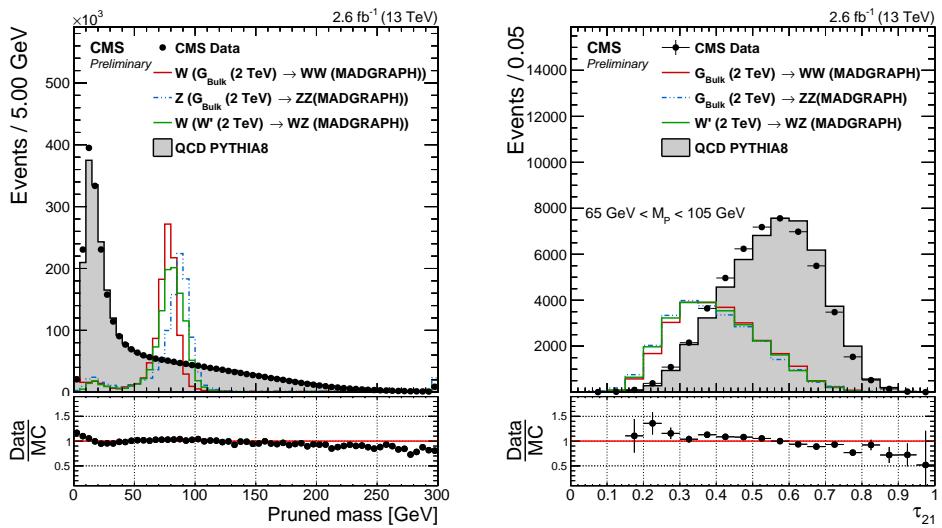


Figure 7.16: Pruned jet mass (left) and τ_{21} (right) distributions for data and simulated samples. Simulated samples are scaled to match the distribution in data. The τ_{21} distribution is shown for jets after a cut on the pruned jet mass of $65 \text{ GeV} < m_p < 105 \text{ GeV}$ has been applied.

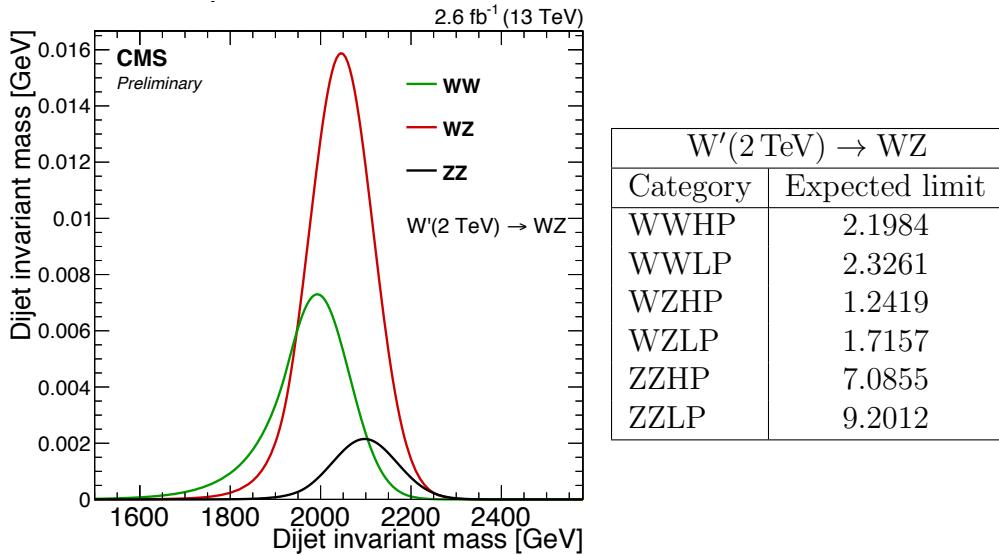


Figure 7.17: The expected signal yield per mass category for a W' (2 TeV) decaying to a W and a Z (left) together with the expected limit per mass category for the same signal (right).

1718 7.4.4 Analysis categorization

1719 As the analysis requires two W/Z -tags, we always require one HP-tagged jet and then divide into
 1720 LP and HP categories depending on whether the other jet is of high or low purity. In addition, in
 1721 order to further enhance the analysis sensitivity, we further split the pruned jet mass window
 1722 into a W and a Z boson window where the W window is defined as $65 \text{ GeV} < m_p < 85 \text{ GeV}$ and
 1723 the Z boson window as $85 \text{ GeV} < m_p < 105 \text{ GeV}$. This has the added benefit of allowing us to
 1724 discriminate between a G_{bulk} decaying to WW or ZZ , and a W' decaying into WZ by separating
 1725 these events into categories. The signal yield will be higher in the WZ category for a W' decaying
 1726 to a W and Z boson than for a G_{bulk} decaying to WW or ZZ . Figure 7.17 shows the relative
 1727 expected signal yield (left) and expected limits (left) in the different mass categories for a W' with
 1728 a mass of 2 TeV. All categories are combined in the end, leading to the same or better sensitivity
 1729 at each resonance mass value than when using the whole pruned mass window. Figure 7.18 shows
 1730 the expected 95% CL upper limits on the production cross section of a W' decaying to WZ (left)
 1731 and a G_{bulk} decaying to WW (right) as a function of the resonance mass in the HP category.
 1732 The blue line corresponds to the expected limits obtained when not splitting into mass categories
 1733 and the red line corresponds to the limit using the combination of two categories. The dotted
 1734 and solid black lines are the limits for the W and Z categories, respectively. The combination of
 1735 two mass categories leads to slightly better (10%) or similar sensitivity as when using one large

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

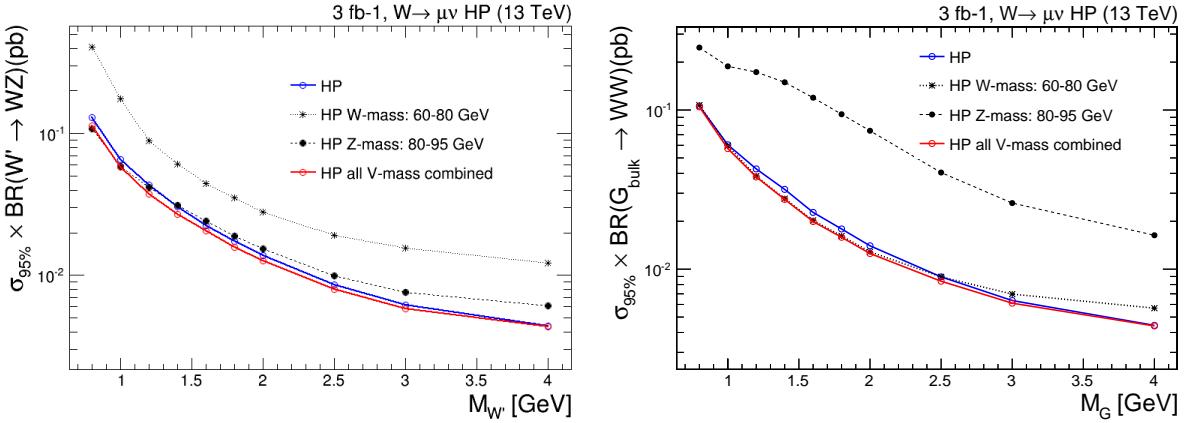


Figure 7.18: Expected 95% CL upper limits on the production cross section of a W' (left) and G_{bulk} (right) signal as a function of the resonance mass for the different mass categories for events passing the high-purity τ_{21} selection.

1736

1737 a test statistics based on the likelihood ratios of each signal hypothesis, $q = -2 \ln(L_{G_{\text{bulk}}}/L_{W'})$,
 1738 shown in Figure 7.19. For a signal with a signal strength corresponding to a 3-4 σ excess, the test
 1739 statistics for each signal hypothesis are well separated ($\sim 3.5\sigma$), allowing us to make a statement
 of whether a possible signal is more likely due to a G_{bulk} or a W' particle. With the high-purity

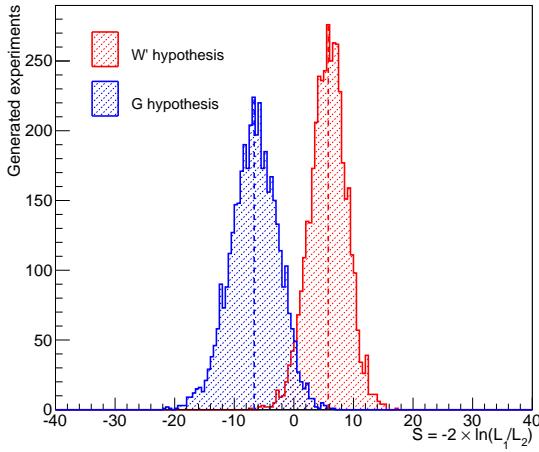


Figure 7.19: Distribution of the test statistic $q = -2 \ln(L_{G_{\text{bulk}}}/L_{W'})$ for a G_{bulk} (blue) and W' signal hypothesis.

1740

1741 and low-purity categories as defined above for each mass window combination, this leaves us with

1742 six different signal categories: HP with 3 mass categories corresponding to WW, WZ, and ZZ, and
1743 the same 3 mass categories for LP. In parallel to the mass-category based analysis, we perform an
1744 analysis without categorization in mass (similar to the 8 TeV analysis) as a cross-check. We found
1745 the sensitivity with mass categories to be higher and hence will not present these studies here.
1746 The final tagging efficiency for different signal hypotheses (top) together with the QCD mistag
1747 rate (bottom) in the different signal categories is shown in Figure 7.20. The solid lines represent
1748 the tagging efficiency in the full mass window ($65 \text{ GeV} < m_p < 105 \text{ GeV}$) before splitting into
1749 mass categories. A lower signal efficiency in the ZZ mass category is observed in all cases. This
1750 can be explained from the pruned jet mass distribution on the left in Figure 7.16, where a lower
1751 pruned jet mass selection of 85 GeV leaves a large fraction of the Z peak in the W mass window.
1752 The main benchmark models under consideration preferably decay to W bosons, since in the
1753 Bulk Graviton model the branching ratio $\text{BR}(G_{\text{Bulk}} \rightarrow \text{WW}) = 2 * \text{BR}(G_{\text{Bulk}} \rightarrow \text{ZZ})$ and in the
1754 HVT model $\text{W}'/\text{Z}' \rightarrow \text{WZ}/\text{WW}$ (but not ZZ). Therefore, the tagging procedure is optimized for
1755 a high efficiency to tag W bosons. In the limit-setting procedure all the categories are combined
1756 and the overall signal efficiency is maintained. For the combined mass-categories (solid line) the
1757 signal efficiency is between 16 and 23% in the double-tag categories, and between 20 and 34 % in
1758 the single-V tag categories. The mistagging rate is below 1% in the high-purity category. The
1759 full analysis selections and final signal categories are listed in Table 7.1, where m_{jet_1} and m_{jet_2}
1760 refers to the jet with the highest and second highest jet p_T , respectively.

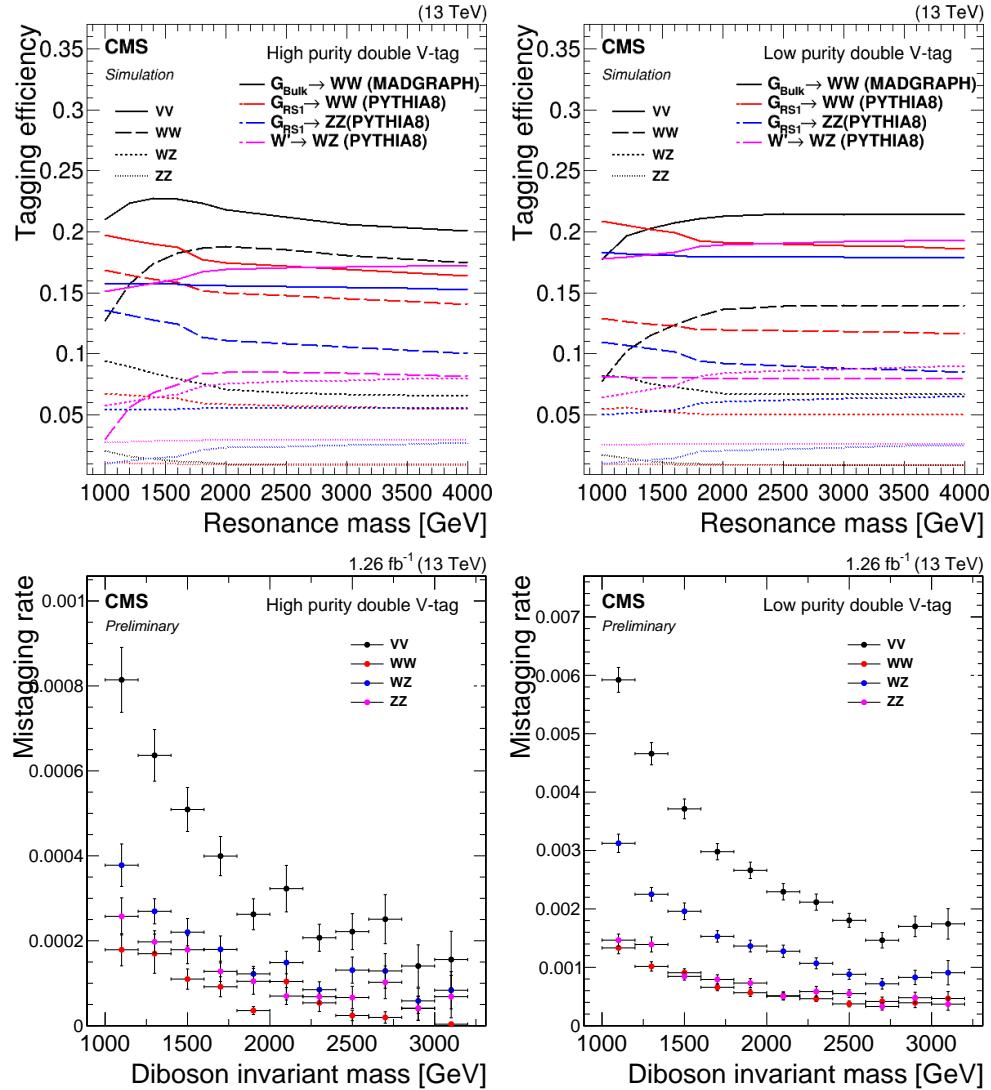


Figure 7.20: Tagging efficiency (top) and mistagging rate (bottom) in the different pruned mass categories in the high-purity category (left) and in the low-purity category (right).

Selection	Value
Boson selections	
$V \rightarrow q\bar{q}$ (2 AK8 jets)	$p_T > 200 \text{ GeV}$ $ \eta < 2.4$
Pruned jet mass	$65 < m_{\text{jet}_1}, m_{\text{jet}_2} < 105 \text{ GeV}$
Topology	$ \Delta\eta_{jj} < 1.3$
Dijet invariant mass	$m_{jj} > 1 \text{ TeV}$
2- to 1-subjettiness ratio	$\tau_{21} < 0.75$
<i>m_{jet}</i> 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	$85 < m_{\text{jet}_1} < 105 \text{ GeV}, 85 < m_{\text{jet}_2} < 105 \text{ GeV}$
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 7.1: The full analysis selections, and the signal categories based on the pruned jet mass and τ_{21} values.

¹⁷⁶¹ 7.5 Background modeling

¹⁷⁶² We assume that the QCD multijets background can be described by a smooth, monotonically
¹⁷⁶³ decreasing function, which is parametrizable. This is similar to what is done in previous CMS
¹⁷⁶⁴ analyses looking for bumps in the dijet invariant mass spectrum [66, 67]. The search is then
¹⁷⁶⁵ performed by fitting the sum of the functions for background and signal to the dijet invariant
¹⁷⁶⁶ mass spectrum in data. Neither data control regions nor simulated samples are used directly by
¹⁷⁶⁷ this method. The background functions are of the following form:

$$\frac{dN}{dm_{jj}} = \frac{P_0}{(m_{jj}/\sqrt{s})^{P_2}} \quad \text{and} \quad \frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}}, \quad (7.1)$$

where m is the dijet invariant mass, \sqrt{s} is the center-of-mass energy, P_0 is a normalization parameter for the probability density function, and P_1 and P_2 describe the shape. The number of fit parameters is decided through a Fisher’s F-test [68]. 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. If the confidence level is less than 10%, we add additional parameters to model the background distribution. The 4- and 5-parameter functions are

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2 + P_3 \times \log(m_{jj}/\sqrt{s})}} \quad \text{and} \quad (7.2)$$

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s})^{P_1}}{(m_{jj}/\sqrt{s})^{P_2 + P_3 \times \log(m_{jj}/\sqrt{s}) + P_4 \times \log(m_{jj}/\sqrt{s})^2}}, \quad (7.3)$$

¹⁷⁶⁸ where P_3 and P_4 are additional free parameters. As an additional crosscheck, an alternative fit
¹⁷⁶⁹ function is also tested:

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/\sqrt{s} + P_3(m_{jj}/\sqrt{s})^2)^{P_1}}{(m_{jj}/\sqrt{s})^{P_2}}. \quad (7.4)$$

¹⁷⁷⁰ The fit range is chosen such that it starts where the trigger efficiency has reached its plateau to
¹⁷⁷¹ avoid bias from trigger inefficiency, and extends to the bin after the highest m_{VV} mass point.
¹⁷⁷² The binning chosen for the fit follows the detector resolution as in Refs. [66, 67], and is referred to
¹⁷⁷³ as the “dijet binning”. Before unblinding the signal region, we check that the QCD dijet invariant
¹⁷⁷⁴ mass spectrum is expected to be smooth from the distribution in QCD MC as well as exercise
¹⁷⁷⁵ the F-test in QCD MC and in a data sideband.

¹⁷⁷⁶ The fits to data in the signal region using the different fit functions are shown in Figure 7.21,

and the corresponding F-test output are given in Tables 7.2 through 7.4. The findings can be summarized as follows: for the WW enriched category, a 2-parameter fit is sufficient to describe the data in both the high- and low-purity categories. In the WZ category, a 2-parameter fit is sufficient in the high-purity category, while three parameters are needed for the low-purity category. For the ZZ category, a 3-parameter fit is needed for both purity categories. The 2- and 3-parameter fit functions as defined in Equation C.2 will therefore be used to model the background component in the simultaneous signal and background fit.

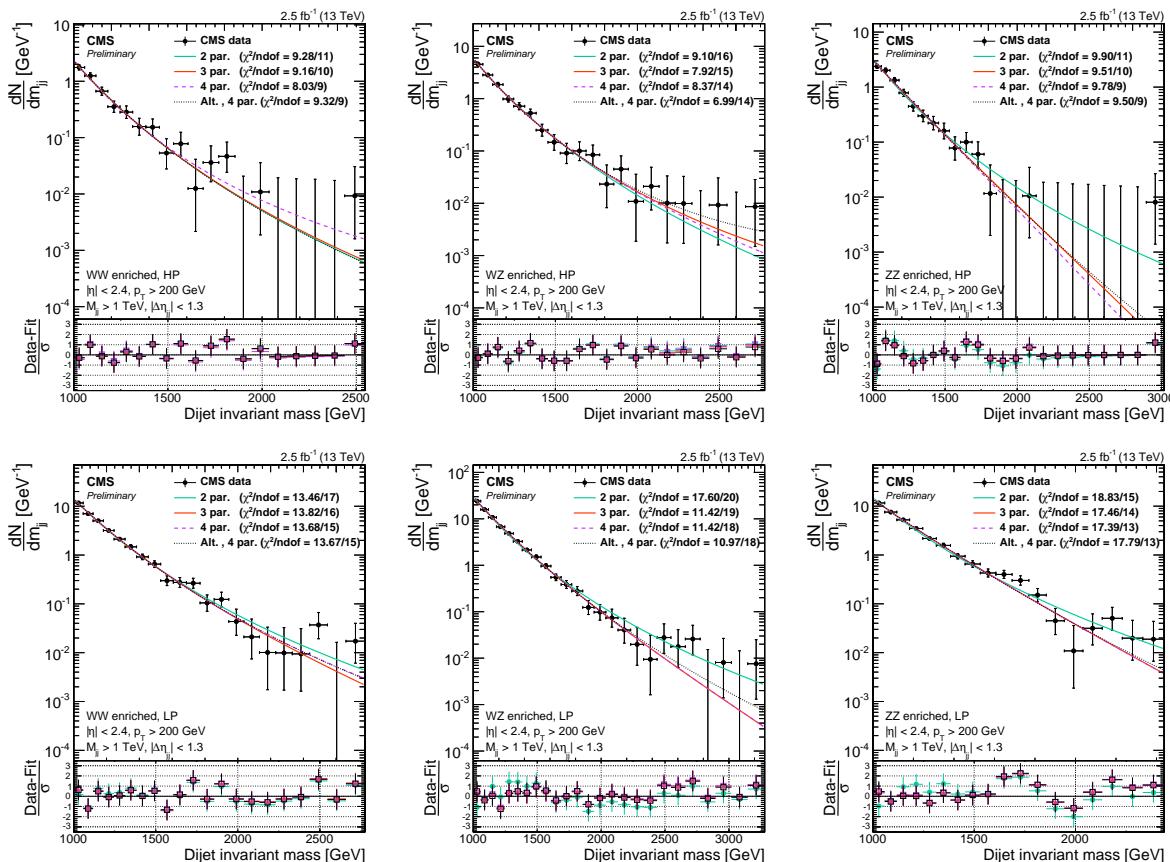


Figure 7.21: Fitted dijet mass spectrum in the different mass and purity categories in data. A 2-parameter fit is sufficient to describe the data in the WW- (HP and LP) and WZ-enriched (LP) categories. For the ZZ- (HP and LP) and WZ-enriched (HP) categories, a 3-parameter fit is needed.

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 7.2: 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 7.3: 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 7.4: 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.

¹⁷⁸⁴ **7.6 Signal modeling**

¹⁷⁸⁵ The signal shape is extracted from signal MC with resonance masses in the range from 1 to 4 TeV.
¹⁷⁸⁶ A linear interpolation provides shapes for the mass points in between in steps of 100 GeV. From
¹⁷⁸⁷ these shapes, signal shape models are constructed as composite models with a Gaussian core due
¹⁷⁸⁸ to detector resolution and an exponential tail to account for parton distribution function effects.
¹⁷⁸⁹ Parametric shape uncertainties due to jet energy scale and resolution uncertainties are inserted by
¹⁷⁹⁰ variations of the Gaussian peak position and width. The dijet invariant mass shape for different
¹⁷⁹¹ benchmark model signals is shown in Figure 7.22. The signal and background components are
then simultaneously fitted to the data points.

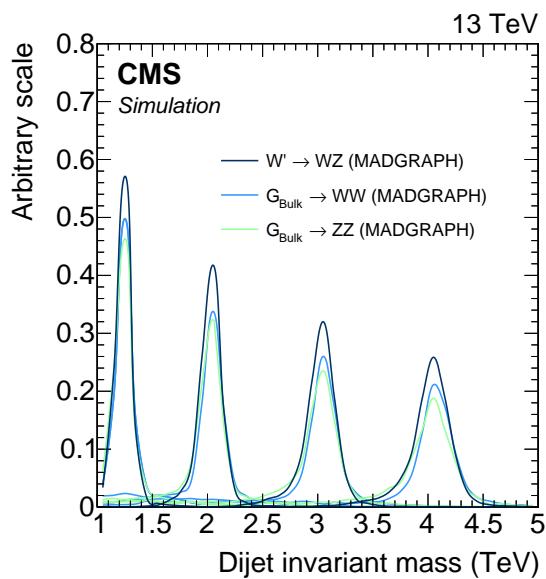


Figure 7.22: Dijet invariant mass from signal MC used to extract the signal shape, shown here for resonances with masses of 1.2, 2, 3, and 4 TeV.

¹⁷⁹³ 7.7 V-tagging scale factors

¹⁷⁹⁴ As seen in Figure 7.16, a discrepancy is observed in the τ_{21} distribution between data and MC.
¹⁷⁹⁵ This could lead to a bias in the signal efficiency estimation so we must measure the real signal
¹⁷⁹⁶ efficiency in data in an orthogonal data sample. The W-tagging efficiency is measured using real
¹⁷⁹⁷ boosted W-jets in a data sample enriched in $t\bar{t}$ decays with a hadronically decaying W boson.
¹⁷⁹⁸ This region is mainly quark-enriched, as opposed to the gluon-enriched QCD region previously
¹⁷⁹⁹ studied, and substructure variables are better described here. The sample is obtained by requiring
¹⁸⁰⁰ a final state compatible with two b-jets and two W bosons, where one of the bosons decays
¹⁸⁰¹ leptonically and the other one hadronically. There are several good reasons to use this channel:
¹⁸⁰² Top-quark pair production events are plentifully produced at the LHC; we can ensure a high
¹⁸⁰³ purity of the sample by requiring a high-energy lepton, b-tag and missing energy requirements;
¹⁸⁰⁴ and lastly, we can ensure that the W jets are boosted by requiring the leptonic leg, together
¹⁸⁰⁵ with the hadronic W boson candidate, to have high transverse momentum. The final state is
¹⁸⁰⁶ illustrated in Figure 7.23, with the object of interest being the AK R=0.8 jet containing the two
quark daughters of the hadronically decaying W boson.

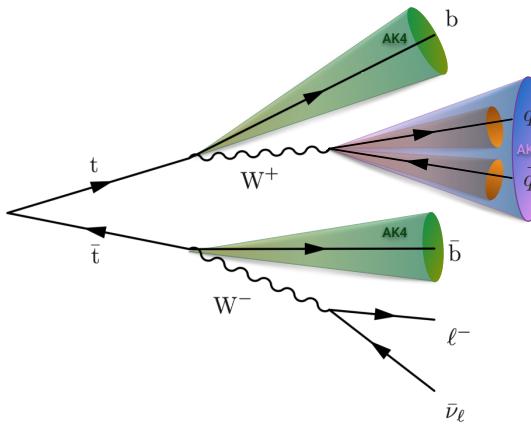


Figure 7.23: A top quark pair decaying into two b quarks and two W bosons, one of which decays leptonically and one of which decays hadronically.

¹⁸⁰⁷

¹⁸⁰⁸ 7.7.1 Event selection

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

p_T above 45 GeV for the muon or 135 GeV for the electron. This requires an offline muon (electron) p_T threshold of 53 (120) GeV. The leptons are further required to pass the lepton requirements defined in Section 6.2.2, and events containing additional leptons (passing the same ID requirements, but looser cuts as defined in Table 7.5) are vetoed. Offline, we further require a high missing energy of 40 (80) GeV in the muon(electron) channel. To ensure a high signal purity of boosted, hadronically decaying W bosons, the four-vector of the leptonically decaying W boson is reconstructed such that we can put tight momentum requirements on the leptonic W boson (ensuring that both tops, and therefore vector bosons, have a high momentum). The leptonic W boson is reconstructed in two steps: First, the unknown z component of the neutrino momentum must be solved for through a second order equation assuming the real W boson mass

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

1809 This results in a completely defined neutrino four-vector, which is then added to the lepton
 1810 four-vector. The sum of the two defines the leptonic W boson four-momentum, and its momentum
 1811 is required to be greater than 200 GeV. Further, we require at least one AK R=0.4 jet to be
 1812 b-tagged with the Combined Secondary Vertex (CSV) algorithm [69, 70]. This algorithm exploits
 1813 the relatively long lifetime and large mass of b hadrons that leads to the presence of a displaced
 1814 vertex in order to distinguish between jets originating from b quarks and those originating from
 1815 light-flavor quarks. More information on the CSV algorithm can be found in [69, 70]. The reason
 1816 for requiring only one b-tagged jet is to ensure a high selection efficiency. Finally, we require at
 1817 least one AK R=0.8 jet in the event with a momentum greater than 200 GeV, which will be the
 1818 hadronic W boson candidate. It's pruned jet mass is required to be between 40 GeV and 150
 1819 GeV. After reconstructing and selecting all our objects, a set of angular selections are applied to
 1820 ensure a diboson-like topology. These are the following:

- 1821 • the ΔR between the lepton and the hadronic W boson candidate must be $< \pi/2$,
- 1822 • the $\Delta\phi$ between the hadronic W boson candidate and the E_T^{miss} must be > 2 , and
- 1823 • the $\Delta\phi$ between the hadronic W boson candidate and the lepton must be > 2 .

1824 With these requirements, we have a nearly pure sample of $t\bar{t}$ events, with a small contamination
 1825 from single top-quark production, W+jets, and VV events. A summary of the final selection
 1826 criteria is presented in Table 7.5. The pruned jet mass and τ_{21} variables in data and in MC are
 1827 shown in Figure 7.24.

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 W reconstruction
Jet η	$ \eta < 2.4$	
AK4 jet selections		
Jet p_T	$p_T > 30 \text{ GeV}$	Used for b-tag jet selection
Jet η	$ \eta < 2.4$	
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	
$\Delta\phi(W_{\text{AK8}}, W_{\text{lept}})$	> 2	

Table 7.5: Summary of the final semi-leptonic $t\bar{t}$ selections.

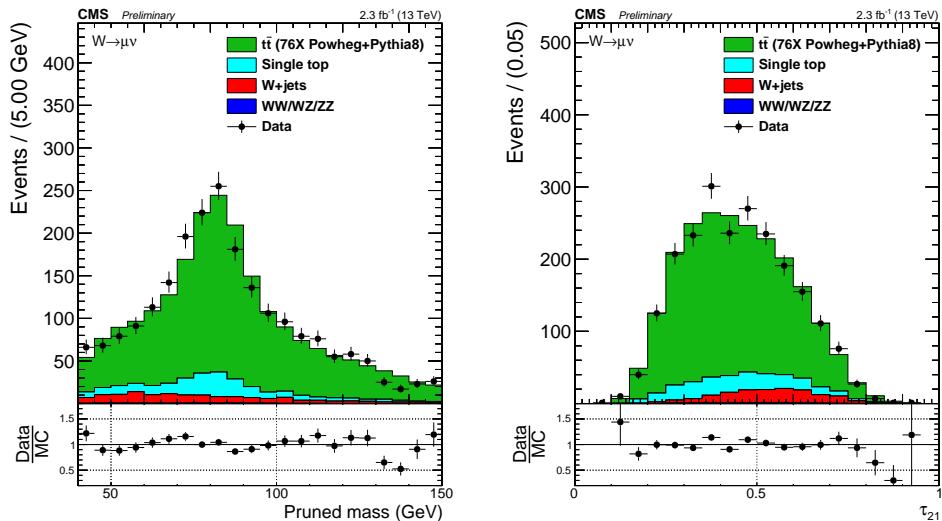


Figure 7.24: Distribution of pruned jet mass (left) and n-subjettiness (right) in the $t\bar{t}$ control sample.

1828 7.7.2 Fitting procedure

1829 For this measurement, what we are interested in is to extract and compare the W-tagging efficiency
 1830 of the combined selection of jet mass and τ_{21} in data and MC. We are additionally interested in
 1831 the difference in jet-mass scale (mean of the W-jet mass peak) and jet-mass resolution (width
 1832 of the W-jet mass peak), as this also affects the shape of the signal jet mass and therefore the
 1833 tagging efficiency. In order to study these variables, we look at the pruned jet mass spectrum
 1834 between 40 and 150 GeV in two regions:

- 1835 • Pass region: $0 < \tau_{21} \leq 0.45 \sim$ high purity , and
- 1836 • Fail region: $0.45 < \tau_{21} \leq 0.75 \sim$ low purity.

1837 Our goal is to understand what the real fraction of merged W jets is in the pass category and in
 1838 the fail category, assuming that the sum of the two correspond to a 100% selection efficiency (the
 1839 amount of W boson jets falling outside of this region is negligible). The strategy is as follows:
 1840 We first derive probability density functions (PDFs) which describe the distribution of fully
 1841 merged W boson jets and non-W boson jets in $t\bar{t}$, both in the pass and in the fail region. The
 1842 PDFs describing real W jets and non-W jets are added with a fraction which is left floating:
 1843 the fit decides what the fraction of real W to non-W jets is in the pass and in the fail region.
 1844 A simultaneous fit of pass and fail regions is then performed (using the two composite PDFs),
 1845 where the fraction of real W jets in both pass and fail regions is constrained such that, if the
 1846 signal efficiency in the pass region is ϵ_S , the signal efficiency in the fail region is $(1 - \epsilon_S)$. This
 1847 is done by letting the normalization of the PDF describing real W jets in the pass category be
 1848 defined as the total real W boson yield in the pass and fail regions combined, multiplied by some
 1849 fraction, ϵ_S . The normalization of the PDF describing real W boson jets in the fail category is
 1850 then the total real W boson yield in the pass and fail regions combined, multiplied by $(1 - \epsilon_S)$.

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

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

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

Figure 7.25 shows the fitted pruned jet mass spectrum for real W jets (upper) and non-W jets (lower) in $t\bar{t}$ for jets that passed (left column) and failed (right column) the N-subjettiness selection $\tau_{21} < 0.45$.

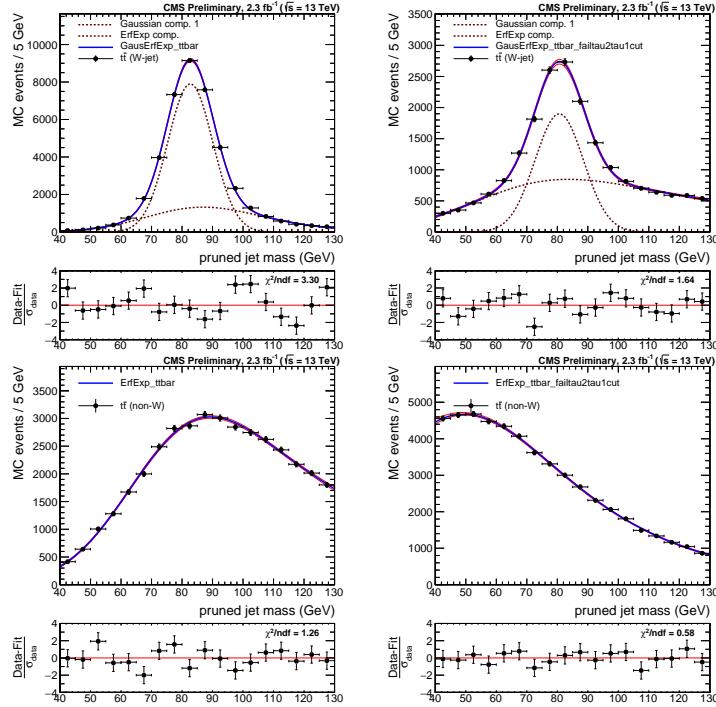


Figure 7.25: Fit to the real W (upper) and non-W (lower) boson pruned jet mass distribution for jets that pass (left) and fail (right) the cut on $\tau_{21} < 0.45$.

These shapes constitute the fit functions used for the simultaneous fit. As can be seen from the fit to real W jets in the pass region, the distribution is not purely Gaussian and has a tail at higher groomed masses. This tail depends on the matching requirements used to define real merged W jets and is unphysical. We therefore assume that the distribution of real W jets can be described by a Gaussian only, allowing the exponential error function used to describe non W-jets to cover the contribution from the tails, thereby taking the number of real W jets as the integral of the Gaussian shape only. This eliminates two additional fit functions, corresponding to six free parameters from the fit. In older estimations of the W-tagging scale factor based on the same procedure [71]), the functions used to describe the tail of the real W-jet distributions were also taken into account as contributing to the real W-jet tagging efficiency. These two calculations test two extremes: the new method assumes a Gaussian peak, absorbing the tails into the background function making the fit more robust, while the old method assumes a Gaussian peak with tails estimated from matched MC. The latter uses a more precise definition of real W jets, but yields a

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

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

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

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

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

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

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

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

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

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

¹⁸⁵⁴ with the corresponding distributions shown in Figure 7.26. The floating parameters of the fit
¹⁸⁵⁵ (besides the PDF shape parameters themselves) are the rates N_W , N_2 and N_3 , and the mean and
¹⁸⁵⁶ 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
¹⁸⁵⁷ simulation efficiencies is then taken as the W-tagging scale factor:

$$SF_{HP} = \frac{\epsilon_{HP}(\text{data})}{\epsilon_{HP}(\text{sim})}. \quad (7.5)$$

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

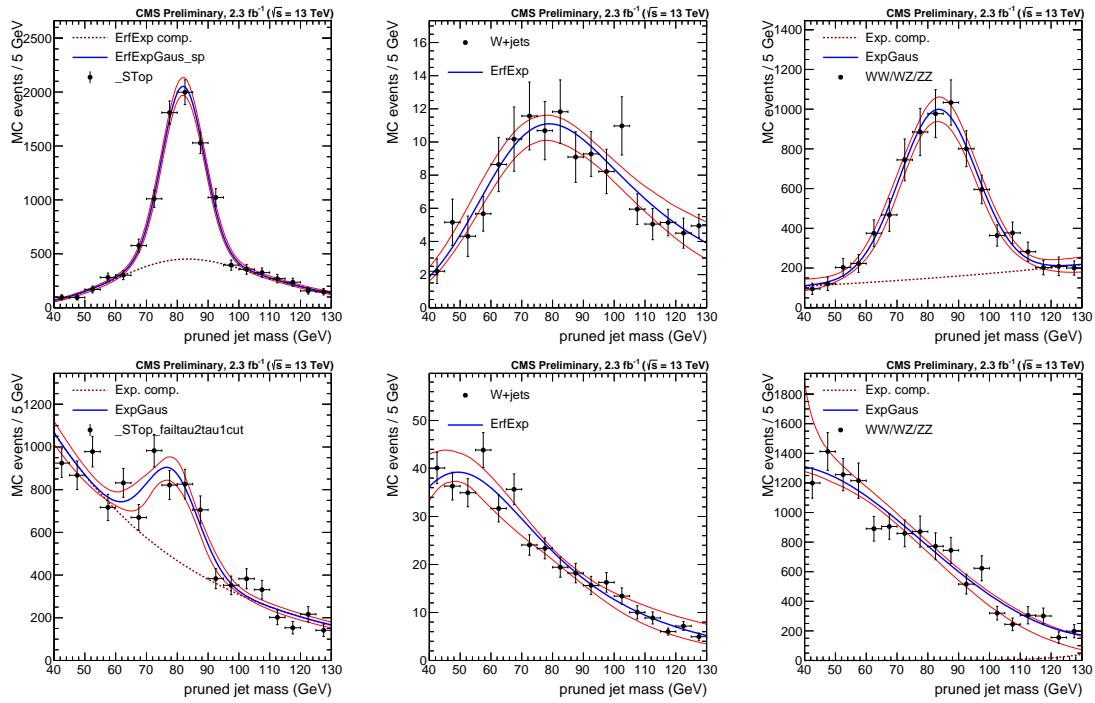


Figure 7.26: Fits to the pruned jet mass spectrum for the non-dominant backgrounds (Single top, W+jets and VV respectively) in the pass (upper) and fail (lower) regions.

purity category can be defined as:

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

1858 where ϵ_{fail} is the ratio between the number of events with $\tau_2/\tau_1 > 0.75$ and the total number of
 1859 events. As mentioned previously, the number of real W jets with $\tau_2/\tau_1 > 0.75$ is negligible and
 1860 the definition of the low purity scale factor simplifies to

$$SF_{LP} = \frac{1 - \epsilon_{HP}(\text{data})}{1 - \epsilon_{HP}(\text{sim})}. \quad (7.6)$$

1861 7.7.3 Systematic uncertainties

1862 As systematic uncertainties, we consider effects due to differences between $t\bar{t}$ simulation as well as
 1863 effects due to the choice of fit method. The former is evaluated by comparing the extracted scale
 1864 factor when using $t\bar{t}$ MC samples produced with three different combinations of matrix element
 1865 (ME) and shower generators: POWHEG (NLO) interfaced with PYTHIA8, MADGRAPH (LO) QCD

interfaced with HERWIG++, and POWHEG interfaced with HERWIG++. The uncertainty due to different ME generators (POWHEG versus MADGRAPH) corresponds to 3(13)% and is listed in Table 7.6 as the first quoted systematic uncertainty. The uncertainty due to parton showering (PYTHIA8 versus HERWIG++) is 8.6%, but is not relevant for analyses where no HERWIG++ based simulation is used, as is the case for the search presented in this chapter. For the systematic uncertainty accounting for effects due to choice of fit method, we compare the estimated extracted efficiency in $t\bar{t}$ MC using the two different fit models described above: the new model, where the signal is modeled by a Gaussian peak and the tails of the distribution are absorbed in the background fit model, and the old model, including the tails when calculating the fraction of real W jets. Figure 7.27 shows the fits obtained in the pass and fail regions using the two different models. With the new model only the Gaussian component of the fit contributes to the W-tagging efficiency while, with the old model, a Chebyshev component is additionally contributing to the total W-tagging efficiency. The estimated efficiencies obtained using both methods, after

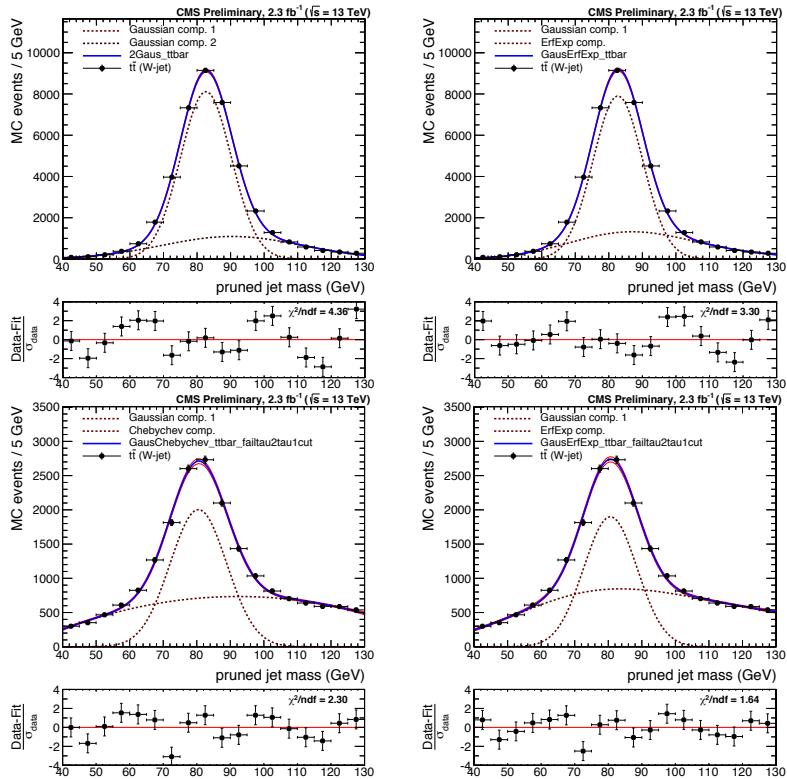


Figure 7.27: Fits obtained in the pass (top) and fail (bottom) regions using two different models: An alternative model with tails (top and bottom, left) where the tail component is contributing to the total W-tagging efficiency. When using the default model (top and bottom, right), only the Gaussian component of the fit contributes to the W-tagging efficiency.

1879 being corrected for the fraction of W jets in the tails, agree within 0.3(0.8)% and are listed as a
1880 systematic uncertainty in Table 7.6.

1881 One additional uncertainty is added. As the W-tagging scale factor is evaluated in a $t\bar{t}$ sample,
1882 the transverse momentum range is rather limited. When the W-jet p_T reaches ~ 400 GeV, the
1883 AK8 jet becomes a fully merged top jet with a mass of 170 GeV and a scale factor measurement
1884 becomes impossible. However, the jets used in the analyses presented in this thesis have very high
1885 transverse momenta, up to 2-3 TeV, and we therefore need an estimate of how the uncertainty
1886 on the W-tagging scale factor changes as a function of p_T . This is estimated by comparing
1887 the difference in tagging efficiency between $G_{\text{bulk}} \rightarrow WW$ signal MC showered by PYTHIA8 as
1888 compared to HERWIG++ as a function of jet p_T , relative to the difference in tagging efficiency
1889 between the two at a $p_T \sim 200$ GeV. This measurement was performed by a separate analysis
1890 team, and found to be $5.90\% \times \ln(p_T/200 \text{ GeV})$. Systematic uncertainties from other sources
1891 (lepton identification, b tagging, etc.) are less than 0.5% and therefore negligible.

1892 7.7.4 Fit results

1893 The simultaneous fit as described above is then performed both for data and for simulation, where
1894 we take the ratio of data and MC efficiencies as efficiency scale factors. The corresponding fits are
1895 shown in Figure 8.17, with the corresponding extracted efficiencies and scale factors summarized
 in Table 7.6. We additionally extract the jet mass scale and jet mass resolution, used to scale

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

Table 7.6: Efficiencies in data and in MC together with the corresponding W-tagging scale factors for the high purity and low purity categories.

1896
1897 and smear the jet mass signal shape in the limit setting procedure. These values are taken from
1898 the mean $\langle m \rangle$ and width σ of the Gaussian component of the simultaneous fit in the pass region
1899 and are summarized in Table 7.7. Both the jet mass scale as well as the jet mass resolution is
1900 larger in simulation than in data with a relative difference of 2 and 10%, respectively. However,
1901 the jet mass resolution scale factor has a large uncertainty attached to it and is statistically
1902 insignificant (in agreement with unity within uncertainty).

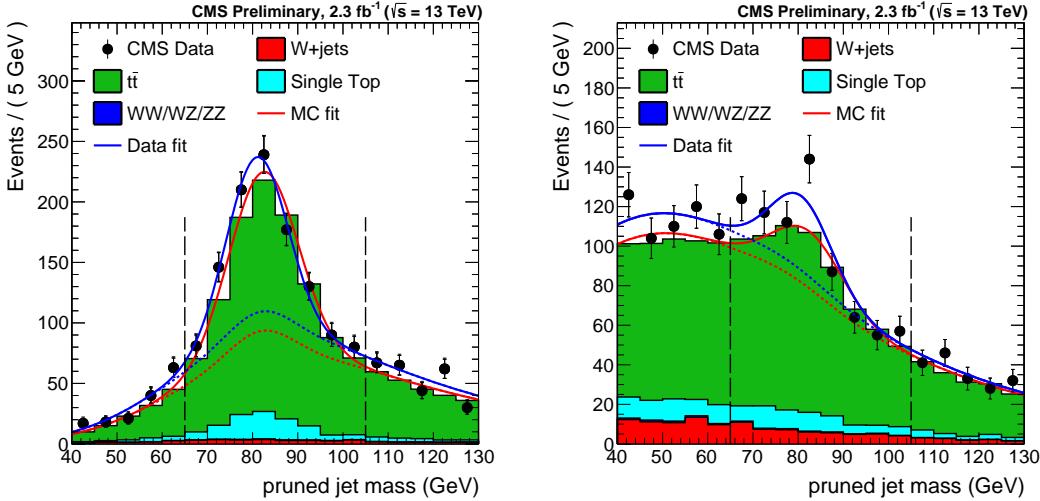


Figure 7.28: Pruned jet mass distribution that pass (left) and fail (right) the $\tau_{21} < 0.45$ selection. Results of both the fit to data (blue) and simulation (red) are shown. The background components of the fit are shown as short-dashed lines.

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

Table 7.7: Jet mass scale and resolution in data and in simulation together with the relevant data-simulation scale factors.

7.7.5 Impact on search variables

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

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

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

¹⁹⁰⁴ The uncertainties on the scale factors are considered as anti-correlated between the HP and the LP categories. The jet mass scale and resolution are used to scale and smear the signal Monte Carlo. An uncertainty on the signal yield based on the uncertainty on jet mass scale and

1907 resolution is also considered by scaling and smearing the jet mass up and down within the quoted
1908 uncertainties and then recomputing the signal efficiency. The results are listed in Table 7.8.

1909 7.8 Systematic uncertainties

1910 The uncertainty on the background parametrization is statistical only and is taken as the
1911 covariance matrix of the dijet fit function. As demonstrated in the F-test, we study different
1912 background parameterizations and we have found these to be within the fit uncertainty of the
1913 nominal fit. The remaining uncertainties concern the signal shape and yield and are listed in
1914 Table 7.8. Jet reconstruction uncertainties affect both the signal yield and the signal shape. These
1915 are evaluated by rescaling the jet four-momenta according to uncertainties on the jet energy scale
1916 and resolution and recomputing the signal efficiency. The difference in efficiency with and without
1917 smearing/scaling is taken as a systematic uncertainty, as described above. The jet mass/energy
1918 scale and resolution also affect the signal shape, and are added as uncertainties in the peak
 position and width of the Gaussian component of the signal PDFs.

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

Table 7.8: Summary of the systematic uncertainties and the quantities they affect. Migration uncertainties result in events switching between the purity/mass categories and change the efficiency in each category, but do not affect the total signal efficiency.

1920 7.9 Results

1921 The background fits for each analysis category in the data signal region are shown in Figure
 1922 7.29. Here a background-only fit is performed while, as described above, a simultaneous fit is
 1923 used for the limit-setting procedure. The filled area corresponds to the 1σ error band of the
 1924 background fit, obtained using linear error propagation. No excess is observed, and we proceed
 1925 by setting upper limits on the cross section times branching ratio of the process $X \rightarrow VV$, using
 1926 the asymptotic CL_S method [72]. The binned likelihood is defined as

$$L = \prod_i \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}, \quad (7.7)$$

1927 with

$$\mu_i = \sigma \cdot N_i(S) + N_i(B). \quad (7.8)$$

1928 Here σ is the signal strength scaling the expected number of signal events in the i -th dijet invariant-
 1929 mass bin $N_i(S)$, $N_i(B)$ is the expected number of background events in dijet invariant-mass bin i
 1930 and n_i is the observed number of events in the i th dijet invariant-mass bin. The background per
 1931 bin $N_i(B)$ is estimated from the background component of the best signal and background fit to
 1932 the data points, with the signal cross section set to zero. The number of signal events in the i -th
 1933 dijet invariant mass bin, $N_i(S)$, is estimated from the signal templates, considering only dijet
 1934 invariant mass in a 20% window around the resonance mass, containing most of the signal
 1935 contribution while making sure to keep a good description of the core.

1936 7.10 Limits

1937 7.10.1 All-hadronic analysis only

1938 As mentioned in Section 7.3, we set limits on three different signal scenarios: $G_{\text{bulk}} \rightarrow WW$,
 1939 $G_{\text{bulk}} \rightarrow ZZ$, and $W' \rightarrow WZ$, with a $\tilde{k} = 0.5$ for the G_{bulk} . Figure 7.30 shows the asymptotic
 1940 limits at 95 % confidence level on the signal cross section as a function of its resonance mass
 1941 obtained with 2.7 fb^{-1} of 13 TeV CMS data after combining all mass and purity categories
 1942 (top). The corresponding p-values are shown in the bottom panel. The statistics are too low to
 1943 exclude the excess around 2 TeV observed in the corresponding Run 1 analysis and, in addition,
 1944 an under-fluctuation in data is present in this region. The largest excess of 2.8σ is observed for a
 1945 $G_{\text{bulk}} \rightarrow ZZ$ hypothesis at a resonance mass of 2.8-3 TeV. This is driven by the high-purity ZZ

mass category, the category with the lowest statistics, and is caused by one event. A 3-parameter fit is the default background fit function for this category, however, a 2-parameter fit could also be used to describe these data. In Figure 7.31 we compare the limits and p-values obtained using a 2-parameter and a 3-parameter fit to describe the background in this category. The significance at 3 TeV is reduced from 2.8 to 1.5 σ with a 2-parameter fit, reflecting the fact that the fit is poorly constrained in the high mass tail due to low statistics. The fit to data using both a 2 and 3-parameter fit in the ZZ high-purity category is shown in Figure 7.32, where we see that the 2-parameter fit lies within the fit uncertainties of the nominal fit. The lack of constraint on the fit in the dijet invariant mass tail when statistics are very low is a drawback of a method relying fully on a parametric fit and reduces the analysis sensitivity in the high- m_{jj} region. In Search II (Section 8) we will keep taking advantage of the dijet fit, however, the integrated luminosity is ~ 15 times higher, resulting in more datapoints in the m_{jj} tail which further constrains the fit. In Search III (Section 9), we will explore alternate methods which allow more control over the background shape across the full mass spectrum.

7.10.2 Semi-leptonic and all-hadronic combination

To maximize the search sensitivity, we combine the results obtained above with those of the corresponding semi-leptonic analysis. We assume the uncertainties on luminosity, V-tagging efficiency, and jet mass scale and resolution to be fully correlated. The obtained exclusion limits in Figure 7.33 show the resulting expected and observed exclusion limits after combining the two analyses. As before, we consider a scenario where only either a W' or Z' resonance is expected, called the singlet hypothesis (upper two plots). In addition, we set limits on the triplet hypothesis, assuming the W' and Z' bosons to be degenerate in mass (bottom left plot). The all-hadronic analysis sets stronger upper limits than the semi-leptonic analysis above 1.7 TeV for a Z' and > 1.3 TeV for a W' . The improvement in sensitivity is due to the higher branching fraction, such that $\mathcal{B}(WW \rightarrow q\bar{q}q\bar{q}) = 44\%$, $\mathcal{B}(WW \rightarrow \ell\nu q\bar{q}) = 31\%$, $\mathcal{B}(WZ \rightarrow q\bar{q}q\bar{q}) = 46\%$, and $\mathcal{B}(WZ \rightarrow \ell\nu q\bar{q}) = 16\%$. The analysis sensitivity for G_{bulk} is too weak to set limits, but cross sections between 3–1200 fb are excluded. For the HVT models A and B, a W' is excluded below 2.0 and 2.2 TeV, respectively, and Z' resonances are excluded below 1.7 (1.6) TeV assuming a HVT model A(B). Assuming a HVT Model A(B) triplet hypothesis, resonances below 2.3 (2.4) TeV are excluded. The combined results would therefore just exclude a W' with a mass around 2 TeV, the favored candidate to explain the diboson excess observed in 8 TeV data. However, Bulk Graviton signals were still far from excluded and, with the expected increase in luminosity of a factor of ten in 2016, we were excited to keep on searching.

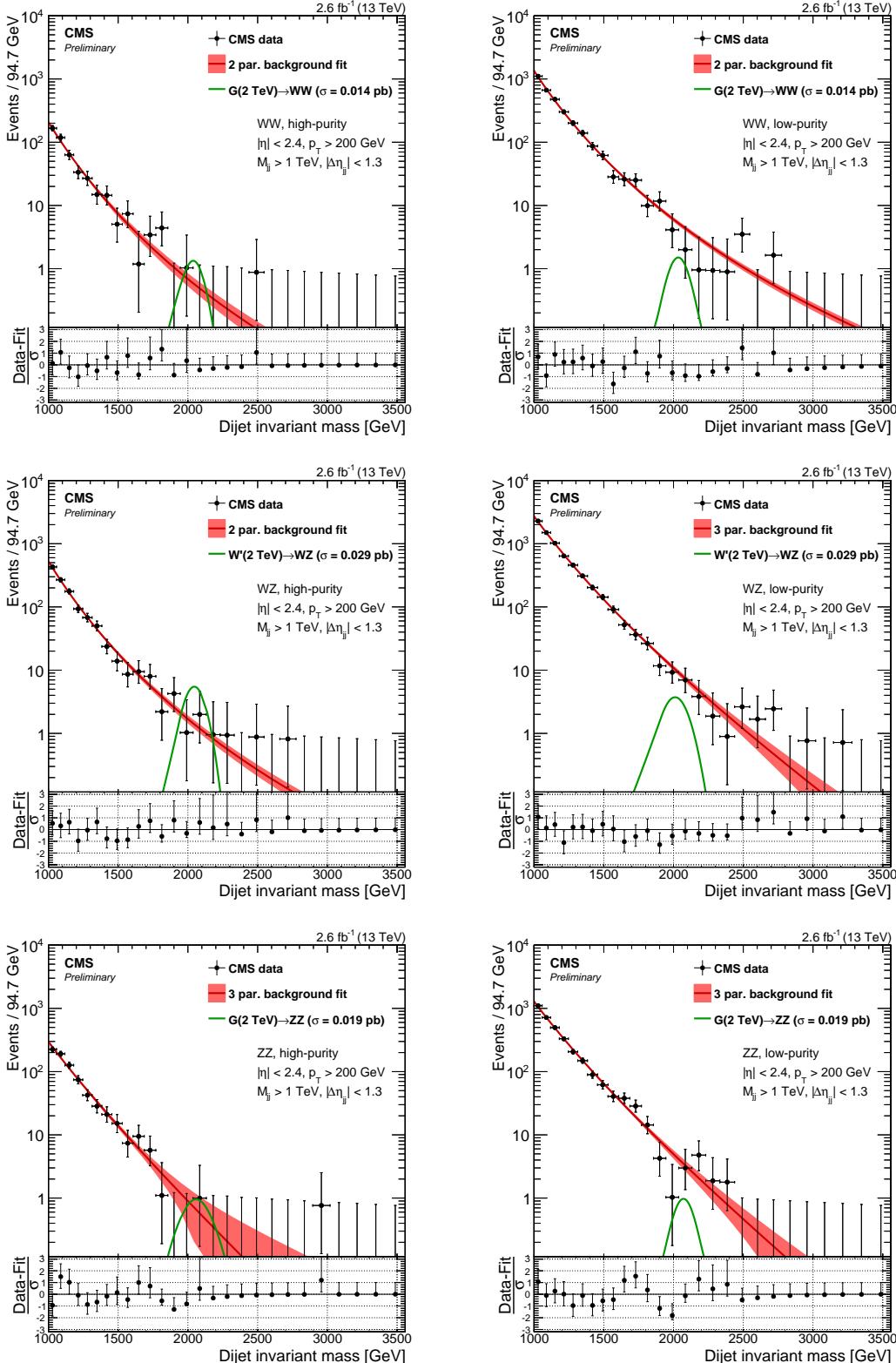


Figure 7.29: Fit to data in the signal region using the background fit only for the different mass and purity categories. The filled red area correspond to the 1σ statistical error of the fit.

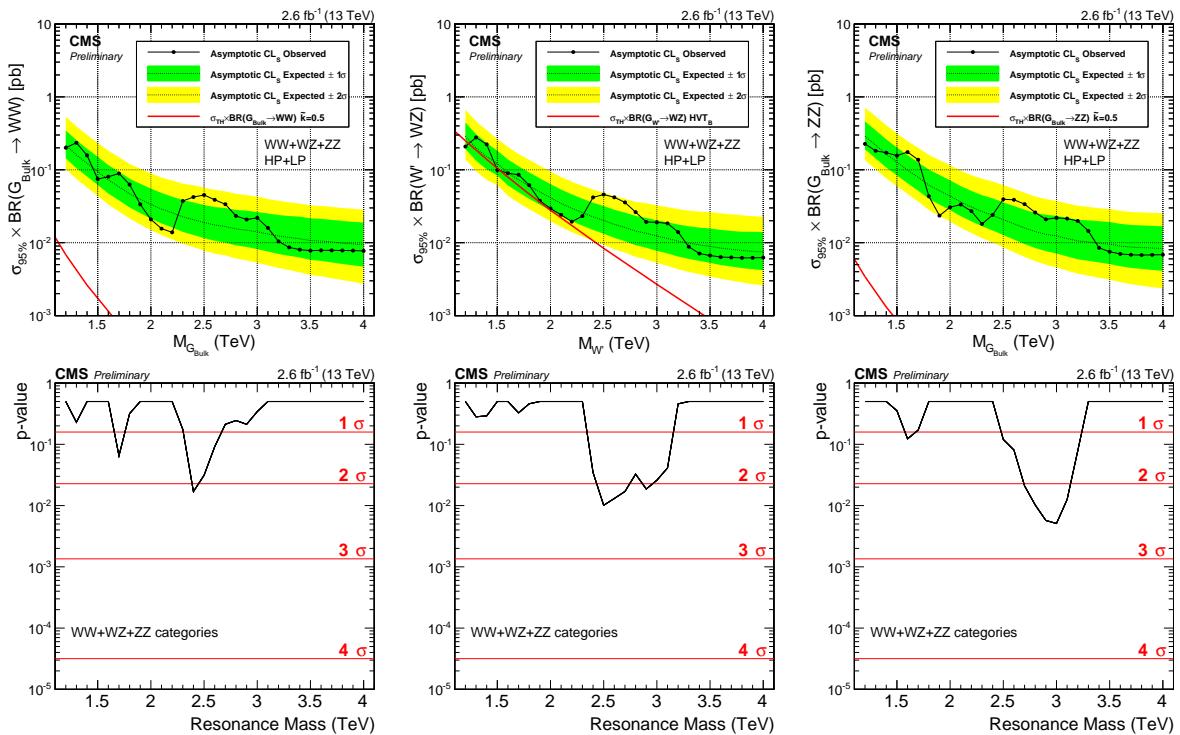


Figure 7.30: Expected and observed limits with corresponding p-values obtained using 2.7 fb^{-1} of CMS data after combining all mass and purity categories. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle), and $G \rightarrow ZZ$ (right) signal.

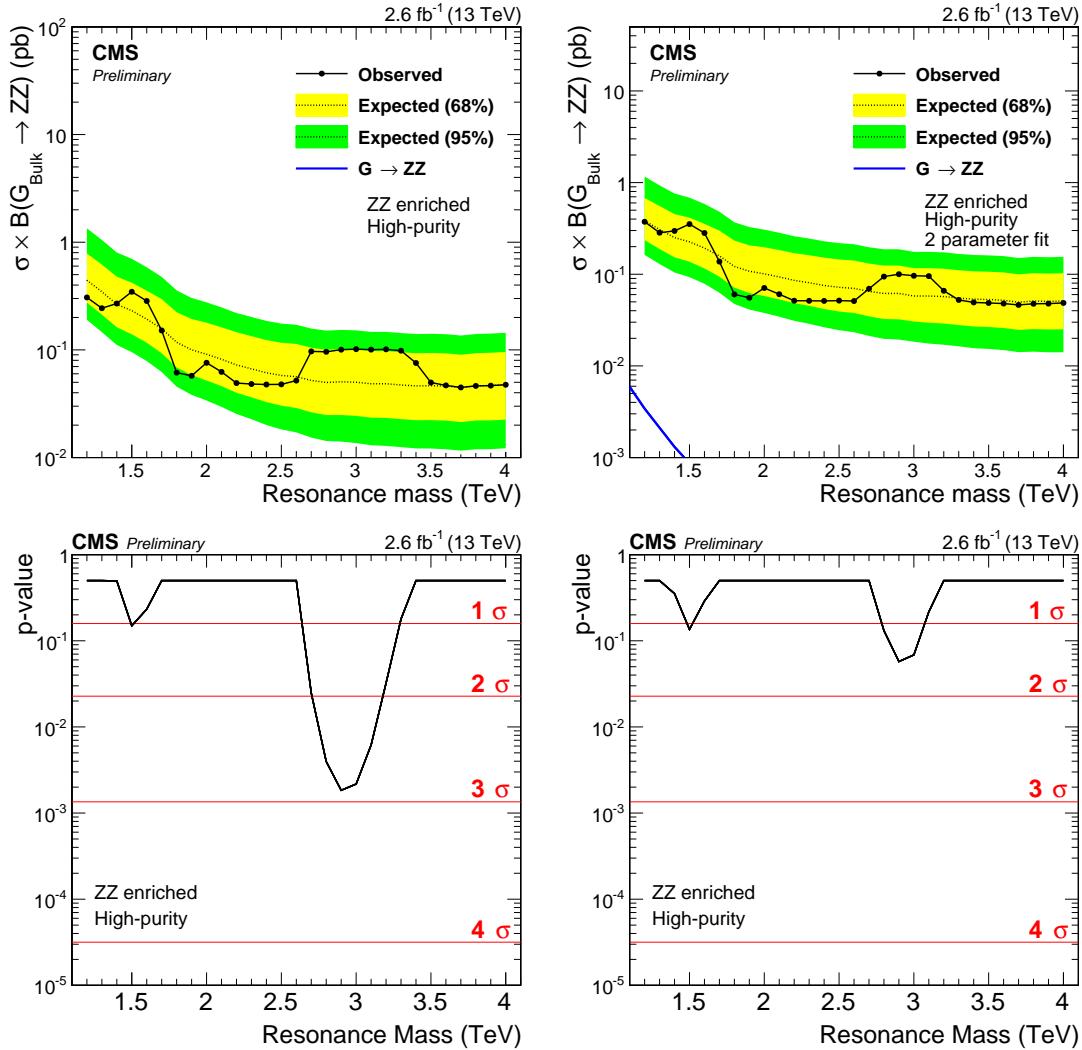


Figure 7.31: Expected and observed limits (top) and the corresponding p-values (bottom) obtained in the ZZ high-purity category using a 3- (left) and 2-parameter (right) fit to describe the background. The significance at 3 TeV is reduced from 2.8 to 1.5 σ .

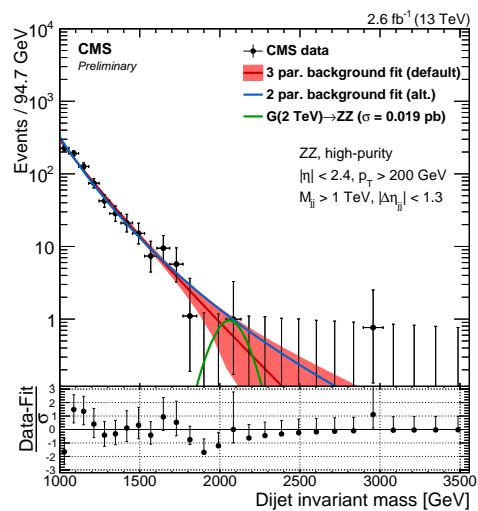


Figure 7.32: Background fit to data in the ZZHP category using the nominal 3- (red) and an alternate 2-parameter (blue) fit to describe the background.

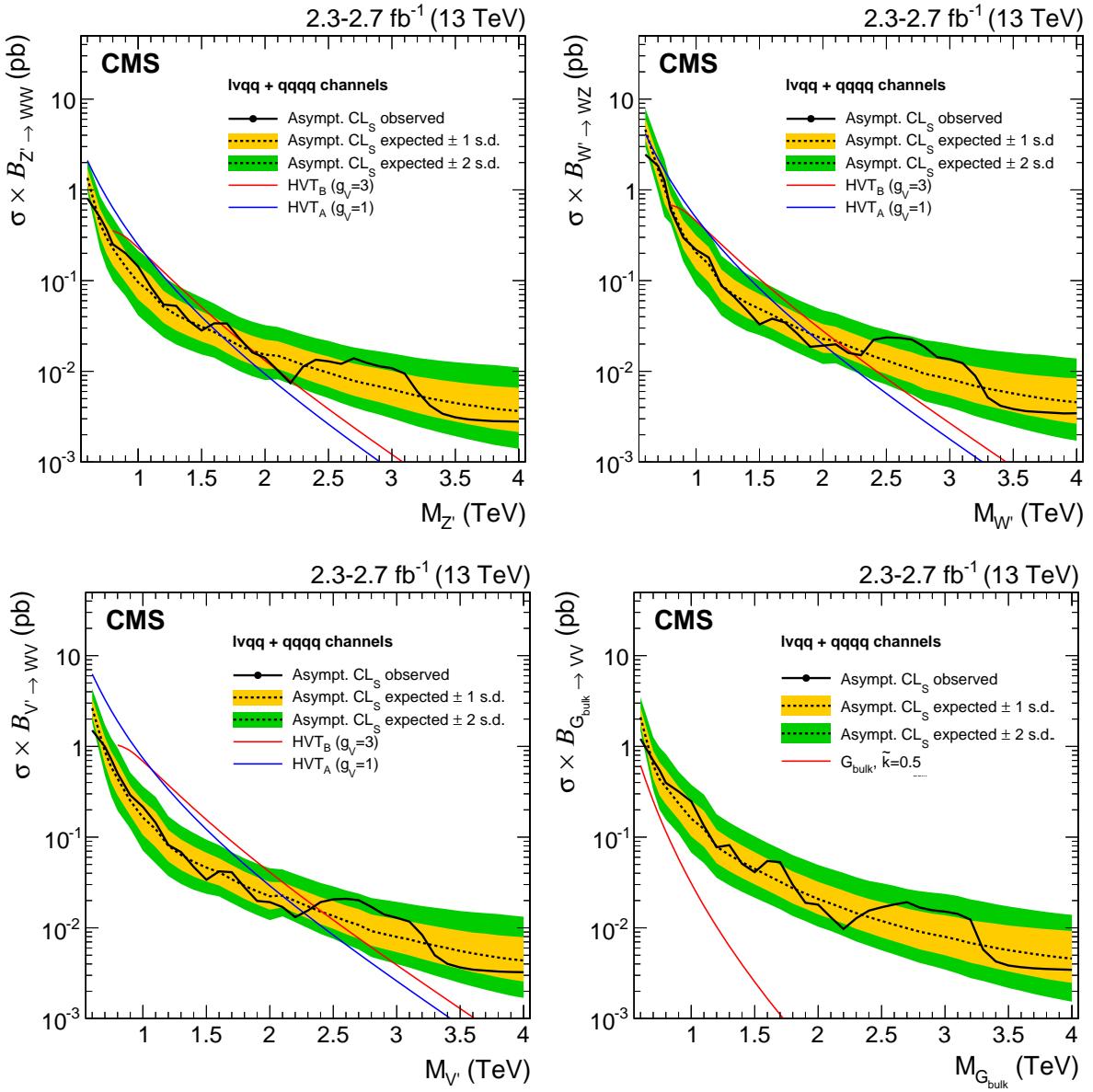


Figure 7.33: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of a narrow-width resonance decaying to a pair of vector bosons for different signal hypotheses. In the upper plots, limits are set in the context of a spin-1 neutral Z' (left) and charged W' (right) resonance, and compared with the prediction of the HVT Models A and B. In the lower left plot, limits are set in the same model under the triplet hypothesis (W' and Z' of degenerate mass). In the lower right plot, limits are set in the context of a bulk graviton with $\tilde{k} = 0.5$ and compared with the prediction.

1979

CHAPTER 8

1980

Search II: A new pileup resistant and perturbative safe tagger

1981

1982 *With the first diboson resonance search with a center-of-mass energy of 13 TeV published, we*
1983 *concluded that more data would be needed in order to fully exclude the excess observed in Run 1.*
1984 *Data from 2016 was right around the corner and, with the LHC planning to reduce β^* from 80 to*
1985 *40 cm, the machine was expected to deliver an instantaneous luminosity three times that of the*
1986 *peak luminosity in 2015. This higher instantaneous luminosity, however, meant double the pileup.*

1987

1988 *We knew that a novel pileup-subtraction algorithm had been developed, which provided far better*
1989 *pileup and underlying event rejection than the current default algorithm that used charged hadron*
1990 *subtraction. We also knew that there had been made progress on the theory side in the development*
1991 *of a groomer which was completely insensitive to correlated soft emissions (non-global logarithms),*
1992 *allowing jet grooming to be accomplished in a theoretically calculable way, softdrop with $\beta = 0$*
1993 *($mMDT$).*

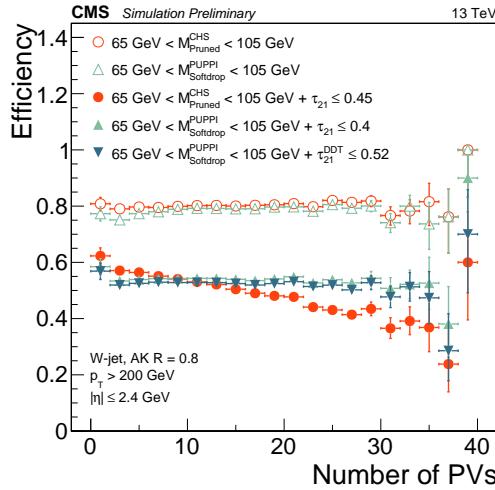
1994

1995 *With more time at hand than in 2015, I therefore decided to pursue a novel W-tagger for*
1996 *this second search. This meant studying and optimizing new approaches for pileup rejection and*
1997 *grooming, developing dedicated jet-mass corrections, as well as validating this new tagger in MC*
1998 *and data. This tagger, together with the mass corrections, became the default and recommended*
1999 *W-tagging algorithm of CMS.*

2000

2001 *Search II was the first published analysis to use a novel combination of PUPPI and softdrop*
2002 *grooming algorithms, now default for W-tagging in CMS. Through this search, the tagger was*
2003 *optimized, commissioned and validated, making it available for several analysis to come. In*

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



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

2008 8.1 Towards robust boosted jet tagging

2009 When we first studied W tagging with 13 TeV data, in the context of the analysis conducted
 2010 with the 2015 dataset, Section 7.4.3, two interesting correlations were observed:

- 2011 1) a strong dependence of the jet mass on the jet p_T , when using the AK8 jet clustering algorithm
 2012 with CHS pileup removal and softdrop with $\beta = 0$ (mMDT) grooming, and
- 2013 2) a strong dependence of the τ_{21} selection efficiency on pileup, when using the AK8 jet clustering
 2014 algorithm with CHS pileup removal.

2015 The reason we studied the softdrop algorithm with $\beta = 0$, hereafter just referred to as “softdrop”,
 2016 as an alternative to the pruning algorithm used in the 2015 analysis was that, besides the
 2017 possibility that it might perform better, we knew it had certain favorable qualities compared to
 2018 other groomers; the softdrop algorithm, in addition to removing sensitivity to the soft divergences
 2019 of QCD as all groomers do (they are IRC safe, see Section 6.4.1), eliminates all correlated soft
 2020 emissions in the jet which are wider than the dominant two-prong substructure, leading to no
 2021 non-global logarithmic terms (NGLs) in the jet mass [44]. NGLs arise from configurations where,
 for instance, a soft gluon is radiated into the jet cone, as illustrated in Figure 8.1. The benefit of

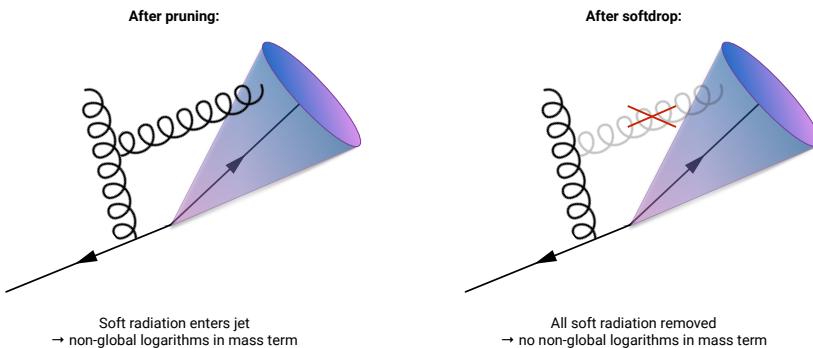


Figure 8.1: The pruning algorithm does not remove all soft emission and therefore has non-global logarithmic terms in the jet mass. Softdrop with $\beta = 0$ removes all soft emissions wider than the dominant two-prong substructure and is therefore free of non-global logarithms.

2022 being NGL-free, is that one can calculate the softdrop jet mass to a significantly higher precision
 2023 than what is possible for other grooming algorithms or for the plain jet mass (NGLs are the main
 2024 reason a full resummation of the plain jet mass beyond NLL accuracy does not exist). There
 2025

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

However, as was discussed in Section 7.4.3, the softdrop jet-mass of signal jets displayed a large dependence on jet p_T , both for reconstructed and generator-level jets. One hypothesis was that this was due to pileup and that perhaps softdrop was more sensitive to contribution from additional interaction vertices than pruning. A study in Ref. [73] had also shown mMDT to be more sensitive to the underlying event, but using a different asymmetry criteria than softdrop (see Appendix A.1).

In parallel to the ongoing theoretical work on groomers, a novel pileup removal algorithm had been proposed: Pileup per particle identification (PUPPI) [35]. Described in detail in Section 6.3.2, PUPPI considers not only charged pileup, but also reweights neutral particles in the jet with its probability of arising from pileup. PUPPI had proven itself far superior to the current CHS algorithm in terms of jet observables for large radius jets, and therefore seemed like a good candidate to address both issues listed above: the softdrop p_T dependence and the strong pileup dependence of τ_{21} . The focus of Search II would therefore be on the commissioning of a novel W-tagger. There are interesting additions to the analysis strategy as well: the inclusion of a $Z' \rightarrow WW$ signal hypothesis and the addition of a completely new analysis, the single V-tag analysis.

8.2 Analysis strategy

The analysis strategy for this search is conceptually the same as for Search I. In addition, we take advantage of the n-subjettiness categorization and do an additional analysis in parallel: a search for excited quark resonances q^* [74, 75] decaying to qW or qZ . We call this the single V-tag analysis, and the analysis selection only differs in that one jet is not required to pass the V-tag selection (groomed mass and n-subjettiness). The VV analysis is hereby referred to as the double V-tag analysis. The difference between the two analyses is illustrated in Figure 8.2. In addition, limits are set on a $Z' \rightarrow WW$ signal hypothesis in the double V-tag analysis, another first with data collected at 13 TeV center-of-mass energy. This analysis was published in two steps: an early CMS physics analysis summary document (PAS) based on 12.9 fb^{-1} of 2016 data [71], describing the new PUPPI softdrop based V tagger as well as the single V-tag analysis, and a second analysis considering the full 2016 dataset [76]. The commissioning of the new V tagger has also been documented in a jet performance PAS [77]. In this chapter, the main emphasis will be on the work presented in CMS-PAS-B2G-16-021 [71] since this was where the new V tagger and single V-tag analysis was first presented. The second part of the results chapter, Section 8.10.2,

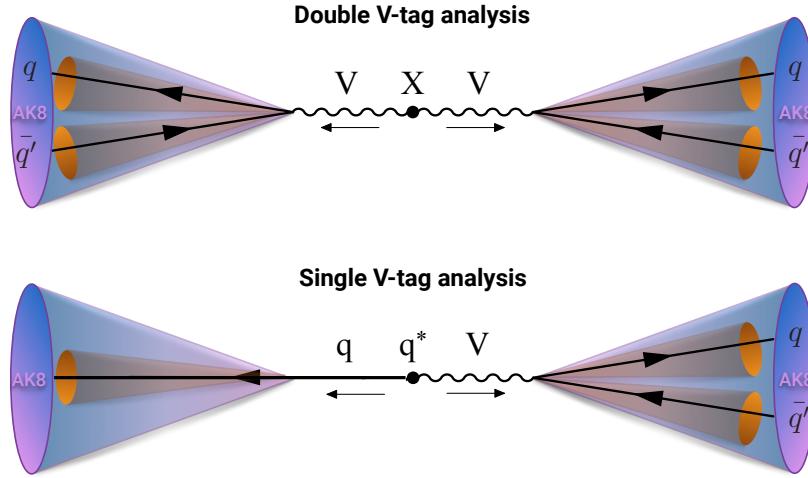


Figure 8.2: The double (top) and single (bottom) W/Z-tag analysis.

2059 includes the results obtained using the full 2016 dataset of 35.9 fb^{-1} .

2060 8.3 Data and simulated samples

2061 As mentioned above, the analysis of the 2016 dataset was done in two steps corresponding
 2062 to two different datasets: one analysis based on 12.9 fb^{-1} of early 2016 data, describing the
 2063 new W-tagger and single V-tag category, and a second paper considering the full 2016 dataset,
 2064 corresponding to 35.9 fb^{-1} . Both were collected at a center-of-mass energy of 13 TeV.

2065 The G_{bulk} and HVT signal samples are modeled in the same way as in 2015. The q^* samples
 2066 are simulated assuming unpolarized bosons and a compositeness scale Λ set equal to the resonance
 2067 mass. These are generated at leading order using PYTHIA version 8.212 [63].

2068 The standard model background processes of QCD, W+jets, and Z+jets are all simulated to
 2069 leading order. The W and Z+jets backgrounds are simulated with MADGRAPH5_AMC@NLO [62, 78].
 2070 Three different combinations of matrix element and shower generators are used for QCD as these
 2071 predictions are known to differ: PYTHIA-only, the leading order mode of MADGRAPH5_AMC@NLO
 2072 matched with PYTHIA, and HERWIG++ 2.7.1 [50] with the CUETHS1 tune [57].

2073 8.4 Event selection

2074 8.4.1 Triggering

2075 The triggers used in this analysis are the same ones as in 2015 (see Section 7.4.1), however, due
 2076 to the new single V-tag analysis, the trigger turn-ons have been reevaluated separately with the
 2077 requirement that either one or two jets have an offline softdrop jet mass above 65 GeV. Figure 8.3
 2078 shows the trigger turn-on curves as a function of dijet invariant mass for jets passing one of the
 2079 H_T -based triggers only (pink markers), one of the grooming triggers only (green markers), and
 2080 when combining all of them (purple markers). The turn-on curves are shown for all jet pairs
 2081 passing loose selections as described in Section 7.4.2 (top), and for jet pairs where one (bottom
 2082 left) or both (bottom right) of the jets additionally are required to have a softdrop mass larger
 2083 than 65 GeV. Including the grooming triggers lowers the 99% trigger efficiency threshold by
 2084 around 50 (80) GeV in the single-tag (double-tag) category once substructure is required on the
 2085 analysis level. Using a combination of either of the triggers, we are safely on the trigger plateau
 2086 for dijet invariant masses above 955 (986) GeV in the double (single) tag category, so that the
 2087 dijet invariant-mass threshold can be set at 955 GeV for the double-tag analysis and 990 GeV for
 2088 the single-tag analysis. For control plots, where no groomed-mass window is applied, a trigger
 2089 threshold of 1020 GeV is used. Figure 8.4 shows the trigger efficiency as a function of the offline
 2090 PUPPI softdrop-jet mass (left) and pruned-jet mass (right) for the trigger requiring an online
 2091 trimmed-jet mass of at least 30 GeV. Here, the jet transverse momentum of one of the jets is
 2092 required to be at least 600 GeV and no other mass cut is applied. The trigger turn-on is sharp
 2093 and the plateau is reached for offline groomed-jet masses around 50 GeV.

2094 8.4.2 Preselection

2095 The same preselections as in Search I, described in Section 7.4.2, are used; we require two AK8 jets
 2096 with CHS pileup subtraction applied, and that are required to pass the tight jet-ID requirement,
 2097 $p_T > 200$ GeV and $|\eta| < 2.5$. The same selection requirement that suppresses t-channel QCD
 2098 production is required, $|\Delta\eta| < 1.3$, together with thresholds on the dijet invariant mass of 955
 2099 GeV for the double-tag and 990 GeV for the single-tag analyses. The jet p_T (top left), η (top
 2100 right), $\Delta\eta_{jj}$ (bottom left) and dijet invariant mass (bottom right) for the two leading jets in the
 2101 event after loose preselections are applied is shown in Figure 8.5. A large difference in slope in
 2102 the jet p_T and dijet invariant mass spectrum depending on the QCD matrix element or shower
 2103 generator is observed. Pure PYTHIA QCD MC describes the data best, while HERWIG++ tends to
 2104 underestimate and MADGRAPH5_AMC@NLO+PYTHIA tends to overestimate the number of high
 2105 p_T/m_{jj} jets. Pure PYTHIA8 QCD MC is therefore used for all background checks in this analysis.

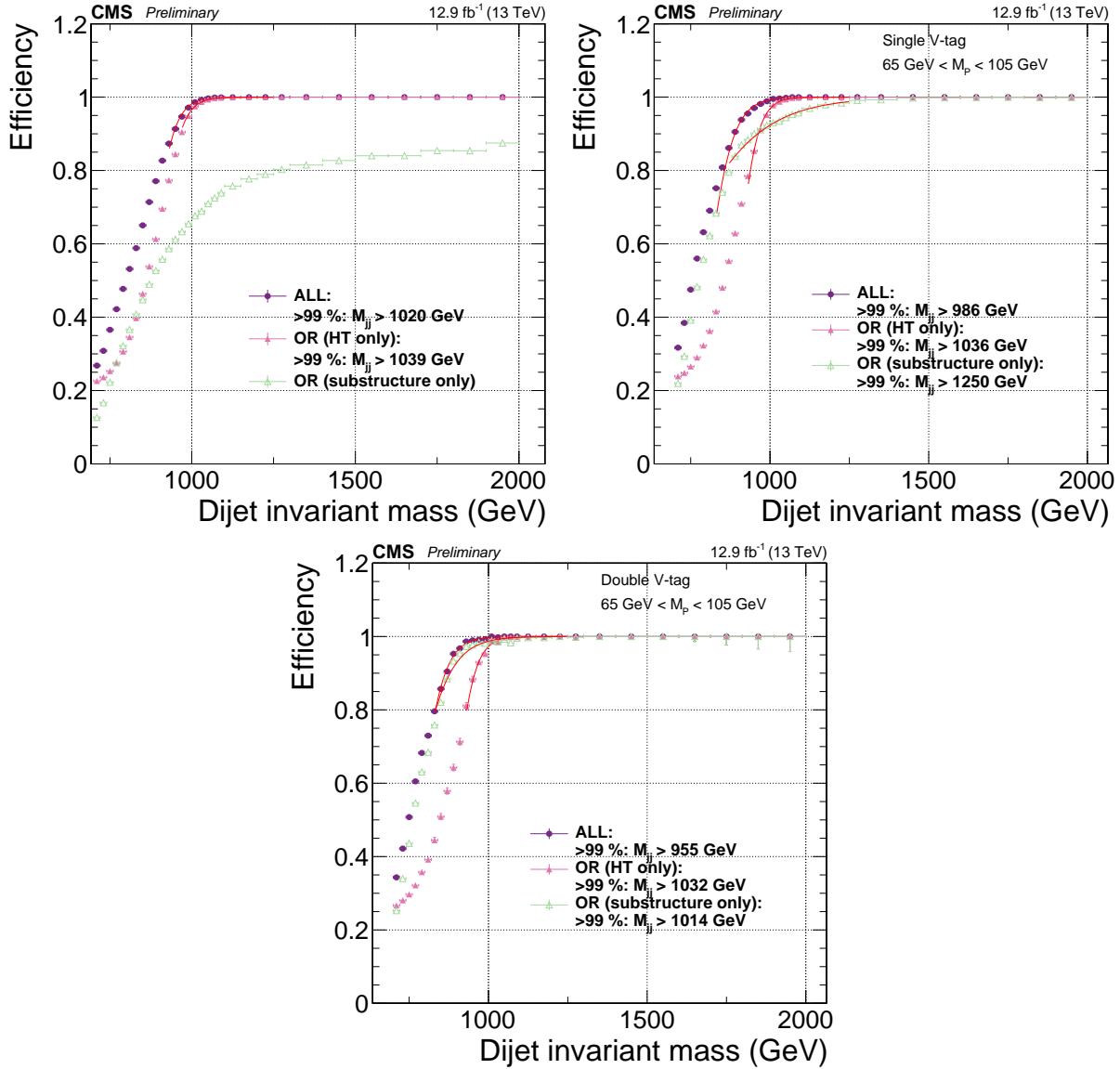


Figure 8.3: Comparison of trigger efficiencies for jets passing one of the HT-triggers only (pink), for jets passing one of the grooming-triggers only (green), and for jets passing one of the HT-triggers or one of the grooming triggers (purple). It is shown here as a function of the dijet invariant mass for all jet pairs passing loose selections (top), where one jet additionally is required to have a softdrop mass larger than 65 GeV (bottom left), and where both jets are required to have a softdrop mass larger than 65 GeV (bottom right).

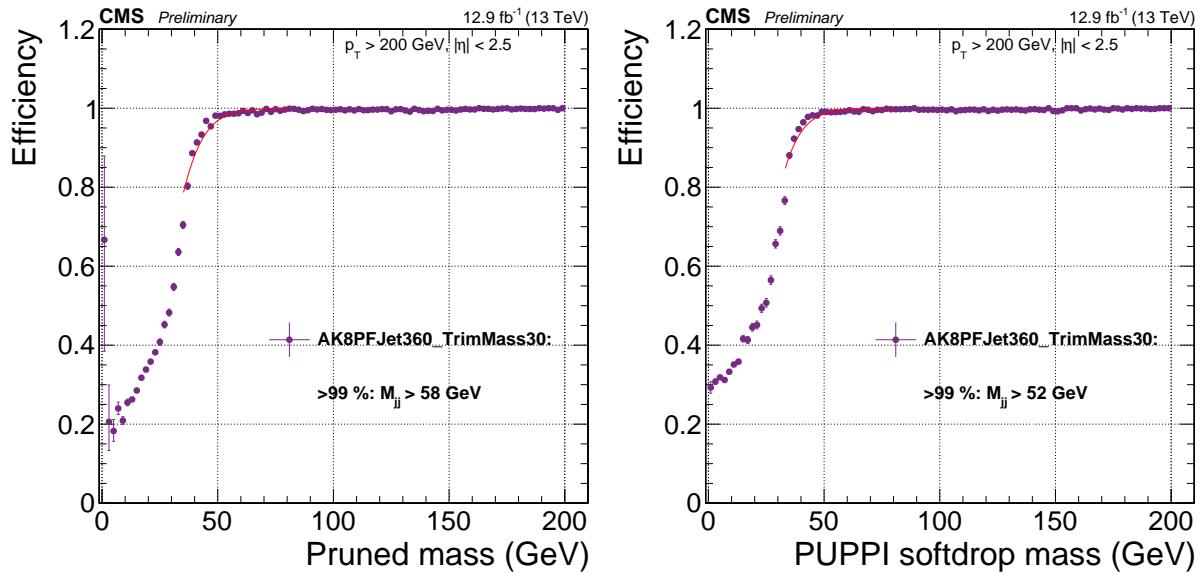


Figure 8.4: Efficiency for the trigger requiring an online trimmed jet mass of at least 30 GeV as a function of the pruned-jet (left) and softdrop-jet (right) mass for jets with $p_T > 600$ GeV.

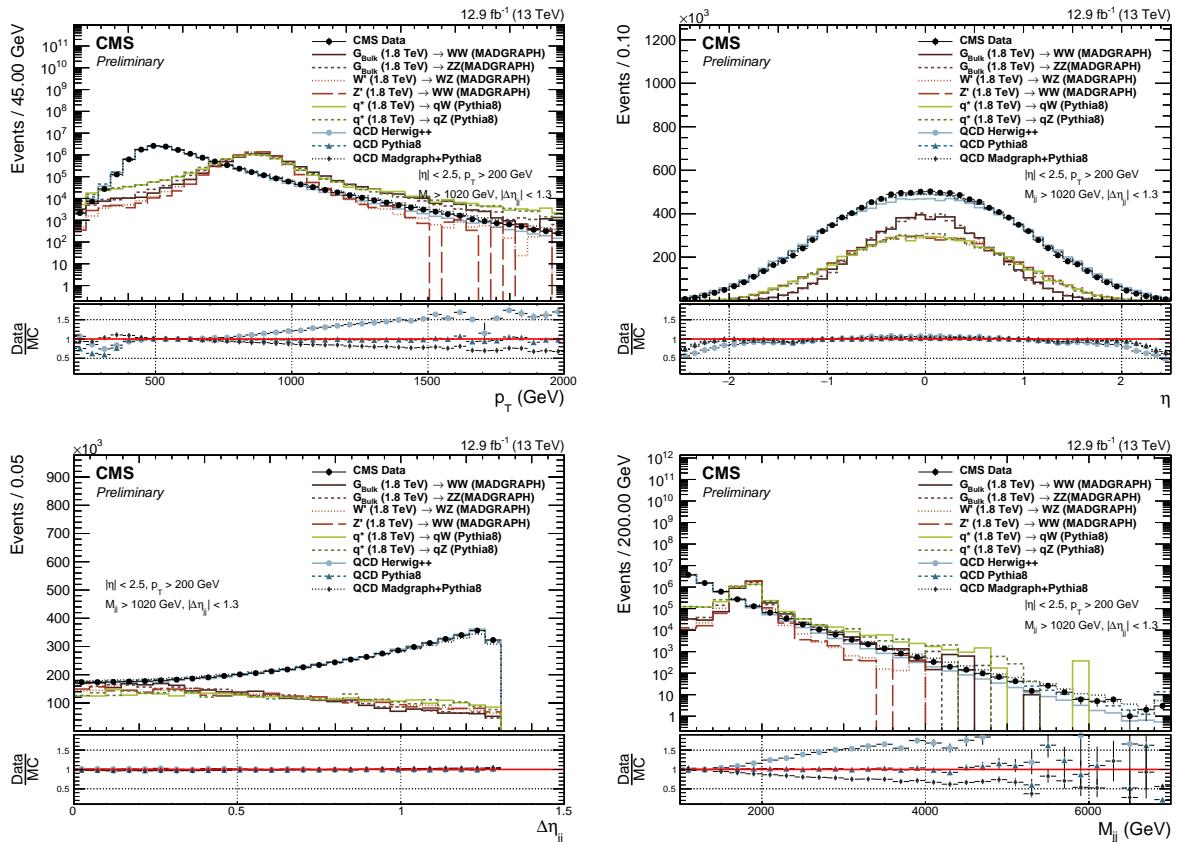


Figure 8.5: Jet p_T (top left), η (top right), $\Delta\eta_{jj}$ (bottom left) and dijet invariant mass (bottom right) for the two leading jets in the event after loose preselections are applied. The signal is scaled by an arbitrary number.

2106 8.5 Developing a new W-tagger

2107 As mentioned in the introduction to this chapter, studies had shown that the PUPPI pileup
 2108 subtraction algorithm yielded superior resolution on large-cone jet observables like the jet mass.
 2109 We therefore wanted to check whether the softdrop jet mass, and its observed sensitivity to the
 2110 underlying event and pileup, would be improved if a better pileup subtraction algorithm was
 2111 applied pre-clustering.

2112 Two interesting observations were made. Softdrop used together with PUPPI pileup subtraction
 2113 displayed a much smaller p_T -dependent shift than the combination of CHS and softdrop.
 2114 Figure 8.6 shows the PUPPI softdrop jet mass for W-jets from a 1 TeV resonance, corresponding
 2115 to a jet $p_T \sim 500$ GeV, and a 4 TeV resonance, corresponding to a jet $p_T \sim 2$ TeV, exhibiting the
 desired reduced p_T dependence on jet mass scale. However, when applying the recommended L2

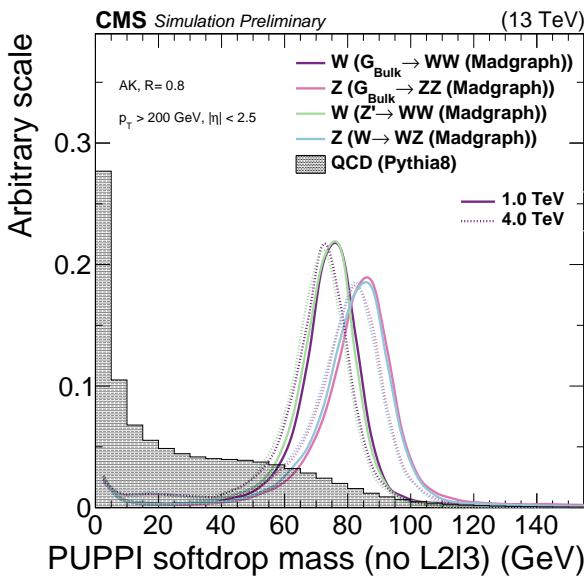


Figure 8.6: The PUPPI softdrop jet mass distribution for W boson jets coming from a 1 and 4 TeV resonance assuming different signal hypothesis. The QCD multijet background is shown in gray. No jet energy corrections have been applied to the jets.

2116 and L3 jet energy corrections (see Section 6.4.3) to the jet groomed mass, a strong p_T dependent
 2117 shift is re-introduced. This effect is not present for the pruned jet mass. Figure 8.7 shows
 2118 the softdrop (top left) and pruned (top right) jet mass distributions with recommended L2L3
 2119 corrections applied. Here, the PUPPI softdrop jet mass shift as a function of p_T is significantly
 2120 increased with respect to what was observed for the uncorrected mass, while the CHS pruned
 2121 jet mass remains stable. This points to the PUPPI jet energy corrections not being optimal for
 2122

PUPPI soft-dropped jets. Indeed, there is no reason to expect the CMS JEC corrections to work for groomed jets at all as they were derived from un-groomed PUPPI or CHS jets. The fact that they work well for CHS pruned jets is not a given. The jet energy corrections derived for CHS and PUPPI jets as a function of jet p_T are shown in the bottom plot in Figure 8.7. A significant slope in JEC as a function of p_T is measured for PUPPI, while not present for CHS.

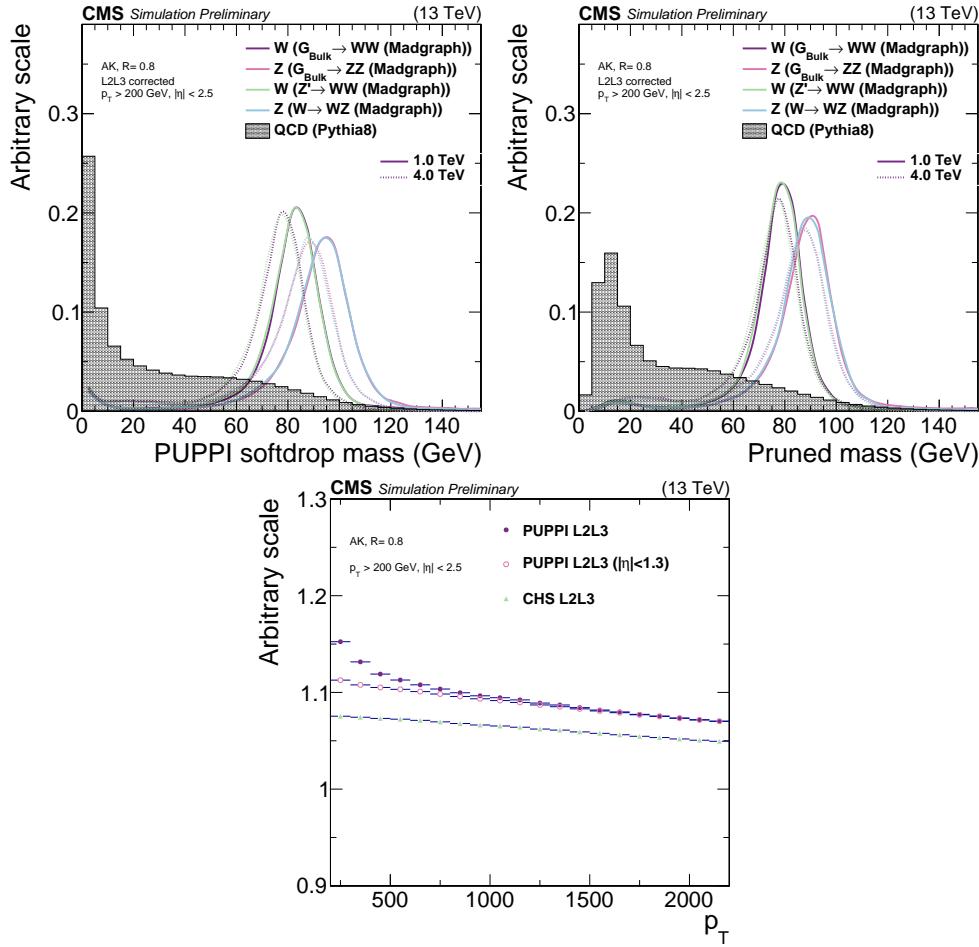


Figure 8.7: Top: PUPPI softdrop jet mass distribution (top left) and pruned jet mass distribution (top right) with L2 and L3 corrections applied. Bottom: The projection of CHS and PUPPI jet energy corrections versus jet p_T .

2127

2128 8.5.1 Dedicated PUPPI softdrop jet mass corrections

In order to remove the PUPPI softdrop jet mass dependence on jet p_T , all jet energy corrections to the softdrop jet mass are removed. However, this still leaves a residual p_T -dependent shift and,

in addition, the uncorrected mass does not peak near the correct W-mass of 80.4 GeV. Figure 8.8 shows the mean of a Gaussian fit to the uncorrected PUPPI softdrop mass as a function of jet p_T in two different η bins (smaller or greater than $|\eta| = 1.3$) for W-jets coming from a Bulk Graviton signal sample. A mass shift both as a function of η and p_T is observed, together with an average mean shifted significantly lower than the W boson mass. In order to use PUPPI softdrop for

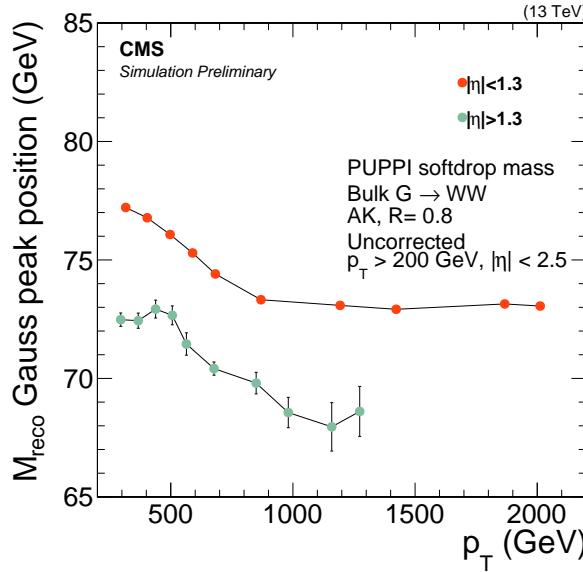


Figure 8.8: The mean of a Gaussian fit to the W-jet PUPPI softdrop mass peak as a function of jet p_T in two different η bins (smaller or greater than $|\eta| = 1.3$). No jet energy corrections have been applied to the softdrop mass.

W tagging, we therefore derive dedicated jet-mass corrections to compensate for two factors: a generator-level p_T -dependence, as first observed in Section 7.4.3, and a reconstruction-level p_T - and η -dependence, most likely due to calibration (recall, the PUPPI jet energy corrections have been removed). Figure 8.9 shows the mean of the generated softdrop jet mass (left) and the normalized difference in reconstructed and generated softdrop jet mass (right) as a function of jet p_T . The shift in generated softdrop mass at lower p_T is of the order of 2-3% while the difference between reconstructed and generated softdrop mass is a 5-10% effect. The mass shift introduced at generator-level is corrected by a fit to $M_{\text{PDG}}/M_{\text{GEN}}$ 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 8.9. To correct for the residual shift between generated and reconstructed softdrop mass, a fit to $(M_{\text{RECO}} - M_{\text{GEN}})/M_{\text{RECO}}$, where M_{RECO} is the reconstructed mass shown in the right plot in Figure 8.9 and M_{GEN} is as defined above, as a function of jet p_T in two η bins (smaller or greater than $|\eta| = 1.3$) is performed. Polynomial fit functions of the following forms

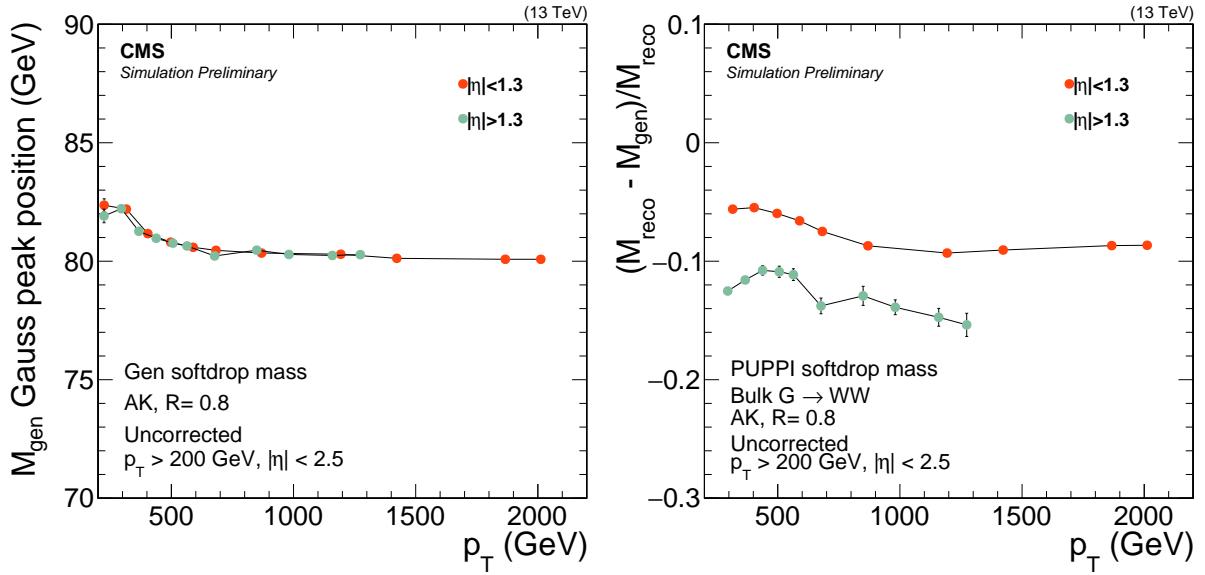


Figure 8.9: The mean of the fitted generator level W-jet softdrop mass distribution as a function of jet p_T (left) and the normalized difference in reconstructed and generated softdrop jet mass (right).

are used:

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

2129 The distribution and corresponding parametrization of the two corrections is shown in Figure 8.10
 2130 for the “gen correction” (left) and “reco correction” (right). The two corrections are then applied
 2131 to the uncorrected PUPPI softdrop mass both in data and in MC as

$$M_{SD} = M_{SD,\text{uncorr}} \times w_{GEN} \times w_{RECO} \quad (8.1)$$

2132 where w_{GEN} and w_{RECO} correspond to the generator and reconstructed level corrections, respec-
 2133 tively, and $M_{SD,\text{uncorr}}$ is the uncorrected PUPPI softdrop mass.

2134 A closure test is performed in order to check that the corrected PUPPI softdrop W-jet mass
 2135 peaks at 80.4 GeV and is stable with respect to the jet p_T and η . The fitted mean of the corrected
 2136 PUPPI softdrop mass peak as a function of jet p_T in two different η bins is shown in Figure 8.11.
 2137 Good closure is observed, with the corrected mass peaking around 80 GeV independent of the jet
 2138 p_T and η . The PUPPI softdrop jet mass peak for W, Z, and Higgs boson jets from different signal
 2139 samples after jet mass corrections have been applied is shown in Figure 8.12, for resonances with

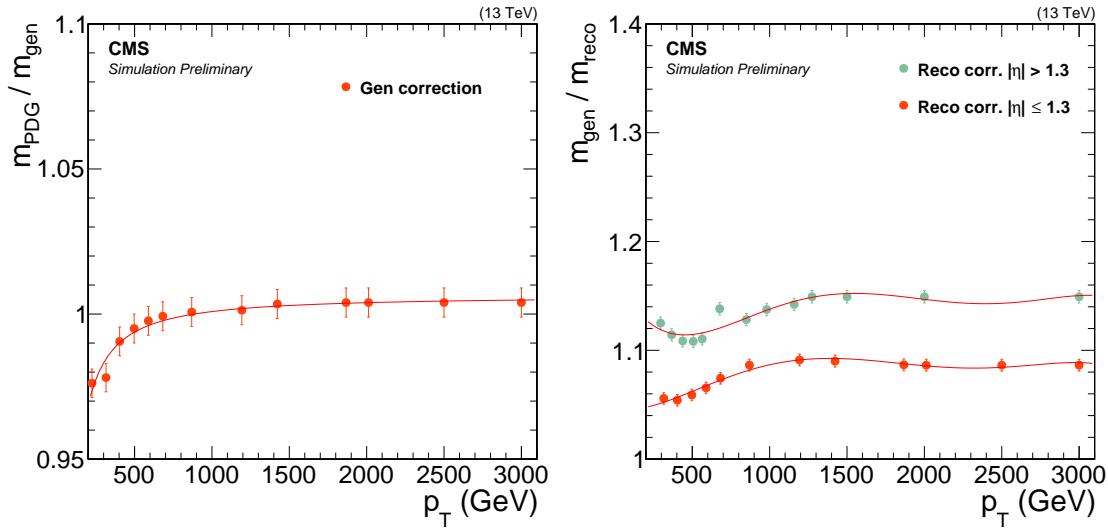


Figure 8.10: Left: fit to $M_{\text{PDG}}/M_{\text{GEN}}$ as a function of jet p_T , where $M_{\text{PDG}} = 80.4$ GeV and M_{GEN} is the fitted mean of the generator level mass. Right: fit to $(M_{\text{RECO}} - M_{\text{GEN}})/M_{\text{RECO}}$, where M_{RECO} is the reconstructed softdrop mass, as a function of jet p_T in two η bins.

2140 a mass of 1 and 4 TeV. The corrections yield a mass stable with p_T , peaking around the boson mass, for all three boson-jets under consideration.

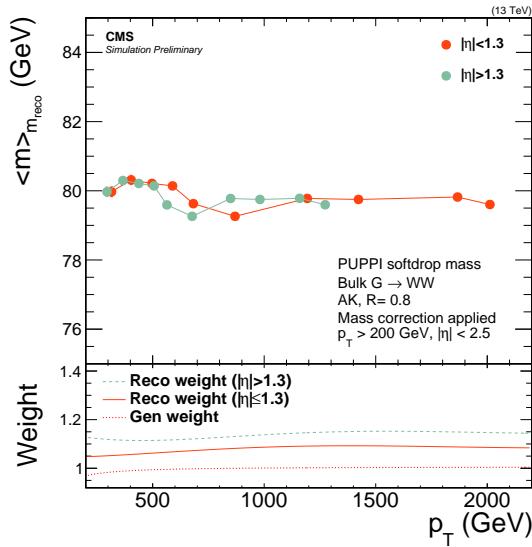


Figure 8.11: The mean of a Gaussian fit to the corrected PUPPI softdrop mass peak for real W-jets as a function of jet p_T in two different η bins.

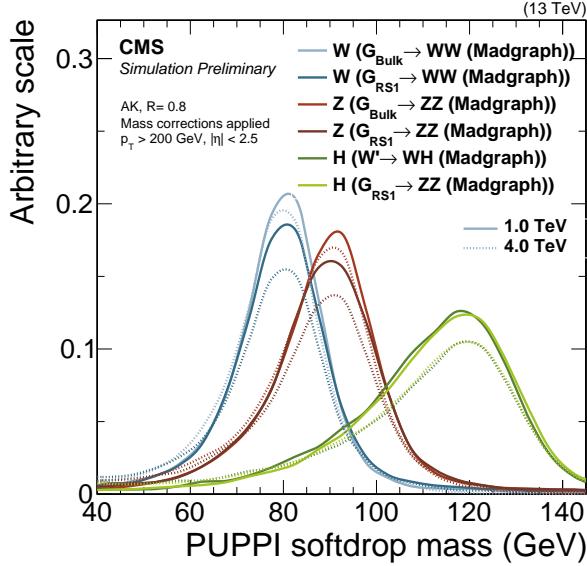


Figure 8.12: The corrected PUPPI softdrop jet mass for W, Z, and H jets from different signal samples with masses of 1 and 4 TeV.

2142 8.5.2 W-tagging performance

2143 The new PUPPI softdrop based V tagger uses a softdrop jet mass window of $65 \text{ GeV} < m_{SD} <$
 2144 105 GeV in combination with a selection of PUPPI $\tau_{21} < 0.4$. The CMS default is to compute the
 2145 τ_{21} variable starting from un-groomed jets, but it can also be computed from already groomed jets.
 2146 I briefly studied calculating τ_{21} starting from soft-dropped jets, summarized in Appendix A.2,
 2147 where I found a small gain in performance when calculating τ_{21} starting from jets groomed with
 2148 softdrop $\beta = 0$. This is something that needs further study, and is left for future work. We
 2149 compare the performance of the PUPPI softdrop based V tagger to that of the CHS pruning-based
 2150 tagger used in Search I, as well as to that of a “DDT-transformed” τ_{21} -based tagger [79]. The
 2151 τ_{21}^{DDT} variable is a linear transformation of τ_{21} given as

$$\tau_{21}^{DDT} = \tau_{21} + M \times \log \left(\frac{m^2}{p_T \times 1 \text{ GeV}} \right) \quad (8.2)$$

2152 where $M = -0.063$ is obtained from a fit of τ_{21} against the variable $\rho' = \log(m^2/p_T/\mu)$ in QCD
 2153 Monte Carlo, where $\mu = 1 \text{ GeV}$. The purpose of this is to decorrelate τ_{21} from the QCD jet
 2154 softdrop mass and p_T , yielding a mass and dijet invariant mass spectrum minimally sculpted by
 2155 a cut on the τ_{21}^{DDT} tagging variable. This tagger will be further explored, and described in detail,
 2156 in the context of Search III in Section 9.6.

2157 The background rejection efficiency for QCD light-flavor jets as a function of W-jet signal
 2158 efficiency is shown in Figure 8.13. The efficiency is measured after requiring a softdrop jet mass
 selection of $65 \text{ GeV} < m_{SD} < 105 \text{ GeV}$, while scanning the cut on τ_{21} . The general performance

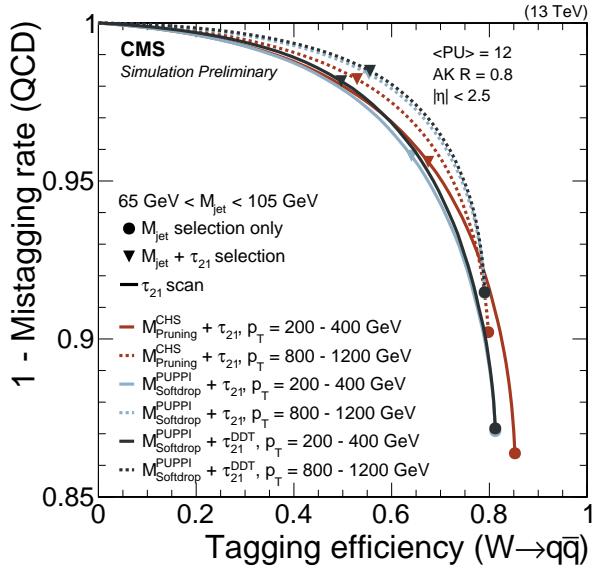


Figure 8.13: The background rejection efficiency for QCD light-flavor jets as a function of W-jet signal efficiency. A cut on CHS pruned or PUPPI softdrop jet mass of $65 < m_{p/sd} < 105 \text{ GeV}$ is applied while scanning the cut on τ_{21} . The cuts corresponding to $\tau_2/\tau_1 < 0.45$ for CHS pruned jets, PUPPI $\tau_{21} < 0.4$ or $\tau_{21}^{DDT} < 0.52$ for PUPPI softdrop jets are indicated with triangles, while the solid circles represent the efficiency and mistag rate for a mass cut only.

2159
 2160 of each tagger is very similar, with the PUPPI softdrop-based taggers displaying a slightly higher
 2161 signal efficiency for a given mistag rate at high p_T and CHS pruning-based taggers performing
 2162 slightly better at low p_T . In order to better understand the difference between each tagger, we
 2163 look at the tagging performance as a function of jet p_T and pileup, as shown in Figure 8.14
 2164 and 8.15. In Figure 8.14 we observe that the signal efficiency for a PUPPI softdrop or a CHS
 2165 pruned jet mass selection is flat and around 80% as a function of jet p_T . The QCD mistagging
 2166 rate, on the other hand, falls off as a function of p_T , with a 1-3% lower mistag rate when using
 2167 PUPPI softdrop jet mass than when using CHS pruned jet mass. Once applying an n-subjettiness
 2168 cut, the signal efficiency as well as the mistag rate for the τ_{21} -based taggers becomes smaller
 2169 as a function of jet p_T , with an average signal efficiency of around 50% for a ~ 2% mistag rate.
 2170 An interesting behavior is observed for the τ_{21}^{DDT} tagger: while the mistagging rate is flat as a
 2171 function of p_T , by construction, the signal efficiency gets higher as the p_T increases, outperforming
 2172 the other taggers above 1 TeV.

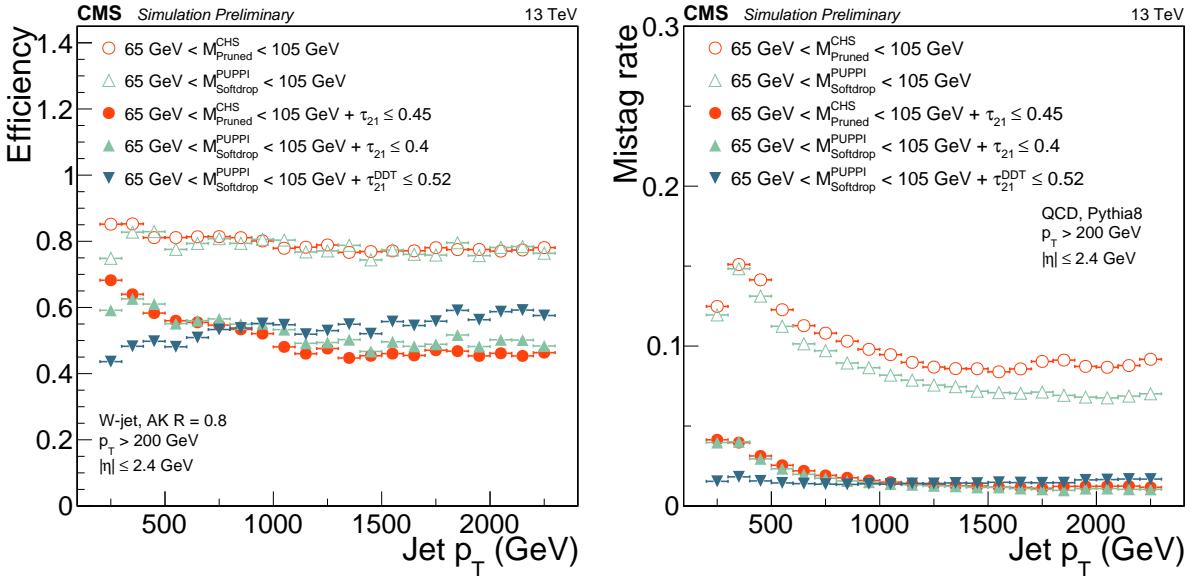


Figure 8.14: W-jet efficiency (left) and QCD light-flavor jet mistag rate (right) for a PUPPI softdrop or CHS pruned jet mass selection only (hollow circles) and the combined $m_{\text{p/sd}}$ + (PUPPI) τ_{21} (DDT) selection (solid circles) as a function of jet p_T .

2173 Figure 8.15 shows the W-jet tagging efficiency (left) and QCD light-flavor jet mistagging rate
2174 (right) as a function of the number of interaction vertices in the event. The tagging efficiency
2175 of the CHS pruned jet tagger (red solid circles) falls off steeply versus the number of primary
2176 vertices in the event, while the PUPPI softdrop taggers (light and dark blue solid circles) are
2177 more or less insensitive to pileup. The per-jet efficiency of the PUPPI softdrop and PUPPI τ_{21}
2178 tagger is around 50-55% for a 1-2% mistag rate. Based on the observed general performance,
2179 tagging stability versus pileup, and due to theoretical considerations, PUPPI softdrop jet mass
2180 with dedicated mass corrections applied, together with PUPPI τ_{21} , is chosen as the W tagger for
2181 this analysis.

2182 8.5.3 Efficiency scale factors, jet mass scale and resolution

2183 In order to measure the W-tagging efficiency, and jet-mass scale and resolution for the new
2184 PUPPI softdrop jet mass tagger, we use the same procedure as outlined in Section 7.7. The
2185 first commissioning of the new tagger was done using 2.3 fb^{-1} of data collected in 2015, and
2186 was published in a jet algorithms performance note [77]. The measurements were then redone
2187 using 12.9 and 35.9 fb^{-1} of data collected in 2016 for the two analyses presented in this chapter
2188 (the latter measurement was performed by a separate analysis team). The results shown in the

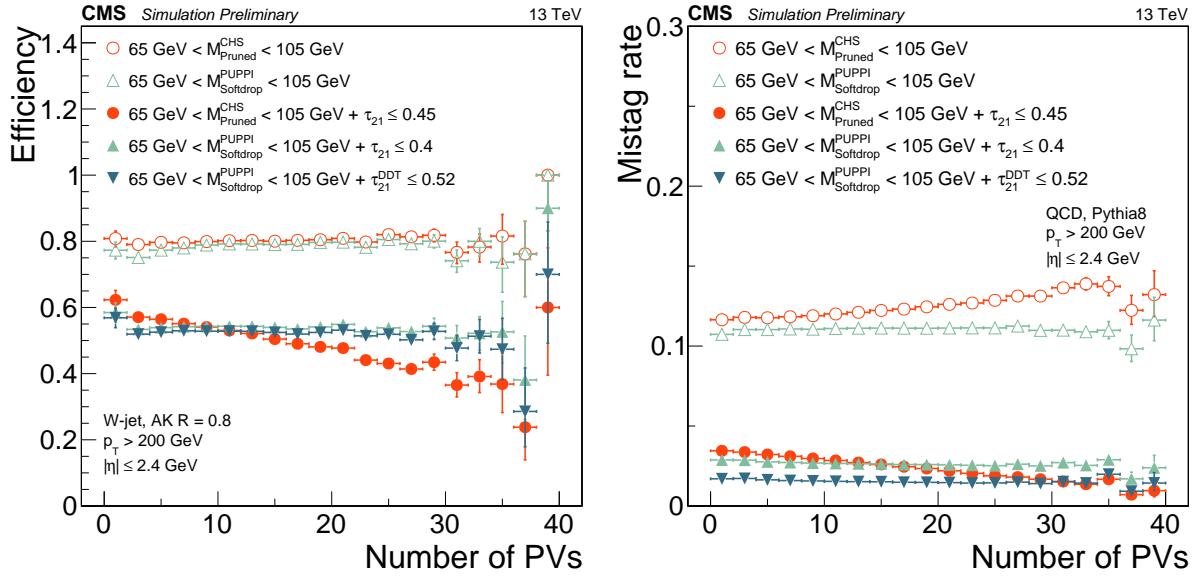


Figure 8.15: W-jet efficiency (left) and QCD light-flavor jet mistag rate (right) for a PUPPI softdrop or CHS pruned jet mass selection only (hollow circles) and the combined $m_{\text{p/sd}} + (\text{PUPPI}) \tau_{21}$ (DDT) selection (solid circles) as a function of the number of interaction vertices in the event.

following will be those obtained when the V tagger was first presented in an analysis context, corresponding to the PAS based on 12.9 fb^{-1} of data collected in the beginning of 2016. Since the fit method is outlined in detail in Section 7.7, fits to matched $t\bar{t}$ MC and minor backgrounds for the PUPPI softdrop based tagger are skipped here and can be found in Appendix C.1.

The PUPPI softdrop jet mass and the PUPPI τ_{21} variable in data and MC are shown in Figure 8.16. These can be compared to the corresponding plots for the CHS pruned jet mass and τ_{21} distributions in Figure 8.16. The softdrop and pruned jet mass distributions, as well as the PUPPI and CHS τ_{21} variables, are very similar and the variables are described equally well in simulation. Following what was done in Section 7.7, we extract and compare the W-tagging efficiency, jet mass scale and jet mass resolution of the combined jet mass and τ_{21} selection in data and in MC. This is done through a simultaneous fit of the the softdrop jet mass spectrum between 40 and 150 GeV in two regions:

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

The fits are shown in Figure 8.17, with the corresponding extracted efficiencies from the Gaussian component of the total fit and scale factors summarized in Table 8.1. The quoted systematic

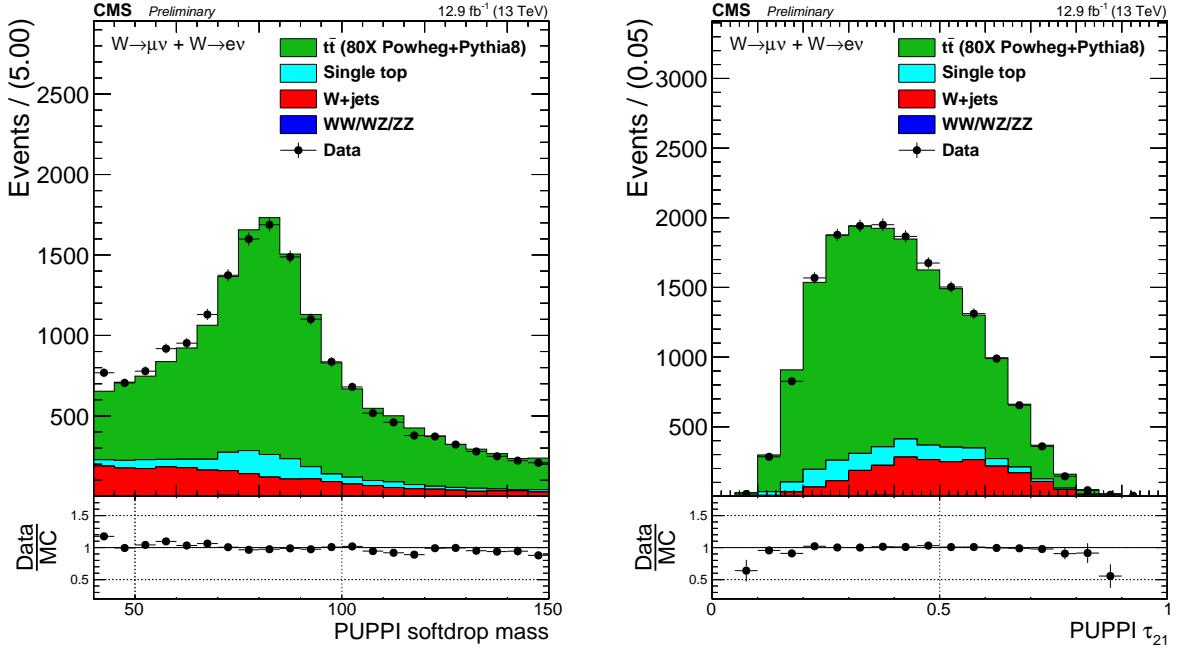


Figure 8.16: Distribution of the PUPPI softdrop mass (left) and PUPPI n-subjettiness (right) distribution in the $t\bar{t}$ control sample.

uncertainties are evaluated in the same way as was described in Section 7.7.3, and correspond to the uncertainty due to the choice of ME generator and due to the 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 8.1: 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} .

Both scale factors are compatible with unity within uncertainties. We additionally extract the jet-mass scale and jet-mass resolution from the mean and width of the Gaussian component of the total fit in the pass region. These are summarized in Table 8.2. We find that the W-jet mass scale for PUPPI softdrop jet mass is identical in simulation and in data, whereas for pruning (Table 7.7) the difference was around 2%, The jet mass resolution, on the other hand, is larger in data for PUPPI softdrop jet mass, by roughly 8%, whereas for pruning the resolution is larger in simulation (11%). However, both are statistically insignificant and compatible with unity within uncertainties. The W-tagging efficiency scale factors, jet-mass and resolution scales affect the signal yield and are included in the analysis in the same way as was described in Section 7.7.5, by

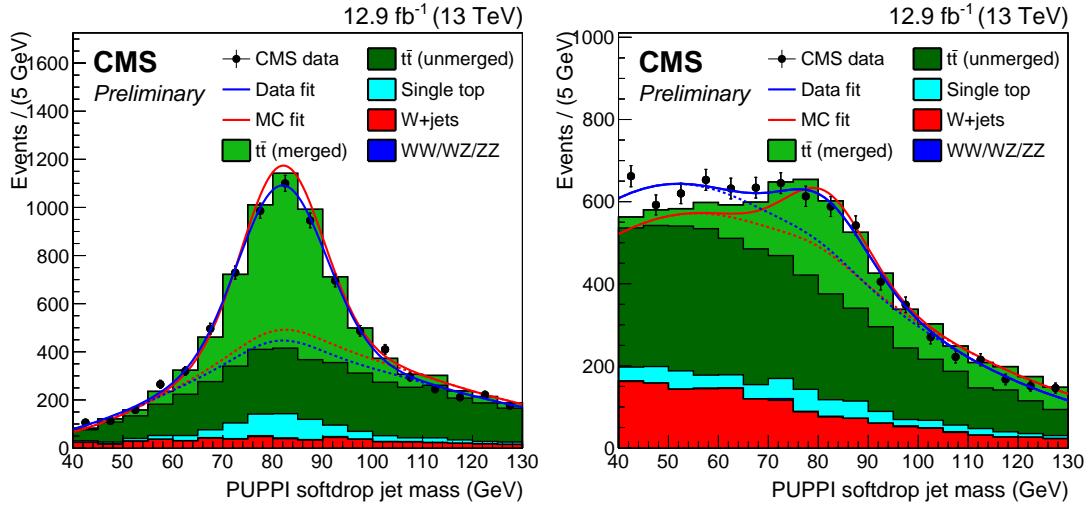


Figure 8.17: PUPPI softdrop jet mass distribution that pass (left) and fail (right) the PUPPI $\tau_{21} < 0.40$ selection. Results of both the fit to data (blue) and simulation (red) are shown and the background components of the fit are shown as short-dashed lines.

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

scaling the total signal yield and incorporating an uncertainty on the signal efficiency due to a shift and broadening of the W-jet mass peak.

8.5.4 W-tagging mistag rate measurement

The W-tagging light-flavor jet mistagging rate is measured in a QCD dijet-enriched region in data and is compared to the prediction from QCD MC using three different combinations of generators: HERWIG++, PYTHIA8, and MADGRAPH+PYTHIA8. Figure 8.18 shows the mistagging rate as a function of jet p_T for three different taggers: CHS pruning and τ_{21} , PUPPI softdrop and τ_{21} , and PUPPI softdrop and τ_{21}^{DDT} . We find a substantial difference in the modeling of substructure variables between the different generators, most likely coming from their very different description of gluon radiation (dominant in QCD multijet events). The best description is obtained with HERWIG++, while all three generators model the p_T -dependence of the tagger well.

We additionally study how a selection on τ_{21} and τ_{21}^{DDT} affect the total quark and gluon content of the QCD background. Figure 8.19 shows the stacked relative quark and gluon content

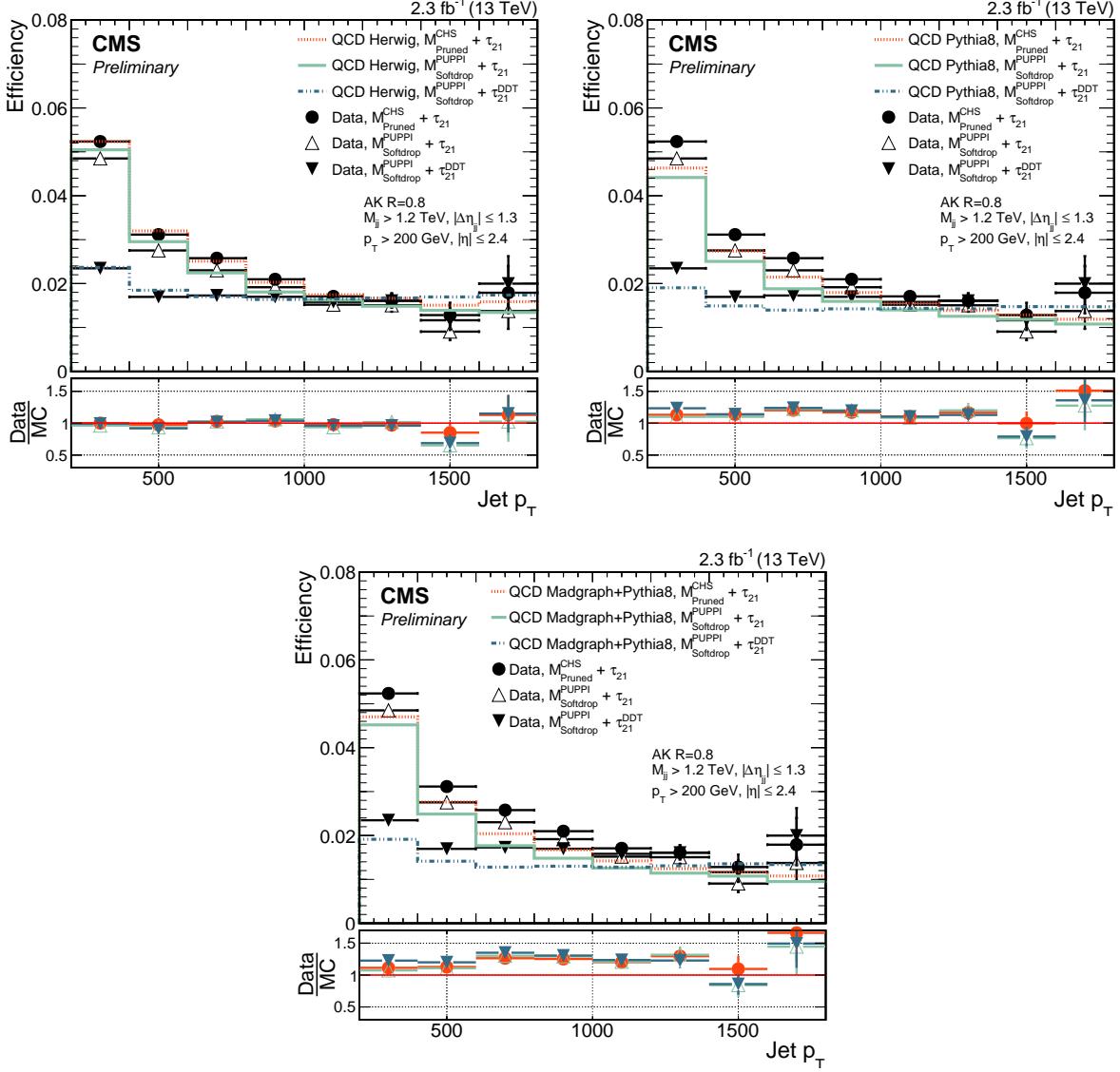


Figure 8.18: The fraction of jets that pass the $m_{p/\text{sd}}$ and τ_{21} selections in a dijet enriched sample for data and for simulation as a function of jet p_T . Here the comparison is between HERWIG++ (top left), PYTHIA8 (top right), and PYTHIA8 with MADGRAPH as the matrix-element generator (bottom).

in a PYTHIA8 QCD dijet sample for selection requirements based on PUPPI τ_{21} and τ_{21}^{DDT} . We see that the quark content increases as a function of jet p_T when applying a selection on τ_{21}^{DDT} , while it decreases when applying a selection on τ_{21} . This can be attributed to the fact that the distribution of jet mass divided by the jet p_T , m/p_T , for quark and gluon jets are very different

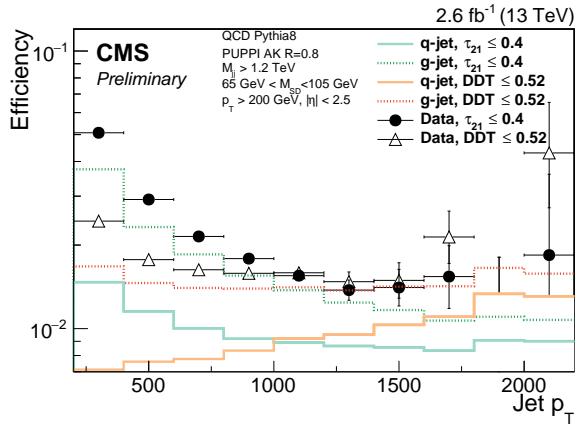


Figure 8.19: The fraction of jets that pass the PUPPI softdrop jet mass selection and a selection on τ_{21} (turquoise) or τ_{21}^{DDT} (orange) in a dijet-enriched sample. The jets from QCD MC are split into two contributions: jets originating from gluons (dotted line) and jets originating from quarks (solid line).

from one another, and these differences increase as the jet p_T increases. Figure 8.20 shows the mass divided by p_T for jets originating from a quark (blue) or a gluon (red), for jets with a p_T of 200 GeV (left) or 1600 GeV (right). We see that the jet mass divided by the jet p_T peaks at significantly higher values for gluon jets than for quark jets. From the definition of the τ_{21}^{DDT} tagger in Equation 8.2, it can therefore be seen that a selection on τ_{21}^{DDT} will act more aggressively on jets with a high m/p_T , effectively removing more gluon jets.

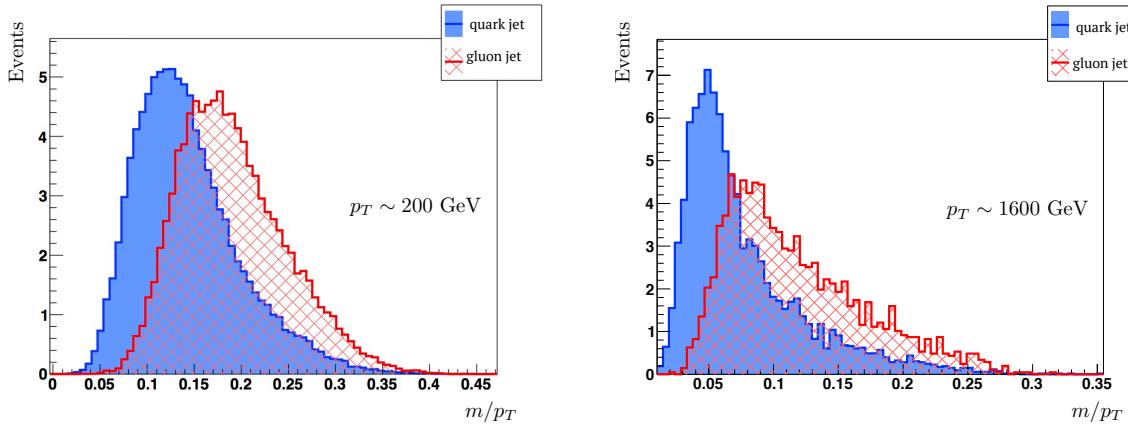


Figure 8.20: The jet mass divided by the jet p_T for quark (blue) and gluon (red) jets for a jet p_T of 200 (left) and 1600 GeV (right). Generated using [80].

8.6 Mass and purity categorization

The PUPPI softdrop jet mass and PUPPI τ_{21} distributions after loose analysis preselections, as outlined in Section 8.3, have been applied are shown in Figure 8.21. We see some disagreement between data and MC, especially in the high-purity region ($\text{PUPPI } \tau_{21}^{DDT} < 0.4$), confirming what was observed in Section 8.5.4. As this analysis is sensitive to both heavy resonances

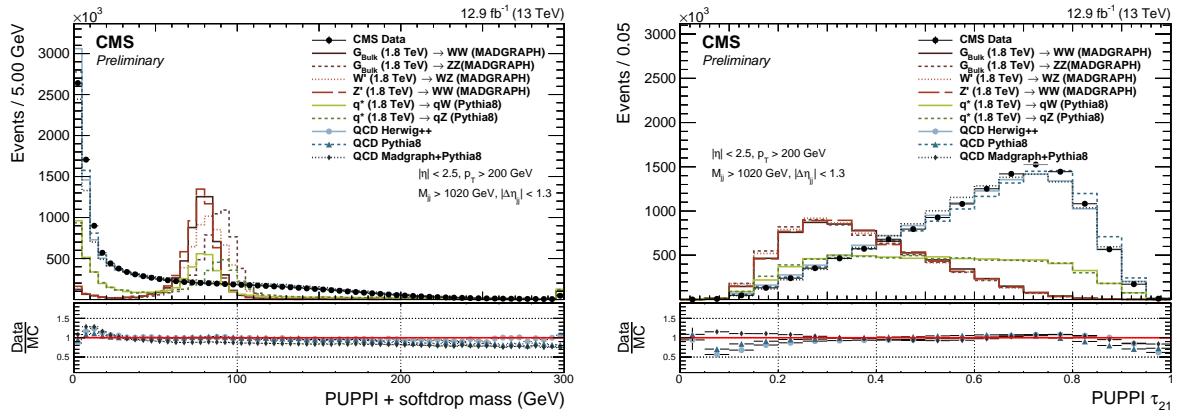


Figure 8.21: PUPPI softdrop jet mass (left) and PUPPI n-subjettiness τ_{21} (right) distribution for data and simulated samples. Simulated samples are scaled to match the distribution in data.

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

2258 8.7 Background modeling

2259 The background is modeled using a parametric fit of the dijet invariant mass spectrum in the
2260 data signal region in the same way as described in Section 7.5. We determine the number of
2261 necessary fit parameters in order to describe the background through a Fishers F-test, comparing
2262 the same fit functions as in Section 7.5. This test is first exercised in QCD MC and then in a
2263 data sideband before the final determination is performed in the data signal region. As the F-test
2264 method was presented in detail in the context of Search I, only a brief summary of the findings
2265 as well as a presentation of the fits in the single-tag categories will be presented here. Fits and
2266 F-test results for all categories can be found in Appendix C.3. A two- or three-parameter fit
2267 is sufficient to describe the background for all the double-tag categories: a two-parameter fit is
2268 sufficient for the “high-purity” WZ and ZZ categories, as well as the “low-purity” WW category,
2269 while the remaining analysis categories require a three-parameter background fit. From the fits of
2270 the single-tag categories, shown in Figure 8.22, a three-parameter fit is sufficient for all categories
2271 except for the “high-purity” qW category. In the qW category, the improvement in fit quality
2272 when increasing the number of parameters is so large that adding an additional fit parameter is
2273 justified, and we continue by using a 5-parameter fit for this category. A summary of what fit
functions are used for each analysis category is listed in Table 8.3.

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

Table 8.3: The number of fit parameters used in the background fit to the dijet invariant mass distribution for each analysis category as determined through an F-test.

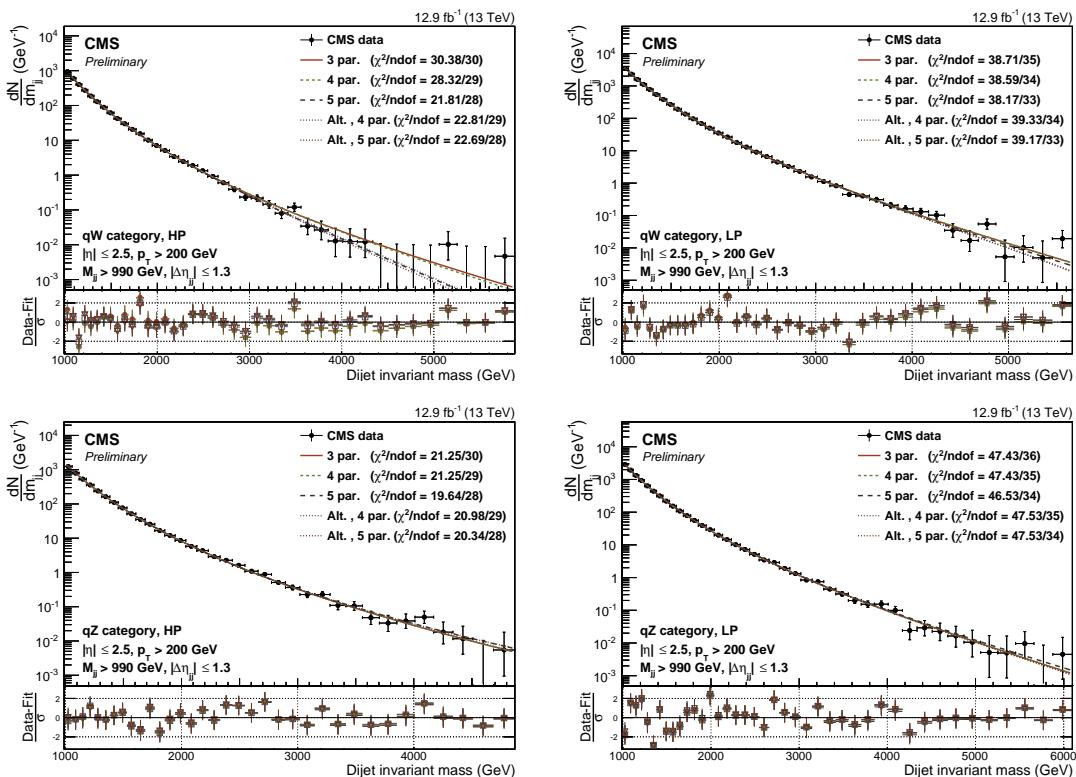


Figure 8.22: Background-only fit of the dijet invariant mass distribution in the data signal region used to establish the number of parameters of the background shape for the single-tag analysis. Here for the high- (left) and low-purity (right) single V-tag categories qW (top) and qZ (bottom).

2275 **8.8 Signal modeling**

2276 The dijet invariant mass distribution for signal is modeled from MC in the same way as in
2277 Section 7.6, where we assume the shape can be described by a Gaussian core and an exponential
2278 tail. The interpolated signal shapes for $q^* \rightarrow qW$ and $q^* \rightarrow qZ$ in their most sensitive analysis
2279 categories (qW and qZ , respectively) are shown in Figure 8.23. The signal shapes for the
 double-tag categories can be compared to those in Figure 7.22.

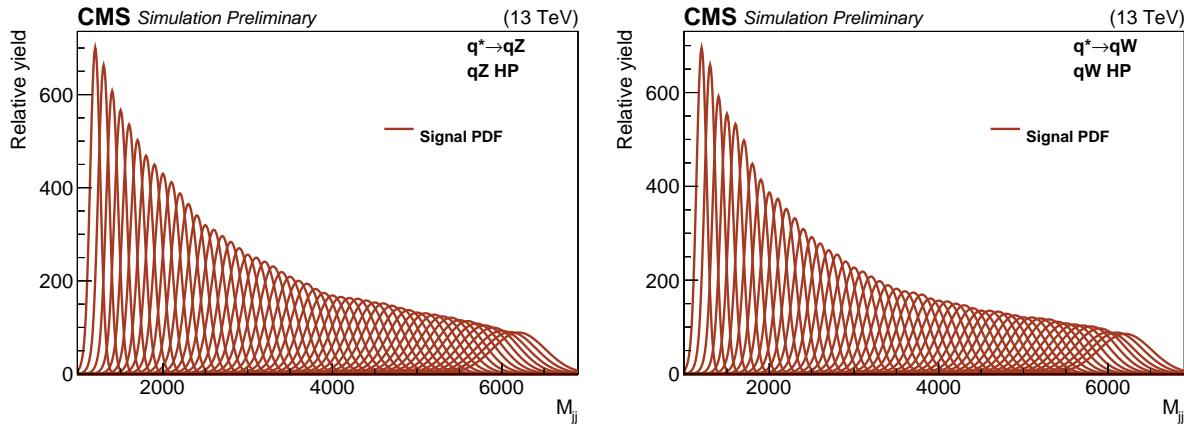


Figure 8.23: Interpolated signal shapes for a $q^* \rightarrow qZ$ (left) and $q^* \rightarrow qW$ (right) signal.

²²⁸¹ 8.9 Systematic uncertainties

²²⁸² The largest sources of systematic uncertainty for this analysis are, as for Search I, related to the
²²⁸³ signal modeling and are due to the uncertainty in the tagging efficiency of the V-tagger, the jet
²²⁸⁴ energy and mass scale, the jet energy and mass resolution and the integrated luminosity. The
²²⁸⁵ V-tagging uncertainty is estimated as described in Section 8.5.3, and yield uncertainties on the
²²⁸⁶ scale factors for the HP and LP tagging categories. The p_T - and η -dependent jet energy scale
²²⁸⁷ and resolution uncertainties on the resonance shape are approximated by a constant 2% and 10%.
²²⁸⁸ The jet energy scale and resolution uncertainty are taken into account as shape uncertainties by
²²⁸⁹ shifting and widening the signal resonance model, while all other signal uncertainties only affect
the yield. The most relevant systematic uncertainties are listed in Table 8.4.

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

Table 8.4: Summary of the systematic uncertainties on the signal and their impact on the event yield in the signal region and on the reconstructed dijet invariant mass shape (mean and width).

2291 8.10 Results

2292 As mentioned in the introduction to this chapter, the analysis of the 2016 dataset was done in
2293 two steps: one based on 12.9 fb^{-1} of data collected early in 2016, demonstrating the new PUPPI
2294 softdrop V-tagger and the new single-tag analysis categories, and one utilizing the full 35.9 fb^{-1}
2295 of data collected in 2016. The results from both will be presented in the following sections.

2296 8.10.1 Early analysis

2297 The dijet invariant mass distributions in the analysis signal regions using 12.9 fb^{-1} of data
2298 collected in 2016, are shown in Figure 8.24 for the double-tag and in Figure 8.25 for the single-tag
2299 analysis. No significant excess is observed and we proceed to set upper limits on the signal cross
2300 section using the method described in Section 7.9. Exclusion limits are set in the context of
2301 the bulk graviton model, the HVT model B, and excited-quark resonance models, assuming the
2302 resonances have a natural width negligible with respect to the experimental resolution (as in
2303 Search I). Figure 8.26 shows the expected and observed upper exclusion limits at 95% confidence
2304 level (CL) on the signal cross section times the branching fraction as a function of the resonance
2305 mass for the different signal hypotheses in the double-tag analysis. The limits are compared
2306 with the theoretically predicted cross section times branching fraction to WW and ZZ for a bulk
2307 graviton with $\tilde{k} = 0.5$, and with the cross section times branching fraction to WZ and WW
2308 for spin-1 particles predicted by the HVT model B. For the HVT model B, we exclude W' (Z')
2309 resonances with masses below 2.7 (2.6) TeV. The signal cross section uncertainties are displayed
2310 as a red checked band and result in an additional uncertainty on the resonance mass limits of
2311 0.05 (0.04) TeV. The cross section limits for $Z' \rightarrow WW$ and $G_{\text{bulk}} \rightarrow WW$ are not identical due
2312 to the different acceptance for those two signal scenarios. Figure 8.31 shows the corresponding
2313 exclusion limits for excited quarks decaying to qW and qZ , excluding a q^* decaying into qW and
2314 qZ with masses below 5.0 and 3.9 TeV, respectively. The signal cross section uncertainties are
2315 displayed as a red checked band and result in an additional uncertainty on the resonance mass
2316 limits of 0.1 TeV.

2317 8.10.2 Full 2016 dataset

2318 Figure 8.28 shows the dijet invariant mass distributions in the analysis signal regions using the
2319 full dataset of 35.9 fb^{-1} collected in 2016 for the double-tag and in Figure 8.29 for the single-tag
2320 analysis. No significant excess is observed and we proceed with setting upper limits on the signal
2321 cross section times branching ratio as in Section 7.9.

2322 For a G_{bulk} we exclude production cross sections in a range from 36.0 fb, at a resonance mass
2323 of 1.3 TeV, to 0.6 fb at a resonance mass of 3.6 TeV. The W' (Z') resonances are excluded with
2324 masses below 3.2 (2.7) TeV for the HVT model B, in addition to W' resonances with a mass
2325 between 3.3 and 3.6 TeV. For excited quark resonances, we can exclude the production of q^*
2326 decaying to qW or qZ for masses below 5.0 and 4.7 TeV, respectively. Figure 8.30 and 8.31 show
2327 the resulting expected and observed upper limits at 95% confidence level on the signal cross
2328 section times branching ratio as a function of the resonance mass for VV and qV resonances,
2329 respectively.

2330 The sensitivity improvement in terms of mass reach with respect to Search I for a W' resonance
2331 of the HVT model B, improved from excluding resonances below 2 TeV to excluding resonances
2332 below 3.2 TeV.

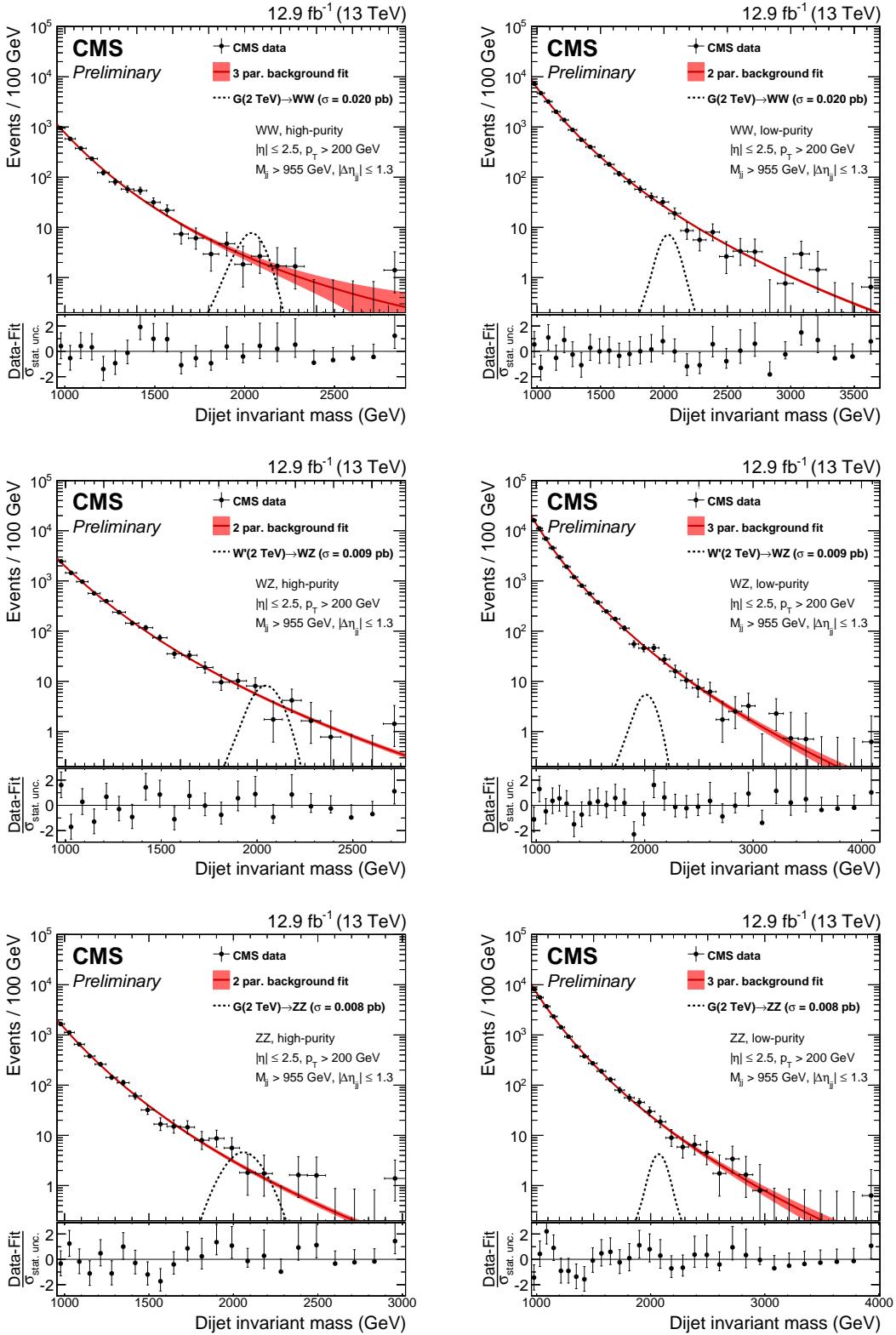


Figure 8.24: Observed dijet invariant mass spectrum for the double-tag analysis in the high-purity (left) and low-purity (right) categories using 12.9 fb^{-1} of data collected in 2016.

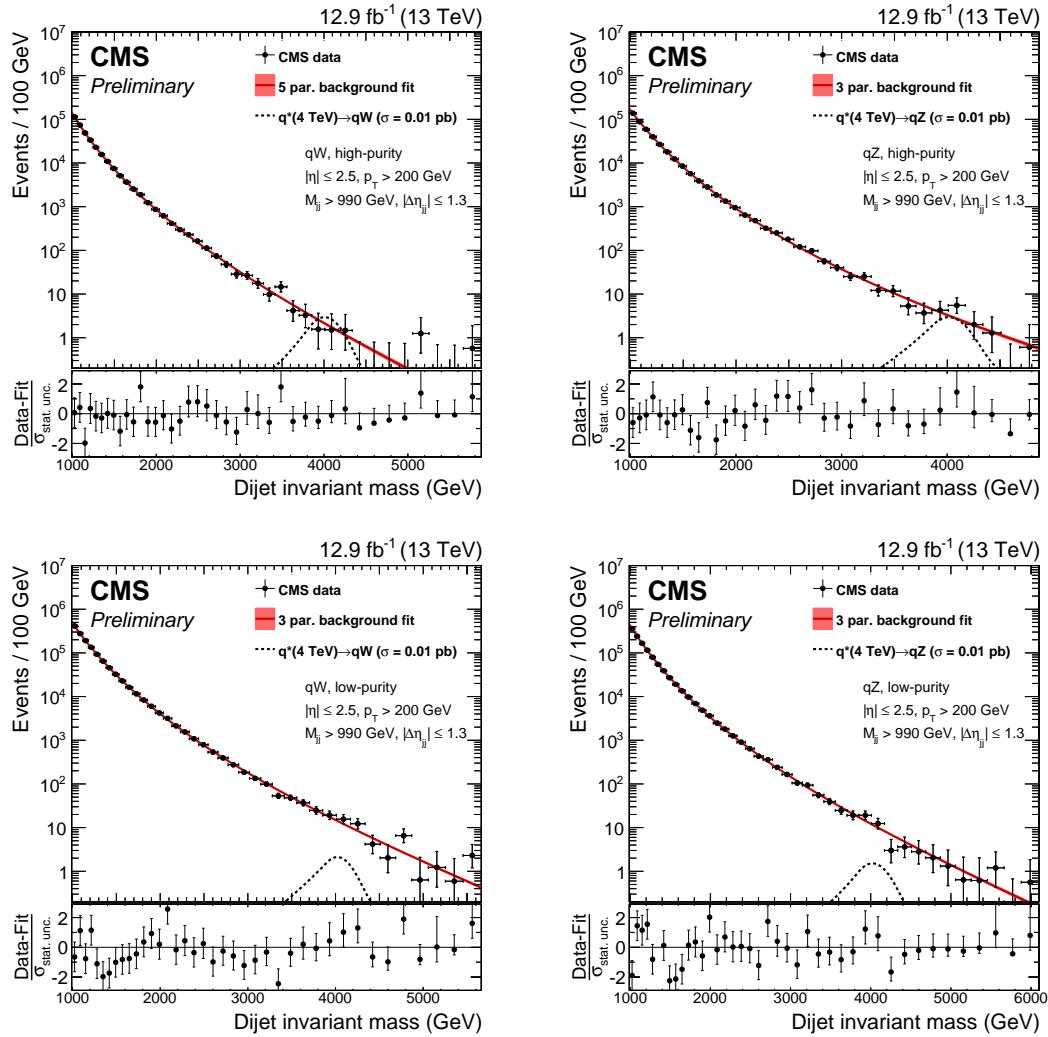


Figure 8.25: Observed dijet invariant mass spectrum for the single-tag analysis in the high-purity (top) and low-purity (bottom) categories using 12.9 fb^{-1} of data collected in 2016.

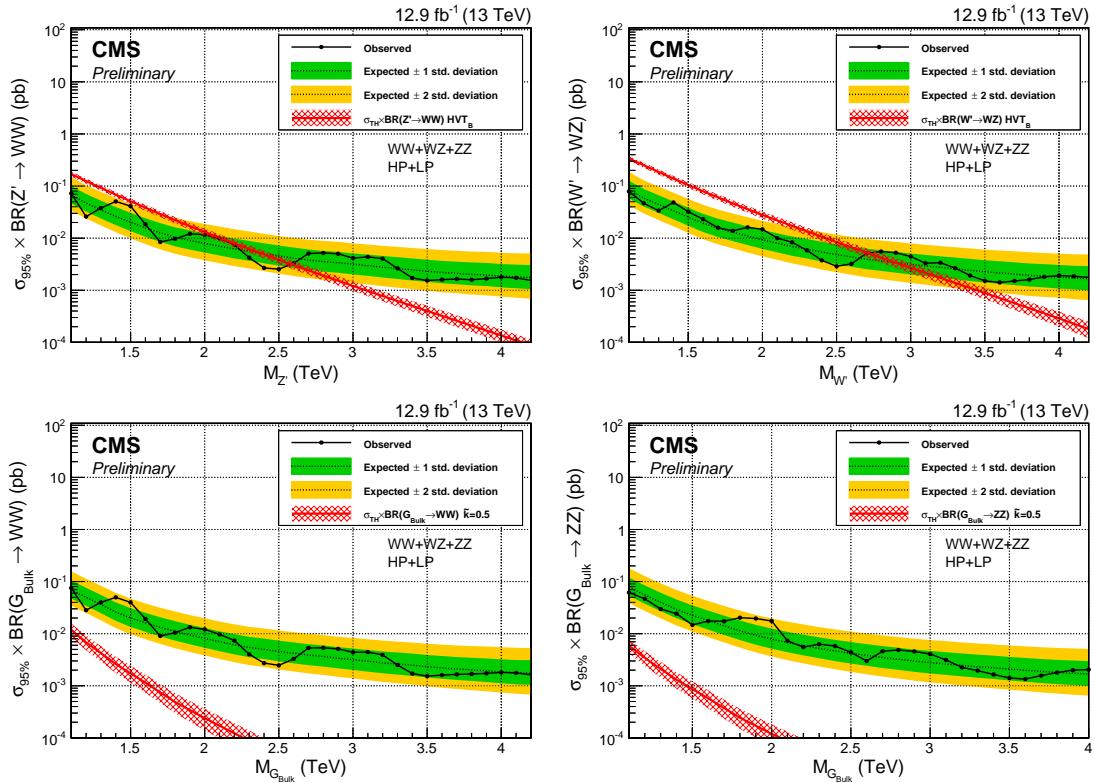


Figure 8.26: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of a narrow-width resonance decaying to a pair of vector bosons for different signal hypotheses. Limits are set in the context of a spin-1 neutral Z' (left) and charged W' (right) resonance, and compared with the prediction of the HVT model B. On the bottom, limits are set in the context of a bulk graviton decaying into WW (left) and ZZ (right) with $\tilde{k}=0.5$ and compared with the model prediction. Signal cross section uncertainties are displayed as a red checked band.

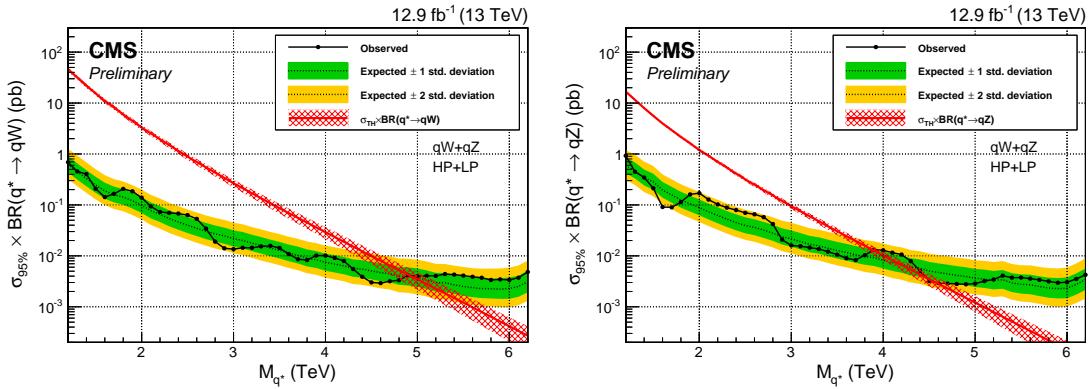


Figure 8.27: Observed (black solid) and expected (black dashed) 95% CL upper limits on the production of an excited quark resonance decaying into qW (left) or qZ (right). Signal cross section uncertainties are displayed as a red checked band.

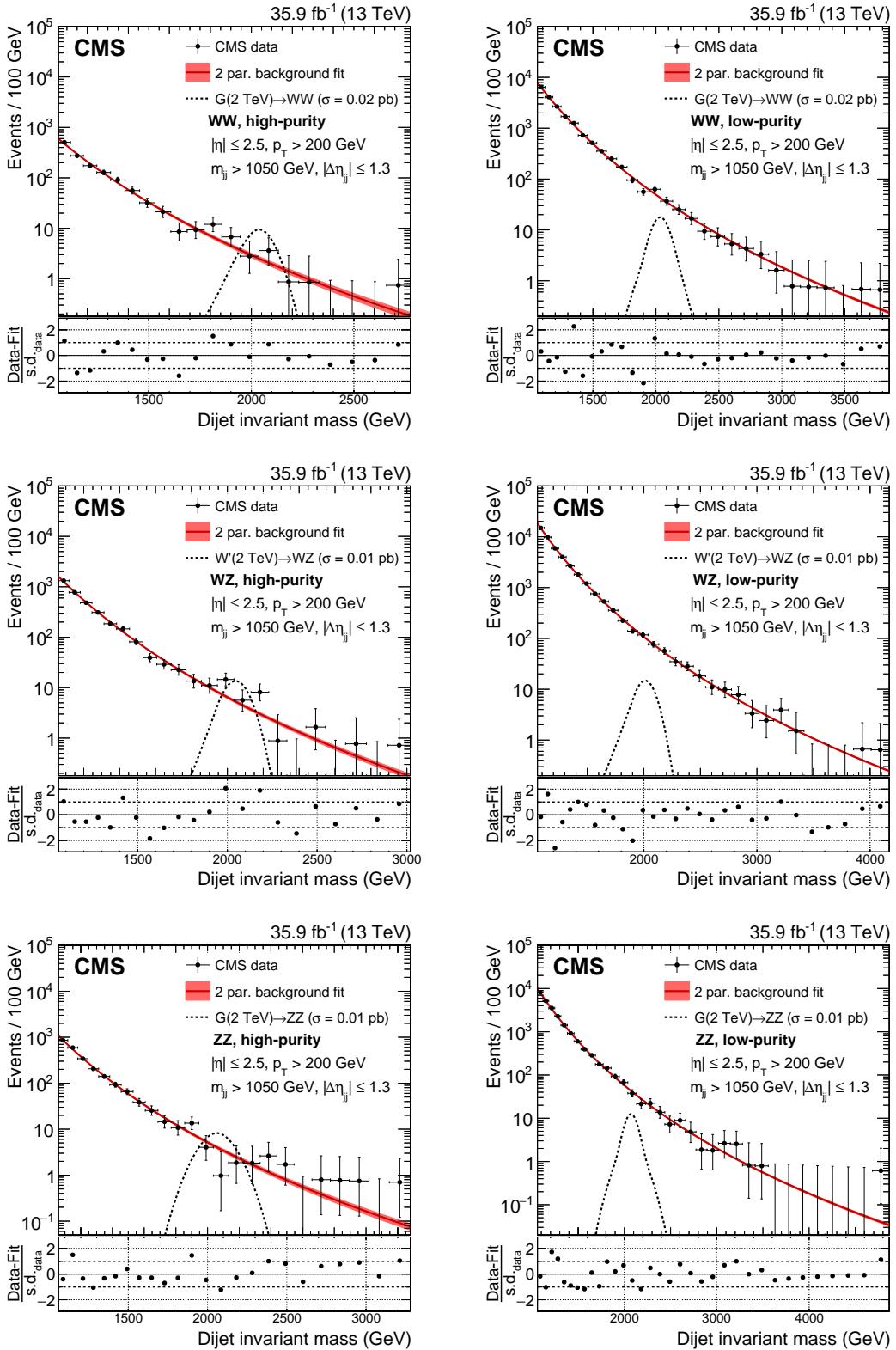


Figure 8.28: Observed dijet invariant mass spectrum for the double-tag analysis in the high-purity (left) and low-purity (right) categories using 35.9 fb^{-1} of data collected in 2016.

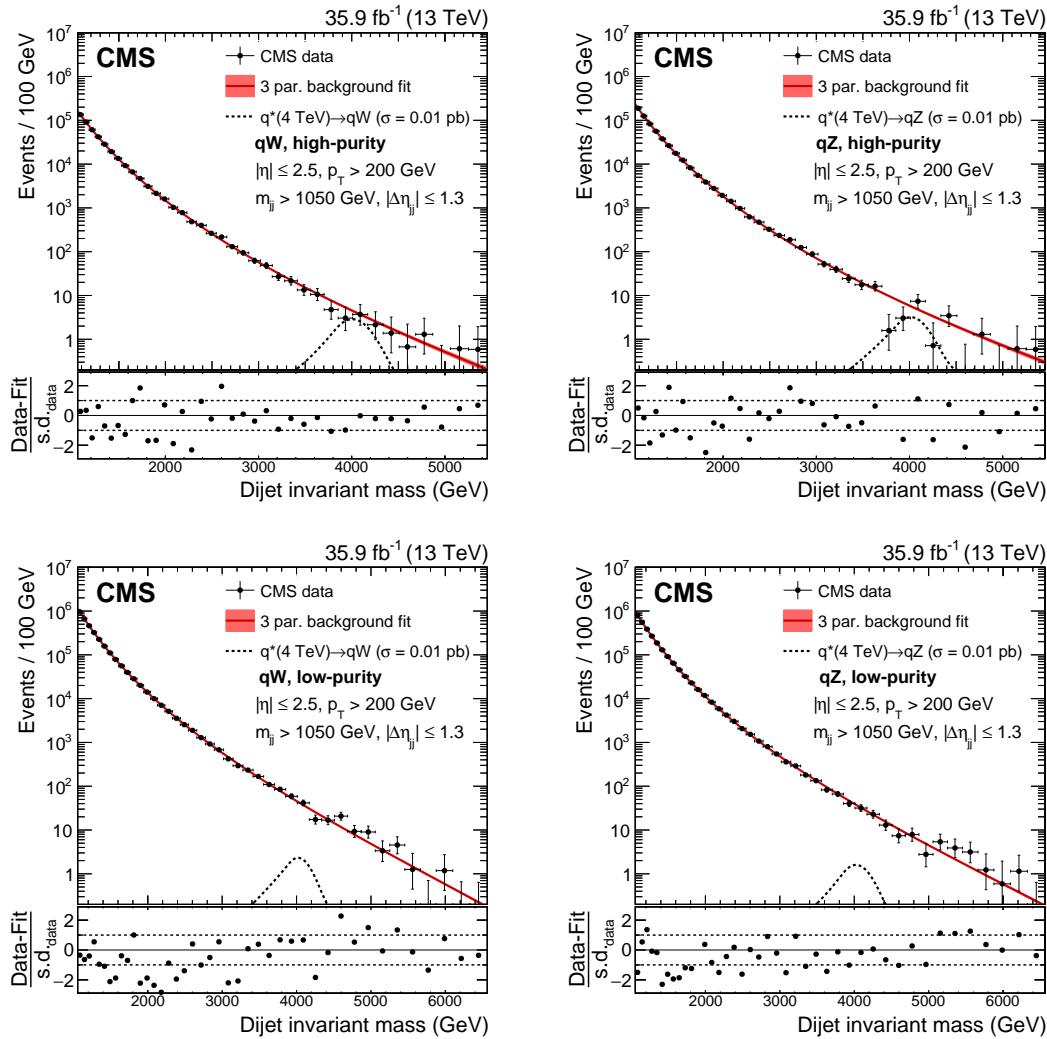


Figure 8.29: Observed dijet invariant mass spectrum for the single-tag analysis in the high purity (top) and low purity (bottom) categories using 35.9 fb^{-1} of data collected in 2016.

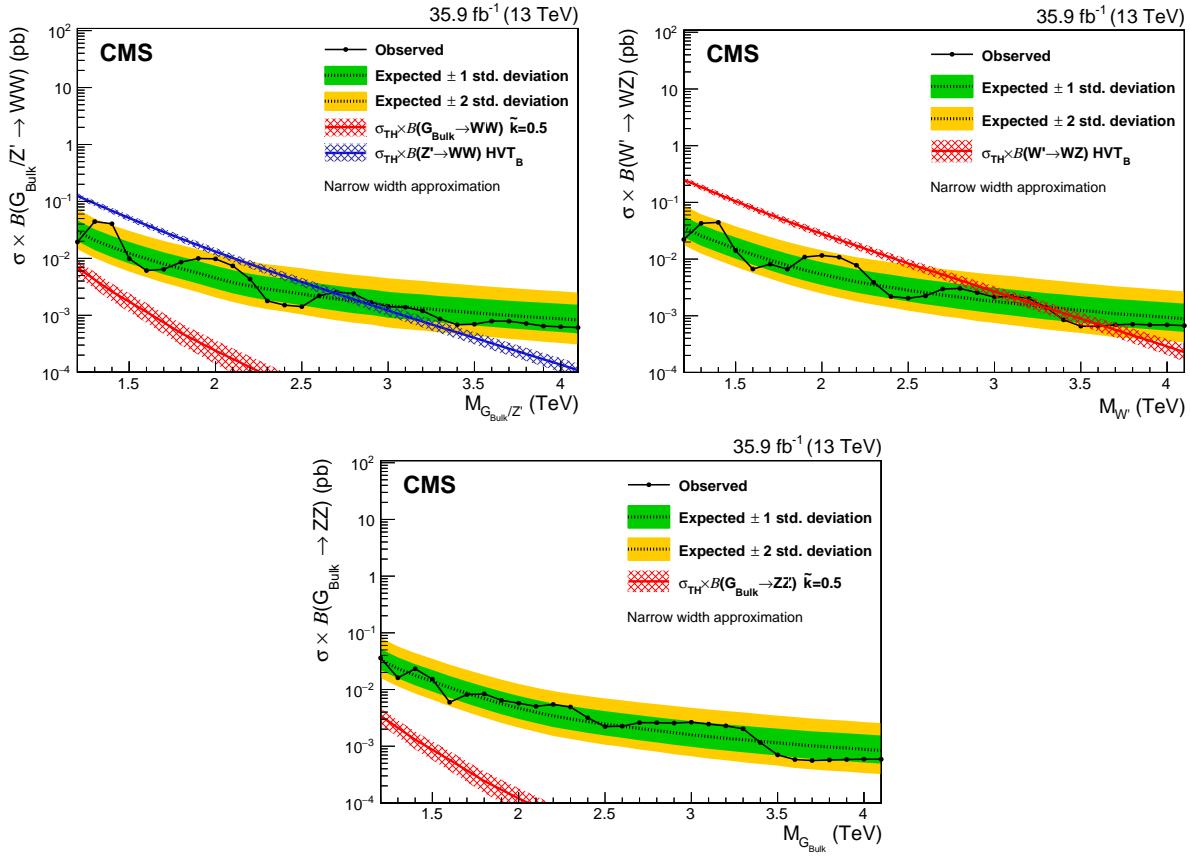


Figure 8.30: Observed (solid line) and expected (dashed line) 95% CL upper limits on the production cross section of a narrow resonance decaying into two vector bosons for different signal hypotheses: a Z' or G_{bulk} resonance decaying into WW (top left), a Z' decaying into WZ (top right) and a bulk graviton decaying into ZZ (bottom).

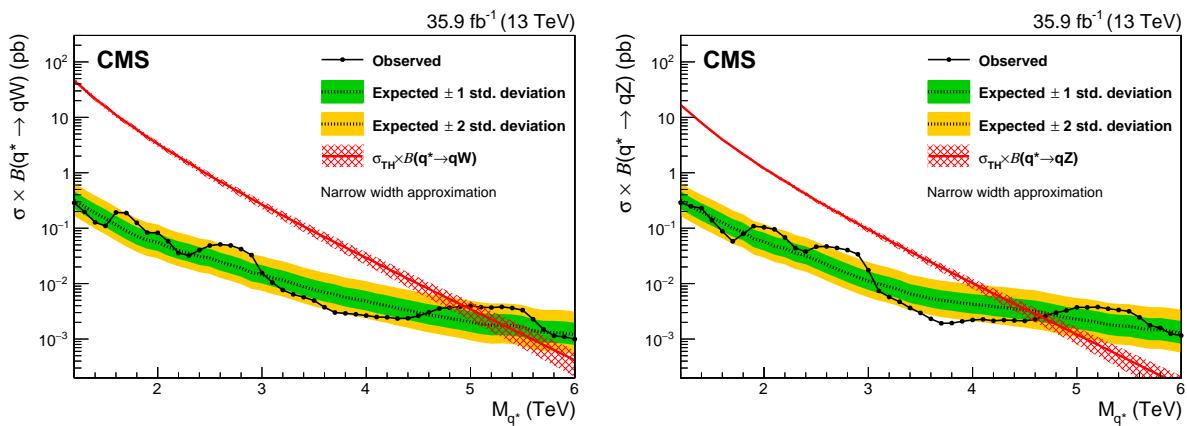


Figure 8.31: Observed (solid line) and expected (dashed line) 95% CL upper limits on the production of an excited quark resonance decaying into qW (left) or qZ (right).

2333

CHAPTER 9

2334

Search III: A novel framework for multi-dimensional searches

2335

2336 After two successful analysis using the LHC data collected at a center-of-mass energy of 13
2337 TeV, there was no confirmation of the excess observed at a center-of-mass energy of 8 TeV,
2338 and the prospect of observing new physics in diboson analyses was bleak. However, this was not
2339 a source of concern to some theorists, who considered whether the small bumps that had been
2340 observed were due to us observing the tail of another type of boson with a mass slightly different
2341 from that of a W or a Z boson, and that perhaps these bosons were not 2-prong, but 4-prong objects.

2342

2343 With no significant excess observed with the 2016 dataset of 36 fb^{-1} , we were expecting to
2344 collect a total dataset of 150 fb^{-1} in Run 2 (2015-2018) that would allow us to probe alternative
2345 BSM models, and we wanted to do so as efficiently as possible. More specifically, we wanted a
2346 way of looking for any heavy resonance decaying to any two jets with substructure, with masses
2347 anywhere in the jet mass spectrum. We decided to do so by taking advantage of the fact that we
2348 are looking for bumps in a three-dimensional space: the mass of the two jets as well as their dijet
2349 invariant mass. The one-dimensional dijet fit can therefore be replaced by a three-dimensional fit,
2350 looking for bumps in all three dimensions simultaneously. This would allow us to easily search for
2351 resonances decaying to any object peaking in jet mass; $W(qq)$, $Z(qq)$, $H(qq)$, as well as non-SM
2352 bosons; in one common framework.

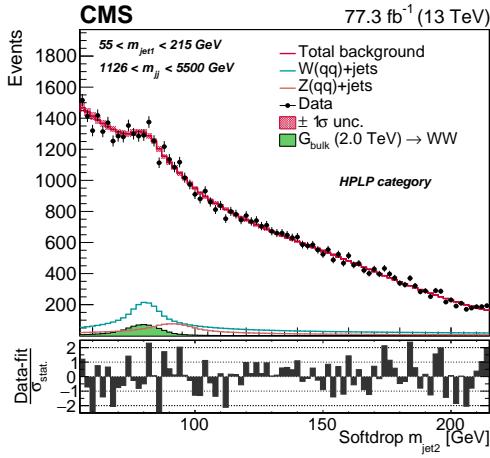
2353

2354 An additional benefit of the method, was that the modeling of the QCD multijet background
2355 would start from simulation rather than relying on a parametric fit to data. This had the benefit
2356 of allowing more control over the background shape across the full dijet invariant mass spectrum,
2357 making the fit less prone to fluctuations in the low statistics tail of the distribution, something we

2358 found to be problematic with the 1D dijet fit method.

2359

2360 Search III introduces a novel three-dimensional fit method that can be used to search for heavy
 2361 resonances decaying to any two jets with substructure peaking in the jet groomed mass spectrum.
 2362 It is validated in the context of a diboson resonance search in the all-hadronic final state, but is
 2363 easily extendable to other signals. The fit method has allowed for the first measurement of the
 2364 jet mass scale and resolution simultaneously from a $W(qq)+jets$ and $Z(qq)+jets$ mass peak, and
 2365 could also allow for the extraction of the SM $W/Z+jets$ cross section. It was published with data
 2366 collected in 2016 and 2017, corresponding to a total integrated luminosity of $\sim 80 \text{ fb}^{-1}$.



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

2367 9.1 Small bumps and tri-bosons

2368 In addition to the observation of a 3.4σ excess in the search for diboson resonances in the
 2369 all-hadronic final state by ATLAS [52], not confirmed by subsequent searches, several little
 2370 bumps near 2 TeV were observed in data collected at center-of-mass energies of 8 and 13 TeV, as
 2371 illustrated by the dijet invariant mass distributions observed by ATLAS and CMS in Figure 9.1.
 2372 These were not statistically significant, and are expected due to statistical fluctuations, however,
 2373 only a coherent analysis would be able to determine if a new physics signal could be partly
 responsible for any of these excesses in the dijet invariant mass of V jets. Due to their small size

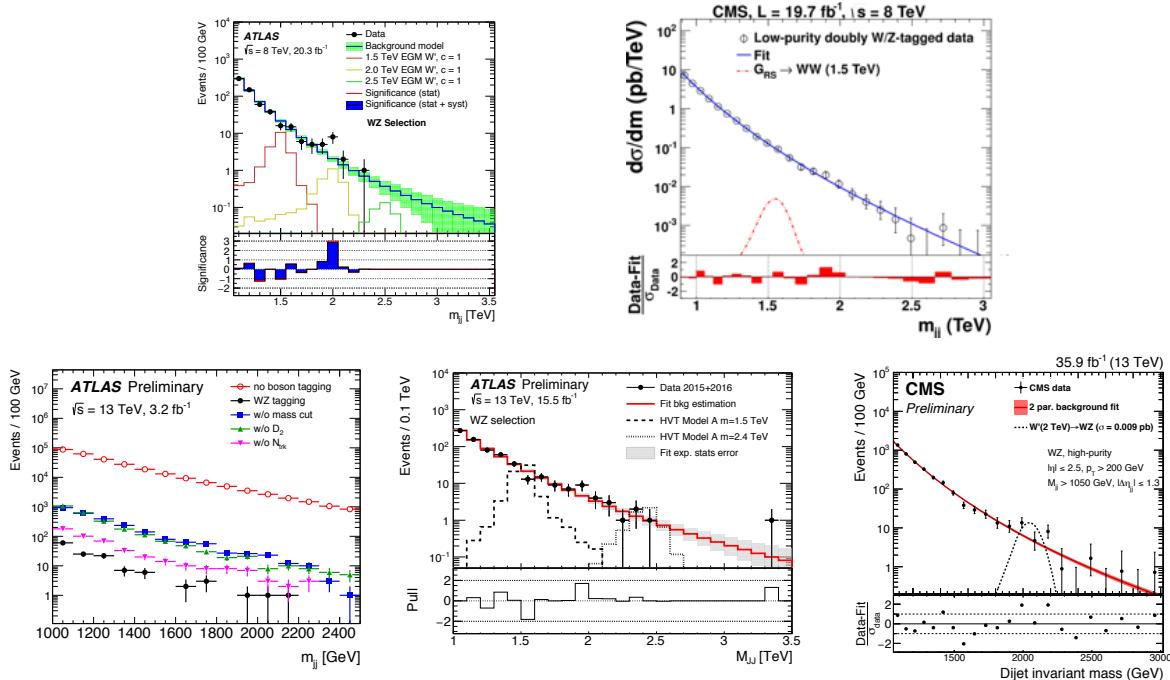


Figure 9.1: Several small bumps have been observed in diboson resonance searches in the all-hadronic final state, both in ATLAS and in CMS [81].

2374 and the way the excesses appeared to slightly shift around, these were obviously not diboson
 2375 resonances. However, a proposal was made that they could be caused by the observation of
 2376 a non-SM boson, such that jets with a mass several standard deviations away from the new
 2377 particle's mean, its so-called *tail*, were being observed, as illustrated in Figure 9.2. Further, these
 2378 could be 4-pronged objects rather than 2-prong, which would cause the excess to vary in size
 2379 depending on the 4-prong efficiency of the analysis specific W-tagger in use. An explanation for
 2380 the observed excesses was proposed in [82]. This paper pointed out that, if particles like W' and

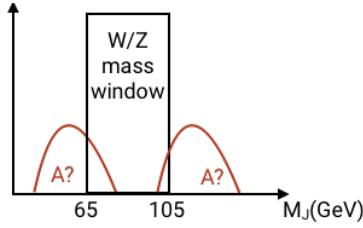


Figure 9.2: The small excesses in diboson searches could be caused by the observation of the tail of a non-SM object, peaking at a groomed jet mass slightly higher or lower than that of a W or Z boson.

Z' exist, an extended scalar sector is needed in order to give mass to the vector bosons. These heavy scalars would decay to lighter bosons, if kinematically allowed, leading to a cascade of decays that produced signals of multiple bosons in a single jet. Some example signatures are illustrated in Figure 9.3. Signatures like these would peak in the groomed-jet mass spectrum and,

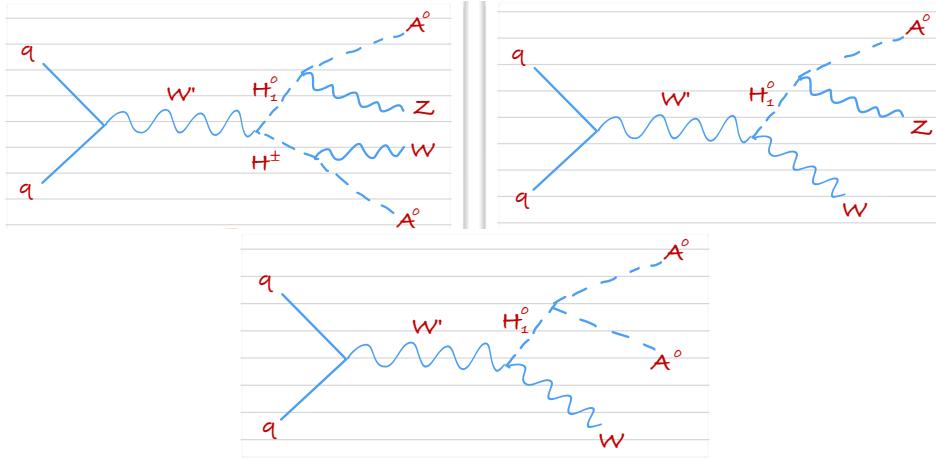


Figure 9.3: A W' decaying to a neutral H^0 and a charged H^\pm scalar particle leading to a quadriboson final state (left), and a W' decaying to a neutral scalar particle H^0 and a W leading to a triboson final state (middle and right) [82].

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

2394 the diboson all-hadronic search since this was an analysis we were familiar with and which would
2395 allow for a straight forward comparison of the obtained results.

2396 **9.2 Analysis strategy**

2397 The background estimation used in Search I and Search II relies on a one-dimensional fit of
2398 the dijet invariant mass of the signal region after a tight selection on the jet mass has been
2399 applied. Since the signal will peak in all three dimensions, mainly the dijet invariant mass
2400 M_{VV} , and the mass of jet 1 and jet 2 (M_{jet1} and M_{jet2}), we will extract the signal from the
 three-dimensional plane of M_{VV} - M_{jet1} - M_{jet2} . The benefits of doing so are that we can search for

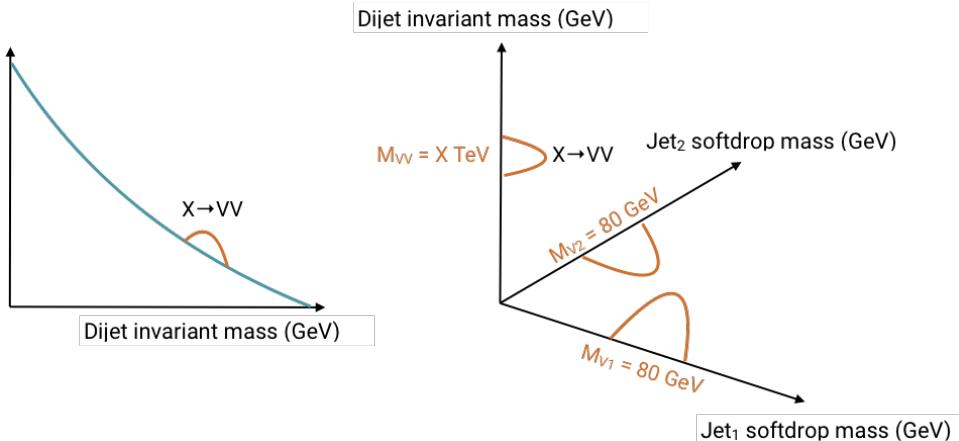


Figure 9.4: The one dimensional diboson analysis versus the three dimensional fit.

2401
2402 resonances decaying to VV ($V=W/Z$), VH ($H=$ Higgs), HH, VX, VH, XX, or XY, where X and
2403 Y are new hypothetical bosons, in the same analysis. Additionally, a jet-mass selection is no
2404 longer needed as we fit the full jet mass line-shape to extract the signal. This effectively increases
2405 the signal statistics since a large fraction of the W and Z signal falls outside the mass window.
2406 Fitting the groomed-jet mass and resonance mass together also allows for the addition of nuisance
2407 parameters that simultaneously affect both in order to fully account for the correlation between
2408 the variables. Finally, we would model the background starting from simulation, rather than
2409 from a dijet fit to data. This allows the background shape to assume non-smooth distributions,
2410 and could allow the search to probe lower dijet masses, something we will discuss further in
2411 Section 9.5. Replacing the parametric fit by a simulation-based model would also reduce the fit
2412 sensitivity to background fluctuations in the extreme tails of the dijet invariant mass spectrum.

2413 This chapter presents a novel three-dimensional fit method in the search for diboson resonances

2414 in the all-hadronic final state, and is based on 77.3 fb^{-1} of data collected in 2016 and 2017. The
2415 data collected in 2016 was already analyzed in the context of Search II, but is reanalyzed here in
2416 order to compare the performance of the three-dimensional fit to that of the one dimensional
2417 search.

2418 9.3 Data and simulated samples

2419 The data analyzed in this search consists of 35.9 fb^{-1} of data collected in 2016 and 41.4 fb^{-1} of
2420 data collected in 2017, yielding a total of 77.3 fb^{-1} .

2421 The simulated samples are the same as those described in Section 8.3.

2422 9.4 Event selection

2423 Events are selected following the same criteria as in Search I and Search II (see Section 7.4.2)
2424 and can be summarized as follows: we require two AK8 jets with PUPPI pileup subtraction
2425 applied, and with $\eta < 2.5$ and $p_T > 200 \text{ GeV}$. The two jets are further required to be separated
2426 by $|\Delta\eta|_{jj} < 1.3$. The softdrop algorithm is applied, and the two jets with the highest jet mass in
2427 the event are selected as potential vector boson candidates. In order to avoid any bias in the
2428 jet mass shapes when performing the three-dimensional fit, the two jets are randomly sorted
2429 after the above selections have been applied. In addition, the dijet invariant mass is required
2430 to be greater than 1126 GeV in order to be on the trigger plateau. As already mentioned in
2431 the introduction, the background modeling used in this analysis is capable of modeling trigger
2432 turn-ons. We explored the possibility of starting the analysis below the trigger plateau, however,
2433 while the background modeling was found to be reliable, the extraction of a signal peaking on
2434 top of a turn-on was difficult, and we therefore had to abandon the modeling of the trigger
2435 turn-on for this first demonstration of the method. More details will be given in Section 9.5. The
2436 dijet invariant mass and $|\Delta\eta|_{jj}$ distribution for the two leading jets in the event after the above
2437 preselections have been applied are shown in Figure 9.5 and the jet p_T and η distributions for
2438 signal and for background is shown in Figure 9.6.

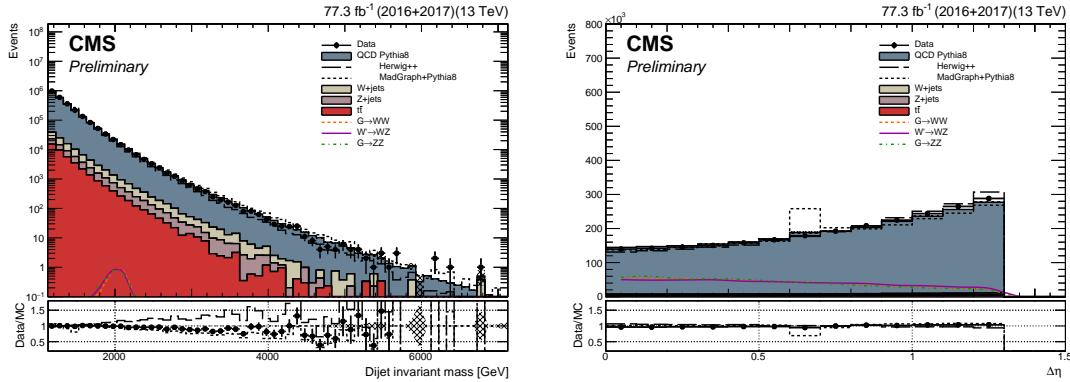


Figure 9.5: The dijet invariant mass (left) and $|\Delta\eta|_{jj}$ (right) for the two leading jets after preselections are applied. The signal is scaled by an arbitrary number to be visible.

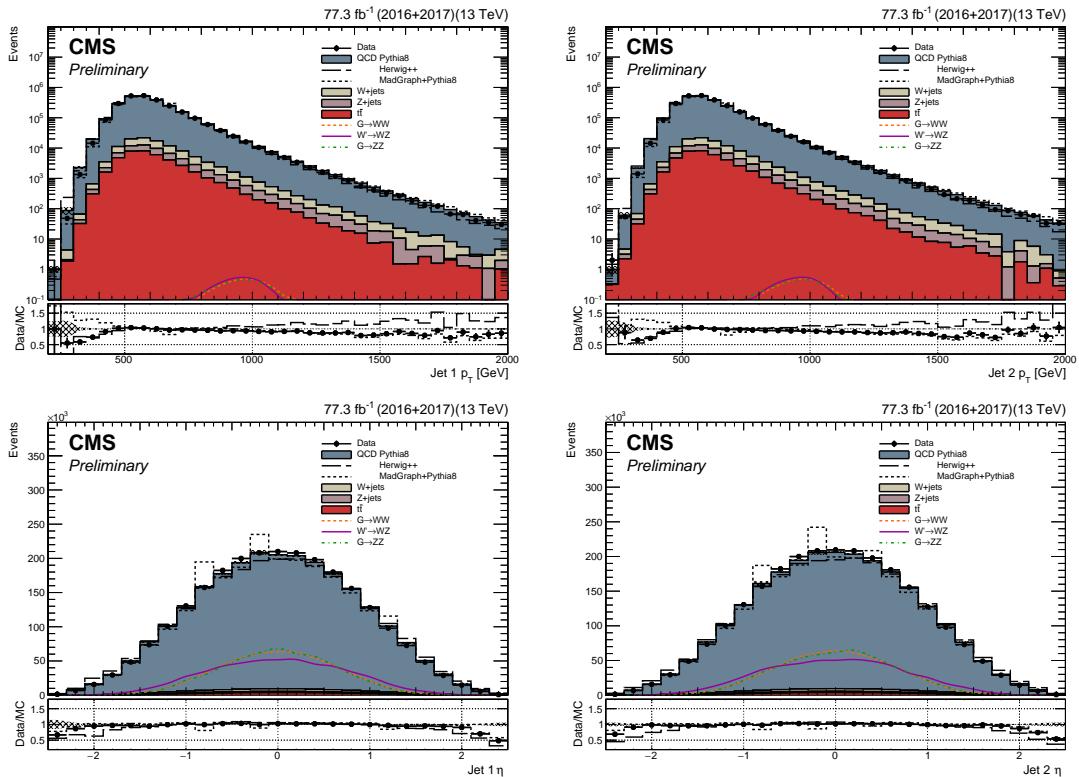


Figure 9.6: Jet p_T (top) and η (bottom) of the first (left) and second (right) selected jet in the event. The signal is scaled by an arbitrary number.

2439 9.5 Triggering

2440 The triggers used for the data collected in 2016 are the same as in Section 7.4.1, while the
 2441 thresholds in 2017 have increased in order to push the trigger rate to a level acceptable for the
 2442 increased luminosity. The triggers used for 2017 data are

- 2443 • HLT_PFHT1050
- 2444 • HLT_AK8PFJet500
- 2445 • HLT_AK8PFJet360/380/400/420_TrimMass30
- 2446 • HLT_AK8PFHT750/800/850/900_TrimMass50.

2447 As in Section 7.4.1, the trigger names indicate the primary selection criteria, with the terms “PF”,
 2448 “HT”, and “AK8” being previously explained. The term “JetXXX” refers to the lower value of
 2449 the p_T selection criteria of the jet, “HTXXX” refers to the HT requirement, and “TrimMassYY”
 2450 refers to the trimmed-jet mass selection applied to least one jet. For the results presented here,
 2451 the analysis threshold is set by the value of M_{VV} such that the combination of all triggers are
 2452 greater than 99% efficient. The trigger turn-on is evaluated in an orthogonal data set in which
 2453 the presence of a single muon with a p_T above 27 or 50 GeV is required. The trigger efficiency as
 2454 a function of dijet invariant mass using a combination of all triggers (left), and as a function of
 2455 the jet soft drop mass for the grooming-based triggers only (right) are shown in Figure 9.7. The
 2456 grooming-based triggers in the 2017 data set do not fully reach the trigger plateau as the triggers
 where unavailable during one data taking period, corresponding to 4.8 fb^{-1} . The corresponding

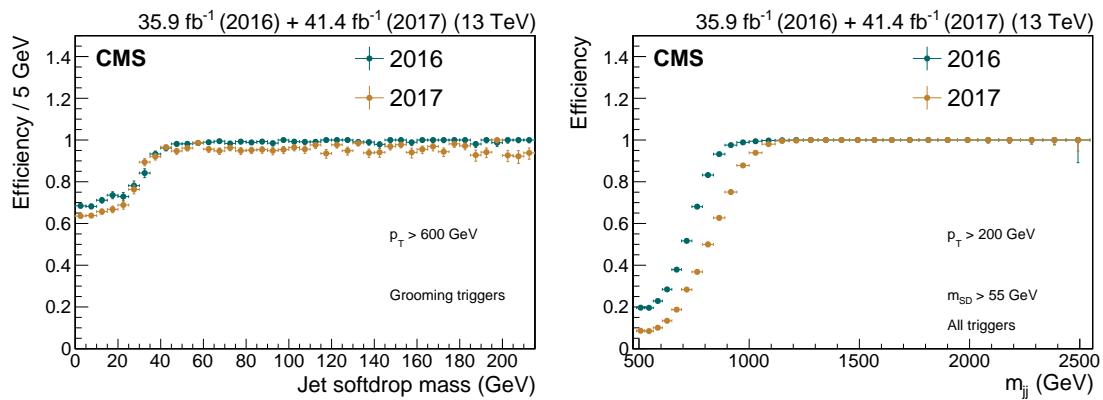


Figure 9.7: Left: trigger efficiency as a function of the dijet invariant mass using a combination of all analysis triggers. Right: trigger efficiency as a function of the jet softdrop mass for triggers requiring an online trimmed mass of at least 30 or 50 GeV.

²⁴⁵⁸ turn-on excluding the data taken without grooming-based triggers included, is shown in Figure 9.8.
²⁴⁵⁹ Here we see the that the trigger efficiency as a function of jet softdrop mass reaches the trigger plateau at the same point as in 2016. The combination of all triggers are > 99% efficient above

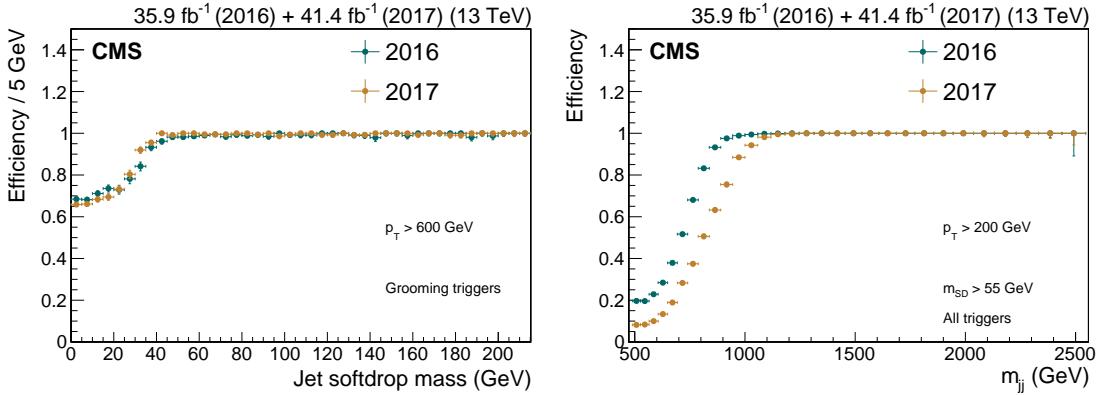


Figure 9.8: Left: trigger efficiency as a function of the dijet invariant mass using a combination of all analysis triggers. Right: trigger efficiency as a function of the jet softdrop mass for triggers requiring an online trimmed mass of at least 30 or 50 GeV. Here after excluding the data taken in 2017 without grooming-based triggers included.

²⁴⁶⁰
²⁴⁶¹ a dijet invariant mass of 1126 (990) GeV for the 2017 (2016) dataset. Since the point of full
²⁴⁶² efficiency is largest in the 2017 dataset, this sets the analysis threshold.

²⁴⁶³ 9.5.1 Trigger turn-on modeling

The threshold in the dijet mass at which resonances can be searched for depends on the value which the trigger can efficiently select events. The threshold used in the previous analyses was required to be high enough so that the dijet mass spectrum could be described with a smoothly falling function. Since the background modeling for this analysis does not depend on a smoothly falling dijet invariant-mass spectrum, as will be described in detail in Section 9.7, and in order to compensate for a loss in acceptance due to the increased trigger thresholds, we sought to model the trigger turn-on directly from data. This would allow us to search for diboson resonances with masses below the value at which the trigger is fully efficient. To model the trigger turn-on we derive a three dimensional histogram of the trigger efficiency versus dijet invariant mass (M_{VV}) and the groomed-jet mass of jet 1 and jet 2 (M_{jet1} and M_{jet2}), where each bin corresponds to the trigger efficiency for a small range of values of M_{VV} , M_{jet1} and M_{jet2} . The procedure is as follows. From the one-dimensional trigger turn-on histograms shown in Figure ??, values of M_{VV} , M_{jet1} and M_{jet2} at which the signal efficiency is above 99% are determined. For every bin above this

threshold, the trigger efficiency is fixed to one, corresponding to a 100% efficiency. For all bins below this threshold, we fit slices of M_{VV} in bins of M_{jet1} and M_{jet2} with a sigmoid function, determine the trigger efficiency to be used as a weight for simulated events to pass the trigger, and set the value from this function, and set the bin content of the three dimensional weight histogram accordingly. Since the trigger efficiency falls below 50 percent around a dijet invariant mass of 800 GeV, as seen from the right-hand plot in Figure ??, searching for resonances with masses below this point is not feasible due to low acceptance. In addition, the full signal shape needs to be contained within the dijet invariant mass spectrum, requiring the search to consider dijet invariant masses 10% lower than the actual value of the lowest resonance mass being probed. These two factors dictate that the minimum resonances mass we can search for is 1 TeV, and that the background modeling must begin around 900 GeV in order to fully contain the signal. Starting from a dijet invariant mass of $M_{VV} = 893$ GeV, the “dijet bin” closest to 900 GeV (where the “dijet binning” is as described in Section 7.5), and a jet mass of $M_{jet} = 40$ GeV, a coarsely binned three-dimensional histogram is filled with the fraction of events that pass one of the signal triggers,

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

From this three-dimensional coarse histogram, a finer binned histogram is achieved through interpolation. In bins of M_{jet1} and M_{jet2} , each slice in M_{VV} is fitted with a sigmoid function,

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

and the trigger weight is extracted in dijet invariant mass bins of 10 GeV. This fine binning in dijet invariant mass is sufficient to yield a smooth distribution after trigger reweighting, as will be demonstrated below, and no additional interpolation is done. Figure 9.9 shows the total projections on the M_{VV} , M_{jet1} and M_{jet2} axes for the full trigger weight histogram. The M_{jet} and

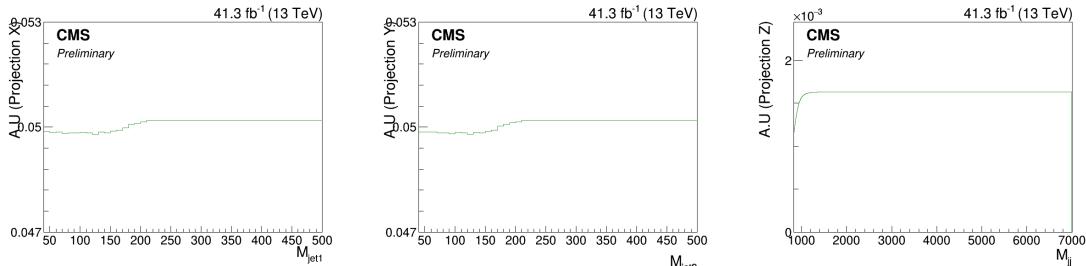


Figure 9.9: One-dimensional projections of the trigger weight histogram for M_{jet1} , M_{jet2} and M_{VV} , respectively.

2468 M_{VV} spectra for the signal sample with the lowest resonance mass, and for the QCD background
 2469 before and after trigger weights have been applied, are shown in Figure 9.10 and are compared to
 2470 data. A greater than 95% the signal efficiency is retained after the trigger weights have been
 applied, and the reweighted QCD simulation agrees well with data. The modeling of the trigger

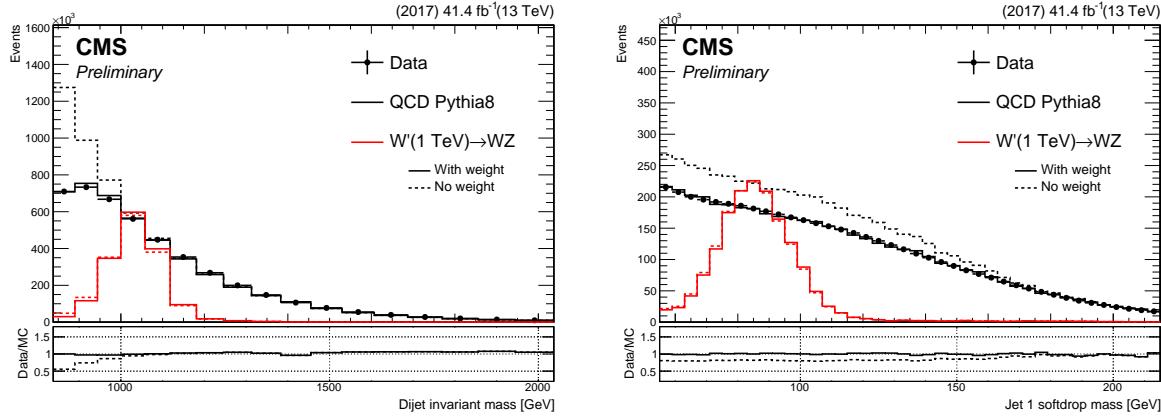


Figure 9.10: The M_{VV} (left) and M_{jet} (right) spectra for signal and background before and after trigger weights have been applied.

2471
 2472 turn-on was then implemented in the background fit method in the 3D analysis and we found
 2473 that the method could model the turn-on well. However, when studying the bias on the extracted
 2474 signal rate for a possible signal in this turn-on region, we found this to be large due to the
 2475 difficulties involved with fitting a signal peak on top of a background that is peaking. Since
 2476 the intention of the analysis presented here was to introduce the 3D fit method and allow it to
 2477 become available as soon as possible, we therefore abandoned the modeling of the trigger turn-on
 2478 for this paper. However, we still wish to pursue this strategy in the future.

²⁴⁷⁹ 9.6 A mass- and p_T -decorrelated tagger

²⁴⁸⁰ In order to identify W-jet and Z-jet candidates we utilize the softdrop algorithm and calculate
²⁴⁸¹ the n-subjettiness ratio τ_{21} for AK8 jets that are clustered with PUPPI constituents. As before,
²⁴⁸² the softdrop jet mass is used to improve the mass resolution of the jet, while n-subjettiness serves
²⁴⁸³ as a discriminant by yielding a probability of how compatible the jet is with having N axes. For
²⁴⁸⁴ this search we require the softdrop jet mass to be between $55 < m_{sd} < 215$ GeV, removing as
²⁴⁸⁵ much as possible of the low-mass QCD background and going high enough to encompass all
²⁴⁸⁶ searches for resonances decaying to W, Z, H or tops (the SM particles known to produce jets
²⁴⁸⁷ with substructure in the final state). Since the analysis is done in bins of jet mass and dijet
²⁴⁸⁸ invariant mass, the background must be modeled for a very wide range of mass and transverse
²⁴⁸⁹ momenta. It is therefore desirable that the QCD background spectrum is minimally sculpted as a
²⁴⁹⁰ function of p_T and mass, such that the background shapes in all regions are similar to one another
²⁴⁹¹ and remains smoothly falling in all three analysis dimensions and in all bins of M_{VV} , M_{jet1} and
²⁴⁹² M_{jet2} . In order to ensure minimal sculpting, we therefore decorrelate the τ_{21} variable from the
²⁴⁹³ jet softdrop mass and the jet p_T following what was done in Ref. [79], and briefly discussed in
²⁴⁹⁴ Section 8.5.2. This decorrelation is performed by flattening the τ_{21} profile dependence on ρ' ,
²⁴⁹⁵ where ρ' is defined by

$$\rho' = \log(m^2/p_T/\mu), \text{ where } \mu = 1 \text{ GeV.} \quad (9.1)$$

²⁴⁹⁶ Figure 9.11 shows the profile distribution of τ_{21} as a function of ρ' for QCD jets after applying
²⁴⁹⁷ the preselections listed above together with a softdrop mass selection of $55 < m_{sd} < 215$ GeV. A

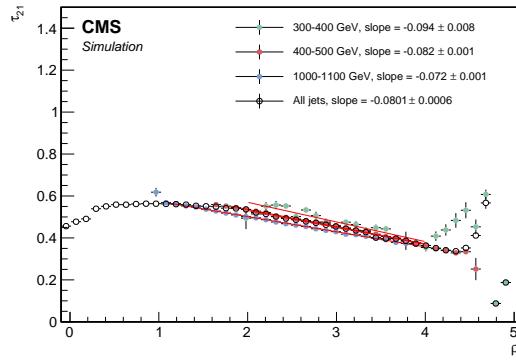


Figure 9.11: Profile distributions of τ_{21} as a function of $\rho' = \log(m^2/p_T/\mu)$ after applying a softdrop mass selection of $55 < m_{sd} < 215$ GeV.

2498 linear transformation is then defined as

$$\tau_{21}^{DDT} = \tau_{21} - M \times \rho', \quad (9.2)$$

2499 where the slope M is fitted from the linear part of the τ_{21} profile versus ρ' for the inclusive p_T
 2500 spectrum, shown in Figure 9.11. The resulting slope is $M = -0.080$. The profile of the retuned
 2501 τ_{21}^{DDT} versus ρ' is shown in Figure 9.12, exhibiting the desired flattened spectra for QCD jets
 versus ρ' . Which selections on τ_{21}^{DDT} to be used are chosen in the following way. First, we

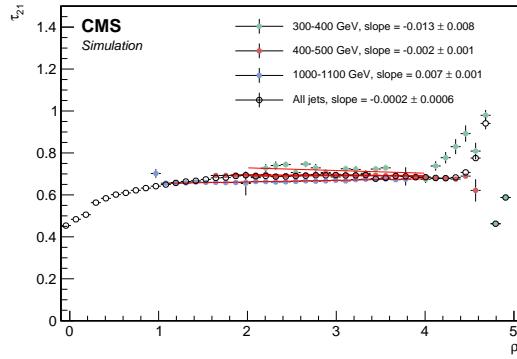


Figure 9.12: Profile distributions of τ_{21}^{DDT} as a function of $\rho' = \log(m^2/p_T/\mu)$, where $\mu = 1$ GeV.

2502
 2503 check which τ_{21}^{DDT} selection corresponds to the highest Punzi significance as a function of the
 2504 resonance mass for different signal samples, as shown in Figure 9.13. All other analysis selections,
 including a selection on the groomed-jet mass of $55 < m_{sd} < 215$ GeV, have been applied. The

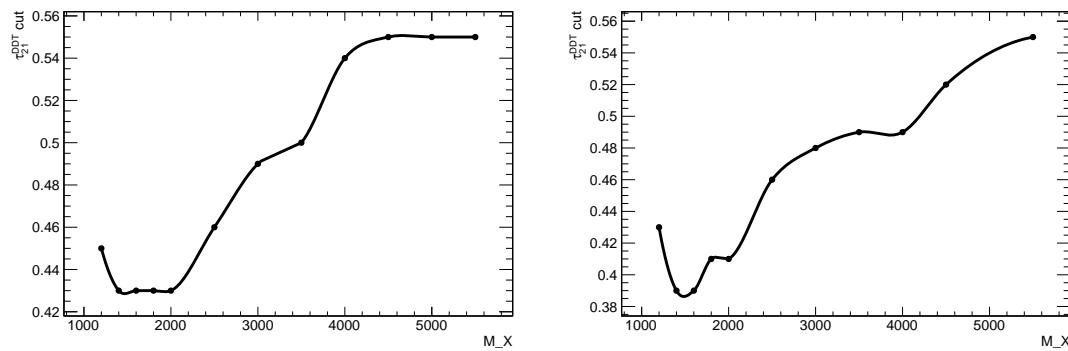


Figure 9.13: The τ_{21}^{DDT} cut corresponding to the highest Punzi significance for a given signal resonance mass, here for a Bulk $G \rightarrow WW$ (left) and a Bulk $G \rightarrow ZZ$ (right) signal.

2506 cut maximizing the Punzi significance at lower resonance mass, where the background is highest,
 2507 is chosen as the “high-purity” (HP) working point. This corresponds to $\tau_{21}^{DDT} \leq 0.43$. We then
 2508 proceed by finding which selections on τ_{21}^{DDT} contains at least 95% of the signal, and out of those
 2509 select the one that optimizes the Punzi significance. This is found to be $0.43 < \tau_{21}^{DDT} \leq 0.79$,
 2510 and is classified as the low purity (LP) category. The purpose of this category is to enhance the
 2511 overall sensitivity of the analysis, especially where the background is low.

2512 Figure 9.14 shows the performance of τ_{21} and τ_{21}^{DDT} (2016 and 2017 tune) in the background-
 2513 signal efficiency plane (top). We observe a significant gain in signal efficiency at a fixed mistag-
 2514 rate with the decorrelated taggers, when not using a tight jet groomed mass selection (recall: a
 2515 window of $55 \text{ GeV} < M_{SD} < 215 \text{ GeV}$ is used in this analysis). That is because we are taking
 2516 advantage of far more information when computing the DDT: subjetteness, and the ratio of
 2517 jet groomed mass and momentum. All these have a distribution which is different between
 2518 quark/gluon jets and W-jets, leading to a larger separation between signal and background,
 2519 as can be seen when comparing the distributions of the different taggers in the left plot of
 2520 Figure 9.15. The distribution of $\log(m_{SD}^2/p_T)$ is shown on the right in Figure 9.15, and it is
 clear that the variable adds discriminating power. In addition to cutting on τ_{21}^{DDT} , a selection on

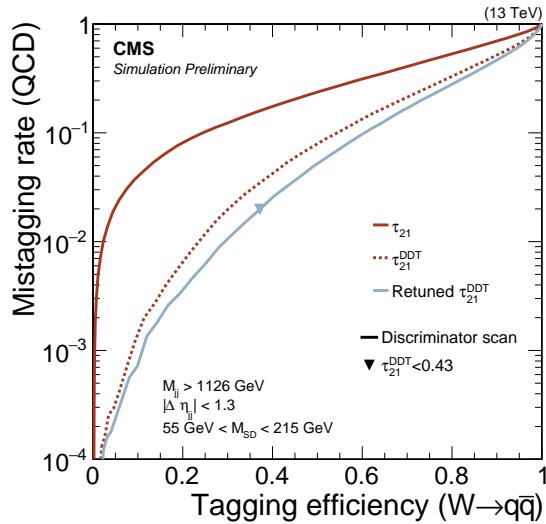


Figure 9.14: Performance of τ_{21} and τ_{21}^{DDT} (2016 and 2017 tune) in the background-signal efficiency plane.

2521
 2522 $\rho = \log(m^2/p_T^2) < -1.8$ is applied. This is in order to avoid configurations where the jet mass is
 2523 high, but the jet transverse momentum is low. In these cases, the AK8 cone size is too small
 2524 to contain the full jet, affecting both the jet mass resolution and the τ_{21}^{DDT} tagging efficiency.
 2525 These edge effects are difficult to model in simulation and we therefore avoid this region in the

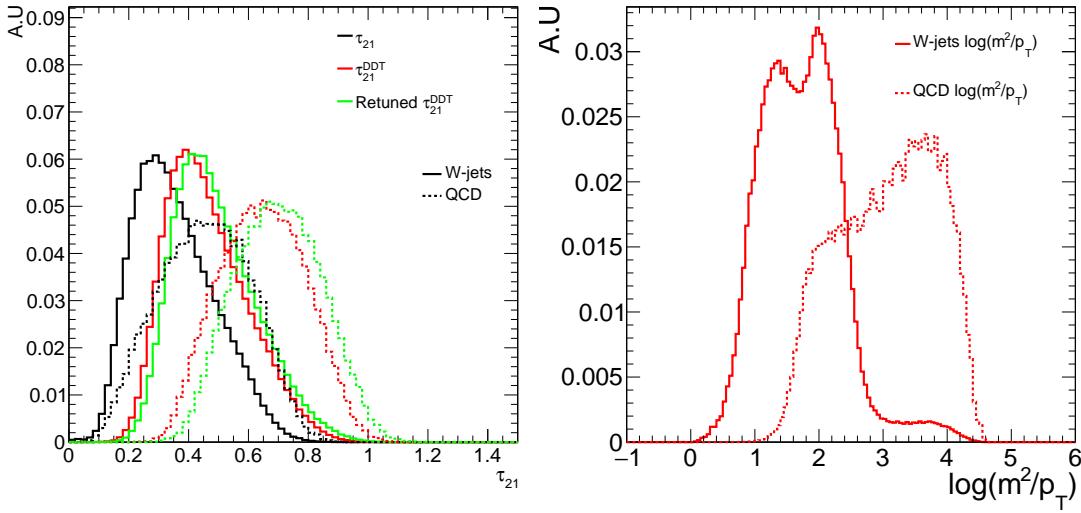


Figure 9.15: Distribution of τ_{21} and τ_{21}^{DDT} with the old and the new tune (left) and comparison of $\log(m^2/p_T)$ for signal and background jets (right).

following. Figure 9.16 shows the profile distribution of τ_{21}^{DDT} versus ρ for QCD jets, where a clear deviation from the observed linear spectra is observed above a ρ of -1.8. Note that this has a

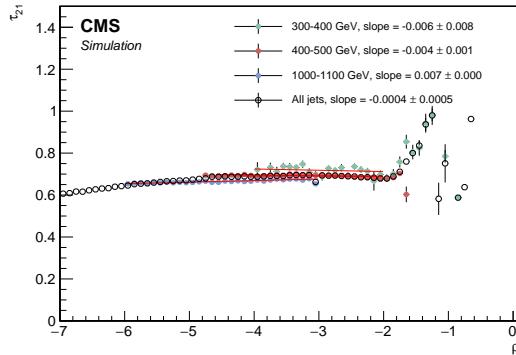


Figure 9.16: Profile distributions of τ_{21}^{DDT} as a function of $\rho = \log(m^2/p_T^2)$, where $\mu = 1$ GeV. The DDT breaks down above $\rho = -1.8$.

negligible effect on signal jets, which mainly peak around 80 GeV and have a high jet transverse momenta. Figure 9.17 shows the distributions of the PUPPI softdrop jet mass, the τ_{21}^{DDT} , and the τ_{21} variable for simulation and for data. The softdrop jet mass in the signal distribution peaks nicely around the W mass, while the multijets background spectrum is peaked at lower softdrop masses.

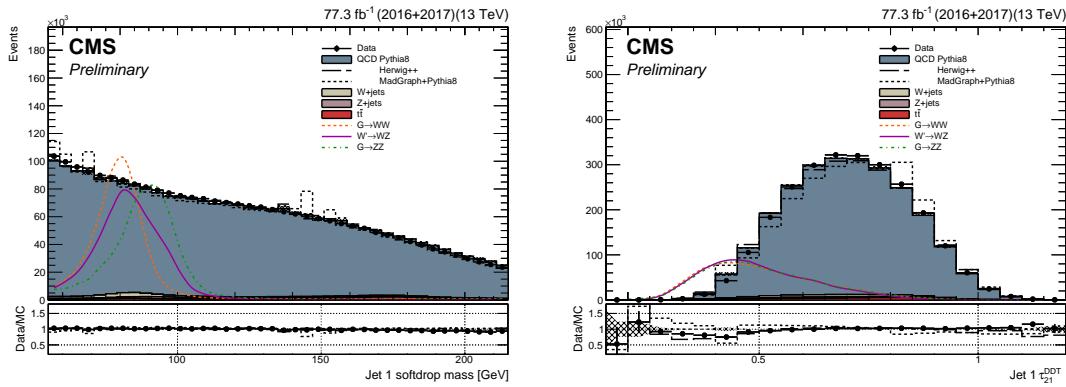


Figure 9.17: The softdrop jet mass (left) and τ_{21}^{DDT} (right) distribution in data and in simulation. Signal is scaled by an arbitrary number.

2533 9.6.1 Data to simulation scale factors

2534 Following what was done in Section 7.7 and 8.5.3, we derive W-tagging scale factors in order to
 2535 correct the efficiency in MC of the selection on τ_{21}^{DDT} , by estimating the ratio of the selection
 2536 efficiencies in data and in simulation. The softdrop jet mass range is fitted from 55 to 215 GeV,
 2537 and the two categories of purity are:

- 2538 • pass region: $0 < \tau_{21}^{DDT} \leq 0.43 \sim$ high-purity,
- 2539 • fail region: $0.43 < \tau_{21} \leq 0.79 \sim$ low purity.

2540 The obtained scale factors, jet-mass scale and resolution together with their statistical uncertainty
 2541 are listed in Tables 9.1 and 9.2 for 2016 and 2017 data, respectively. The corresponding
 2542 simultaneous fits are shown in Figure 9.18. The W-tagging efficiency in this region is lower
 2543 than what is expected for signal jets, around 7% compared to 35%. This is expected as the
 2544 measurement in $t\bar{t}$ is dominated by W-jets with a p_T of around 200 GeV, just where the W decay
 2545 products merge into a single jet, whereas the signal jets mostly have a transverse momentum
 2546 above 600 GeV. The signal efficiency for τ_{21}^{DDT} increases at a function of jet p_T , as can be seen
 2547 from Figure 9.26, whereas the background efficiency is flat by design. In addition to the statistical
 2548 uncertainty, two systematic uncertainties are evaluated: one due to generator differences and one
 2549 due to NNLO corrections. The former is evaluated by comparing the resulting scale factors when
 2550 using $t\bar{t}$ simulation produced with different generators. The latter is evaluated by comparing the
 2551 extracted efficiency with and without top- p_T reweighting. This weight is derived from data in
 2552 order to better describe the observed p_T distribution in $t\bar{t}$ and is calculated for each top jet as
 2553 $w = \exp^{0.0615 - 0.0005 * p_{T,top}}$. The W tagging scale factors, jet mass scale, and jet mass resolution,
 2554 with their total uncertainty after adding systematics, are listed in Table 9.3. As before, the scale
 2555 factor is used to scale the signal yield and the jet mass scale and resolution are used to smear the
 2556 MC. The difference in jet mass scale between data and simulation is around 2%, and the jet mass
 2557 resolution scale factor is roughly 8%.

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

Table 9.1: Jet mass scale, jet mass resolution, and τ_{21}^{DDT} scale factors, as evaluated in the full 2016 single-muon dataset.

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

Table 9.2: Jet mass scale, jet mass resolution, and τ_{21}^{DDT} scalefactors as evaluated in the full 2017 single-muon dataset.

	SF $\pm \sqrt{\text{Stat.} + \text{Sys}_{\text{Generator}} + \text{Sys}_{\text{NNLO}}}$	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 9.3: Final jet mass scale, jet mass resolution, and τ_{21}^{DDT} scalefactors with statistical and systematic uncertainties.

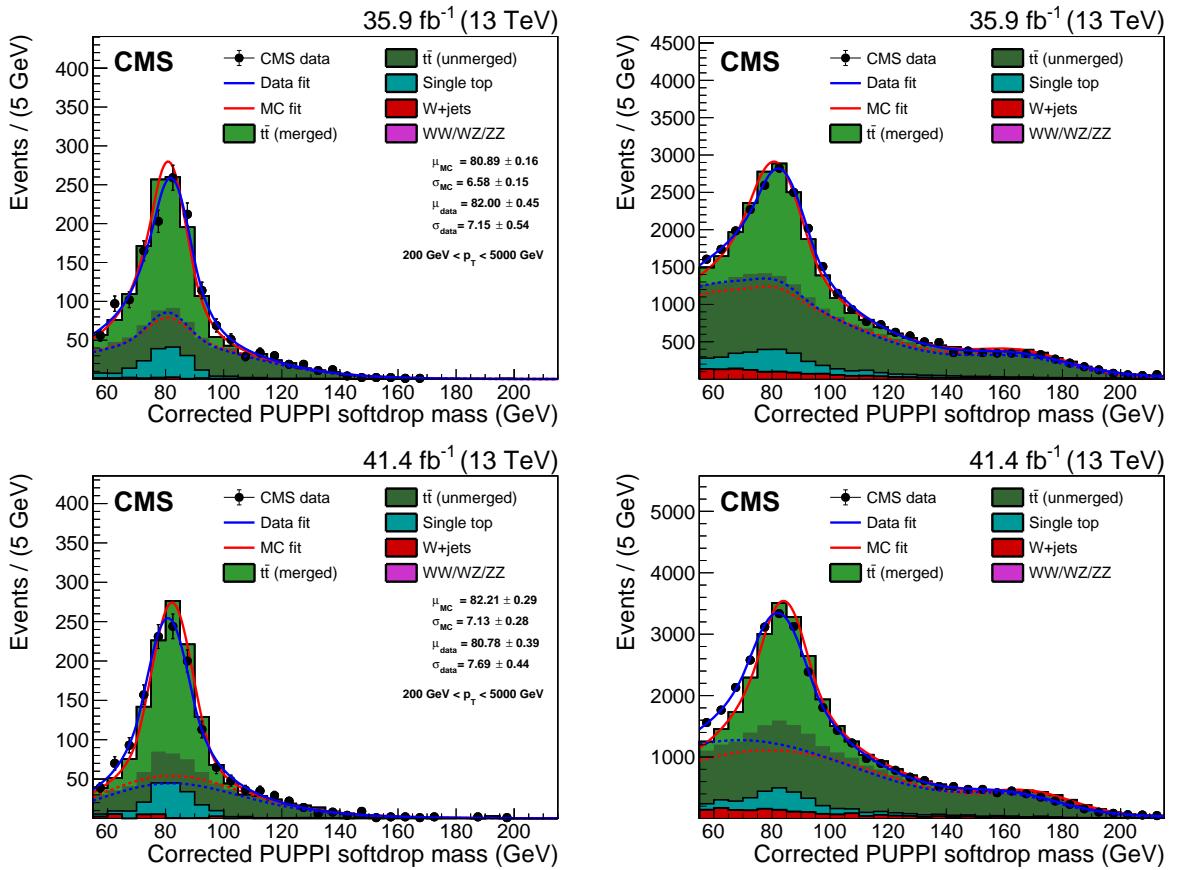


Figure 9.18: PUPPI softdrop jet mass distribution for events that pass (left) and fail (right) the $\tau_{21}^{DDT} < 0.43$ selection in the $t\bar{t}$ control sample. The result of the fit to data and simulation is shown by the solid blue and solid red line, respectively. The background components of the fit are shown as dashed-dotted lines. The fit to 2016 data is shown in the upper panels and the fit to 2017 data in the lower panels.

2558 9.6.2 p_T -dependence

2559 The τ_{21}^{DDT} selection criteria used for this search is restrictive to favor better signal significance, but
 2560 results in small event samples of the data and simulation needed to produce the scale factors. In
 2561 order to determine the dependence of the W-tagging efficiency, and the PUPPI softdrop jet-mass
 2562 scale and resolution on the transverse momentum of the jets, we divide the sample into 4 different
 2563 jet- p_T bins using the looser softdrop + τ_{21} tagger, and compute the scale factors per bin. The
 2564 systematic uncertainties are evaluated the same way as above (one due to top- p_T reweighting and
 2565 one comparing different $t\bar{t}$ samples). Figure 9.19 shows the extracted W-tagging efficiency for
 2566 data (black markers) and for simulation (red markers) using a tagger based on PUPPI softdrop
 2567 with a requirement of $\tau_{21} < 0.4$, as a function of jet p_T . The inclusive efficiency measurement
 2568 is marked with triangles. The lower panel shows the efficiency ratio of data over simulation,
 2569 corresponding to the W-tagging scale factor. All scale factors are compatible with unity, but
 2570 the uncertainty on the measurement grows as statistics decrease. The corresponding extracted
 scale factors are listed in Table 9.4. With the observation of this clear trend of an uncertainty

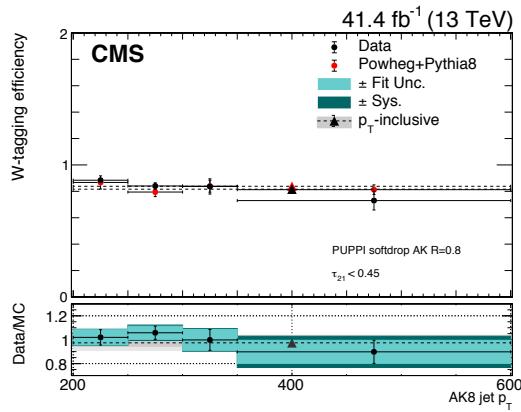


Figure 9.19: The W-tagging efficiency in data (black circles) and in simulation (red circles) as a function of jet p_T . The p_T -inclusive measurement is marked with triangles. The lower panel shows the efficiency in data divided by the efficiency in simulation, corresponding to the W-tagging uncertainty. The blue bands mark the fit and systematic uncertainties.

2571
 2572 increase as a function of p_T , we evaluate a p_T -dependent W-tagging scale factor uncertainty in the
 2573 following way: using signal Monte Carlo generated with two different shower generators, PYTHIA8
 2574 and HERWIG++, we compute the difference in tagging efficiency between the two samples at
 2575 low- p_T , where we have a real measurement in data, and compare that to the difference in tagging

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

Table 9.4: The data to simulation scale factors in bins of jet p_T for a tagger based on PUPPI softdrop with a requirement of $\tau_{21} < 0.4$. All scalefactors are compatible with unity.

efficiency between the two at high- p_T . In other words, we take a double ratio

$$\sigma_{p_T, \text{Bin}=i} = \frac{\left(\frac{\epsilon_{\text{HERWIG}}}{\epsilon_{\text{PYTHIA}}}\right)_{p_T, \text{Bin}=i}}{\left(\frac{\epsilon_{\text{HERWIG}}}{\epsilon_{\text{PYTHIA}}}\right)_{p_T=300 \text{ GeV}}} - 1, \quad (9.3)$$

where $\sigma_{p_T, \text{Bin}=i}$ is the uncertainty on the scale factor on p_T bin i . In contrast with what was found for the τ_{21} -based tagger in Section 7.7.3, where this uncertainty grew logarithmically with p_T , we find that the corresponding double ratio stabilizes around 1 TeV in the high-purity category when using τ_{21}^{DDT} , and it can be approximated as a flat 28%. The uncertainty for the low-purity category is still well described by a logarithmic function. The uncertainty σ_{p_T} and resulting parametrization is shown in Figure 9.20. In my personal opinion, this way of defining

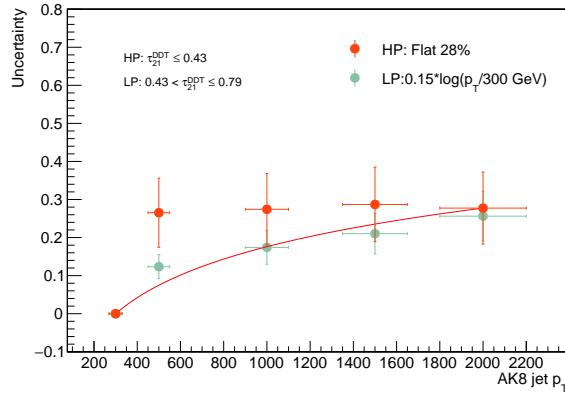


Figure 9.20: The parametrized uncertainty on the W-tagging efficiency scale factor as a function of the jet p_T .

an uncertainty is rather random and ad hoc. We already know there is a large difference in the modeling of substructure variables between generators, and there is no good reason to believe the difference between PYTHIA8 and HERWIG++ provides a reasonable description of the magnitude

2586 of the uncertainty. In an analysis using two jets this is a very large and overly conservative
2587 systematic uncertainty that is impossible to constrain from data. I would rather prefer to use
2588 the measured scale factor uncertainties in data and extrapolate these, but the above method
2589 was used in previous searches and is therefore the method of choice for the Jet Physics Object
2590 Group at CMS. I will therefore continue using these uncertainties for the presented analysis, but
2591 will attempt to change this for future searches. In addition to measuring the dependence of the
2592 tagging efficiency on jet p_T , we also extract the change in softdrop jet-mass scale and resolution
2593 as a function of p_T . We find that the jet-mass scale in data divided by MC ranges between 0.5
2594 and 2.5% for the p_T bins considered, and the jet-mass resolution varies between 4 and 10%, the
2595 latter measurement not being statistically significant as the uncertainties are large, around 10%.
2596 For this analysis, we therefore use a fixed uncertainty of 2 and 10% for the softdrop jet-mass scale
2597 and resolution, respectively, to cover any broadening or shift at high p_T not described by the
2598 simulation.

2599 9.7 The multidimensional fit

2600 The three-dimensional fit method is based on the assumption that the signal peaks in three
 2601 dimensions: the dijet invariant mass (M_{VV}), the groomed jet mass of the first jet (M_{jet1}), and
 2602 the groomed jet mass of the second jet (M_{jet2}), in order to extract a possible signal from the
 2603 three-dimensional M_{jet1} - M_{jet2} - M_{VV} (x,y,z) plane.

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

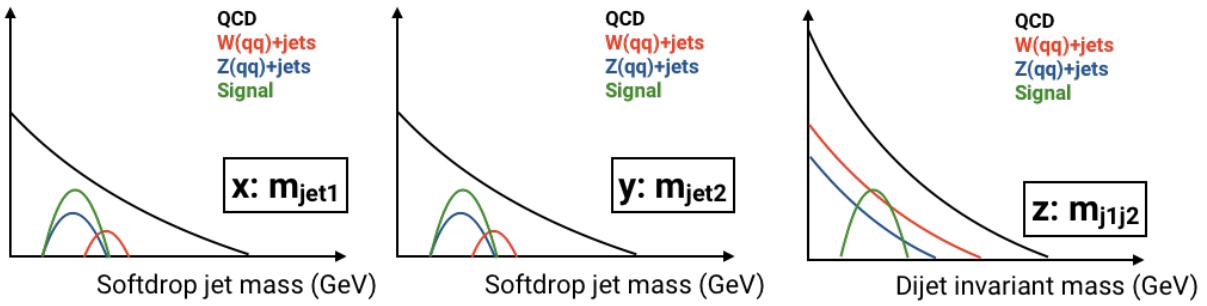


Figure 9.21: An illustration of the shape of the signal and the relative background contributions in the three dimensions $M_{jet1}(x)$, $M_{jet2}(y)$ and $M_{VV}(z)$.

2604 functions (PDF's) need to be derived:

- 2606 • **signal:** a PDF resonant in M_{jet1} , M_{jet2} and M_{VV} ,
- 2607 • **non-resonant background:** PDF for the dominant QCD multijet background. Non-
2608 resonant in M_{jet1} , M_{jet2} and M_{VV} ,
- 2609 • **resonant background:** PDF describing of W+jets and Z+jets. Resonant in M_{jet1} and
2610 M_{jet2} , but smoothly falling in M_{VV} , and
- 2611 • **alternative shapes:** five additional shape uncertainties for the non-resonant background.

2612 These are illustrated in Figure 9.21 and will be described in detail in the following.

2613 9.7.1 Modeling of the signal

2614 The signal shape in three dimensions is defined as the product of the resonance mass shape and
 2615 the jet mass shapes:

$$P_{sig}(M_{VV}, M_{jet1}, M_{jet2} | \theta(M_X)) = P_{VV}(M_{VV} | \theta_1(M_X)) \times P_{j1}(M_{jet1} | \theta_2(M_X)) \times P_{j2}(M_{jet2} | \theta_2(M_X)). \quad (9.4)$$

2616 The shapes for M_{VV} , M_{jet1} and M_{jet2} all depend on the hypothesized mass of the new particle
 2617 (M_X) and a set of parameters $\theta = (\theta_1, \theta_2)$ that in principle depend on M_X . The modeling of M_{jet}
 2618 is done separately for the high-purity and the low-purity category, while the modeling of M_{VV}
 2619 is done with no selection on τ_{21}^{DDT} applied (M_{VV} was found to be decorrelated from τ_{21}^{DDT} for
 2620 signal jets).

2621 The signal is parametrized by fitting the resonance mass and jet-mass line shapes for each
 2622 resonance mass value, extracting the fitted parameters and then interpolating these as a function
 2623 of the resonance-mass hypothesis. For the resonance mass M_{VV} , the sum of a crystal-ball
 2624 function and a Gaussian shape is used for each mass point, following the shapes used in Search
 2625 II. Figure 9.22 shows the derived parameters and interpolation as a function of resonance mass,
 and the final M_{VV} shapes as extracted from the parametrization are shown in Figure 9.23.

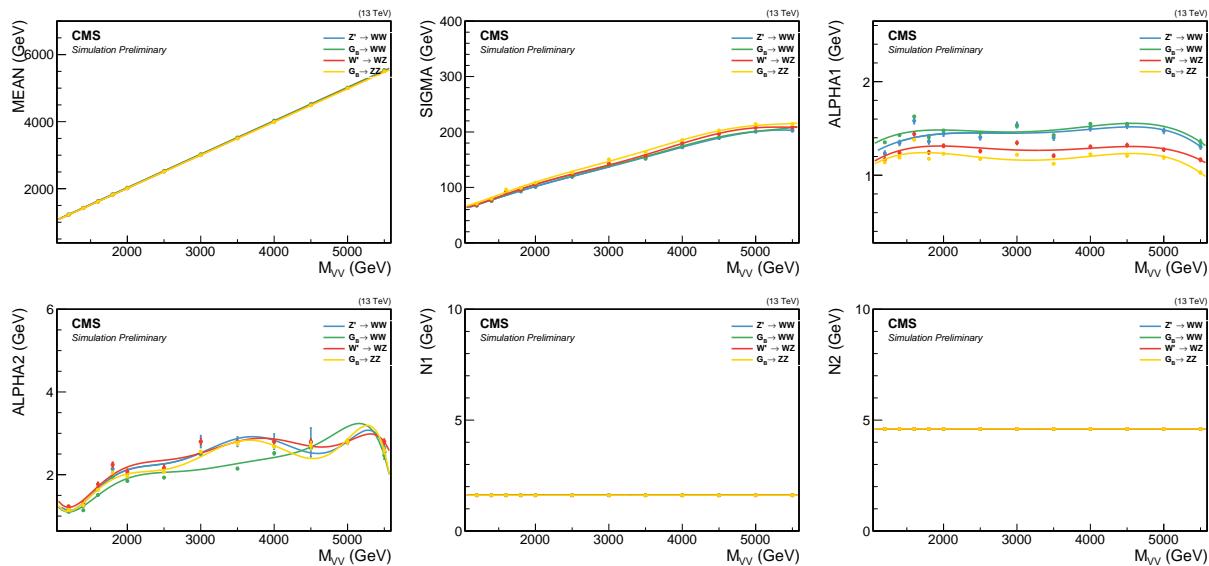


Figure 9.22: The interpolated fit parameters of a crystal-ball fit to the dijet invariant mass as a function of M_X . The small variations for ALPHA2 have been shown to have no effect on the overall modeling.

2626

2627 The same procedure is used to model the jet mass; the M_{jet} spectrum for each resonance mass

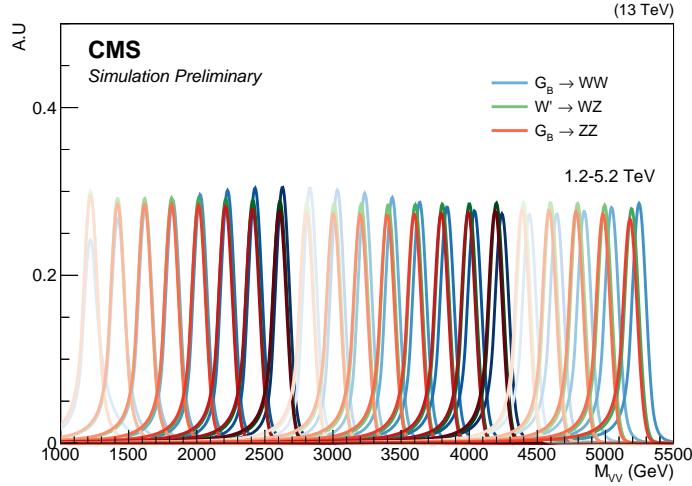


Figure 9.23: Final M_{VV} signal shapes extracted from the parametrization. Here for a G_{bulk} decaying to WW (blue), ZZ (red), and for a W' decaying to WZ (green).

2628 hypothesis is fitted using a double Crystal-ball function, and the fitted parameters are extracted
 2629 and interpolated as a function of the resonance mass. This is done separately for $M_{\text{jet}1}$ and $M_{\text{jet}2}$.
 2630 The fitted parameters and interpolations are shown in Figure 9.24 for $M_{\text{jet}1}$, and the corresponding
 distributions for $M_{\text{jet}2}$ can be found in Appendix D.1. The final $M_{\text{jet}1}$ shapes, as extracted from

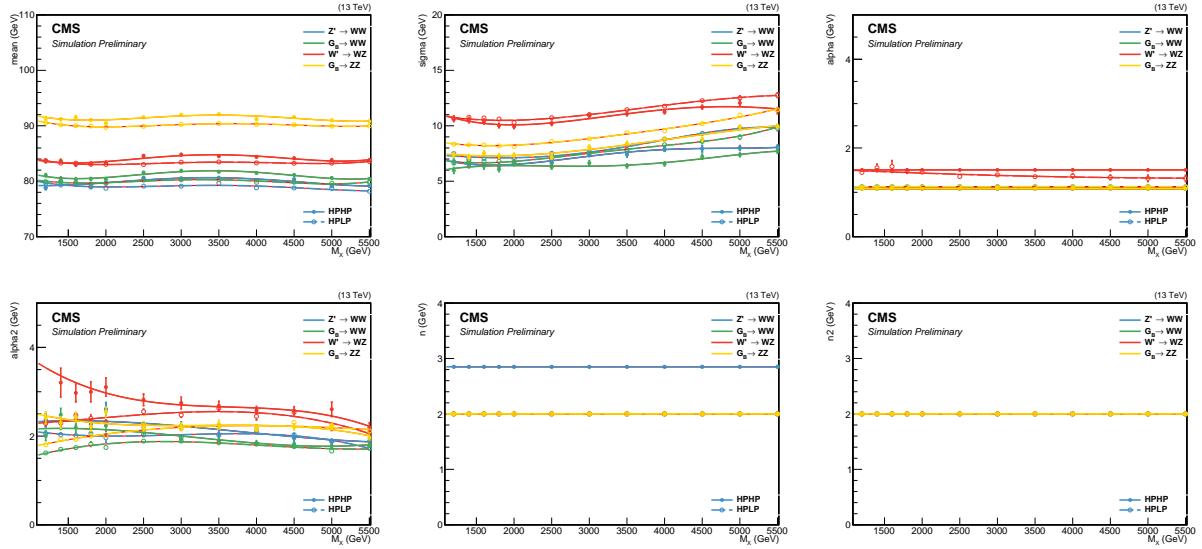


Figure 9.24: The interpolated double Crystal-ball parameters for the softdrop jet mass as a function of M_X . To improve the stability of the fit some parameters are set constant.

the parametrization are shown in Figures 9.25. Finally, the signal yield is parametrized as a

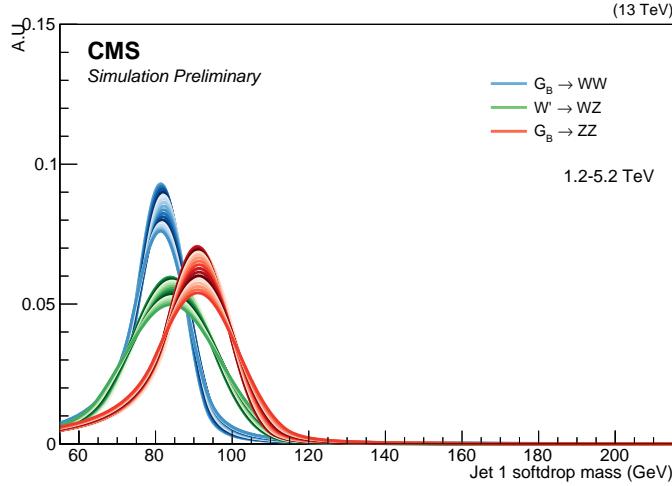


Figure 9.25: Final M_{jet} signal shapes extracted from the parametrization for a G_{bulk} decaying to ZZ (red), a G_{bulk} decaying to WW (blue), and for a W' decaying to WZ (green).

2632 function of the resonance mass. For each mass point M_X and each purity category, the signal
 2633 function of the resonance mass. For each mass point M_X and each purity category, the signal
 2634 yield per picobarn of cross section is calculated as the integral of the resulting Monte Carlo
 2635 histogram after all analysis selections have been applied. The yields are then interpolated as a
 function of M_X . The signal efficiency as a function of resonance mass is shown in Figure 9.26.

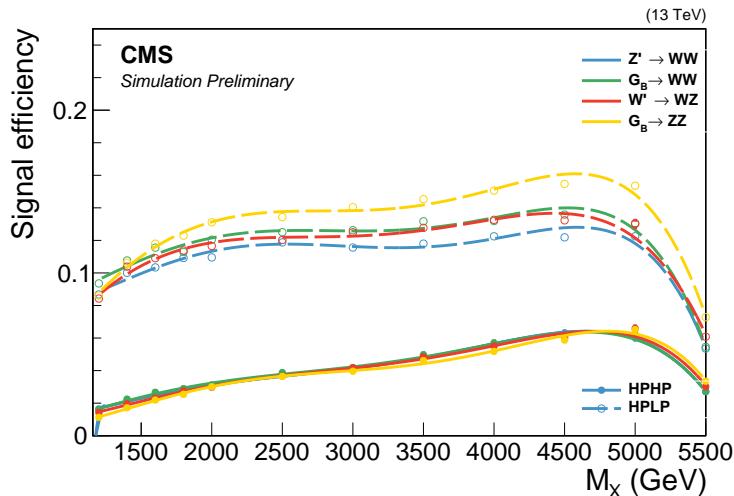


Figure 9.26: Signal efficiency as a function of resonance mass.

2637 9.7.2 Modeling of the non-resonant background

2638 In order to model the QCD multijets background in the three-dimensional M_{VV} - M_{jet1} - M_{jet2}
 2639 plane, we use the following conditional product:

$$P(M_{VV}, M_{jet1}, M_{jet2}) = P_{VV}(M_{VV}|\theta_1) \times P_{cond,1}(M_{jet1}|M_{VV}, \theta_2) \times P_{cond,2}(M_{jet2}|M_{VV}, \theta_2). \quad (9.5)$$

2640 This probability density requires a computation of the conditional two-dimensional shapes of
 2641 M_{jet1} or M_{jet2} , given M_{VV} , as well as a one dimensional shape of the M_{VV} distribution. The
 2642 reason for requiring a conditional jet-mass shape, is due to the high correlation between the jet
 2643 mass and the dijet invariant mass for quark and gluon jets. The modeling is done separately for
 2644 the high-purity and the low-purity category.

2645 The following fit ranges are used for the three axes: the jet masses M_{jet1} and M_{jet2} are fitted
 2646 from 55 to 215 GeV using 2 GeV bins, and the dijet invariant mass M_{VV} is fitted from 1126 to
 2647 5500 GeV. The lower M_{VV} bound is chosen to avoid complications in the fitting procedure due
 2648 to trigger turn-on effects, while the upper M_{VV} bound is chosen by considering the highest dijet
 2649 invariant mass event found in data. For M_{VV} , the “dijet binning”, corresponding to the dijet
 2650 mass resolution, is used. In units of GeV, the binning is as follows:

2651
 2652 Dijet binning = 1126, 1181, 1246, 1313, 1383, 1455, 1530, 1607, 1687, 1770, 1856, 1945,
 2653 2037, 2132, 2231, 2332, 2438, 2546, 2659, 2775, 2895, 3019, 3147, 3279, 3416, 3558, 3704,
 2654 3854, 4010, 4171, 4337, 4509, 4686, 4869, 5058, 5253, 5500.

2655
 2656 The background model is built starting from simulation using a “forward-folding” approach. To
 2657 build the one-dimensional template for the dijet invariant mass, $P_{VV}(M_{VV}|\theta_1)$, a 1D Gaussian
 2658 kernel is built starting from generator level quantities where, for each MC event, a Gaussian con-
 2659 tributes to the total one-dimensional probability density function. To build the two-dimensional
 2660 conditional templates, $P_{cond,1}(M_{jet1}|M_{VV}, \theta_2)$ and $P_{cond,2}(M_{jet2}|M_{VV}, \theta_2)$, a two-dimensional
 2661 Gaussian kernel is created for each MC event, where each 2D kernel contributes to the to-
 2662 tal conditional PDF.

2663 In order to define the Gaussian kernel, the dijet invariant mass scale and resolution, and the
 2664 softdrop jet-mass scale and resolution, must be derived. This is done by comparing the generated
 2665 jet-mass $M(\text{gen})$ to the reconstructed jet-mass $M(\text{reco})$. The M_{jet} and M_{VV} scale and resolution
 2666 are extracted from a Gaussian fit to either $M_{jet}(\text{reco})/M_{jet}(\text{gen})$ or $M_{VV}(\text{reco})/M_{VV}(\text{gen})$, in
 2667 bins of generator-level jet p_T . Figure 9.27 shows the fit to $M_{jet}(\text{reco})/M_{jet}(\text{gen})$ (left) and
 2668 $M_{jj}(\text{reco})/M_{jj}(\text{gen})$ (right) for an arbitrary p_T bin. The mean of the Gaussian yields the jet-mass
 2669 scale and the width yields the jet-mass resolution, for a given bin in generator-level jet p_T .

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

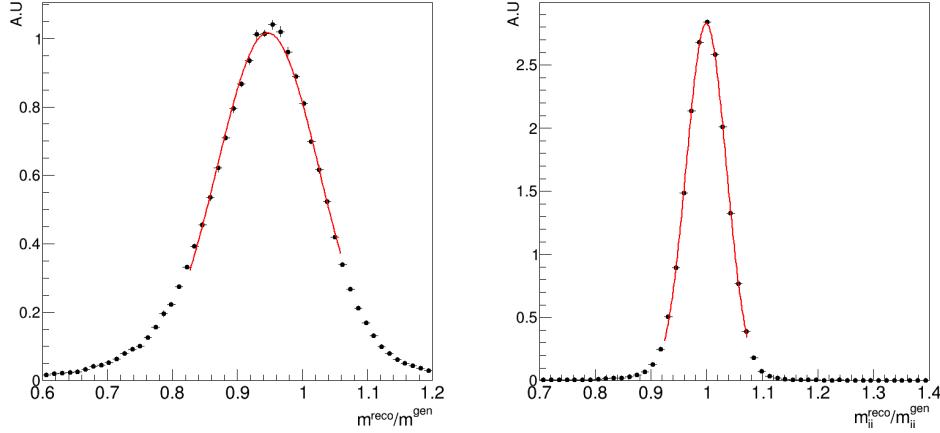


Figure 9.27: Left: fit to $M_{\text{jet}}(\text{reco})/M_{\text{jet}}(\text{gen})$. Right: fit to $M_{\text{VV}}(\text{reco})/M_{\text{VV}}(\text{gen})$. The mass resolution is taken as the width of the fitted Gaussian, while the Gaussian mean yields the mass scale.

2670
2671 Figure 9.28, and the total projection of these resolution functions are shown in Figure 9.29.
2672 The jet-mass and the dijet invariant mass scale and resolution are then used to populate the
2673 conditional 2D histogram as follows. Each generated event is smeared with a 2D Gaussian kernel
2674 of the form

$$k(M_{\text{jet}}, M_{\text{VV}}) = \frac{w_i}{\sqrt{2\pi} r_{M_{\text{VV}},i} \cdot r_{M_{\text{jet}},i}} \exp \left(-\frac{1}{2} \left(\frac{M_{\text{VV}} - s_{M_{\text{VV}},i}}{r_{M_{\text{VV}},i}} \right)^2 - \frac{1}{2} \left(\frac{M_{\text{jet}} - s_{M_{\text{jet}},i}}{r_{M_{\text{jet}},i}} \right)^2 \right), \quad (9.6)$$

2675 where s_i, r_i are the scale and the resolution derived in the previous step and w_i is the event weight
2676 product due to effects such as PU-reweighting and cross-sections. The resulting kernel values are
2677 then filled into a 2D histogram. This procedure is performed separately for $M_{\text{jet}1}$ and $M_{\text{jet}2}$. To
2678 build the one-dimensional template for the dijet invariant mass, the same procedure as above is
2679 used with the exception that the smearing is done with a one-dimensional Gaussian kernel only
2680 depending on M_{VV} . The three templates are then added together to form a three-dimensional
2681 PDF, and this PDF is fitted to QCD MC in order to remove any residual bias. The result is a
2682 smooth shape, covering the entire search range, that can be used in place of the prediction from
2683 simulation.

2684 As the high-purity category is limited by statistics, we instead build this template starting
2685 from the low-purity templates. This is done by fitting the low-purity 3D template, obtained using
2686 the method described above, to QCD MC in the high-purity category. Figure 9.30 shows the final

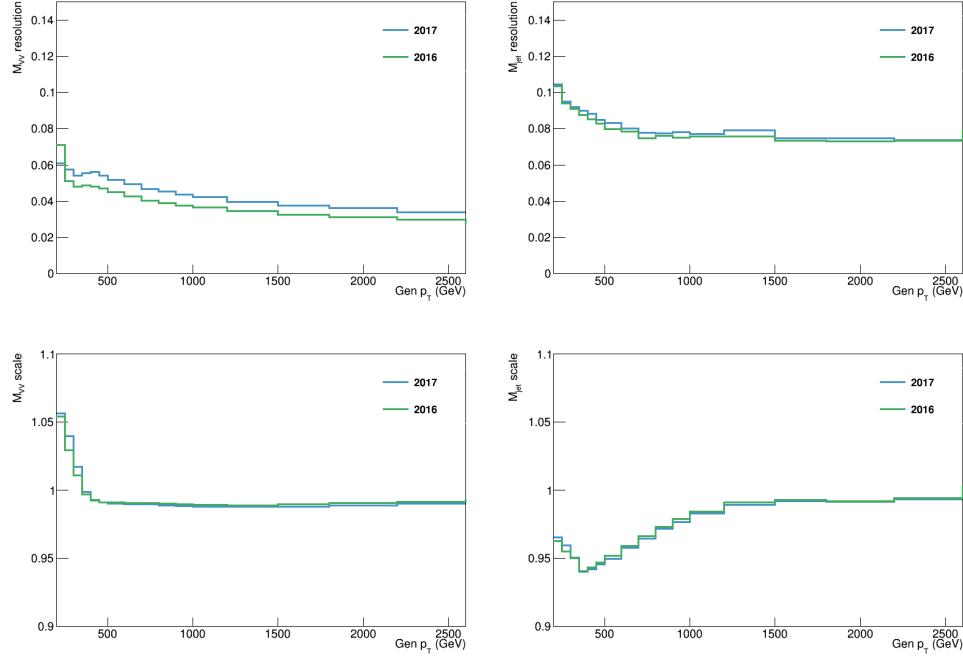


Figure 9.28: The M_{VV} resolution (top left) and scale (bottom left), and the M_{jet} resolution (top right) and scale (bottom right) as a function of generator-level jet p_T .

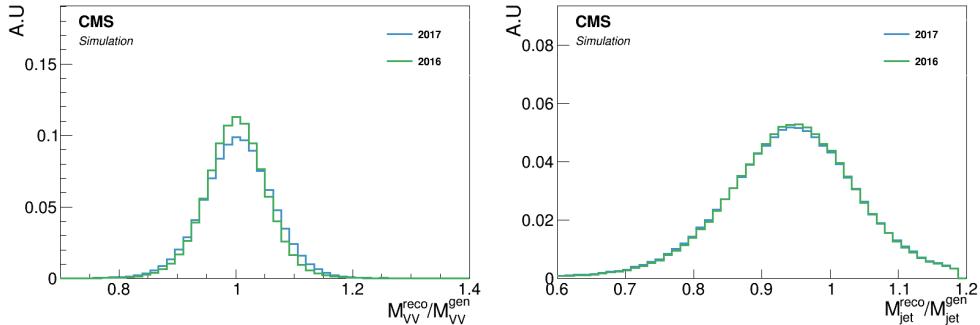


Figure 9.29: Projections of the resolution functions for all generator-level jet p_T bins for M_{VV} (left) and M_{jet} (right).

2687 templates (solid lines) together with the QCD MC (data points) derived from the 2017 MC in
 2688 the low-purity (top) and high-purity (bottom) category. Good agreement between simulation and
 2689 templates in all three dimensions is observed, within statistical uncertainties. The corresponding
 2690 distributions for 2016 MC can be found in Appendix D.2.

2691 In order to validate the kernel-transfer method, we check that we can fit a higher-statistics

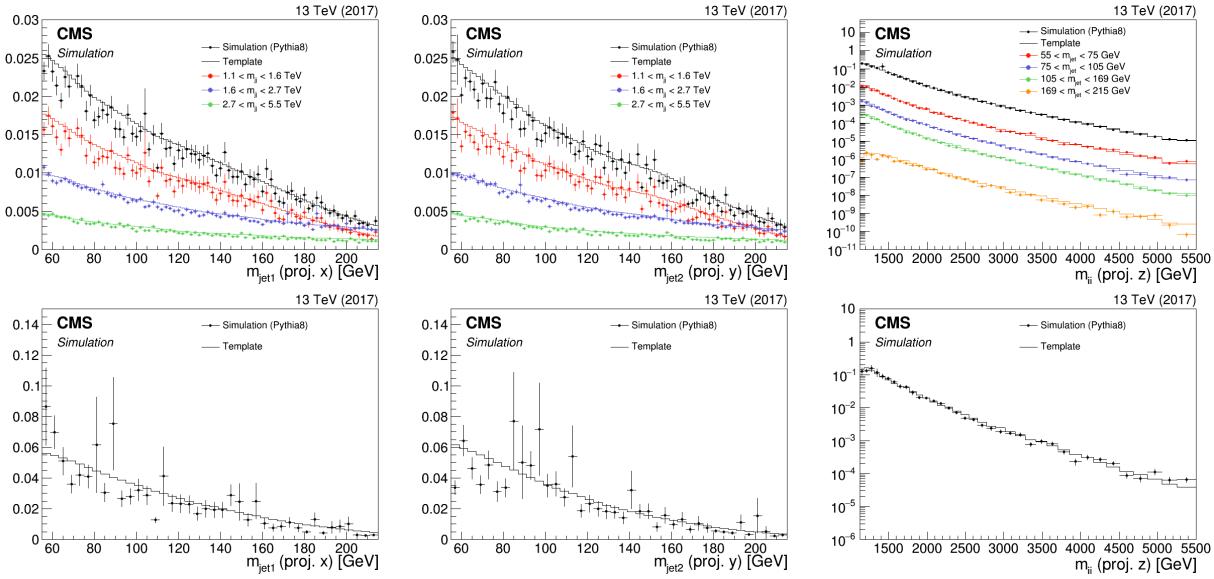


Figure 9.30: Comparison between QCD MC simulation (markers) and kernels derived from generator-level quantities (lines) for the HP (top) and LP (bottom) categories using 2017 MC. The kernels are shown for $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right).

2692 high-purity region by loosening the τ_{21}^{DDT} cut to 0.49. This results in 12 times more background
 2693 events, and should uncover whether any degeneracy is present in the fits themselves and whether
 2694 the low-purity kernel indeed is capable of modeling the high-purity region. The resulting kernel-
 2695 versus-MC spectra are shown in Figure 9.31. Good closure is observed in all three dimensions,
 2696 demonstrating that the LP kernels adapt well to the HP MC data points even when statistics are
 sufficient, and we therefore consider the method to be sound.

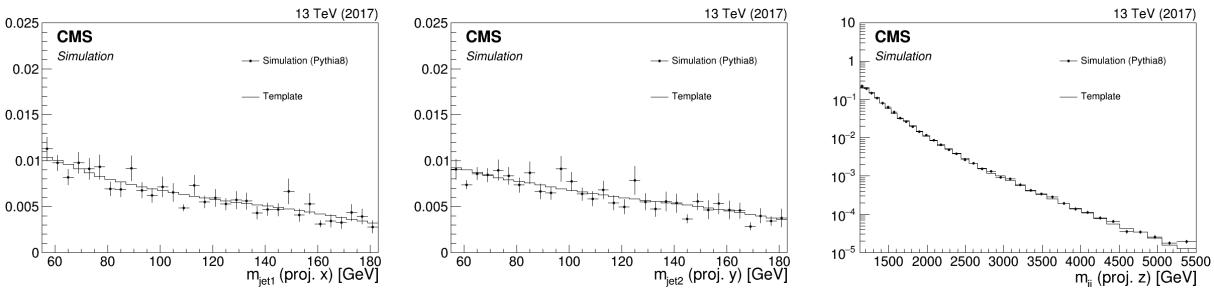


Figure 9.31: Comparison between QCD MC simulation (markers) and kernels derived from generator-level quantities (lines) in the HP category, using a looser cut on τ_{21}^{DDT} .

2698 9.7.3 Modeling of the resonant background

In addition to the QCD multijet background, there are a few sub-dominant processes which contain one real vector boson and at least one QCD-jet that must be accounted for. These are W+jets, Z+jets and events from $t\bar{t}$ processes. These events are resonant in M_{jet1} and M_{jet2} and must therefore be treated differently than the non-resonant QCD background. Figure 9.32 shows the projections on M_{jet1} (left), M_{jet2} (middle) and M_{VV} (right) for the resonant backgrounds under consideration. As the jets are randomly sorted, each jet mass distribution contains two

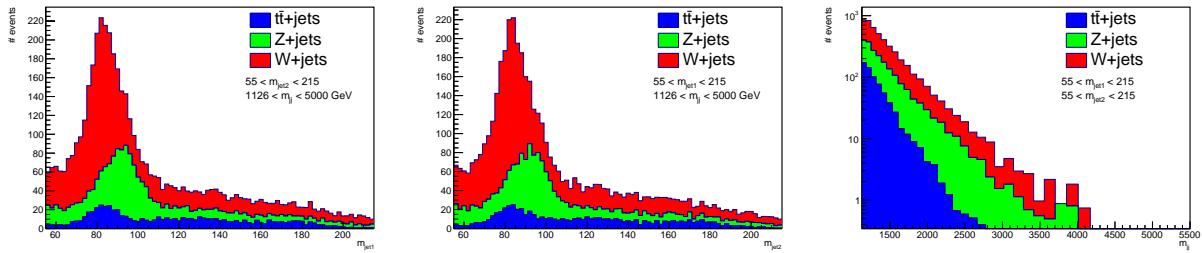


Figure 9.32: Projections of the sub-dominant backgrounds on the jet mass axes M_{jet1} (left) and M_{jet2} (middle), as well as on the dijet invariant mass M_{VV} (right). Here in the low-purity category.

contributions: a resonant part consisting of real vector-boson jets, peaking around the W or Z boson mass; and a non-resonant part composed of jets originating from a quark or a gluon, similar to the QCD-multijets background. These two contributions are modeled separately since we know that the non-resonant part of the jet mass spectrum is correlated with the dijet invariant mass (as was the case for QCD), while the resonant part is not (as was the case for the signal PDF). We additionally want to encode the information that these backgrounds in reality have only one real vector boson and so only peak in one jet-mass dimension, by requiring the PDFs to consist of a resonant part on one axis, and a non-resonant part on the other axis. A three-dimensional PDF for the resonant backgrounds is built as a product of three one-dimensional PDFs as follows:

$$P_{res}(M_{jet1}, M_{jet2}, M_{VV}) = P_{VV}(M_{VV}) \times P_{jet1}(M_{jet1}, M_{jet2}) \times P_{jet2}(M_{jet2}, M_{jet1}), \quad (9.7)$$

where

$$P_{jet1}(M_{jet1}, M_{jet2}) = f \times R(M_{jet2}) \times P_{res,1}(M_{jet1}) + (1 - f)P_{non-res,1}(M_{jet1}) \text{ and} \quad (9.8)$$

$$P_{jet2}(M_{jet2}, M_{jet1}) = (1 - f) \times R(M_{jet1}) \times P_{res,2}(M_{jet2}) + fP_{non-res,2}(M_{jet2}). \quad (9.9)$$

2699 Here, $R(M_{\text{jet}})$ parametrizes the correlation between $M_{\text{jet}1}$ and $M_{\text{jet}2}$ and f is a fit parameter
 2700 that is used to adjust the fraction of real vector boson jets in $M_{\text{jet}1}$ compared to $M_{\text{jet}2}$. Its value
 2701 can vary by 10% around $f = 0.5$, the expected median when using a random jet sorting. $P_{\text{res},1}$
 2702 and $P_{\text{res},1}$ are PDFs describing the resonant part of the jet mass spectrum, while $P_{\text{non-res},1}$ and
 2703 $P_{\text{non-res},2}$ describe the non-resonant part. The functional form of these will be described below.
 2704 As the contribution of background from $t\bar{t}$ is much smaller than the one coming from V+jets
 2705 (less than 2%), this background is modeled together with the W+jets contribution. Therefore
 2706 only two PDFs are built for the resonant background: one for the W+jets and $t\bar{t}$ processes, and
 2707 one for the Z+jets process.

2708 The available MC statistics in the high-purity category are very low and therefore the parametriza-
 2709 tion of the resonant background is done for the low-purity category only, and the resulting shapes
 2710 are then used for both purity categories. The uncertainties for the different purity categories are,
 2711 however, included as separate nuisance parameters in the fit.

2712 The non-resonant M_{VV} PDF is constructed using the same kernel approach as is used to
 2713 model the QCD multijet background with one minor difference: due to the low statistics in the
 2714 high- M_{VV} tail, an additional smoothing of the jet mass tail using a leveled exponential of the
 2715 form

$$\frac{dN}{dm_{jj}} = \frac{P_0(1 - m_{jj}/s)^{P_1}}{(m_{jj}/s)^{P_2}} \quad (9.10)$$

2716 is performed. Here, s is the center-of-mass energy. The function is fitted to the spectrum starting
 2717 from a dijet mass of 1.1 (2.1) TeV for the high-purity (low-purity) category. Two uncertainties
 2718 on the shape are added in order to accommodate possible MC mis-modeling due to higher order
 2719 QCD and electroweak corrections: one proportional to M_{VV} and one proportional to $1/M_{\text{VV}}$. The
 2720 resulting M_{VV} kernels (solid lines) for the W+jets background are shown in Figure 9.33 and are
 2721 compared to MC simulation (markers). The blue line corresponds to the nominal shape, while the
 2722 red and green lines correspond to the uncertainties proportional to M_{VV} and $1/M_{\text{VV}}$, respectively.
 2723 The corresponding distributions for the Z+jets background are shown in Appendix D.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. The uncertainty on the mean and width of the M_{jet} distribution
 are fully correlated between signal and the resonant background, as these uncertainties affect
 all jets generated by a 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

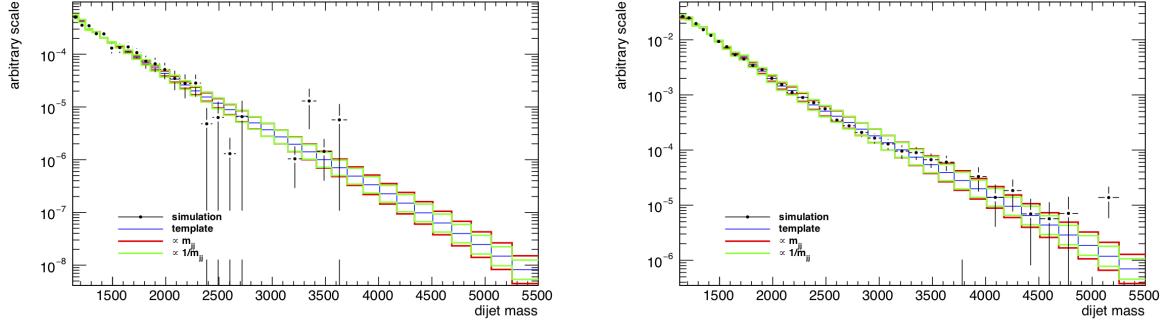


Figure 9.33: One-dimensional M_{VV} kernels (solid line) compared to MC (markers) for the W+jets background in the HP (left) and LP (right) categories. The blue line corresponds to the nominal shape, while the red and green lines correspond to uncertainties proportional to M_{VV} and $1/M_{VV}$, respectively.

jet mass spectrum for M_{jet1} is shown in Figure 9.34 for W+jets and $t\bar{t}$ (left), and Z+jets (right). The small enhancement around 170 GeV is caused by fully merged top jets, but is so small (< 2% in the low-purity and < 0.5% in the high-purity category) that we do not take it into account in the final PDF. The non-resonant part of the V+jets background is modeled using a simple

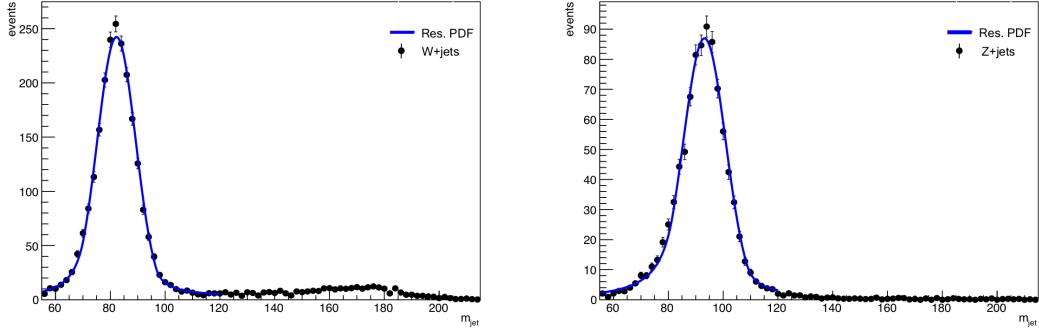


Figure 9.34: Fit to the resonant part of the V+jets and $t\bar{t}$ spectrum for W+jets and $t\bar{t}$ (left) and Z+jets (right).

Gaussian fit to the spectrum of jets not matched to real vector bosons, as shown in Figure 9.35. 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 originate from a real boson peaking around the V mass, the other is bound to be a quark jet. Therefore, the fraction of real

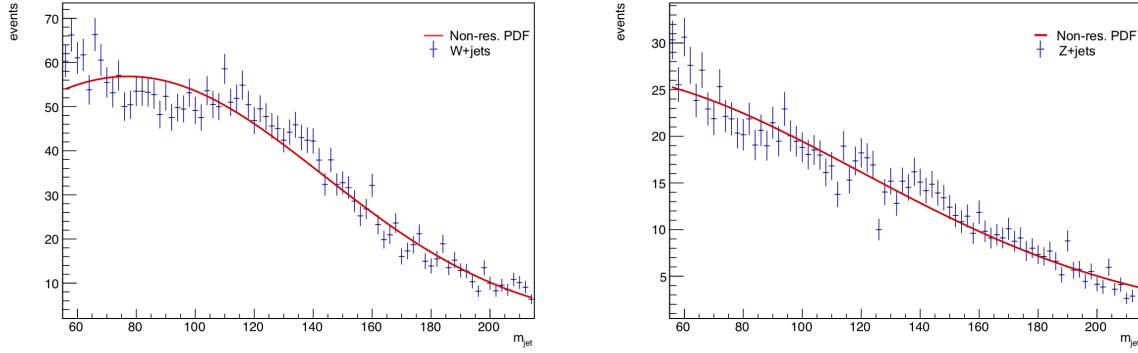


Figure 9.35: Fit to the non-resonant part of the $V+jets$ and $t\bar{t}$ spectrum for $W+jets$ and $t\bar{t}$ (left) and $Z+jets$ (right).

boson jets contained in M_{jet1} affects the fraction of real boson jets in M_{jet2} . To account for this, the fraction of real V jets versus quark or gluon jets is parametrized as a function of the jet mass. As the jets are randomly sorted, we define the parametrization as

$$R(M_{jet}) = \frac{N_{res,jet1}(M_{jet2}) + N_{res,jet2}(M_{jet1})}{N_{non-res,jet1}(M_{jet2}) + N_{non-res,jet2}(M_{jet1})}, \quad (9.11)$$

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 9.36. As a closure test, the three-

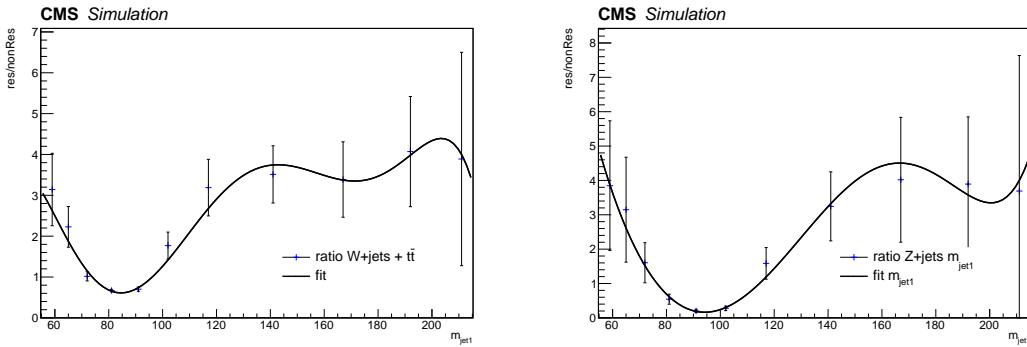


Figure 9.36: Ratio of resonant to non-resonant events in $W+jets$ (left) and $Z+jets$ (right) as a function of jet mass.

dimensional kernel as defined in Equation 9.7 is fitted to the $V+jets$ and $t\bar{t}$ simulation. Figure 9.37 shows the fitted kernel (red) together with the MC data points (black markers) in the low-(top) and high-purity (bottom) categories. Mostly good agreement is observed along all dimensions,

2729 with some deviations in the very high M_{VV} tail in the high-purity category. This is, however,
 2730 a region populated with very few events and is completely overwhelmed by the QCD multijets
 background that has the same shape.

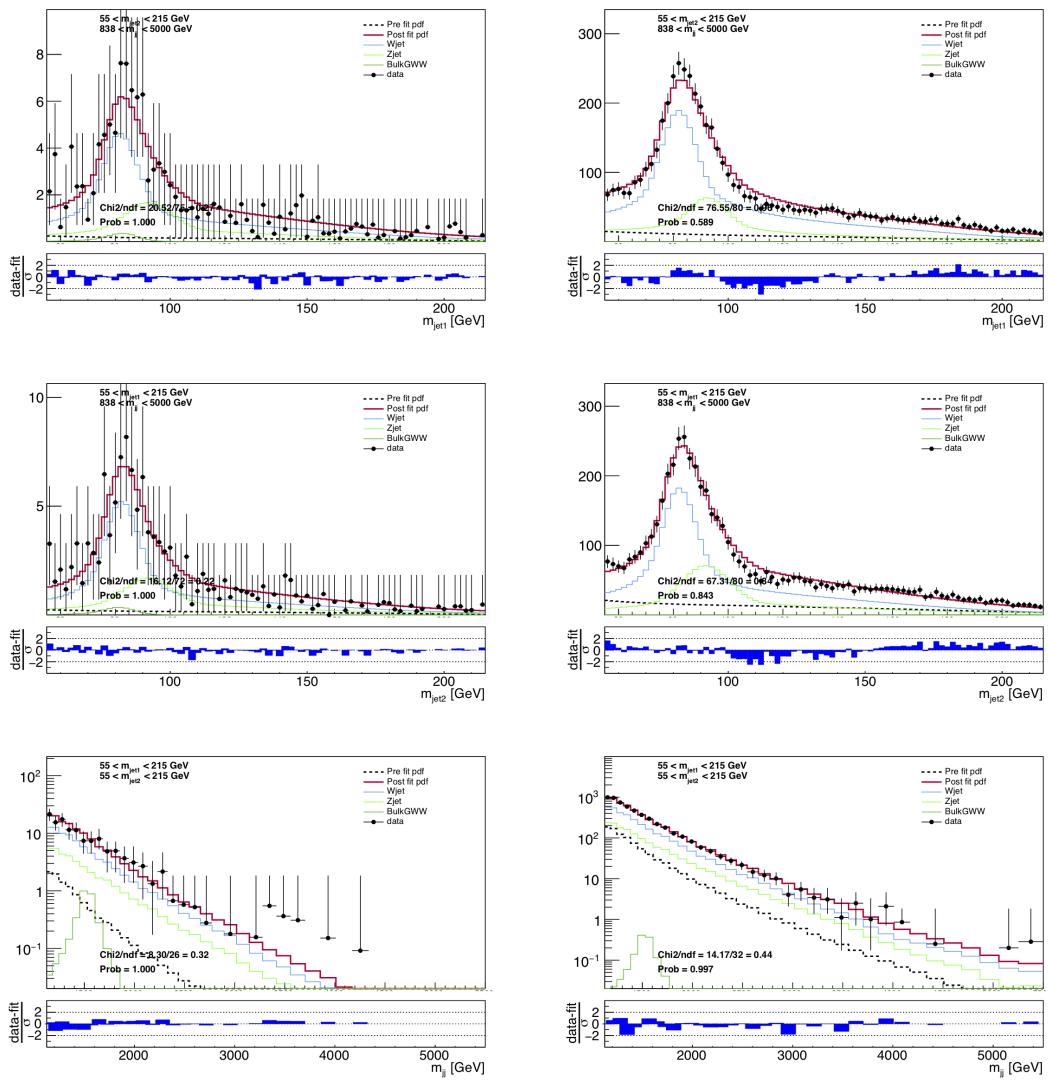


Figure 9.37: Fits using the complete resonant background model (red line) to the V+jets MC simulation (black markers) for the high- (left) and low-purity (right) category. Here for $M_{\text{jet}1}$ (top), $M_{\text{jet}2}$ (middle) and M_{VV} (bottom).

2732 9.8 Systematic uncertainties

2733 The background estimate for each signal mass hypothesis is obtained by fitting the signal and
 2734 background three-dimensional probability density functions obtained above, to the observed
 2735 data in each analysis category. We follow the modified frequentist prescription (CLs) criterion,
 2736 evaluated using asymptotic expressions described in Ref. [72]. Systematic uncertainties are
 2737 treated as nuisance parameters and profiled in the statistical interpretation using log-normal
 2738 priors, while Gaussian priors are used for shape uncertainties. The uncertainties entering the fit
 are listed in Table 9.5. Here, V+jets ratio describes the ratio between W+jets and Z+jets events

Source	Relevant quantity	HPHP unc. (%)	HPLP unc. (%)
PDFs	Signal yield	3	
W-tagging efficiency	Signal+ V+jets yield	25 (21)	13 (11)
W-tagging p_T -dependence	Signal+ V+jets yield	5-60	5-44
Integrated luminosity	Signal+ V+jets yield	2.3 (2.6)	
QCD normalization	Background yield	50	
V+jets normalization	Background yield	10	
V+jets ratio	Migration	10	
PDFs	Signal M_{VV}/M_{jet} shape	< 1	
Jet energy scale	Signal M_{VV} shape	2	
Jet energy resolution	Signal M_{VV} shape	5	
Jet mass scale	Signal + V+jets M_{jet} shape	1	
Jet mass resolution	Signal + V+jets M_{jet} shape	8	
QCD HERWIG++	QCD shape	33	
QCD MADGRAPH+PYTHIA8	QCD shape	33	
p_T -variations	QCD shape	33	
Scale-variations	QCD shape	33	
High- M_{jet} turn-on	QCD shape	33	
p_T -variations	V+jets M_{VV} shape	33	

Table 9.5: Summary of the systematic uncertainties and the quantities they affect. Numbers in parenthesis correspond to uncertainties for the 2016 analysis, when these differ from 2017.

2739 and is allowed to float by 10%. The W-tagging efficiency scale factor is fully anti-correlated
 2740 between the HP and LP categories (3–10%), and fully correlated between signal and +jets. The
 2741 p_T -dependence uncertainty of the scale factor arises from the extrapolation of the W-tagging
 2742 efficiency scale factors, which are measured in $t\bar{t}$ events where the jet has a p_T around 200 GeV,
 2743 towards higher transverse momenta. This uncertainty is estimated in signal MC and is based on
 2744 the difference in tagging efficiency between PYTHIA and HERWIG++ as a function of p_T relative
 2745 to the difference at 200 GeV. This is considered as correlated between the τ_{21}^{DDT} categories and is
 2746 defined in Section 9.6.2. The shape uncertainties on M_{jet} are considered fully correlated between
 2747 signal and V+Jets, allowing for the data to constrain these parameters. These affect the mean
 2748

2749 and the width of the signal and V+jets PDFs.

2750 Uncertainties on the background shape are added as alternate PDFs to the fit through vertical
2751 template morphing. This creates nuisance parameters for each shape that simultaneously affect all
2752 three dimensions. We define five shape nuisance parameters in total. The first effect corresponds
2753 to a variation of the underlying transverse momentum spectrum and the corresponding template is
2754 obtained by simultaneously varying jet masses and M_{VV} by a quantity proportional to M_{VV} and
2755 M_{jet} . The second effect is a variation of the scale and is taken into account through an alternate
2756 shape obtained by simultaneously varying the jet masses and M_{VV} by a quantity proportional
2757 to $1/M_{VV}$ and $1/M_{jet}$. Three additional alternate shapes that simultaneously affect resonance
2758 mass and jet groomed mass are also added in order to take into account differences in MC
2759 generation and modeling of parton shower. Since the choice of QCD MC used to generate the
2760 nominal template is random, we insert additional templates derived using all available QCD
2761 MC: HERWIG++, MADGRAPH+PYTHIA8 and POWHEG. This allows us to include all background
2762 knowledge we have available into the fit. Finally, in order to account for an expected M_{VV}
2763 turn-on in the extreme large- M_{jet} (> 175 GeV) and low- M_{VV} (< 1200 GeV) region, an additional
2764 shape uncertainty parametrizing any discrepancy between the three-dimensional template and
2765 QCD MC is added to the fit. Note that the latter shape uncertainty only affects the region
2766 above $M_{jet} > 175$ GeV and below $M_{VV} < 1200$ GeV, far from where signal is expected. The
2767 above nuisance parameters are assigned very large pre-fit values, allowed to float within 33–50%,
2768 effectively allowing the simulation enough flexibility to adapt to data. The alternate shapes
2769 described above are shown in Figure 9.38.

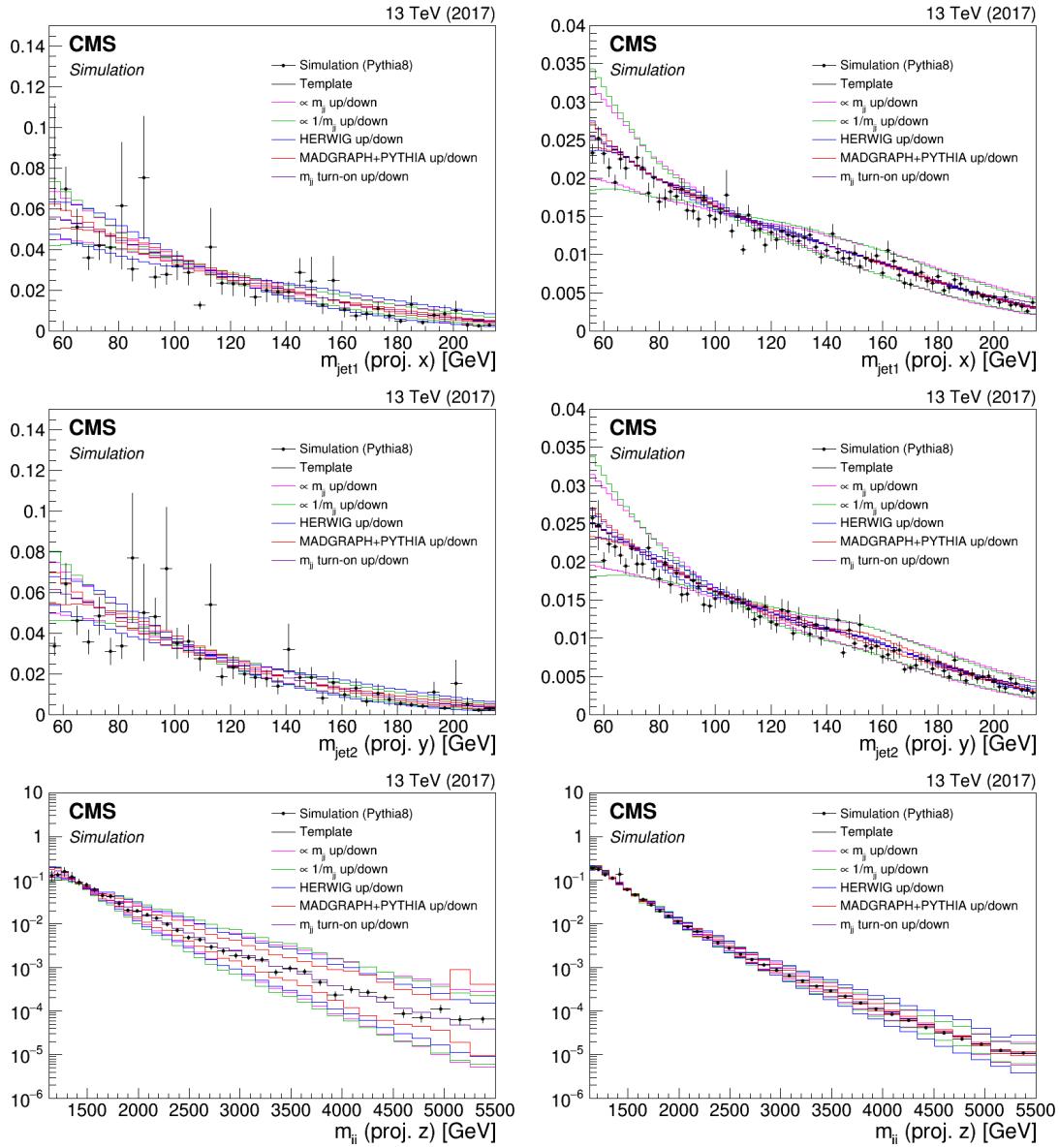


Figure 9.38: The nominal MC data (markers) and smooth nominal kernel obtained from PYTHIA8 (black line), together with the five alternate shapes added to the fit as nuisance shape parameters. Here for the high- (left) and low-purity (right) categories.

2770 9.9 Background model validation

2771 In order to test the robustness of the fit, several checks are performed in simulation and in a data
 2772 control region before unblinding the data signal region.

2773 9.9.1 Variations of QCD multijet predictions

2774 First, we check that the main three-dimensional background template (generated starting from
 2775 QCD PYTHIA8 MC) can fit predictions from alternate QCD multijets samples in order to
 2776 demonstrate that systematic uncertainties in the modeling of parton showers are covered by
 2777 the relevant nuisance parameters. Fits to three different QCD samples are compared: PYTHIA8
 2778 (nominal), HERWIG++ (alternate shape 1) and MADGRAPH+PYTHIA8 (alternate shape 2). To
 2779 ensure a smooth data distribution not affected by low statistics present in the MC spectra, we
 2780 generate a toy dataset from the three-dimensional templates rather than using MC events directly.
 2781 The post-fit distributions after fitting each toy is shown in Figures 9.39-9.41 for the high- (left)
 2782 and low-purity (right) category. The fit quality of these are quantified using a goodness-of-fit
 2783 (based on a saturated model, which is valid when data are non-Gaussian), shown in Figure 9.42.
 2784 The test statistics are Gaussian distributed and the toy dataset is in good agreement with the
 2785 background only hypothesis, demonstrating the fits ability to account for differences in QCD
 2786 multijet predictions.

2787 9.9.2 Fit to data control region

2788 To validate the method in data, the 3D fit procedure is tested in a data control region where
 2789 both jets are required to have $0.43 < \tau_{21}^{DDT} \leq 0.79$, the so-called LPLP category. The templates
 2790 are built following the procedure described in Section 9.7.2. In this category, the contribution of
 2791 the resonant background is negligible with respect to the dominant QCD multijet background
 2792 and is removed from the fit. The post-fit distributions in the data control region are shown in
 2793 Figures 9.43–9.45 for different jet mass and dijet invariant mass bins. A goodness-of-fit check is
 2794 performed, shown on the left plot in Figure 9.46, and we again find that the toys are Gaussian
 2795 distributed and the LPLP data is in good agreement with the background-only hypothesis. The
 2796 post-fit value and uncertainty of each nuisance parameter involved in the fit is also studied, where
 2797 deviations from the pre-fit value is quantified through the “pull”, defined as $p_\theta = (\theta - \theta_{in})/\sigma_\theta$,
 2798 where, θ and σ_θ are the post-fit value of the nuisance parameter and its uncertainty, and θ_{in} the
 2799 pre-fit value. The pulls for all nuisance parameters are shown in the right plot in Figure 9.46,
 2800 where the uncertainty is defined as the ratio between the post- and pre-fit uncertainties. The

2801 post-fit values show a reasonable deviation from the chosen pre-fit values of 0. The uncertainties
 2802 are strongly constrained by data, as expected given the large pre-fit uncertainty assigned to let
 2803 the shapes adjust to the real data. In addition to the goodness-of-fit check in the LPLP data
 2804 sideband region, we perform the same test on the data signal region. This is shown in Figure 9.47.
 2805 Good agreement between the observed fit result in data and the background only hypothesis from
 2806 MC simulation is observed.

2807 9.9.3 Bias test in pseudodata

2808 Finally, we study whether there is any bias on the extracted signal rate due to the background
 2809 model. A $G_{\text{bulk}} \rightarrow WW$ signal is injected on top of a toy dataset generated under either the
 2810 PYTHIA8 or the HERWIG++ template, and is fit using the full background model. The test is done
 2811 for four different signal mass values and with a signal strength chosen such that it corresponds
 2812 to a significance of 4-4.5 standard deviations for each signal mass which is tested. An example
 2813 distribution after injecting a 3 TeV signal is shown in Figure 9.48 for the high-purity category. For
 2814 each tested signal mass point, a signal plus background fit is performed. The signal normalization
 2815 is free to float in the fit, which determines the signal strength μ and its uncertainty σ_μ . For
 2816 each toy, the pull of the signal strength $p_\mu = (\mu - \mu_{in})/\sigma_\mu$ is calculated, where μ_{in} is the pre-fit
 2817 value of the signal strength. The procedure is repeated ~ 1000 times for each category, and the
 2818 cumulative distribution of the pulls is fitted with a Gaussian function to determine the mean and
 2819 its uncertainty, which represent the bias. This is shown in Figure 9.49. Ideally, the distribution
 2820 should be Gaussian distributed with a mean around 0 and a width of 1. The bias for a toy
 2821 dataset generated under the PYTHIA8 (red markers) or the HERWIG++ (black markers) template,
 2822 as shown in Figure 9.50, is consistently below 50% and therefore no additional correction/bias
 2823 term is introduced. In addition, we calculate the pull for each nuisance parameter and fit the
 2824 cumulative distribution with a Gaussian, in order to determine the mean and its uncertainty,
 2825 representing the shift of the parameters θ with respect to the pre-fit values. The results are shown
 2826 in Figure 9.51. The mean of the cumulative distributions of the pulls is found to be zero for all
 2827 nuisance parameters, as expected. The width of the distribution can range from very small values
 2828 to about 1, depending on the assigned pre-fit uncertainty and the power of the data to constrain
 2829 it.

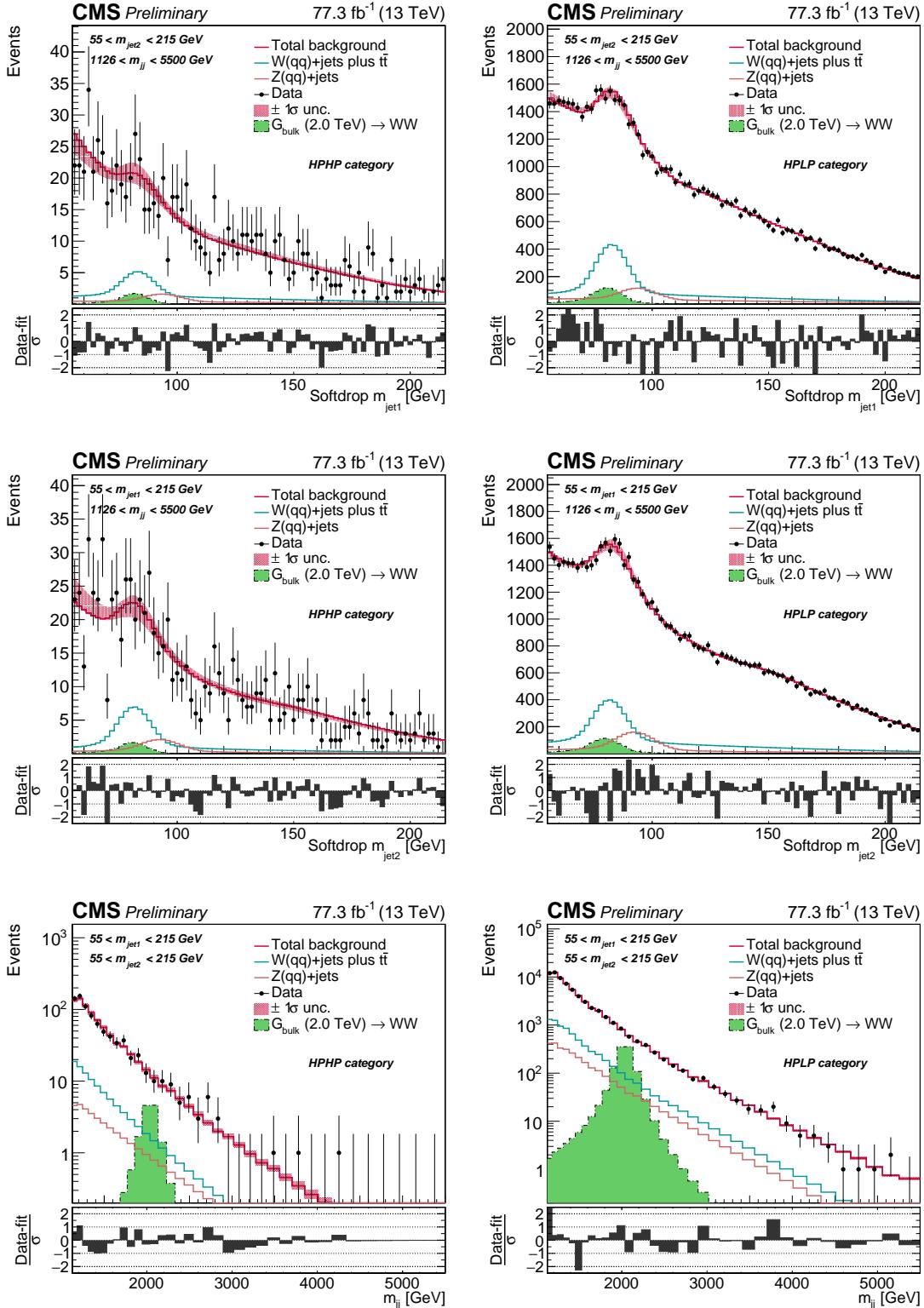


Figure 9.39: Postfit distributions after a combined fit to a toy dataset generated under the QCD PYTHIA8 template. The projections of M_{jet1} (top), M_{jet2} (middle) and M_{ll} (bottom) are shown for the high- (left) and low-purity (right) category.

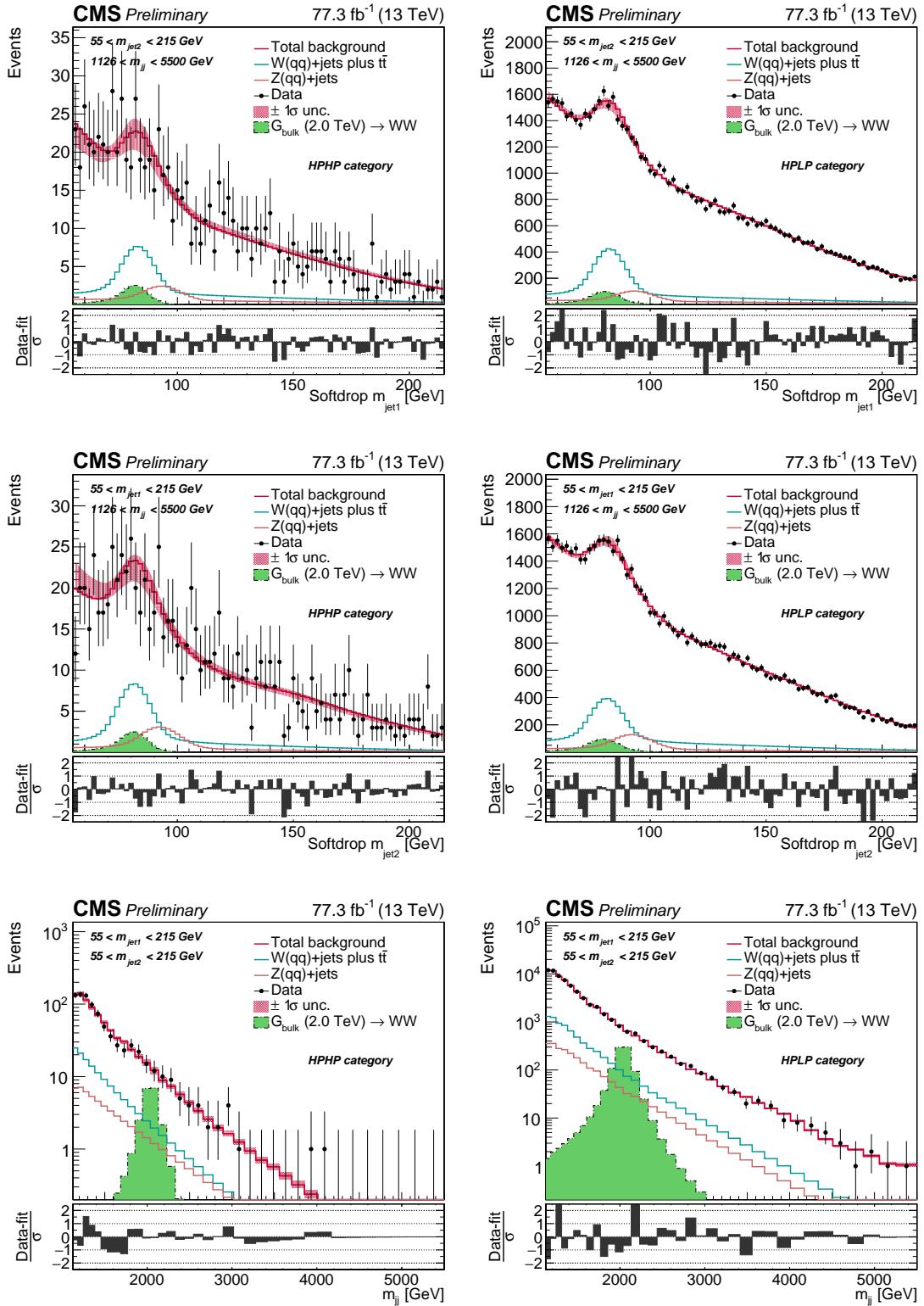


Figure 9.40: Postfit distributions after a combined fit to a toy datasets generated under the QCD HERWIG++ template. The projections of $M_{\text{jet}1}$ (top), $M_{\text{jet}2}$ (middle) and M_{vv} (bottom) are shown for the high- (left) and low-purity (right) category.

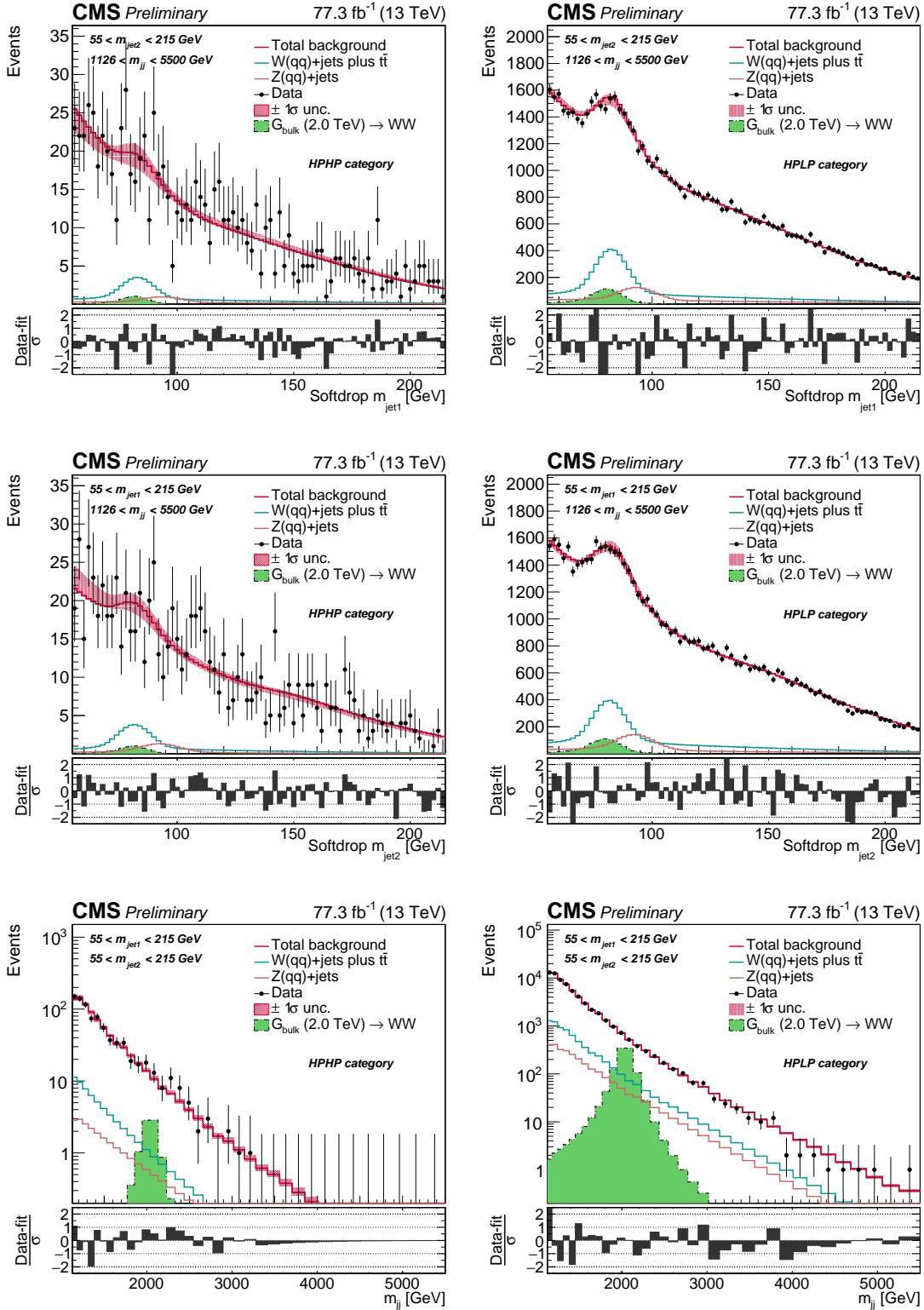


Figure 9.41: Postfit distributions after a combined fit to a toy dataset generated under the QCD QCD MADGRAPH+PYTHIA8 template. The projections of $M_{\text{jet}1}$ (top), $M_{\text{jet}2}$ (middle) and M_{VV} (bottom) are shown for the high- (left) and low-purity (right) category.

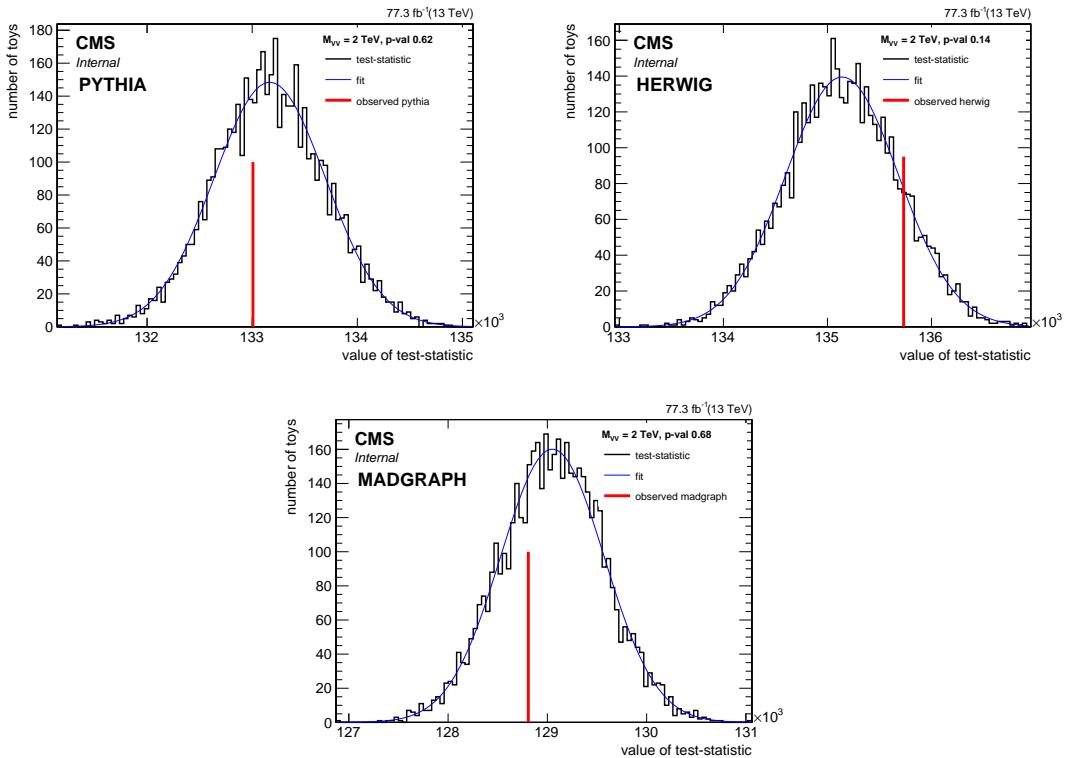


Figure 9.42: The likelihood for toys generated around the background-only hypothesis compared to the likelihood value of a toy dataset generated under the PYTHIA (top left), HERWIG++ (top right) and MADGRAPH+PYTHIA8 template (bottom).

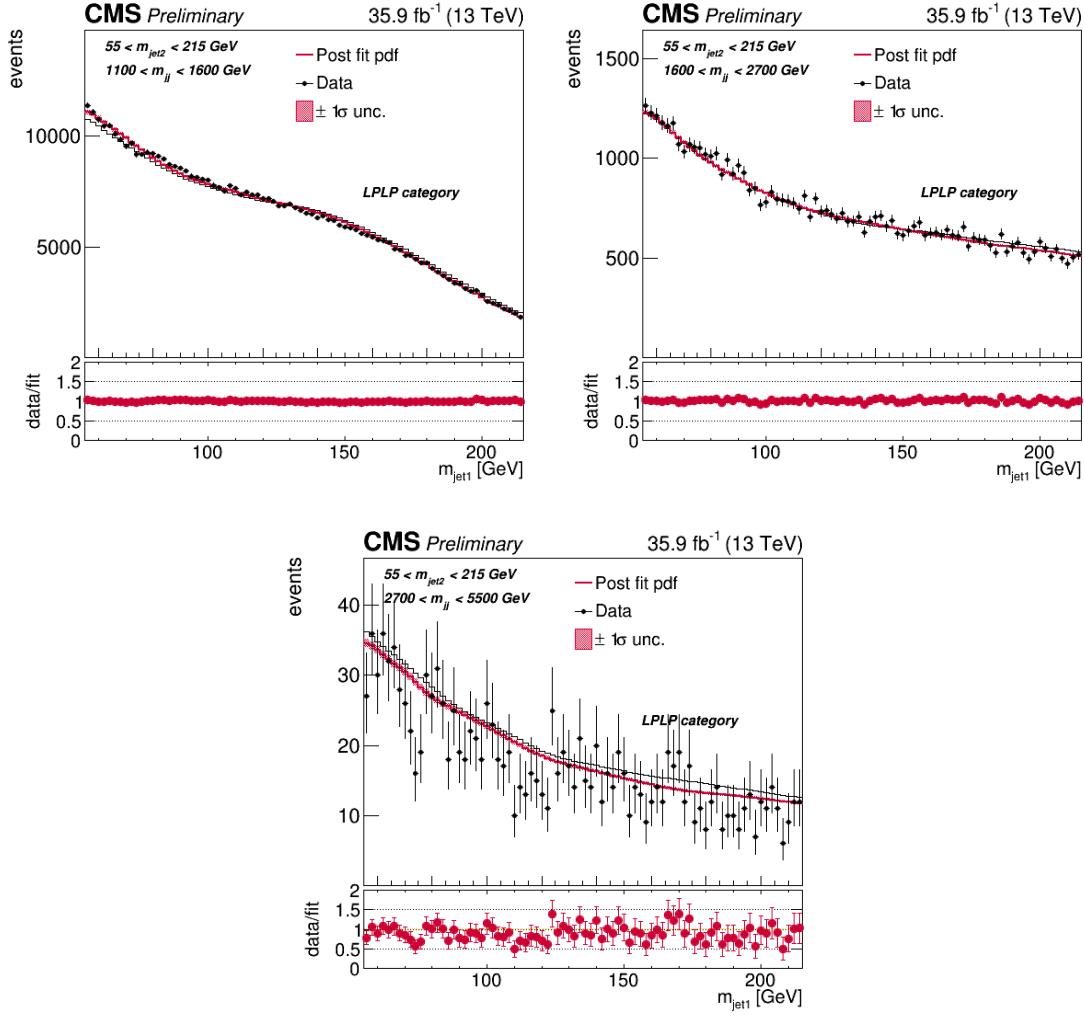


Figure 9.43: Distributions obtained from the fit to 2016 LPLP data. Here the projections of $M_{\text{jet}1}$ are shown for several ranges of M_{VV} , as labelled in the top left corner of each plot, and for the full $M_{\text{jet}2}$ range.

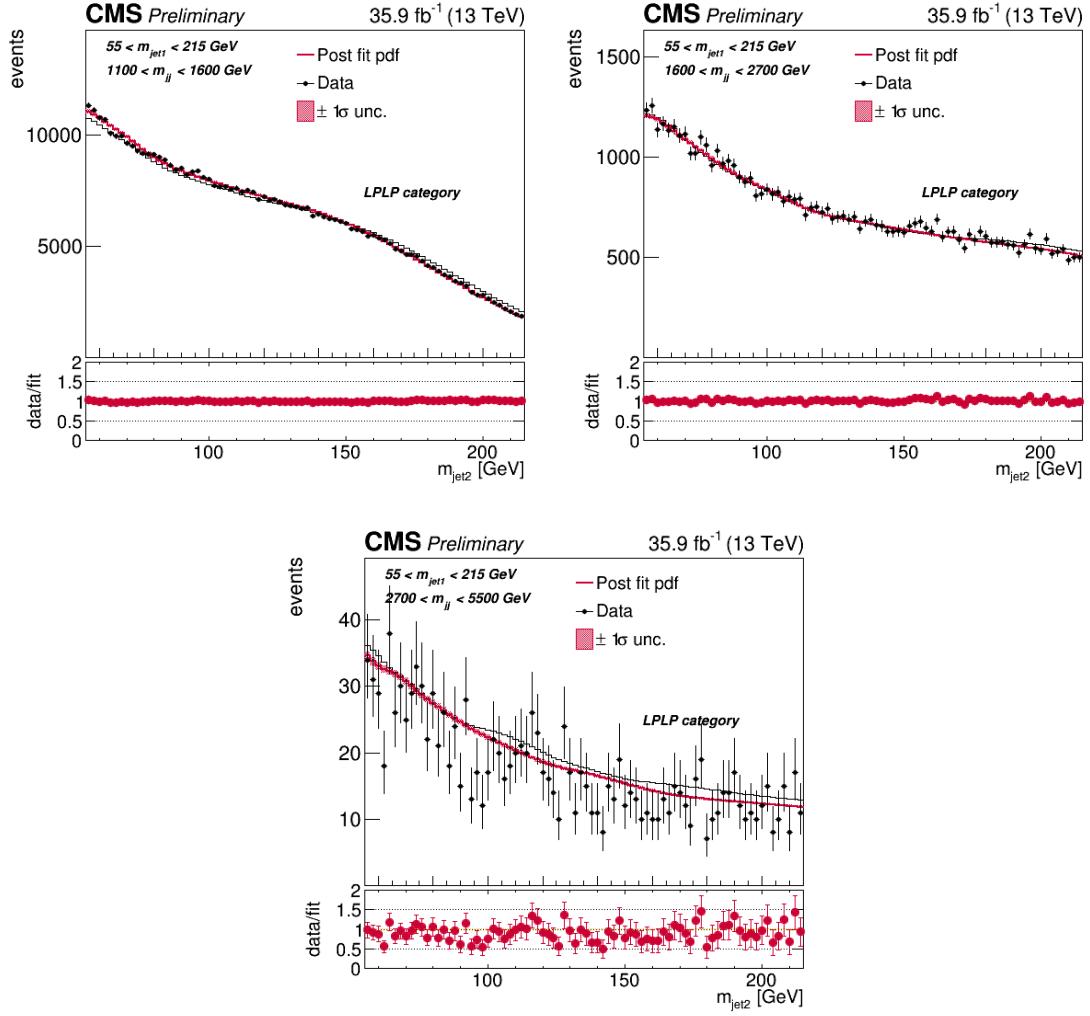


Figure 9.44: Distributions obtained from the fit to 2016 LPLP data. Here the projections of M_{jet2} are shown for several ranges of M_{VV} , as labelled in the top left corner of each plot, and for the full M_{jet1} range.

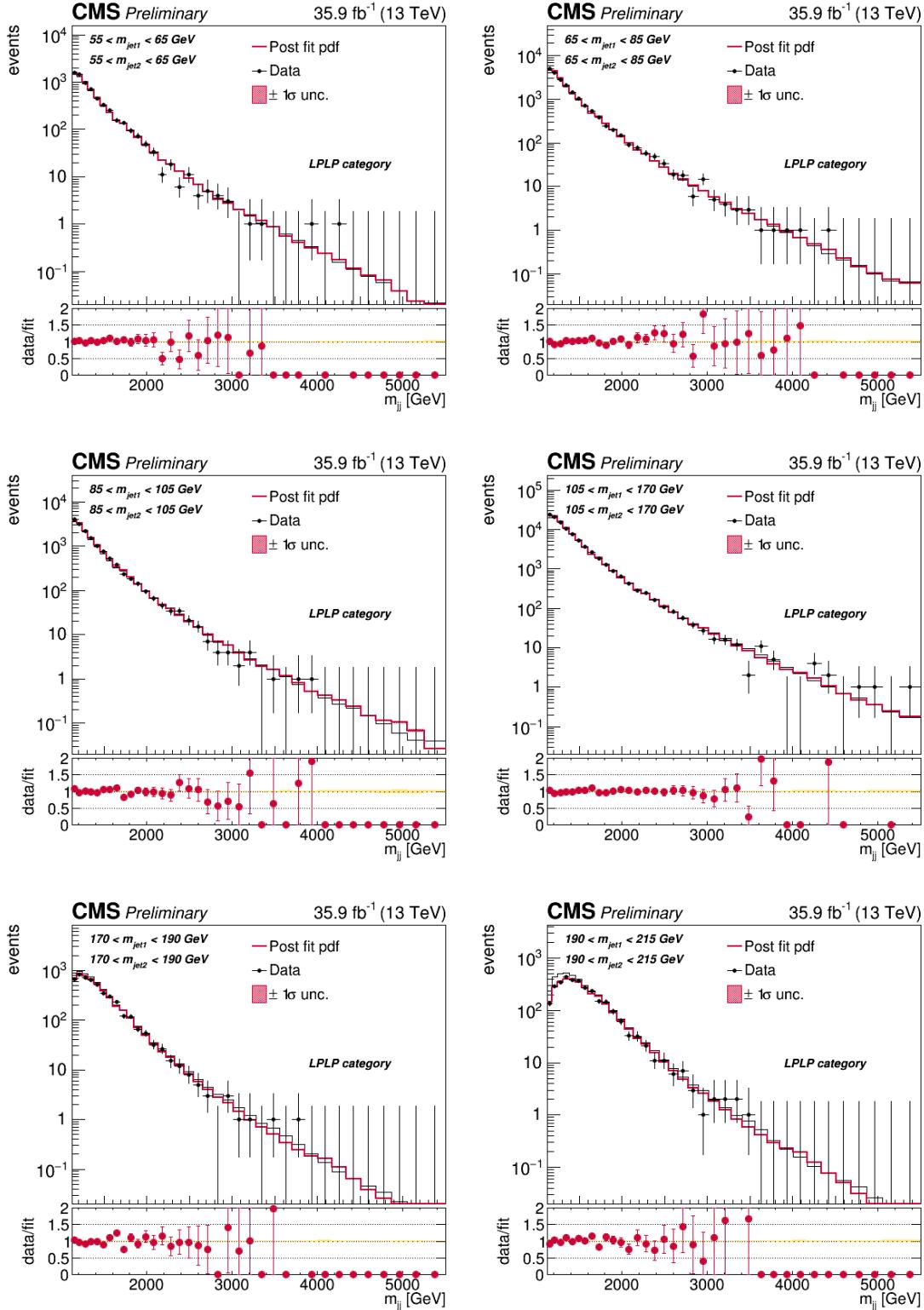


Figure 9.45: Distributions obtained from the fit to 2016 LPLP data. Here the projections of M_{VV} are shown for several equal ranges of M_{jet1} and M_{jet2} , as labelled in the top left corner of each plot.

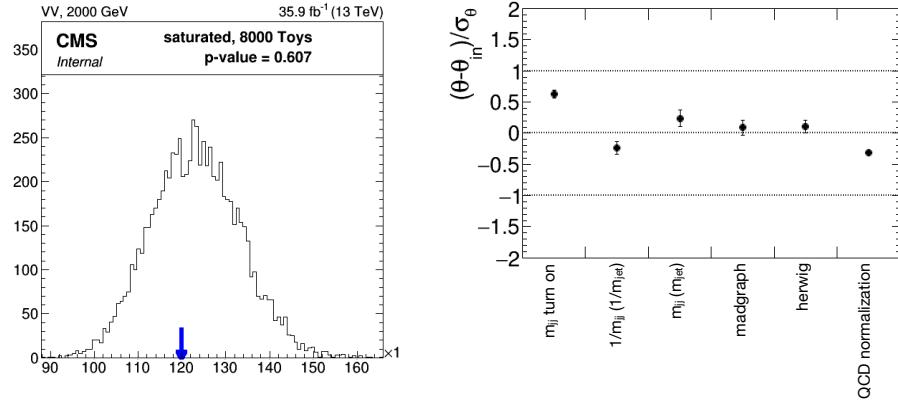


Figure 9.46: Left: the likelihood for toys generated around the background-only fit to LPLP data, compared to the likelihood on data. Data is in good agreement with the background-only hypothesis. Right: pulls of the nuisance parameters for a background-only fit.

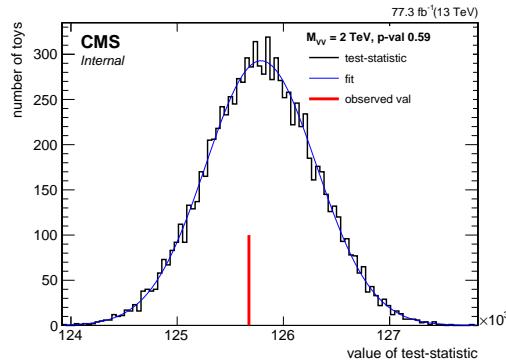


Figure 9.47: Distribution of the likelihood for toys generated around the background-only hypothesis compared to the likelihood on data for the full combination of 2016 and 2017 data.

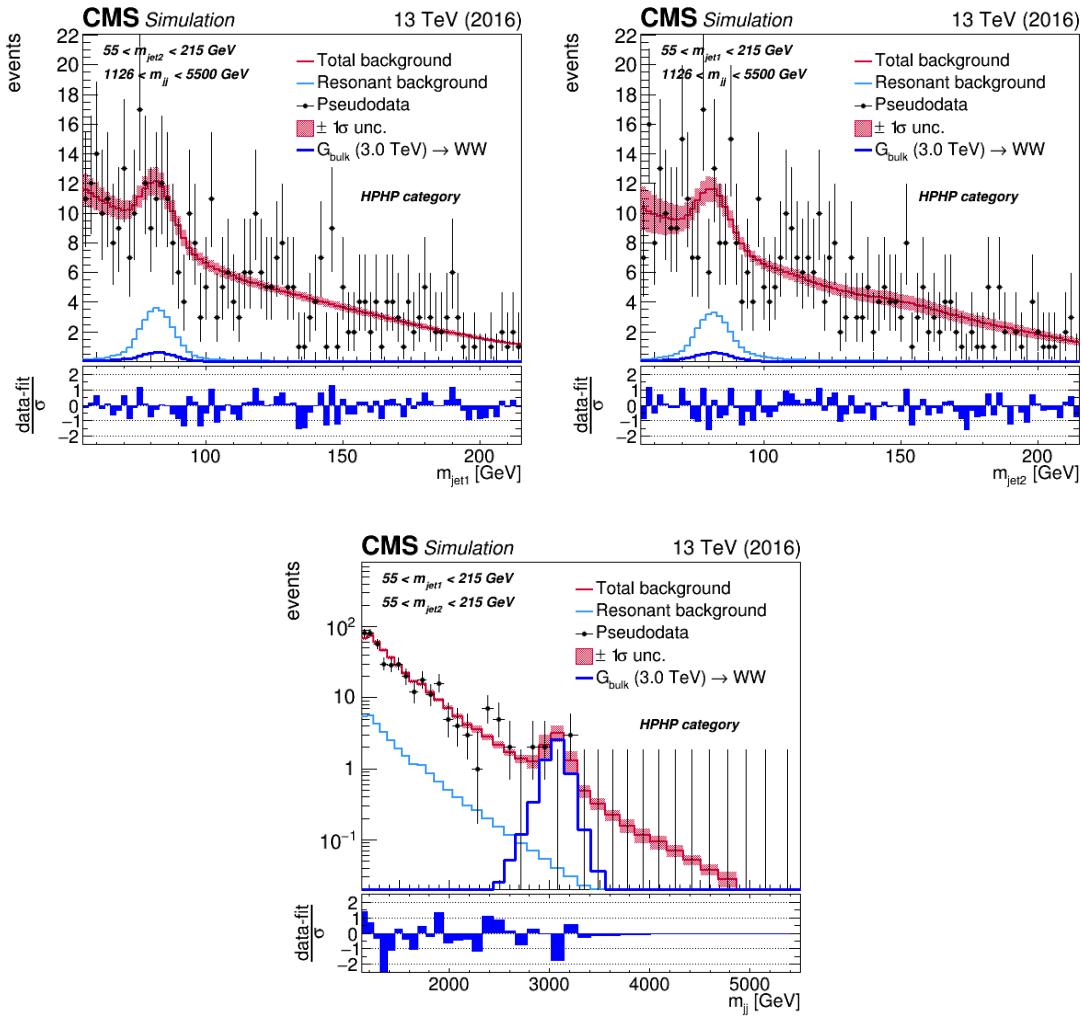


Figure 9.48: The post-fit distribution after injecting a signal with a mass of 3 TeV on top of the background in the HP category.

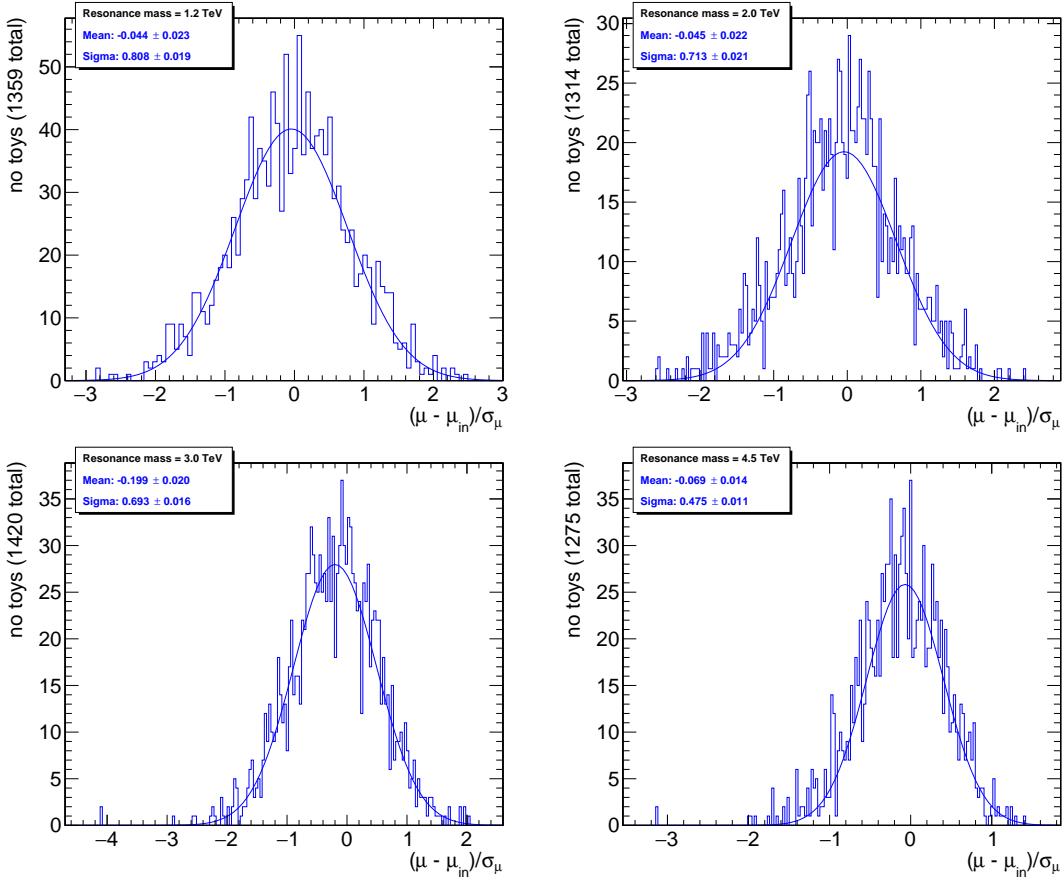


Figure 9.49: Cumulative distributions of the pulls for the signal strength for 4 different signal mass-points.

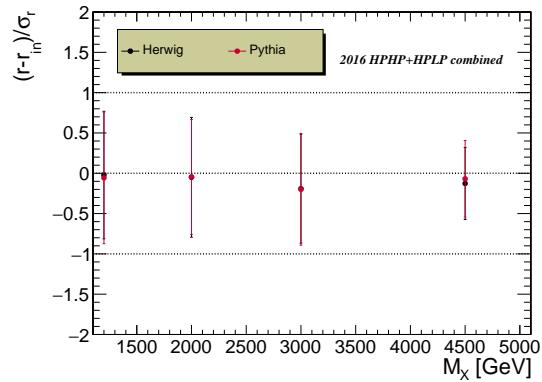


Figure 9.50: Estimated bias as a function of the resonance mass.

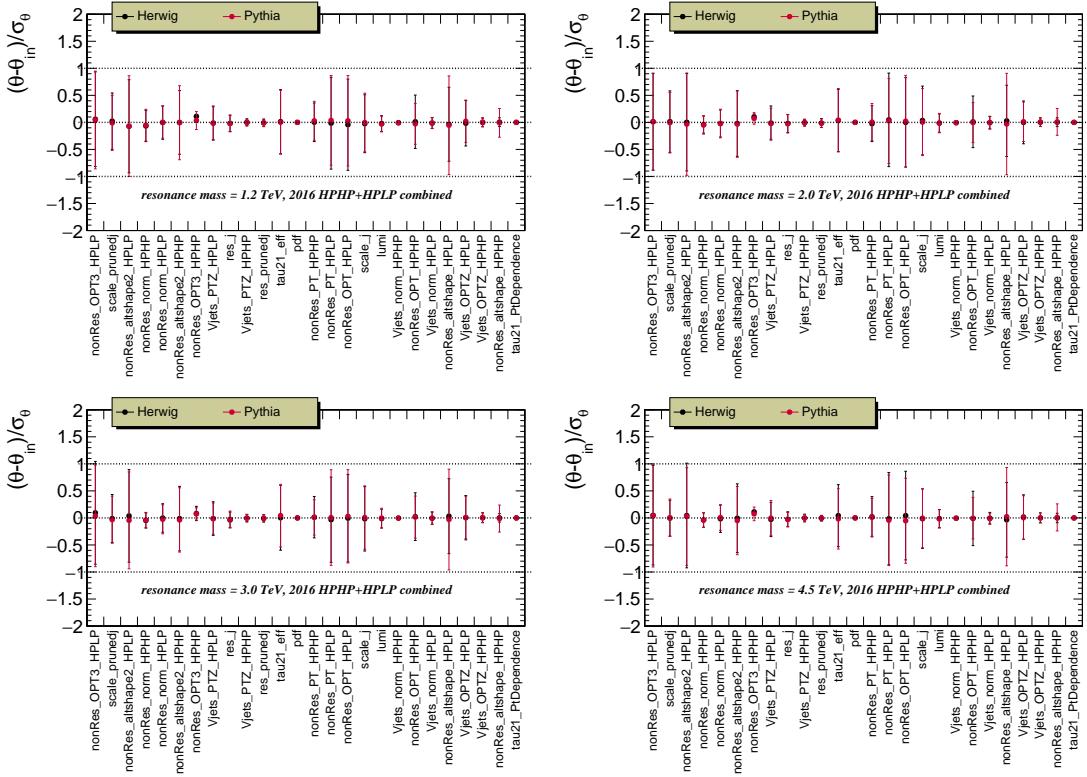


Figure 9.51: Pulls of the fit nuisance parameters after injecting signals of different mass and signal strength on top of a toy datasets generated from the nominal PYTHIA8 template (red markers) and for a toy generated under the HERWIG++ template (black markers).

2830 **9.10 Results**

2831 The distributions obtained from a combined fit to the observed data in 2016 and 2017 are shown
 2832 in Figure 9.52 and 9.53, with the corresponding predicted and observed number of background
 2833 events in the signal region summarized in Table 9.6. We observe a beautiful double peak from
 2834 the $W(q\bar{q})$ and $Z(q\bar{q})+jets$ background, especially visible in the low-purity category. This allows
 2835 us to, for the very first time, constrain the softdrop jet mass scale and resolution simultaneously
 2836 from a $W(q\bar{q})+jets$ and $Z(q\bar{q})+jets$ SM background, which we will discuss in Section 9.10.2. No
 2837 excess is observed and we proceed by setting upper limits on the signal cross section times the
 branching ratio in the same way as in Section 7.9.

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

Table 9.6: Expected and observed yields and their total uncertainty (stat.+sys.) in the two purity categories.

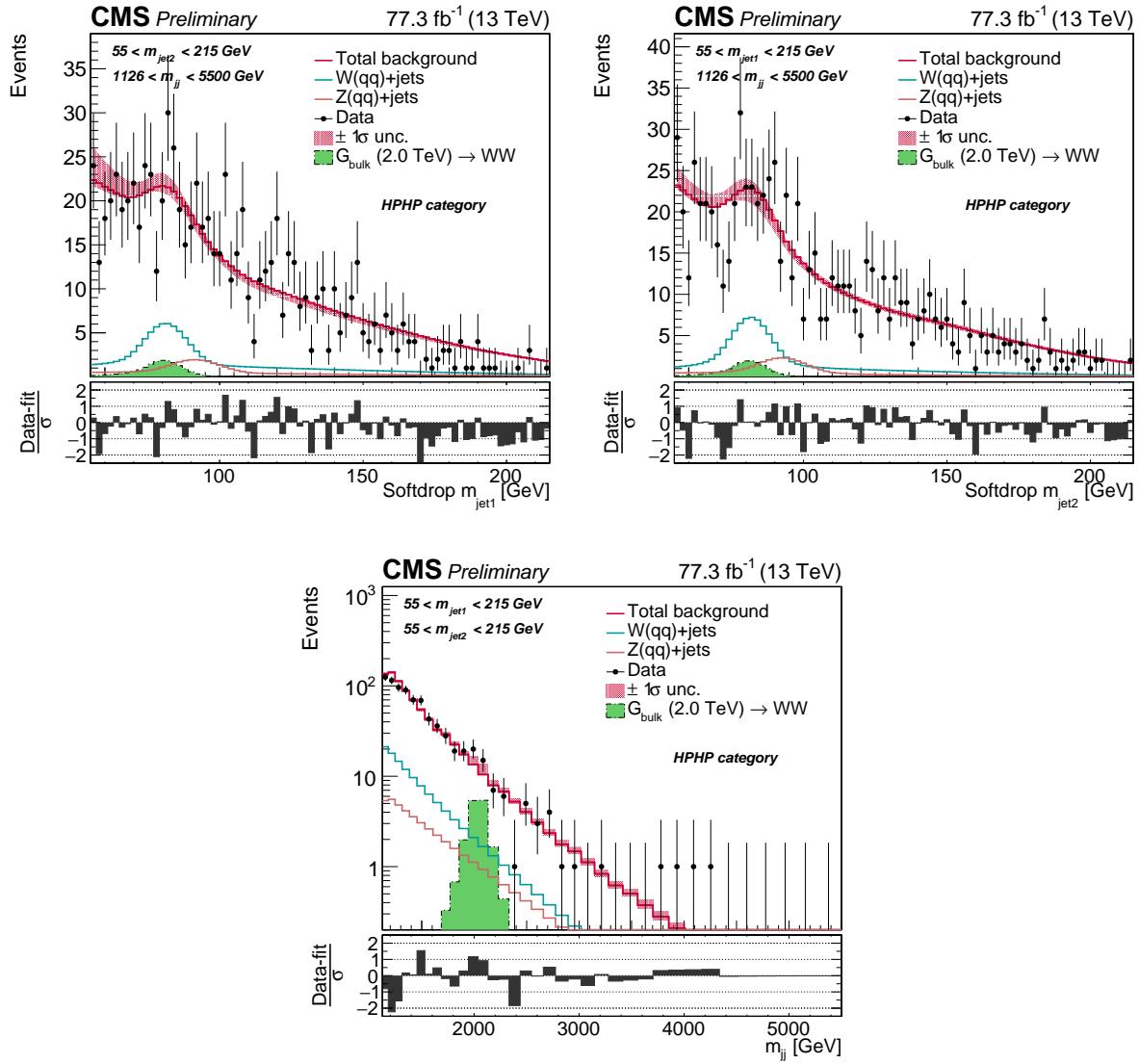


Figure 9.52: Postfit distribution after a fit to 2016 and 2017 data projected onto the M_{jet1} (left), M_{jet2} (middle), and M_{VV} (right) axis for the high-purity category. The background shape uncertainty is shown as a red shaded band, and the statistical uncertainties of the data are shown as vertical bars. The overlaid signal distribution is arbitrarily normalized.

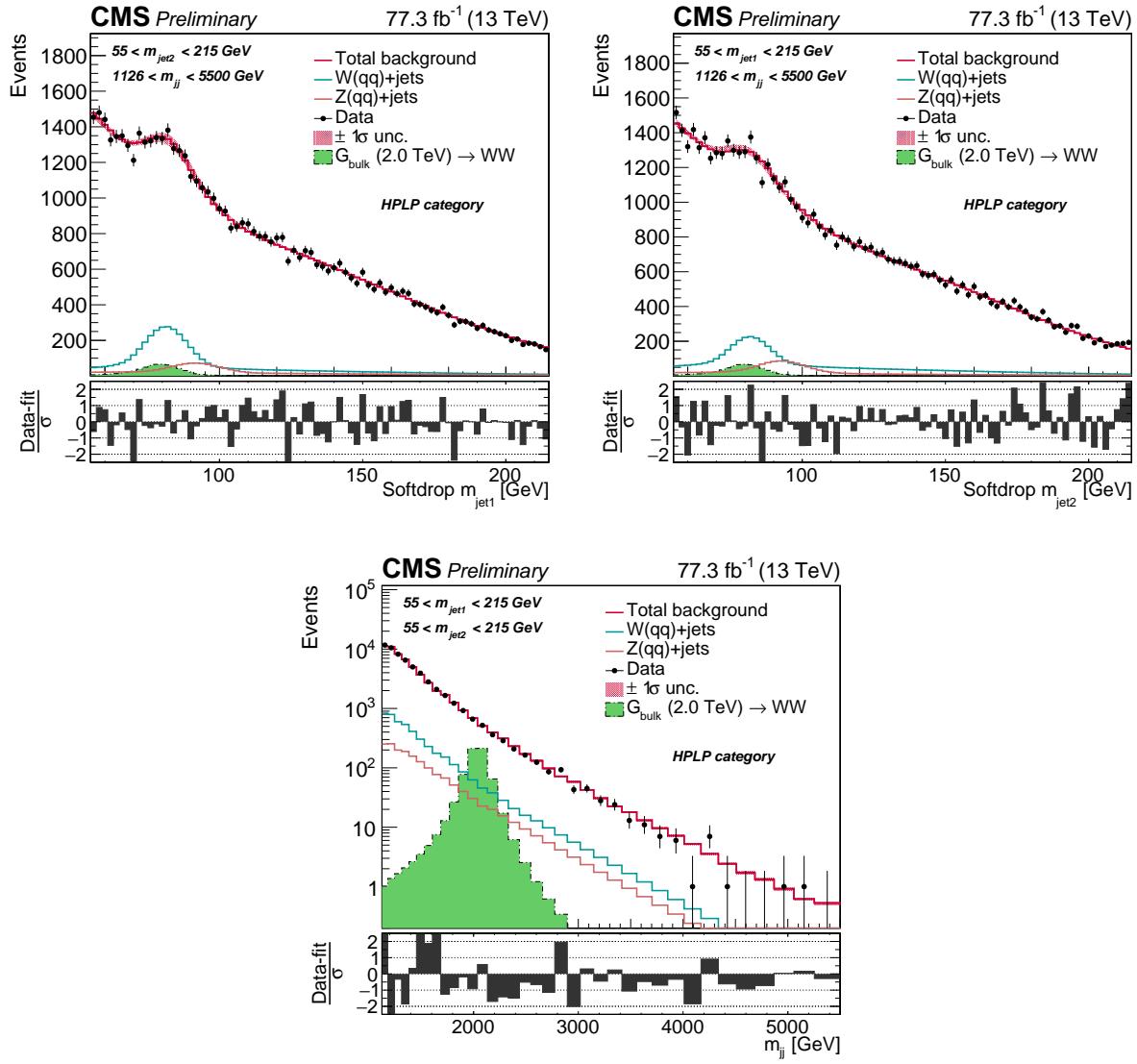


Figure 9.53: Postfit distribution after a fit to 2016 and 2017 data projected onto the $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle), and M_{ll} (right) axis for the low-purity category. The background shape uncertainty is shown as a red shaded band, and the statistical uncertainties of the data are shown as vertical bars. The overlaid signal distribution is arbitrarily normalized.

2839 9.10.1 Limits

2840 As for Search I and Search II, exclusion limits on the cross section of the process $X \rightarrow VV$ are set
 2841 in the context of the Bulk Graviton model and the HVT model B scenario (again obtained using
 2842 the asymptotic CL_S method). Figure 9.54 show the resulting expected and observed exclusion
 2843 limits at 95 % confidence level on the signal cross section times branching fraction as a function of
 2844 the resonance mass for a $G_{\text{bulk}} \rightarrow WW$ (top left), $G_{\text{bulk}} \rightarrow ZZ$ (top right), $W' \rightarrow WZ$ (bottom
 2845 left) and $Z' \rightarrow WW$ (bottom right) signal. The obtained limits are compared with the resonance's
 2846 production cross section times the branching fraction to WW, ZZ and WZ. For the heavy vector
 2847 triplet model B, we exclude at 95% confidence level W' and Z' spin-1 resonances with masses
 2848 below 3.7 and 3.5 TeV, respectively. In the narrow-width bulk graviton model, lower limits on the
 2849 production cross sections are set in the range from 9 fb for a resonance mass of 1.2 TeV, down
 2850 to 0.2 fb for high resonance masses above 5.2 TeV for $G_{\text{bulk}} \rightarrow WW$. In the case of $G_{\text{bulk}} \rightarrow ZZ$,
 2851 lower cross section limits are 13 fb and 0.2 fb for bulk graviton masses of 1.2 and 5.2 TeV,
 2852 respectively.

2853 Comparison with 1D fit

2854 The analysis of the 2016 dataset and corresponding limits obtained using the 3D fit method,
 2855 can be compared to the limits using the 1D fit in Search II in order to estimate whether there
 2856 is a sensitivity gain. One caveat of this comparison is that the analysis performed using the
 2857 1D method does not apply jet energy resolution smearing to the jets, and hence uses an overly
 2858 optimistic assumption on the dijet invariant mass resolution. The difference in resonance mass
 2859 resolution with and without jet energy resolution smearing is shown in Figure 9.55 for the 2016
 2860 (left) and 2017 (right) signal samples. With JER smearing applied, the resonance mass resolution
 2861 is worse and the peak broadens by 5%. Figure 9.56 shows the expected limits based on analyses
 2862 of the 2016 dataset, either using the new 3D fit method, where the jets have been smeared (blabla
 2863 line) or not (blabla line), or using the 1D dijet method. We obtain X–X% improvement with
 2864 the 3D method, which is reduced to X% due to JER smearing being applied and about 35–40%
 2865 improvement when combining the two datasets with respect to the individual results.

2866 Comparison between datasets

2867 We additionally compare the sensitivity between the 2016 and 2017 datasets. Also here there is
 2868 one important caveat: for the 2016 analysis, pileup was removed from the jets using the charged
 2869 hadron subtraction algorithm, whereas in 2017 the PUPPI algorithm was used. In both cases,

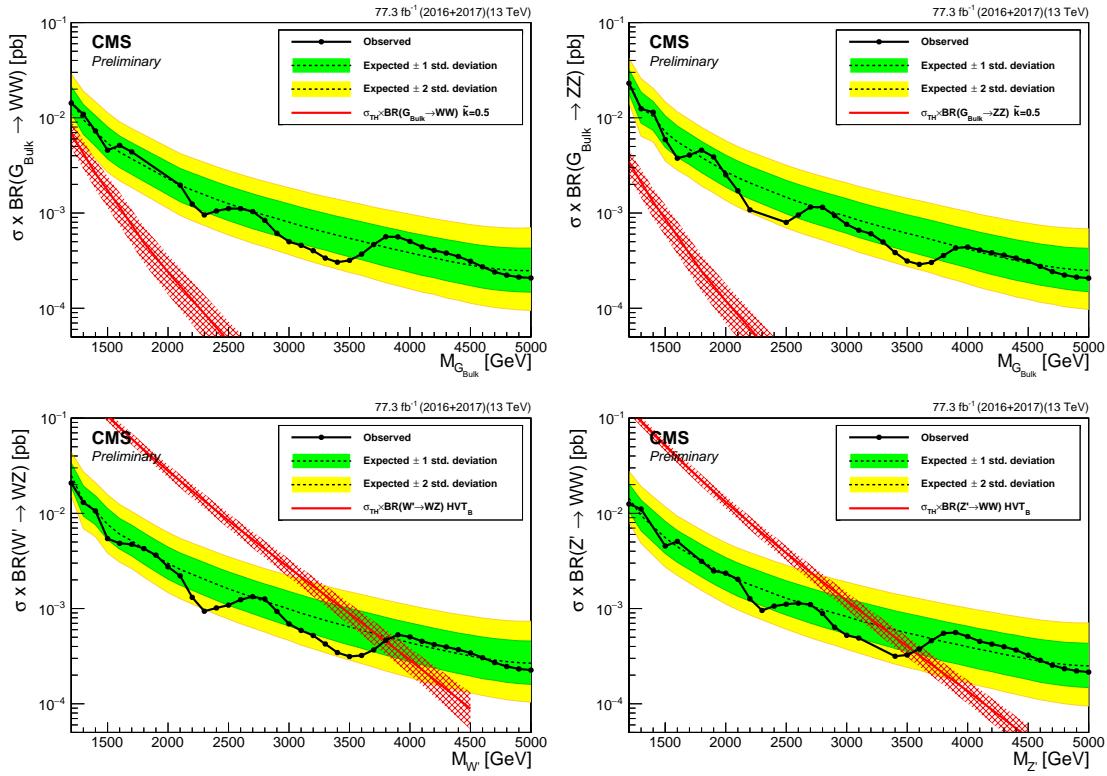


Figure 9.54: Expected limits obtained combining 35.9 fb^{-1} and 41.4 fb^{-1} of data of data after combining all purity categories for Bulk $G \rightarrow WW$ (top left), Bulk $G \rightarrow ZZ$ (top right), $W' \rightarrow WZ$ (bottom left) and $Z' \rightarrow WW$ (bottom right) signals.

the softdrop jet mass was obtained from a PUPPI cleaned soft-dropped jet (in the 2016 analysis, the CHS jet was matched to a PUPPI soft-dropped jet in order to obtain the mass. The jet four-vector still relies on CHS). The reason for this switch is that, while for a pileup between 20-30 the CHS and PUPPI algorithms have a similar performance, once the average number of interactions per event is above 30 the PUPPI algorithm has been shown to achieve a better resolution for large-radius jet observables. In 2016 the average PU was 27 whereas in 2017 it was 38, a 40 % increase. However, despite being better in term of groomed jet-mass resolution, PUPPI pileup subtraction has a worse dijet invariant mass resolution than CHS pileup subtracted jets. This is most likely due to PUPPI removing far more (PU) tracks than CHS and, if these PU tracks happen to be associated to a large calorimeter energy deposit, some of the energy might be lost. This creates a tail at lower dijet invariant masses, as shown in Figure 9.57. Figure 9.58 shows the expected limits obtained with the new 3D fit method using 35.9 fb^{-1} of 2016 data (beige dashed line), 41.4 fb^{-1} of 2017 data (pink dashed line) and when combining the two yielding a

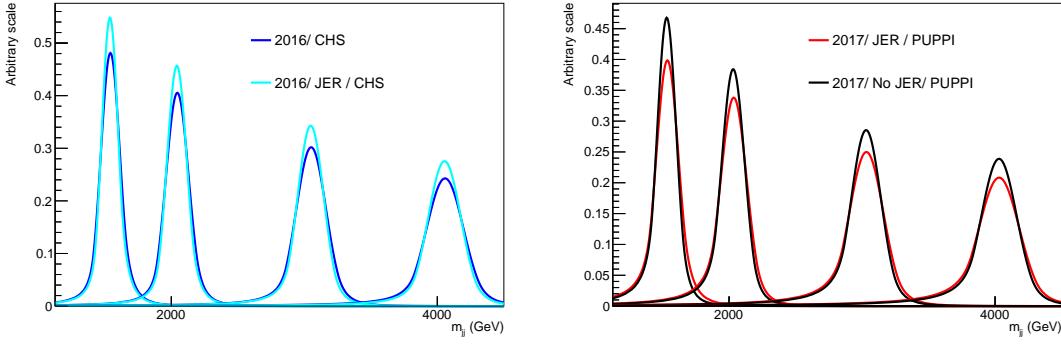


Figure 9.55: Right: A comparison of the resonance mass resolution with (turquoise) and without (blue) jet energy resolution (JER) smearing using 2016 signal samples. Left: A comparison of the resonance mass resolution with (red) and without (black) jet energy resolution (JER) smearing using 2017 signal samples. In both cases, the mass resolution is worse after smearing, and has a roughly 8–10% larger width.

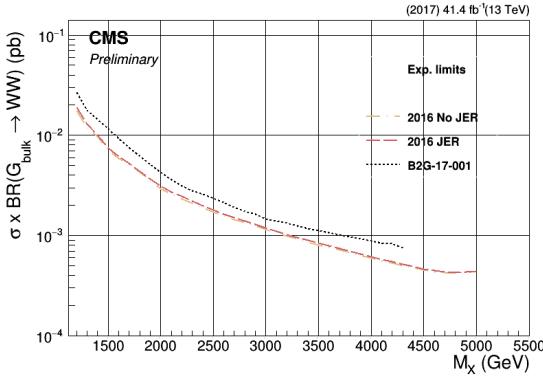


Figure 9.56: Expected limits using 35.9 fb^{-1} of data for a Bulk $G \rightarrow WW$ signal using the 3D fit method, where the jets have had JER smearing applied (red line) or not (yellow line), and using the 1D fit method (dotted black line).

total of 77.3 fb^{-1} of data (dotted black line). These are compared to the results obtained using the 1D fit in Search II (solid purple line). A 35–40% improvement is obtained when combining the two datasets with respect to the individual results, and we observe that the 2017 limits are slightly worse than the 2016 ones due to the above mentioned switch to PUPPI. For the full Run 2 legacy paper combining 2016, 2017, and 2018 data, we will perform a thorough study comparing CHS to PUPPI.

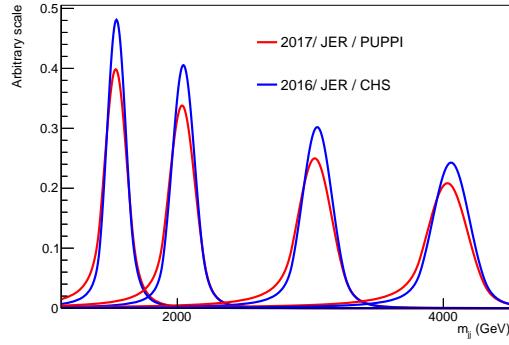


Figure 9.57: A comparison of the resonance mass resolution obtained from 2017 signal samples based on PUPPI jets (red) and 2016 signal samples based on CHS jets (blue).

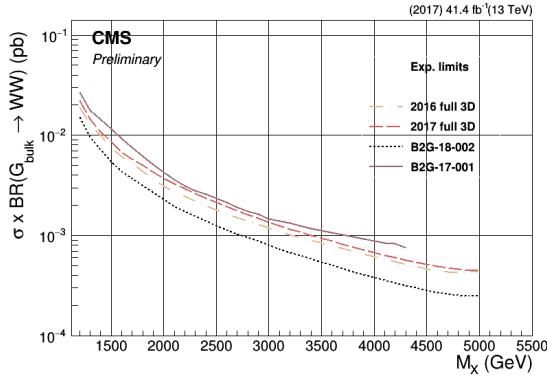


Figure 9.58: Expected limits for a Bulk $G \rightarrow WW$ signal using 35.9 (beige dashed line), 41.4 (pink dashed line), and 77.3 fb^{-1} of data using the 3D fit method (dotted black line), and using 35.9 fb^{-1} of 2016 data and the 1D fit method (solid purple line).

2889 Comparison with ATLAS

2890 Figure 9.59 shows the comparison with the expected limits obtained by the ATLAS collaboration
 2891 in a similar search for a VV resonance decaying to hadrons [83], for W' and Z' signal hypotheses.
 2892 For both signal hypotheses we obtain similar, or slightly better results by up to 35 %.

2893 Limits per category

2894 In Fig. 9.60 we also show the expected limits for a Bulk $G \rightarrow WW$ signal separately for 2016 and
 2895 2017, and for the two HPHP and HPLP categories. As a result of the jet substructure selections
 2896 optimization, the HPHP category dominates at low masses while the HPLP adds sensitivity at
 2897 higher values of the signal mass.

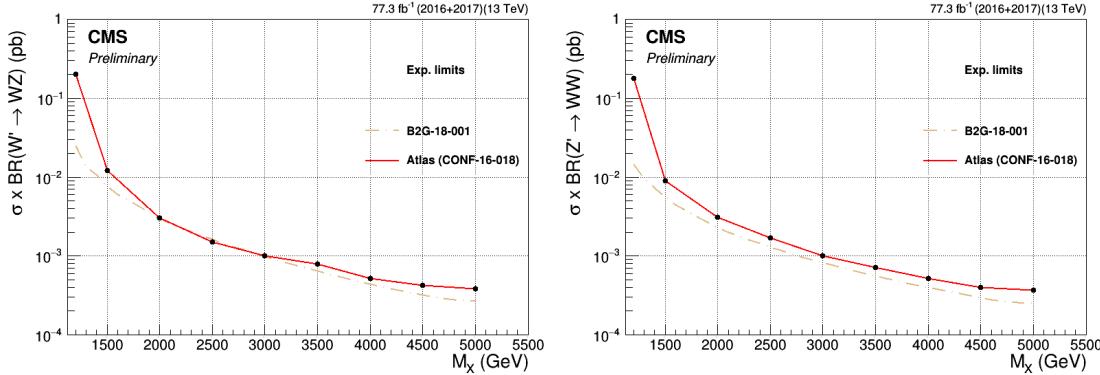


Figure 9.59: Comparison with the expected limits obtained by the ATLAS collaboration in a similar search for a VV resonance decaying to hadrons [83], for W' (left) and Z' (right) signal hypotheses.

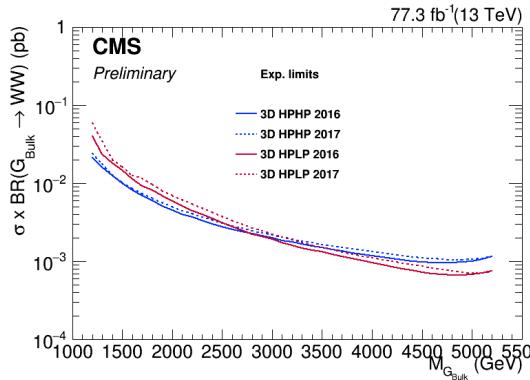


Figure 9.60: Comparison of the sensitivity of the two HPHP and HPLP categories for a Bulk $G \rightarrow WW$ signal hypothesis.

2898 9.10.2 Pulls of nuisance parameters

2899 As summarized in Section 9.8, we add a list of systematic uncertainties to the fit as nuisance
2900 parameters. To quantify the difference between the expected values of the systematic uncertainties
2901 and the observed values, we compute the pull of the nuisance parameters,

$$2902 p_\theta = (\theta - \theta_{in})/\sigma_\theta, \quad (9.12)$$

2902 where θ_{in} is the pre-fit value of the nuisance parameter under consideration, θ the corresponding
2903 parameter post-fit, and σ_θ the post-fit error. The error on p_θ is calculated as the ratio of post-
2904 and pre-fit uncertainties. Figure 9.61 shows the pulls after a signal and background fit to the

2905 combined 2016+2017 dataset (left) and when fitting the two separately (right), here using a
2906 signal hypothesis corresponding to a 2 TeV G_{bulk} . We observe that the W-tagging efficiency
2907 (“CMS_VV_JJ_tau21_eff”), the softdrop jet mass scale (“CMS_scale_prunedj”) and the resolution
2908 (“CMS_res_prunedj”) gets pulled and constrained by the W+jets and Z+jets mass peaks. In
2909 addition, the QCD shape parameters (“CMS_VV_JJ_nonRes_”*) are significantly pulled and
2910 constrained by data because of their large pre-fit uncertainty and unknown a-priori pre-fit value
2911 (again, we do not know if Nature is PYTHIA8, HERWIG++ or MADGRAPH+PYTHIA8. Though
2912 from this measurement, HERWIG++ seems to be favored by data (“altshape2”)).

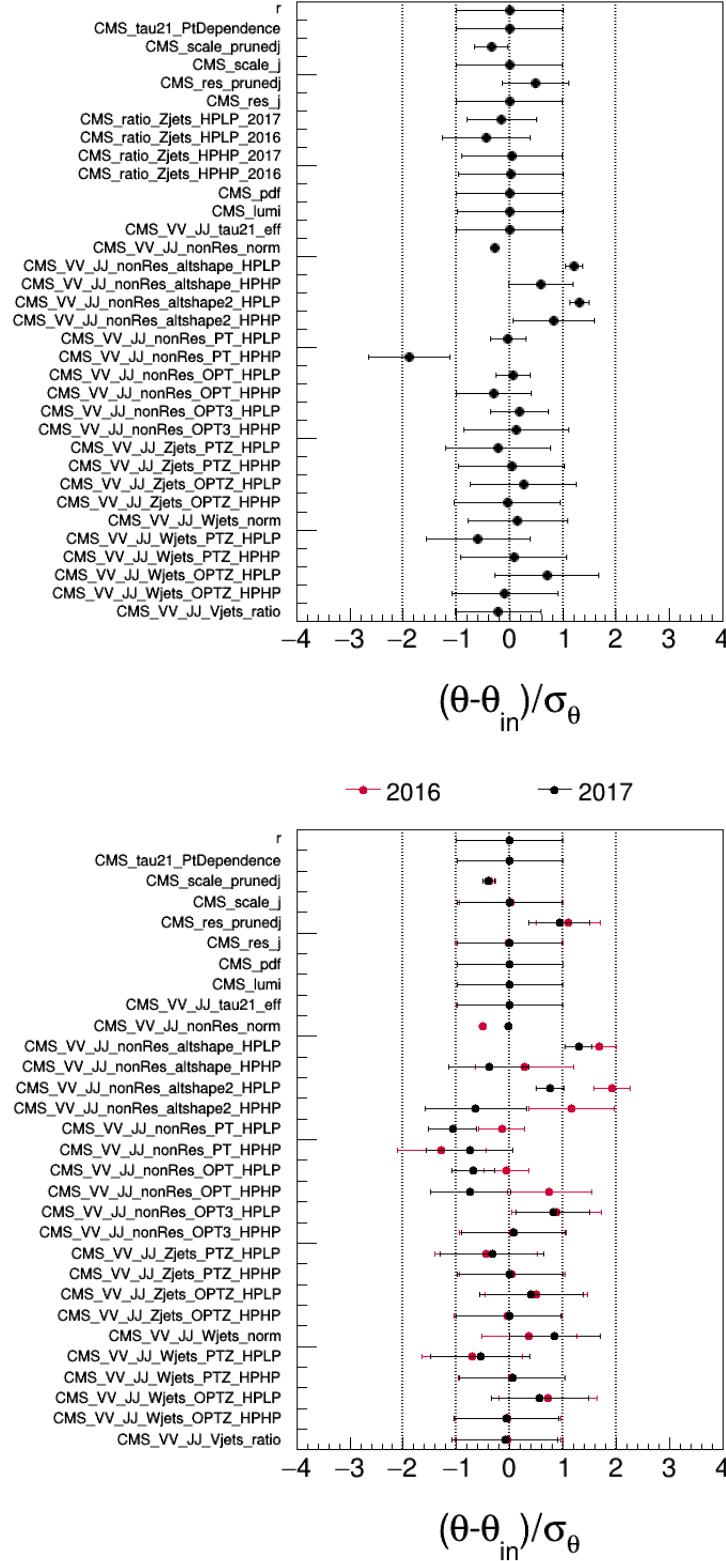


Figure 9.61: Pulls of each nuisance parameter for a combined signal+background fit to the combined 2016+2017 dataset (top) and when fitting the two separately (bottom).

2913

CHAPTER 10

2914

Summary

2915 In chapters 7, 8, and 9, we have followed the search for VV resonances in the all-hadronic final state
2916 through three stages: from being one of the first-ever analyses in the “boosted” final state with a
2917 center-of-mass energy of 13 TeV, and the very first CMS analysis to take advantage of substructure
2918 at the trigger level, through leading the development of a new W-tagging algorithm and mass
2919 corrections now default in CMS, and finally ending with the development of a multi-dimensional
2920 fit for generic searches in the groomed jet mass and dijet invariant mass.

2921 Each analysis has built on findings and improvements that came from the analysis before it:
2922 the substructure triggers and mass-corrected softdrop jet mass are both used for the 3D fit; the
2923 early discovery of the softdrop signal efficiency dependence on p_T led us to derive corrections for
2924 it; the observed lack of a constraint on the fit in the dijet invariant mass tail when statistics were
2925 low led to us exploring alternate methods that utilize simulation rather than data to create a
2926 more robust background prediction in the 3D fit; and the observation of large differences between
2927 predictions from different QCD MC generators led us to derive a clever way in which we can
2928 incorporate all of these predictions into one single fit. Now the question which remains is: *What*
2929 *comes next?*.

2930 A few ideas were already mentioned in the introduction to Search III, Section 9. The natural
2931 next step for this search is an incorporation of the VH(bb) and H(bb)H(bb) searches into the
2932 three-dimensional fit. Orthogonality between the three is guaranteed through b-tagging categories,
2933 as illustrated in Figure 10.1. This process is already underway, and we are aiming for a publication
2934 based on the full Run 2 dataset (data collected in 2016, 2017 and 2018) in one common framework.

2935 Secondly, there will be an increase in center-of-mass energy to 14 TeV in Run 3 (2020-2022).
2936 After this, there is not planned an increase in the collision energy. That means that better
2937 sensitivity will be achieved only through increases in integrated luminosity or better analysis
2938 technique, such as a better procedure for background estimation, a better fit procedure, or a

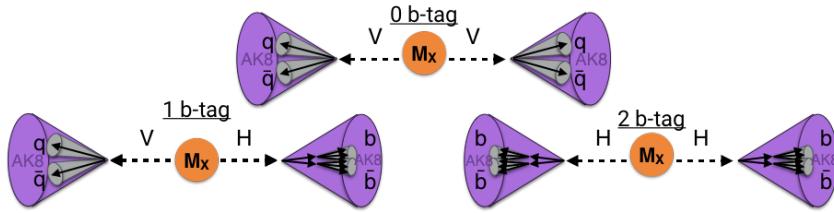


Figure 10.1: The VV , $VH(bb)$, and $H(bb)H(bb)$ analyses can all be incorporated into the multidimensional framework. Orthogonality between the analyses is ensured through b-tagging categories.

2939 better V tagger.

2940 Beyond that, and perhaps more interestingly, is the search for generic resonances peaking
 2941 anywhere in the jet mass and dijet invariant mass spectrum, where the jets themselves could
 2942 have other compositions other than two subjets (for instance, a scalar decaying to two vector
 2943 bosons, whose decay products are merged into one jet, resulting in a 4-prong object). One caveat
 2944 of the current analysis setup is that it only works for two-pronged signals due to its n-subjettiness
 2945 cut. In order for the multi-dimensional framework to be truly generic, the τ_{21} tagger needs to be
 2946 replaced by a generic anti-QCD tagger. Such a tagger works as an anomaly detector by encoding
 2947 the probability density function for quark or gluon jets as a function of certain variables, variables
 2948 for which signal jets are assumed to have a different probability density function. In our case,
 2949 good variables would for instance be groomed jet mass or substructure, as any generic signal is
 2950 assumed to be peaking in jet mass and have some (unknown) substructure. Such a tagger would
 2951 not require information about what a potential signal would look like; it would return only the
 2952 probability of any particular jet being a QCD jet. Taggers like these are usually based on deep
 2953 neural networks (DNN), where the quark and gluon jet PDFs are obtained through training of
 2954 the network, as demonstrated in Refs. [84, 85].

2955 In order for such an encoding to work, the DNN needs access to the features distinguishing
 2956 quark and gluon jets from signal jets without these features being biased towards any signal in
 2957 particular. The network must learn how to encode “non-substructure”.

2958 As a side project in parallel to working on the multi-dimensional fit, I spent the last half year of
 2959 my PhD working on a deep neural network capable of discriminating quark and gluon jets from
 2960 W jets in order to improve the W-tagging performance and improve the search sensitivity for
 2961 VV analyses to come. Based solely on the four vectors of jet constituents, the idea is to let the
 2962 neural network itself compute variables based on grooming and substructure, without feeding
 2963 it any high-level features (like softdrop mass and τ_{21}). This type of architecture is, in addition
 2964 to improving W-tagging performance, ideal for the purpose described above: encoding QCD in

²⁹⁶⁵ terms of substructure-like features. The final chapter of this thesis is therefore dedicated to the
²⁹⁶⁶ two last points: how to improve W-tagging in CMS for future analyses and how to design a neural
²⁹⁶⁷ network capable of learning jet substructure in an unbiased way.

2968

Part IV

2969

Encoding jet substructure in a deep neural network

2970

2971 Infusing deep neural networks with physics

2972 The previous chapter ended by mentioning two ingredients that will become important for future
 2973 searches with the multi-dimensional fit: a better vector-boson tagger, and a generic anti-QCD
 2974 tagger for signal-independent searches. As a side project during my final PhD semester, I worked
 2975 on a solution for the first, which has the added benefit of being a stepping stone towards the latter.
 2976 This is what I will cover in the final chapter of this thesis.

2977

2978 When applying machine learning to particle physics problems, the input has historically con-
 2979 sisted of pre-computed high-level features (quantities based on lower-level variables and certain
 2980 theoretical assumptions). With the rise of deep learning however, computational graphs have
 2981 achieved an increased capability to find even the smallest correlations in datasets, allowing them
 2982 to construct complex features on their own. The deep neural network (DNN) I will present in
 2983 the following is based on the assumption that, given sufficient instructions about the laws of
 2984 Nature, a neural network should be capable of reconstructing its own high-level features based on
 2985 lower-level quantities only. In addition, if smartly designed, the network should be capable of
 2986 finding novel correlations and physical features, *a-priori* unknown, by allocating a physical meaning
 2987 to the training weights deep within the network. The deep neural network I will present here is
 2988 trained to discriminate quark and gluon jets from W -jets. However, as I will discuss in the final
 2989 section of this chapter, it is also the perfect starting point for developing a generic anti-QCD tagger.

2990

2991 The work presented in the following has not been published and still qualifies as work in progress.
 2992 However, I believe developing taggers such as these is of great importance for future versions of
 2993 the searches presented here, and it is something I hope to continue working on in the future.

The notes show a diagram of four vectors E_1, E_2, P_1, P_2 forming a parallelogram. Below the diagram, the text 'COLA:' is written above a matrix equation:

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

Below this, another matrix equation is shown:

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

Further down, a diagram shows a vector k_{M} as a sum of vectors E_1, E_2, P_1, P_2 , with the text '(4,4)' above it. To the right, a detailed calculation of the components of k_{M} is shown:

$$\begin{aligned} k_{\text{M},1} &= E_1 + E_2 + W_{11}P_1^2 + W_{12}P_2^2 \\ k_{\text{M},2} &= P_1^2 + P_2^2 + W_{21}P_1^y + W_{22}P_2^y \\ k_{\text{M},3} &= P_1^y + P_2^y + W_{31}P_1^z + W_{32}P_2^z \\ k_{\text{M},4} &= P_1^z + P_2^z + W_{41}E_1 + W_{42}E_2 \end{aligned}$$

Below these equations, the text 'linear combinations of momenta' is written, followed by:
 $\sum E$
 $\sum P_1$
 $\sum P_2$
 $\sum P_3$

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

Work in progress.

2994

CHAPTER 11

2995

A Lorentz invariance based Deep Neural Network for W-tagging

2996

2997

11.1 LoLa

2998 LoLa is a deep neural network architecture which was first introduced for top tagging [86]. It is
 2999 based on the idea that, given enough information about the laws of Nature, a neural network should
 3000 be capable of calculating jet substructure observables on its own given only low-level information.
 3001 The network is designed to discriminate between AK R=0.8 jets originating from W bosons from
 3002 those originating from quarks or gluons, solely based on the jet constituent four-vectors (variables
 with little discriminating power on their own) as illustrated in Figure 11.1. Rather than being

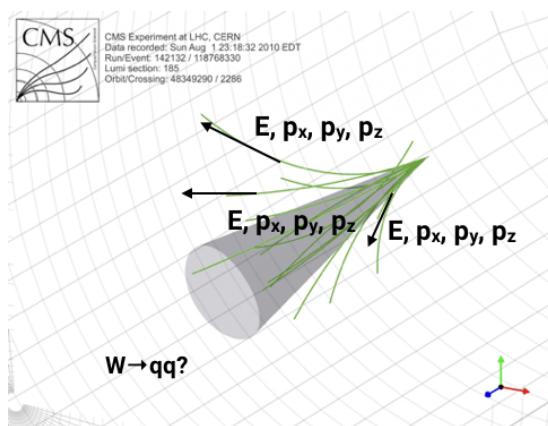


Figure 11.1: LoLa uses only jet constituent four-vectors as input to discriminate W jets from quark and gluon jets.

3003

3004 input high-level features, the neural network is given tools to perform calculations on Lorentz
 3005 vectors using the Minkowski metric. Through two novel layers, linear combinations similar to jet
 3006 clustering and jet substructure algorithms are performed, allowing the algorithm to create its
 3007 own substructure variables. Additionally, training weights deep within the network correspond
 3008 to physical quantities reconstructed by the algorithm; distance between particles, masses and
 3009 energies, linear combinations of particle four-vectors, etc. Besides the end goal of discriminating
 3010 W bosons from quarks and gluons, one could therefore hope to learn of new correlations separating
 3011 QCD from vector-boson jets.

3012 11.1.1 Architecture

3013 The LoLa architecture is designed as a four-layer deep, feed-forward sequential network doing
 3014 supervised learning on fixed-size input vectors. Two novel layers are introduced, the Combination
 3015 Layer (CoLa) and the Lorentz Layer (LoLa), which perform basic jet clustering and substructure
 3016 calculations as well as implement the Minkowski metric. These two layers are followed by two fully
 3017 connected layers, consisting of 100 and 50 nodes respectively, before the final output is computed
 3018 yielding two output probabilities between 0 and 1 (using a Softmax activation function [87]).
 3019 The loss function to be minimized is “categorical crossentropy” (or log loss) with two categories
 3020 corresponding to the probability of a given jet to be W versus non-W like. Only the W-jet
 3021 probability is stored. The optimizer used in the training is the ADAM optimizer [88], which
 3022 adapts the learning rate of the model parameters during training. The code itself is written using
 3023 the Keras [89] interface with a TensorFlow [90] backend. The full architecture with input and
 3024 output dimension per layer is shown in Figure 11.2. The three first boxes, representing the input,
 3025 CoLa and LoLa layers, correspond to tensors of rank 2, while the final four boxes correspond to
 3026 tensors of rank 1. In the following, each layer will be explained in detail.

3027 11.1.2 Input

3028 This algorithm is trained to discriminate between fully merged hadronic W-jets coming from the
 3029 process $G_{\text{bulk}} \rightarrow WW \rightarrow q\bar{q}q\bar{q}$ (where $M_{G_{\text{bulk}}} = 0.6 - 4.5 \text{ TeV}$), and quark or gluon jets from a
 3030 QCD sample generated with PYTHIA8. All jets are clustered with the anti- k_T algorithm with a
 3031 distance parameter of $R=0.8$, with the PUPPI pileup removal algorithm applied. In addition, they
 3032 are required to have $p_T > 200 \text{ GeV}$ and $|\eta| < 2.5$. Jets are defined as W jets if they are matched
 3033 to a generator-level, hadronically decaying W boson, with the following matching criteria. The
 3034 generated vector boson needs to be within $\Delta R < 0.6$ of the jet axis, and the quark decay products
 3035 need to be within $\Delta R < 0.8$ of the jet axis. The p_T and η distribution of signal and background

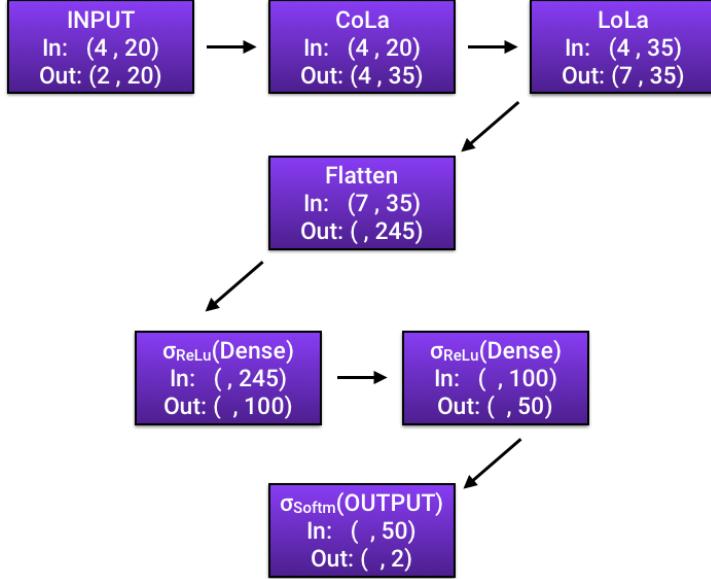


Figure 11.2: The full LoLa architecture. “In” denotes the dimension of the input tensor to the given layer, “Out” is the dimension of the output tensor.

3036 jets, is shown in Figure 11.3. From these signal and background jets, only the four-vectors of the
 3037 $N = 20$ highest- p_T jet constituents are used as input to the deep neural network: E , p_x , p_y and
 3038 p_z . I use 20 constituents since any larger number has a negligible effect on the performance, while
 3039 performance tends to drop once going below 15. The input is therefore a $4 \times (N = 20)$ matrix for
 3040 each signal and background jet, with one four-vector for each of the 20 jet constituents:

$$x_{\mu,i} = \begin{pmatrix} E^1 & E^2 & \dots & E^N \\ p_x^1 & p_x^2 & \dots & p_x^N \\ p_y^1 & p_y^2 & \dots & p_y^N \\ p_z^1 & p_z^2 & \dots & p_z^N \end{pmatrix}. \quad (11.1)$$

3041 The total number of jet constituents are shown in Figure 11.4, and the input variables (here
 3042 for all constituents) are shown in Figure 11.5. It is clear that the input variables provide
 3043 little discriminating power on their own. Therefore, the network must learn how to derive other
 3044 physical quantities where the signal and background PDFs differ to a larger extent. This is
 3045 achieved through the two custom layers described in the following.

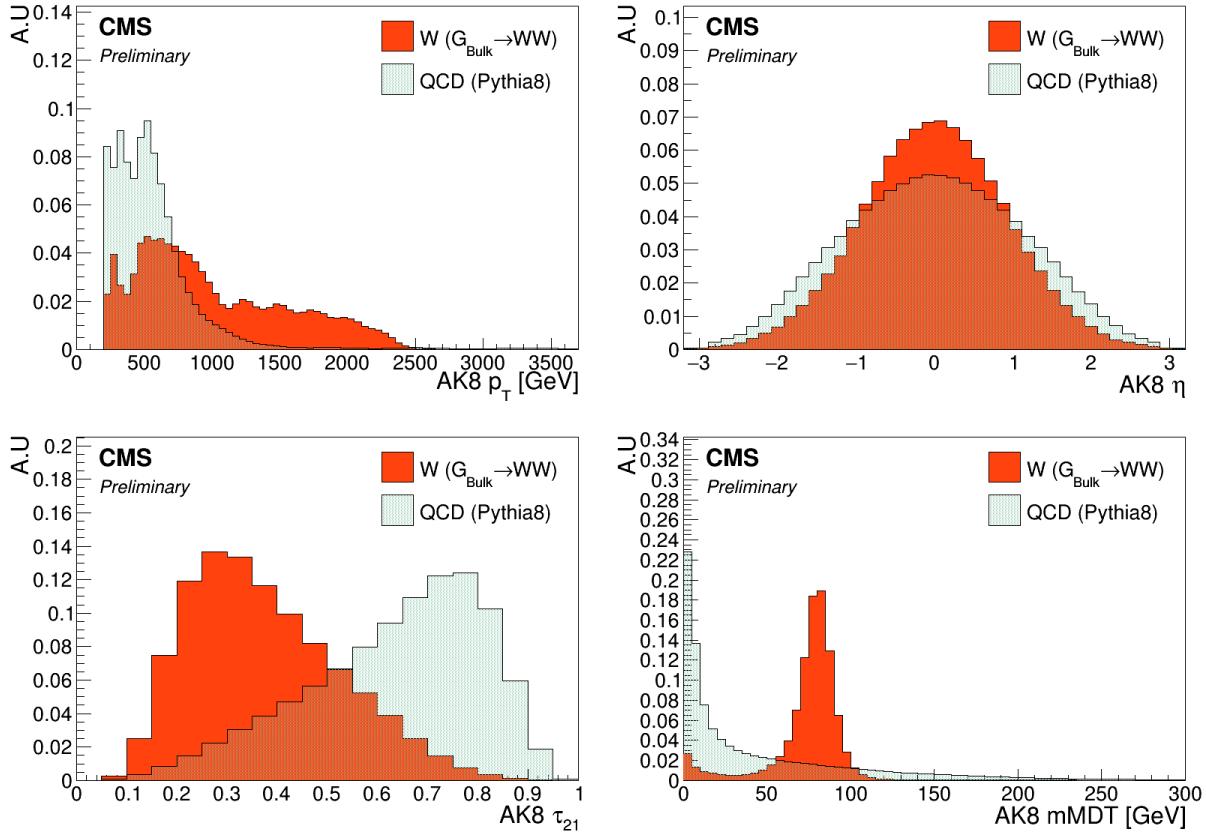


Figure 11.3: Jet p_T (top left), η (top right), τ_{21} (bottom left) and softdrop jet mass (bottom right) for signal and background jets.

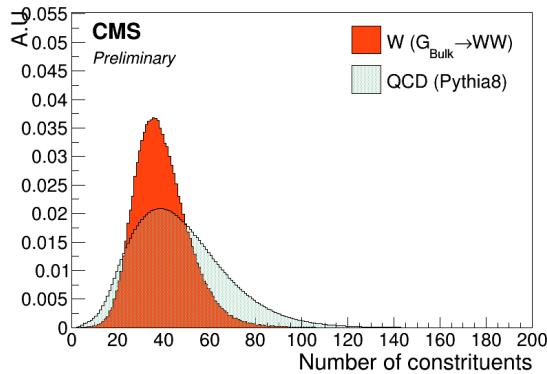


Figure 11.4: The number of jet constituents for signal (red) and background (blue). Only the 20 highest- p_T constituents are used during training.

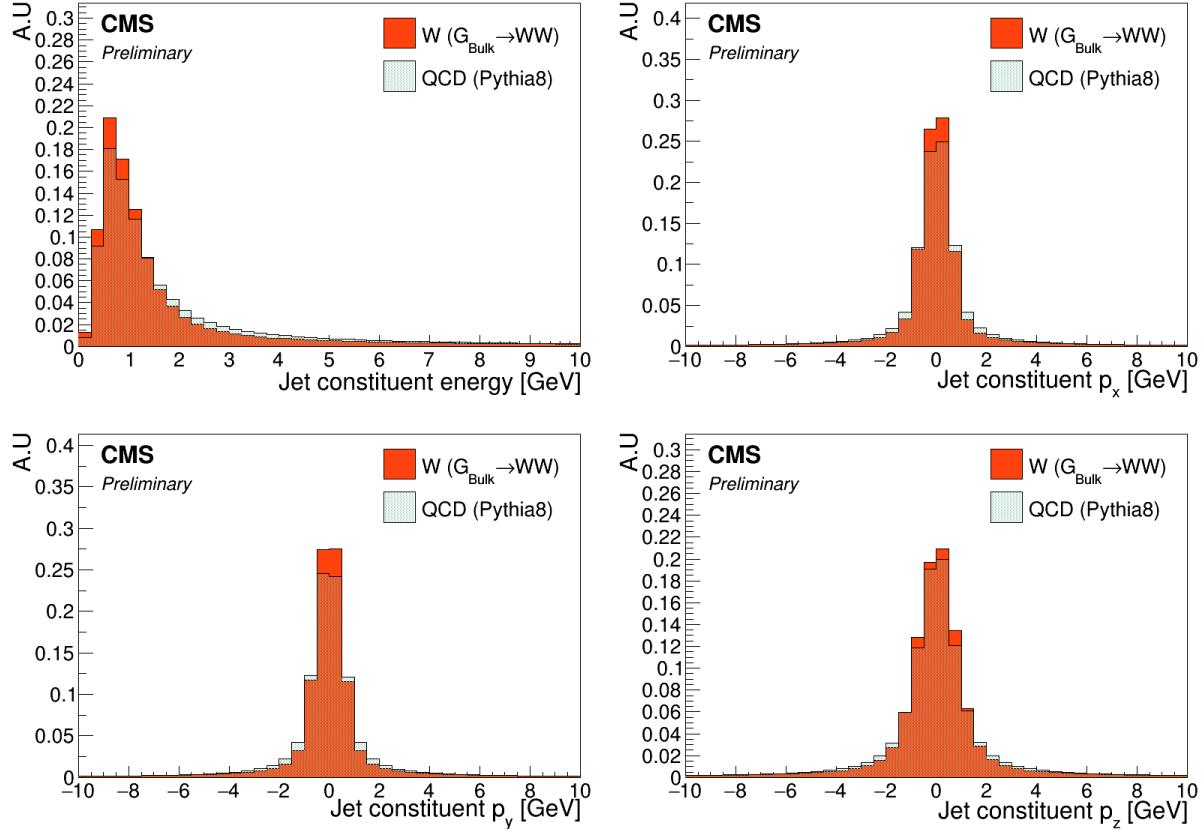


Figure 11.5: Energy (top left), p_x (top right), p_y (bottom left), and p_z (bottom right) for all jet constituents. These values are used as input to the neural network training.

11.1.3 The Combination Layer

The Combination Layer (CoLa) consists of a matrix which, when taking the scalar product with the input matrix, compute linear combinations of the jet constituents, similar to what is done in recombination jet algorithms. The main goal is to create additional four-vectors as input for the next layer. The CoLa matrix is a concatenation of the following: a vector of 1's of length N , the $N \times N$ identity matrix, and a matrix of $N \times M$ trainable weights:

$$C_{i,j} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 & w_{1,N+2} & w_{1,N+3} & \dots & w_{1,(N+2)+M} \\ 1 & 0 & 1 & \dots & 0 & w_{2,N+2} & w_{2,N+3} & \dots & w_{2,(N+2)+M} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & 0 & 0 & 0 & 1 & w_{N,N+2} & w_{N,N+3} & \dots & w_{N,(N+2)+M} \end{pmatrix}. \quad (11.2)$$

3052 When performing the following multiplication

$$x_{\mu,i}^C = x_{\mu,i} C_{i,j}, \quad (11.3)$$

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:

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

3053 In the two last columns, the neural network makes two “subjet” four-vectors by weighting the
 3054 relative contribution of each particle as it sees fit. This is similar to jet grooming (Section 6.5.1)
 3055 or PUPPI pileup subtraction (Section 6.3.2), and should allow the network to learn which
 3056 constituents are part of the hard scatter and which are not. The $x_{\mu,i}^C$ matrix is finally passed on
 3057 to the next layer, the Lorentz Layer.

3058 11.1.4 The Lorentz Layer

3059 The Lorentz Layer (LoLa) is responsible for encoding how particles move in space-time through
 3060 a simple set of rules. Each column (four-vector) of $x_{\mu,i}^C$, is used to compute, and afterwards is
 3061 replaced by, the following $k = 7$ features:

$$x_{k,i}^L = \begin{pmatrix} m^2(x_{\mu,i}^C) \\ p_T(x_{\mu,i}^C) \\ w_{ij}^E E(x_{\mu,j}^C) \\ w_{ij}^{s1} \sum d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{s2} \sum d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{m1} \min d^2(x_{\mu,i}^C, x_{\mu,j}^C) \\ w_{ij}^{m2} \min d^2(x_{\mu,i}^C, x_{\mu,j}^C) \end{pmatrix} \quad (11.4)$$

3062 Going through from top to bottom, these are:

- 3063 • the invariant mass and p_T of each four-vector,
- 3064 • a linear combination of all four-vector energies where each is scaled by a trainable weight,
- 3065 • two times the sum of distances between the four-vector under consideration and every other
 3066 column reweighted with a trainable weight, and
- 3067 • two times the minimum distance between the four-vector under consideration and every
 3068 other column where each distance again is reweighted by a trainable weight.

3069 The Minkowski metric enters explicitly in the first and in the last four calculations, where the
 3070 neural network is told to abide by the rules

$$m^2(x_{\mu,i}^C) = g^{\mu\nu} x_{\mu,i}^C x_{\nu,i}^C \quad (11.5)$$

3071 and

$$d^2(x_{\mu,i}^C, x_{\mu,j}^C) = (x_{\mu,i}^C - x_{\mu,j}^C)_\mu g^{\mu\nu} (x_{\mu,i}^C - x_{\mu,j}^C)_\nu \quad (11.6)$$

3072 with $g^{\mu,\nu} = [-1, 1, 1, 1]$, when calculating the invariant mass and distance between all four-vectors
 3073 produced by CoLa. This tells the neural network to use a space-time geometry in all its calculations
 3074 and to respect Lorentz invariance. The four final rows of LoLa are the most interesting: Here
 3075 the network computes quantities similar to n-subjettiness by summing up the distances between
 3076 all constituents, the jet axis and the subjets produced by CoLa. If, for instance, the network

3077 was able to reconstruct the two hard subjets in the final columns of CoLa, which create linear
3078 combinations of particles, it can create its own “ τ_2 ” variable by taking the distance between those
3079 subjets and all the jet constituents, and weighting down the column corresponding to the full
3080 jet four-vector in column one. Then it can do the same by calculating the distance between the
3081 full jet four-vector and all constituents, now weighing down the linear combinations made by
3082 CoLa, and compute “ τ_1 ”. The two custom layers, CoLa and LoLa, therefore come together in
3083 order to encode jet clustering and substructure in a novel way. They provide the network with
3084 the necessary tools in order to create its own physical quantities, through linear combinations
3085 with trainable weights, which then again are used to produce other physical quantities with
3086 new trainable weights. This allows the network full freedom to explore all interesting particle
3087 correlations, where the resulting output features have a physical meaning that can be probed.
3088 LoLa turns the question “What can we teach the machine?” around to “What can we learn from
3089 the machine?”.

3090 11.2 Performance

3091 The deep neural network is trained on 320,000 signal and background jets for up to 100 epochs,
 3092 but allow for an early stopping after ten epochs if the loss is stable. The test sample consists
 3093 of 60,000 W jets and 60,000 quark and gluon jets. To quantify the performance, we look at
 3094 the signal efficiency versus mistagging rate comparing the performance of LoLa to that of the
 3095 taggers used previously in this thesis: PUPPI softdrop with τ_{21} and PUPPI softdrop with τ_{21}^{DDT} .
 3096 The performance of these three different taggers, is shown in Figure 11.6. The point where the
 3097 blue curves end, represent the signal efficiency for a mass cut of $65 < \text{Softdrop jet mass} < 105$
 GeV, here roughly 70%. We clearly see that LoLa performs significantly better than the current

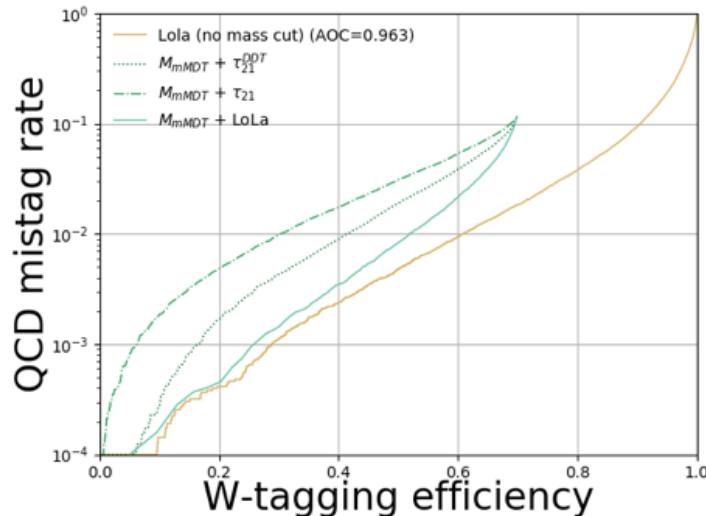


Figure 11.6: Performance of LoLa compared to other W-tagging discriminants in the background-signal efficiency plane: PUPPI softdrop + τ_{21} (dashed blue), PUPPI softdrop + τ_{21}^{DDT} (dotted blue), LoLa with a softdrop mass window applied (solid blue) and the nominal LoLa tagger with no mass cut applied.

3098 baseline W-taggers based on τ_{21} and τ_{21}^{DDT} , with a roughly 20% higher signal efficiency at a given
 3099 mistagging rate. LoLa also has a higher signal acceptance, as it can be used without a mass
 3100 window applied. If LoLa were to replace the tagger used in Search II (a better comparison than
 3101 Search III as the latter uses a rather unconventional mass window), which has a signal efficiency
 3102 of $\sim 42\%$ at a 2% mistagging rate for a single jet, the signal efficiency for the same mistagging
 3103 rate would be 65%, a 55% increase. For an analysis requiring two tagged jets, that would imply
 3104 going from an 18 to a 43% total signal efficiency, a significant gain.
 3105

3106 11.3 Dependence on jet mass and p_T

3107 Despite being a key feature, absolute performance is not all that quantifies how good one tagger
 3108 is compared to another. To use a tagger in a physics analysis, there are three key questions one
 3109 needs to consider:

- 3110 • Is the absolute performance better (compared to common methods)?
- 3111 • Is the tagger p_T -dependent?
- 3112 • Does the tagger sculpt the mass spectrum?

3113 These three measures are important in quantifying performance and, in the following, I will
 3114 attempt to explain why this is the case and which approaches are used here in order to tackle
 3115 them.

3116 Any deep neural network trained to distinguish W jets from quark and gluon jets will naturally
 3117 learn that p_T and mass are discriminating features unless it is penalized for it. Figure 11.7 shows
 the LoLa discriminant as a function of jet p_T and softdrop jet mass. A strong correlation between

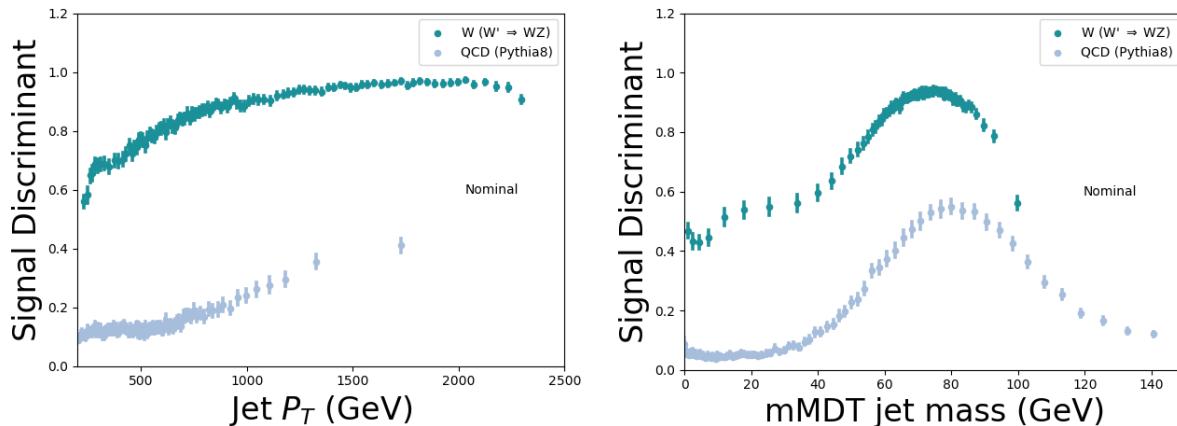


Figure 11.7: The LoLa discriminant as a function of jet p_T (left) and softdrop jet mass (right). A strong correlations with both variables is observed.

3118
 3119 the LoLa discriminant and jet mass and p_T , is observed both for signal and for background jets
 3120 (closer to 1 means more signal like), with a rising slope as a function of p_T (meaning the network
 3121 interprets a higher jet p_T as being more signal-like) and a bump around the W mass for both
 3122 signal and for background.

3123 11.3.1 p_T -decorrelation

3124 A tagger which is p_T -dependent is problematic due to the following reasons: firstly, the signal
3125 efficiency is variable, which requires a working point that scales with p_T . That in itself is not
3126 problematic and can easily be computed. However, it implies that when computing efficiency
3127 scale factors from data, a range of different scale factors for different working points is required.
3128 In addition, the scale factors are measured at low p_T , a region where the tagging efficiency can
3129 be substantially different from the analysis signal region due to the strong p_T -correlation present.
3130 Finally, the dijet invariant mass is intrinsically linked to the p_T -spectrum, meaning that any
3131 p_T -dependence can introduce sculpting of the dijet invariant mass spectrum. From the top left
3132 plot in Figure 11.3, one clearly sees that the jet p_T distribution is very different for signal and for
3133 background. In order to avoid that the network learns jet p_T to be a discriminating feature, I
3134 therefore compute a jet-by-jet weight intended to flatten the jet p_T spectrum. This weight is passed
3135 as a sample weight to the training set, reweighting the contribution from each jet to the total loss
3136 (making high-mass QCD jets and low-mass signal jets count more). Figure 11.8 shows the jet
3137 p_T -distribution without any p_T -reweighting applied (solid lines) and after applying a p_T -weight
 (dashed lines). The training is then repeated, this time passing a sample weight with each jet,

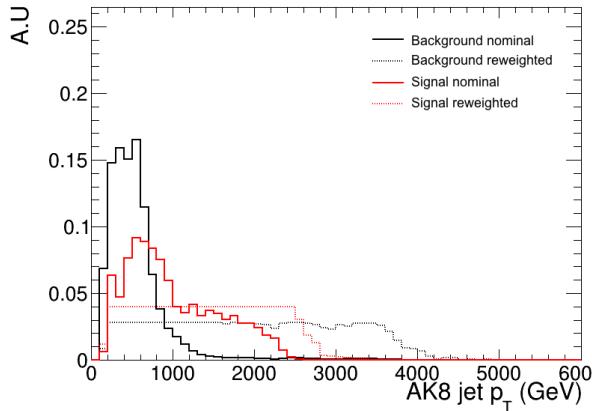


Figure 11.8: Jet p_T distribution before (solid lines) and after (dashed line) applying a weight intended to flatten the jet p_T spectrum.

3138
3139 and the final discriminant compared to the nominal training. Figure 11.9 shows the performance
3140 of the same taggers as above but with one additional line, which is LoLa p_T -reweighted. A
3141 clear drop in performance is observed, as expected when removing information from the training.
3142 However, when we again look at the discriminant output as a function of jet p_T in Figure 11.10,
3143 the correlation we observed before has vanished and we are left with a tagger not depending

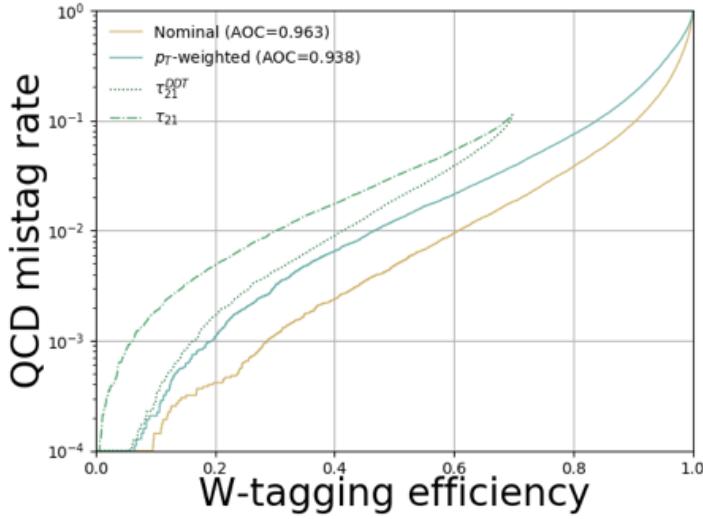


Figure 11.9: Performance of the p_T -reweighted LoLa tagger (solid blue) and the nominal LoLa tagger (solid yellow).

on the jet p_T . For completeness, Figure 11.11 shows the τ_{21} and τ_{21}^{DDT} discriminants versus jet

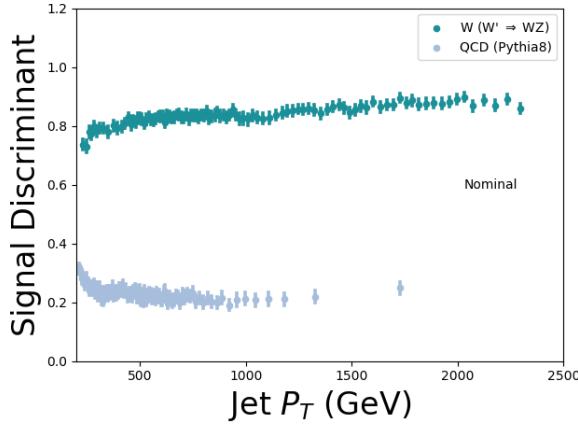


Figure 11.10: The LoLa discriminant as a function of jet p_T after training with a weight intended to flatten the sample p_T spectrum.

3144

3145 p_T . Whereas the nominal LoLa discriminant had a much larger correlation with jet p_T than
 3146 the τ_{21} -based taggers, the p_T -reweighted version is as decorrelated from p_T as the τ_{21} variables
 3147 while still exhibiting a better absolute performance than the baseline taggers. In summary,
 3148 reweighting strategies such as the one described above yield a loss in overall performance, as
 3149 expected when removing information from the training. However, the p_T dependence of the

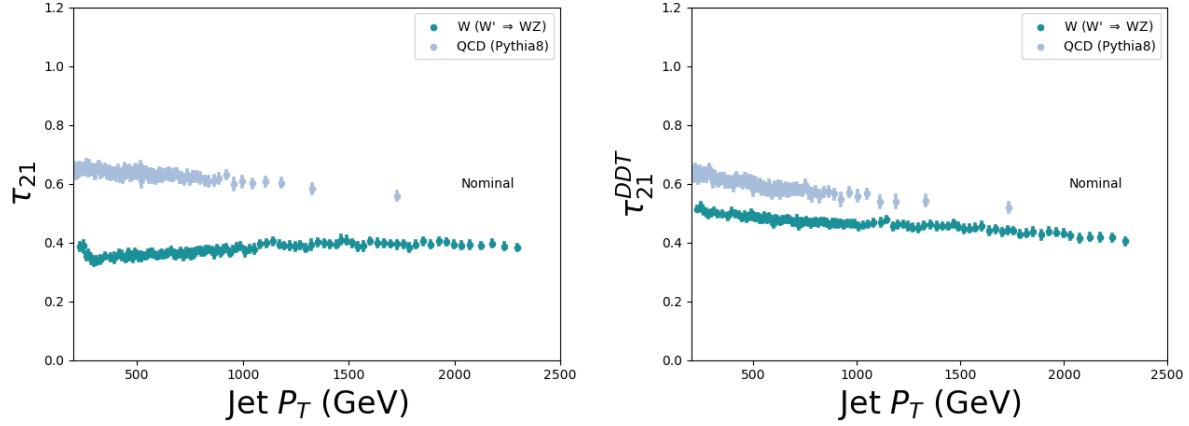


Figure 11.11: The τ_{21} (left) and τ_{21}^{DDT} (right) discriminant as a function of jet p_T .

tagger is strongly reduced, meaning that it might perform better overall in physics analysis when systematic uncertainties are taken into account. There is therefore no clear answer as to which method is better before running a full analysis including systematics for p_T -dependent tagging.

11.3.2 Mass sculpting

Any smart deep neural network intended to separate W bosons from quarks and gluons will inevitably learn the W boson mass as it clearly is very different from the quark or gluon mass. Unfortunately, as these taggers are meant to be used in physics analyses where we often estimate the background in mass sidebands, this has some undesired side effects. If a deep neural network has learned the mass then, after applying a cut on the discriminant, the background jet mass distribution becomes severely sculpted and difficult to constrain.

After applying a cut on the LoLa discriminant corresponding to a 1% mistagging rate, we see in the left plot of Figure 11.12 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 11.12, where all histograms are normalized to unit area, we see that the bulk of the remaining 1% QCD jets is right below the W-boson mass peak and has been sculpted to look exactly like the signal. This mass sculpting is in and on its own not a problem, the tagger still manages to get rid of most of the background. However, in many physics analyses, in order to evaluate the background rate in the data signal region, mass sidebands are used. If the background distribution is peaky rather than smoothly falling, the shape and consequently the expected yield is very difficult to constrain. That leads to large uncertainties

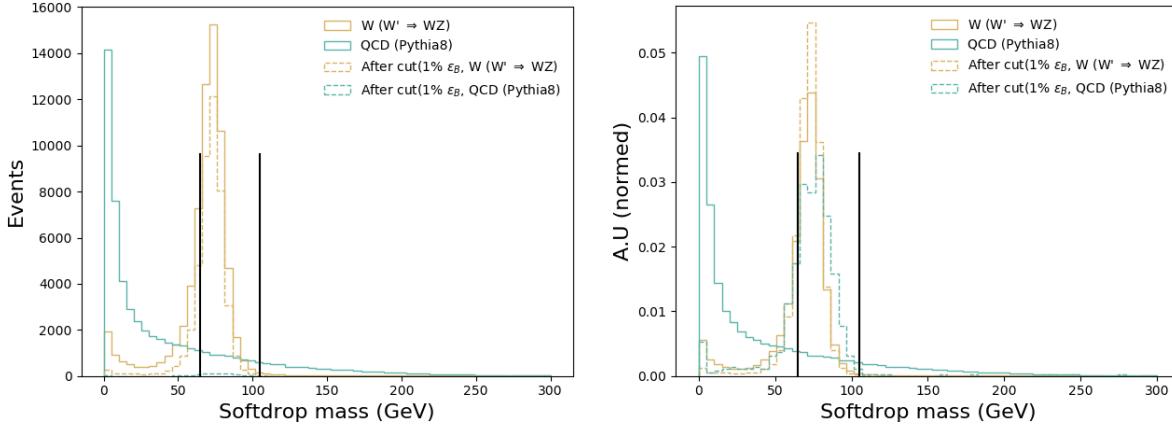


Figure 11.12: The softdrop jet mass distribution before (solid lines) and after (dotted lines) a cut on the LoLa discriminant corresponding to a 1% mistagging rate has been applied. The left plot shows the real number of events left after the cut, the right is normalized to area.

3171 on the background rate and might eventually make an analysis less sensitive than when using
 3172 a tagger with a worse absolute performance, but reduced mass correlation. In addition, if one
 3173 were to search for peaks in the softdrop jet mass, as is the case for the multidimensional fit, this
 3174 becomes increasingly difficult when attempting to fit a potential signal peak on top of a peaking
 3175 background.

3176 It should again be mentioned, that also for the baseline taggers based on τ_{21} , mass sculpting
 3177 is a known feature. Figure 11.13 shown the same softdrop jet mass spectrum before and after
 3178 a cut corresponding to a 1% mistagging rate on τ_{21} (left) and τ_{21}^{DDT} (right). Here τ_{21} clearly
 3179 exhibits mass sculpting, but is not as peaky as was the case for LoLa. The τ_{21}^{DDT} exhibits the
 3180 least amount of sculpting, but is also the tagger with the worst absolute performance.

3181 I have not yet had the chance to implement a mass decorrelation strategy for LoLa, but I
 3182 see two ways going forward: the first is, following the example of what was done to decorrelate
 3183 LoLa from jet p_T , to pass a mass-dependent sample weight to the training. LoLa would then be
 3184 trained with a weight derived to flatten the two dimensional jet mass - jet p_T plane. Another
 3185 option would be to train LoLa together with an adversarial, a dedicated deep neural network
 3186 running in parallel to LoLa and attempting to learn the jet mass from the LoLa output. The
 3187 total loss function would then be a sum of the two, where the better the adversarial is in learning
 3188 the mass, the worse the total loss function gets. Both of these options are something I would like
 3189 to explore in the future.

3190 In summary, mass- and p_T -dependence are in their own right not a problem for a tagger.
 3191 The problem occurs when using these taggers in actual physics analyses where background rate

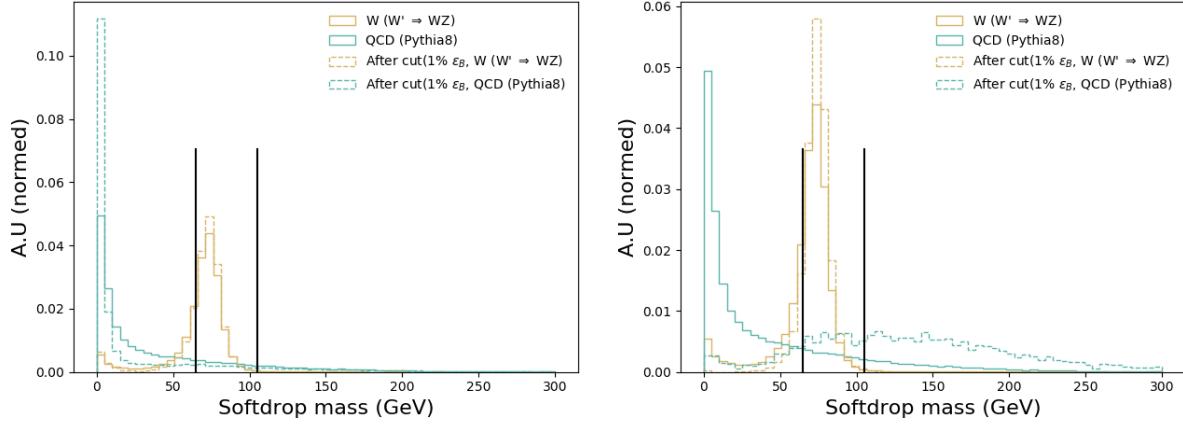


Figure 11.13: The softdrop jet mass distribution before (solid lines) and after (dotted lines) a cut on τ_{21} (left) and τ_{21}^{DDT} (right). All spectra are normalized to unit area.

uncertainties and tagging p_T -dependence uncertainties have a large impact on the final sensitivity.
 There is a trade-off between signal efficiency and (analysis-dependent) systematics. For LoLa,
 rather than choosing, I would like to provide two different taggers: a nominal tagger, where no
 mass/ p_T -decorrelation is attempted, and a decorrelated version. That would allow for both to be
 tested in a full analysis chain before deciding on which tagger to use when looking at data.

11.4 Validation

LoLa is additionally validated on independent samples as an unbiased measure of performance, allowing the comparison of different CMS W-tagging algorithms. The sample is based on a $Z' \rightarrow WW$ sample with $M_{Z'} = 3\text{ TeV}$ produced with MadGraph and a QCD PYTHIA 8 sample. Here, only jets with $1000\text{ GeV} < p_T < 1400\text{ GeV}$ and $|\eta| < 1.5$ are used. The signal efficiency versus mistagging rate for LoLa compared to the baseline PUPPI softdrop + τ_{21} tagger is shown in Figure 11.14. As was pointed out in Section 11.2, a mass selection is not necessary when using LoLa, but has been added to this plot for completeness. A significant improvement in tagging efficiency is observed for LoLa compared to the default tagger, also when being validated on a sample completely independent from the training sample. A selection corresponding to a 30% signal efficiency is used as a reference working point when studying the tagging performance as a function of jet p_T and pileup in the following, and is marked by triangles in Figure 11.14. The signal efficiency and mistagging rate as a function of jet p_T is shown in Figure 11.15. Again we observe the strong correlation between LoLa tagging efficiency and jet transverse momenta.

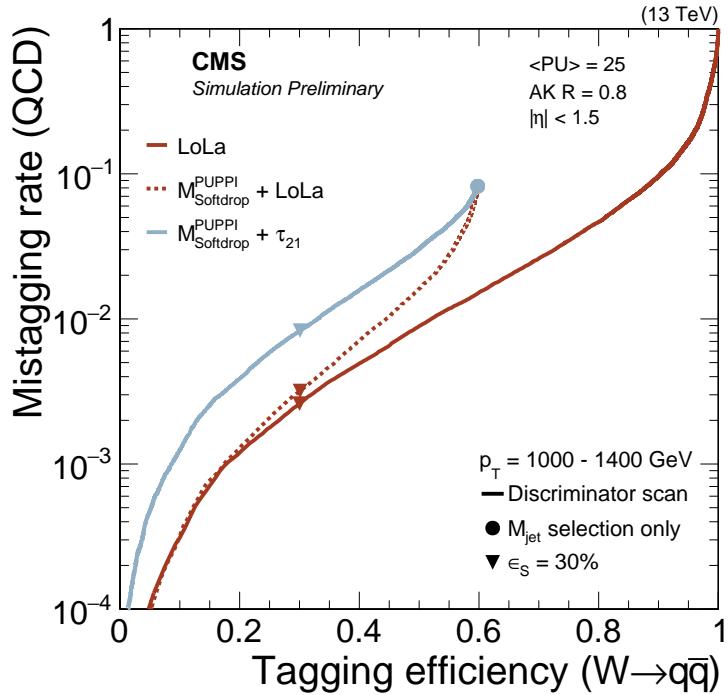


Figure 11.14: Performance of LoLa and PUPPI softdrop + τ_{21} in the background-signal efficiency plane. The PUPPI softdrop jet mass selection of $65 < M_{SD} < 105 \text{ GeV}$, and the 30 percent efficiency points are indicated with symbols.

3211 There is, however, no value for which the τ_{21} tagger has a higher signal over background ratio
 3212 than LoLa. LoLa performs its worst at very high jet p_T , but in this region the background is very
 3213 small (dijet invariant masses around 2.5-3 TeV), so the absolute performance here matters less
 3214 than at lower p_T . Figure 11.16 shows the tagging efficiency and mistagging rate as a function of
 3215 pileup. Both taggers under study are more or less decorrelated from pileup, with a flat efficiency
 3216 for up to 50 reconstructed primary vertices. In Run 3, this number is of course expected to
 3217 be significantly higher and the study should be redone up to a higher number of reconstructed
 3218 primary vertices.

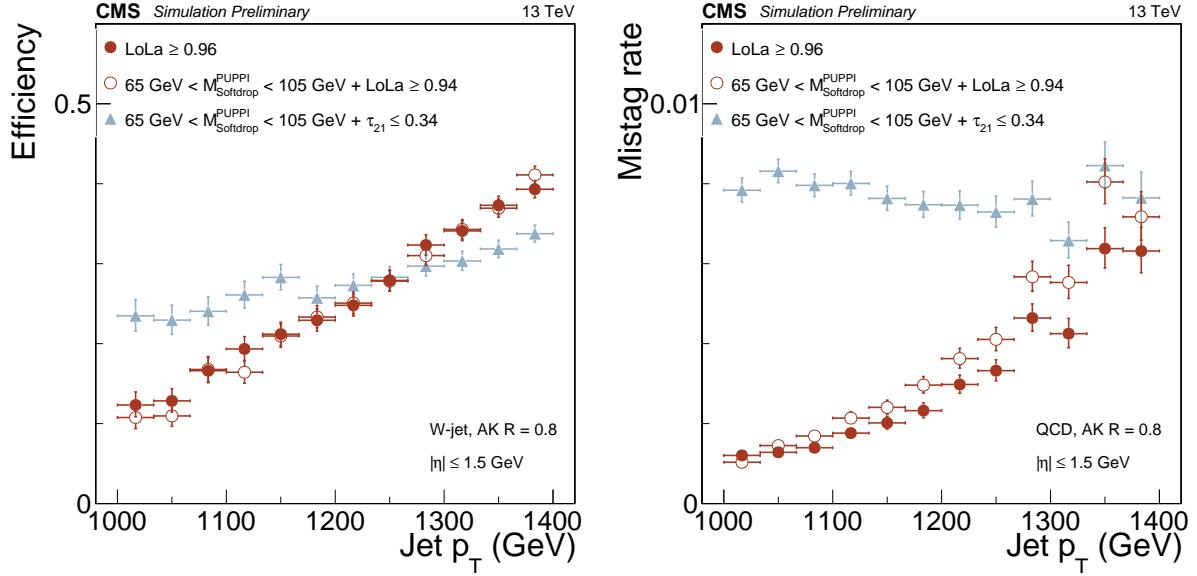


Figure 11.15: Efficiency (left) and mistagging rate (right) of the LoLa selection corresponding to a 30 percent signal efficiency as a function of jet p_T .

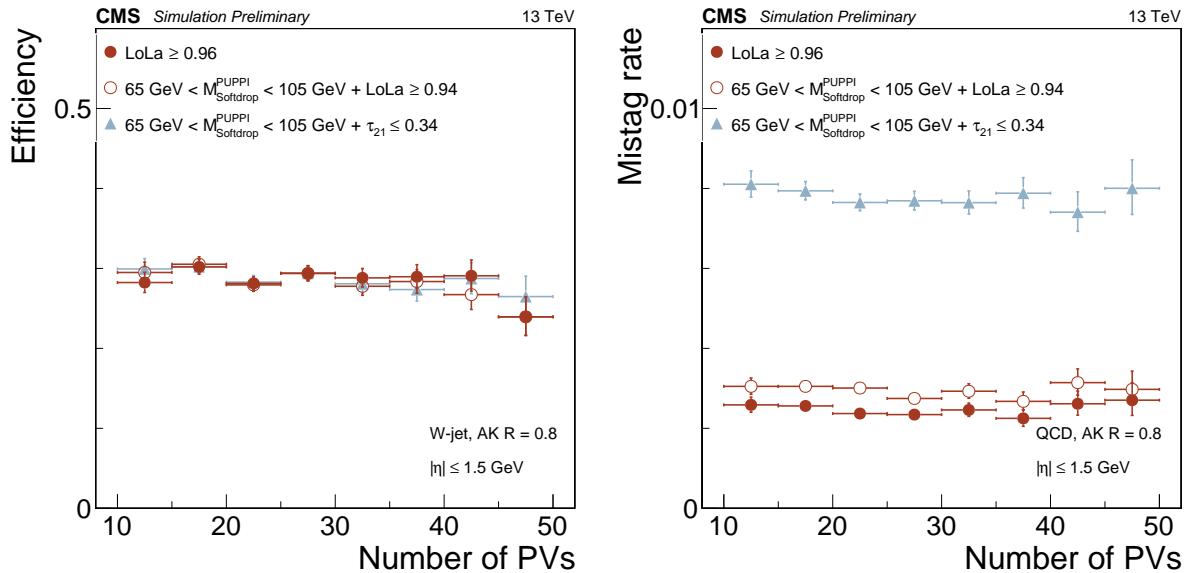


Figure 11.16: Efficiency (left) and mistagging rate (right) of the LoLa selection corresponding to a 30 percent signal efficiency as a function of the number of reconstructed vertices.

³²¹⁹ **11.5 What does LoLa learn?**

³²²⁰ As a first study of what features LoLa learns, we remove features from the LoLa and CoLa layers.
³²²¹ The first thing to investigate is what happens after removing variables sensitive to the jet mass
³²²² and p_T . This is shown in Figure 11.17. The light blue line, no $\sum p_T$, corresponds to removing
³²²³ the first column of CoLa which passes the sum of all particle four-momenta to LoLa. This has
³²²⁴ little impact on the total performance, meaning the algorithm does not rely on the jet axis
³²²⁵ being defined. Removing the Lola variables that correspond to the invariant mass or p_T of the
³²²⁶ particles/pseudo-jets (green line) reduce the performance significantly, which explains the large
³²²⁷ drop in tagging efficiency that was observed in Section 11.3.1 when decorrelating the tagger from
³²²⁸ jet p_T . The worst performance is obtained when removing both the first CoLa column, which
³²²⁹ roughly corresponds to the jet axis, as well as the invariant mass and p_T LoLa variables (pink
³²³⁰ line). This implies that a significant part of what LoLa learns is, as suspected, related to the jet
mass and p_T . We can also go through the CoLa and LoLa variables and systematically remove

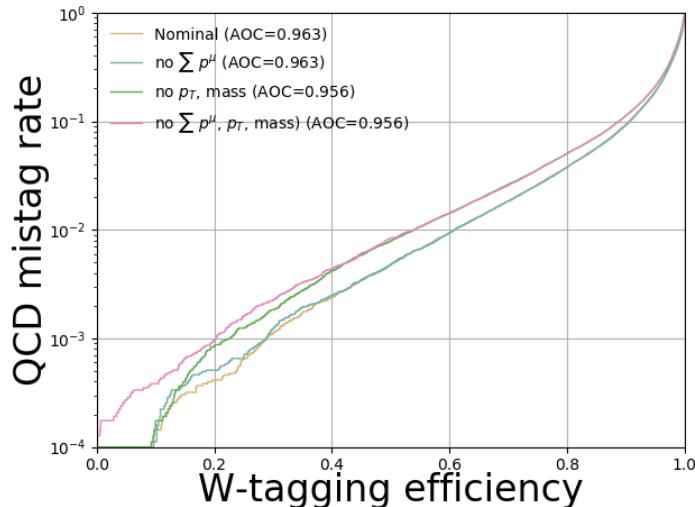


Figure 11.17: Performance after removing jet mass momentum and mass.

³²³¹
³²³² one after the other. This is shown in Figure 11.18. The first entry in the legend correspond to the
³²³³ full nominal training (not p_T -weighted) using all features of CoLa and LoLa (blue). The following
³²³⁴ four entries correspond to removing columns of the CoLa matrix: the first column corresponding
³²³⁵ to the full jet four-vector (orange), removing the identity matrix which passes through each and
³²³⁶ every jet constituent (green), allowing the construction of five rather than 15 pseudo-jets (pink)
³²³⁷ and a combination of all of the above (brown). Removing any of these has very little impact

on the absolute performance as quantified through the area under the curve (AUC) given in parenthesis. This indicates that the CoLa layer has little impact on the absolute performance in terms of tagging efficiency, though it might become important when attempting to train a signal-independent tagger. The next five entries correspond to removing quantities from the LoLa layer: the invariant mass (purple), transverse momentum (gray), energy (red), minimum of all distances (yellow), and the sum of all distances (turquoise). The largest drop in performance is observed when removing the transverse momenta from the training, as was also observed above, whereas removing any other quantity has a minor effect as long as the others are kept. The final legend entry, corresponding to the light green curve, is the performance after removing everything but the sum and minimum of distances between four-vectors. This leads to a large drop in efficiency. Except for the jet p_T , no single variable or CoLa column seems to contain a

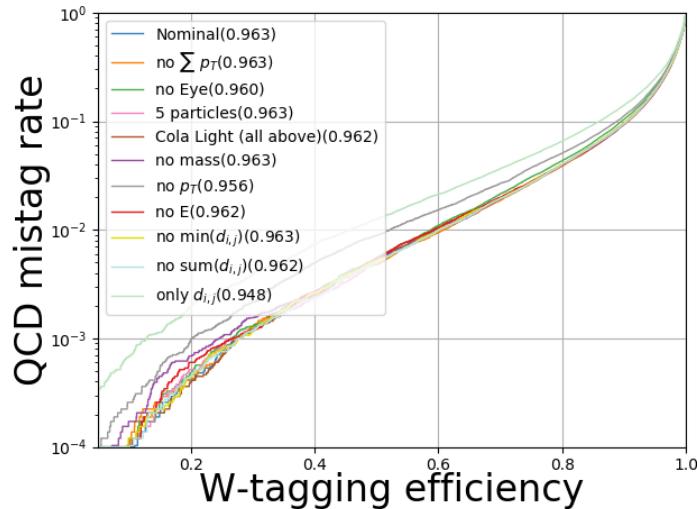


Figure 11.18: Performance after removing components of CoLa or LoLa.

significant amount of information that cannot be replaced by any of the other variables. However, when removing all information except for the distances between four-vectors, a significantly larger performance drop is observed than when removing p_T only. This implies that there is information in LoLa besides the jet p_T that improves the performance, but that this information can be replaced by one of the other variables. Different combinations of LoLa variables therefore need to be studied in the search for a clear performance drop or gain. A complimentary strategy is to look at the LoLa and CoLa weights post-training to check which ones have the largest magnitude, which would in a more direct way quantify the important features of the network. In order to really understand what LoLa learns, further studies are therefore needed and are, as of this writing, currently ongoing.

3259

CHAPTER 12

3260

Summary

3261 In this part of the thesis, we have seen a promising new W-tagging algorithm for future VV
3262 searches. Its absolute performance is better than that of the baseline PUPPI softdrop + τ_{21}
3263 tagger up to a jet p_T of at least 1400 GeV, roughly corresponding to a dijet invariant mass of
3264 2.5-3 TeV, and could lead to an increase in total signal efficiency from 18 to 43 % for the searches
3265 presented here. With a p_T decorrelation method already in place, it could already now be used
3266 for the one dimensional VV search presented in Search I and Search II. However, if to be used in
3267 the multidimensional search framework, a mass decorrelation method needs to be established. I
3268 have already outlined two possibilities of how to achieve this in Section 11.3.2, where one of these
3269 has already been shown to work in the context of p_T -decorrelation. This is, as of this writing, left
3270 to future studies.

3271

3272 When discussing the future of the multidimensional search, I mentioned how a deep neural
3273 network such as the one presented here could be used to encode jet substructure in a way useful in
3274 order to develop a generic anti-QCD tagger. This has already been achieved by a parallel analysis
3275 team through the use of auto-encoders, published ten days before this writing and documented
3276 in [84], and has shown very promising results. However, this strategy is to my knowledge, after
3277 discussing with the authors, no longer pursued after observing that the auto-encoder version of
3278 LoLa was very difficult to decorrelate from the jet mass. It is my belief that this can be overcome
3279 by changing some of the features calculated in the Lorentz Layer (in [84], only the invariant mass
3280 is calculated and the other features listed in Equation 11.4 are removed) and this is an idea I
3281 would like to pursue in future studies in order to achieve a truly generic search for boosted dijet
3282 resonances in the $M_{\text{jet}1}$ - $M_{\text{jet}2}$ - M_{VV} plane.

3283

Part V

3284

Final conclusions and outlook

3285 In this doctoral thesis I have presented three searches for new, heavy resonances decaying to two
 3286 vector bosons in the all-hadronic final state, using datasets collected by the CMS experiment
 3287 corresponding to an integrated luminosity of 2.7 (2015), 35.9 (2016) and 77.3 (2016+2017) fb^{-1} .
 3288 Due to the high energy (“boost”) of the vector bosons, their decay products are so collimated
 3289 that they get merged into one single jet, leading to a dijet final state topology. Dedicated jet
 3290 grooming and jet substructure techniques are therefore explored in order to discriminate vector
 3291 bosons from the overwhelming QCD multijet background.

3292 Each of the analyses presented has provided original contributions to the field: the first search
 3293 was the first of its kind to ever be performed at $\sqrt{s} = 13 \text{ TeV}$, following an observed excess
 3294 of $3.4 (1.3) \sigma$ by ATLAS (CMS) in the 8 TeV dataset, and the first time CMS demonstrated
 3295 the efficiency of using jet-grooming techniques at trigger level. It, at the time, set the most
 3296 stringent limits to date for the signal scenarios under consideration. The second search introduces
 3297 a novel pileup-resistant and perturbative-safe vector boson tagging algorithm based on using
 3298 PUPPI for pileup subtraction, and softdrop for jet grooming, ensuring a high and stable signal
 3299 efficiency up to a pileup of at least 50 interactions per event. The optimization, validation, and
 3300 full commissioning of the tagger was performed in the context of this search. Dedicated jet mass
 3301 corrections, in order to account for an observed p_T and η dependence in PUPPI softdrop jet mass,
 3302 due to the nature of the softdrop algorithm, were also developed. The tagger based on PUPPI
 3303 with softdrop, together with the jet mass corrections, became, and still is, the recommended
 3304 algorithm for W-tagging in CMS.

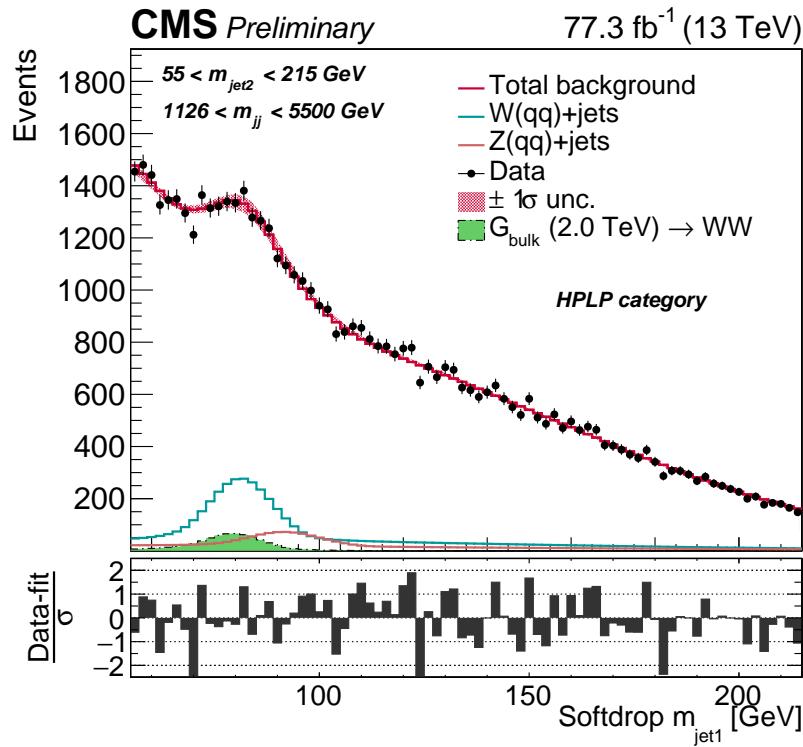
3305 The final analysis introduces a brand new way of doing diboson resonance searches through a
 3306 three-dimensional fit of the dijet invariant mass and the masses of the two jets. By optimizing
 3307 the W-tagging algorithm used in this search and due to the nature of the search method, we have,
 3308 for the first time, been able to constrain the jet mass scale and resolution from a boosted $W(\bar{q}q)$
 3309 and $Z(\bar{q}q)$ peak produced by the V+jets standard model process. The method itself leads to a
 3310 20-30 % higher sensitivity than the default search method.

3311 The benefit of using a three-dimensional fit based on dijet invariant mass and the masses of the
 3312 two jets, is that one can look for resonances peaking anywhere in the jet mass and dijet invariant
 3313 mass spectrum. The natural next step for this search is therefore the incorporation of $VH(bb)$
 3314 and $H(bb)H(bb)$ final states into the three-dimensional fit framework, where orthogonality is
 3315 guaranteed through b-tagging categories. This process is ongoing and planned for the full Run 2
 3316 dataset (including the data collected in 2018).

3317 Going further, one can incorporate searches for generic resonances peaking anywhere in the
 3318 softdrop jet mass and dijet invariant mass spectrum in the multi-dimensional fit, where the jets
 3319 themselves can have substructure compositions other than two subjets. This type of model-

3320 independent search requires a generic anti-QCD tagger in order to be truly model independent.

3321 In the final chapter of this thesis I presented ongoing work on a W-tagging algorithm
 3322 using a deep neural network for future searches, capable of more than doubling the analysis signal
 3323 efficiency by incorporating jet substructure algorithms within the deep layers. As there will be
 3324 no center-of-mass energy increase after the LHC reaches 14 TeV, achieving the best possible
 3325 analysis sensitivity for the dataset to come will be of key importance. I also showed how such a
 3326 deep neural network model is the ideal starting point for building a signal-independent anti-QCD
 3327 tagger. With $\sim 80 \text{ fb}^{-1}$ of 13 TeV data analyzed and no excess observed, the future of this
 3328 search therefore lies in increasing the analysis sensitivity through novel taggers, as the one I have
 3329 presented here, and, making the search strategy as generic as possible through multi-dimensional
 3330 scans and generic anti-QCD taggers, both of which have a foundation now built within this
 3331 doctoral work.



3333

APPENDIX A

3334

V-tagging

3335 A.1 mMDT and underlying event

3336 Signal jets groomed with mMDT using a y_{cut} definition, $\min(p_{T,s_1}^2, p_{T,s_2}^2) \Delta R(s_1, s_2) / m_j^2 > y_{cut}$,
 3337 rather than a z_{cut} definition (as is default in CMS), $\frac{\min(p_{T,i}, p_{T,j})}{p_{T,P}} > z_{cut}$, were found to be far more
 3338 sensitive to the underlying event than pruned jets [73]. Figure A.1 shows the signal efficiency for
 3339 pruning (left) and softdrop (right) as a function of jet transverse momenta when including final
 3340 state radiation (FSR) only, final and initial state radiation (ISR), hadronization, and hadronization
 with an underlying event component added. At parton level, as well as after hadronization, the

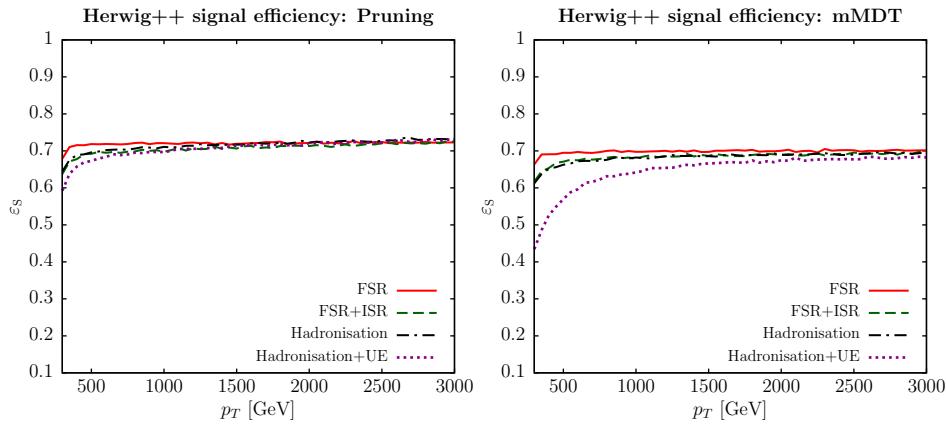


Figure A.1: The signal efficiency for pruning (left) and softdrop (mMDT) (right) as a function of jet p_T when adding FSR, ISR, hadronization and UE. The UE has a severe impact on the softdrop efficiency for signal jets [73].

3341
 3342 two algorithms perform very similar as a function of p_T . However, once UE contamination is
 3343 added, the softdrop tagging efficiency is severely affected. This can be explained by the larger
 3344 effective radius considered by the mMDT algorithm ($\propto m_V/p_T \sqrt{z_{cut}(1 - z_{cut})}$) in comparison
 3345 to pruning ($\propto m_V/p_T$), as can be seen from Equations 6.9 and 6.8. This observation corresponds
 3346 very well with the shift in jet mass we observed for softdrop as a function of p_T in Section 7.4.3:
 3347 as the jet p_T decreases, the mMDT effective radius increases and the jet mass mean shifts to
 3348 higher values, due to absorbing more background radiation.

3349 A.2 N-subjettiness on groomed jets

3350 In CMS, n-subjettiness is by default calculated on un-groomed jets. A proposal was made in
 3351 Ref. [91] to, when calculating n-subjettiness, calculate τ_2 on softdrop $\beta = 2$ jets (less aggressive
 3352 groomer) τ_1 on softdrop $\beta = 0$ jets (more aggressive groomer). This would emphasize the jet
 3353 hard substructure when calculating 1-subjettiness and the full color radiation pattern when
 3354 calculating 2-subjettiness. In the above mentioned reference, this was shown to increase the signal
 3355 significance by 25% (at 2 TeV) for a given mistagging rate, as shown in Figure A.5 and also
 reduce non-perturbative effects by a factor of 2 to 3. In this section, we study the performance

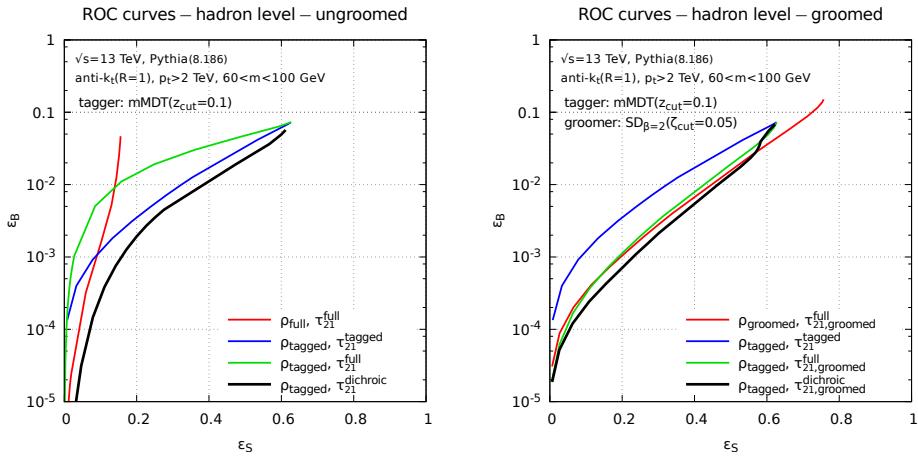


Figure A.2: Using a dichroic tagger has been to increase the signal significance by 25% at a jet p_T of 2 TeV and reduce non-perturbative effects by a factor of 2 to 3 [91].

3356
 3357 of n-subjettiness when used in combination with different versions of softdrop, varying the
 3358 β -parameter. Different values of β results in the algorithm removing more or less radiation. In
 3359 this section we study three groomer definitions: mMDT (softdrop with $\beta = 0$), softdrop with
 3360 $\beta = 2$ and the un-groomed (plain) mass. The mass distribution after applying these algorithms
 3361 is shown in Figure A.3. The current CMS default for V-tagging at the time of this study was
 3362 softdrop $\beta = 0$ and τ_{21} , where τ_{21} was calculated on the un-groomed jet. N-subjettiness can
 3363 be calculated starting from un-groomed jets or jets groomed with different algorithms. Going
 3364 from the most to the least aggressive, the ones studies here are τ_{21} on softdrop $\beta = 0$ jets, τ_{21}
 3365 on softdrop $\beta = 2$ jets, dichroic τ_{21} (τ_2 from softdrop $\beta = 2$ and τ_1 from softdrop $\beta = 0$ jets)
 3366 or τ_{21} calculated on un-groomed jets. The distributions of these are shown in Figure A.4. The
 3367 performance in the signal efficiency versus mistagging rate plane for different versions of τ_{21} , is
 3368 shown in Figure A.5. A cut on the softdrop $\beta = 0$ jet-mass of 65-105 GeV is applied by default

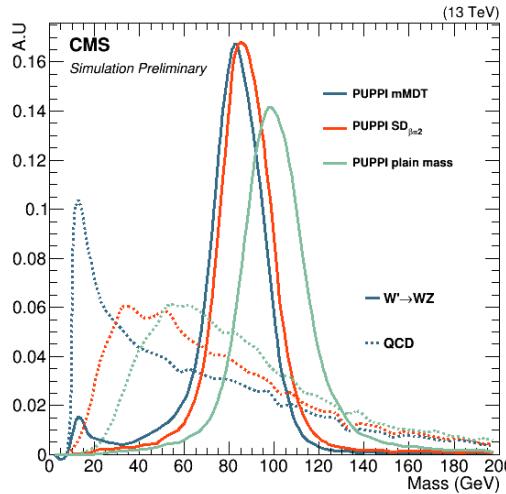


Figure A.3: The bare PUPPI (turquoise), softdrop $\beta = 0$ (blue), and softdrop $\beta = 2$ (red) jet-mass.

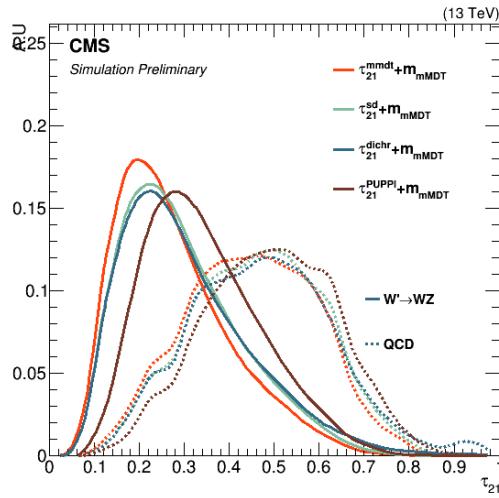


Figure A.4: N-subjettiness distributions after the algorithm has been applied to un-groomed jets (brown, CMS default), or on jets groomed with softdrop $\beta = 0$ (red), softdrop $\beta = 2$ (turquoise) or a mix of both, so-called dichroic (blue), where softdrop $\beta = 0$ is used for τ_1 and softdrop $\beta = 2$ is used for τ_2 .

for all taggers. The CMS default tagger, τ_{21} in combination with softdrop $\beta = 0$ has the lowest signal efficiency for a given mistagging rate. The best performance is achieved when calculating n-subjettiness on softdrop $\beta = 0$ groomed jets. No huge gain in performance is observed when using a dichroic tagger.

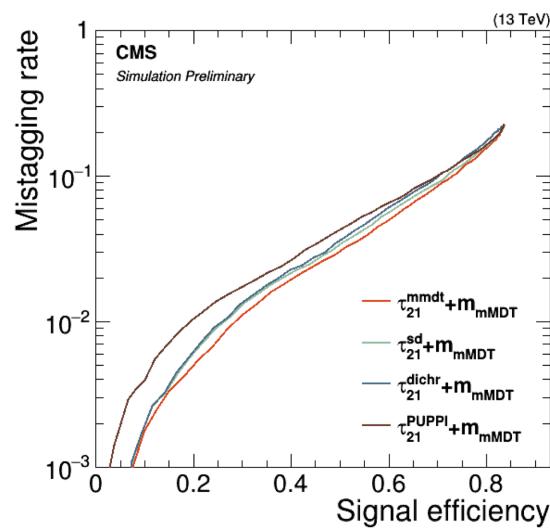


Figure A.5: Performance of a tagger consisting of a selection on the softdrop $\beta = 0$ jet mass of 65-105 GeV and a cut on n-subjettiness computed on groomed or un-groomed jets

3373

APPENDIX B

3374

Search I: Limits per mass category

3375 B.1 Limits per mass category

3376 The asymptotic limits obtained with 2.6 fb^{-1} of 13 TeV CMS data per mass and purity category.

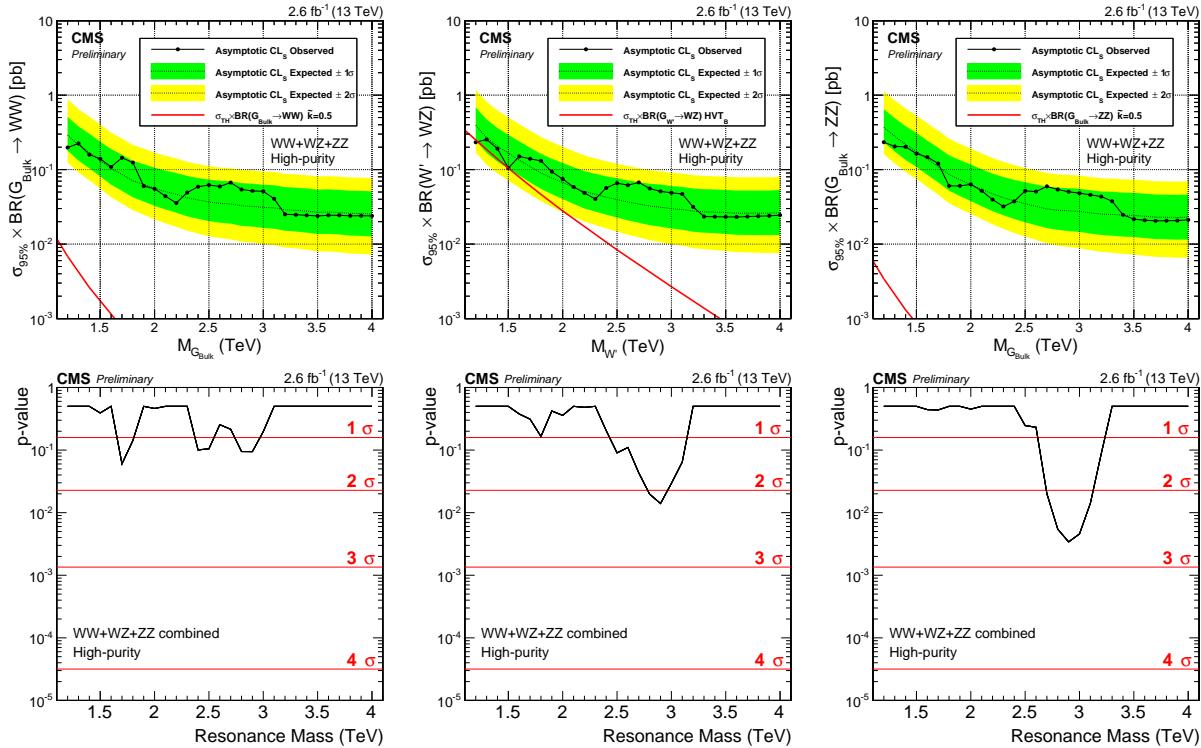


Figure B.1: Expected and observed limits at 95% CL and corresponding p-values obtained in the high purity category using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle) and $G \rightarrow ZZ$ (right) signal.

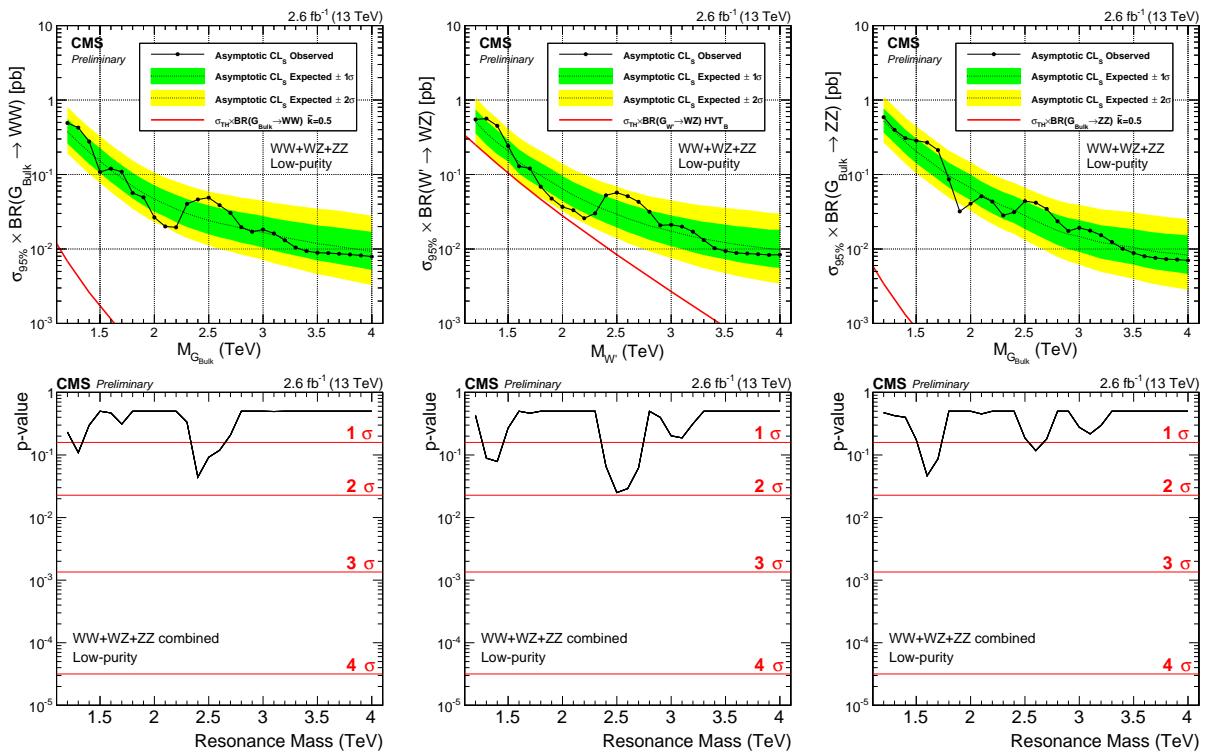


Figure B.2: Expected and observed limits at 95% CL and corresponding p-values obtained in the low purity category using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ (left), $W' \rightarrow WZ$ (middle) and $G \rightarrow ZZ$ (right) signal.

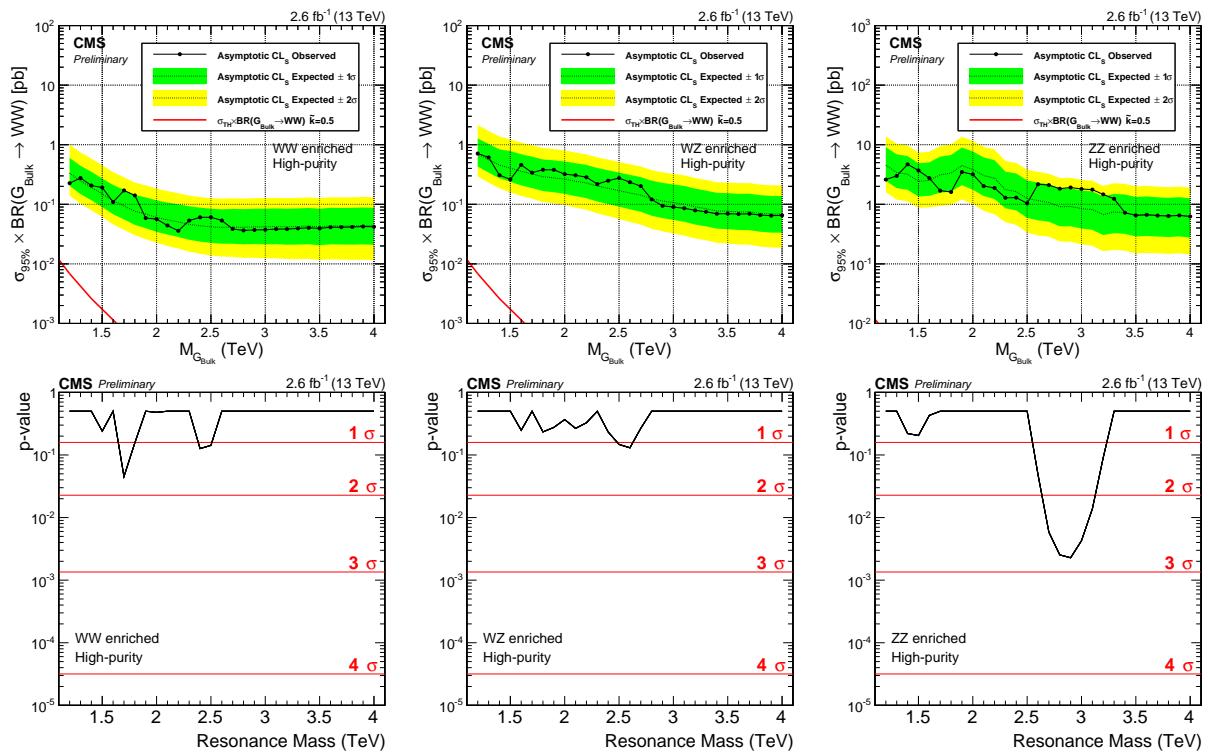


Figure B.3: Expected and observed limits at 95% CL and corresponding p-values obtained for the different mass categories using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ signal in the HP category

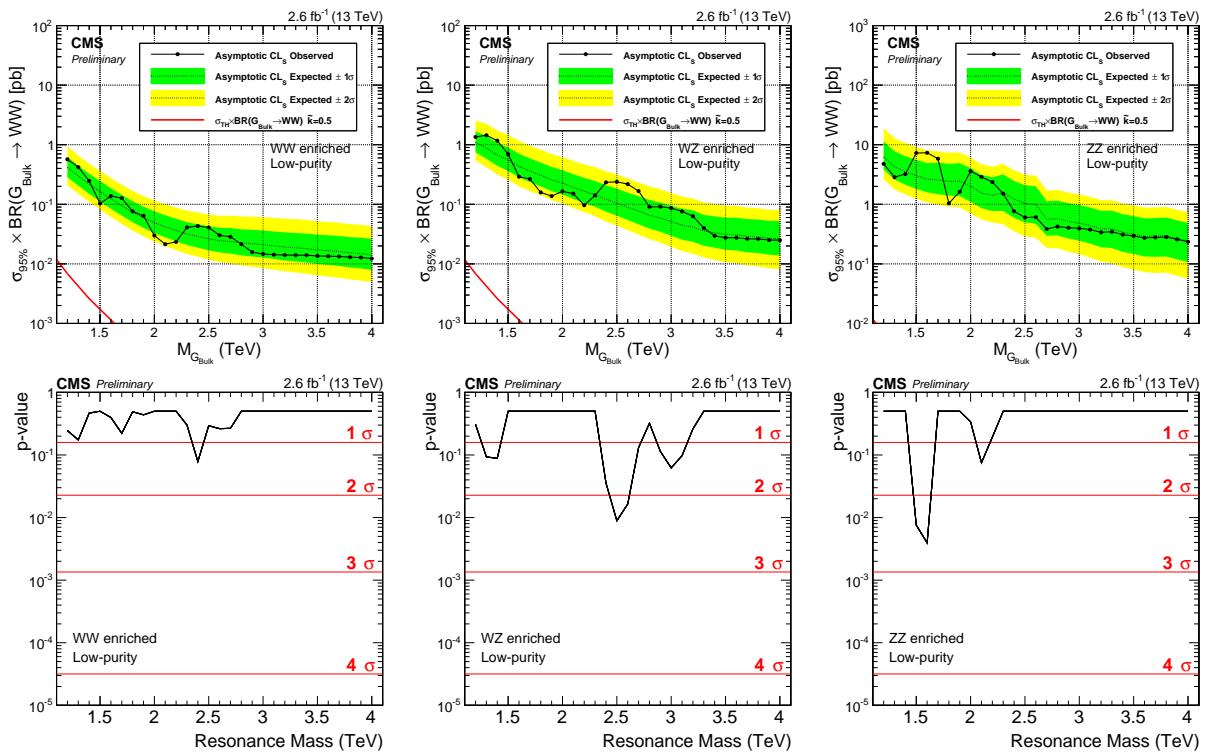


Figure B.4: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories using 2.6 fb^{-1} of CMS data. Here for a Bulk $G \rightarrow WW$ signal in the LP category.

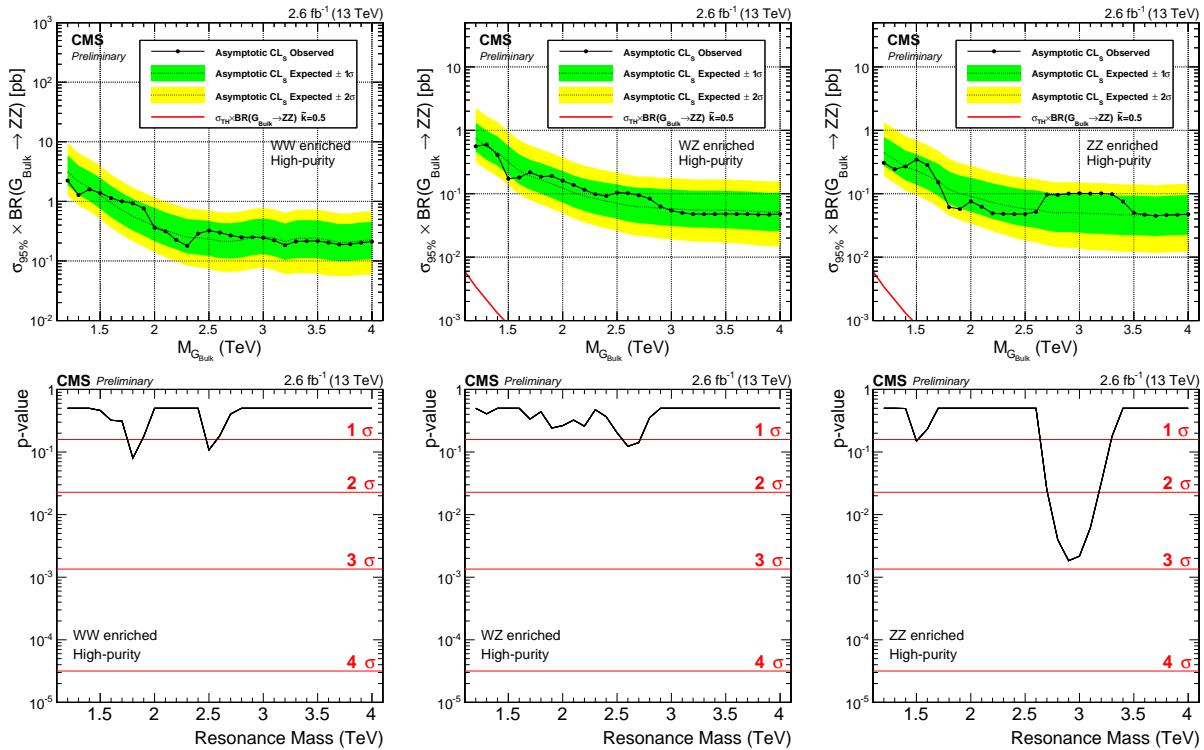


Figure B.5: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories. Here for a $G \rightarrow ZZ$ signal in the HP category.

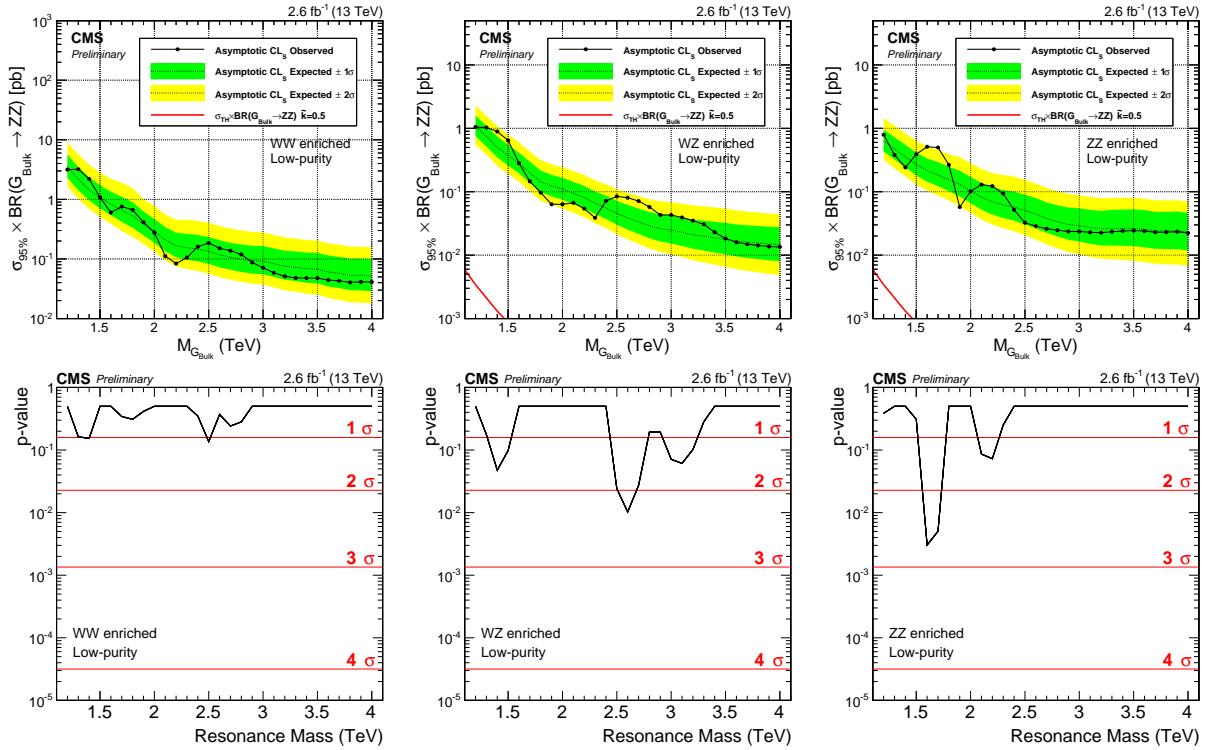


Figure B.6: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories. Here for a $G \rightarrow ZZ$ signal in the LP category.

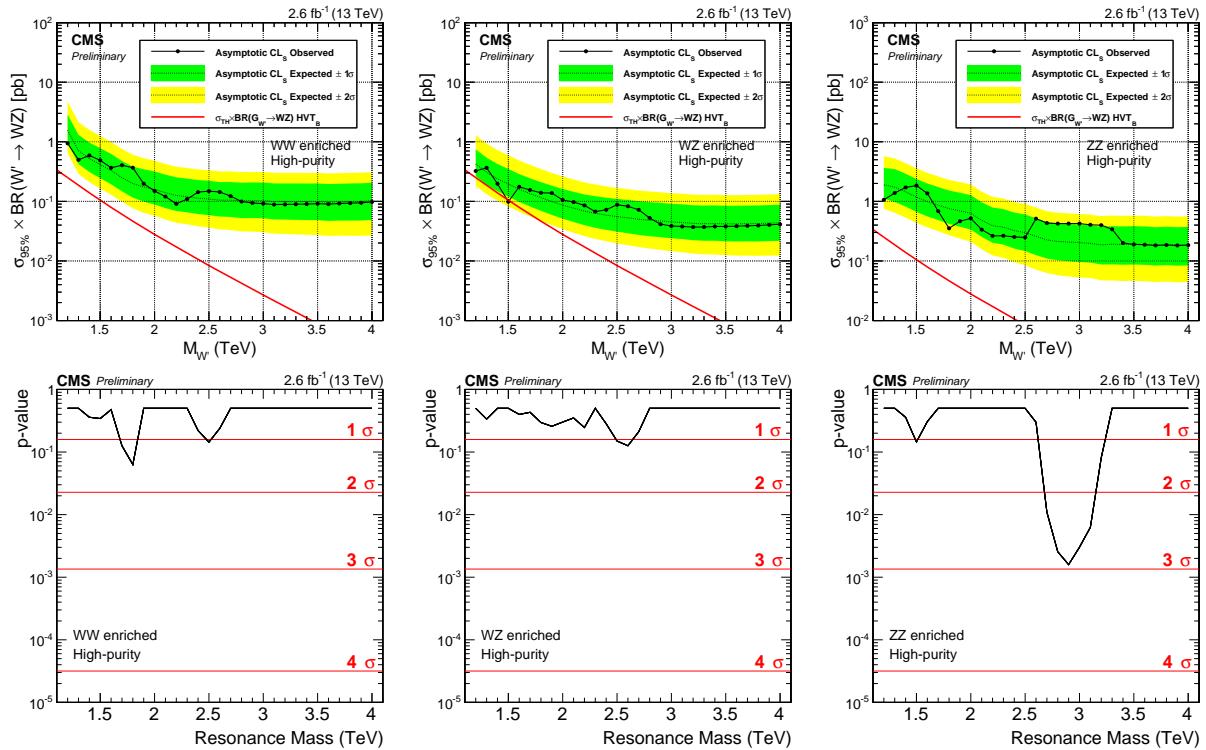


Figure B.7: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories. Here for a $W' \rightarrow WZ$ signal in the high-purity category.

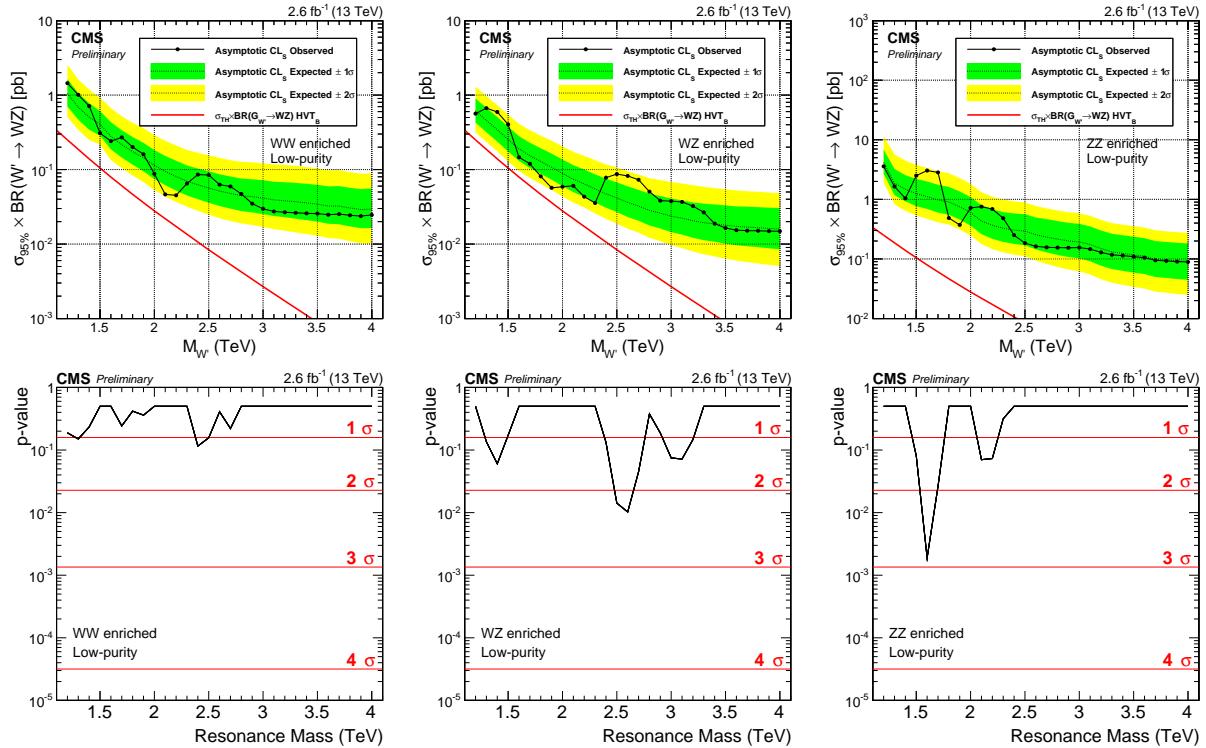


Figure B.8: Expected and observed limits at 95% CL and corresponding p-values obtained in the different mass categories. Here for a $W' \rightarrow WZ$ signal in the low purity category.

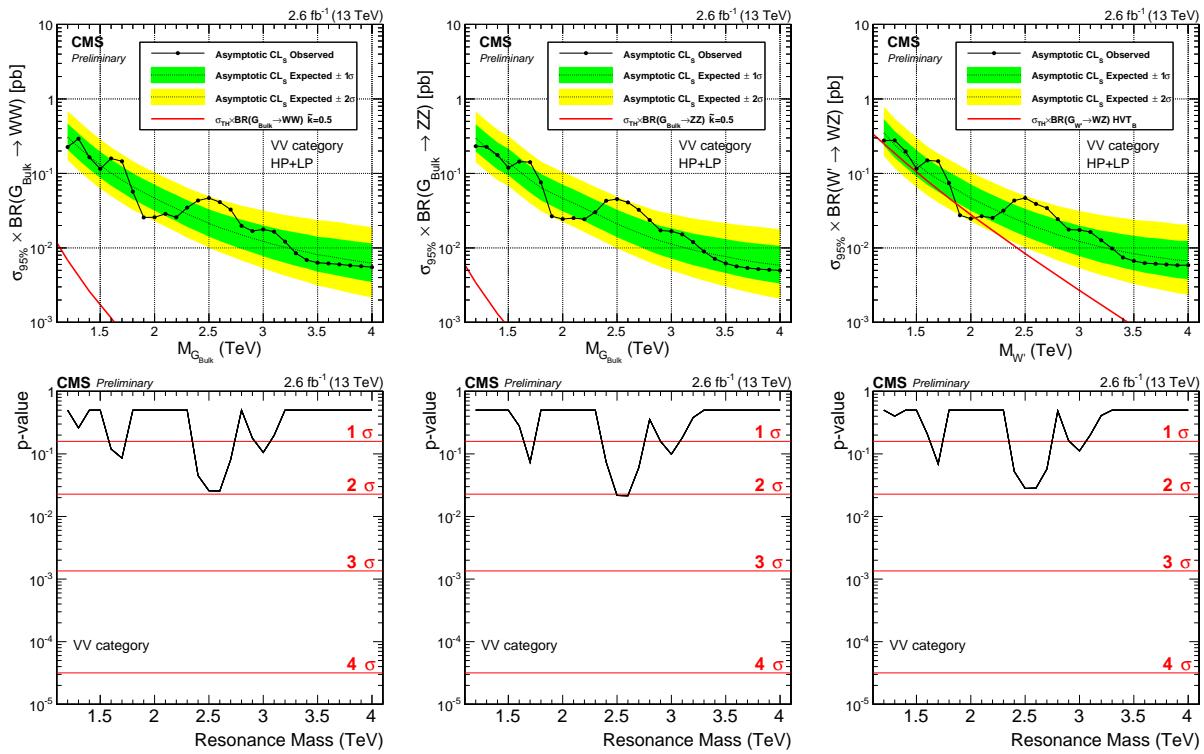


Figure B.9: Expected and observed limits at 95% CL and corresponding obtained without splitting into mass categories. This analysis is performed as a cross check analysis and directly compares with the method used in the corresponding Run 1 analysis. Here for a Bulk $G \rightarrow WW$ (left), $G \rightarrow ZZ$ (middle) and $W' \rightarrow WZ$ signal (right).

3378

APPENDIX C

3379

Search II

3380 C.1 W-tagging scale factor: Additional plots

3381 Figures C.1 shows the $t\bar{t}$ real W (top) and non-W (bottom) PUPPI softdrop jet mass distributions
 3382 for jets that passed (left) and failed (right column) the N-subjettiness selections $\text{PUPPI } \tau_{21} < 0.40$.
 Figures C.2 shows the fitted PUPPI softdrop jet mass distributions for the non-dominant

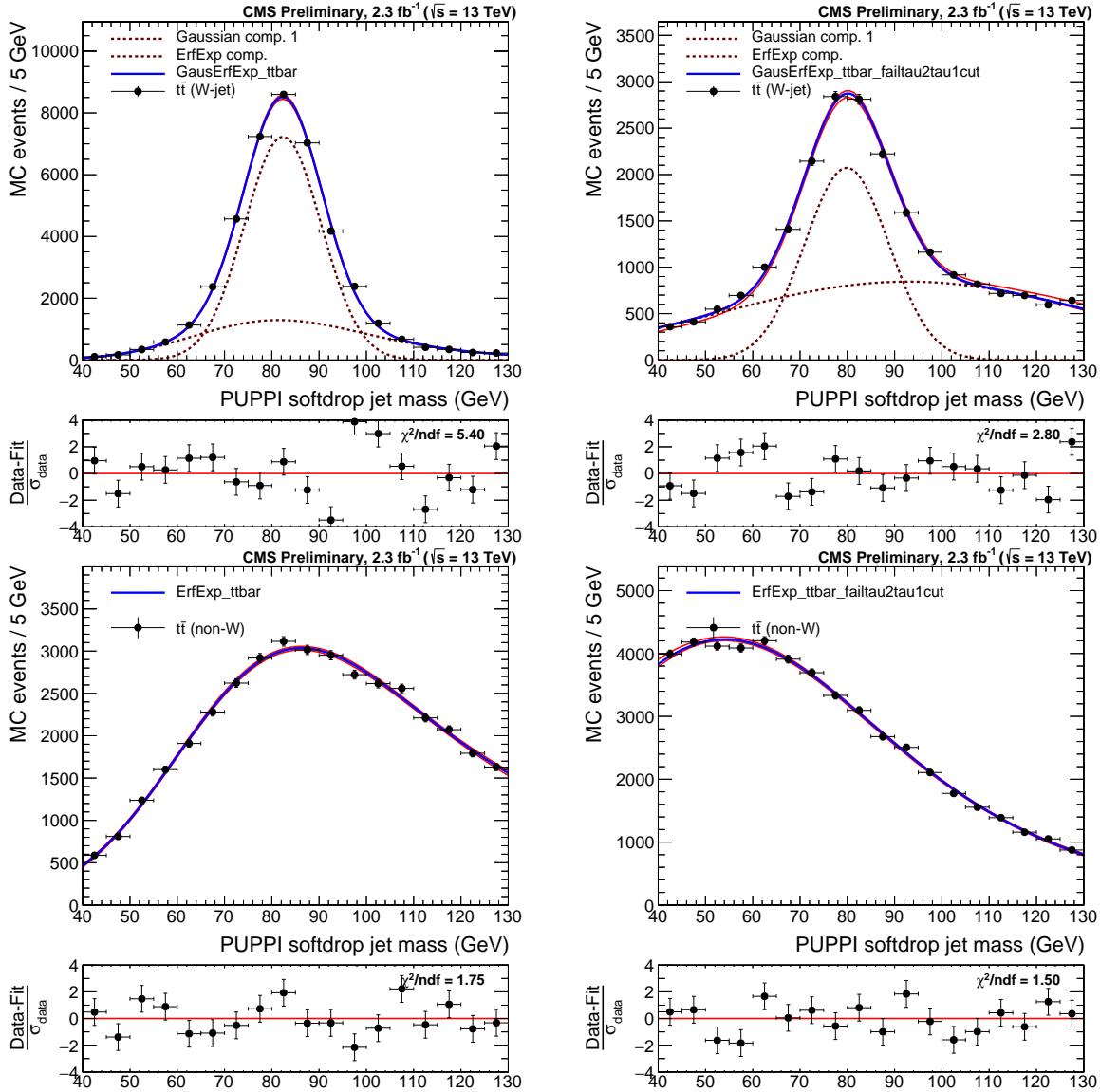


Figure C.1: Fit to the real W (top) and non-W (bottom) softdrop jet mass distribution for jets that pass (left) and fail (right) the cut on $\text{PUPPI } \tau_{21} < 0.4$.

3383 backgrounds in the evaluation of the W-tagging scale factors. Here it is shown for jets that pass
 3384

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

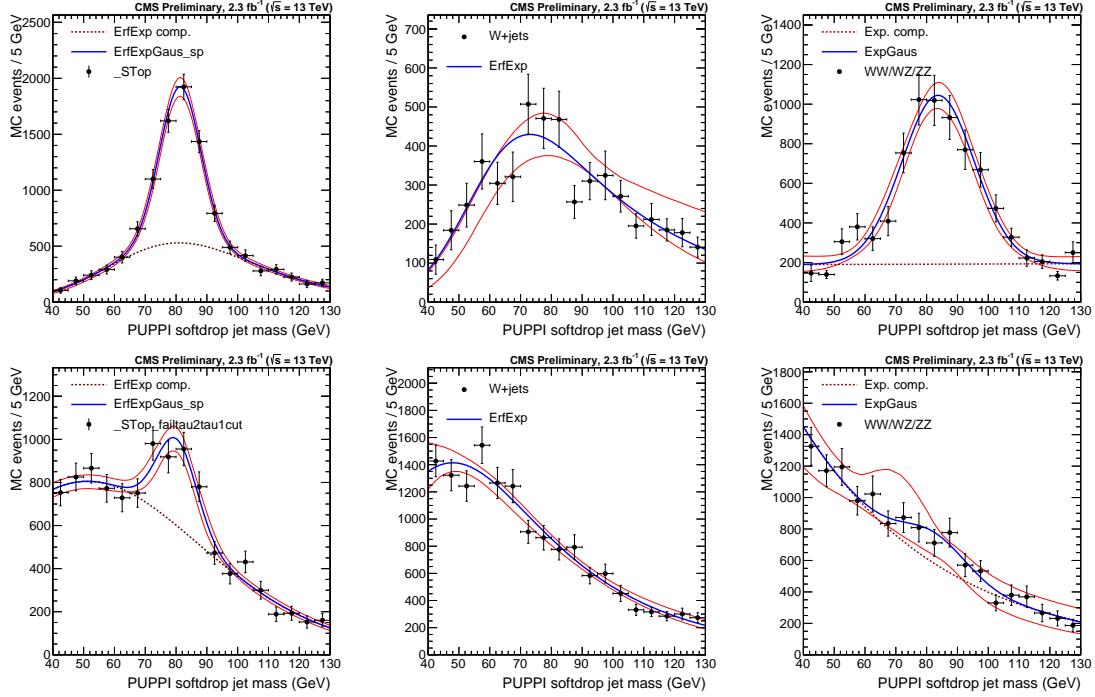


Figure C.2: Fits to the PUPPI softdrop jet mass spectrum for the non-dominant backgrounds (Single top, W+jets and VV respectively) in the pass (top) and fail (bottom) regions.

3386 **C.2 Efficiency scale factors for 2.5 fb^{-1}**

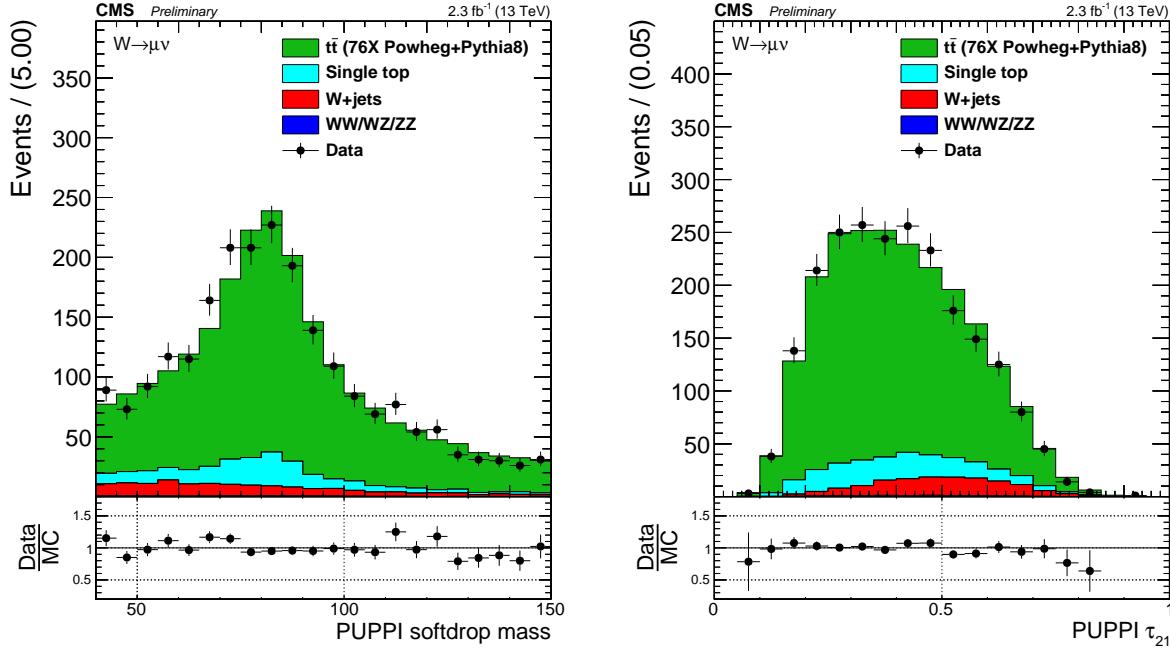


Figure C.3: Distribution of the PUPPI softdrop mass (left) and PUPPI n-subjettiness (right) distribution in the $t\bar{t}$ control sample.

Category	Working point	Eff. data	Eff. simulation	Scale factor
HP	$\tau_2/\tau_1 < 0.4$	0.785 ± 0.045	0.81 ± 0.01	$0.97 \pm 0.06 \text{ (stat)} \pm 0.04 \text{ (sys)} \pm 0.06 \text{ (sys)}$
LP	$0.4 < \tau_2/\tau_1 < 0.75$	0.215 ± 0.057	0.204 ± 0.041	$1.13 \pm 0.24 \text{ (stat)} \pm 0.17 \text{ (sys)} \pm 0.12 \text{ (sys)}$

Table C.1: W-tagging scale factors for high purity and low purity categories for a tagger based on PUPPI softdrop jet mass and τ_{21} .

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

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

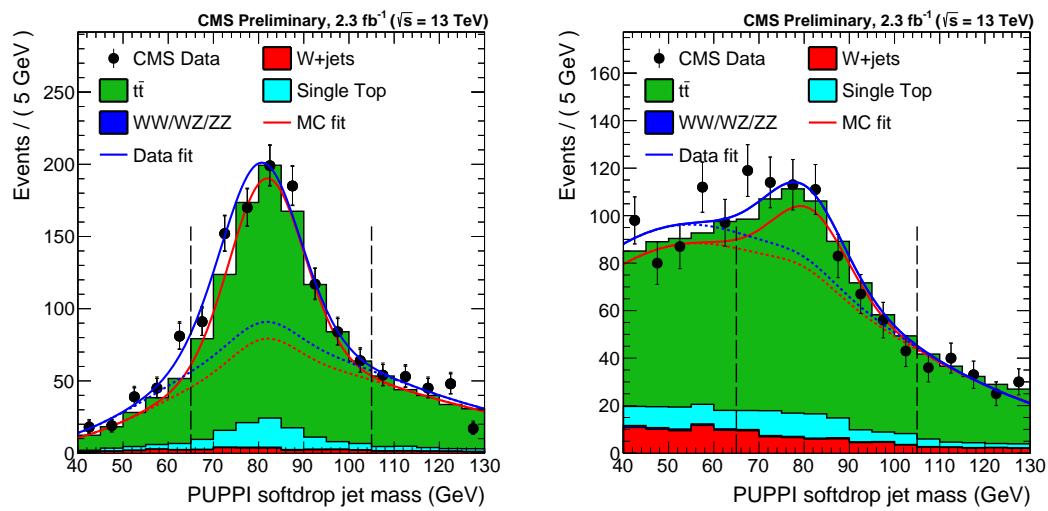


Figure C.4: PUPPI softdrop jet mass distribution for events that pass (left) and fail (right) the PUPPI $\tau_2/\tau_1 < 0.40$ selection. Results of both the fit to data (blue) and simulation (red) are shown and the background components of the fit are shown as short-dashed lines.

³³⁸⁷ C.3 Background fit checks

³³⁸⁸ The background from QCD multijet events is modeled by a smoothly falling distribution in each
³³⁸⁹ analysis category. The method consists of a smoothness test of the observed data where the
³³⁹⁰ background is assumed to be described by the following empirical probability density function:

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

³³⁹¹ where m is the dijet invariant mass, \sqrt{s} the centre-of-mass energy, P_0 is a normalization parameter
³³⁹² for the probability density function, and P_1 and P_2 describe the shape. To ensure that this
³³⁹³ function is sufficient to describe the data in all the different analysis categories, we first perform
³³⁹⁴ a test to check that no additional parameters are needed and to check the systematic uncertainty
³³⁹⁵ due to choice of fit function. For these studies we use a data sideband, where one of the two jets
³³⁹⁶ is required to have a mass between $20 < M_{\text{Softdrop}} < 65$ GeV. In order to quantify how many
³³⁹⁷ parameters are necessary, a Fishers F-test is performed for the fits to data in the data sideband.
³³⁹⁸ The critical value that the test statistic must exceed is chosen to be $\alpha > 10\%$. If the returned
³³⁹⁹ Confidence Level is larger than α , the simpler fit is preferred. The three parameter fit is compared
³⁴⁰⁰ with the following 2, 4 and 5 parameter functions:

$$\frac{dN}{dm} = \frac{P_0}{(m/\sqrt{s})^{P_2}}, \quad (\text{C.2})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2 + P_3 \times \log(m/\sqrt{s})}}, \text{ and} \quad (\text{C.3})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2 + P_3 \times \log(m/\sqrt{s}) + P_4 \times \log(m/\sqrt{s})^2}}. \quad (\text{C.4})$$

³⁴⁰¹ Additionally, fits with an alternative fit function have also been performed (for the single-tag
³⁴⁰² categories we try both 4 and 5 parameter versions):

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s} + P_3(m/\sqrt{s})^2)^{P_1}}{(m/\sqrt{s})^{P_2}}, \text{ and} \quad (\text{C.5})$$

$$\frac{dN}{dm} = \frac{P_0(1 - m/\sqrt{s} + P_3(m/\sqrt{s})^2)^{P_1}}{(m/\sqrt{s})^{P_2 + P_4 \times \log(m/\sqrt{s})}} \quad (\text{C.6})$$

3403 C.3.1 Background fit checks in data sideband

3404 We perform a test in a data sideband to make sure the fit functions work on real data and to
 3405 exercise the estimation of number of necessary fit parameters via an F-test. The sideband is
 3406 constructed by requiring one of the two jets two have a mass in the low softdrop jet mass sideband,
 3407 between $20 \text{ GeV} < M_{\text{Softdrop}} < 65 \text{ GeV}$, while the full W/Z-tag selections are applied to the other
 3408 jet. The low-mass jet is also required to pass the τ_{21} selection corresponding to the given category.
 3409 For the single-tag category, the sideband is constructed by requiring one of the two jets to have a
 3410 mass in the low softdrop jet mass sideband, between $20 \text{ GeV} < M_{\text{Softdrop}} < 65 \text{ GeV}$ and the other
 3411 in a high-mass sideband, between $105 < M_{\text{Softdrop}} < 200 \text{ GeV}$. One of the jets is also required to
 3412 pass the τ_{21} selection corresponding to the given category. We first check whether the sideband
 3413 can be used to exercise the F-test by checking whether or not there are features introduced in the
 3414 dijet mass spectrum that may be hard to cover with the fit using QCD MC. To do so we look at
 3415 the dijet invariant mass spectrum in the signal region divided by the distribution in the sideband.
 3416 The obtained distributions are shown in Figure C.5 for three different generators, where the pure
 3417 Pythia8 QCD samples (top left) have the highest number of events. The distributions are mostly
 3418 smooth, but we do see features introduced in the “WW HP” and “ZZ HP” categories which might
 3419 prove difficult to fit, as well as in the tail of the single-tag categories. These kinks shift around
 3420 depending on what MC generator is used and do not seem to be a systematic feature caused
 3421 by the selection requirements, but rather due to limited statistics. We proceed with exercising
 3422 the F-test in a data sideband, with the caveat that there might be features introduced in the
 3423 spectrum where statistics are low.

3424 Figure C.6 shows the fit to data in the data sideband for the WW and ZZ mass categories,
 3425 both in the HP and in the LP n-subjettiness categories. The corresponding residuals, χ^2 and
 3426 F-test results are shown in Table C.7 through C.10. For the double-tag categories, a two or three
 3427 parameter function is sufficient to describe the data and we conclude that the function as defined
 3428 in C.1 is sufficient for all mass categories. For the single-tag category a five parameter fit seems
 3429 to be required in order to describe the data and the fit quality is not optimal. To ensure that the
 3430 fit functions with sufficient number of parameters are able to describe the shape in the single-tag
 3431 categories, we have additionally looked at the fit quality in QCD MC (see below). Here, the
 3432 default dijet function seems sufficient to describe the distributions. The sideband in the single-tag
 3433 categories in QCD MC do not show the same features as the data sideband as shown in Figure
 3434 C.7.

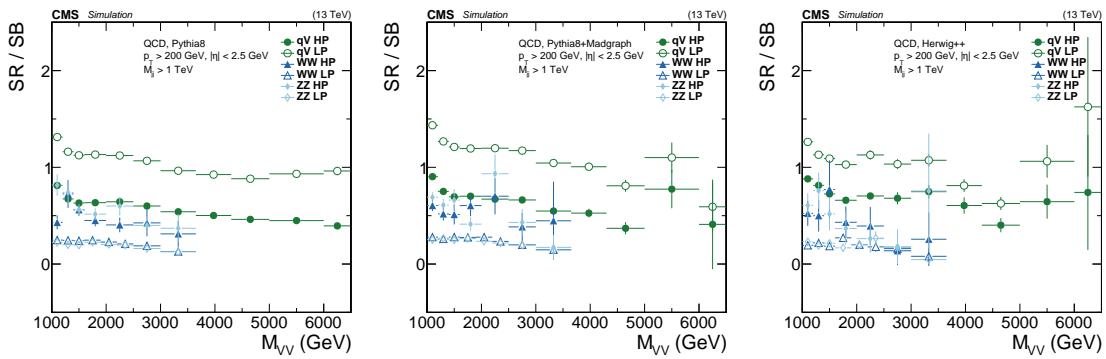


Figure C.5: Dijet mass spectrum in the signal region divided by the dijet mass spectrum in the sidebands using QCD Pythia8 (left), QCD Pythia8+Madgraph (middle) and QCD Herwig++ (right) simulated samples. Here for the double W/Z-tag and the single V-tagged HP and LP categories. Some jumps are observed in the high-mass tail of the dijet invariant mass distribution in the high-purity WW/ZZ categories, but otherwise no strange features seem to be induced by using the sideband.

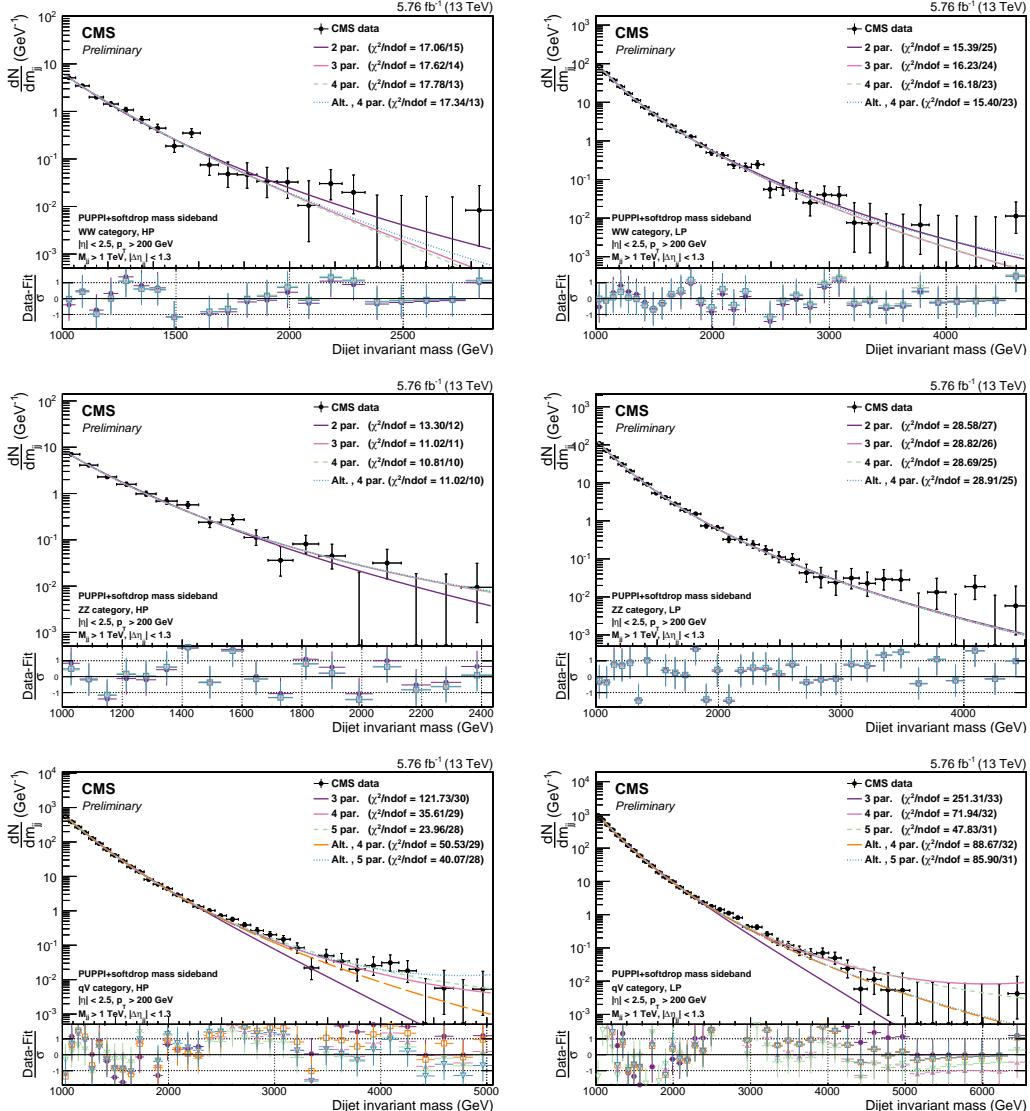


Figure C.6: Fitted dijet mass spectrum in the different mass and purity categories in a data sideband: WW high-purity (top left) and low-purity (top right), ZZ high-purity (middle left) and low-purity (middle right), qV high-purity (bottom left) and low-purity (bottom right).

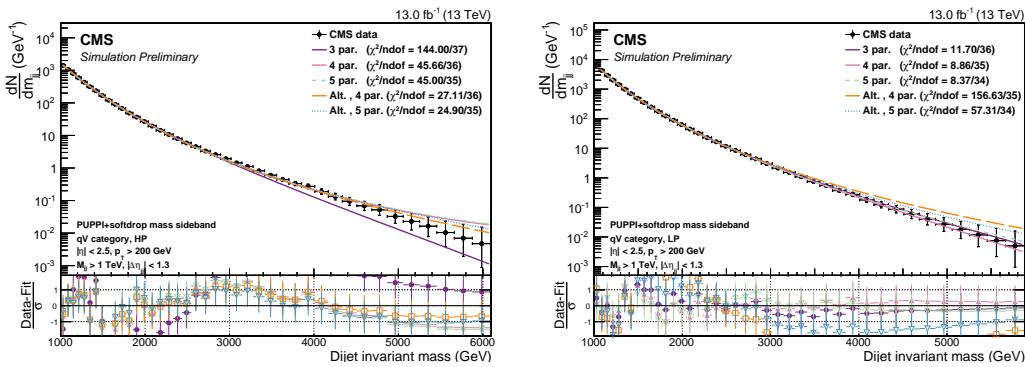


Figure C.7: Fitted dijet mass spectrum in the QCD MC sideband: qV high-purity (left) and low-purity (right).

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 C.3: Residuals, χ^2 , and degrees of freedom for the WW category, HP category. A 2-parameter fit is needed to describe these data.

WW category, LP			
Function	Residuals	χ^2	ndof
2 par	0.908	15.388	25
3 par	0.279	16.225	24
4 par	0.263	16.178	23
Fishers23	56.395	CL	0.000
Fishers34	1.406	CL	0.247

Table C.4: Residuals, χ^2 , and degrees of freedom for the WW category, LP category. A 3-parameter fit is needed to describe these data.

ZZ category, HP			
Function	Residuals	χ^2	ndof
2 par	0.215	13.296	12
3 par	0.133	11.022	11
4 par	0.119	10.810	10
Fishers23	7.465	CL	0.018
Fishers34	1.304	CL	0.278

Table C.5: Residuals, χ^2 , and degrees of freedom for the ZZ category, HP category. A 3-parameter fit is needed to describe these data.

ZZ category, LP			
Function	Residuals	χ^2	ndof
2 par	2.459	28.583	27
3 par	2.363	28.817	26
4 par	2.175	28.694	25
Fishers23	1.107	CL	0.302
Fishers34	2.244	CL	0.146

Table C.6: Residuals, χ^2 , and degrees of freedom for the ZZ category, LP category. A 2-parameter fit is needed to describe these data.

qV category, HP			
Function	Residuals	χ^2	ndof
3 par	128.276	121.731	30
4 par	29.113	35.606	29
5 par	7.036	23.962	28
Alt. 4 par	37.232	50.528	29
Alt. 5 par	30.948	40.068	28
Fishers34	102.185	CL	0.000
Fishers45	90.988	CL	0.000
FishersAlt4Alt5	5.888	CL	0.022

Table C.7: Residuals, χ^2 , and degrees of freedom for the qV category, HP category. A 5-parameter fit is needed to describe these data.

qV category, LP			
Function	Residuals	χ^2	ndof
3 par	671.341	251.311	33
4 par	171.593	71.942	32
5 par	80.801	47.830	31
Alt. 4 par	215.431	88.666	32
Alt. 5 par	214.766	85.896	31
Fishers34	96.109	CL	0.000
Fishers45	35.957	CL	0.000
FishersAlt4Alt5	0.099	CL	0.755

Table C.8: Residuals, χ^2 , and degrees of freedom for the qV category, LP category. A 5-parameter fit is needed to describe these data.

3435 C.3.2 Background fit checks in QCD MC

3436 As an additional check, we look at the fit functions in the different signal categories using QCD
3437 MC. This is shown in Figure C.8 for the double and Figure C.9 for the single tag categories. Here
3438 all fit functions and their pull distributions are plotted. The fits are performed in a mass range
3439 corresponding to the expected distribution in the different categories for 13 fb^{-1} of data. We
3440 have adapted the error bars to correspond to the maximum of the expected Poisson error for 13 fb^{-1}
3441 of data and the pure simulation error (accounting for the different weights assigned to the
3442 p_T -binned QCD MC sample). The reason for this choice is to get an estimate of whether the
3443 set of fit functions we plan to use to fit the background distribution in data, and plan to use in
3444 order to understand the systematic uncertainty due to our choice of fit function, are appropriate
3445 and do not produce fake bumps/kinks. As this distribution is the pure MC simulation curve,
3446 whose variation at high masses is much smaller than the expected poisson error for 13 fb^{-1} of
3447 data, we expect the χ^2/ndof to be lower than one. In order to protect against the fact that
3448 the MC simulation at lower dijet masses does not have more statistics than the expected data
3449 for 13 fb^{-1} of data, we use the largest of Poisson and the MC errors. The resulting errors are
3450 therefore a mixture of Poisson and the MC errors and the χ^2/ndof for the QCD MC fits should
3451 not be considered. Fit quality in the form of χ^2/ndof should only be estimated from the fits to
3452 data sideband where pure Poisson errors are used. Overall the fits to QCD MC in the different
3453 categories describe the data well, with a two or three parameter function sufficient to describe the
3454 distributions. However, due to an under fluctuation of the first bin in the “ZZHP” category, the
3455 higher parameter fits are steered by the first bin leading to discrepancies in the tail. This is the
3456 lowest statistics category and the danger for under-fluctuations does exist. We have investigated
3457 the dijet invariant mass distribution down to a dijet invariant mass threshold of 800 GeV to
3458 make sure we are not seeing a turn-on effect. This is shown in Figure C.10. The fits in the data
3459 sideband for the same category (Figure C.6, middle left) do not show the same trend.

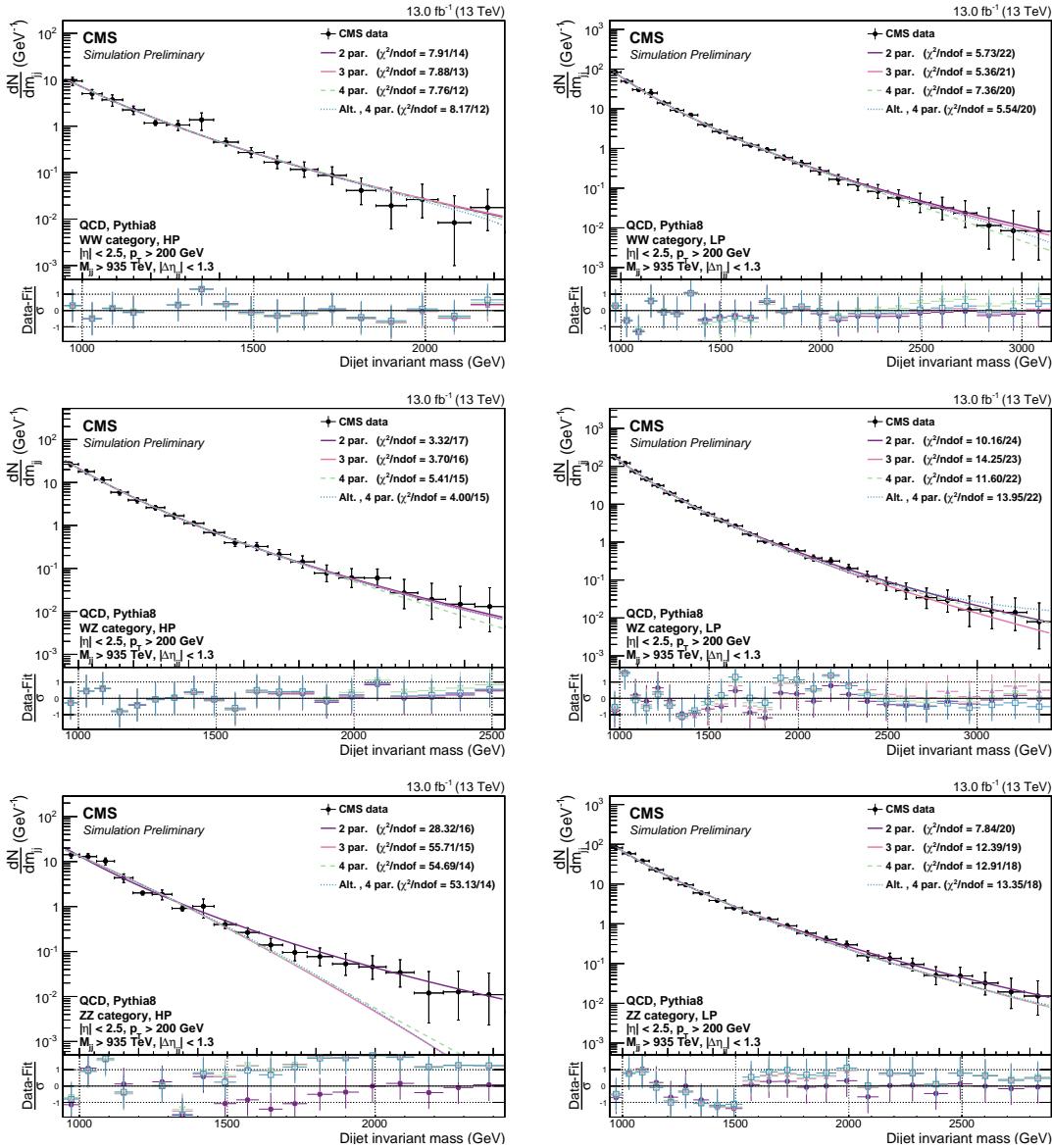


Figure C.8: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Shown here for the high- and low-purity double W/Z-tag category for the three different mass categories: WW category (top), WZ category (middle) and ZZ category (bottom).

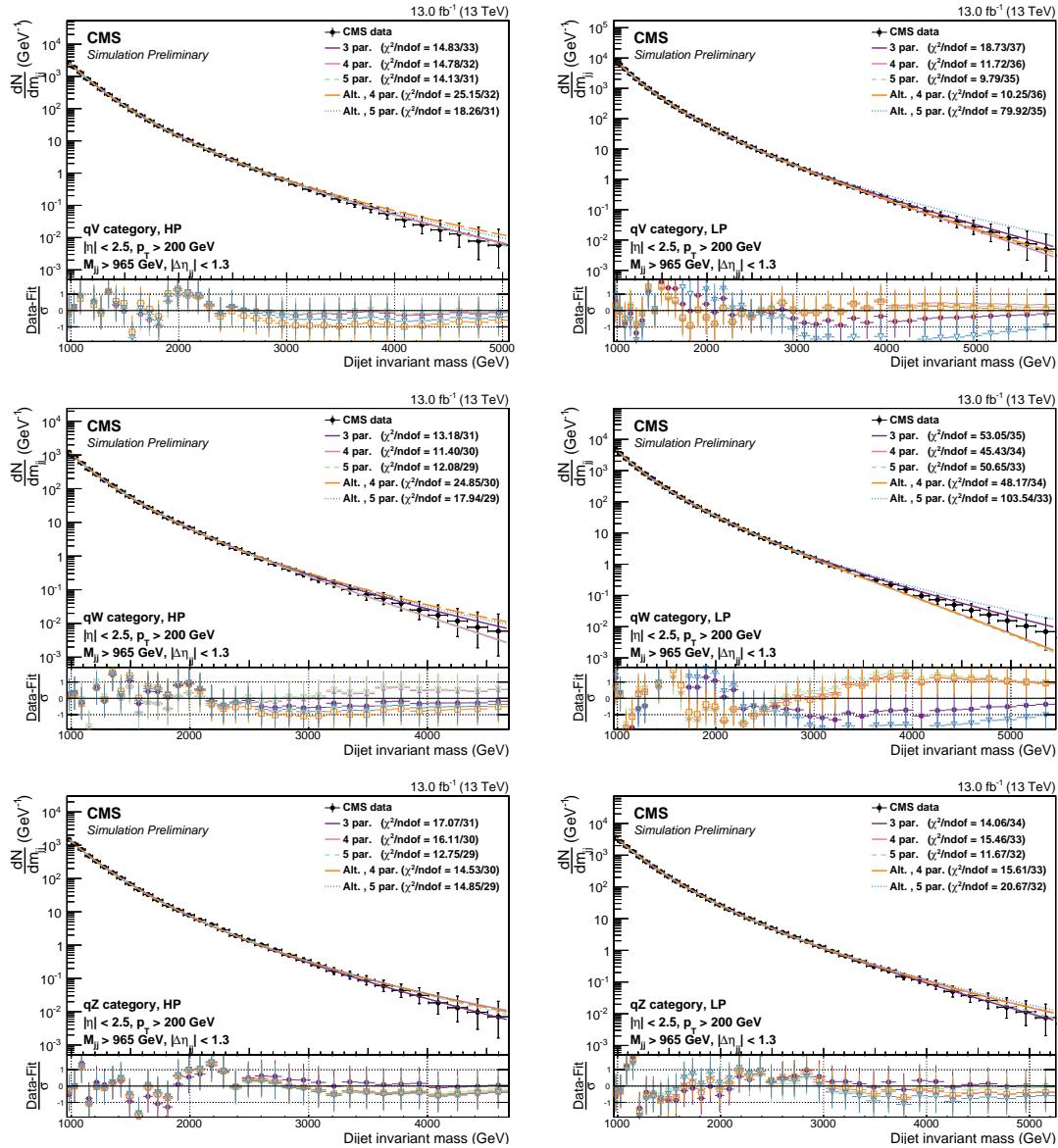


Figure C.9: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Shown here for the high- and low-purity single W/Z-tag category for the two different mass categories: qW category (top) and qZ category (bottom).

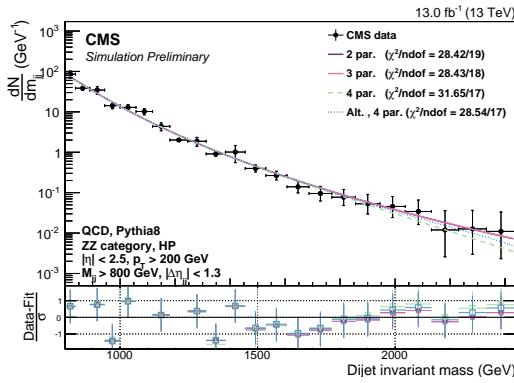


Figure C.10: Background fit for the M_{jj} distribution in QCD MC corresponding to 13 fb^{-1} . Shown here for the high-purity double Z-tag category using a dijet invariant mass threshold of 800 GeV. No turn-on effect at low invariant masses is observed.

3460 C.4 F-test in signal region

3461 The final F-test performed in order to define the number of fit parameters to be used to fit the
 3462 background in each analysis category is performed in the signal region. The resulting fits and
 3463 F-test values for the double tag categories are shown in Figure C.11 and Tables C.9 to C.14. The
 3464 F-test results for the single-tag category are listed in Tables C.15 to C.18. Here, a three-parameter
 3465 fit is sufficient for all categories except the “high-purity” qW category where a 5-parameter fit is
 preferred.

WW category, HP			
Function	Residuals	χ^2	ndof
2 par	0.251	17.673	16
3 par	0.187	14.863	15
4 par	0.183	14.618	14
Fishers23	5.454	CL	0.033
Fishers34	0.391	CL	0.541

Table C.9: Residuals, χ^2 , and degrees of freedom for the WW category, HP category. A 3-parameter fit is needed to describe these data.

WW category, LP			
Function	Residuals	χ^2	ndof
2 par	2.974	13.997	23
3 par	3.082	14.775	22
4 par	3.080	14.768	21
Fishers23	-0.805	CL	1.000
Fishers34	0.015	CL	0.905

Table C.10: Residuals, χ^2 , and degrees of freedom for the WW category, LP category. A 2-parameter fit is needed to describe these data.

WZ category, HP			
Function	Residuals	χ^2	ndof
2 par	2.333	17.562	17
3 par	2.158	16.952	16
4 par	2.114	16.842	15
Fishers23	1.372	CL	0.258
Fishers34	0.338	CL	0.569

Table C.11: Residuals, χ^2 , and degrees of freedom for the WZ category, HP category. A 2-parameter fit is needed to describe these data.

WZ category, LP			
Function	Residuals	χ^2	ndof
2 par	12.301	21.368	25
3 par	6.827	20.715	24
4 par	6.521	20.419	23
Fishers23	20.046	CL	0.000
Fishers34	1.126	CL	0.299

Table C.12: Residuals, χ^2 , and degrees of freedom for the WZ category, LP category. A 3-parameter fit is needed to describe these data.

ZZ category, HP			
Function	Residuals	χ^2	ndof
2 par	0.634	17.919	17
3 par	0.662	17.400	16
4 par	0.716	17.096	15
Fishers23	-0.720	CL	1.000
Fishers34	-1.197	CL	1.000

Table C.13: Residuals, χ^2 , and degrees of freedom for the ZZ category, HP category. A 2-parameter fit is needed to describe these data.

ZZ category, LP			
Function	Residuals	χ^2	ndof
2 par	9.293	19.452	22
3 par	6.884	20.118	21
4 par	6.598	20.076	20
Fishers23	7.701	CL	0.011
Fishers34	0.909	CL	0.351

Table C.14: Residuals, χ^2 , and degrees of freedom for the ZZ category, LP category. A 3-parameter fit is needed to describe these data.

qW category, HP			
Function	Residuals	χ^2	ndof
3 par	69.757	30.375	30
4 par	59.677	28.318	29
5 par	25.298	21.815	28
Alt. 4 par	35.610	22.810	29
Alt. 5 par	25.634	22.687	28
Fishers34	5.067	CL	0.032
Fishers45	39.409	CL	0.000
FishersAlt4Alt5	11.285	CL	0.002

Table C.15: Residuals, χ^2 , and degrees of freedom for the qW category, HP category. A 5-parameter fit is needed to describe these data.

qW category, LP			
Function	Residuals	χ^2	ndof
3 par	153.869	38.713	35
4 par	156.715	38.586	34
5 par	201.767	38.167	33
Alt. 4 par	189.434	39.327	34
Alt. 5 par	192.782	39.170	33
Fishers34	-0.636	CL	1.000
Fishers45	-7.592	CL	1.000
FishersAlt4Alt5	-0.590	CL	1.000

Table C.16: Residuals, χ^2 , and degrees of freedom for the qW category, LP category. A 3-parameter fit is needed to describe these data.

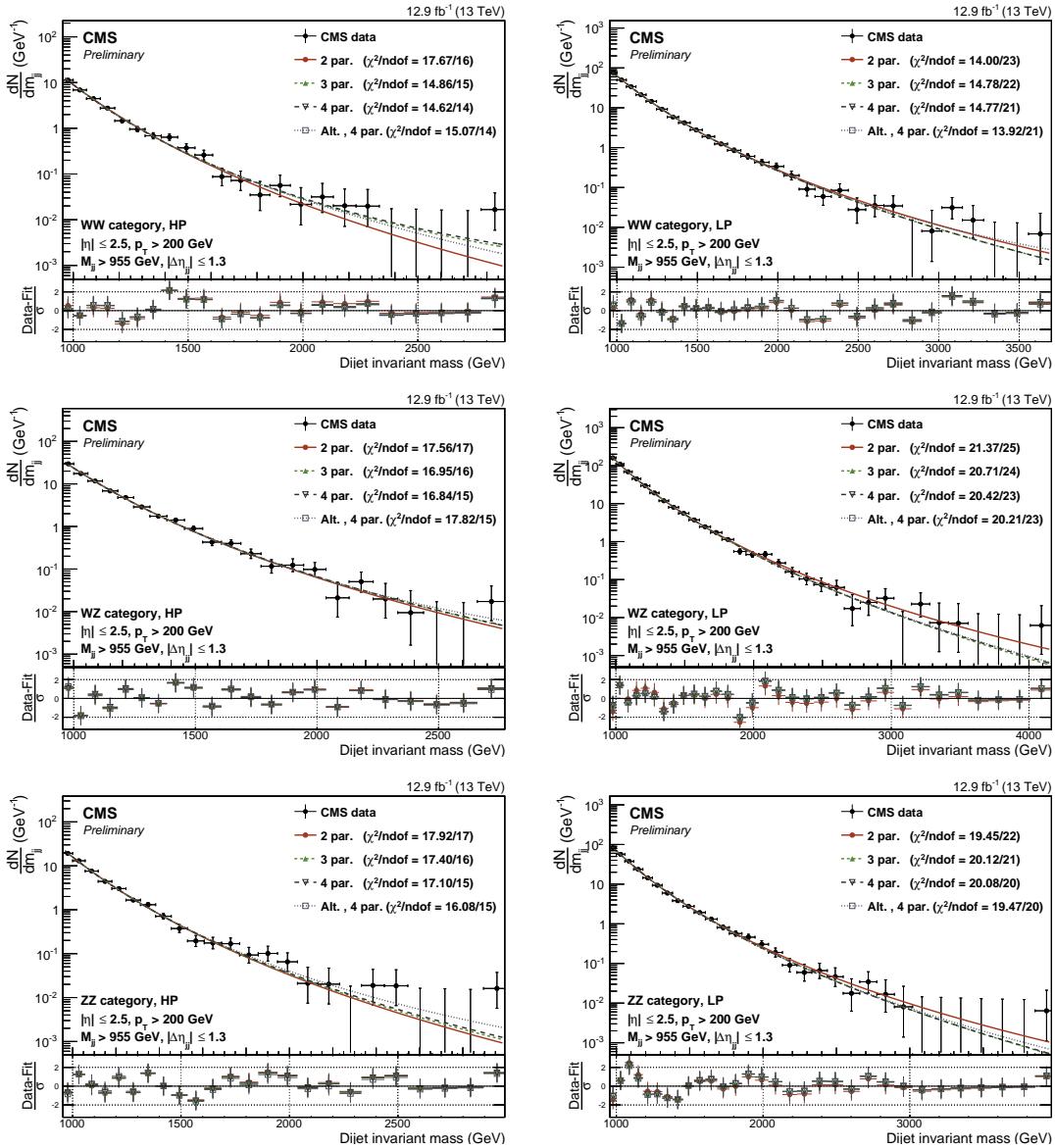


Figure C.11: Background fit for the M_{jj} distribution in the data signal region. Shown here are the high- (left) and low-purity (right) double W/Z-tag categories WW (top), WZ (middle) and ZZ (bottom).

qZ category, HP			
Function	Residuals	χ^2	ndof
3 par	12.963	21.252	30
4 par	12.961	21.252	29
5 par	9.256	19.644	28
Alt. 4 par	13.931	20.977	29
Alt. 5 par	9.739	20.344	28
Fishers34	0.004	CL	0.948
Fishers45	11.609	CL	0.002
FishersAlt4Alt5	12.484	CL	0.001

Table C.17: Residuals, χ^2 , and degrees of freedom for the qZ category, HP category. A 3-parameter fit is needed to describe these data.

qZ category, LP			
Function	Residuals	χ^2	ndof
3 par	369.554	47.426	36
4 par	369.554	47.426	35
5 par	298.358	46.525	34
Alt. 4 par	379.111	47.531	35
Alt. 5 par	379.120	47.531	34
Fishers34	0.000	CL	0.994
Fishers45	8.352	CL	0.007
FishersAlt4Alt5	-0.001	CL	0.000

Table C.18: Residuals, χ^2 , and degrees of freedom for the qZ category, LP category. A 3-parameter fit is needed to describe these data.

3467

APPENDIX D

3468

Search III

3469 D.1 Signal fits

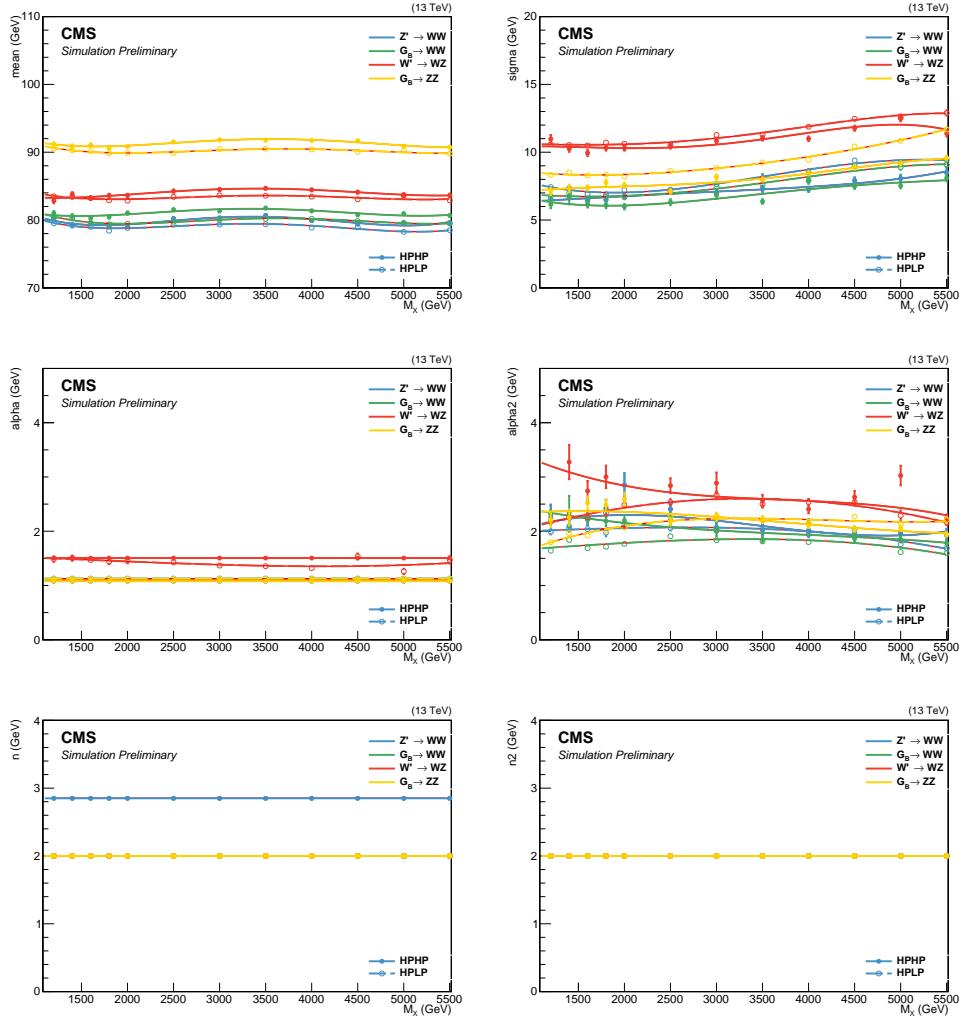


Figure D.1: The interpolated double Crystal-ball parameters for the softdrop jet mass of M_{jet2} as a function of M_X . To improve the stability of the fit some parameters are set constant.

3470 D.2 2016 kernels

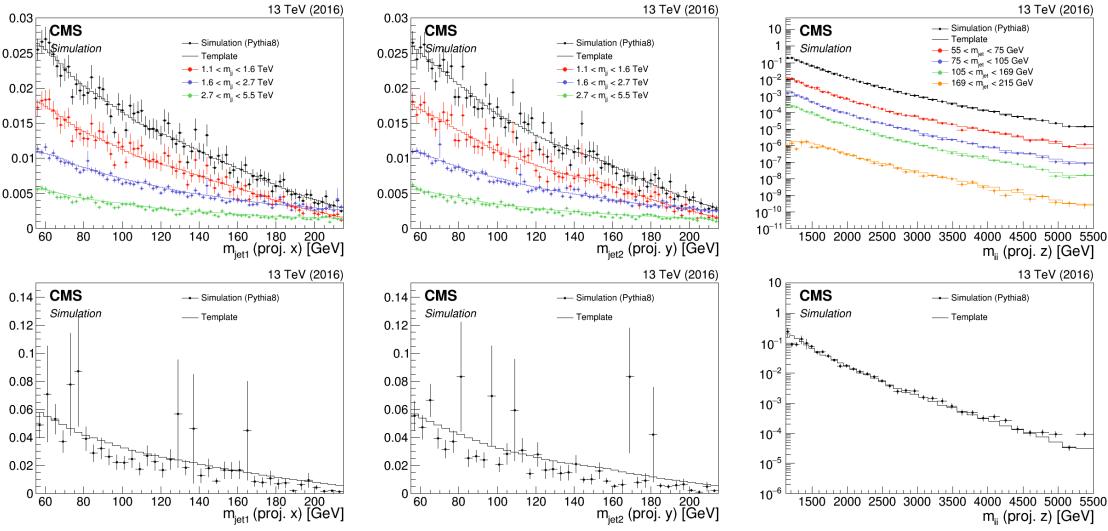


Figure D.2: Comparison between QCD MC simulation (markers) and kernels derived from generator-level quantities (lines) for the HP category (top) and the LP category (bottom) for 2016 MC. The kernels are shown for $M_{\text{jet}1}$ (left), $M_{\text{jet}2}$ (middle) and M_{VV} (right).

3471 D.3 Resonant background

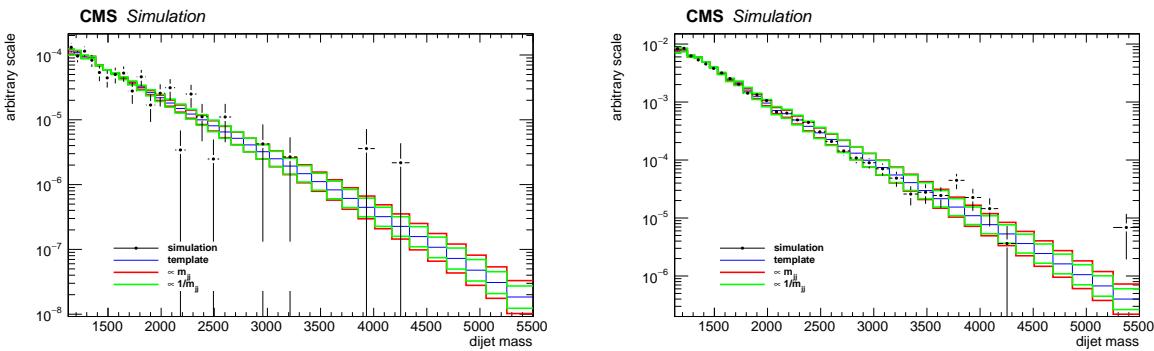


Figure D.3: One-dimensional M_{VV} kernels (solid line) compared to MC (markers) for the Z+jets background in the HP (left) and LP (right) categories. The nominal shape derived from the smoothing procedure can be seen as a blue line, alternative shapes derived from varying the slope of the M_{VV} spectrum are shown in green and red.

3472

Bibliography

- 3473 [1] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys.* **C40** (2016),
 3474 no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.
- 3475 [2] P. Collaboration, “Planck intermediate results - XLVI. Reduction of large-scale systematic
 3476 effects in HFI polarization maps and estimation of the reionization optical depth”, *A&A*
 3477 **596** (2016) A107, doi:10.1051/0004-6361/201628890.
- 3478 [3] M. Gell-Mann, “Symmetries of Baryons and Mesons”, *Phys. Rev.* **125** (Feb, 1962)
 3479 1067–1084, doi:10.1103/PhysRev.125.1067.
- 3480 [4] W. Pauli, “Mathematical contributions to the theory of Dirac’s matrices”, *Ann. Inst. H.
 3481 Poincare Phys. Theor.* **6** (1936) 109–136.
- 3482 [5] R. Tumulka, “Pauli Spin Matrices”, pp. 470–472. 07, 2009.
 3483 doi:10.1007/978-3-540-70626-7_142.
- 3484 [6] S. Duplij and F. Klinkhamer, “Levi-Civita Symbol”, pp. 227–227. Springer Netherlands,
 3485 Dordrecht, 2004. doi:10.1007/1-4020-4522-0_298.
- 3486 [7] S. Weinberg, “Conceptual Foundations of the Unified Theory of Weak and Electromagnetic
 3487 Interactions”, *Rev. Mod. Phys.* **52** (1980) 515–523, doi:10.1103/RevModPhys.52.515.
 3488 [543(1979)].
- 3489 [8] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys.
 3490 Rev. Lett.* **13** (Aug, 1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- 3491 [9] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13**
 3492 (Oct, 1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- 3493 [10] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless
 3494 Particles”, *Phys. Rev. Lett.* **13** (Nov, 1964) 585–587, doi:10.1103/PhysRevLett.13.585.

- 3495 [11] P. W. Anderson, “Coherent Excited States in the Theory of Superconductivity: Gauge
 3496 Invariance and the Meissner Effect”, *Phys. Rev.* **110** (May, 1958) 827–835,
 3497 doi:10.1103/PhysRev.110.827.
- 3498 [12] J. Goldstone, “Field Theories with Superconductor Solutions”, *Nuovo Cim.* **19** (1961)
 3499 154–164, doi:10.1007/BF02812722.
- 3500 [13] J. Ellis, “Higgs Physics”, arXiv:1312.5672.
- 3501 [14] B. Bellazzini, C. Csáki, and J. Serra, “Composite Higgses”, *Eur. Phys. J.* **C74** (2014),
 3502 no. 5, 2766, doi:10.1140/epjc/s10052-014-2766-x, arXiv:1401.2457.
- 3503 [15] R. Contino, D. Marzocca, D. Pappadopulo, and R. Rattazzi, “On the effect of resonances in
 3504 composite Higgs phenomenology”, *JHEP* **10** (2011) 081, doi:10.1007/JHEP10(2011)081,
 3505 arXiv:1109.1570.
- 3506 [16] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, “Heavy Vector Triplets: Bridging
 3507 Theory and Data”, *JHEP* **09** (2014) 060, doi:10.1007/JHEP09(2014)060,
 3508 arXiv:1402.4431.
- 3509 [17] L. Randall and R. Sundrum, “Large Mass Hierarchy from a Small Extra Dimension”, *Phys.
 3510 Rev. Lett.* **83** (Oct, 1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370.
- 3511 [18] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, “Warped gravitons at the CERN LHC
 3512 and beyond”, *Phys. Rev. D* **76** (Aug, 2007) 036006, doi:10.1103/PhysRevD.76.036006.
- 3513 [19] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, “Searching for the Kaluza-Klein
 3514 Graviton in Bulk RS Models”, *JHEP* **09** (2007) 013,
 3515 doi:10.1088/1126-6708/2007/09/013, arXiv:hep-ph/0701150.
- 3516 [20] CERN, “Accelerators and Schedules”, 2018 (accessed November 12, 2018),
 3517 <https://beams.web.cern.ch/content/accelerators-schedules>.
- 3518 [21] CMS, “CMS Detector Design”, 2018 (accessed November 21, 2018),
 3519 <https://cms.cern/news/cms-detector-design>.
- 3520 [22] T. Lenzi, “Development and Study of Different Muon Track Reconstruction Algorithms for
 3521 the Level-1 Trigger for the CMS Muon Upgrade with GEM Detectors”, Master’s thesis, U.
 3522 Brussels (main), 2013.
- 3523 [23] A. Dominguez et al., “CMS Technical Design Report for the Pixel Detector Upgrade”,
 3524 Technical Report CERN-LHCC-2012-016. CMS-TDR-11, CERN, Sep, 2012.

- 3525 [24] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004,
3526 doi:10.1088/1748-0221/3/08/S08004.
- 3527 [25] P. Adzic et al., “Energy resolution of the barrel of the CMS electromagnetic calorimeter”,
3528 *JINST* **2** (2007) P04004, doi:10.1088/1748-0221/2/04/P04004.
- 3529 [26] C. Biino, “The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1
3530 and future projections”, *Journal of Physics: Conference Series* **587** (2015), no. 1, 012001.
- 3531 [27] L. Veillet, “Assembly of the 7th wedge”, (Aug, 2000). CMS Collection.
- 3532 [28] S. Sharma, “Understanding the performance of CMS calorimeter”, *Pramana* **69** (Dec,
3533 2007) 1069–1074, doi:10.1007/s12043-007-0229-8.
- 3534 [29] CMS Collaboration, “CMS reconstruction improvement for the muon tracking by the RPC
3535 chambers”, *PoS RPC2012* (Sep, 2012) 045. 9 p. Presented by Minsuk Kim at the XI
3536 workshop on Resistive Plate Chambers and Related Detectors - RPC2012, INFN Laboratori
3537 Nazionali di Frascati Italy, February 5-10, 2012.
- 3538 [30] CMS, “CMS Luminosity - public results”, 2018 (accessed December 11, 2018),
3539 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- 3540 [31] P. Billoir, “Progressive track recognition with a Kalman-like fitting procedure”, *Computer
3541 Physics Communications* **57** (1989), no. 1, 390 – 394,
3542 doi:[https://doi.org/10.1016/0010-4655\(89\)90249-X](https://doi.org/10.1016/0010-4655(89)90249-X).
- 3543 [32] A. S. et. Al., “Particle-flow reconstruction and global event description with the CMS
3544 detector”, *Journal of Instrumentation* **12** (2017), no. 10, P10003.
- 3545 [33] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with
3546 the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear
3547 and Particle Physics* **31** (2005), no. 9, N9.
- 3548 [34] T. C. collaboration, “Performance of CMS muon reconstruction in pp collision events at s =
3549 7 TeV”, *Journal of Instrumentation* **7** (2012), no. 10, P10002.
- 3550 [35] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup per particle identification”, *Journal
3551 of High Energy Physics* **2014** (2014), no. 10, doi:10.1007/JHEP10(2014)059.
- 3552 [36] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering
3553 algorithms”, *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001,
3554 arXiv:[hep-ph/9707323](https://arxiv.org/abs/hep-ph/9707323).

- 3555 [37] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”,
 3556 *Phys. Rev.* **D48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160,
 3557 arXiv:hep-ph/9305266.
- 3558 [38] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04**
 3559 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- 3560 [39] J. POG, “Jet Identification: Recommendations for 13TeV data analysis”, 2018 (accessed
 3561 December 18, 2018), <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID>.
- 3562 [40] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum
 3563 Resolution in CMS”, *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002,
 3564 arXiv:1107.4277.
- 3565 [41] D. Krohn, J. Thaler, and L.-T. Wang, “Jet Trimming”, *JHEP* **02** (2010) 084,
 3566 doi:10.1007/JHEP02(2010)084, arXiv:0912.1342.
- 3567 [42] CMS Collaboration Collaboration, “Pileup Removal Algorithms”, Technical Report
 3568 CMS-PAS-JME-14-001, CERN, Geneva, 2014.
- 3569 [43] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, “Jet substructure as a new
 3570 Higgs search channel at the LHC”, *Phys. Rev. Lett.* **100** (2008) 242001,
 3571 doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.
- 3572 [44] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet
 3573 substructure”, *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.
- 3574 [45] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05** (2014) 146,
 3575 doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.
- 3576 [46] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP*
 3577 **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.
- 3578 [47] J. Thaler and K. Van Tilburg, “Maximizing boosted top identification by minimizing
 3579 N-subjettiness”, *Journal of High Energy Physics* **2012** (Feb, 2012) 93,
 3580 doi:10.1007/JHEP02(2012)093.
- 3581 [48] S. Bolognesi et al., “Spin and parity of a single-produced resonance at the LHC”, *Phys.
 3582 Rev. D* **86** (Nov, 2012) 095031, doi:10.1103/PhysRevD.86.095031.
- 3583 [49] CMS Collaboration, “Identification techniques for highly boosted W bosons that decay into
 3584 hadrons”, *JHEP* **12** (2014) 017, doi:10.1007/JHEP12(2014)017, arXiv:1410.4227.

- 3585 [50] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J.* **C58** (2008) 639–707,
3586 [doi:10.1140/epjc/s10052-008-0798-9](https://doi.org/10.1140/epjc/s10052-008-0798-9), arXiv:0803.0883.
- 3587 [51] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191**
3588 (2015) 159–177, [doi:10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024), arXiv:1410.3012.
- 3589 [52] The ATLAS collaboration, “Search for high-mass diboson resonances with boson-tagged jets
3590 in proton-proton collisions at $s=8$ TeV with the ATLAS detector”, *Journal of High Energy*
3591 *Physics* **2015** (Dec, 2015) 1–39, [doi:10.1007/JHEP12\(2015\)055](https://doi.org/10.1007/JHEP12(2015)055).
- 3592 [53] S. Frixione and B. R. Webber, “Matching NLO QCD computations and parton shower
3593 simulations”, *JHEP* **06** (2002) 029, [doi:10.1088/1126-6708/2002/06/029](https://doi.org/10.1088/1126-6708/2002/06/029),
3594 arXiv:hep-ph/0204244.
- 3595 [54] S. Frixione, P. Nason, and B. R. Webber, “Matching NLO QCD and parton showers in
3596 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),
3597 arXiv:hep-ph/0305252.
- 3598 [55] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton
3599 Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070,
3600 [doi:10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070), arXiv:0709.2092.
- 3601 [56] S. A. et. al., “Geant4—a simulation toolkit”, *Nuclear Instruments and Methods in Physics*
3602 *Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506**
3603 (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).
- 3604 [57] CMS Collaboration, “Event generator tunes obtained from underlying event and
3605 multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155,
3606 [doi:10.1140/epjc/s10052-016-3988-x](https://doi.org/10.1140/epjc/s10052-016-3988-x), arXiv:1512.00815.
- 3607 [58] CMS Collaboration, “Search for massive resonances in dijet systems containing jets tagged
3608 as W or Z boson decays in pp collisions at $\sqrt{s}=8$ TeV”, *JHEP* **08** (May, 2014) 173. 37 p.
- 3609 [59] F. Dias et al., “Combination of Run-1 Exotic Searches in Diboson Final States at the LHC”,
3610 *JHEP* **04** (2016) 155, [doi:10.1007/JHEP04\(2016\)155](https://doi.org/10.1007/JHEP04(2016)155), arXiv:1512.03371.
- 3611 [60] CMS Collaboration, “Search for massive WH resonances decaying into the $\ell\nu b\bar{b}$ final state
3612 at $\sqrt{s}=8$ TeV”, *Eur. Phys. J.* **C76** (2016), no. 5, 237,
3613 [doi:10.1140/epjc/s10052-016-4067-z](https://doi.org/10.1140/epjc/s10052-016-4067-z), arXiv:1601.06431.

- 3614 [61] CERN, “Collider Reach”, 2019 (accessed January 12, 2019),
 3615 <http://collider-reach.web.cern.ch/>.
- 3616 [62] J. Alwall et al., “The automated computation of tree-level and next-to-leading order
 3617 differential cross sections, and their matching to parton shower simulations”, *JHEP* **07**
 3618 (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- 3619 [63] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “A brief introduction to PYTHIA 8.1”, *Comput.
 3620 Phys. Commun.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.
- 3621 [64] S. C. et al., “Search for resonances in the dijet mass spectrum from 7 TeV pp collisions at
 3622 CMS”, *Physics Letters B* **704** (2011), no. 3, 123 – 142,
 3623 doi:<https://doi.org/10.1016/j.physletb.2011.09.015>.
- 3624 [65] G. Punzi, “Sensitivity of searches for new signals and its optimization”, *eConf* **C030908**
 3625 (2003) MODT002, arXiv:physics/0308063. [,79(2003)].
- 3626 [66] CMS Collaboration, “Search for heavy resonances in the W/Z-tagged dijet mass spectrum
 3627 in pp collisions at 7 TeV”, *Phys. Lett.* **B723** (2013) 280–301,
 3628 doi:10.1016/j.physletb.2013.05.040, arXiv:1212.1910.
- 3629 [67] CMS Collaboration Collaboration, “Search for Narrow Resonances using the Dijet Mass
 3630 Spectrum with 19.6fb-1 of pp Collisions at sqrt{s}=8 TeV”, Technical Report
 3631 CMS-PAS-EXO-12-059, CERN, Geneva, 2013.
- 3632 [68] D. J. Hand, “Statistical Concepts: A Second Course, Fourth Edition by Richard G. Lomax,
 3633 Debbie L. Hahs-Vaughn”, *International Statistical Review* **80** (2012), no. 3, 491–491.
- 3634 [69] T. C. collaboration, “Identification of b-quark jets with the CMS experiment”, *Journal of
 3635 Instrumentation* **8** (2013), no. 04, P04013.
- 3636 [70] T. C. collaboration, “Identification of heavy-flavour jets with the CMS detector in pp
 3637 collisions at 13 TeV”, *Journal of Instrumentation* **13** (2018), no. 05, P05011.
- 3638 [71] CMS Collaboration Collaboration, “Search for massive resonances decaying into WW, WZ,
 3639 ZZ, qW and qZ in the dijet final state at $\sqrt{s} = 13$ TeV using 2016 data”, Technical Report
 3640 CMS-PAS-B2G-16-021, CERN, Geneva, 2016.
- 3641 [72] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for likelihood-based
 3642 tests of new physics”, *The European Physical Journal C* **71** (Feb, 2011)
 3643 doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727.

- [73] M. Dasgupta, A. Powling, and A. Siodmok, “On jet substructure methods for signal jets”, *JHEP* **08** (2015) 079, doi:10.1007/JHEP08(2015)079, arXiv:1503.01088.
- [74] U. Baur, I. Hinchliffe, and D. Zeppenfeld, “Excited quark production at hadron colliders”, *Int. J. Mod. Phys. A* **2** (1987) 1285, doi:10.1142/S0217751X87000661.
- [75] U. Baur, M. Spira, and P. M. Zerwas, “Excited-quark and -lepton production at hadron colliders”, *Phys. Rev. D* **42** (1990) 815, doi:10.1103/PhysRevD.42.815.
- [76] CMS Collaboration Collaboration, “Search for massive resonances decaying into WW , WZ , ZZ , qW , and qZ with dijet final states at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **97** (Apr, 2018) 072006, doi:10.1103/PhysRevD.97.072006.
- [77] CMS Collaboration Collaboration, “Jet algorithms performance in 13 TeV data”, Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.
- [78] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J.* **C53** (2008) 473–500, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- [79] J. Dolen et al., “Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure”, arXiv:1603.00027.
- [80] J. Gallicchio and M. D. Schwartz, “Quark and Gluon Tagging at the LHC”, *Phys. Rev. Lett.* **107** (2011) 172001, doi:10.1103/PhysRevLett.107.172001, arXiv:1106.3076.
- [81] J. A. Aguilar-Saavedra, “Stealth bosons and where to find them”, 2018, <https://indico.cern.ch/event/649482/contributions/2993323/attachments/1688557/2716149/SBAWTFD.pdf>.
- [82] J. A. Aguilar-Saavedra, “Running bumps from stealth bosons”, *Eur. Phys. J.* **C78** (2018), no. 3, 206, doi:10.1140/epjc/s10052-018-5717-0, arXiv:1801.08129.
- [83] ATLAS Collaboration Collaboration, “Search for diboson resonances in hadronic final states in 79.8 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, Technical Report ATLAS-CONF-2018-016, CERN, Geneva, Jun, 2018.
- [84] T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson, “QCD or What?”, arXiv:1808.08979.
- [85] J. A. Aguilar-Saavedra, J. H. Collins, and R. K. Mishra, “A generic anti-QCD jet tagger”, *JHEP* **11** (2017) 163, doi:10.1007/JHEP11(2017)163, arXiv:1709.01087.

- 3674 [86] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, “Deep-learned Top Tagging with a
3675 Lorentz Layer”, *SciPost Phys.* **5** (2018), no. 3, 028,
3676 doi:10.21468/SciPostPhys.5.3.028, arXiv:1707.08966.
- 3677 [87] I. Goodfellow, Y. Bengio, and A. Courville, “Deep Learning”. The MIT Press, 2016.
- 3678 [88] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, *CoRR*
3679 abs/1412.6980 (2014) arXiv:1412.6980.
- 3680 [89] F. Chollet et al., “Keras”. <https://keras.io>, 2015.
- 3681 [90] M. A. et al, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”, 2015.
3682 Software available from tensorflow.org.
- 3683 [91] G. P. Salam, L. Schunk, and G. Soyez, “Dichroic subjettness ratios to distinguish colour
3684 flows in boosted boson tagging”, *JHEP* **03** (2017) 022, doi:10.1007/JHEP03(2017)022,
3685 arXiv:1612.03917.