

1

---

2

# Search for diboson resonances with CMS 3 and 4 Pixel Barrel Detector Calibration and Upgrade

5

---

6

Dissertation

7

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

10

vorgelegt der Mathematisch-naturwissenschaftlichen Fakultät  
11 der  
12 Universität Zürich

13



14

von  
15 Jennifer Ngadiuba

16

Promotionskomitee

17

Prof. Dr. Benjamin Kilminster  
18 Prof. Dr. Florencia Canelli  
19 Prof. Dr. Laura Baudis  
20 Prof. Dr. Ueli Straumann

21

Zürich 2016

# Contents

22

23	<b>1 Introduction</b>	1
24	<b>2 The Standard Model and beyond</b>	2
25	2.1 The Standard Model . . . . .	2
26	2.1.1 Particles and interactions . . . . .	2
27	2.1.2 Spontaneous symmetry breaking . . . . .	2
28	2.1.3 The Higgs mechanism . . . . .	2
29	2.1.4 The Higgs boson discovery at LHC . . . . .	2
30	2.2 The hierarchy problem and other SM limitations . . . . .	2
31	2.3 Theories of new physics . . . . .	2
32	2.3.1 Warped Extra dimensions . . . . .	2
33	2.3.2 Compositeness . . . . .	2
34	2.3.3 Heavy vector triplet . . . . .	2
35	<b>3 The CMS Experiment at the LHC</b>	3
36	3.1 The Large Hadron Collider . . . . .	3
37	3.2 The CMS Detector . . . . .	6
38	3.2.1 Tracking detectors . . . . .	7
39	3.2.2 Calorimetry . . . . .	10
40	3.2.3 Muon detectors . . . . .	12
41	3.2.4 The trigger system . . . . .	14
42	<b>I Search for diboson resonances with CMS</b>	18
43	<b>4 Diboson resonances as signature for new physics</b>	19
44	<b>5 Data sets and simulated samples</b>	22
45	5.1 Simulation of proton-proton collisions . . . . .	22
46	5.1.1 Monte Carlo event generators . . . . .	22
47	5.1.2 CMS detector simulation . . . . .	27
48	5.2 Simulated samples . . . . .	27
49	5.2.1 Simulation of signal processes . . . . .	27
50	5.2.2 Simulation of background processes . . . . .	30
51	5.3 Data sets . . . . .	31
52	<b>6 Object and event reconstruction</b>	34
53	6.1 Tracks and vertices . . . . .	34
54	6.1.1 Track reconstruction . . . . .	34
55	6.1.2 Vertex reconstruction . . . . .	36
56	6.2 Electrons . . . . .	37
57	6.2.1 Electron reconstruction . . . . .	37
58	6.2.2 Electron trigger . . . . .	39
59	6.2.3 Electron identification . . . . .	42
60	6.3 Muons . . . . .	46
61	6.3.1 Muon reconstruction . . . . .	46

---

62	6.3.2 Muon trigger . . . . .	49
63	6.3.3 Muon identification . . . . .	51
64	6.4 Jets . . . . .	52
65	6.4.1 Jet clustering algorithms . . . . .	53
66	6.4.2 Jet reconstruction and calibration . . . . .	56
67	6.4.3 Identification of b jets . . . . .	61
68	6.5 Missing transverse energy . . . . .	67
69	6.6 $W \rightarrow \ell\nu$ reconstruction . . . . .	69
70	<b>7 Identification of highly boosted <math>W/Z \rightarrow q\bar{q}^{(\prime)}</math> and <math>H \rightarrow b\bar{b}</math></b>	<b>71</b>
71	7.1 Jet substructure observables . . . . .	71
72	7.1.1 Pruned jet mass . . . . .	71
73	7.1.2 N-subjettiness . . . . .	72
74	7.2 The V tagging algorithm . . . . .	74
75	7.3 The H tagging algorithm . . . . .	76
76	<b>8 Analysis strategy</b>	<b>79</b>
77	8.1 Final event selection and categorization . . . . .	79
78	8.2 $W+jets$ background estimate with $\alpha$ ratio method . . . . .	83
79	8.2.1 Description . . . . .	83
80	8.2.2 Extraction of the $W+jets$ normalization . . . . .	84
81	8.2.3 Extraction of the $W+jets$ shape . . . . .	84
82	8.2.4 Validation of the $\alpha$ method . . . . .	88
83	8.3 Modelling of top quark production . . . . .	90
84	8.4 Signal modeling . . . . .	91
85	8.5 Systematic uncertainties . . . . .	95
86	8.5.1 Systematic uncertainties in the background estimation . . . . .	95
87	8.5.2 Systematic uncertainties in the signal prediction . . . . .	97
88	8.6 Testing new resonance hypothesis . . . . .	99
89	8.6.1 Limit setting procedure . . . . .	100
90	8.6.2 The asymptotic approximation . . . . .	102
91	8.6.3 Quantifying an excess of events . . . . .	104
92	<b>9 Results with 8 TeV data</b>	<b>106</b>
93	9.1 Final $m_{WH}$ distribution . . . . .	106
94	9.2 Significance of the data . . . . .	106
95	9.3 Cross section limits . . . . .	108
96	<b>10 Results with 13 TeV data</b>	<b>110</b>
97	10.1 Final $m_{WV}$ distribution . . . . .	110
98	10.2 Cross section limits . . . . .	111
99	<b>11 Combination of searches for diboson resonances at <math>\sqrt{s} = 8</math> and 13 TeV</b>	<b>113</b>
100	11.1 Inputs to the combination . . . . .	114
101	11.1.1 Reinterpretations . . . . .	114
102	11.2 Combination procedure . . . . .	116
103	11.3 Results . . . . .	117
104	11.3.1 Limits on $W'$ and $Z'$ singlets . . . . .	117
105	11.3.2 Limits on heavy vector triplet ( $W' + Z'$ ) . . . . .	119
106	11.3.3 Limits on bulk graviton . . . . .	120

---

107	11.3.4 Significance at 2 TeV . . . . .	120
108	<b>12 Conclusions</b>	<b>122</b>
109	<b>II Calibration and upgrade of the CMS pixel barrel detector</b>	<b>123</b>
110	<b>13 Introduction</b>	<b>124</b>
111	<b>14 The CMS pixel barrel detector</b>	<b>126</b>
112	14.1 Design . . . . .	126
113	14.2 Detector modules . . . . .	127
114	14.2.1 Sensor . . . . .	127
115	14.2.2 Readout chip . . . . .	130
116	14.2.3 Token bit manager . . . . .	132
117	14.3 Detector readout and control . . . . .	132
118	14.3.1 Readout of the analog signal . . . . .	133
119	14.3.2 Detector control and programming . . . . .	134
120	14.3.3 Supply tubes . . . . .	135
121	14.4 Pixel online software . . . . .	136
122	<b>15 Optimization and commissioning for LHC Run 2</b>	<b>138</b>
123	15.1 Effects of radiation damage in LHC Run 1 . . . . .	138
124	15.2 Optimization for LHC Run 2 . . . . .	141
125	15.3 Overview of pixel calibrations . . . . .	143
126	15.3.1 Adjustment of readout chain settings . . . . .	143
127	15.3.2 Optimization of signal performance . . . . .	151
128	15.4 Re-commissioning for LHC Run 2 . . . . .	155
129	15.4.1 Installation into CMS . . . . .	155
130	15.4.2 Calibrations at -10 °C . . . . .	157
131	15.5 Performance at the start of Run 2 . . . . .	158
132	<b>16 Phase I upgrade of the CMS pixel barrel detector</b>	<b>161</b>
133	16.1 Motivations . . . . .	162
134	16.2 Detector layout . . . . .	164
135	16.3 Pixel modules . . . . .	165
136	16.3.1 The digital ROC . . . . .	165
137	16.3.2 The TBM and readout . . . . .	166
138	16.4 Supply tubes . . . . .	169
139	16.5 The test stand . . . . .	169
140	16.6 Testing and calibration . . . . .	172
141	16.6.1 Delay adjustments . . . . .	173
142	16.6.2 POH bias and gain . . . . .	174
143	16.6.3 Read back test . . . . .	174
144	16.6.4 ROC analog current and digital voltage . . . . .	176
145	<b>17 Conclusions</b>	<b>178</b>

<sup>146</sup>	<b>III Summary</b>	<b>179</b>
<sup>147</sup>	<b>A Studies on track reconstruction problems</b>	<b>180</b>
<sup>148</sup>	<b>Bibliography</b>	<b>184</b>

149

## CHAPTER 1

150

# Introduction

151

---

# The Standard Model and beyond

## 155 2.1 The Standard Model

### 156 2.1.1 Particles and interactions

### 157 2.1.2 Spontaneous symmetry breaking

### 158 2.1.3 The Higgs mechanism

### 159 2.1.4 The Higgs boson discovery at LHC

## 160 2.2 The hierarchy problem and other SM limitations

## 161 2.3 Theories of new physics

### 162 2.3.1 Warped Extra dimensions

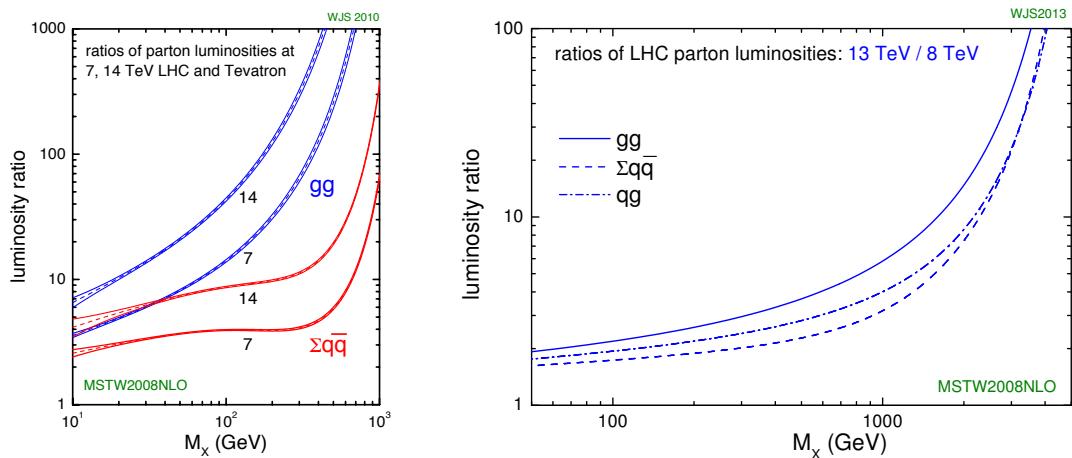
### 163 2.3.2 Compositeness

### 164 2.3.3 Heavy vector triplet

# The CMS Experiment at the LHC

## 168 3.1 The Large Hadron Collider

169 The Large Hadron Collider (LHC) [1] is a proton-proton (pp) collider located at the European  
 170 Particle Physics Laboratory (CERN) near Geneva, Switzerland. It is situated in the former  
 171 CERN Large Electron-Positron Collider (LEP) tunnel with a circumference of 27 km about  
 172 100 m under ground crossing the border between France and Switzerland. A hadron collider  
 173 has been chosen to allow higher center-of-mass energies ( $\sqrt{s}$ ) compared to electron-positron  
 174 colliders, the latter limited by synchrotron radiation due to the low mass of the particles to  
 175 be accelerated. High center-of-mass energies are required for the production of heavy SM  
 176 particles such as the top quark and the Higgs boson, and to search for new BSM interactions  
 177 at the TeV scale. For this purpose, the LHC is designed to produce pp collisions up to a  
 178  $\sqrt{s} = 14$  TeV, superseding previous high energy hadron colliders, such as Tevatron, by a  
 179 factor of 7. Higher center-of-mass energies lead to larger cross sections for the production  
 180 of the physics processes of interest in parton-parton interactions (Fig. 3.1), maximizing the  
 181 sensitivity to new discoveries. In addition to colliding protons, the LHC is also capable of  
 182 accelerating and colliding heavy nuclei, which is, however, not considered in this work.



**Figure 3.1:** (left) Parton luminosity ratios of pp collisions at LHC at  $\sqrt{s} = 7, 14$  TeV and p $\bar{p}$  collisions at Tevatron at  $\sqrt{s} = 1.96$  TeV. (right) Parton luminosity ratios of pp collisions at LHC at  $\sqrt{s} = 13$  TeV and at  $\sqrt{s} = 8$  TeV [2].

184 The LHC is the final element in a succession of machines that accelerate protons to  
 185 increasingly higher energies. Protons, obtained from a hydrogen source, are first accelerated  
 186 by a linear accelerator (LINAC 2) to energies of 50 MeV. The beam is then injected into the  
 187 Proton Synchrotron Booster (PSB), which accelerates the protons to 1.4 GeV, followed by the  
 Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the

188 Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Finally, the beam is  
189 injected in the LHC ring, where it completes several revolutions to reach the targeted energy.  
190 The LHC ring and the acceleration chain are sketched in Fig. 3.2.

191 Inside the ring, the two proton beams circulate in opposite directions in two tubes kept at  
192 ultrahigh vacuum, referred as beam pipes. The acceleration of protons inside LHC is made  
193 by radio-frequency cavities (400 MHz), giving a 492 keV energy gain per revolution, with a  
194 7 keV loss per turn due to synchrotron radiation. It takes 4 minutes and 20 seconds to fill  
195 each LHC ring, and 20 minutes for the protons to reach their maximum energy of 7 TeV. The  
196 maximum energy of the protons is limited by the strength of the magnetic field required for  
197 keeping the protons inside the ring. For 7 TeV-protons a magnetic field of 8.3 T has to be  
198 produced, which can only be reasonably obtained by superconducting magnets. The ring is  
199 equipped with 1232 dipole magnets for bending and 392 quadrupole magnets for focussing  
200 made of niobium-titanium (NbTi), which are cooled down to a temperature of 1.9 K with the  
201 help of super-fluid helium. After acceleration the protons move through the ring in separate  
202 bunches of protons with a fixed spatial separation.

203 The LHC ring has four interaction points at which the two counter rotating beams are  
204 made to cross and located in the center of the four LHC experiments. Just prior to collision,  
205 particles from the incoming beams must be squeezed closer together in order to maximize  
206 the chances of interaction. For this purpose, a system of three quadrupole magnets, so-called  
207 inner triplet, is located at both sides of each interaction point, which squeeze the beams and  
208 lead them to collisions in the center of the detector. Inner triplets tighten the beam, making  
209 it 12.5 times narrower from 0.2 mm down to 16  $\mu\text{m}$  across.

210 Besides the high center-of-mass energy required for the production of heavy particles, a  
211 high event rate has to be obtained to allow the discovery of processes with low production  
212 cross sections. The instantaneous luminosity  $\mathcal{L}$  characterizes the interaction rate. For a  
213 process with a cross section  $\sigma$ , the interaction rate is given by

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

214 The instantaneous luminosity depends only on the beam parameters and can be written  
215 for a Gaussian beam distribution as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \sigma_x \sigma_y}, \quad (3.2)$$

216 where  $N_b$  is the number of particles per bunch,  $n_b$  the number of bunches per beam,  $f_{\text{rev}}$   
217 the revolution frequency,  $\gamma_r$  the relativistic gamma factor, while  $\sigma_x$  and  $\sigma_y$  characterize the  
218 widths of the transverse beam profiles in the horizontal and vertical direction, respectively.  
219 The number of interaction events in a period of running time of the collider can be derived as

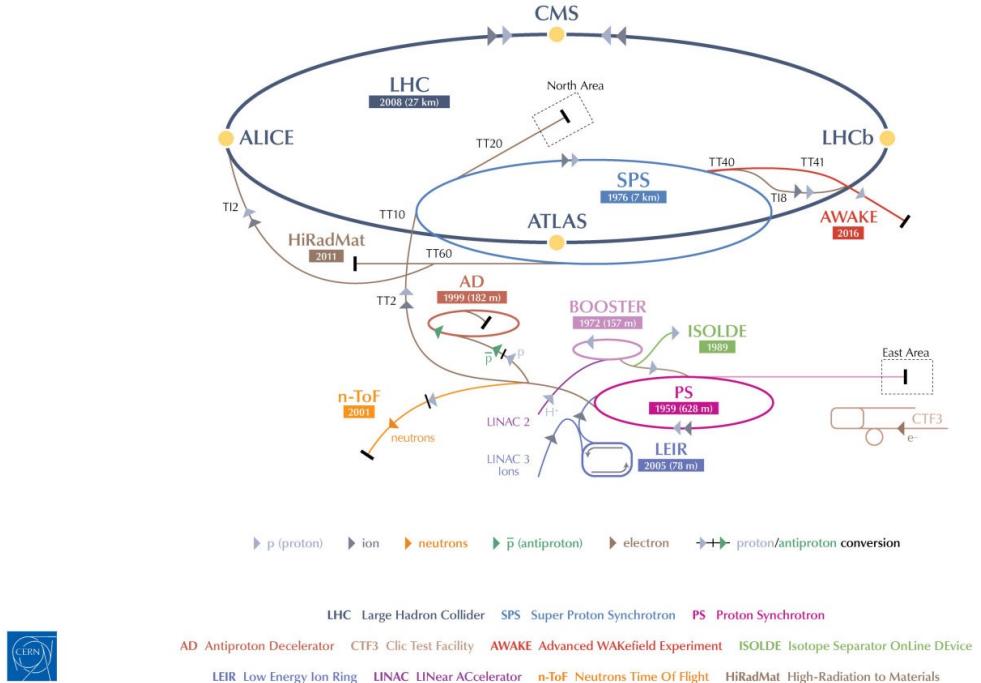
$$N_{ev} = \sigma \int \mathcal{L} dt = \sigma L, \quad (3.3)$$

220 where  $L$  is called the integrated luminosity. It is a measurement of the collected data size  
221 and it is usually expressed in inverse of cross section.

222 The LHC beams can reach very high luminosity with a high frequency bunch crossing  
223 and a high density of protons per bunch. In the ring, 2808 bunches of  $1.15 \times 10^{11}$  protons are  
224 circulated, with an average length of 7.5 cm, a width of about 16  $\mu\text{m}$  and a bunch spacing of  
225 25 ns (collision frequency of 40 MHz). This corresponds to the design instantaneous luminosity  
226 of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  for pp collisions, which supersedes by a factor of 100 the luminosity reached  
227 by previous hadron colliders.

228 Proton collisions take place in four points of the LHC tunnel where the four main experiments are located: ATLAS (*A Toroidal LHC ApparatuS*) [3], CMS (*Compact Muon*  
 229 *Solenoid*) [4], LHCb (*LHC beauty experiment*) [5] and ALICE (*A Lead Ion Collider Experiment*) [6]. ATLAS and CMS are general purpose experiments, designed to get an extensive  
 230 study of SM and BSM physics and to operate at the design luminosity. The LHCb experiment  
 231 is instead optimized for bottom quark physics studies while the ALICE experiment is  
 232 dedicated to the study of the lead-lead collisions at the design luminosity of  $10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ .  
 233

235



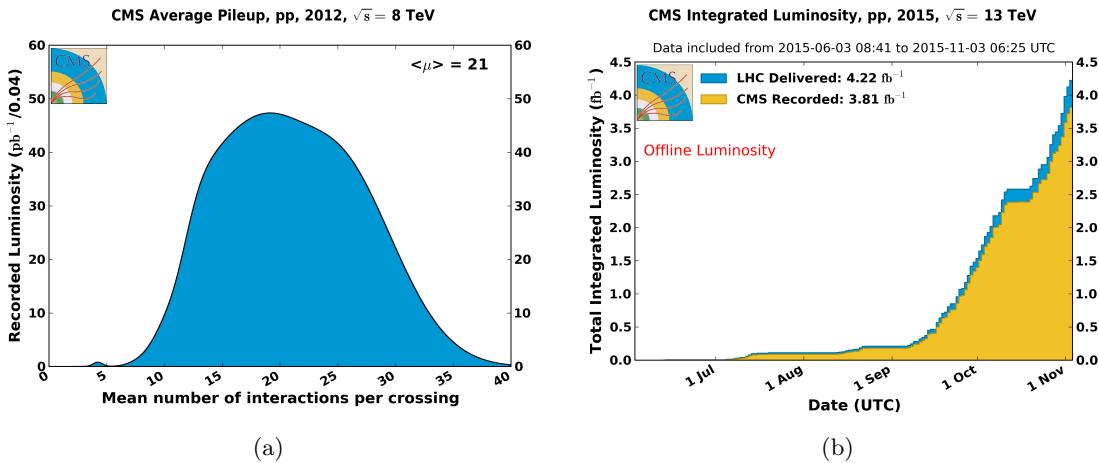
**Figure 3.2:** The CERN accelerator complex showing the chain of injection of protons into the LHC ring and the locations of the four main experiments ATLAS, CMS, LHCb and ALICE [7].

236 LHC operation officially started at the beginning of September 2008 but it was interrupted  
 237 after a short period, due to the breakdown of superconducting magnets. The collider has  
 238 been reactivated in November 2009 with first pp collisions at  $\sqrt{s} = 900 \text{ GeV}$ , officially starting  
 239 a new era in the particle physics experiments. The operating center-of-mass energies in pp  
 240 collisions have so far been 7 TeV in 2010-2011, 8 TeV in 2012 and 13 TeV in 2015-2016. The 7  
 241 and 8 TeV periods together make out the *LHC Run 1*, while the 13 TeV period is called the  
 242 *LHC Run 2*. The work presented in this document is based on data sets collected with pp  
 243 collisions at 8 TeV in 2012 and at 13 TeV in 2015.

244 During the whole Run 1, the LHC operated with a 50 ns bunch spacing. The peak  
 245 of instantaneous luminosity in 2011 has been  $\approx 0.4 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  with a total delivered  
 246 integrated luminosity of  $6.1 \text{ fb}^{-1}$  [8]. In 2012 the beam energy increased to 4 TeV per beam  
 247 with a peak luminosity of  $\approx 0.8 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  and  $23.3 \text{ fb}^{-1}$  delivered integrated luminosity  
 248 by the end of that year [8]. The increment of the instantaneous luminosity leads to a no more  
 249 negligible number of simultaneous interactions per bunch crossing, the so-called *pileup* (PU)  
 250 events. It depends on the cross section of inelastic collisions (75 mb at  $\sqrt{s} = 8 \text{ TeV}$  [9]) and it  
 251 is directly linked to the instantaneous luminosity. The average PU of the data collected in  
 252 2012 is equal to 21 (Fig. 3.3) while it has been around 15 in 2011 [8].

A long shut-down period for the LHC (LS1) occurred during the whole 2013 and 2014, where upgrades and technical improvements have been performed in order to reach the designed instantaneous luminosity and center-of-mass energy. On March, 21st 2015 the first pp collisions at  $\sqrt{s} = 13$  TeV has been obtained, a new record-breaking energy. For the first three months the machine operated with 50 ns bunch spacing while, from August 2015, it has been reduced to the designed 25 ns and the number of bunches per beam has been increased. The first part of this Run 2 phase ended on November 2015 with a total delivered integrated luminosity of  $4.2 \text{ fb}^{-1}$  and a peak luminosity of  $\approx 0.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  with an average pileup of 12 [8].

The LHC Run 2 has been restarted in April 2016, after an end-of-the-year technical stop, reaching a peak luminosity of  $\approx 1.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ . The machine has remained in operation at  $\sqrt{s} = 13$  TeV for the whole year with a total delivered integrated luminosity of  $40 \text{ fb}^{-1}$ . Accordingly to the current LHC schedule, the Run 2 will proceed up to the end of 2018 with a total expected integrated luminosity of  $\approx 150 \text{ fb}^{-1}$ . The data collected in 2016 are not considered in this work.



**Figure 3.3:** (a) Number of simultaneous interactions per bunch crossing in data collected in 2012 by the CMS experiment at LHC. (b) Cumulative luminosity versus day delivered by LHC (blue) in 2015; the offline luminosity recorded by the CMS experiment is also reported (orange). [8]

## 268 3.2 The CMS Detector

The CMS detector is a general purpose detector installed 100 m underground at the LHC interaction point 5 (P5) near the village of Cessy in France. It has been designed to exploit the different properties of the wide range of particles and phenomena produced in high-energy collisions at the LHC.

The design of this detector is driven by the challenges of a physics experiment in the LHC environment. Many of the physics processes of interest have a small cross section and the background from QCD jet production is overwhelmingly dominant. In order to achieve high rejection power with an optimal efficiency for rare channels, the detector has to be able to reconstruct the primary interaction entirely and to reduce the influence of overlapping events. Therefore, one needs to collect all possible information on the particles passing through the detector. Since these have different properties, a mixture of subdetectors is required

for a complete event reconstruction. The reconstruction of lepton signatures is essential for the extraction of rare processes and an excellent muon and electron identification and momentum resolution is desired. A precise measurement of secondary vertices and track impact parameters is fundamental for an efficient identification of heavy flavor quarks and  $\tau$  leptons. Moreover, a large hermetic geometric coverage is preferred, which allows for a precise estimate of the transverse momentum carried away by invisible particles through the sum of all visible particles.

The high peak luminosities of LHC lead to large pileup imposing further challenges to the design. As a consequence of pileup, the products of an interaction under study may be confused with those from other interactions in the same bunch crossing. This effect can be reduced by using high-granularity detectors resulting in low occupancy. In addition, the short bunch crossing requires fast response time and good time resolution of each detector element. Hence, a large number of detector channels and an excellent synchronization among them are necessary. Another challenge arises from the large flux of particles near the interaction point which leads to high radiation levels and the need of radiation hard detectors and front-end electronics.

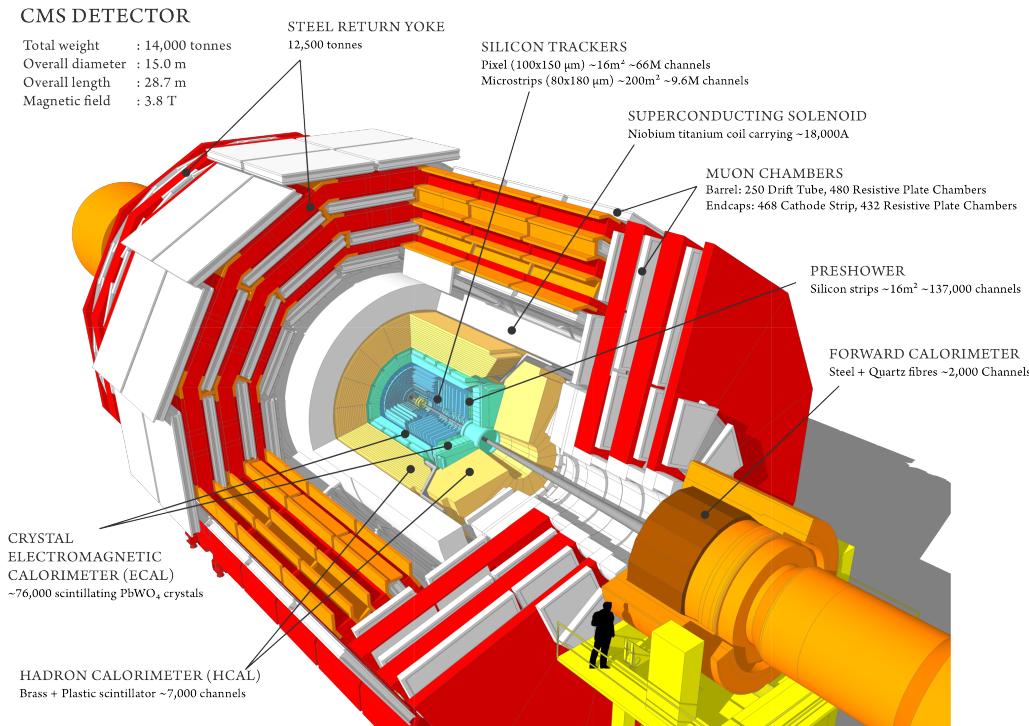
Figure 3.4 shows the layout of the CMS detector. The detector is built in a cylindrical structure composed of a barrel in the center and endcaps at both sides. This structure is 21.6-m-long, 14.6 m in circumference and 12500-tons-heavy. The detector design and layout was driven by the choice of the magnetic field configuration. Large bending power is needed for a precise measurement of the momentum of high-energy charged particles. Within the CMS detector this is achieved by a superconducting solenoid with a length of 12.9 m and an inner diameter of 5.9 m generating a magnetic field of 4 T. The bore of the magnet coil is large enough to accommodate the inner tracker and the calorimetry inside. The inner tracker consists of a pixel and a strip detector both made out of Silicon, representing the key component of CMS to measure the momenta of charged particles and identify primary and secondary vertices. The calorimetry system comprises a crystal electromagnetic calorimeter (ECAL) and a brass and scintillator hadronic calorimeter (HCAL), which provide information on the energies and directions of all charged and neutral particles. Outside the magnet are the large muon detectors, which, integrated inside the return yokes of the magnet, provide identification of muons and measurement of their momenta.

For the description of the CMS detector the following coordinate system is used. The origin is centered at the nominal collision point inside the experiment with the  $y$ -axis pointing vertically upward, the  $x$ -axis pointing radially inward toward the center of the LHC, and the  $z$ -axis points along the beam direction. The azimuthal angle  $\phi$  is measured from the  $x$ -axis in the  $x$ - $y$  plane. The polar angle  $\theta$  is measured from the  $z$ -axis. Pseudorapidity is defined as  $\eta = -\ln \tan(\theta/2)$ . Thus, the momentum and energy measured transverse to the beam direction, denoted by  $p_T$  and  $E_T$ , respectively, are computed from the  $x$  and  $y$  components.

In the following sections the three main components of the CMS detector will be described together with a section on the triggering system.

### 3.2.1 Tracking detectors

The tracking system of CMS (Fig. 3.5) is designed to provide a precise and efficient measurement of the trajectories of charged particles emerging from the LHC collisions, as well as a precise reconstruction of secondary vertices [10]. It surrounds the interaction point and has a length of 5.8 m and a diameter of 2.5 m providing coverage up to  $|\eta| < 2.5$ . In order



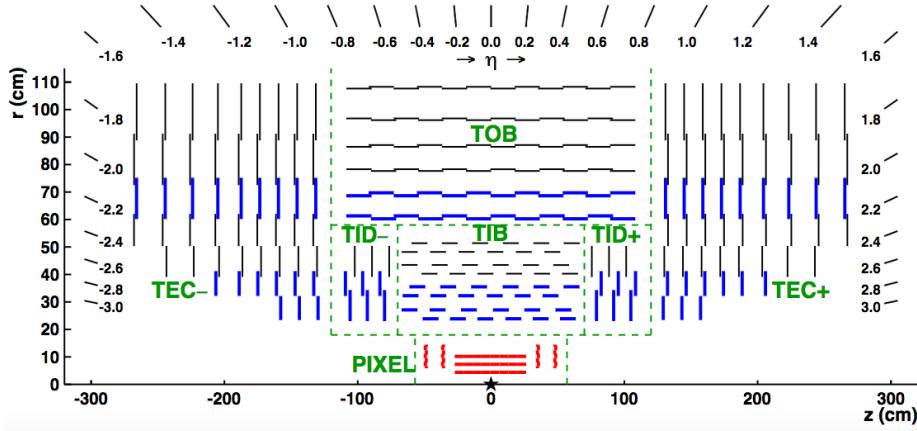
**Figure 3.4:** Layout of the CMS experiment and its subdetectors.

329 to achieve high tracking efficiency at the high luminosities of LHC, a detector technology  
 330 featuring granularity, speed and radiation hardness is required. Furthermore, the material  
 331 budget of the tracking system has to be as low as possible in order to avoid a worsening of the  
 332 tracking efficiency and resolution due to material interaction effects of the charged particle,  
 333 such as multiple scattering, bremsstrahlung, photon conversion or nuclear interactions. These  
 334 requirements lead to a tracker design entirely based on silicon detector technology. With  
 335 about  $200 \text{ m}^2$  of active silicon area the CMS tracker is the largest silicon tracker ever built.  
 336 It is divided into a pixel detector close to the interaction region and a strip detector in the  
 337 outer region. The motivations for this layout are explained in what follows.

338 At LHC design luminosity more than 1000 particles are hitting the tracking volume in  
 339 each bunch crossing. This leads to a hit rate density of  $1 \text{ MHz/mm}^2$  at a radius of 4 cm  
 340 which imposes severe challenges to the design of the tracking detectors. With a pixel size  
 341 of  $100 \times 150 \mu\text{m}^2$  in  $r\phi$  and  $z$ , respectively, an occupancy of the order of  $10^{-4}$  per pixel  
 342 and LHC bunch crossing can be achieved. The hit rate density falls with the distance from  
 343 the interaction point to  $60 \text{ kHz/mm}^2$  at a radius of 22 cm and to  $3 \text{ kHz/mm}^2$  at a radius of  
 344 115 cm. Therefore, at intermediate radii (20–55 cm), silicon micro-strip detectors are used,  
 345 with a typical cell length of 10 cm and a pitch of  $80 \mu\text{m}$ . At the outermost radii (55–110 cm)  
 346 the strip size can be further increased to  $25 \text{ cm} \times 180 \mu\text{m}$ . With this choice an occupancy of  
 347 less than 3% is maintained in the strip detector. However, the strip capacitance scales with  
 348 its length and therefore the electronics noise is a linear function of the strip length as well,  
 349 becoming not negligible in the outermost region where the strip size is the largest. In order  
 350 to maintain a good signal to noise ratio well above 10, CMS uses thicker silicon sensors for  
 351 the outer tracker region ( $500 \mu\text{m}$  thickness as opposed to the  $320 \mu\text{m}$  in the inner tracker)  
 352 with correspondingly higher signal. To mitigate the radiation damage effects and prolong  
 353 the lifetime of the detector modules, the tracking detectors are designed to run at subzero

temperatures. The cooling is established using a mono-phase liquid cooling system with  $\text{C}_6\text{F}_{14}$  as cooling fluid. The whole tracker system operated at  $+4^\circ \text{ C}$  during Run 1. After this phase, several improvements have been implemented and an operative temperature of  $-15^\circ \text{ C}$  is currently maintained for Run 2.

358

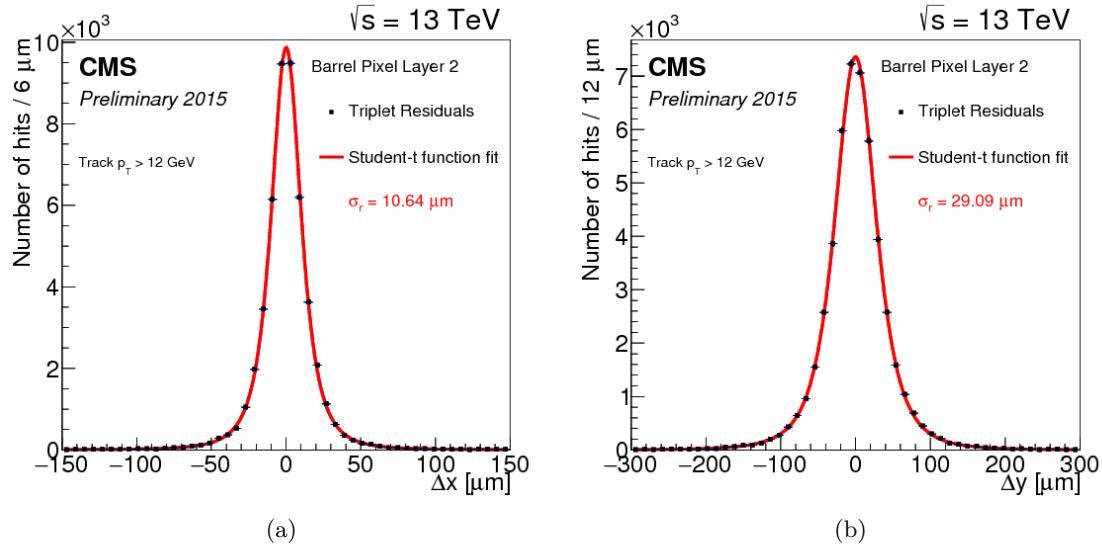


**Figure 3.5:** Longitudinal section of half of the CMS silicon tracker system; the different detector types are indicated.

359     The pixel detector is built from 3 barrel layers at radii of 4.4, 7.3 and 10.2 cm (BPix) and  
360     two end disks (FPix) on each side at a distance of  $z = \pm 34.5, \pm 46.5 \text{ cm}$  from the interaction  
361     point. It consists of 1440 segmented silicon sensor modules with a total of 66 million readout  
362     channels covering an area of about  $1 \text{ m}^2$ . The pixel detector is essential for the reconstruction  
363     of secondary vertices from the decay of bottom quarks and  $\tau$  leptons. It provides precise  
364     track point measurements in  $r$ - $\phi$  and  $z$  and therefore guarantees a small impact parameter  
365     resolution important for good secondary vertex reconstruction. This is achieved thanks to  
366     the readout of the analog pulse height information. The sensor surface in the barrel layers is  
367     parallel to the magnetic fields, hence the charge carriers produced by a particle traversing  
368     experience a Lorentz drift, which leads to charge spreading over more than one pixel. The  
369     analog pulse height information can be used to calculate a center of gravity of the charge  
370     distribution improving the hit information. The forward detectors are tilted at  $20^\circ$  in a  
371     turbine-like geometry to induce charge-sharing. As shown in Fig. 3.6, a spatial resolution of  
372      $10 \mu\text{m}$  in the transverse plane and  $30 \mu\text{m}$  in the longitudinal plane can be achieved for BPix.  
373     For FPix a spatial resolution of  $20 \mu\text{m}$  is obtained. A detailed description of the design and  
374     the functioning of the CMS pixel barrel detector is given in Chapter 14.

375

376     The strip detector occupies the radial region between 20 cm and 1.16 m. As illustrated in  
377     Fig. 3.5, it is composed of four subsystems: the four-layer Tracker Inner Barrel (TIB), the  
378     six-layer tracker outer barrel (TOB) and on each side three-disk Tracker Inner Disks (TID)  
379     and nine-disk Tracker Endcaps (TEC). The silicon micro-strip sensors have strips parallel  
380     to the beam axis in the barrel and radial on the disks. The modules in the first two layers  
381     and rings, respectively, of TIB, TID, and TOB as well as rings 1, 2, and 5 of the TECs carry  
382     a second micro-strip detector module which is mounted back-to-back with a stereo angle  
383     of 100 mrad in order to provide a measurement of the second coordinate ( $z$  in the barrel  
384     and  $r$  on the disks). This tracker layout ensures at least 9 hits in the silicon strip tracker in  
385     the full range  $|\eta| < 2.4$  with at least 4 of them being two-dimensional measurements. The



**Figure 3.6:** Distributions of the hit residuals on the pixel barrel layer 2 in the transverse (a) and longitudinal (b) direction with respect to the beam. The distributions are fitted with a Student's t-function. The fitted width parameter  $\sigma_r$  is reported on the plot [11].

386 total number of silicon sensors in the strip tracker is 24244, making up a total active area of  
 387  $198 \text{ m}^2$ , with about 9.3 million of strips.

### 388 3.2.2 Calorimetry

389 The calorimeter measures the energies and directions of all neutral and charged particles  
 390 traversing the detector, with the exception of muons and neutrinos. It consists of two parts,  
 391 the electromagnetic calorimeter (ECAL) [12] and the hadronic calorimeter (HCAL) [13].  
 392

393 The goal of ECAL is to measure precisely the energy of electrons and photons which  
 394 generate electromagnetic showers inside it. It is a hermetic and homogeneous calorimeter  
 395 with a large pseudorapidity coverage up to  $|\eta| < 3$ . As illustrated in Fig. 3.7, ECAL is  
 396 divided into barrel and endcap detectors consisting of scintillation crystals made from lead  
 397 tungstate ( $\text{PbWO}_4$ ). The choice of this material is motivated by its high density ( $8.28 \text{ g/cm}^3$   
 398 ), short radiation length ( $X_0 = 0.89 \text{ cm}$ ) and small Molière radius ( $2.2 \text{ cm}$ ), resulting in a high  
 399 stopping power, fine granularity and therefore a compact calorimeter able to fit inside the  
 400 solenoid. The ECAL comprises 61200 crystals in the barrel and 7324 crystals in each of the 2  
 401 endcaps, for a total volume of  $8.14 \text{ m}^3$  and  $2.9 \text{ m}^3$ , respectively. The crystals have a tapered  
 402 shape and are mounted in a quasi-projective geometry. The barrel extends radially between  
 403 1.29 and 1.75 cm covering the region  $|z| < 3.05 \text{ m}$  and  $|\eta| < 1.479$ . The crystals have a front  
 404 face cross-section of  $22 \times 22 \text{ mm}^2$  and a length of 2.3 cm ( $25.8 X_0$ ). They are organized in 36  
 405 identical supermodules each covering  $20^\circ$  in  $\phi$ . The crystals are contained in a thin-walled  
 406 glass-fibre alveola structures (“submodules”) with  $2(\phi) \times 5(\eta)$  crystals per each resulting in a  
 407 granularity 360-fold in  $\phi$  and  $2 \times 85$ -fold in  $\eta$ . The endcaps are placed at a distance of 3.14 m  
 408 from the interaction point and they extend radially between 3.16 and 17.11 cm, covering the  
 409 region  $1.479 < |\eta| < 3.0$ . The crystals have a front face cross section of  $28.6 \times 28.6 \text{ mm}^2$  and  
 410 a length of 2.2 cm ( $24.7 X_0$ ). A preshower detector with a thickness of  $3 X_0$  is placed in front  
 411 of the endcaps ( $1.653 < |\eta| < 2.6$ ) to guarantee a reliable discrimination of single photons

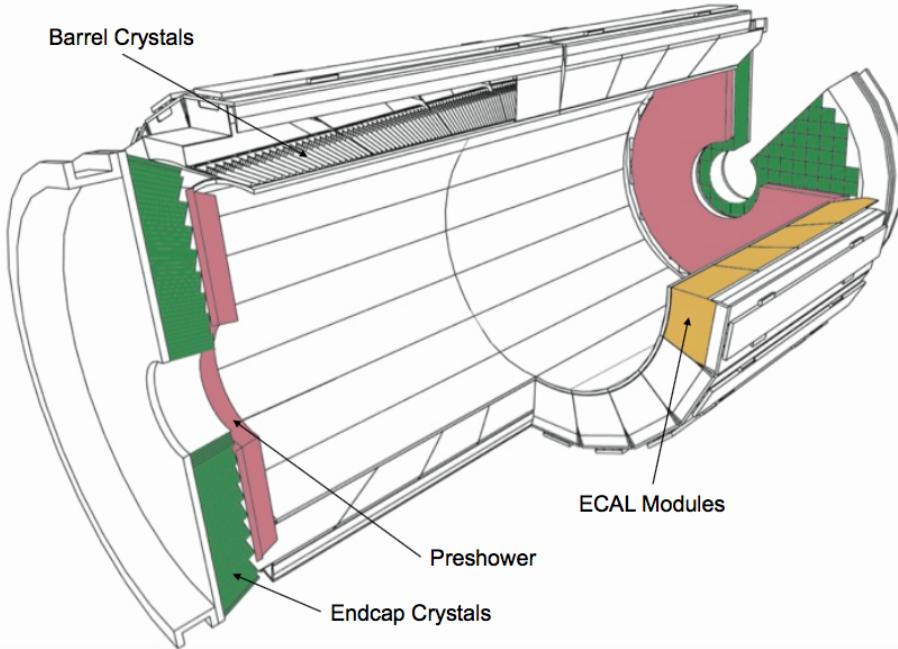
and photons produced in pairs in neutral pion decays. The relatively low light yield of the crystals ( $30 \gamma/\text{MeV}$ ) requires use of photodetectors with intrinsic gain that can operate in a magnetic field. Silicon avalanche photodiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in the endcaps. The light output and the amplification have a strong temperature dependence. The response to an incident electron changes by  $(3.8 \pm 0.4)\%/\text{^\circ C}$  which in turn means that the temperature has to be closely monitored and kept stable to a precision of  $\pm 0.05\text{^\circ C}$ . The nominal operating temperature of the ECAL is  $18\text{^\circ C}$  and is provided by a water cooling system.

The energy resolution of the electromagnetic calorimeter can be parametrized by the following expression:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E(\text{GeV})}} \oplus \frac{N}{E(\text{GeV})} \oplus C. \quad (3.4)$$

The first term is stochastic including contributions from the shower containment, the number of photoelectrons and the fluctuations in the gain process. The second contribution corresponds to the noise term, which includes noise in the readout electronics and fluctuations in pileup. The third term is a constant dominating the energy resolution for high-energy electron and photon showers. It depends on non-uniformity of the longitudinal light collection, energy leakage from the back of the calorimeter, single-channel response uniformity and stability. The value of the three coefficients were determined by measurements with electron beam in a matrix of  $3 \times 3$  crystals to be  $S = 2.8\%$ ,  $N = 12\%$  and  $C = 0.3\%$  [14].

430



**Figure 3.7:** Schematic view of the CMS electromagnetic calorimeter [4].

431 The energy measurement of the ECAL is complemented by the measurement of the  
 432 hadronic calorimeter. The HCAL is designed to be as near to hermetic around the interaction  
 433 region as possible to allow events with missing energy to be identified. It is a sampling  
 434 calorimeter composed of layers of brass absorber interlaced with tiles of plastic scintillators

as active material to detect the showers generated by the hadrons in the brass. The energy released in the scintillator tiles causes them to emit blue-violet light, a fraction of which is absorbed and re-emitted by embedded wavelength-shifting fibres in the green region of the spectrum. The green light is then carried by special fibre-optic waveguides to the readout system. The photodetection readout is based on multi-channel hybrid photodiodes (HPDs), which are photodetectors configured especially for CMS that can provide gain and operate in a high magnetic field.

Figure 3.8 shows a schematic cross section of the HCAL detectors. The hadron barrel (HB) and endcaps (HE) calorimeters sit behind the tracker and the electromagnetic calorimeter as seen from the interaction point. The HB is radially restricted between the outer extent of the electromagnetic calorimeter ( $r = 1.77\text{ m}$ ) and the inner extent of the magnet coil ( $r = 2.95\text{ m}$ ). This constrains the total amount of material which can be put in to absorb the hadronic shower. Therefore, an Outer Hadron (HO) calorimeter is placed outside the solenoid complementing the barrel calorimeter. The HO uses the solenoid as additional absorbing material and provides sufficient containment for hadronic showers with a thickness of 11.8 interaction lengths ( $\lambda_l$ ). The first scintillators are placed in front of the first absorber plate in order to sample showers developing in the material between the ECAL and the HCAL, while the last scintillators are installed after the last absorber plate to correct for late developing showers leaking out. A total amount of 70000 and 20916 scintillator tiles are installed in the HB and the HE, respectively. The HB and HE cover the region  $|\eta| < 1.3$  and  $1.3 < |\eta| < 3.0$ , respectively. Beyond  $|\eta| = 3$ , the Hadron Forward (HF) calorimeter placed at 11.2 m from the interaction point extends the pseudorapidity coverage down to  $|\eta| = 5.2$ . The HF is a sampling calorimeter made from steel absorber plates composed of 5 mm thick grooved plates with quartz fibers inserted as active medium. The signal is generated when charged shower particles above the threshold generate Cherenkov light in the quartz fibres, thereby rendering the calorimeter mostly sensitive to the electromagnetic component of showers. The calorimeter is segmented and arranged in towers as summarized in Table 3.1.

The HCAL energy resolution is

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E(\text{GeV})}} \oplus 5\% \quad (3.5)$$

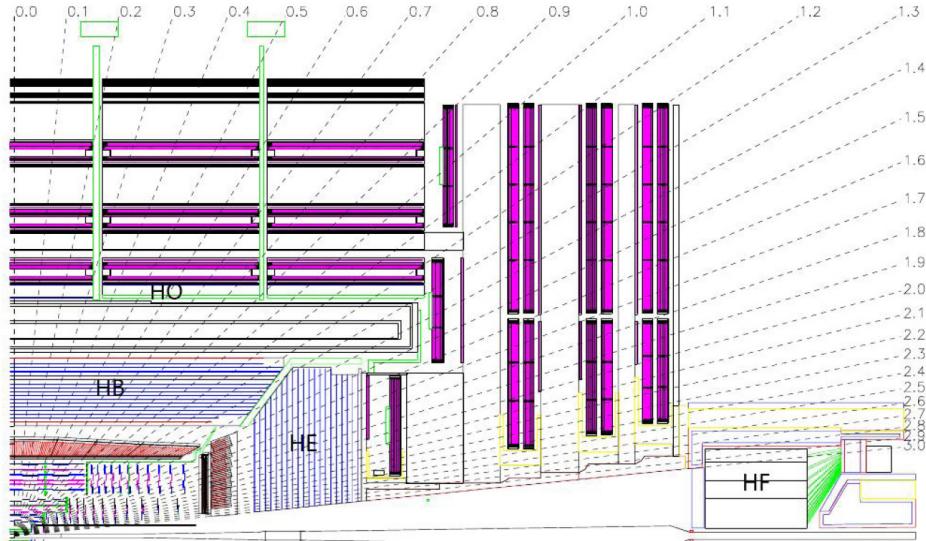
where  $a$  is 65% in the barrel, 85% in the endcaps and 100% in the forward calorimeter.

**Table 3.1:** Tower segmentation in azimuthal and polar angle for the hadronic barrel, endcap and forward calorimeters.

	HB/HO	HE ( $ \eta  \leq 2.5$ )	HE ( $ \eta  > 2.5$ )	HF ( $ \eta  \leq 4.7$ )	HF ( $ \eta  > 4.7$ )
$\Delta\phi \times \Delta\eta$	$0.087 \times 0.087$	$0.087 \times 0.087$	$0.175 \times 0.175$	$0.175 \times 0.175$	$0.175 \times 0.35$

### 3.2.3 Muon detectors

The muon system is the outermost part of the CMS detector. It is located in the steel return yoke of the solenoid, covering the pseudorapidity region  $|\eta| < 2.4$ . This is possible because muons are hardly affected by this large material budget. The coil acts as a shield from electromagnetic and hadronic particles escaping the calorimeters and the yoke provides a magnetic field between consecutive muon stations, allowing a momentum measurement independent from the inner tracker. The muon system is designed for three major functions: robust and fast identification of muons, good resolution of momentum measurement, and integration to a fast and reliable trigger system. The gaseous detectors have been chosen



**Figure 3.8:** Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [4].

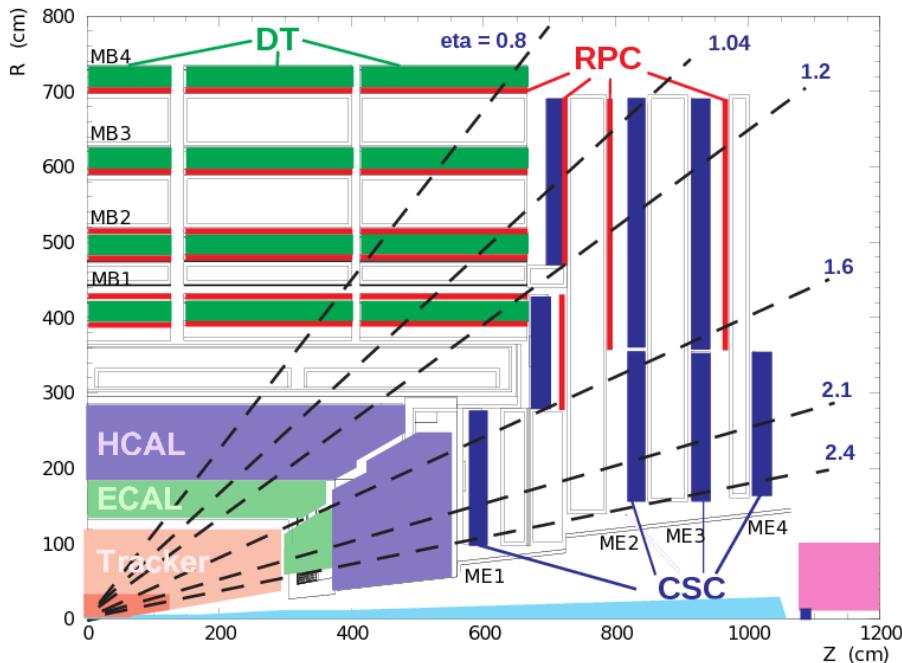
since they are robust and with a relative fast response. Furthermore, the area to be covered is extremely wide and a gaseous detector system allows to reduce the cost and the amount of readout channels. The muon system is thus composed of three types of gaseous detectors arranged in barrel and endcap sections, as shown in Fig. 3.9: Drift Tubes (DTs), Resistive Plate Chambers (RPCs) and Cathode Strip Chambers (CSCs). The choice of different detector topologies lies essentially in the different expected particle rates.

In the barrel region, where the neutron-induced background is small, the muon rate is low, and the 4-T magnetic field is uniform, DTs with standard rectangular drift cells are used covering the pseudorapidity region  $|\eta| < 1.2$ . A DT cell is a 4 cm wide gas tube with a positively charged stretched wire inside. The barrel DT chambers are organized in five separate wheels. Each wheel is divided into 12 sectors, each covering a  $30^\circ$  azimuthal angle. In each of the 12 sectors there are 4 chambers per wheel which are concentric around the beam line and separated by the iron return yoke. Each DT chamber, on average  $2\text{ m} \times 2.5\text{ m}$  in size, consists of 12 layers of DT cells, arranged in three groups of four. For the first 3 stations in each wheel, the middle group measures the  $z$  coordinate while the two outside groups measure the  $r\phi$  coordinate. The fourth and outermost station does not contain the  $z$ -measuring planes. Each one of the 250 DT chambers has a resolution of  $\approx 100\text{ }\mu\text{m}$  in  $r\phi$  and up to  $150\text{ }\mu\text{m}$  in  $z$ , and can measure the particle direction with 1 mrad accuracy.

In the two endcap regions of CMS, where the muon rates and background levels are high and the magnetic field is large and non-uniform, CSCs are used with their fast response time, fine segmentation, and radiation resistance, covering the pseudorapidity region between 0.9 and 2.4. Each CSC is trapezoidal shaped multiwire proportional chambers which consists of 6 gas gaps, each gap having a plane of radial cathode strips and a plane of anode wires running almost perpendicularly to the strips. The gas ionization and subsequent electron avalanche caused by a charged particle traversing each plane of a chamber produces a charge on the anode wire and an image charge on a group of cathode strips. Thus, each CSC provides a two-dimensional position measurement, where the  $r$  and  $\phi$  coordinates are determined by the cathode strips and the anode wires, respectively. A total amount of 540 CSC are arranged in 4 disks per endcaps divided in concentric rings (3 rings in the innermost station, 2 in the

502 others). Each chamber has a spatial resolution of about 200 mm in  $r$ , and  $75 \times 150 \mu\text{m}$  in the  
503  $r\phi$  coordinate.

504 In addition, there is a total of 610 RPCs added in both the barrel and endcap regions  
505 to provide a fast, independent, and highly-segmented trigger over a large portion of the  
506 rapidity range ( $|\eta| < 1.6$ ) of the muon system. They produce a fast response, with good  
507 time resolution ( $\approx 2$  ns) but coarser position resolution than the DTs or CSCs. RPCs are  
508 made from two high resistive plastic plates with a voltage applied and separated by a gas  
509 volume. The signal generated by the muon when passing through the gas volume is detected  
510 by readout strips mounted on top of one of the plastic plates. Six layers of RPCs are installed  
511 in the barrel muon system, two layers in each of the first two stations and one layer in each  
512 of the last two stations. One layer of RPCs is built into each of the first three stations of the  
513 endcap.



**Figure 3.9:** A longitudinal view of one quarter of the CMS experiment; the three muon detectors  
detector types are highlighted.

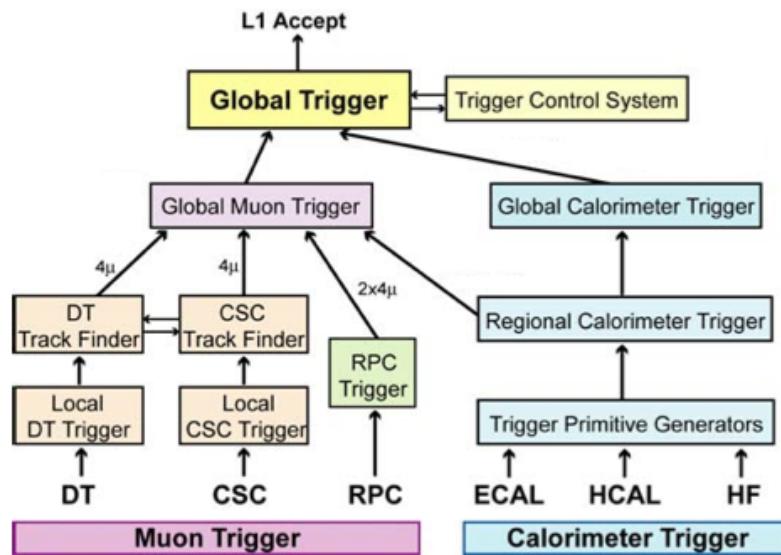
#### 514 3.2.4 The trigger system

515 The LHC provides proton-proton and heavy-ion collisions at unprecedented high luminosity  
516 and interaction rates. Given the high segmentation of the CMS detector, about 100 million  
517 readout channels are present and this corresponds to an enormous volume of data at the  
518 detector front-ends. At the design luminosity and collision frequency, each crossing produces  
519 approximately 1 MB of zero-suppressed data resulting in a raw data rate of about 40 TB  
520 per second. These figures are many orders of magnitude larger than the archival storage  
521 capability of  $\approx 1$  kHz at data rates of  $\mathcal{O}(10^2)$  MB/s. Technical difficulties in handling, storing  
522 and processing such extremely large amounts of data impose a reduction factor on the rate  
523 of events that can be written to permanent storage. This task is performed by the trigger  
524 system, which is the baseline of the physics event selection process. The key point of the  
525 trigger system is a fast time rejection of all the “non-interesting” events. This can be done

by exploiting event topologies common to group of physics processes, such as the presence of one or more leptons in the event. The trigger system needs to be as inclusive as possible, in order to collect data for all the physics searches that can be performed with pp collisions, but it has also to operate within the CMS time restriction and avoid the saturation of the storage capability. The required rejection power of  $\mathcal{O}(10^5)$  is too large to be achieved in a single processing step, if a high efficiency has to be maintained for the physics phenomena that CMS plans to study. For this reason, the full selection task is split into two steps. The first step (Level-1 Trigger) is designed to reduce the rate of events accepted for further processing to less than 100 kHz. The second step (High-Level Trigger or “HLT”) is designed to reduce this maximum L1 accept rate of 100 kHz to a final output rate of 1 kHz.

536

The L1 Trigger is built from custom designed, programmable electronics and is housed partly on the detectors, partly in the underground control room located at a distance of approximately 90 m from the experimental cavern. It is designed to take a fast accept/reject decision every bunch crossing, on the basis of a rough reconstruction of the event. The detector information used at L1 are coarsely segmented data from the calorimeters and the muon system only. Within a time budget of  $3.2 \mu\text{s}$ , it has to decide if an event is discarded or kept, and transfer this decision back to the subdetectors, which in the meantime keep the high resolution data in the front-end electronics. Figure 3.10 shows the L1 Trigger architecture: it has local, regional and global trigger components.



**Figure 3.10:** Architecture of the Level-1 Trigger [4].

Trigger primitives are generated by calculating the transverse energy of a trigger tower and assigning it to the correct bunch crossing. A regional calorimeter trigger then determines regional electron, photon and jet candidates and information relevant for muon and  $\tau$  lepton identification. The global calorimeter trigger provides information about the jets, the total transverse energy and the missing energy in the event and identifies the highest-ranking trigger candidates.

In the muon system all three types of detectors take part in the trigger decision. The DT chambers provide track segments in the projection and hit pattern in  $\eta$ , while the CSC determine three-dimensional track segments. The track finders in the DT chambers and the

555 CSCs calculate the transverse momentum of a track segment and its location and quality.  
 556 The RPCs deliver an independent measurement derived from regional hit patterns. The  
 557 global muon trigger receives up to four candidates from each subsystem (DT, barrel RPC,  
 558 CSC and endcap RPC) together with the isolation information from the global calorimeter  
 559 trigger. The aim is to improve the efficiency and to reduce the rate by making use of the  
 560 complementarity and the redundancy of the subsystems. In the end, the global muon trigger  
 561 selects a maximum of four muon trigger candidates and determines their momentum, charge,  
 562 position and quality.

563 The trigger objects extracted by the global calorimeter trigger and the global muon trigger  
 564 are sent to the global trigger where the decision to accept or reject an event is taken and  
 565 distributed to the subdetectors. The simplest triggers are in general those based on the  
 566 presence of one object with an  $E_T$  or  $p_T$  above a predefined threshold (single-object triggers)  
 567 and those based on the presence of two objects of the same type (di-object triggers) with  
 568 either symmetric or asymmetric thresholds. Other requirements are those for multiple objects  
 569 of the same or different types (“mixed” and multiple-object triggers). Up to 128 algorithms  
 570 can be executed in parallel. The decision is also based on the readiness of the subdetectors  
 571 and the data acquisition system (DAQ), which is supervised by the Trigger Control System  
 572 (TCS). The Level-1 Accept (L1A) decision is communicated to the subdetectors through the  
 573 Timing, Trigger and Control (TTC) system.

574

575 If an event is accepted by the L1 trigger, the full detector information ( $\approx 1$  MB) is readout  
 576 by the DAQ system and passed to the HLT system for further analysis. The HLT is a special  
 577 part of the CMS software which runs on a farm of several thousand processors performing  
 578 high-level object reconstruction and analysis. Each processor works on the reconstruction of  
 579 one event at a time, to get to a trigger decision within on average 100 ms. Since the time  
 580 budget for one event is much larger than at the L1 trigger, more complicated algorithms,  
 581 including tracking, can be executed at the HLT. Once an event is accepted, it is stored on  
 582 disk and fully reconstructed offline at a later time.

583 The full detector readout is available at HLT, but in order to meet the timing requirements  
 584 given by the input rate from L1, events are discarded before being fully reconstructed, as  
 585 soon there is enough reconstructed information to take the decision. Therefore the selection is  
 586 organized in a sequence of logical steps. The Level-2 uses the full information from calorimeters  
 587 and muon detectors to reconstruct the physical objects and to reduce the event rate by roughly  
 588 one order of magnitude. The data from the silicon tracker represent almost 80% of the event  
 589 size and require complex and time consuming algorithms for the reconstruction. For this  
 590 reason this information is used only during the Level-3 selection.

591 The HLT consists of approximately 400 trigger paths. Each trigger path starts from the  
 592 seed provided by the L1 trigger and it is built from reconstruction modules and filter modules.  
 593 After some parts of the data are reconstructed, a filter module decides if the reconstructed  
 594 objects pass the thresholds and the next step in reconstruction is started, or if the event is  
 595 not accepted by the path. In the later case, the execution of the path is stopped and the  
 596 following reconstruction steps and filter steps are not performed to save computation time. If  
 597 an event is not accepted by a path, it can still be accepted by a different path.

598 If, for some paths with low thresholds, the acceptance rate is too high, they can be  
 599 prescaled to lower the rate. A prescale value of ten means, for example, that the path is  
 600 executed only for every tenth event that was accepted by the L1 trigger, and, consequently,  
 601 the trigger rate for that path is ten times smaller. The prescale value for one trigger path  
 602 has several predefined levels, depending on the instantaneous luminosity of the LHC machine.  
 603 During an LHC fill, the instantaneous luminosity decreases, and the prescale values can be

604 changed during a CMS run to keep the global trigger rate at an optimal level.

605

## Part I

606

# Search for diboson resonances with CMS

607

# Diboson resonances as signature for new physics

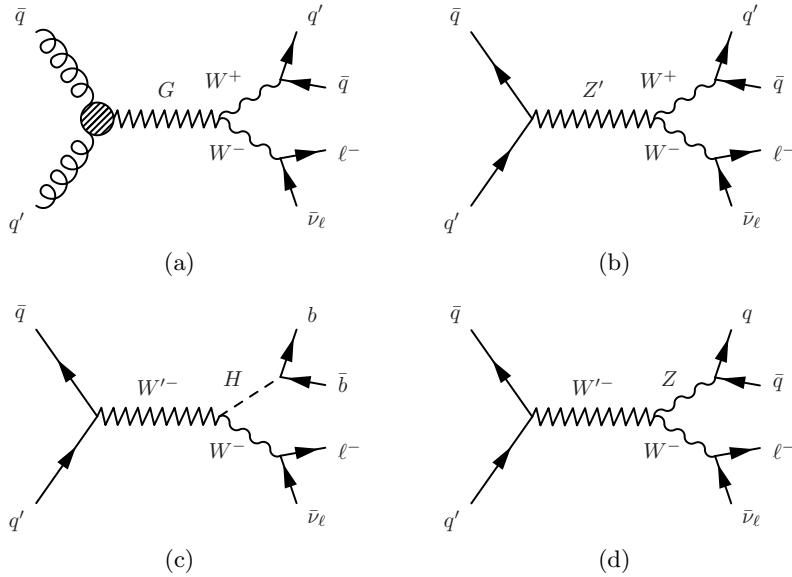
This part of the thesis is dedicated to the description and discussion of searches for new physics in proton-proton collision data collected with the CMS experiment at LHC. As pointed out in Chapter 2, the remarkable compatibility of the discovered scalar resonance by the ATLAS and CMS collaborations with the SM predictions for the Higgs boson, force physicists to deeply understand the role of naturalness in the dynamics of this particle. Several theoretical extensions to the SM have been proposed offering a concrete realization of naturalness, where new particles with masses in the TeV range generate loop corrections with the necessary cancellations to stabilize the Higgs boson mass. This means that the attention can be restricted to direct experimental manifestations of new physics represented by the production of reasonably narrow new particles. More natural solutions can therefore be probed at the LHC through the direct discovery of these new particles in final states with SM objects with known properties. The research described in this work follows exactly this approach and it is focused on the direct search for new massive resonances ( $M_X > 800 \text{ GeV}$ ) decaying to pairs of vector bosons (WW, WZ, or ZZ) or to a vector boson and a Higgs boson (WH or ZH). These decay modes can have large branching fractions in several BSM models. Popular examples include the bulk scenario of the Randall–Sundrum warped extra-dimensions described in Section 2.3.1, as well as the composite Higgs and Littlest Higgs models discussed in Section 2.3.2. Furthermore, the HVT model (Section 2.3.3) generalizes a large class of explicit theories that predict new heavy spin-1 vector bosons, adopting a simplified model strategy. The properties of the above benchmark models studied in this thesis are summarized in Table 4.1.

**Table 4.1:** Summary of the properties of the heavy resonance models considered in this work. The polarization of the produced W/Z boson in all considered models is mostly longitudinal.

model	particles	spin	charge	main production	main decay
HVT model A, $g_V = 1$	W' singlet	1	$\pm 1$	$q\bar{q}^{(\prime)}$	$q\bar{q}^{(\prime)}$
HVT model A, $g_V = 1$	Z' singlet	1	0	$q\bar{q}$	$q\bar{q}$
HVT model A, $g_V = 1$	W' + Z' triplet	1	0, $\pm 1$	$q\bar{q}/q\bar{q}^{(\prime)}$	$q\bar{q}/q\bar{q}^{(\prime)}$
HVT model B, $g_V = 3$	W' singlet	1	$\pm 1$	$q\bar{q}^{(\prime)}$	WZ, WH
HVT model B, $g_V = 3$	Z' singlet	1	0	$q\bar{q}$	WW, ZH
HVT model B, $g_V = 3$	W' + Z' triplet	1	0, $\pm 1$	$q\bar{q}/q\bar{q}^{(\prime)}$	WW, WZ, ZH
RS bulk scenario, $\bar{k} = 0.5$	$G_{\text{bulk}}$	2	0	$gg$	WW, ZZ

The signal under investigation is a narrow resonance, referring to the assumption that the resonance's natural width is smaller than the experimental resolution, covering a large fraction of the parameter space of the reference models considered. This assumption allows a “model-independent” type of search, where the description of the resonance mass distribution can be restricted to the detector effects only and hence, independently from the chosen benchmark model.

The semi-leptonic final states are considered, where one of the two bosons is a W decaying into a charged lepton ( $\ell$ ) and a neutrino ( $\nu$ ). The lepton can be either a muon ( $\mu$ ) or an electron (e), however, the results include the  $W \rightarrow \tau\nu$  contribution from the decay  $\tau \rightarrow \ell\nu\bar{\nu}$ . However, the gain in sensitivity from  $\tau$  leptons is limited by the small branching fractions involved. The second boson in the final state decays into hadrons, and can be either a vector boson  $V = W$  or  $Z$ , or a Higgs boson. In the first case, the final state is labelled as  $\ell\nu q\bar{q}$  including  $W \rightarrow q\bar{q}^{(\prime)}$  and  $Z \rightarrow q\bar{q}$  decays (Figures 4.1(a), 4.1(b) and 4.1(c)). For the Higgs boson, the final state is labeled as  $\ell\nu b\bar{b}$  referring to the Higgs boson decay into a bottom quark-antiquark pair (Fig. 4.1(d)).



**Figure 4.1:** Feynman diagrams for the production of a neutral spin-2  $G$  (a), and a neutral  $Z'$  (b) and charged  $W'$  (c and d) spin-1 resonances. All resonances decay to a pair of bosons (WW, WZ, or WH) with their subsequent semi-leptonic decay.

The search in the  $\ell\nu b\bar{b}$  final state is based pp collision data at  $\sqrt{s} = 8$  TeV collected in 2012 and corresponding to an integrated luminosity of  $19.7 \text{ fb}^{-1}$ . The second analysis described in this thesis and focused on the  $\ell\nu q\bar{q}$  final state is instead based on the pp collision data at  $\sqrt{s} = 13$  TeV collected in 2015 and corresponding to an integrated luminosity of  $2.3 \text{ fb}^{-1}$ . Although different algorithms are used for the reconstruction and identification of the hadronically decaying boson, the analysis strategy is similar in the two searches.

The key challenge of these analyses is the reconstruction of the highly energetic decay products. Since the resonances under study have masses of  $\approx$  TeV, their decay products, i.e. the bosons, have on average transverse momenta of several hundred GeV or more. As a consequence, the particles emerging from the boson decays are very collimated. In particular, the jet-decay products of the bosons cannot be resolved using the standard algorithms, but are instead reconstructed as a single jet object. Dedicated techniques, so-called jet “V tagging” and “H tagging” techniques, are applied to exploit the substructure of such jet objects, and can help resolve jet decays of massive bosons. These techniques also help to suppress SM background, which mainly originates from the production of W bosons in association with jets (W+jets). Further discrimination is achieved in the  $\ell\nu b\bar{b}$  analysis channel exploiting the specific characteristics of jets arising from the hadronization of bottom quarks.

668     The aim is to reconstruct the full event to be able to search for a localized enhancement in  
 669     the invariant mass of the WV or WH system on the top of a smoothly falling SM background  
 670     distribution. The background mainly comprises W+jets production, although another signifi-  
 671     cant contribution is represented by events involving pair produced top quarks ( $t\bar{t}$ ). Other  
 672     minor backgrounds are represented by single top quark and SM diboson (WW, WZ or ZZ)  
 673     production processes.

674     The invariant mass of the WV and WH system is determined by estimating the neutrino  
 675     transverse momentum with the measured missing transverse energy in the event, while an  
 676     estimate of the neutrino longitudinal momentum is derived by imposing the constraint of  
 677     the W mass on the invariant mass of the  $\ell\nu$  system. In the following, the diboson invariant  
 678     mass will be labelled either  $m_{\ell\nu+\text{jet}}$ , or  $m_{\text{WV}}$  and  $m_{\text{WH}}$  for the  $\ell\nu q\bar{q}$  and  $\ell\nu b\bar{b}$  final states,  
 679     respectively. The mass spectrum for the dominant W+jets background is determined from  
 680     observed events with a reconstructed jet mass not compatible with the V or H hypothesis.  
 681     This strategy partially relies on the simulation of the background processes. Furthermore,  
 682     simulated events are used for the optimization of the analysis selection aimed at maximiz-  
 683     ing the discrimination of the signal against the background and hence the analysis sensitivity.  
 684

685     This part of the thesis is organized as follows. Chapter 5 gives an overview of the methods  
 686     used to simulate the physics processes happening in pp collisions at the LHC together with a  
 687     description of the specific simulated background and signal events used in this analysis, as  
 688     well as a discussion about the data sets analyzed. Chapter 6 provides a detailed description  
 689     of the algorithms used in CMS for the reconstruction of the event and of the physics objects  
 690     expected in the semi-leptonic final states under investigation. Particular attention is given to  
 691     the V and H tagging algorithms representing the key feature of this analysis and therefore,  
 692     separately discussed in Chapter 7. The analysis strategy, already outlined here, is explained  
 693     in details in Chapter 8. This includes the final event selection and categorization optimized  
 694     to enhance the analysis sensitivity, as well as the strategy for the estimation of the expected  
 695     background, the modelling of the signal and the related systematic uncertainties which will  
 696     be used as input to the statistical analysis of the diboson invariant mass distribution observed  
 697     in data. The final results are discussed in Chapters 9 and 10 for the 8 and 13 TeV data  
 698     analysis, respectively. Eventually, these results are combined with limits derived in companion  
 699     CMS searches for resonances decaying to a pair of bosons in several different final states,  
 700     with data collected in both LHC Run 1 and Run 2. These analyses use the same V and  
 701     H tagging techniques as presented here to separate the signal from the large multijet or  
 702     V+jets background. The statistical combination represent the last piece of this work and it  
 703     is presented in Chapter 11.

# Data sets and simulated samples

708 The simulation of pp collisions is usually performed by means of Monte Carlo (MC) event  
 709 generators, providing an accurate modelling of the event kinematics and topology at parton  
 710 and hadron level. The hard inelastic scattering has to be fully calculated: from the hard  
 711 interaction between the partons inside the protons, where perturbative QCD calculations  
 712 (**FIXME: point to theory**) can be used, to the formation of particle jets from the outgoing  
 713 partons. Furthermore, it is fundamental to understand the exact response of the detector to  
 714 the outgoing particles produced in pp collisions. Consequently, the stable outgoing particles  
 715 are fed to a full detector simulation that models the interaction of those particles with the  
 716 detector material and the corresponding detector response. The raw detector data are then  
 717 subject to the same reconstruction algorithms that are also used for real data. In this chapter,  
 718 MC event generators are described in detail, followed by a brief description of the CMS  
 719 detector simulation. Finally, few details are given in the last section on the pp collision data  
 720 sets used to perform the searches described in this thesis.

721 **5.1 Simulation of proton-proton collisions**

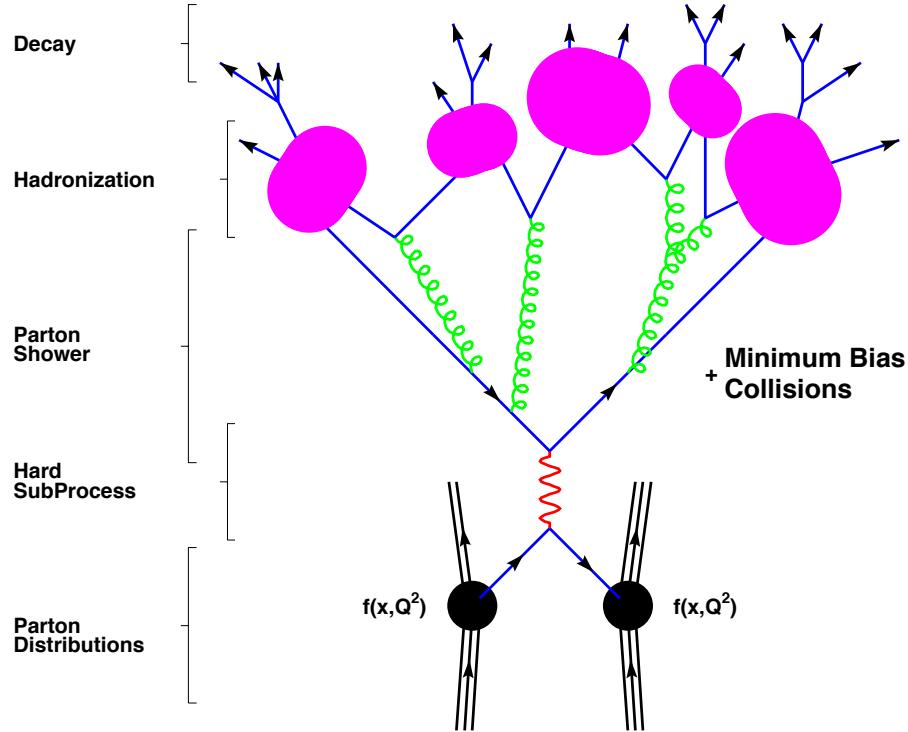
722 **5.1.1 Monte Carlo event generators**

723 The generation of hard inelastic pp collisions is factorized into different steps ordered by the  
 724 timescale on which they happen, as illustrated in Fig. 5.1, and described in the following.

726 The basis of theoretical event generation at the LHC is a parametrisation of the incoming  
 727 partons (quarks, anti-quarks and gluons) stemming from the proton, which is given by the  
 728 parton density functions (PDF). They describe the probability to find a quark or gluon with  
 729 a given proton momentum fraction  $x$  in the pp collision taking place at the LHC. In QCD  
 730 the PDFs depend on a factorization scale  $\mu_F^2$  at which the proton is probed. All interactions  
 731 between quarks and gluons happening at scales below the scale  $\mu_F^2$  are absorbed into the  
 732 PDFs. Therefore at small  $\mu_F^2$  the proton is observed basically as a combination of its three  
 733 valence quarks  $uud$ . At higher scales, however, it is dominated by sea quarks and gluons.

734 A collision between two partons, one from each side, gives the hard process of interest,  
 735 which can be due to physics within or beyond the standard model. Using the incoming  
 736 partons as input, the simulation of the hard process is performed by the event generator. It  
 737 produces hypothetical events with the distributions and rates predicted by theory based on  
 738 the cross section formulae of the physics process of interest. Using the cross section formula  
 739 the phase space is sampled and candidate events are defined by choosing values for the degrees  
 740 of freedom from a uniformly distributed random number generator.

741 The cross section can be calculated by means of the so called *factorization theorem* [16].  
 742 According to the theorem, the hadron itself is described by the whole particle composition  
 743 interacting on a soft binding energy scale, whereas the collisions occur between the partons  
 744 on a hard energy scale with large transverse momenta. The cross section for the process is  
 745 then given by the convolution of the PDF  $f_i(x, Q^2)$ , integrated over the proton momentum



**Figure 5.1:** Steps of Monte Carlo event generation as described in the text evolving in time from bottom to top [15].

fraction  $x$ , for the colliding protons (A, B) at an energy scale  $Q^2$ , and the hard parton-parton cross sections  $\hat{\sigma}_{ij} \rightarrow X$  for all combinations of two partons i and j:

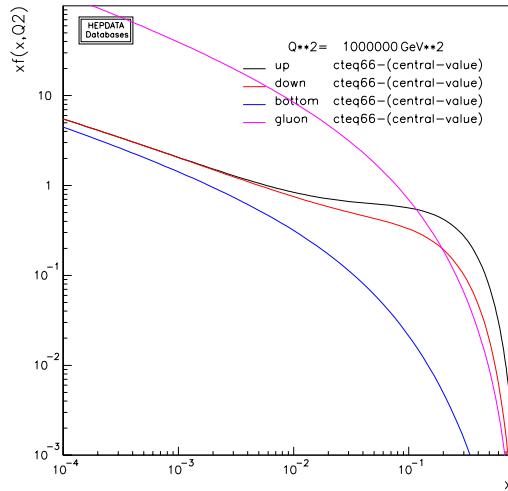
$$\sigma(AB \rightarrow X) = \sum_{q,g=0}^n \alpha_S^n(\mu_R^2) \sum_{ij} \int dx_i dx_j f_{i,A}(x_i, \mu_F^2) f_{j,B}(x_j, \mu_F^2) \cdot \hat{\sigma}_{ij \rightarrow X}^{(n)}(s; x_i, x_j, \mu_R^2, \mu_F^2). \quad (5.1)$$

In this equation the index  $n$  runs over the perturbative order and  $s$  is the squared center-of-mass energy of the collision. The tree-level process, where no emission of gluons or quarks happens, is called “Leading Order” (LO) and takes place when  $n = 0$ . Further orders are called “Next-to-Leading Order” (NLO,  $n = 1$ ), “Next-to-Next-to-Leading Order” (NNLO,  $n = 2$ ) and so on.

As it can be seen from the formula, the PDFs play a fundamental role in the description of the hard process, and it is very important to have several experimental tests to access their values. In fact, perturbative QCD cannot predict the PDFs, since they contain also the low energy (non-perturbative) information about the scattering. As a consequence, PDFs distributions are extracted from data, in deep-inelastic scattering experiments. Most of the parametrizations of proton PDFs now used for the LHC have been extracted from the ZEUS [17] and H1 [18] experiments in electron-proton collisions at the HERA collider and fixed target experiments. The more recent parametrizations also take into account vector boson production and single-inclusive jet production from the Tevatron experiments, as well as LHC data. Once measured for a certain momentum fraction  $x_i$  at an energy scale  $Q^2$ , they can be extrapolated to another scale using the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Parisi) evolution equation [19]. The PDF sets used for the simulation of signal

samples in these analyses are provided by the CTEQ/CT group [20, 21]. This set especially incorporates the effects of Tevatron Run I jet production data on the gluon distribution and is therefore expected to describe the mainly gluon based LHC processes realistically. The CT sets additionally include measurements from HERA-1 data, new data on the asymmetry in the rapidity distribution of the charged lepton from W boson decay from CDF, and rapidity distributions of Z bosons from both CDF and DØ. The NNPDF sets [22] are calculated with an approach based on neural network and the newest versions include LHC data as well. An example of the most important parton distributions inside the proton is shown in Fig. 5.2.

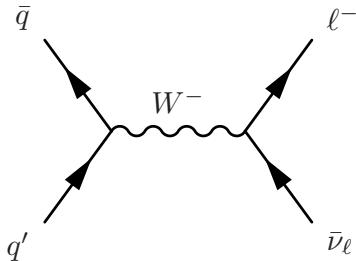
773



**Figure 5.2:** CTEQ6.6 central value parton distribution functions at the typical mass scale of a new diboson resonance ( $Q^2 = (1000 \text{ GeV})^2$ ) for up, down and bottom quarks, and gluons in the proton in double-logarithmic scale.

An accurate description of the process must take into account radiative corrections to the tree-level or LO description of the process of interest. In particular, one has to include the effects of real and virtual higher-order corrections in perturbation theory. This is achieved by computing the matrix element between the initial and final states as the sum of contributions with increasing powers of  $\alpha_S$ . For instance, the LO contribution to the W boson production process can be calculated from the diagram in Fig. 5.3. The diagrams contributing at NLO to this process and corresponding to the real and virtual radiative corrections at the first order are shown in Fig. 5.4.

782

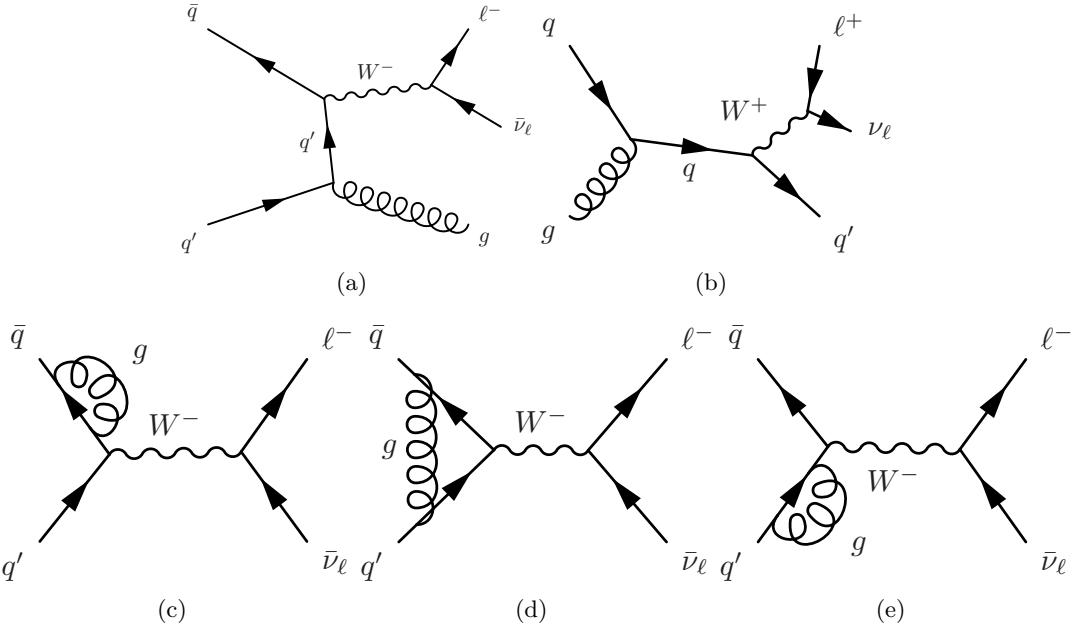


**Figure 5.3:** (top) Feynman diagram contributing to the W boson production at leading order. The charge conjugate production mode is implied. Only the leptonic decay of the W boson is considered.

Perturbative calculations in QCD are limited to processes in which the coupling constant  $\alpha_S$  is small, and by the complexity of higher order calculations preventing their evaluation.

783

784



**Figure 5.4:** Feynman diagrams contributing at next-to-leading order to the W boson production and corresponding to the first order real (top) and virtual (bottom) radiative corrections. The charge conjugate production modes are implied. Only the leptonic decay of the W boson is considered.

Consequently, the current generators are only able to treat a limited number of partons in the final state. Parton showering algorithms extend the fixed order calculations beyond these limiting factors by calculating emissions of additional partons from the in- and outgoing partons of the main interaction. This approach in principle takes into account emissions of an unlimited number of partons, but, as opposed to full higher order calculations, does not take into account loop diagrams. Parton showering algorithms start from the hard process allowing the partons to split (or branch) into pairs of other partons. These again may also branch and so on, so that an event then consists of a large number of elementary particles, including quarks and gluons. The cascade of splittings is stopped once the energy scale reaches values where the coupling constant  $\alpha_S$  becomes large.

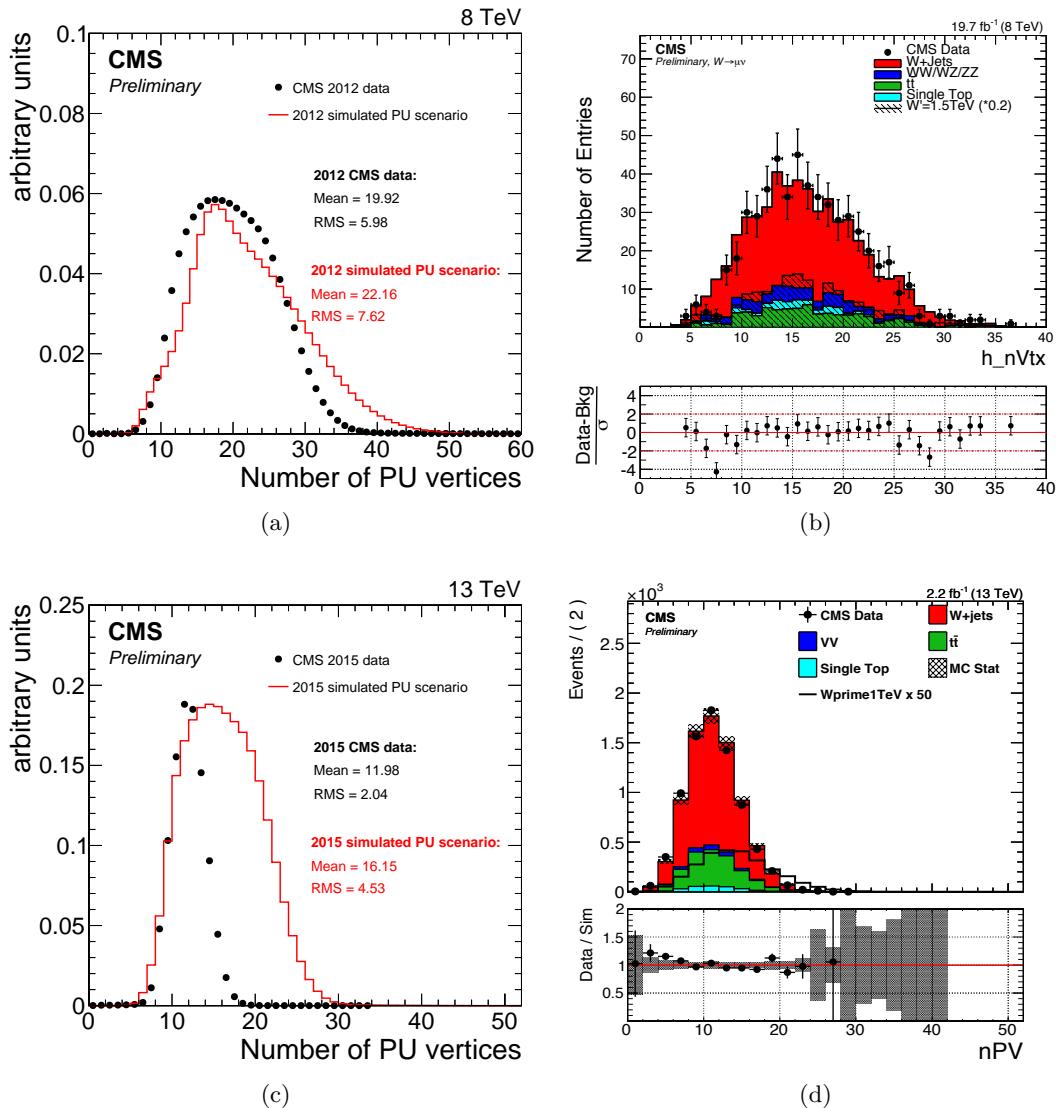
At this stage, quarks and gluons, which carry colour, cannot be considered as free anymore and recombine to form neutral hadrons, through the so called *hadronization* process. The formation of color-neutral hadrons from the colored partons is treated in phenomenological non-perturbative models. Eventually, many short-lived resonances will be present after hadronization which are then decayed.

The showering and hadronization programs often bring along the possibility to add underlying events. The underlying event arises from the colored remains of the protons that did not take part in the hard collisions, the so-called beam remnants. They are usually included in the hadronization process, because they might be colour-connected to the hard subprocess. The produced hadrons will however carry a very small transverse momentum and will be very forward. The probability for colour reconnection to take place between two partons can also be adjusted based on experimental data. It is also possible that more than one parton interact with the other proton. This phenomenon, called multiple parton interaction, and it is usually added to the description of the process.

As last step the pileup is also accounted for. Additional simulated minimum-bias interactions are added to the generated events to match the additional particle production due to

pileup. The exact number of average collisions per bunch crossing in the data is estimated by multiplying the instantaneous luminosity, continuously monitored, by the total inelastic cross section. One can then reconstruct the distribution of the number of pileup interactions in the data for the complete data taking. The corresponding distributions for the 2012 and 2015 data are shown in Figs. 5.5(a) and 5.5(c), respectively, together with the corresponding simulated pileup scenarios. Simulated events are then reweighted such that they match the data distribution. The description of the pileup by the simulation can be verified by counting the number of reconstructed vertices in the event as illustrated in Figs. 5.5(b) and 5.5(d).

820



**Figure 5.5:** Distributions of the estimated average number of pileup collisions in the full data set of pp collisions recorded at  $\sqrt{s} = 8 \text{ TeV}$  in 2012 (a) and at  $\sqrt{s} = 13 \text{ TeV}$  (c), together with the corresponding simulated pileup scenarios. Also shown are the distributions of the number of reconstructed primary vertices in 8 TeV (b) and 13 TeV (d) data (black dots) and in various simulated samples after pileup reweighting, for lepton+jet events.

Currently, NLO available calculations included in MC event generators cover a wide range of physics processes, starting with two particles annihilation to a maximum of five

821

822

final state objects. A popular generator is PYTHIA [23, 24], a general purpose program which, in addition to the hard process, also takes care of the parton showering, the hadronization, and the description of the underlying event. For the matrix element calculation, PYTHIA only considers the leading order hard subprocess (diagram in Fig. 5.3 for the W production case), and higher order effects are added by “evolving” the event using the parton shower. A more accurate approach is followed by MADGRAPH [25] where the hard real radiative corrections are included in the matrix element (Fig. 5.4). This generator is well suited to study processes such as W or Z produced in association with hard jets. Since it does not completely simulate the events, it needs an additional program, typically PYTHIA, to perform the parton shower after the calculation of the matrix element. It has to be noted that matrix element generators as well as shower and hadronization generators are usually treated independently: the matrix element generators compute the hard process at fixed-order and the parton shower processes the soft and collinear emissions. However, this fails to correctly represent higher order processes in which an additional parton is emitted at the hard scale because parts of this process overlap with the soft one. Combining an NLO matrix element program with a parton shower program therefore leads to double-counting of events. The NLO matrix element generators, such as POWHEG [26] and MC@NLO [27], take special care of the merging of soft and collinear emissions and hard ones.

### 5.1.2 CMS detector simulation

For a detailed understanding on how interactions in pp collisions at the LHC are observed by the CMS detector, a dedicated simulation of the whole detector is needed. Both the propagation of particles through the detector material as well as the response of the active detector components and their digital output need to be simulated. The input to the detector simulation are collections of particles produced by MC event generators. The output is the digital signal from all detector components in the same format that is used for real data.

The CMS simulation is based on the GEANT4 [28] toolkit. The program calculates the trajectory of the various particles generated during the collision, simulates their electromagnetic and hadronic interaction with the crossed material and the signal they will produce in the various subdetectors. The detector geometry is given as an input to the program, and to obtain a description as close as possible to the reality, any available information such as the existence of insensitive materials or dead channels and their position, is included. The electronic readout of the hits produced by particles is simulated, taking into account resolution and detector response effects. The same algorithms as for real data are then used to reconstruct the various physical objects (Chapter 6)

## 5.2 Simulated samples

### 5.2.1 Simulation of signal processes

For the 8 TeV data analysis, the signal hypothesis has been simulated at LO accuracy with a W' boson produced via quark-antiquark annihilation and decaying into W and Higgs bosons in the  $\ell\nu q\bar{q}$  final state with  $q = b, c$  or  $g$  and  $\ell = e, \mu$  or  $\tau$ . Resonance masses in the range 0.8–2.5 TeV are considered in this analysis. The events are generated at parton level using a model of a generic narrow spin-1 W' resonance implemented with MADGRAPH. Showering and hadronization are performed using PYTHIA6 using the Z2\* tune [29, 30]. It has been verified that the kinematic distributions obtained with the implementation of the generic model agree with those predicted by implementations of the LH, composite Higgs and HVT

models in MADGRAPH. The resonance width differs in the three models, but in each case it is found to be negligible with respect to the experimental resolution.

The full simulation of the detector has been done privately following the standard CMS procedure described in Section 5.1.2. This emulation has been validated comparing the private production with samples from the MC production campaign carried out centrally for the whole collaboration.

The following parameters are used to compute the cross sections:  $g_V = 3$ ,  $c_H \simeq -1$ , and  $c_F \simeq 1$  in the HVT model B (**FIXME: point to theory**) and  $\cot 2\theta = 2.3$ ,  $\cot \theta = -0.20799$  in the LH model, where  $\theta$  is a mixing angle parameter that determines  $W'$  couplings (**FIXME: point to theory**) such that  $\cot 2\theta$  and  $\cot \theta$  can be directly related to  $c_H$  and  $c_F$ .

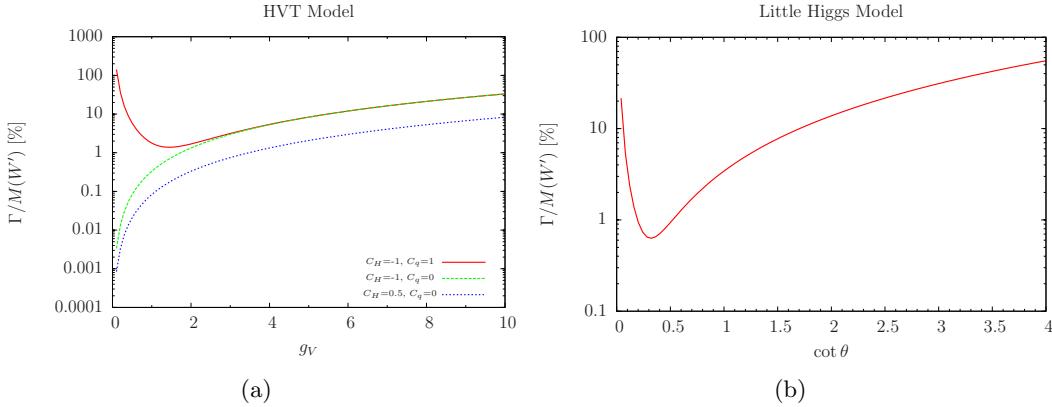
The intrinsic width and cross section for both models are listed in Table 5.1 for the resonance masses considered. The widths for the HVT model B are computed by means of Equation (2.31) in Ref. [31], while the cross sections were obtained using the online tools provided by the authors of Ref. [31]. (**FIXME: point to theory**)

**Table 5.1:** Intrinsic total widths ( $\Gamma$ ) and cross sections for  $\sqrt{s} = 8$  TeV ( $\sigma$ ) for the LH model and HVT model B for different masses of a resonance  $W'$  decaying to  $WH$ . The  $WH \rightarrow \ell\nu b\bar{b}$  branching fraction is not included in the calculation.

Resonance mass [TeV]	LH model		HVT model B	
	$\Gamma$ [GeV]	$\sigma$ [pb]	$\Gamma$ [GeV]	$\sigma$ [pb]
0.8	7.22	$5.09 \times 10^{-1}$	24.1	$3.37 \times 10^{-1}$
0.9	8.12	$3.03 \times 10^{-1}$	27.1	$2.48 \times 10^{-1}$
1.0	9.02	$1.87 \times 10^{-1}$	30.1	$1.71 \times 10^{-1}$
1.1	9.92	$1.18 \times 10^{-1}$	33.1	$1.16 \times 10^{-1}$
1.2	10.8	$7.65 \times 10^{-2}$	36.1	$8.05 \times 10^{-2}$
1.3	11.7	$5.06 \times 10^{-2}$	39.1	$5.59 \times 10^{-2}$
1.4	12.6	$3.39 \times 10^{-2}$	42.2	$3.88 \times 10^{-2}$
1.5	13.5	$2.29 \times 10^{-2}$	45.2	$2.51 \times 10^{-2}$
1.6	14.4	$1.56 \times 10^{-2}$	48.2	$1.87 \times 10^{-2}$
1.7	15.3	$1.08 \times 10^{-2}$	51.2	$1.30 \times 10^{-2}$
1.8	16.2	$7.43 \times 10^{-3}$	54.2	$9.03 \times 10^{-3}$
1.9	17.1	$5.17 \times 10^{-3}$	57.2	$6.27 \times 10^{-3}$
2.0	18.0	$3.61 \times 10^{-3}$	60.2	$4.25 \times 10^{-3}$
2.1	19.0	$2.53 \times 10^{-3}$	63.2	$3.02 \times 10^{-3}$
2.2	19.8	$1.76 \times 10^{-3}$	66.2	$2.10 \times 10^{-3}$
2.3	20.8	$1.24 \times 10^{-3}$	69.2	$1.46 \times 10^{-3}$
2.4	21.6	$8.67 \times 10^{-4}$	72.2	$1.01 \times 10^{-3}$
2.5	22.6	$6.07 \times 10^{-4}$	75.3	$7.31 \times 10^{-4}$

Figure 5.6 shows the ratio between the resonance's natural width and mass for a  $W'$  in the LH and the HVT model B. The width is less than 5% for the following parameter values:  $0.95 < g_V < 3.76$ ,  $c_H = -1$ , and  $c_F = 1$ ;  $g_V < 3.9$ ,  $c_H = -1$ , and  $c_F = 0$ ; or  $g_V < 7.8$ ,  $c_H = 0.5$ , and  $c_F = 0$ . The widths for the LH model have been computed by means of Eq. (15) in Ref. [32], and they are less than 5% for values of  $0.084 < |\cot \theta| < 1.21$ . Hence, in both models the resonance's natural width can be considered to be negligible compared to the experimental resolution.

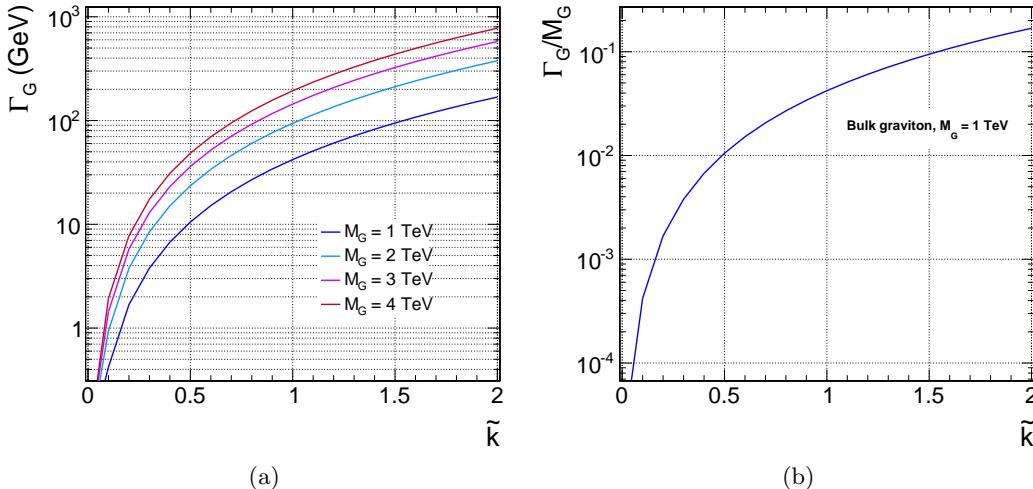
For the 13 TeV data analysis, the bulk graviton model and HVT models are used as benchmark signal processes. In these models, a resonance is simulated which decays only to pairs of vector gauge bosons in the  $\ell\nu q\bar{q}^{(\prime)}$  final state, with  $\ell = e, \mu$ , and  $\tau$ . The vector gauge



**Figure 5.6:** Ratio between the resonance's natural width and mass for a  $W'$  in the LH and the HVT model B.

bosons are produced with a longitudinal polarization in more than 99% of the cases. For each resonance hypothesis, masses are considered in the range 0.6 to 4.0 TeV. Simulated signal events are generated at LO accuracy with MADGRAPH5\_AMC@NLO with a relative resonance width of 0.1%.

The natural width of a bulk graviton as a function of the curvature parameter  $\tilde{k}$  and for different mass hypotheses is shown in Fig. 5.7. For cases in which  $\tilde{k} \leq 0.5$  the relative width of the graviton resonance ( $\Gamma_G/M_G$ ) is predicted to be below 1%. Hence, it can be neglected when compared to the detector resolution over the whole explored mass range.

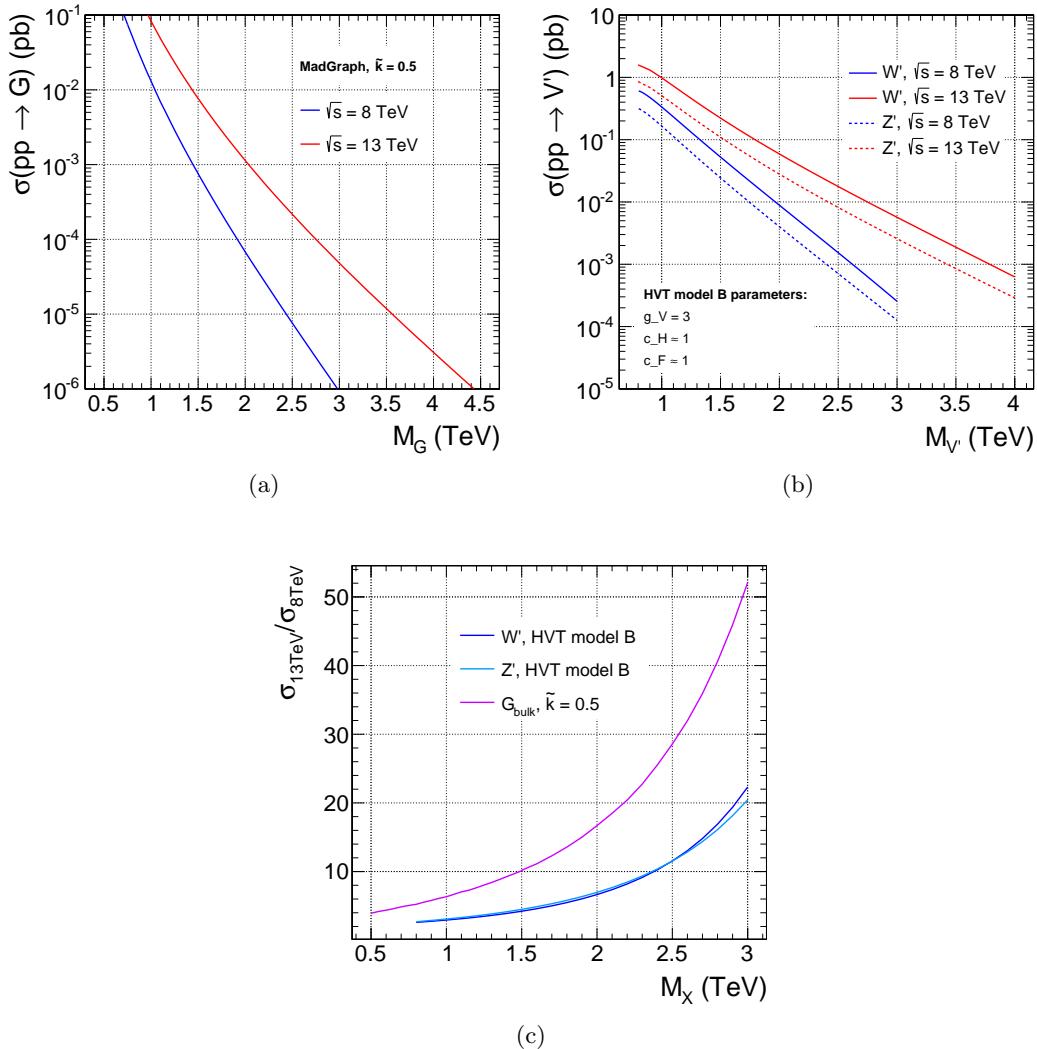


**Figure 5.7:** (a) Natural width of a bulk graviton as a function of the coupling constant  $\tilde{k}$  and for various mass hypotheses. (b) The same dependence is expressed as relative fraction of the signal width with respect to a reference graviton mass of 1 TeV.

Figure 5.8 compare the production cross sections  $\sigma(pp) \rightarrow X$  of the resonance for  $\sqrt{s} = 8$  and 13 TeV, for the bulk graviton with  $\tilde{k} = 0.5$ , and  $W'$  and  $Z'$  in the HVT model B, as a function of the resonance mass. Cross sections for the bulk graviton model are computed with MADGRAPH with the model used for the even generation, while values for the HVT model B are obtained using the online tools provided by the authors of Ref. [31] using the

same parameters as for the 8 TeV data analysis.

For a resonance mass of 2 TeV, the production rates at for  $\sqrt{s} = 13$  TeV are expected to increase of a factor  $\approx 17$  for a resonance produced via gluon-gluon fusion such as the graviton; a smaller factor of  $\approx 7$  is expected instead for resonances produced via quark-antiquark annihilation such as  $W'$  and  $Z'$ .



**Figure 5.8:** Comparison of the production cross sections of the resonance for  $\sqrt{s} = 8$  and 13 TeV for the bulk graviton (a), and  $W'$  and  $Z'$  in the HVT model B (b), as a function of the resonance mass. (c) Ratio of the production cross sections for  $\sqrt{s} = 8$  and 13 TeV for all models.

### 5.2.2 Simulation of background processes

For the 8 TeV data analysis, the background is modelled using the MADGRAPH5 v1.3.30 event generator to simulate the production of  $W$  boson in association with jets at LO, the POWHEG 1.0 r1380 package to generate  $t\bar{t}$  and single top quark events at NLO accuracy, and PYTHIA6 v424 for SM diboson ( $WW$ ,  $WZ$ , and  $ZZ$ ) production at LO. All simulated event samples are generated using the CTEQ6L1 PDF set with  $\alpha_S$  also at LO, except for the POWHEG  $t\bar{t}$  sample, for which the CT10 NNLO PDF set is used. All the samples are then processed

further by PYTHIA6, using the Z2\* tune for simulation of parton showering and subsequent hadronization, and for simulation of the underlying event. All simulated background samples are normalized to the integrated luminosity of the recorded data, using inclusive cross sections determined at NLO, or NNLO when available, calculated with the cross section integrators MCFM [33–36] and FEWZ [37], except for the  $t\bar{t}$  sample, for which TOP++ [38] is used. The NNLO cross section for the W+jets process is obtained by rescaling the LO value given by the generator to the NNLO cross section derived from the inclusive production by means of a flat  $k$ -factor = NNLO/LO = 1.3. The simulated samples used in the 8 TeV data analysis described in this work are listed in Table 5.2 together with the corresponding cross sections.

926

**Table 5.2:** Summary of the MC generated samples for background processes used for the 8 TeV data analysis. The cross sections used to normalize the samples are also indicated.

Process	Cross section (pb)	Generator	PDF set
W+jets, $W \rightarrow \ell\nu$ , $p_T^W > 180$ GeV	29.0 (NNLO)	MADGRAPH	CTEQ6L1
$t\bar{t}$ (inclusive)	252.9 (NNLO+NNLL)	POWHEG	CT10
single t quark (t-channel, inclusive)	54.9 (NNLO)	POWHEG	CTEQ6L1
single $\bar{t}$ quark (t-channel, inclusive)	29.7 (NNLO)	POWHEG	CTEQ6L1
single t quark (tW-channel, inclusive)	11.2 (NNLO)	POWHEG	CTEQ6L1
single $\bar{t}$ quark (tW-channel, inclusive)	11.2 (NNLO)	POWHEG	CTEQ6L1
single t quark (s-channel, inclusive)	3.8 (NNLO)	POWHEG	CTEQ6L1
single $\bar{t}$ quark (s-channel, inclusive)	1.8 (NNLO)	POWHEG	CTEQ6L1
WW (inclusive)	54.8 (NLO)	PYTHIA6	CTEQ6L1
WZ (inclusive)	33.2 (NLO)	PYTHIA6	CTEQ6L1
ZZ (inclusive)	8.1 (NLO)	PYTHIA6	CTEQ6L1

927 For the 13 TeV analysis, the W+jets SM process is simulated with MADGRAPH5\_AMC@NLO  
 928 at LO accuracy. The  $t\bar{t}$ , single top quark and diboson events are generated with both POWHEG  
 929 and MADGRAPH5\_AMC@NLO at NLO accuracy. Parton showering and hadronization are  
 930 implemented through PYTHIA8 using the CUETP8M1 tune [39, 40]. The NNPDF 3.0 PDFs  
 931 with  $\alpha_S$  at NLO, are used for all simulated samples. The simulated background is normalized  
 932 using inclusive cross sections calculated at NLO, or NNLO order in QCD where available, using  
 933 MCFM and FEWZ, except for the  $t\bar{t}$  sample, for which TOP++ [38] is used. A  $k$ -factor = 1.21  
 934 is used to rescale the W+jets simulation to the NNLO cross section.

935 The simulated samples used in the 13 TeV data analysis described in this work are listed  
 936 in Table 5.3 together with the corresponding cross sections.

### 937 5.3 Data sets

938 Two independent data sets are analyzed in this work to search for diboson resonances decaying  
 939 to two different final states.

940

941 The analysis focused on the  $\ell\nu b\bar{b}$  final state is performed with the complete set of data  
 942 recorded in 2012 by the CMS detector and corresponding to an integrated luminosity of  
 943  $19.7 \text{ fb}^{-1}$  of pp collisions at  $\sqrt{s} = 8$  TeV. The recorded events are divided into 4 run periods  
 944 (runs A, B, C, D).

945

946 The second analysis described in this work is focused on the  $\ell\nu q\bar{q}'$  final state and it is  
 947 performed with only the largest part of the full set of data recorded in 2015 by the CMS

**Table 5.3:** Summary of the MC generated samples for background processes used for the 13 TeV data analysis. The cross sections used to normalize the simulated events are also indicated. The NNPDF 3.0 PDFs are used for all simulated samples

Process	Cross section (pb)	Generator
W+jets, $W \rightarrow \ell\nu$ , $100 < H_T < 200$ GeV	1627.5 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$ , $200 < H_T < 400$ GeV	435.2 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$ , $400 < H_T < 600$ GeV	59.2 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$ , $600 < H_T < 800$ GeV	14.6 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$ , $800 < H_T < 1200$ GeV	6.7 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$ , $1200 < H_T < 2500$ GeV	1.6 (NNLO)	MADGRAPH5_AMC@NLO
W+jets, $W \rightarrow \ell\nu$ , $H_T > 2500$ GeV	0.04 (NNLO)	MADGRAPH5_AMC@NLO
$t\bar{t}$ (inclusive)	831.8 (NNLO+NNLL)	POWHEG
single t quark (t-channel), $W \rightarrow \ell\nu$	44.5 (NNLO)	POWHEG
single $\bar{t}$ quark (t-channel), $W \rightarrow \ell\nu$	26.5 (NNLO)	POWHEG
single t quark (tW-channel, inclusive)	35.9 (NNLO)	POWHEG
single $\bar{t}$ quark (tW-channel, inclusive)	35.9 (NNLO)	POWHEG
single $t+\bar{t}$ quark (s-channel), $W \rightarrow \ell\nu$	3.7 (NNLO)	MADGRAPH5_AMC@NLO
$WW \rightarrow \ell\nu q\bar{q}'$	50.0 (NNLO)	POWHEG
$WZ \rightarrow \ell\nu q\bar{q}$	10.7 (NLO)	MADGRAPH5_AMC@NLO
$ZZ \rightarrow \ell\ell q\bar{q}$	3.22 (NLO)	MADGRAPH5_AMC@NLO

948 detector corresponding to an integrated luminosity of  $2.3 \text{ fb}^{-1}$  of pp collisions at  $\sqrt{s} = 13 \text{ TeV}$ .  
 949 During 2015, there have been three running periods labeled from B to D. In fact, after a short  
 950 period of 50 ns operation (period B), the machine collected data with a bunch spacing of  
 951 25 ns (period C and D). However, since the first two periods only add a tiny contribution to  
 952 the total integrated luminosity of 2015 collisions, the decision was made to base the analysis  
 953 on period D only, corresponding to the largest data set.

954  
 955 All events that are accepted by a specific set of high level triggers enter one specific data  
 956 set, so that the choice of a trigger for the analysis defines which data set has to be used. As  
 957 discussed in the next chapter (Sections 6.3.2 and 6.2.2), events are collected with a trigger  
 958 requiring either one muon or one electron passing given  $p_T$  and  $\eta$  selections. Hence, the data  
 959 sets used in these analyses are the so called "SingleMuon" and "SingleElectron" primary data  
 960 sets listed in Table 5.4.

961 Even though stable run periods are chosen for the analyses, not all runs can be used.  
 962 This analysis requires the whole detector to be functional since the objects employed are  
 963 reconstructed from all parts of the detector as described in the next chapter. Therefore, only  
 964 data-taking runs and luminosity blocks during which the detector was in a state sufficiently  
 965 good for further analysis are used.

**Table 5.4:** Data sets used in this analysis.

$\sqrt{s}$	Year	Data set	Run period	Run range	$\mathcal{L} [\text{pb}]^{-1}$
8 TeV	2012	SingleMuon	A	190456–193621	889.362
			B	193833–196531	4424
			C	198022–203742	7144
			D	203777–208686	7307
		Total	190456–208686	19764	
	2015	SingleElectron	A	190456–193621	889.362
			B	193833–196531	4422
			C	198022–203742	7080
			D	203777–208686	7314
		Total	190456–208686	19705	
13 TeV	2015	SingleMuon	D	256630–260627	2320
		SingleElectron	D	256630–260627	2320

# Object and event reconstruction

969 In the pp collisions at the LHC a large number of particles are produced which must be  
 970 properly reconstructed and identified. These particles travel through the CMS detector and  
 971 they are classified in objects depending on their specific signature in each subdetector. This  
 972 chapter covers the reconstruction of physics objects that are needed for the identification of  
 973 signal events in the lepton plus jet event topology described in Chapter 4.

974 The measurement of a track in the tracker detector for charged particles and the recon-  
 975 struction of the primary vertices represent key aspects of the reconstruction of the various  
 976 objects and are detailed in Section 6.1. In this analysis  $\tau$  leptons are reconstructed as  
 977 electrons (Section 6.2) or muons (Section 6.3) and accounted to the respective channel if  
 978 they decay leptonically, or as jets (Section 6.4) if they decay hadronically. However, only the  
 979 leptonic decay mode contributes to the analysis since at least one muon or electron has to be  
 980 reconstructed in the event. In addition to leptons and jets, the last type of particle present  
 981 in the final state is the neutrino, whose presence can be inferred from an imbalance of the  
 982 transverse energy (Section 6.5). The identified lepton and the missing transverse energy in  
 983 the event are associated with the  $W \rightarrow \ell\nu$  candidate which is entirely reconstructed through  
 984 the algorithm described in Section 6.6.

985 **6.1 Tracks and vertices**

986 The reconstruction of tracks of charged particles allows for their momentum measurement  
 987 and aids in particle identification as described in the following. The reconstruction of the  
 988 tracks' vertices is important to distinguish the primary interaction, i.e. the hard interaction,  
 989 from additional interactions that might take place in the event and also for the identification  
 990 of secondary vertices of jets that contain c or b quarks called c-/b-tagging (see Sec. 6.4.3).

991 **6.1.1 Track reconstruction**

992 The track reconstruction at CMS [41] is based on information coming from the silicon tracker  
 993 system. A charged particle passing through a tracker layer can in general induce a signal  
 994 in more than one pixel or more than one strip. The first step of the tracking procedure is  
 995 the assembly of nearby tracker channels into one hit cluster. The particle position and its  
 996 uncertainty is then inferred from the relative signal amplitudes in each channel.

997 Because of the magnetic field, charged particles travel through the tracking detectors on  
 998 a helix trajectory which is described by 5 parameters: the curvature  $k$ , the track azimuthal  
 999 angle  $\phi$  and polar angle  $\theta$ , the signed transverse impact parameter  $d_0$  and the longitudinal  
 1000 impact parameter  $z_0$ . The transverse (longitudinal) impact parameter of a track is defined as  
 1001 the transverse (longitudinal) distance of closest approach of the track to the primary vertex.

1002 The trajectories of charged particles are reconstructed through a iterative procedure  
 1003 consisting in multiple iterations of the *Combinatorial Track Finder algorithm* (CTF) [42],  
 1004 which uses the reconstructed hits in the silicon detectors to determine the track parameters.  
 1005 In the first iterations the algorithm searches for tracks of relative large  $p_T$  and produced near  
 1006 the interaction region. Then, hits associated to high quality tracks are iteratively removed

from the input list to reduce the combinatorial complexity of the next iterations, and to allow the more difficult reconstruction of low  $p_T$  or displaced tracks. Each iteration of the CTF algorithm is made of four steps: track seeding, track finding and track fitting.

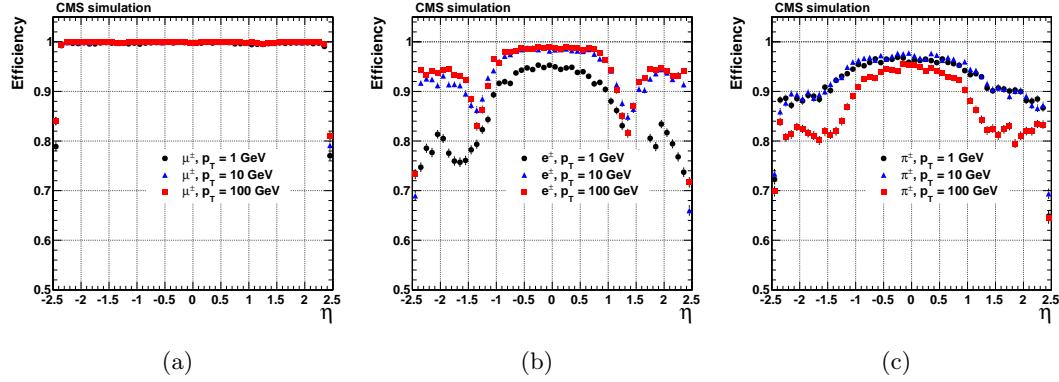
In the first step, a first estimate of the helix parameters and of its covariance matrix is provided using only pairs or triplets of hits compatible with the hypothesis of a track coming from the pp interaction region. Track candidates are best seeded from hits in the pixel detector because of the low occupancy, high efficiency and unambiguous 3-dimensional position information.

The track finding stage associates new hits in the next tracker layers to the trajectory obtained from seeds using a standard Kalman Filter (KF) pattern recognition approach [43,44], which takes into account the effect of multiple scattering in the tracker layers. The current trajectory is extrapolated to the next tracker layer and compatible hits are assigned to the track on the basis of the  $\chi^2$  between the predicted and measured positions. In case multiple compatible hits are found when extrapolating the helix to a single layer, the algorithm creates one trajectory candidate for each hit and they are propagated independently. Furthermore, in order to take into account possible inefficiencies, one additional candidate is created without including any hit information. A quality index is assigned to the tracks, based on the  $\chi^2$ , the number of missing hits, and how compatible they are with originating from a primary interaction vertex. Only the best quality tracks are kept for further propagation and ambiguities are resolved between tracks during and after track finding. In case two tracks share more than 50% of their hits, the lower quality track is discarded. The fake rate, defined as the fraction of reconstructed tracks not associated with a charged particle, is substantially reduced by these quality requirements.

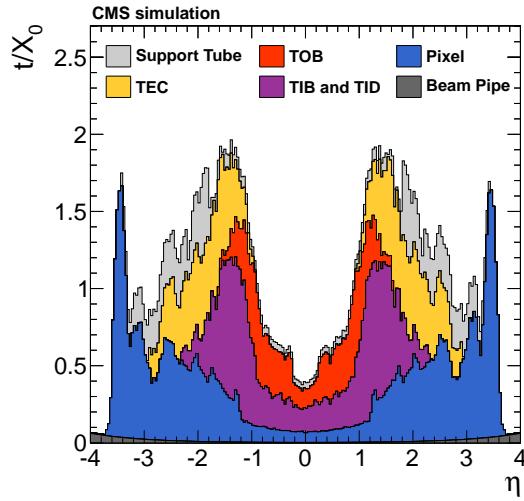
For each trajectory the finding stage results in an estimate of the track parameters. However, since the full information is only available at the last hit and constraints applied during trajectory building can bias the estimate of the track parameters, all valid tracks are refitted using the KF to determine the most accurate estimate of the helix parameters. The usual fit starting from the interaction point to the end of the tracker is complemented with a second fit running backward from the outermost tracker layer to the interaction point. This approach is found to improve the accuracy of the  $p_T$  and impact parameter measurement by 0.5% and 1%, respectively.

The performance of the track reconstruction is shown in Fig. 6.1 for simulated muons, electrons and pions. For isolated muons with  $1 < p_T < 100 \text{ GeV}$ , the track reconstruction efficiency is  $> 99\%$  over the full  $\eta$ -range of tracker acceptance, and does not depend on  $p_T$  (Fig. 6.1(a)). The fake rate is completely negligible. For pions and electrons the efficiency is in general lower along with a higher fake rate because of interactions with the material in the tracker. The material budget of the CMS tracker in units of radiation length is presented in Fig. 6.2.

In Fig. 6.3(a) the transverse momentum resolution for muon tracks with  $p_T = 1, 10$ , and  $100 \text{ GeV}$  is shown. At high transverse momentum ( $100 \text{ GeV}$ ), the resolution is 2–3% up to  $|\eta| = 1.6$ . The material of the tracker accounts for 20–30% of the transverse momentum resolution. At lower momenta, the resolution is dominated by multiple scattering and its distribution reflects the amount of material traversed by the track. The resolutions of the track impact parameter in the transverse and longitudinal plane are also shown in Fig. 6.3. At high momentum the transverse impact parameter resolution is fairly constant and is dominated by the hit resolution in the first pixel layer. It is progressively degraded by multiple scattering at lower momenta. The same applies to the longitudinal impact parameter resolution. The improvement of the  $z_0$  resolution up to  $|\eta| = 0.5$  is due to the charge sharing



**Figure 6.1:** Track reconstruction efficiency for simulated muons (a), electrons (b), and pions (c) passing the high-purity quality requirements as a function of  $\eta$  and for  $p_T = 1, 10$ , and  $100$  GeV [41].



**Figure 6.2:** Material budget of the CMS tracker in units of radiation length  $X_0$  as a function of pseudorapidity divided into the contributions of the different subdetectors [41].

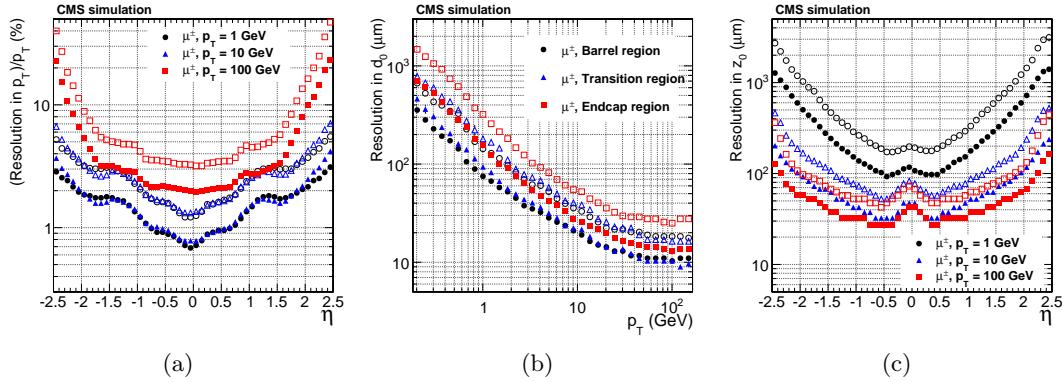
1056 effects among neighboring pixels.

### 1057 6.1.2 Vertex reconstruction

1058 The identification of vertices is essential to distinguish the primary vertex associated with  
 1059 the hard interaction from additional pileup vertices that might be present in the event. This  
 1060 became even more important at the highest LHC luminosity reached at the end of 2016 where  
 1061 on average up to **FIXME: number?** pp interactions took place simultaneously.

1062

1063 In the primary-vertex reconstruction [45], the measurements of the location and uncertainty  
 1064 of an interaction vertex are computed from a given set of reconstructed tracks. The prompt  
 1065 tracks originating from the primary interaction region are selected based on the transverse  
 1066 impact parameter significance with respect to the beam line, number of strip and pixel hits,  
 1067 and the normalized track  $\chi^2$  from a fit to the trajectory. The selected tracks are then clustered  
 1068 on the basis of their  $z$ -coordinates at their point of closest approach to the center of the  
 1069 beam spot using a *deterministic annealing* (DA) algorithm [46]. This clustering allows for the



**Figure 6.3:** Resolution of track transverse momentum (a), transverse (b) and longitudinal (c) impact parameter for simulated muons passing the high-purity quality requirements as a function of  $\eta$  and for  $p_T = 1, 10$ , and  $100$  GeV [41].

reconstruction of any number of pp interactions in the same LHC bunch crossing. Vertices are resolved with separations of about 1 mm, appropriate for a multiplicity of interactions per bunch crossing up to 20, as the longitudinal RMS spread of the luminous region is about 6 cm.

After identifying candidate vertices based on the DA clustering in  $z$ , those candidates containing at least two tracks are then fitted using an *adaptive vertex fitter* [47], to compute the best estimate of vertex parameters, including its  $x$ ,  $y$ , and  $z$  position, and covariance matrix. This algorithm addresses the issue of secondaries and fake tracks in the cluster by iteratively down-weighting the tracks which are not compatible with the fitted common vertex. The primary vertex, where the hard process of interest takes place, is chosen as the vertex with the highest sum of  $p_T^2$  of the clustered tracks.

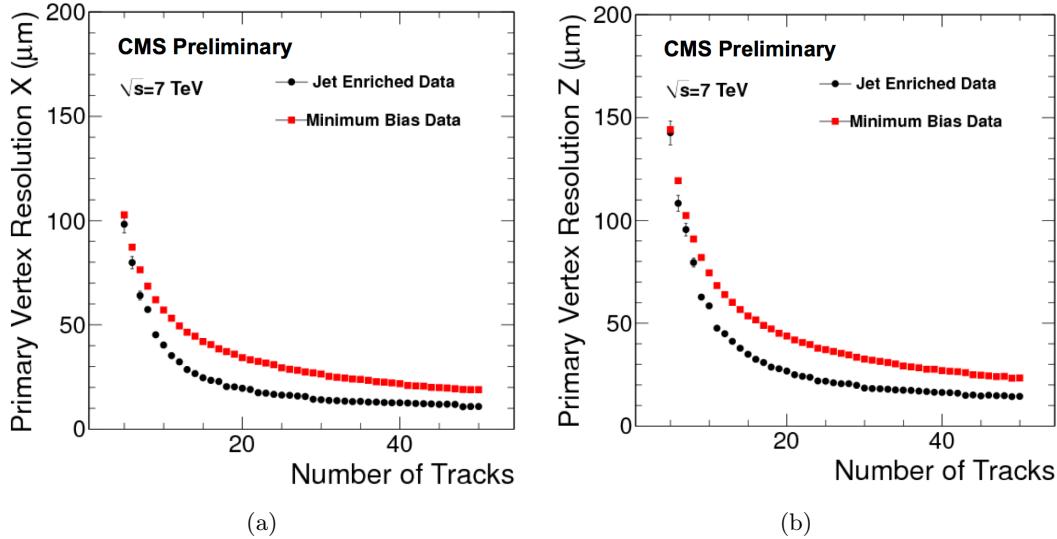
The primary vertex spatial resolution depends on the event topology and on the number of tracks related to the vertex, as shown in Fig. 6.4. For minimum-bias events, the resolutions in  $x$  and  $z$  are, respectively, less than  $20\ \mu\text{m}$  and  $25\ \mu\text{m}$ , for primary vertices reconstructed using at least 50 tracks. The resolution is better for the jet-enriched sample where tracks have significantly higher mean  $p_T$  resulting in higher resolution in the track impact parameter, and consequently better vertex resolution. For these events, the resolutions approach  $10\ \mu\text{m}$  in  $x$  and  $12\ \mu\text{m}$  in  $z$  for primary vertices using at least 50 tracks.

In the analysis described in this work, all events are required to have at least one primary vertex reconstructed within a 24 cm window along the beam axis, with a transverse distance from the nominal pp interaction region of less than 2 cm.

## 6.2 Electrons

### 6.2.1 Electron reconstruction

The electron reconstruction in CMS [49] is based on the association of an energy deposit in the ECAL with a track reconstructed in the silicon tracker system. Electrons lose energy primarily through bremsstrahlung when interacting with the tracker layers, and consequently they suffer from large energy losses. Given the non-Gaussian properties of the energy loss distributions, the standard track reconstruction algorithm based on the KF is not appropriate



**Figure 6.4:** Primary-vertex resolution in  $x$  (a) and  $z$  (b) as a function of the number of tracks at the fitted vertex, for two kinds of events with different average track  $p_T$  values. The results in  $y$  are almost identical to the one in  $x$  [48].

and leads in general to a reduced hit-collection efficiency, as well as to a poor estimation of track parameters. A better performance for electron reconstruction is achieved by using dedicated techniques that make use of information, not only from the tracker, but also from the ECAL, as described in the following.

The electron reconstruction starts by searching for clusters of energy in the ECAL. As the electrons are degraded in energy, the effect of the magnetic field is to enhance the bending of their trajectories, resulting in a spread of irradiated photons along the  $\phi$  coordinate. To recover this radiated energy, ECAL superclusters are formed, by merging clusters of similar  $\eta$  over some range of  $\phi$ . Because of the different geometry of the detector in barrel and endcap, different clustering algorithms are used in different regions.

For the electron track reconstruction two approaches are used. In the first one, referred to as “ECAL driven”, the supercluster energy and position, and the assumption that the electron originated near the center of the beam spot, are used to extrapolate the electron trajectory in the tracker. Tracker seeds compatible with the predicted trajectory are sought in the first or second layer of the pixel detector (and also in the TEC to improve efficiency in the forward region). This method is designed for isolated electrons with  $p_T > 5 \text{ GeV}$ .

A second approach, referred to as “tracker driven”, complements the electron track reconstruction, especially for low- $p_T$  or non isolated electrons, as well as for electrons in the barrel-endcap transition region. This method is developed as part of the particle-flow (PF) reconstruction algorithm [50, 51] described in Section 6.4.2. It takes the standard track collection reconstructed with the KF algorithm and attempts to identify a subset of these tracks that are compatible with being electrons. Electrons that suffer only little bremsstrahlung loss can be identified by searching for tracks extrapolated to the ECAL that pass close to an ECAL PF cluster. Electrons that suffer large bremsstrahlung loss can be identified by the fact that the fitted track will often have poor  $\chi^2$  or few associated hits. The track seeds originally used to generate these electron-like tracks are retained.

The seed collections obtained by using these two methods are merged, and used to initiate electron track finding. This procedure is similar to that used in standard tracking, except

that the  $\chi^2$  threshold, used by the KF to decide whether a hit is compatible with a trajectory, is weakened. This is to accommodate tracks that deviate from their expected trajectory because of bremsstrahlung.

To obtain the best estimate of the track parameters, the final track fit is performed using a modified version of the KF method, called the Gaussian Sum Filter (GSF) [52]. The fractional energy loss of an electron, as it traverses a layer of material, follows a Bethe–Heitler distribution. This distribution is non-Gaussian, making it unsuitable for use in a conventional KF algorithm. The GSF technique solves this by approximating the Bethe–Heitler energy-loss distribution as the sum of several Gaussian functions. This method is then a generalization of the KF where the trajectory in each tracker layer is described by a weighted sum of KF components for which the energy loss follows a Gaussian law with a given width. The propagation of each component is done separately from one layer to another and the weights are then updated given the measurement in the new site. The allowed window to search for a hit in the next tracker layer is larger than for the usual KF track. This procedure is iterated until the last tracker layer, unless no hit is found in two subsequent layers. A minimum of five hits is finally required to create a track. A GSF electron candidate is finally built by associating an ECAL supercluster with a GSF track with compatible  $\eta$  and  $\phi$  positions.

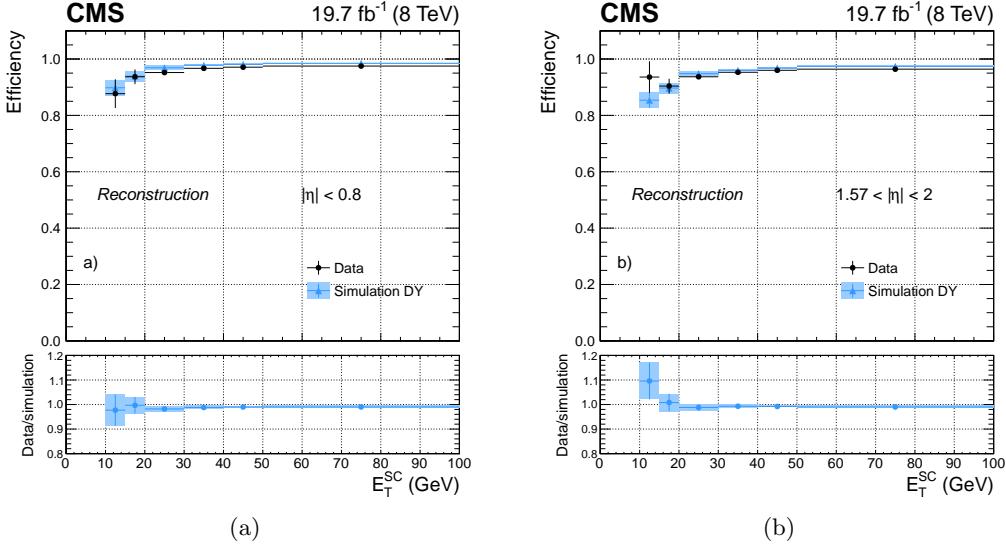
The electron transverse energy  $E_T$  is equal to the transverse energy of the correspondent ECAL energy deposit (or supercluster)  $E_T^{SC}$ , and defined as  $E_T = E \sin \theta$ , where  $\theta$  is the polar angle of the supercluster (ST) relative to the beam axis, and  $E$  the energy measured in the supercluster.

The performance of the GSF electron reconstruction are studied using a “tag-and-probe” (T&P) method [53]. The method uses a known SM resonance mass and decay (e.g.  $Z \rightarrow e^+ e^-$ ) to select particles of the desired type and probe the efficiency of a particular selection criterion on those particles. In general the “tag” is an object that passes a set of very tight selection criteria designed to isolate the required particle type (in this case an electron, though the method is not strictly limited to this case). A generic set of the desired particle type (i.e. with potentially very loose selection criteria) known as “probes”, is selected by pairing these objects with tags such that the invariant mass of the combination is consistent with the mass of the resonance. Combinatorial backgrounds are usually eliminated through a variety of background subtraction methods. The definition of the probe object depends on the specifics of the selection criterion being examined. The efficiency itself is measured by counting the number of “probe” particles that pass the desired selection criteria. It is found that the estimated efficiencies are almost insensitive to any specific definition of the tag. The GSF electron reconstruction efficiency measured with this method is above 95% for electrons in the ECAL barrel with  $E_T > 35$  GeV, as shown in Fig. 6.5(a). Slightly lower efficiencies are obtained for electrons reconstructed in the ECAL endcaps (Fig. 6.5(b)). A good agreement is found between data and simulation, resulting in scale factors consistent with unity almost in the entire range. The performance are presented here for the electron reconstruction in Run 1 but similar results are obtained in CMS for Run 2.

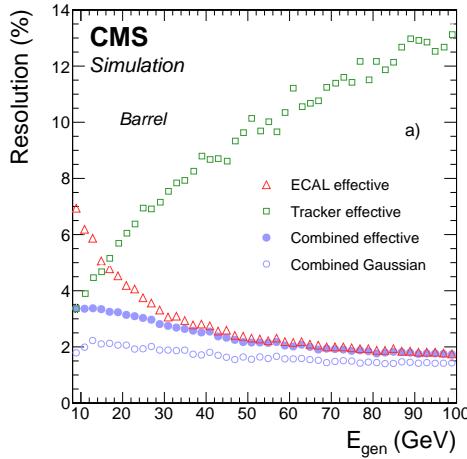
Once a GSF electron candidate is reconstructed, the energy measurement provided by electromagnetic calorimeter can be combined with the tracker momentum measurement to improve the estimate of electrons with energies below 35 GeV as shown in Fig. 6.6. At energies above 35 GeV however, the momentum measurement is completely driven by the supercluster.

### 6.2.2 Electron trigger

As explained in Section 3.2.4, the events of interest for physics analyses are selected by the trigger system in two steps, namely, the L1 and HLT. At the L1, where the tracker information



**Figure 6.5:** Electron reconstruction efficiency measured in dielectron events in data (dots) and Drell-Yan simulation (triangles), as a function of the  $E_T$  for electrons reconstructed in the ECAL barrel (a) and endcaps (b). The bottom panels show the corresponding data-to-simulation scale factors [54].



**Figure 6.6:** Expected resolution in  $E_T$  for isolated electrons in the ECAL barrel as a function of the electron generated energy, obtained from the ECAL, the tracker and the combined estimates [54].

is not available, electrons and photons are indistinguishable and based on calorimeter trigger towers, consisting, in the barrel, of a  $5 \times 5$  matrix of ECAL crystals and the corresponding HCAL tower, while a more complex definition of the tower is used in the endcaps. A L1 candidate is formed combining the highest-energy central trigger tower together with its next-highest adjacent tower. At this stage, the trigger choice is based on the energy distribution among the central and neighbouring towers, on the amount of energy in the HCAL downstream the central tower, and on the  $E_T$  of the  $e/\gamma$  candidate. Events passing L1 are then filtered by the HLT. Here, the pixel tracker information is used to separate electrons from photons. The starting point of any electron HLT selection consists of building

1186 a supercluster and a trajectory as described in Section 6.2.1. Many different triggers involving  
 1187 electrons are designed at the HLT level and various additional identification and isolation  
 1188 requirements on the electrons are made for each of them. They consist of conditions on:

- 1189 • transverse profile of the cluster of energy in the ECAL;
- 1190 • the amount of energy in the HCAL downstream the ECAL cluster;
- 1191 • the existence of a KF or GSF track matching the supercluster position;
- 1192 • quality of association between the track and the ECAL cluster;
- 1193 • activity in the ECAL, HCAL, or tracker around the candidate.

1194 The conditions used and their severity depend on the number of electrons requested by  
 1195 the trigger and their transverse energy threshold, each trigger being designed to have a rate  
 1196 of accepting events of 50 Hz or less. Practically, all the HLT steps and criteria involving  
 1197 only calorimeters information are done first, while the time consuming steps involving track  
 1198 reconstruction are only performed at the end for events passing the previous criteria. The  
 1199 L1 and HLT triggers used to collect the data analyzed in this thesis are listed in Tables 6.1  
 1200 and 6.2 for the 8 and 13 TeV data sets, respectively. The tables also detail the conditions  
 1201 imposed on several variables described in Section 6.2.3. Figure 6.7 shows the L1 trigger  
 1202 efficiencies for different  $E_T$  thresholds as a function of the electron  $E_T$ . The curves exhibit  
 1203 the typical turn on behaviour in correspondence of the imposed  $E_T$  threshold.

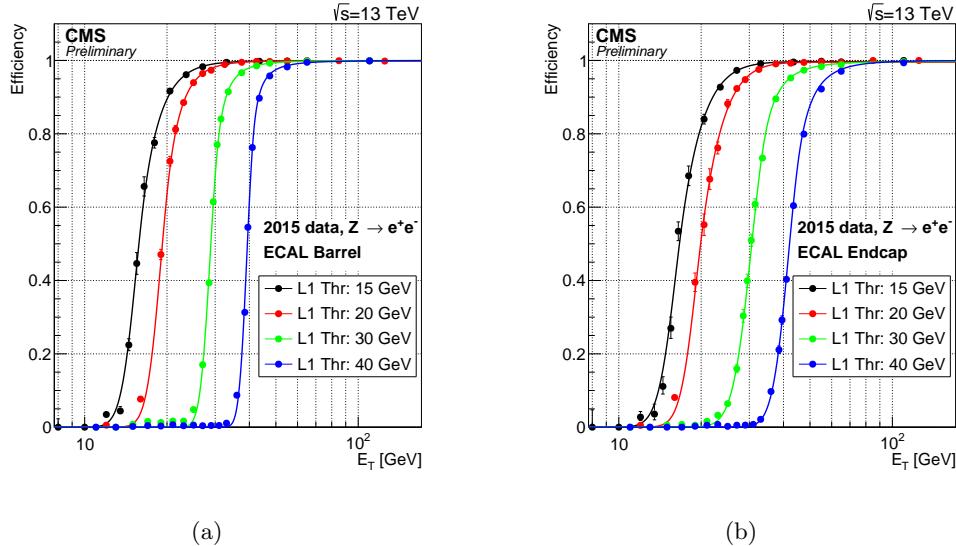
**Table 6.1:** The L1 and HLT single-electron triggers used to collect the 8 TeV data analyzed in this thesis together with the imposed requirements on the electron candidate.

Trigger	Name	Selections
Level 1	L1_SingleEG20	1 e/ $\gamma$ candidate $E_T > 20$ GeV
HLT	HLT_Ele80_CaloIdVT_GsfTrkIdT	1 GSF electron: $E_T > 80$ GeV $ \Delta\eta_{in}  < 0.008$ $ \Delta\phi_{in}  < 0.07$ (barrel) or 0.05 (endcaps) $H/E < 0.05$ $\sigma_{i\eta i\eta} < 0.011$ (barrel) or 0.031 (endcaps)

**Table 6.2:** The L1 and HLT single-electron triggers used to collect the 13 TeV data analyzed in this thesis together with the imposed requirements on the electron candidate.

Trigger	Name	Selections
Level 1	L1.SingleEG35 OR L1.SingleEG40	1 e/ $\gamma$ candidate $E_T > 35$ GeV OR $E_T > 40$ GeV
HLT	HLT_Ele105_CaloIdVT_GsfTrkIdT OR HLT_Ele115_CaloIdVT_GsfTrkIdT	1 GSF electron: $E_T > 105$ GeV OR $> 115$ GeV $ \Delta\eta_{in}  < 0.008$ $ \Delta\phi_{in}  < 0.07$ (barrel) or 0.05 (endcaps) $H/E < 0.05$ $\sigma_{i\eta i\eta} < 0.011$ (barrel) or 0.031 (endcaps)

1204 Both the L1 and HLT triggers require one electron (or  $\gamma$ ) candidate. The  $E_T$  thresholds  
 1205 imposed for the data collected in pp collisions at 13 TeV are higher compared to the one used



**Figure 6.7:** L1 electron triggering efficiency in ECAL barrel (a) and endcaps (b) as a function of the offline reconstructed electron  $E_T$ . The efficiency is shown for the 15, 20, 30, 40 GeV EG trigger thresholds [55].

in Run 1, in order to keep low trigger rates given the higher production rates of low-energy multijet background expected in Run 2. The chosen HLT triggers require a reconstructed GSF track whose association to the ECAL cluster has to pass tight quality criteria ( $|\Delta\eta_{in}|$  and  $|\Delta\phi_{in}|$ ). Requirements are also applied at this level on the transverse profile of the cluster of energy in the ECAL ( $\sigma_{in\eta\eta}$ ) and on the amount of energy in the HCAL downstream the ECAL ( $H/E$ ). There are no requirements imposed on the electron candidate isolation. In general, this results in high fake rates of jets misreconstructed as electrons from multijet background, and, as a consequence, in high trigger rates which would require a prescale. However, the high  $E_T$  threshold allows for an unprescaled trigger, as jets from multijet background are characterized by low momentum. In addition, the kinematic region of the analyses presented in this thesis is located at very high lepton  $p_T$  and the signal efficiency is mainly affected at very low resonance masses ( $< 1 \text{ TeV}$ ) with a loss in efficiency of 20–25%.

1218

The efficiency for an electron passing the high- $E_T$  selections described in Sec. 6.2.3 to fire the HLT triggers of Tables 6.1 and 6.2 have been measured in data with T&P method and are found to be 98–99% for electrons with  $E_T$  in the trigger plateau, with data-to-simulation scale factors close to unity.

### 1223 6.2.3 Electron identification

All the physics analyses in CMS involving one or two electrons in the final state start with the general electron reconstruction algorithm presented in Section 6.2.1. A high efficiency in any kinematical conditions is therefore needed and, as a consequence, the probability for other particles to be reconstructed as electrons is sizeable. For instance, a charged pion can mimic the signature of an electron if it interacts early and leaves most of its energy in the ECAL. Moreover electrons can emerge in a jet through the weakly decay of a hadron containing a c or b quark. Finally, in addition to jets, photons can also lead to GSF electron candidates. This happens if the photon converts into a dielectron pair in one of the first layers of the tracker detector. If one of the electron takes most of the photon momentum, a GSF electron

candidate is likely to be reconstructed. An analysis dependent selection, which takes into account the specific kinematics and background level, has therefore to be applied on top of the electron reconstruction. This thesis focuses on the search for massive resonances decaying to pairs of SM bosons where one of the bosons is a W decaying leptonically, with a highly energetic electron or muon in the final state. A high and stable selection efficiency for  $E_T$  above 100 GeV is therefore an important requirement. Since this is a common feature of many searches for new physics, a specific cut based selection has been developed in CMS [56], consisting of requirements on several variables that exploit the characteristics of high- $E_T$  electrons. Only GSF electron candidates with  $E_T > 35$  GeV and well reconstructed in the tracker and ECAL sensitive regions are selected. Candidates in the ECAL transition region ( $1.442 < |\eta_{SC}| < 1.56$ ) and beyond the  $\eta$  coverage ( $|\eta_{SC}| > 2.5$ ) of the tracker are therefore discarded. A different selection is applied for candidates reconstructed in the ECAL barrel ( $|\eta_{SC}| < 1.442$ ) and endcaps ( $1.56 < |\eta_{SC}| < 2.5$ ). For Run 2 the values of  $\eta_{SC}$  have been slightly adjusted to match the acceptance of the detector more accurately. The selections are summarized in Tables 6.3 and 6.4, for the 8 and 13 TeV data analysis, respectively, and discussed in the following.

**Table 6.3:** List of the variables used in the high- $E_T$  electron selections for the 8 TeV data analysis, together with the corresponding requirements for electrons reconstructed in the ECAL barrel and endcaps.

Variable	ECAL barrel	ECAL endcaps
ECAL-driven	yes	yes
$E_T$	$> 35$ GeV	$> 35$ GeV
$ \eta_{SC} $	$< 1.442$	$1.56\text{--}2.5$
$ \Delta\eta_{in} $	$< 0.005$	$< 0.007$
$ \Delta\phi_{in} $	$< 0.06$	$< 0.06$
Relative track isolation	5%	5%
Calorimeter isolation	$< 2 + 0.03E_T + 0.28\rho$	$< 2.5 + 0.28\rho$ if $E_T < 50$ $< 2.5 + 0.03(E_T - 50) + 0.28\rho$ if $E_T \geq 50$
Transverse shower shape	$E_{2\times 5}/E_{5\times 5} > 0.94$ OR $E_{1\times 5}/E_{5\times 5} > 0.83$	$\sigma_{in\eta} < 0.03$
$H/E$	$< 0.05$	$< 0.05$
$ d_{xy} $	$< 0.02$	$< 0.05$
Inner layer lost hits	$\leq 1$	$\leq 1$

As a starting point, electrons are selected if the reconstruction was seeded in the ECAL (Section 6.2.1). In fact, while useful for low-energy and non-isolated electrons, the PF algorithm is less suitable for high-energy electrons.

The difference in  $\eta$ ,  $\Delta\eta_{in}$ , and in  $\phi$ ,  $\Delta\phi_{in}$ , between the track position as measured in the inner layers, extrapolated to the interaction vertex and to the calorimeter, and the position of the supercluster, are required to be  $< 0.005$  and  $< 0.06$ , respectively. In fact, for jets, the position of the center of the ECAL deposit can be far from the track position, as all of the constituents can leave an energy deposit in the ECAL. The  $\Delta\phi_{in}$  distribution is however much broader than  $\Delta\eta_{in}$ , because of the wider spread of the energy in  $\phi$  due to photons from bremsstrahlung, resulting in a looser requirement. The distributions of  $\Delta\phi_{in}$  and  $\Delta\eta_{in}$  become narrower with increasing  $E_T$ , and therefore a higher discrimination power can be achieved with a tighter requirement at high  $E_T$  compared to the usual selections for low or intermediate energetic electrons. The reason of this behaviour comes from the fact that bremsstrahlung photons are more collinear to the electron at higher  $E_T$ . The definition of  $\Delta\eta_{in}$  has been changed for Run 2 to use instead the  $\eta$  of the seed cluster of the supercluster

**Table 6.4:** List of the variables used in the high- $E_T$  selections for the 13 TeV data analysis, together with the corresponding requirements for electrons reconstructed in the ECAL barrel and endcaps.

Variable	ECAL barrel	ECAL endcaps
ECAL-driven	yes	yes
$E_T$	$> 35 \text{ GeV}$	$> 35 \text{ GeV}$
$ \eta_{SC} $	$< 1.4442$	$1.566\text{--}2.5$
$ \Delta\eta_{in} $	$< 0.004$	$< 0.006$
$ \Delta\phi_{in} $	$< 0.06$	$< 0.06$
Relative track isolation	5%	5%
Calorimeter isolation	$< 2 + 0.03E_T + 0.28\rho$	$< 2.5 + 0.28\rho \text{ if } E_T < 50$ $< 2.5 + 0.03(E_T - 50) + 0.28\rho \text{ if } E_T \geq 50$
Transverse shower shape	$E_{2\times 5}/E_{5\times 5} > 0.94$ OR $E_{1\times 5}/E_{5\times 5} > 0.83$	$\sigma_{i\eta i\eta} < 0.03$
$H/E$	$< 1/E + 0.05$	$< 5/E + 0.05$
$ d_{xy} $	$< 0.02$	$< 0.05$
Inner layer lost hits	$\leq 1$	$\leq 1$

1264 which is found to provide a more accurate indication of the  $\eta$  of the original electron before  
 1265 bremsstrahlung.

1266 To suppress the misidentification of jets as electrons, the sum of the  $p_T$  of all other tracks  
 1267 in a cone of  $\Delta R < 0.3$  around the track of the electron candidate is required to be less  
 1268 than 5 GeV, imposing an isolation condition on the electron candidate track. To be used in  
 1269 the calculation of the isolation of the candidate track, the tracks have to be within 0.2 cm,  
 1270 in the  $z$  direction, of the primary vertex with which the electron candidate is associated.  
 1271 This requirement reduces the impact of pileup and it does not show a dependency with the  
 1272 electron  $E_T$  for values above 100 GeV. For electrons with  $E_T$  much lower than 100 GeV, the  
 1273 efficiency decreases up to 10% depending on the region of the detector in which the electrons  
 1274 are detected.

1275 A calorimeter-based isolation is applied and defined as the sum of:

- 1276 • ECAL isolation: sum of the  $E_T$  of the energy deposits in the ECAL calorimeter in a  
 1277 cone of  $\Delta R < 0.3$  around the track of the electron candidate excluding those associated  
 1278 with the candidate;
- 1279 • HCAL1 isolation: sum of the  $E_T$  of the energy deposits in the first layer of the HCAL  
 1280 calorimeter in a cone of  $\Delta R < 0.3$  around the track of the electron candidate excluding  
 1281 those associated with the candidate.

1282 The isolation variable so defined, is required to be less than 3% (plus a small  $\eta$ -dependent  
 1283 offset) of the candidate  $E_T$ . This sum, which allows a selection on the isolation of the  
 1284 electron candidate, is corrected for the average energy density in the event,  $\rho$ , to minimize  
 1285 the dependence of the efficiency of this selection criterion on pileup. This requirement differs  
 1286 from the selection usually applied for electrons of low or intermediate  $E_T$ . For these cases, a  
 1287 PF-based isolation is generally used, which merges the information of the tracker, the ECAL  
 1288 and the HCAL allowing to measure the contribution to the isolation from charged hadrons,  
 1289 neutral hadrons and photons separately. One of the main advantage of the PF-based isolation  
 1290 is that the energy deposit in the calorimeters associated to a charged hadron produced in  
 1291 another interaction, characterized by a different primary vertex, can be removed from the  
 1292 isolation sum. For very high energy ( $> 1 \text{ TeV}$ ) electrons, however, the PF algorithm might  
 1293 fail to recognize an electron from a GSF electron candidate and assigns all its energy deposit

to the photon isolation. Furthermore, the PF isolation is generally required to be below a fixed fraction of the electron  $E_T$  independently on its value. However, for high  $E_T$  values the background rejection can be improved while keeping an acceptable efficiency by following the  $E_T$  dependence of the ECAL+HCAL1 isolation variable. In fact, this isolation tends to increase for high- $E_T$  electrons due to the extension of the shower.

Further suppression of the misidentification of jets as electrons is achieved by requiring that the ratio  $H/E$  of the energy in the HCAL towers in a cone of  $\Delta R < 0.15$  centered on the electron candidate position, to the electromagnetic energy of the electron candidate supercluster is required to be less than 5%. This requirement is tighter compared with the threshold applied for low- or medium-energy electrons, where it becomes quite inefficient for a high number of pileup interactions. For Run 2, the selection on this variable has been increased. Additionally, the transverse profile of the energy deposition in the ECAL is required to be consistent with that expected for an electron, being defined by the following variables:

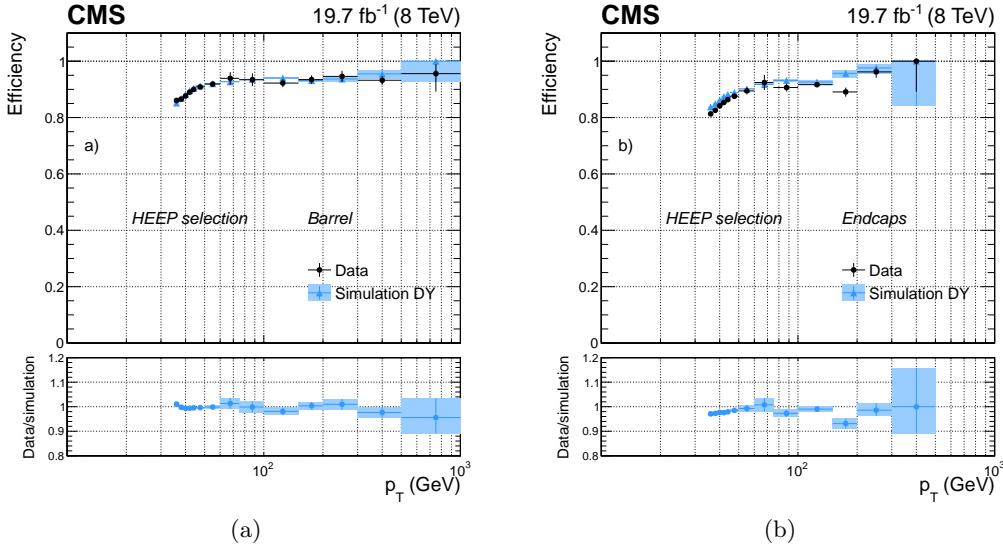
- $E_{1\times 5}/E_{5\times 5}$ : ratio of the energy contained in the  $1\times 5$  matrix in  $\eta \times \phi$  in the barrel ( $x \times y$  in the endcaps) centered on the seed crystal of the supercluster over the energy of the  $5 \times 5$  matrix centered on the seed crystal;
- $E_{2\times 5}/E_{5\times 5}$ : ratio of the energy contained in the most energetic  $2\times 5$  matrix in  $\eta \times \phi$  in the barrel ( $x \times y$  in the endcaps) centered on the seed crystal of the supercluster over the energy of the  $5 \times 5$  matrix centered on the seed crystal;
- $\sigma_{in\eta}$  : measure of the spread in  $\eta$  in units of crystals of the electrons energy in the  $5 \times 5$  block centered on the seed crystal.

In the barrel, the best performance are obtained applying a selection on both  $E_{1\times 5}/E_{5\times 5}$  and  $E_{2\times 5}/E_{5\times 5}$ . The two variables are indeed complementary: while  $E_{1\times 5}/E_{5\times 5}$  is well designed for electrons hitting the center of a crystal,  $E_{2\times 5}/E_{5\times 5}$  allows to recover electrons that hit the crystal close to its edge. Combining the two variables instead of using just one of them allows to set a tight requirement on both and thus well reject background while keeping a high efficiency on simulated electrons. The distributions of these variables are much broader for electrons in the endcaps and a higher discrimination power is obtained applying a selection on the variable  $\sigma_{in\eta}$ .

Two additional requirements are applied to reject photons that convert into a electron-positron pair in the tracker. First, the track associated with the cluster is required to have no more than one hit missing in the pixel layers. IN fact, the signature arising from photon conversion process is very similar to the one from real electrons, and the gain in discrimination using shower shape variables is limited. However, one of the main differences is the absence of hits in the first layers of the tracker, before the conversion happens. Furthermore, the transverse impact parameter  $d_{xy}$ , defined as the closest distance, in the transverse plane, between the primary vertex and the track of the electron candidate, is required to be  $< 0.02$  cm (barrel) or  $0.05$  cm (endcaps). The distribution of the transverse impact parameter is usually wider in the endcaps due to the poorer resolution of the track position in that region.

The efficiency of the high- $E_T$  electron selection measured with the T&P method in pp collisions at  $\sqrt{s} = 8$  TeV and in simulation as a function of the electron  $p_T$  is shown in Fig. 6.8, for electrons reconstructed in the ECAL barrel and endcaps. Similar results are obtained using 13 TeV data. The efficiencies and data-to-simulation scale factors are summarized in Tables 6.5 and 6.6, as measured in 8 and 13 TeV data and simulation, respectively. The scale

<sup>1340</sup> factors are close to unity, indicating a good agreement between data and simulation. They  
<sup>1341</sup> are used in the analysis presented in this thesis to correct the normalization of simulations.



**Figure 6.8:** Efficiency of the high- $E_T$  electron selection as a function of electron  $p_T$  for dielectron events in pp collisions at  $\sqrt{s} = 8$  TeV (dots) and in DY simulation (triangles) for electrons reconstructed in the ECAL barrel (a), and endcaps (b) [54].

**Table 6.5:** Efficiencies and data-to-simulation scale factors for the high- $E_T$  electron selection, as measured in pp collisions at  $\sqrt{s} = 8$  TeV for electrons with  $E_T > 90$  GeV. The quoted uncertainties are statistical.

	ECAL barrel	ECAL endcaps
Efficiency simulation	$90.2\% \pm 0.2\%$	$92.2\% \pm 0.5\%$
Efficiency data	$88.7\% \pm 0.2\%$	$90.7\% \pm 0.6\%$
Data/simulation scale factor	$0.983 \pm 0.004$	$0.984 \pm 0.010$

**Table 6.6:** Efficiencies and data-to-simulation scale factors for the high- $E_T$  electron selection as measured in pp collisions at  $\sqrt{s} = 13$  TeV for electrons with  $E_T > 120$  GeV. The quoted uncertainties are statistical.

	ECAL barrel	ECAL endcaps
Efficiency simulation	$91.4\% \pm 0.10\%$	$84.4\% \pm 0.3\%$
Efficiency data	$91.6\% \pm 0.04\%$	$82.3\% \pm 0.1\%$
Data/simulation scale factor	$1.002 \pm 0.001$	$0.975 \pm 0.004$

## <sup>1342</sup> 6.3 Muons

### <sup>1343</sup> 6.3.1 Muon reconstruction

<sup>1344</sup> The CMS detector is specifically designed for the optimization of muon detection, as its name  
<sup>1345</sup> clearly states. In general, muons will not be absorbed by the calorimeters, as it happens with  
<sup>1346</sup> electrons, so a specific muon detection system (Section 3.2.3) is needed in order to identify

and correctly measure its momentum.

In the standard CMS reconstruction [57], tracks are first reconstructed independently in the inner tracker (tracker track) and in the muon system (standalone-muon track). A standalone-muon track is reconstructed from pre-built track segments (i.e. a set of aligned DT or CSC hits) in the muon chambers. The state vector associated to the segments found in the innermost chambers is used to seed the muon trajectory, from inside out, using the KF technique: the predicted state vector at the next measurement surface is compared with existing hits and updated accordingly. A suitable  $\chi^2$  cut is applied to reject bad hits and the procedure is iterated until the outermost surface of the muon system is reached. Finally, the track is extrapolated to the nominal interaction point and a vertex-constrained fit is performed. The magnetic field, the multiple scattering inside the steel yoke, and the energy loses are taken into account.

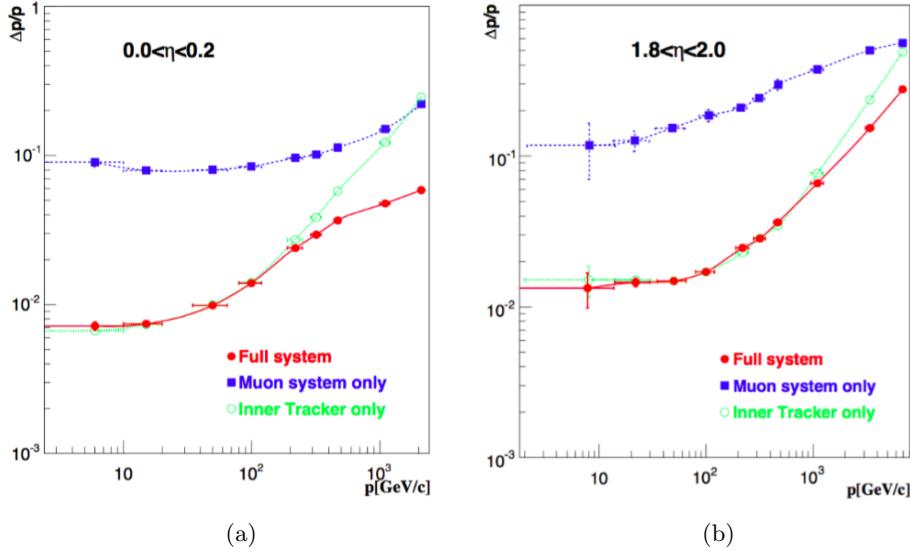
Based on reconstructed standalone-muon and tracker tracks, two reconstruction approaches are then used:

- **global-muon reconstruction (outside-in)**: each standalone-muon track is extrapolated to the tracker and a search is performed in a cone around it to match a tracker track; a global-muon track is fitted combining hits from the tracker track and standalone-muon track, using the KF technique;
- **tracker-muon reconstruction (inside-out)**: all tracker tracks with  $p_T > 0.5 \text{ GeV}$  are considered as possible muon candidates and are extrapolated to the muon system while searching for a match with at least one muon segment.

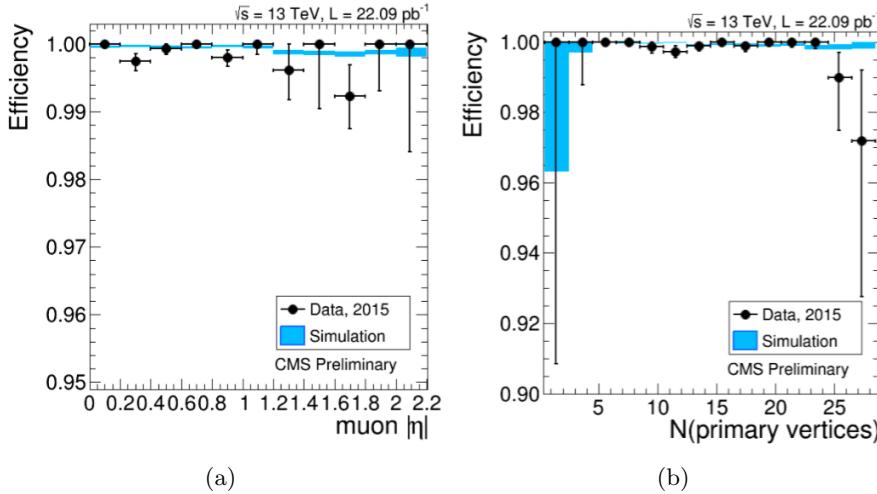
Tracker-muon reconstruction is more efficient than the global-muon reconstruction at low momenta,  $p_T \leq 5 \text{ GeV}$ , because it requires only a single muon segment in the muon system, whereas global-muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station, and typically requires segments in at least two muon stations. However, given the high efficiency of both the tracker track and muon segments reconstruction, about 99% of muons produced within the geometrical acceptance of the muon system and having sufficiently high momentum ( $p_T \geq 5 \text{ GeV}$ ) are reconstructed by both methods. As shown in Fig. 6.9 the additional information provided by the muon system is precious for the momentum reconstruction of high-energy muons ( $p_T \geq 200 \text{ GeV}$ ), for which the tracker-only momentum measurement degrades. In fact, as a particle's momentum increases and the curvature of its corresponding track decreases, the momentum resolution in the tracker becomes limited by position measurement resolution. One can then benefit from the large lever arm and 3.8 T magnetic field in the region between the tracker and the muon system by including hits in the muon chambers. For lower momenta, instead, the resolution of the tracking system is dominating.

Figure 6.10 shows the muon tracking efficiency as a function of the  $\eta$  of the probe muon and the number of primary vertices for 13 TeV data and simulation, evaluated using the T&P method described in Section 6.2.1. In the region  $|\eta| < 2.2$  and for events with number of reconstructed primary vertices lower than 25, the measured tracking efficiency for isolated muons is  $> 99\%$  in both data and simulation. The efficiency is constant as a function of the number of vertices in the event, hence it does not depend on the pileup.

The combination of different algorithms provides robust and efficient muon reconstruction. After the completion of both algorithms, the reconstructed stand-alone, global, and tracker muons are merged into a single software object, with the addition of further information, like



**Figure 6.9:** Relative resolution of the muon momentum measurement for the reconstruction with the inner tracker only, the muon system only and for the combination of the inner tracker and the muon system, for simulated muons emitted in the central (a) and forward (b) regions [58].



**Figure 6.10:** Tracking efficiency measured with a T&P technique, for muons from  $Z$  decays, as a function of the muon  $\eta$  (a) and the number of primary vertices (b), for 2015 data (black dots) and simulation (blue bands) [59].

isolation and energy collected in matching calorimeter towers. This information can be used for further identification, in order to achieve a balance between efficiency and purity of the muon sample as described in Section 6.3.3.

1397

1398 The performance of the reconstruction for high- $p_T$  muons is strongly affected by radiative  
 1399 processes and by the muon detector alignment. Electromagnetic showers and large energy  
 1400 losses can arise as the muon traverses the steel layers of the magnet return yoke, producing  
 1401 additional segments in the muon chambers. These events can affect the measurement done  
 1402 in the muon detectors. Therefore, specialized reconstruction algorithms for high- $p_T$  muons,  
 1403 known as “TeV-muon” refits, have been developed in CMS as described in the following.

The *tracker-plus-first-muon-station* fit (TPFMS) only uses hits from the tracker and the innermost muon station with hits, to reduce the sensitivity to possible showering starting deeper in the muon system. The *Picky* fit uses all tracker hits, while a selection is applied to muon hits. Hits from chambers with a high probability of shower contamination (determined from the hit occupancy) are required to be compatible with the extrapolated trajectory by applying a  $\chi^2$  cut. The *dynamic truncation* algorithm (DYT) starts from the idea that the muon track reconstruction should be stopped after a large energy loss, as hits produced after that can only bias the momentum measurement. For every global muon trajectory the algorithm starts from the corresponding tracker track and propagates it out to the muon stations. Compatible segments (or hits) in the muon chambers are found by using an estimator which takes into account the propagation of the tracker covariance matrix through the material and the magnetic field, and the covariance matrices of the candidate muon segments (or hits).

Momentum assignment is then performed by the *Cocktail* algorithm which combines the above methods to further improve the resolution at high  $p_T$  reducing the tails of the momentum resolution distribution. In particular, the algorithm chooses, on track-by-track basis, the best muon reconstruction. For Run 1, the Cocktail-algorithm decision is taken between the tracker-only, TPFMS, and Picky fits. This version of the algorithm is also known as the *Tune P* algorithm. It starts with the Picky fit, then switches to the tracker-only fit if the goodness of fit ( $\chi^2/\text{n.d.f.}$ ) of the latter is significantly better. Then it compares the  $\chi^2/\text{n.d.f.}$  of the chosen track with that of TPFMS; TPFMS is chosen if it is found to be better. For high- $p_T$  muons, TPFMS and Picky algorithms are selected by Tune P in most of the cases, in approximately equal amounts, while the tracker-only fit is selected only in a few percent of events.

For Run 2, the Tune P algorithm was extended to include also the DYT fit. The selection is still made on a track-by-track basis, but using both the  $\chi^2/\text{n.d.f.}$  of the track and the relative error of the  $p_T$  measurement. The algorithm starts with the Picky fit, then switches to DYT if the DYT track has a lower relative  $p_T$  error. It then compares the  $\chi^2/\text{n.d.f.}$  of the chosen track with that of the tracker-only fit and picks tracker-only if its  $\chi^2/\text{n.d.f.}$  is significantly better. Then the  $\chi^2/\text{n.d.f.}$  of the chosen track and TPFMS are compared and the one giving the best result is kept. At the end, if the final candidate track has  $p_T$  lower than 200 GeV or the tracker-only  $p_T$  is lower than 200 GeV, the tracker-only track is selected.

The momentum resolution obtained with the Tune P algorithm for muons with  $p_T$  in the range  $350 < p_T < 2000$  GeV is found to be  $\approx 6\%$ , as measured with cosmic-ray muon data [57, 60].

### 6.3.2 Muon trigger

The Level-1 muon trigger uses signals from all three CMS muon detector systems: DT, CSC, and RPC. It has a latency of  $3.2 \mu\text{s}$  and reduces the rate of the readout of events with muon candidates at the detector front-end electronics to a few kHz by applying selections on the estimated muon  $p_T$  and quality. In the muon HLT, first a Level-1 trigger object is used as a seed to reconstruct a standalone-muon track in the muon system, leading to an improved  $p_T$  estimate. At this point,  $p_T$  threshold filters are applied to the standalone-muon (also called Level-2 muon). Then seeds in the inner tracker are generated in the region around the extrapolated Level-2 muon, and tracker tracks are reconstructed. If a successful match is made between a tracker track and the Level-2 muon, a global fit combining tracker and muon hits is performed, yielding a Level-3 muon track on which the final  $p_T$  requirements are applied. In this way, the rate of recorded inclusive muon events is reduced to a few tens of Hz. The average processing time of the HLT reconstruction is about 50 ms.

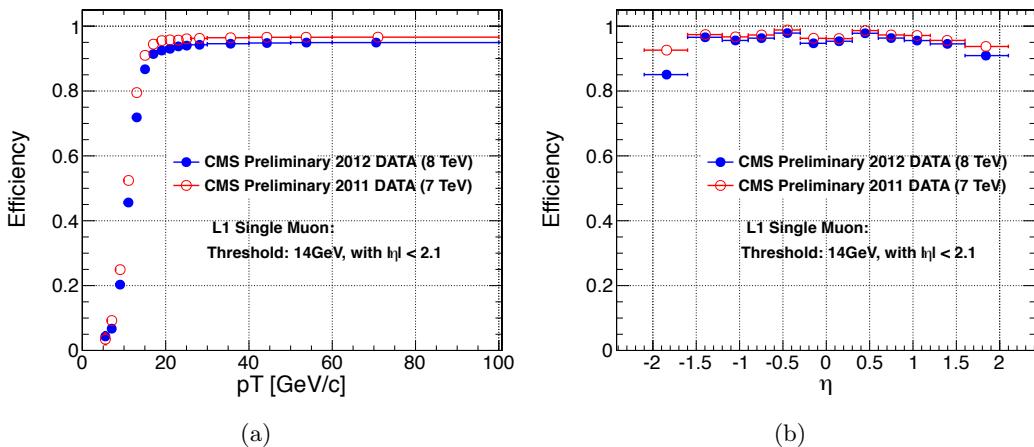
The L1 and HLT trigger used to collect the data analyzed in this thesis are listed in Tables 6.7 and 6.8 for the 8 and 13 TeV data analysis, respectively. For both analyses the HLT used to select the events is the unprescaled single-muon trigger with the lowest  $p_T$  threshold that does not include muon isolation requirements. In fact, although muons produced by the leptonic decay of a high- $p_T$  W boson tend to be isolated, their high momentum enhances the production of electromagnetic showers, that can mimic a non-isolated muon candidate. Therefore, only requirements on the muon  $p_T$  and  $\eta$  are applied at this stage. The efficiency of the L1 single-muon trigger with the 16 GeV threshold is shown in Fig. 6.11 as a function of the offline reconstructed muon  $p_T$  and  $\eta$ . In 2012 the efficiency for this trigger was greater than 90%. A similar result is obtained in 2015.

**Table 6.7:** The L1 and HLT single-muon triggers used to collect the 8 TeV data analyzed in this thesis together with the imposed requirements on the muon candidate.

Trigger	Name	Selections
Level 1	L1_SingleMu16_eta2p1	1 global muon with: $p_T > 16 \text{ GeV}$ $ \eta  < 2.1$
HLT	HLT_Mu40_eta2p1	1 global muon with: $p_T > 40 \text{ GeV}$ $ \eta  < 2.1$

**Table 6.8:** The L1 and HLT single-muon triggers used to collect the 13 TeV data analyzed in this thesis together with the imposed requirements on the muon candidate.

Trigger	Name	Selections
Level 1	L1_SingleMu25	1 global muon with: $p_T > 25 \text{ GeV}$
HLT	HLT_Mu45_eta2p1	1 global muon with: $p_T > 45 \text{ GeV}$ $ \eta  < 2.1$



**Figure 6.11:** Efficiency of the L1 single-muon trigger with a threshold of 14 GeV on the muon  $p_T$  as a function of the muon  $p_T$  (a) and  $\eta$  (b) [61].

1462 The efficiency for a muon passing the high- $p_T$  selections described in Section 6.3.3 to  
1463 fire the HLT single-muon triggers have been measured in data with T&P method and are  
1464 summarized in Tables 6.9 and 6.10.

**Table 6.9:** Efficiencies and scale factors for the single-muon HLT trigger used in the 8 TeV analysis for muons with  $p_T > 50 \text{ GeV}$ ,  $|\eta| < 2.1$ , and satisfying the high- $p_T$  and isolation selections described in Section 6.3.3. The quoted uncertainties are statistical.

	$0 <  \eta  < 0.9$	$0.9 <  \eta  < 1.2$	$1.2 <  \eta  < 2.1$
Efficiency simulation	$95.10\% \pm 0.03\%$	$87.01\% \pm 0.03\%$	$81.56\% \pm 0.03\%$
Efficiency data	$92.90\% \pm 0.02\%$	$83.14\% \pm 0.06\%$	$80.27\% \pm 0.05\%$
Data/simulation scale factor	$0.9768 \pm 0.0004$	$0.956 \pm 0.001$	$0.984 \pm 0.001$

**Table 6.10:** Efficiencies and scale factors for the single-muon HLT trigger used in the 13 TeV analysis for muons with  $p_T > 53 \text{ GeV}$ ,  $|\eta| < 2.1$ , and satisfying the high- $p_T$  and isolation selections described in Section 6.3.3. The quoted uncertainties are statistical.

	$0 <  \eta  < 0.9$	$0.9 <  \eta  < 1.2$	$1.2 <  \eta  < 2.1$
Efficiency simulation	$97.6\% \pm 0.1\%$	$93.4\% \pm 0.4\%$	$94.8\% \pm 0.2\%$
Efficiency data	$94.6\% \pm 0.2\%$	$89.7\% \pm 0.4\%$	$91.8\% \pm 0.2\%$
Data/simulation scale factor	$0.969 \pm 0.002$	$0.961 \pm 0.006$	$0.968 \pm 0.003$

### 6.3.3 Muon identification

1465 1466 The standard CMS muon reconstruction provides additional information for each muon, useful  
1467 for muon quality selection and identification in physics analyses [57]. In general, particles  
1468 detected as muons are produced in pp collision from different sources which lead to different  
1469 experimental signatures. The so-called *prompt muons* arise either from decays of W, Z, and  
1470 promptly produced quarkonia states, or other sources such as Drell-Yan processes or top  
1471 quark production. Real muons are also produced in the decay of heavy flavour particles,  
1472 such as beauty or charmed mesons, as well as in light hadron (pions or kaons) decays. Less  
1473 frequently, muons might be originated from a calorimeter shower or a product of a nuclear  
1474 interaction in the detector. Furthermore, the so called “punch-through” effect, i.e. hadron  
1475 shower remnants penetrating through the calorimeters and reaching the muon system, can  
1476 lead to the reconstruction of a muon candidate. Most of the physics analyses in CMS studying  
1477 SM processes or searching for BSM signals use prompt muons, while all the other categories  
1478 constitute the background. These analyses exploit the same set of information, although  
1479 the applied selections might be different depending on the interesting signature and the  
1480 expected background. In this section only the specific selection developed for high- $p_T$  muons  
1481 are described. One of the main difference with respect to the low- and medium- $p_T$  muon  
1482 selection is that this particular identification procedure does not use the PF algorithm. It is  
1483 aimed at the best reconstruction of the muon track parameters without relying on external  
1484 information on the event. Moreover, the goodness of the global-muon track fit selection, based  
1485 on the  $\chi^2$  of the track, is not requested, but an additional selection based on the relative  $p_T$   
1486 resolution for the track used for momentum determination is applied.

1487 The high- $p_T$  muon selection criteria are described in the following and they have not been  
1488 changed since Run 1:

- 1489 • The muon must be reconstructed both as a tracker- and a global-muon. This is

1490 effective against decays-in-flight, punch-through and accidental matching (with noisy or  
 1491 background tracks or segments).

- 1492 • Number of pixel hits in the tracker track  $\geq 1$ . To further suppress muons from decays  
   1493 in flight.
- 1494 • Number of tracker layers involved in the measurements  $\geq 6$ . This guarantees a good  $p_T$   
   1495 measurement, for which some minimal number of measurement points in the tracker is  
   1496 needed. It also suppresses muons from decays in flight.
- 1497 • Number of muon-chamber hits included in the global-muon track fit  $\geq 1$ . This require-  
   1498 ment assures that the global muon is not an accidental match between the information  
   1499 from the muon system and the tracker. This could happen in particular for non-prompt  
   1500 muons or fake muons from punch-through.
- 1501 • The muon track is required to have muon segments in at least 2 muon stations. To further  
   1502 suppress punch-through and accidental track-to-segment matches. This selection is  
   1503 furthermore consistent with the logic of the single-muon trigger, which requires segments  
   1504 in at least two muon stations to obtain a meaningful estimate of the muon  $p_T$ .
- 1505 • Transverse impact parameter of the muon track  $< 2$  mm. This assures the compatibility  
   1506 of the muon track with the interaction point hypothesis and it is effective against cosmic  
   1507 background and further suppress muons from decays in flight.
- 1508 • Longitudinal impact parameter of the muon track  $< 5$  mm. To further suppress cosmic  
   1509 muons, muons from decays in flight and tracks from pileup.
- 1510 • Relative  $p_T$  error  $< 30\%$ . To further suppress mis-reconstructed muons.

1511 In addition to these identification criteria, an isolation requirement is applied to the  
 1512 well-identified muons. In particular, the muon must pass a relative tracker-only isolation  
 1513 selection: the scalar sum of the  $p_T$  of all other tracks in a cone of  $\Delta R < 0.3$  around but not  
 1514 including the muon tracker track must be less than 10% of the muon  $p_T$ , also as measured  
 1515 by the tracker. To be used in the calculation of the tracker-based isolation, tracks have to  
 1516 be within 2 mm, in the  $z$  direction, of the primary vertex with which the muon candidate is  
 1517 associated. These additional criteria help suppress the effect of tracks originating from pileup  
 1518 on the reconstructed quantities.

1519 The efficiency and data-to-simulation scale factors for the high- $p_T$  muon identification  
 1520 and isolation criteria measured with the T&P method in 8 and 13 TeV data are summarized,  
 1521 respectively, in Tables 6.11 and 6.12. The scale factors are close to unity, indicating a good  
 1522 agreement between data and simulation. They are used in the analyses presented in this  
 1523 thesis to correct the normalization of simulations.

## 1524 6.4 Jets

1525 Particles carrying a color charge, such as quarks, cannot exist in free form because of QCD  
 1526 confinement which only allows for colorless states (**FIXME: point here to theory chapter**).  
 1527 Quarks and gluons interact with pairs of quarks and anti-quarks produced from the vacuum  
 1528 until the formation of stable colourless hadrons. The ensemble of the final colourless objects  
 1529 is called a *jet* and it is reconstructed in the detector from energy depositions and charged  
 1530 particle momenta. The jets point back to the primary interaction, i.e. to the partons  
 1531 the jets originated from, but a correction for hadronization and detector effects is needed.

**Table 6.11:** Efficiencies and scale factors for the high- $p_T$  muon identification and isolation criteria used in the 8 TeV data analysis for muons with  $p_T > 50$  GeV and  $|\eta| < 2.1$ .

Muon $ \eta $	$0 <  \eta  < 0.9$	$0.9 <  \eta  < 1.2$	$1.2 <  \eta  < 2.1$
High- $p_T$ muon identification			
Efficiency simulation	$96.51\% \pm 0.02\%$	$96.61\% \pm 0.04\%$	$95.54\% \pm 0.03\%$
Efficiency data	$95.54\% \pm 0.02\%$	$95.87\% \pm 0.04\%$	$95.06\% \pm 0.03\%$
Data/simulation scale factor	$0.9900 \pm 0.0003$	$0.992 \pm 0.001$	$0.9949 \pm 0.0004$
Tracker-based muon isolation			
Efficiency simulation	$99.49\% \pm 0.01\%$	$99.58\% \pm 0.01\%$	$99.59\% \pm 0.01\%$
Efficiency data	$99.46\% \pm 0.01\%$	$99.51\% \pm 0.01\%$	$99.56\% \pm 0.01\%$
Data/simulation scale factor	$0.9996 \pm 0.0001$	$0.9994 \pm 0.0001$	$0.9997 \pm 0.0001$

**Table 6.12:** Efficiencies and scale factors for high- $p_T$  muon identification and isolation criteria used in the 13 TeV data analysis for muons with  $p_T > 53$  GeV and  $|\eta| < 2.1$ .

Muon $ \eta $	$0 <  \eta  < 1.2$	$1.2 <  \eta  < 2.1$
High- $p_T$ muon identification		
Efficiency simulation	$97.6\% \pm 0.2\%$	$99.81\% \pm 0.2\%$
Efficiency data	$96.7\% \pm 0.4\%$	$1.0\% \pm 0.7\%$
Data/simulation scale factor	$0.991 \pm 0.005$	$1.002 \pm 0.007$
Tracker-based muon isolation		
Efficiency simulation	$99.8\% \pm 0.1\%$	$99.6\% \pm 0.1\%$
Efficiency data	$99.7\% \pm 0.1\%$	$99.7\% \pm 0.1\%$
Data/simulation scale factor	$0.999 \pm 0.001$	$1.001 \pm 0.001$

1532 Jet clustering algorithms have been developed to cluster particles (at parton, particle or  
 1533 detector level) into jets and reconstruct the energy and direction of the original parton. The  
 1534 task of a jet clustering algorithm is to allow comparisons between theoretical predictions,  
 1535 which are usually described by perturbative calculations, and experimental data. This is  
 1536 achieved reducing the complex structure of particle jets from a scattered parton to a simple  
 1537 four-momentum, which represents the main property of particle jets. In order to guarantee  
 1538 a meaningful calculation of theory predictions, jet clustering algorithms are characterized  
 1539 by two important properties. Clustering algorithms need to be infrared-safe, which means  
 1540 that the emission of infinitesimally-low-energy partons from partons inside a jet does not  
 1541 affect the jet properties. Furthermore, they need to be collinear-safe, which means that jet  
 1542 properties are not affected by the splitting of a parton inside a jet into two collinear partons.  
 1543 Jet algorithms for hadron colliders can be divided into two classes: cone [62] and sequential  
 1544 clustering [63–67] algorithms. The main algorithms used by LHC experiments belong to  
 1545 the second class and are the anti- $k_t$  [67] (AK) and the Cambridge–Aachen (CA) [63, 65]  
 1546 algorithms. In fact, they are found to fulfil theory requirements and to exhibit good properties  
 1547 for experimental measurements. For this work both algorithms are used and described in the  
 1548 following.

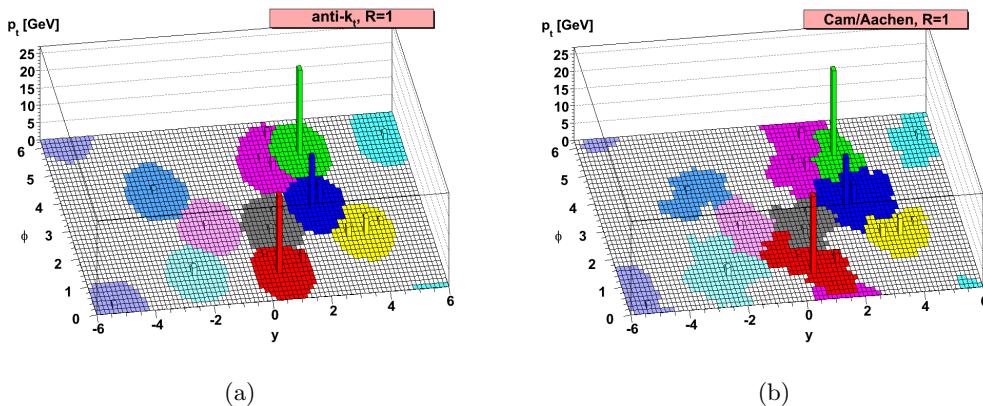
#### 1549 6.4.1 Jet clustering algorithms

1550 In sequential jet clustering algorithms, jets are defined through sequential, iterative procedures  
 1551 that combine four-vectors of input pairs of particles until certain criteria are satisfied and jets  
 1552 are formed. In particular, for each pair of particles  $i$  and  $j$ , a distance variable between the  
 1553 two particles ( $d_{ij}$ ), and the so-called “beam distance” for each particle ( $d_{iB}$ ), are computed:

$$d_{ij} = \min(p_{Ti}^{2n}, p_{Tj}^{2n}) \frac{\Delta R^2 ij}{R^2} , \quad d_{iB} = p_{Ti}^{2n} , \quad (6.1)$$

where  $p_{Ti}$  and  $p_{Tj}$  are the transverse momenta of particles  $i$  and  $j$ , respectively, “ $\min$ ” refers to the smaller of the two  $p_T$  values, the integer  $n$  depends on the specific jet algorithm,  $\Delta R^2 ij$  is the distance between  $i$  and  $j$  in  $\eta$  and  $\phi$ , and  $R$  is a free distance parameter, with all angles expressed in radians. The particle pair  $(i, j)$  with smallest  $d_{ij}$  is combined into a single object. All distances are recalculated using the new object, and the procedure is repeated until, for a given object  $i$ , all the  $d_{ij}$  are greater than  $d_{iB}$ . Object  $i$  is then classified as a jet and not considered further in the algorithm. The process is repeated until all input particles are clustered into jets.

The distance parameter  $R$  is responsible for defining the angular size of the jet. The parameter  $n$  governs the topological properties of the jets and depending on its value three different classes of clustering algorithms are distinguished. For  $n = 1$  the procedure is referred to as the  $k_t$  algorithm (KT), which clusters soft objects before harder ones are added to the final jet. The KT jets tend to have irregular shapes and are especially useful for reconstructing jets of lower momentum [67]. For this reason, they are also sensitive to the presence of low- $p_T$  pileup contributions. For  $n = 0$ , the procedure corresponds to the CA algorithm. This relies only on angular information, and, like the KT algorithm, provides irregularly-shaped jets. The CA algorithm is useful in identifying jet substructure as described in Chapter 7. For  $n = -1$ , the procedure corresponds to the AK algorithm, which compares the inverse square of the transverse momenta. The AK algorithm is used extensively in LHC experiments and by the theoretical community for finding well-separated jets. The use of inverse square of the  $p_T$  as a weight in the  $d_{ij}$  distances has the advantage that hard objects collect adjacent soft ones before these are clustered among themselves into harder objects, figuratively reproducing in reverse the parton fragmentation and gluon emission processes. This property makes the algorithm independent on soft radiation preserving infrared-safety. The AK algorithm is also collinear-safe as the clustering is driven by the angular distance between two particles. Gluons emitted at small angles are picked up by the algorithm in early steps of the iteration and therefore do not affect the jet properties. Furthermore, this algorithm tends, by construction, to form almost circular jets allowing for straight-forward calibration and understanding of the detector acceptance. The behaviours of the CA and AK jet algorithms are illustrated in Fig. 6.12.



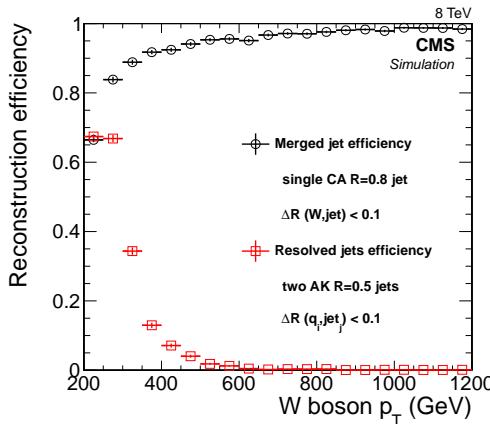
**Figure 6.12:** An example of jet clustering with the AK (a) and CA (b) algorithms. The reconstructed jets are shown as colored regions [67].

The choice of the distance parameters  $R$ , generally depends on the analysis. While large cone size jets collect all energy from the scattered parton, they also pick up a large contribution of background energy from the underlying event or pileup interactions. Small cone size jets pick up little contamination, but may not collect all energy from the scattered parton. The default choice in CMS for physics analyses in Run 1 and Run 2 uses the KT algorithm with  $R = 0.5$  (AK5) and  $R = 0.4$  (AK4), respectively, since more collimated jets are expected at higher  $\sqrt{s}$ . However, a larger value of  $R$  increases the efficiency to entirely reconstruct the highly energetic products in the decays into hadrons of boosted V and Higgs bosons. In fact, the average angular distance between the decay products is inversely proportional to the  $p_T$  of the mother particle. The default choice in CMS for physics analyses involving boosted V or Higgs bosons decaying hadronically is  $R = 0.8$ . In particular, CA8 and AK8 jets are used for Run 1 and Run 2 analyses, respectively. The chosen value of  $R$  provides a high efficiency for V or Higgs bosons with small boost and ensures that no efficiency is lost in the transition from the classical reconstruction in two small jets at low boson  $p_T$  to the reconstruction as a single large-cone jet at higher values. Another point to consider when choosing the value of  $R$ , is the  $t\bar{t}$  data sample available for validating highly boosted W jets (Section 7.2). If  $R$  is chosen too large, the b quark from the

$mathrm{t} \rightarrow W$

$mathrm{mb}$  decay tends to merge into the W jet. The chosen value of  $R$  is the result of a compromise between high efficiency for V or Higgs bosons with small boost and a sufficiently large sample of W jets in  $t\bar{t}$  data for validating the boosted boson jet identification procedure. Figure 6.13 shows the  $p_T$  range of W bosons for which the CA8 algorithm is efficient and compares this to the efficiency for reconstructing W bosons from two AK5 jets. Above a  $p_T$  of 200 GeV, the CA8 jet algorithm, used to identify W jets, becomes more efficient than the reconstruction of a W boson from two AK5 jets.

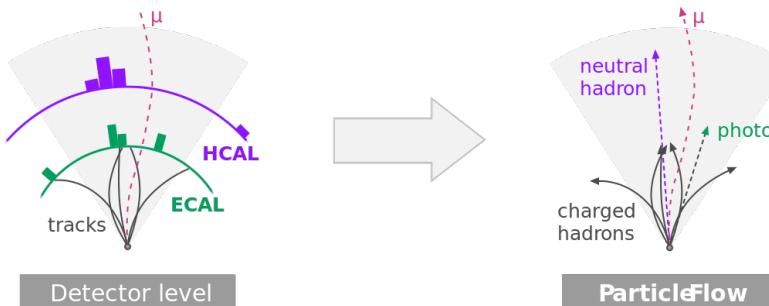
The AK5 or AK4 algorithms are used in this analysis to put requirements on additional b jets in the event selection (Section 8.1), along with the b tagging algorithm described in Section 6.4.3.



**Figure 6.13:** Efficiency to reconstruct a CA8 jet within  $\Delta R < 0.1$  of a generated W boson, and the efficiency to reconstruct two AK5 jets within  $\Delta R < 0.1$  of the generated quarks in the W boson decay, as a function of the  $p_T$  of the W boson [68].

### 6.4.2 Jet reconstruction and calibration

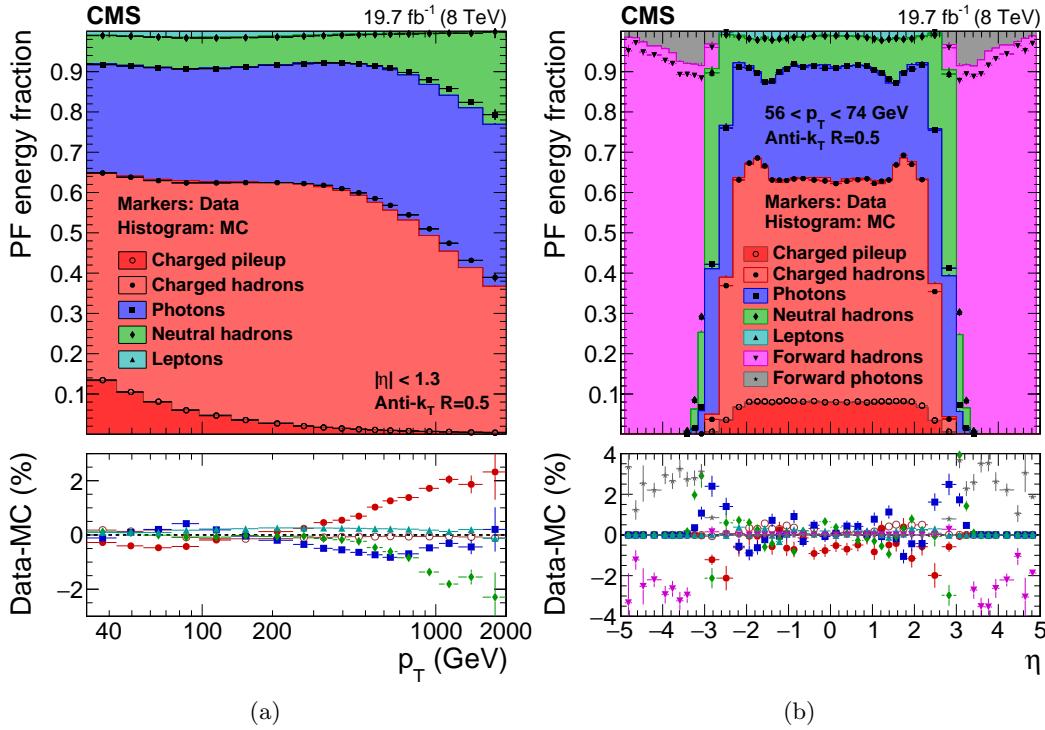
In CMS several standard methods for jet reconstruction are available which make use of different detector components, e.g. the tracker and the calorimeters, and give different reconstructed objects as input to the above explained jet clustering algorithms. In this work, only jets reconstructed with the PF algorithm are used and referred to as “PF jets”. As sketched in Fig. 6.14, the PF algorithm aims at reconstructing all the stable particles produced in an event, combining the information coming from all CMS sub-detectors to optimize particle identification, direction and energy determination. These particles are classified in several types: charged hadrons, photons, neutral hadrons, electrons and muons. Jets are typically composed by 65% charged hadrons, 25% photons, 10% neutral hadrons (Fig. 6.15). The PF algorithm is optimized to identify all these different components inside the jet, contrary to a calorimetric-only reconstruction. Typically, photons correspond to ECAL deposits not compatible with a tracker track. Charged hadrons correspond to HCAL and/or ECAL deposits matched to a inner track and not compatible with an electron, whereas neutral hadrons are identified as HCAL deposits not matched to any track. While the momentum of neutral particles is measured in the calorimeters, the momentum of charged particles is measured by the tracker with a better resolution. Hence, both the position and energy measurements are greatly improved with respect to calorimeter jets as this algorithm makes use of the tracking detectors and high granularity of the ECAL which is much higher than that of the HCAL. Once all the PF candidates in the event are reconstructed, they are used as input to the jet clustering algorithms described in the previous section and a PF jet is formed.



**Figure 6.14:** Sketch of the CMS particle-flow algorithm.

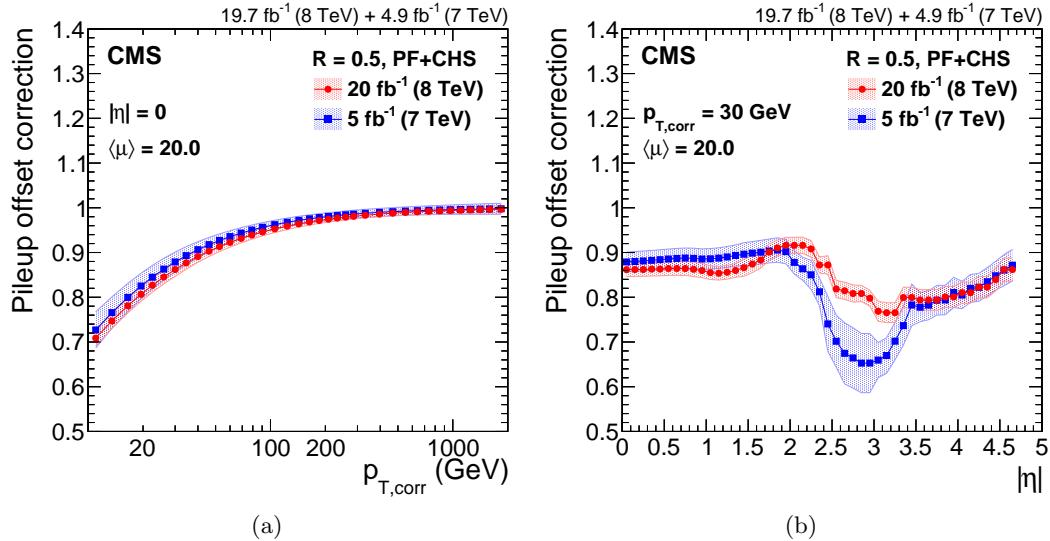
The jet momentum is determined as the vectorial sum of all PF candidates in the formed jet providing its “raw” estimate. At this stage, the reconstructed jet energy has a large uncertainty due to the several intrinsic limitations of the system, such as the non-linear response of the calorimeters, the detector segmentation, the presence of material in front of calorimeters, electronic noise and pileup. The raw jet energy and resolution are thus corrected for several factors in order to obtain the energy value as close as possible to the true energy of the initial parton. CMS has adopted a factorized approach [70] to the problem of jet energy corrections, where each level of correction takes care of a different effect as described in the following.

The first step in this approach is a correction to the jet energies to mitigate pileup effects. The additional pp collisions occurring within the same bunch-crossing as the primary hard interaction produce additional tracks in the tracker and deposit energy in the calorimeters. This contribution is usually referred to as in-time pileup. Due to the finite signal decay time



**Figure 6.15:** PF jet composition in data and simulation as a function of jet  $p_{\text{T}}$  for jets with  $|\eta| < 1.3$  (a), and as a function of  $\eta$  for jets with  $p_{\text{T}}$  in the range  $56 < p_{\text{T}} < 74 \text{ GeV}$  (b) [69].

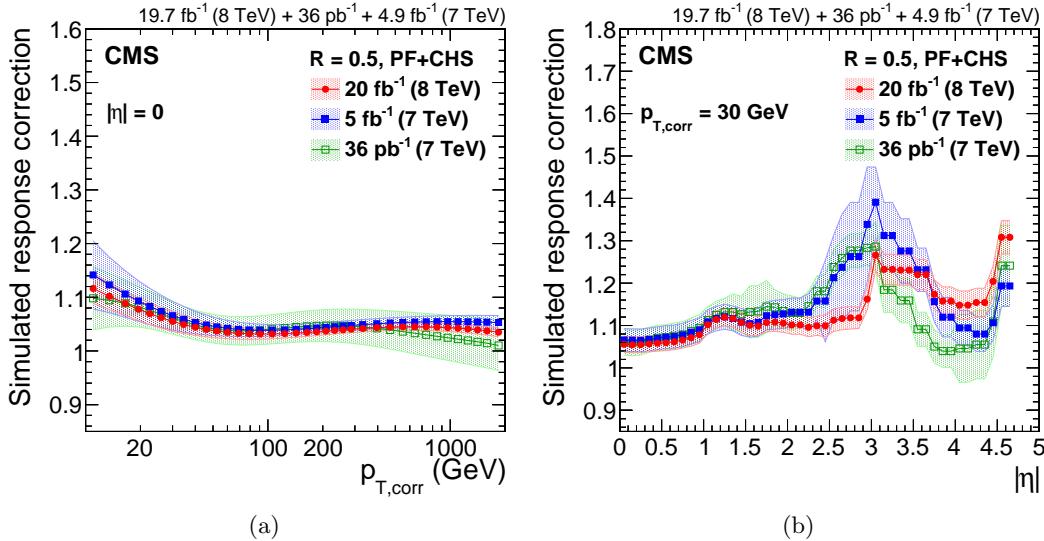
in the calorimeters, the pp collisions occurring in the previous and subsequent beam crossings also contribute to calorimetric energy in the same time window as the primary hard interaction. This contribution is called out-of-time pileup. The out-of-time contribution is mitigated at the level of signal processing, while the in-time one is partially removed using tracking information. This is achieved identifying which vertex the charged PF candidates originate from, and removing those unambiguously associated with pileup vertices before clustering jets. This method is referred to as *charged-hadron subtraction* (CHS), and represents the reference standard method for jet reconstruction in CMS for Run 1 and beginning of Run 2. For the second part of Run 2, other pileup mitigation techniques in addition to CHS have been developed and tested in CMS [71–73] but they are not used in this work. The CHS jets are then corrected to subtract residual contributions from neutral pileup particles, overlapping inside the jet cone. These corrections are determined from the simulation of a sample of QCD dijet events processed with and without pileup contaminations. This correction is usually parametrized as a function of the pileup energy density ( $\rho$ ) [74, 75], the jet area ( $A$ ) [76], jet  $p_{\text{T}}$  and  $\eta$ . The pileup offset corrections, defined as the mean value of the difference between the  $p_{\text{T}}$  of the reconstructed jet in events with and without pileup contamination, for AK5 CHS jets as a function of the corrected jet  $p_{\text{T}}$  and  $\eta$  are shown in Fig. 6.16, estimated for typical 2012 (8 TeV) conditions with an average number of additional pileup interactions  $\langle \mu \rangle = 20$ . The typical offset correction for a AK5 jet without CHS is 0.75 for a corrected jet  $p_{\text{T}}$  of 30 GeV, while a correction of 0.85 is obtained for AK5 CHS jets with same  $p_{\text{T}}$  value. This indicates that CHS removes approximately half of this offset before jet clustering by matching tracks to pileup vertices, reducing the residual offset correction. Roughly one third of the remaining pileup is from PF charged hadrons that have not been matched to good pileup vertices, and much of the rest is from PF photons.



**Figure 6.16:** Pileup offset correction for AK5 CHS jets estimated for the typical 2012 condition of  $\langle\mu\rangle = 20$ . Corrections are shown for jets at  $|\eta| = 0$  as a function of the corrected jet  $p_T$  (a), and for jets with  $p_T = 30 \text{ GeV}$  as a function of the jet  $|\eta|$  (b) [69].

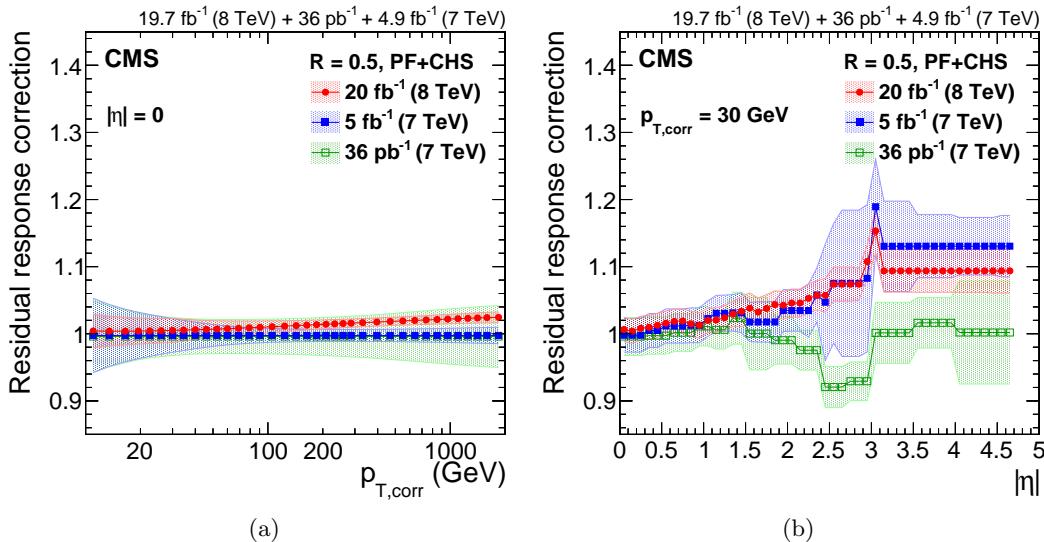
1671 Secondly, a simulation driven jet energy response correction is applied. The detector  
 1672 simulation takes into account effects due to particles deflected by the magnetic field, energy lost  
 1673 when traversing the detector material, particle conversions, and a detailed detector geometry.  
 1674 In this step the aim is to correct for non-uniformities in the different CMS subdetectors, by  
 1675 comparing the reconstructed jet  $p_T$  to the particle-level one using simulated events only. The  
 1676 corrections are derived as a function of jet  $p_T$  and  $\eta$  and make the response uniform over  
 1677 these two variables. The simulated particle response corrections are summarized in Fig. 6.17  
 1678 for 7 and 8 TeV data. The response is quite flat at  $p_T > 50 \text{ GeV}$ , where the competing  
 1679 effects of increasing calorimeter response and falling tracking efficiency within the jet core  
 1680 compensate each other. In the barrel and endcap regions, the corrections rise with  $|\eta|$ , due to  
 1681 the increasing amount of material located in front of the calorimeters, which leads to effects  
 1682 such as an increased rate of nuclear interactions in the tracker. The corrections are higher  
 1683 around  $|\eta| = 1.3$  and 3.0 due to the degradation of the response in the transition regions.

1684 Finally data-driven residual corrections are applied to correct for any measurable difference  
 1685 between the detector simulation and the jets measured in data. This correction is done in  
 1686 two steps. At first, an additional correction for the non homogeneous response of the detector  
 1687 with  $\eta$  is derived from dijet events, in which the  $p_T$  response of a probe jet, outside the barrel  
 1688 region, is balanced to the one in the reference tag region ( $|\eta| < 1.3$ ) as a function of the  
 1689 average  $p_T$  of the dijet system. Only events with back-to-back dijets and little additional  
 1690 activity in the event are used, to avoid any impact from unbalanced events. The jet energy  
 1691 is calibrated as a function of transverse momentum using a combination of  $Z(\rightarrow \ell\ell) + \text{jet}$ ,  
 1692  $\gamma + \text{jet}$ , and multijet events for jets in the reference barrel region ( $|\eta| < 1.3$ ). The basic idea,  
 1693 in all the considered topologies, is to exploit the transverse momentum balance between  
 1694 the jet to be calibrated and a well reconstructed and calibrated reference object ( $Z$  or  $\gamma$ ).  
 1695 The jet energy response is studied using two approaches. In one method the jet response is  
 1696 evaluated by comparing the reconstructed jet momentum ( $p_{T,\text{jet}}$ ) directly to the momentum  
 1697 of the reference object ( $p_{T,\text{ref}}$ ), while the second, more advanced, method takes into account  
 1698 the missing energy measured in the calorimeters to balance the reference object and jet  
 1699 momenta. In this method the additional event activity is taken into account by the missing



**Figure 6.17:** Detector response correction factors for AK5 CHS jets estimated for the 8 TeV data collected in 2012. Corrections are shown for jets at  $|\eta| = 0$  as a function of the corrected jet  $p_T$  (a), and for jets with  $p_T = 30 \text{ GeV}$  as a function of the jet  $|\eta|$  [69].

1700 energy. Therefore, additional jets in the event have only a small impact on the measurement.  
 1701 The residual corrections are summarized in Fig. 6.18 for 8 TeV data. The residual response  
 1702 corrections are less than 3% in the barrel, less than 10% in the endcaps, and about 10% in  
 1703 the forward detector.

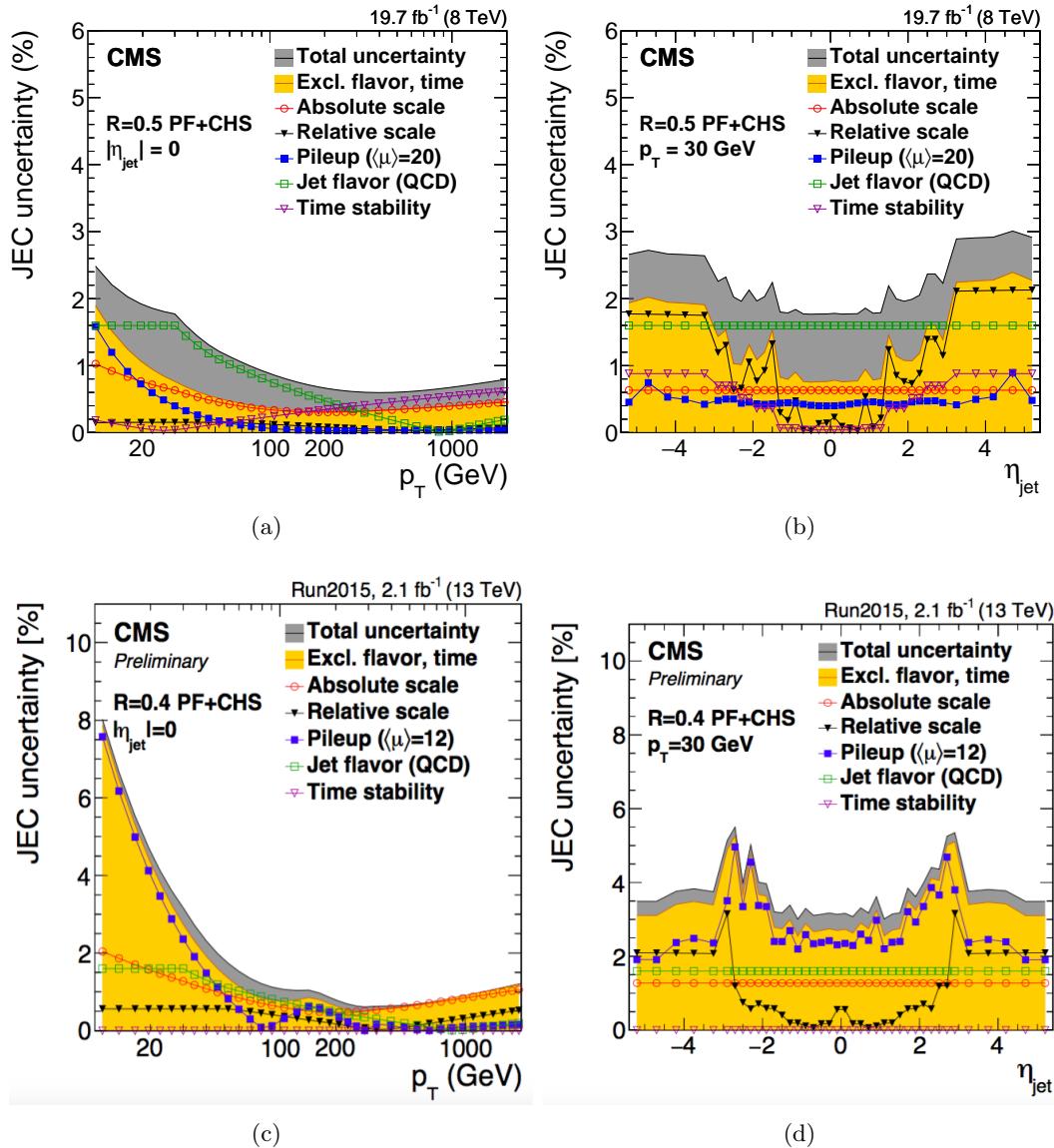


**Figure 6.18:** Residual data/simulation response correction factors for AK5 CHS jets for the 8 TeV data collected in 2012. Corrections are shown for jets at  $|\eta| = 0$  as a function of the corrected jet  $p_T$  (a), and for jets with  $p_T = 30 \text{ GeV}$  as a function of the jet  $|\eta|$  [69].

1704 The fully calibrated PF jets are finally obtained in both data and simulation by multiplying  
 1705 all the above correction factors to the raw jet  $p_T$  as follows:

$$p_{T,\text{corr}} = p_{T,\text{raw}} \times C_{\text{pu}}(p_{T,\text{raw}}, \eta, \rho, A) \times C_{\text{sim}}(C_{\text{pu}} \cdot p_{T,\text{raw}}, \eta) \times C_{\text{res}}(C_{\text{pu}} \cdot C_{\text{sim}} \cdot p_{T,\text{raw}}, \eta) \quad (6.2)$$

where  $C_{\text{pu}}$  represents the pileup correction,  $C_{\text{sim}}$  is the simulated response correction and  $C_{\text{res}}$  is the global residual correction applied only on jets in data. Figure 6.19 shows the overall uncertainty on the corrections to the jet energy scale for AK5 and AK4 CHS jets for 8 and 13 TeV data, respectively. In both cases, the final uncertainties are below 3% across the phase space of this analysis.



**Figure 6.19:** Summary of jet energy scale systematic uncertainties for the 8 TeV data collected in 2012 for AK5 CHS jets (upper plots) and for the 13 TeV data collected in 2015 for AK4 CHS jets (lower plots). Uncertainties are shown for jets at  $|\eta| = 0$  as a function of the corrected jet  $p_T$  (left), and for jets with  $p_T = 30$  GeV as a function of the jet  $|\eta|$  (right) [69, 77].

The energy resolution of jets is relatively poor compared to the resolution of other physics objects (electrons, muons, photons), and the biases caused by jet resolution smearing

is important for steeply falling spectra and for resonance decays. Hence, calibrations are evaluated to correct the jet energy resolution in addition to the corrections to the jet energy scale described above. The measurements are performed with methods which are extensions of the methods used for measuring jet energy scales, but instead of looking at the mean of the response distribution, the width is the interesting parameter. Furthermore, corrections have to compensate for effects that do not produce an overall shift in the mean, but that can widen the distribution. As shown in Fig. 6.20, the jet energy resolution in data is worse than in the simulation by 10–20% depending on  $\eta$ , and the jets in simulation need to be smeared accordingly.

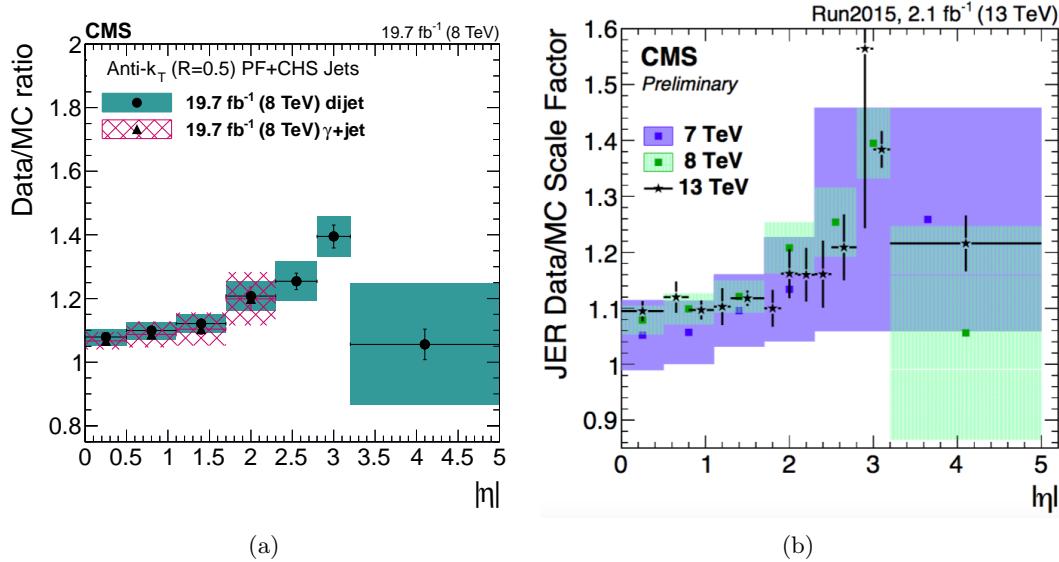
Jets used in this analysis are requested to pass loose identification criteria, in order to reject spurious jet-like features originating from isolated noise patterns in the calorimeters or the tracker. The efficiency of these requirements is above 99% for real jets.

For the 13 TeV data analysis described in this work, all AK4 and AK8 jets must have corrected  $p_T > 30 \text{ GeV}$  and  $> 200 \text{ GeV}$ , respectively, and  $|\eta| < 2.4$  to be considered in the subsequent steps of the analysis. Furthermore, the AK4 and AK8 jets are required to be separated from any well-identified muon or electron (Sections 6.3 and 6.2) by  $\Delta R > 0.3$  and  $> 0.8$ , respectively. This requirement is applied to clean the jet collection used in the analysis from leptons mis-identified as jets. Finally, AK4 jets are required to be separated from the AK8 jet representing the  $V \rightarrow q\bar{q}^{(\prime)}$  candidate by  $\Delta R > 0.8$  since an overlap is expected between the two reconstructions. The same selections are applied for AK5 and CA8 jets in the 8 TeV data analysis. For this case an additional selection is applied to the pseudorapidity of CA8 jets. In particular, CA8 jets are not used in the analysis if their pseudorapidity falls in the region  $1.0 < |\eta| < 1.8$ , thus overlapping the barrel-endcap transition region of the silicon tracker. In fact, in Run 1 it has been found that in this region, ‘noise’ can arise when the tracking algorithm reconstructs many fake displaced tracks associated with the jet. This issue in the reconstruction has been studied in details in the context of this work. The studies, presented and discussed in Appendix A, resulted in the choice of the  $\eta$  region to be excluded. In particular, the simulation does not sufficiently describe the full material budget of the tracking detector in that region, thus it does not accurately describe this effect. Without this requirement, a bias can be introduced in the b tagging, jet substructure and missing energy information, making this analysis systematically prone to that noise. As a consequence of these results, other analyses involving similar kinematic cuts and identification algorithms have been affected [78]. However, this problem has been fixed for Run 2 and this additional fiducial cut does not have to be applied in 13 TeV data analyses.

#### 6.4.3 Identification of b jets

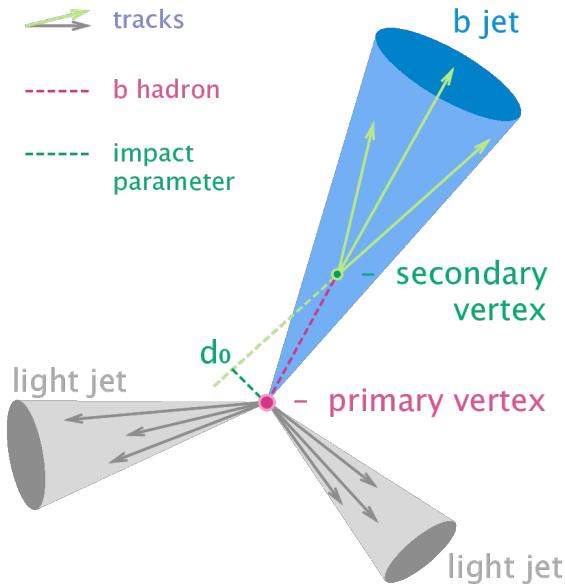
The identification of jets originating from b quarks (“b jets”) is one of the key ingredients of the analysis described in this work, which aims at isolating events of new physics with H bosons decaying to  $b\bar{b}$ . The ability to identify b jets (“b tagging”) plays a crucial role in reducing background coming from processes involving jets from gluons and light-flavor quarks (u, d, s), and from c quark fragmentation.

Identifying b jets relies on the properties of the production and the weak decay of b hadrons. The most important property is the relatively long lifetime of b hadrons of about 1.5 ps ( $c\tau \equiv 450\mu\text{m}$ ) corresponding to a flight distance that is observable with high resolution tracking detectors. A b hadron with  $p_T = 50 \text{ GeV}$  covers, on average, almost half a centimetre ( $Lc \sim \gamma\tau$ ) before decaying. As shown in Fig. 6.21, this leads to secondary vertices displaced from the primary event vertex and charged particle tracks incompatible with the primary vertex with sizeable impact parameter. In addition, b hadrons have a large mass and large



**Figure 6.20:** Data/MC scale factors for the jet  $p_T$  resolution for AK5 CHS jets as a function of  $|\eta|$  determined from 8 TeV data collected in 2012 (a) and for AK4 CHS jets in 13 TeV data collected in 2015 (b) [69, 77].

1761 multiplicity of charged particles in the final state (about five charged particles on average  
 1762 per b hadron decay). Because of the hard b-fragmentation function, the b hadron in a b jet  
 1763 carries a large fraction of the jet energy. Since b and c hadrons may decay semileptonically,  
 1764 in about 20% (per lepton species) of the cases an electron or muon is produced inside a b jet,  
 1765 if both direct and cascade decays are taken into account.



**Figure 6.21:** Representation of a b hadron decay and reconstructed b jet in the transverse plane.

1766 A variety of algorithms has been developed in CMS [79] to identify b jets based on the the  
 1767 b hadron properties described above. These algorithms use low-level physics objects, mainly

1768 jets and charged tracks. Only the tracking detectors offer the spatial resolution needed to  
 1769 measure the properties of b hadron decays such as their significant flight path. Efficient track  
 1770 reconstruction, and in particular precise spatial reconstruction close to the interaction point,  
 1771 is thus the key ingredient. Some of these algorithms use just a single observable, while others  
 1772 combine several of these objects to achieve a higher discrimination power. Each of these  
 1773 algorithms yields a single discriminator value for each jet. The minimum thresholds on these  
 1774 discriminators define loose (“L”), medium (“M”), and tight (“T”) operating points with a  
 1775 misidentification probability for light-flavor jets of 10%, 1%, and 0.1%, respectively, at an  
 1776 average jet  $p_T$  of about 80 GeV.

1777 The jets used for b tagging are reconstructed with the PF algorithm and calibrated as  
 1778 described in Section 6.4.2. A sample of well-reconstructed tracks of high purity is required  
 1779 as input to each of the b tagging methods. In addition to selection applied in the iterative  
 1780 tracking procedure described in Section 6.1.1, specific requirements are imposed:

- 1781 • the fraction of misreconstructed or poorly reconstructed tracks is reduced by requiring  
 $p_T > 1 \text{ GeV}$ ;
- 1783 • at least 8 tracker hits (including pixel) must be associated with the track;
- 1784 • at least 2 hits are required in the pixel system since track measurements in the innermost  
 $1785 \text{ layers provide most of the discriminating power;}$
- 1786 • the normalised  $\chi^2$  is required to be  $< 5$  to ensure a good-quality fit;
- 1787 • the absolute value of the transverse and longitudinal impact parameter of the track must  
 $1788 < 0.2$  and  $< 17 \text{ cm}$ , respectively, to reject charged particle tracks having their origin  
 $1789 \text{ from sources with large displacement from the primary vertex (e.g. photon conversions}$   
 $1790 \text{ and nuclear interactions in the beam pipe or the first layers of the pixel detector);}$
- 1791 • tracks are associated to jets in a cone  $\Delta R < 0.3$  around the jet axis, where the jet axis  
 $1792 \text{ is defined by the primary vertex and the direction of the jet momentum;}$
- 1793 • in order to reject tracks from pileup the distance to jet axis, defined as the distance of  
 $1794 \text{ closest approach of the track to the axis, is required to be } < 700 \mu\text{m};$
- 1795 • the point of closest approach between the track trajectory and the jet axis, must be  
 $1796 \text{ within 5 cm of the primary vertex.}$

1797 Two algorithms for reconstructing secondary vertices are exploited. For the first algorithm,  
 1798 the tracks associated to jets and fulfilling the above selection requirements are used in the  
 1799 *adaptive vertex reconstruction* (AVR) algorithm [80] based on the adaptive vertex fitter  
 1800 described in Section 6.1.1. This is the secondary vertex reconstruction algorithm used for b  
 1801 tagging methods in CMS during Run 1. A number of selection criteria are applied to remove  
 1802 vertices that are less likely to originate from a b hadron decay.

- 1803 • at least 2 tracks must be associated to the secondary vertex;
- 1804 • the fraction of tracks shared with the primary vertex is required to be  $< 65\%$ ;
- 1805 • the distance between the primary vertex to the secondary vertex in the transverse plane,  
 $1806 \text{ the 2D flight distance, must be in the range } 0.1\text{--}25 \text{ mm;}$
- 1807 • the 2D flight distance divided by its uncertainty or so-called 2D flight distance signifi-  
 $1808 \text{cance has to be } > 3;$

- 1809     • the invariant mass of charged particles associated to the vertex is required to be  
1810         $< 6.5 \text{ GeV}$  and not compatible with the mass of the  $K_S^0$  hadron in a window of 50 MeV;
- 1811     • the angular distance  $\Delta R$  between the jet axis and the secondary vertex flight direction  
1812        is required to be less than the jet distance parameter;

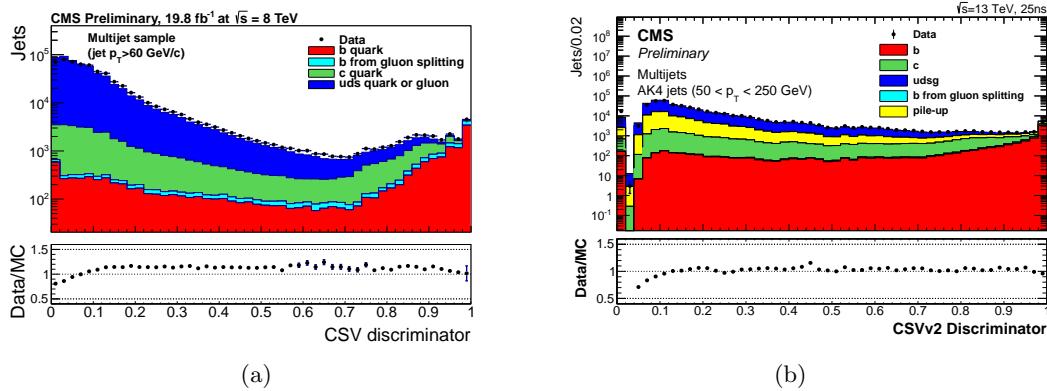
1813     In contrast with the AVR algorithm, the *inclusive vertex finder* (IVF) [81], is not seeded  
1814     from tracks associated to the reconstructed jets. The IVF algorithm uses as input the  
1815     collection of reconstructed tracks in the event and looser quality criteria are applied. The  
1816     selected tracks are then used to identify clusters of nearby tracks based on their minimum  
1817     distance and the angles between them. The clusters are fitted with the adaptive vertex fitter  
1818     and a cleaning procedure is applied. At this stage, tracks can appear in multiple vertices and  
1819     therefore, one of the vertices is removed based on the number of shared tracks and distance  
1820     between the vertex and another one. Furthermore, tracks in the secondary vertex compatible  
1821     with the primary vertex are removed. When there are at least 2 tracks associated to the  
1822     secondary vertex after the track arbitration, the vertex is refitted and selection criteria similar  
1823     to the case of the AVR vertices are applied.

1824     The efficiency to reconstruct a secondary vertex for b (c) jets using the IVF algorithm is  
1825     about 10% (15%) higher compared to the efficiency to reconstruct a secondary vertex with  
1826     the AVR algorithm. However, for light-flavour jets the probability to find a secondary vertex  
1827     also increases by about 8%. Independently of the jet flavour, around 60% of the jets with an  
1828     AVR vertex also have an IVF vertex.

1829     In this analysis the *Combined Secondary Vertex* (CSV) b tagging algorithm is used, which  
1830     combines the information of displaced tracks with the information of secondary vertices  
1831     associated to the jet. This allows the algorithm to avoid limitations due to inefficiencies in  
1832     the secondary vertex reconstruction. Jets are divided in three vertex-dependent exclusive  
1833     categories: the presence of a reconstructed secondary vertex; at least two tracks with impact  
1834     parameter significance larger than 2; none of the previous. The following set of variables with  
1835     high discriminating power and low correlations are considered:

- 1837     • the secondary vertex category;
- 1838     • the 2D flight distance significance of the secondary vertex;
- 1839     • the number of tracks in the jet
- 1840     • the number of tracks associated to the secondary vertex;
- 1841     • the secondary vertex mass;
- 1842     • the ratio of the energy carried by tracks at the vertex with respect to all tracks in the  
1843        jet;
- 1844     • the  $\eta$  of the tracks at the vertex with respect to the jet axis;
- 1845     • the 2D impact parameter significance of the first track that raises the invariant mass  
1846        above the charm threshold of 1.5 GeV when subsequently summing up tracks ordered  
1847        by decreasing impact parameter significance;
- 1848     • the 3D signed impact parameter significance for each track in the jet.

Two likelihood ratios are built from these variables used to discriminate between b and c jets and between b and light-flavor jets and combined with prior weights of 0.25 and 0.75, respectively. Figure 6.22(a) shows the distribution of the CSV discriminator value in a multijet sample for 8 TeV data and for simulation, for jets clustered with the AK5 algorithm.



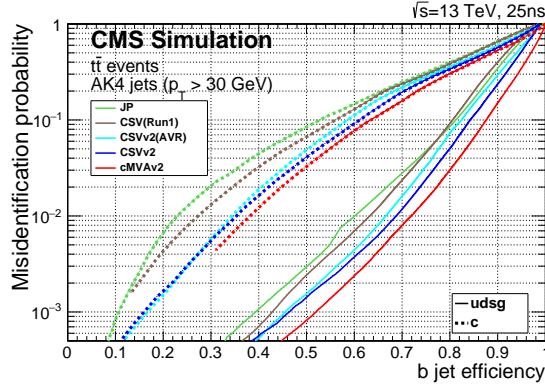
**Figure 6.22:** (a) Distribution of the CSV discriminator value in a multijet sample for data collected at 8 TeV and for simulation [82], for jets reconstructed with the AK5 algorithm. (b) Distribution of the CSVv2 discriminator value in a multijet sample for data collected at 13 TeV and for simulation, for jets reconstructed with the AK4 algorithm [83].

The CSV algorithm was further optimized for Run 2 and the new version is referred to as CSV version 2 (CSVv2) [83]. The main differences with respect to the Run 1 version of the CSV algorithm are the different vertex reconstruction algorithm used, the number of input variables and the way those are combined. In the newest version the input variables are combined using a multivariate technique. In fact, the likelihood-based method previously used, limited the amount of input variables since correlation between those could not be taken into account properly. In addition, the secondary vertex information is obtained with the IVF method described above. Figure 6.22(b) shows the distribution of the CSVv2 discriminator value in a multijet sample for 13 TeV data and for simulation, for jets clustered with the AK4 algorithm.

The performance of the CSVv2 tagger is presented in Fig. 6.23 as the b jet identification efficiency versus the misidentification probability for jets in simulated  $t\bar{t}$  events requiring jet  $p_T > 30$  GeV. A comparison is shown with the Run 1 version of the CSV algorithm trained for 8 TeV pp collisions using AK5 jets. The absolute improvement of the CSVv2 algorithm with respect to the CSV is of the order of 2 to 4% in b jet identification efficiency when comparing at the same misidentification probability for light-flavour jets. The improvement of using IVF vertices with respect to using AVR vertices in the CSVv2 algorithm is of the order of 1 to 2%.

The value of the discriminator threshold for the b tagging algorithms used in this analysis and the corresponding efficiencies are presented in Table 6.13. In this analysis the medium working point is used to identify and reject  $t\bar{t}$  events where a real b jet is expected in addition to the large-cone jet used to reconstruct the  $V \rightarrow q\bar{q}'$  or  $H \rightarrow b\bar{b}$  candidate, representing instead the signal. The same b tagging algorithm but together with the loose working point is used to identify whether the CA8 jet comes from a H boson decaying into bottom quarks, as described in Section 7.3.

The mismodelling of the b tagging variables in simulation is taken into account by

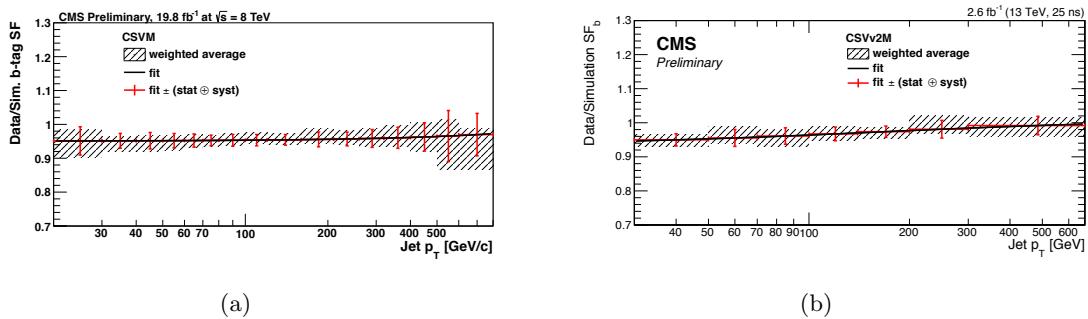


**Figure 6.23:** Performance of the CSVv2 algorithm showed as the probability for non-b jets to be misidentified as b jet as a function of the efficiency to correctly identify b jets. The improvement of this algorithm with respect to the Run 1 version is also shown [83].

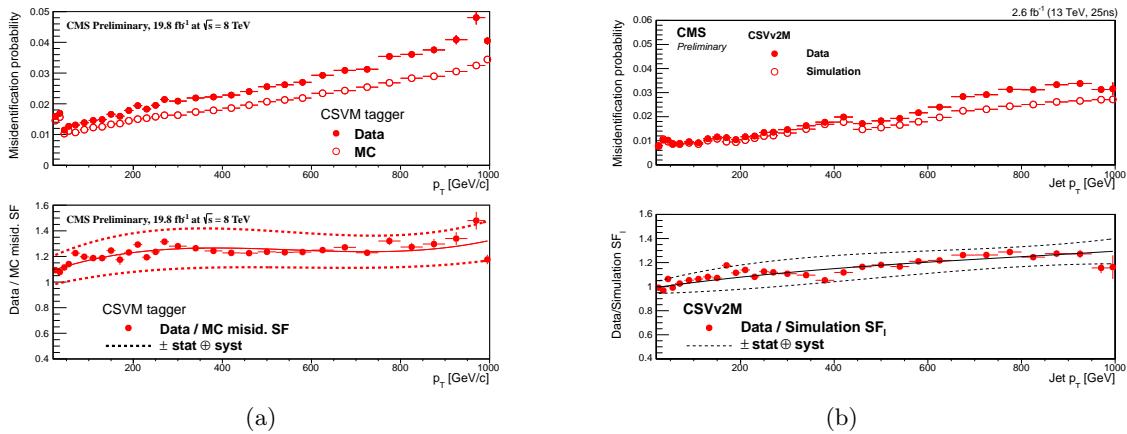
**Table 6.13:** B taggers and discriminator threshold used in CMS for Run 1 and Run 2 and corresponding efficiency for b jets with  $p_T > 30$  GeV in simulated  $t\bar{t}$  events.

Algorithm	operating point	discriminator value	b tagging efficiency (%)
CSV (Run 1)	CSVL	0.244	84
	CSVM	0.679	68
	CSVT	0.898	52
	CSVv2L	0.460	83
CSVv2 (Run 2)	CSVv2M	0.800	69
	CSVvsT	0.935	49

1880 reweighting simulation event-by-event with the ratio of the b tagging efficiency in data and  
 1881 simulation, determined in a sample enriched with b jets and depending on the jet  $p_T$  and  $\eta$ .  
 1882 The correction factors as a function of the b jet  $p_T$  are shown in Fig. 6.24(a) and 6.24(b) for  
 1883 the CSVM and CSVv2M operating points respectively, as measured in 8 and 13 TeV data. In  
 1884 a similar way, correction factors are also derived and applied to correct the misidentification  
 1885 probability in simulation. These factors are shown in Fig. 6.25(a) and 6.25(b) as a function  
 1886 of the jet  $p_T$  for the CSVM and CSVv2M operating points.



**Figure 6.24:** Data-to-simulation correction factors for the b tagging efficiency for the CSVM (a) and CSVv2M (b) algorithms as a function of the b jet  $p_T$  as measured in 8 and 13 TeV data [82, 83].



**Figure 6.25:** Data-to-simulation correction factors for the misidentification probability for the CSVM (a) and CSVv2M (b) algorithms as a function of the jet  $p_T$  as measured in 8 and 13 TeV data [82, 83].

## 6.5 Missing transverse energy

CMS is a full coverage hermetic detector which identifies and reconstructs almost all stable or long-lived particles produced in pp collisions. The only exceptions are neutrinos and hypothetical neutral weakly interacting particles. Although these particles do not leave a signal in the detector, their presence can be inferred from the momentum imbalance in the transverse plane, a quantity known as missing transverse momentum and denoted by  $\vec{p}_T^{\text{miss}}$ .

Several standard methods are available in CMS for the reconstruction of  $\vec{p}_T^{\text{miss}}$ , which, as for the jet reconstruction, can be based on calorimeter information only, include also tracker information, or use the PF algorithm [84]. In this analysis, the PF  $\vec{p}_T^{\text{miss}}$  is used along with PF jets and it is calculated as the negative vector sum of the transverse momenta of all reconstructed PF candidates in a given event

$$\vec{p}_T^{\text{miss}} = - \sum_i^N \vec{p}_{T,i}. \quad (6.3)$$

Its magnitude is referred to as missing transverse energy and denoted by  $E_T^{\text{miss}}$ . The  $E_T^{\text{miss}}$  is an important variable in many searches for physics beyond the standard model such as the ones described in this thesis where a real highly energetic neutrino is expected in the final state. In addition, the precise measurement of  $E_T^{\text{miss}}$  plays a crucial role for measurements of standard model physics involving W bosons and top quarks. The  $\vec{p}_T^{\text{miss}}$  reconstruction is sensitive to pileup, detector malfunctions and to various reconstruction effects and a precise calibration of all reconstructed physics objects is crucial for its performance. The level of mismeasurement is significantly reduced after jet energy calibration, described in Section 6.4.2. A correction to the  $\vec{p}_T^{\text{miss}}$  is derived by propagating the jet energy scale corrections as described in the following.

The raw missing transverse momentum can be written as:

$$\vec{p}_T^{\text{miss,raw}} = - \sum_i^{N_{\text{jets}}} \vec{p}_{T,i}^{\text{raw}} - \sum_i^{N_{\text{uncl}}} \vec{p}_{T,i}, \quad (6.4)$$

where the first and second sum runs over the  $p_T$  of the PF candidates clustered as jets and unclustered, respectively, and the superscript ‘‘raw’’ indicates the uncorrected value. The

correction to the  $\vec{p}_T^{\text{miss}}$  is then obtained by replacing the first sum with the vector sum of the transverse momenta of the jets to which jet energy scale corrections (JEC) are applied:

$$\vec{C}_T^{\text{JEC}} = \sum_i^{N_{\text{jets}}} \vec{p}_{T,i}^{\text{raw}} - \sum_i^{N_{\text{jets}}} \vec{p}_{T,i}^{\text{JEC}}, \quad (6.5)$$

where the sum is performed over all jets with corrected  $p_T > 10 \text{ GeV}$ .

Further corrections improve the performance of the  $\vec{p}_T^{\text{miss}}$  reconstruction in events with large numbers of pileup interactions. This is achieved as explained in the following.

The raw  $\vec{p}_T^{\text{miss}}$  can be written as a sum of the two contributions due to particles produced in the primary vertex (PV) and in pileup interactions (PU)

$$\vec{p}_T^{\text{miss,raw}} = - \sum_{i \in \text{PV}} \vec{p}_{T,i} - \sum_{i \in \text{PU}} \vec{p}_{T,i}. \quad (6.6)$$

Particles produced in the pileup interactions can be further classified into neutral (PUneu) and charged (PUch) particles so that the equation above can be expressed as

$$\vec{p}_T^{\text{miss,raw}} = - \sum_{i \in \text{PV}} \vec{p}_{T,i} - \sum_{i \in \text{PUch}} \vec{p}_{T,i} - \sum_{i \in \text{PUneu}} \vec{p}_{T,i}. \quad (6.7)$$

The contribution to the genuine  $\vec{p}_T^{\text{miss}}$  from such interactions is close to zero, as the probability to produce neutrinos is small in inelastic pp scattering interactions (e.g. neutrinos from Kaon decays). The vectorial  $\vec{p}_T$  sum of charged particles is therefore expected to be well balanced by that of neutral particles. However, the nonlinearity and minimum energy thresholds in the calorimeters cause  $\vec{p}_T^{\text{miss}}$  to point on average in the direction of the vectorial  $\vec{p}_T$  sum of neutral particles. Nevertheless, it can be assumed that the directions of neutral pileup particles is measured with high precisions from the positions of the calorimeter cells in which we observe the energy deposits, while their energies are systematically off by the same factor. At the same time, the CMS tracker can also measure very well the charged pileup particles from their large curvature due to the low  $p_T$  characterizing this type of processes. With these assumptions, the total contribution from pileup can be estimated as

$$\vec{\Delta}_{\text{PU}} = \sum_{i \in \text{PUch}} \vec{p}_{T,i} + \sum_{i \in \text{PUneu}} \vec{p}_{T,i} = \sum_{i \in \text{PU}} f(\vec{v}) \vec{v}, \quad (6.8)$$

where  $\vec{v}$  represents the sum of the transverse momenta of charged particles for each pileup interaction. The correction  $f(\vec{v})$  is parametrized as  $f(\vec{v}) = c_1(1.0 + \text{erf}(-c_2|\vec{v}^{c_3}|))$ , where the coefficients  $c_1$ ,  $c_2$ , and  $c_3$  are extracted from simulated minimum bias events. The corrected  $\vec{p}_T^{\text{miss}}$  is then obtained removing the additional contribution  $\vec{\Delta}_{\text{PU}}$  from Eq. 6.6

$$\vec{p}_T^{\text{miss,PUcorr}} = \vec{E}_T^{\text{miss,raw}} + \vec{\Delta}_{\text{PU}}. \quad (6.9)$$

Another type of correction is derived and applied to correct for a modulation in  $\phi$  in the  $\vec{p}_T^{\text{miss}}$  present not only in data but also in simulation. The distribution of genuine  $\vec{p}_T^{\text{miss}}$  is instead independent of  $\phi$  because of the rotational symmetry of the collisions around the beam axis. The possible causes of the modulation include imperfect detector alignment, inefficiencies, a residual  $p_T$  dependence of the calibration, and a shift between the center of the detector and the beam line. The correction for this effect can be expressed as a shift in the  $\vec{p}_T^{\text{miss}}$  components along the  $x$  and  $y$  detector coordinates, which increases approximately linearly with the number of reconstructed vertices. This correlation is used for a correction procedure as follows

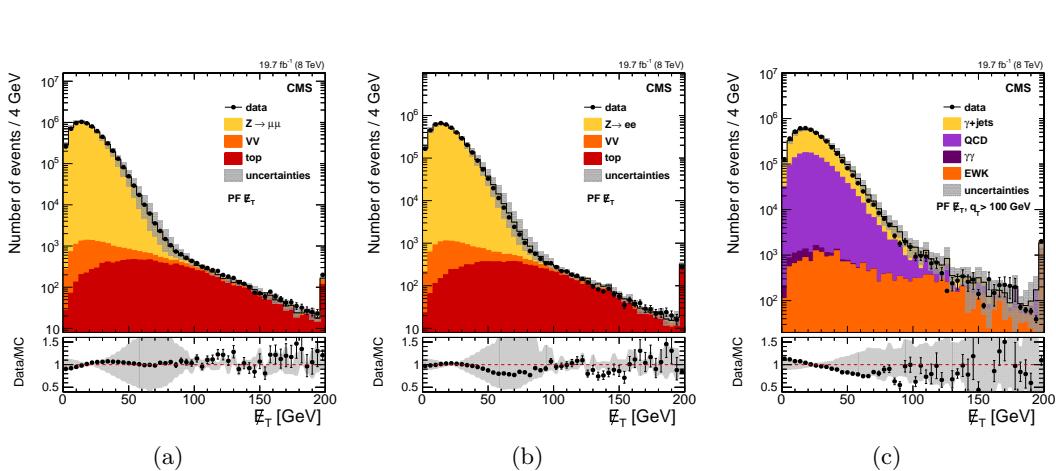
$$\vec{E}_{T,x}^{\text{miss,corr}} = \vec{E}_{T,x}^{\text{miss,raw}} - (c_{x_0} + c_{x_s} N_{\text{vtx}}), \quad \vec{E}_{T,y}^{\text{miss,corr}} = \vec{E}_{T,y}^{\text{miss,raw}} - (c_{y_0} + c_{y_s} N_{\text{vtx}}), \quad (6.10)$$

where the coefficients are determined separately for data and simulated events.

Other more sophisticated missing energy determinations aimed at improving the resolution have been developed in CMS [85, 86] but will not be discussed in this section since they are not used in this work.

The distributions of the PF  $E_T^{\text{miss}}$ , obtained after applying all the corrections described above, in  $Z \rightarrow \mu^+\mu^-$ ,  $Z \rightarrow e^+e^-$ , and prompt photon events are presented in Fig. 6.26 as measured in 8 TeV data and for simulation. Good agreement between data and simulation is observed in all distributions.

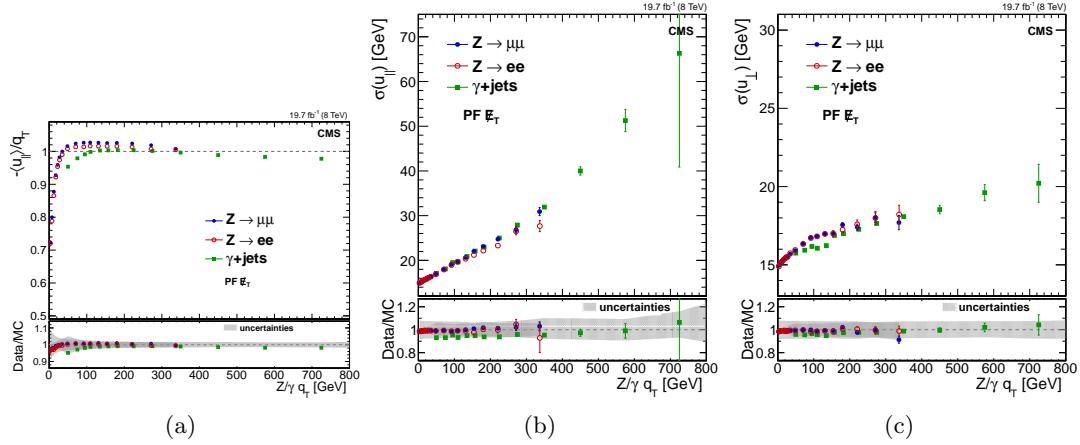
These events contain no genuine  $\vec{p}_T^{\text{miss}}$ , and thus a balance exists between the well-measured vector boson transverse momentum, denoted as  $\vec{q}_T$ , and the hadronic recoil, denoted as  $\vec{u}_T$ , which dominates the  $\vec{p}_T^{\text{miss}}$  measurement. The  $q_T$  can therefore be used as a reference to measure the scale and resolution of  $\vec{p}_T^{\text{miss}}$ . The hadronic recoil can be projected to the axis defined by  $q_T$ , yielding two signed components, parallel ( $u_{||}$ ) and perpendicular ( $u_{\perp}$ ) to this axis. The parallel component is typically negative as the observed hadronic system is usually in the hemisphere opposite the boson. The scalar quantity  $-\langle u_{||} \rangle / \vec{q}_T$  is referred to as the  $\vec{p}_T^{\text{miss}}$  response. The response curves, extracted from the data as a function of the vector boson boost  $\vec{q}_T$ , are shown in Fig. 6.27(a), where deviations from unity indicate a bias on the hadronic recoil energy scale which is fully recovered for  $\vec{q}_T > 40. The resolution curves,  $\sigma(u_{||})$  and  $\sigma(u_{\perp})$  as a function of  $q_T$ , are shown in Fig. 6.27(b) and 6.27(c), respectively, for each control sample. The resolution increases with increasing  $q_T$ , while the data and simulation curves are in good agreement for each control sample.$



**Figure 6.26:** The PF  $E_T^{\text{miss}}$  distribution in  $Z \rightarrow \mu^+\mu^-$  (a),  $Z \rightarrow e^+e^-$  (b), and prompt photon (c) events for 8 TeV data and for simulation. The points in the lower panel of each plot show the ratio between data and simulation describing their agreement [85].

## 6.6 $W \rightarrow \ell\nu$ reconstruction

The identified muon or electron (see Section 6.2.3 and 6.3.3) is associated with the  $W \rightarrow \ell\nu$  candidate. The  $\vec{p}_T$  of the undetected neutrino is assumed to be equal to the  $\vec{p}_T^{\text{miss}}$ . The



**Figure 6.27:** (a) Response curves for PF  $\vec{p}_T^{\text{miss}}$  in events with a Z-boson or prompt photon. Also shown are the resolution curves of the parallel (b) and perpendicular (b) recoil components as a function of the  $Z/\gamma q_T$ . In each plot the upper frame shows the response in 8 TeV data, while the lower one shows the ratio between data and simulation. [85].

longitudinal momentum of the neutrino ( $p_z$ ) is obtained by solving a quadratic equation that sets the  $\ell\nu$  invariant mass to the known W boson mass [87]:

$$M_W^2 = m_\ell^2 + 2(E_\ell E_\nu - p_{x_\ell} p_{x_\nu} - p_{y_\ell} p_{y_\nu} - p_{z_\ell} p_{z_\nu}) = (80.4)^2 \quad (6.11)$$

In the case of two real solutions, the one with the smaller absolute value is chosen. If the discriminant becomes negative, or equivalently the W boson transverse mass  $M_T$  is larger than  $M_W$  used in the constraint, the solutions have an imaginary part. This happens because of the finite resolution of  $E_T^{\text{miss}}$ . Several schemes exist to deal with this situation. One technically simple method consists of taking the real part of the complex solutions but it leads to the wrong W boson mass. This method is used for the reconstruction of the  $W \rightarrow \ell\nu$  candidate in the 13 TeV data analysis described in this work. A second method has been studied, which eliminates the imaginary component by modifying the components of the missing transverse energy such to give  $M_T = M_W$ , still respecting equation 6.11 [88]. This method is used in the 8 TeV data analysis for the reconstruction of the  $W \rightarrow \ell\nu$  candidate and for the reconstruction of the mass of the leptonically decaying top quark in  $t\bar{t}$  events. The performance of the two methods are equivalent in terms of resolution of the reconstructed diboson or top quark invariant mass.

The four-momentum of the neutrino is used to reconstruct the four-momentum of the  $W \rightarrow \ell\nu$  candidate. The same procedure holds also for the cases where the W boson decays to  $\tau\nu$  and the  $\tau$  decays to one muon or electron and two neutrinos. In this case, the  $\vec{p}_T^{\text{miss}}$  represents the  $\vec{p}_T$  of the three-neutrino system.

# Identification of highly boosted $W/Z \rightarrow q\bar{q}^{(\prime)}$ and $H \rightarrow b\bar{b}$

1992

1993 Large-cone jets (Section 6.4), also referred to as “fat jets”, are used to reconstruct the  $W$  jet,  
 1994  $Z$  jet, and  $H$  jet candidates resulting after the hadronization of the two quarks from the decay  
 1995 of highly boosted  $W$ ,  $Z$ , and Higgs boson, respectively. In fact, for the resonance mass range  
 1996 considered in this search, the two quarks from the  $V$  or Higgs boson decay would be separated  
 1997 by a small angle, resulting in the detection of a single “merged” jet after hadronization,  
 1998 instead of two separated jets. To discriminate against multijet backgrounds, the analysis  
 1999 exploits both the reconstructed jet mass, which is required to be close to the boson mass,  
 2000 and the jet substructure arising from the two jet cores that correspond to the two high- $p_T$   
 2001 decay quarks. The techniques used to identify jets arising from the merged decay products of  
 2002 a single  $V$  or Higgs boson are referred to as “ $V$  tagging” or “ $H$  tagging”, respectively. They  
 2003 employ novel jet substructure algorithms, which are described in Section 7.1. The features of  
 2004 the  $V$  tagging algorithm are described in Section 7.2 and its performance in both data and  
 2005 simulation are discussed. Finally, in Section 7.3, a procedure tuned to the specific properties  
 2006 of the Higgs boson decay into a bottom quark-antiquark pair is presented.

## 2007 7.1 Jet substructure observables

### 2008 7.1.1 Pruned jet mass

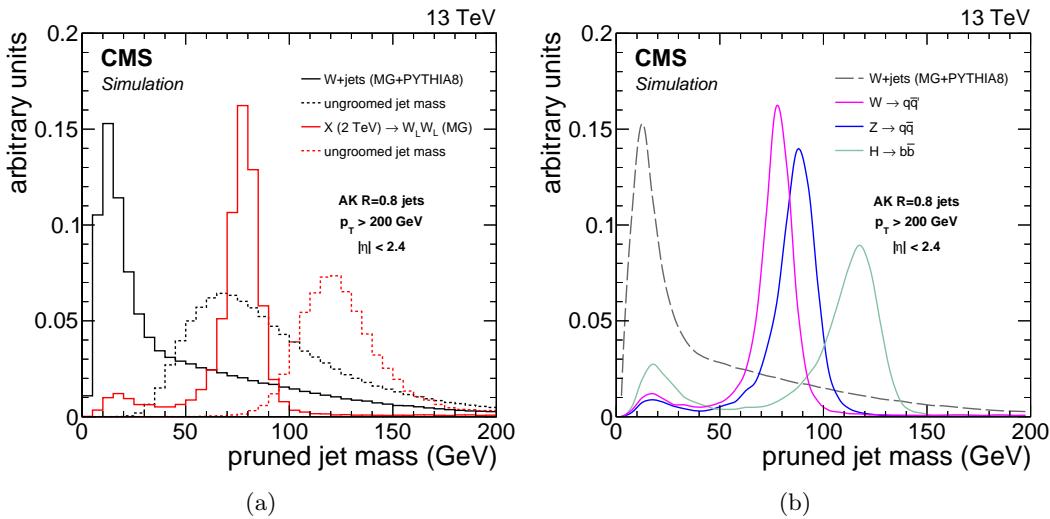
2009 As the mass of the  $V$  or Higgs boson is larger than the mass of a typical QCD jet, the jet mass  
 2010 is the primary observable that distinguishes them from a QCD jet. The bulk of the signal jet  
 2011 mass arises from the kinematics of the two jet cores that correspond to the two decay quarks.  
 2012 In contrast, the QCD jet mass arises mostly from large-angle and soft gluon radiation. As  
 2013 a first step in exploring potential substructure, the jet constituents are subjected to a jet  
 2014 grooming algorithm that improves the resolution in the jet mass and reduces the effect of  
 2015 pileup [68, 89]. The goal of jet grooming is to recluster the jet constituents, while applying  
 2016 additional requirements that eliminate soft, large-angle QCD radiation. This procedure shifts  
 2017 the jet mass of QCD jets to smaller values, while maintaining the mass for signal jets close  
 2018 to the boson mass. Furthermore, soft contributions from the underlying event and pileup,  
 2019 usually present in all jets, are removed. Different jet grooming algorithms have been explored  
 2020 at CMS and their performance on jets in multijet processes has been studied in detail [68, 89].  
 2021 In this analysis, the *jet pruning* algorithm [90, 91] is used, as it was found to provide the best  
 2022 discrimination against QCD background as discussed in Ref. [68, 89].

2023

2024 Jet pruning reclusters each fat jet starting from all its original constituents, through the  
 2025 implementation of the CA algorithm, but applying two additional conditions beyond those  
 2026 given in 6.1. In particular, the softer of the two particles  $i$  and  $j$  to be merged is removed  
 2027 when the following conditions are met:

$$z_{ij} \equiv \frac{\min(p_{Ti} + p_{Tj})}{p_{Ti} + p_{Tj}} < z_{cut}, \quad \Delta R_{ij} > D_{cut} \equiv \alpha \frac{m_j}{p_T} \quad (7.1)$$

where  $m_j$  and  $p_T$  are the mass and transverse momentum of the originally-clustered jet, and  $z_{cut}$  and  $\alpha$  are parameters of the algorithm, chosen to be 0.1 and 0.5, respectively. In this particular choice of parameters, the algorithm removes the largest number of jet constituents, and can therefore be regarded as the most aggressive jet grooming technique. The resulting jet is the *pruned jet*. The pruned jet mass,  $m_{jet}$ , is computed from the sum of the four-momenta of the constituents that survive the pruning; it is then corrected by the same factor used to correct the jet  $p_T$  (Section 6.4). Figure 7.1(a) illustrates the effect of pruning on AK8 jets: the  $m_{jet}$  spectrum of the  $W$  jet candidate from the decay of highly boosted and longitudinally polarized  $W$  bosons is shown together with the distribution in  $m_{jet}$  for the simulated background of  $W+jets$ . Dashed and solid lines correspond to the distributions before and after the application of the pruning algorithm, respectively. Fully merged jets reconstructed from the  $W$  boson decay generate a distinctive peak around the  $W$  boson mass, which is narrowed by the pruning, while background jets acquire a smaller mass on average, enhancing the discrimination. Figure 7.1(b) compares the distributions in  $m_{jet}$  for  $W$ ,  $Z$  and  $H$  jet candidates from the decay of highly boosted  $W$ ,  $Z$  and Higgs bosons, respectively. The distribution in  $m_{jet}$  for the  $W+jets$  background is also shown. Not-full-merged signal jets give rise to a peak at low masses.



**Figure 7.1:** (a) Distribution in pruned jet mass  $m_{jet}$  for simulated events of highly boosted  $W$  bosons and inclusive QCD jets expected in the  $W+jets$  process. The ungroomed jet mass is shown as dotted lines to illustrate the effect of pruning. MG denotes the MADGRAPH generator. (b) Comparison of the distributions in  $m_{jet}$  for simulated events of highly boosted  $V$  and Higgs bosons.

### 7.1.2 N-subjettiness

In addition to the pruned jet mass, additional information about the jet shape is used to discriminate the signal against jets from gluon and single-quark hadronization. This information can be obtained from the quantity called *N-subjettiness* [92]. It takes advantage of the multi-body kinematics in the decay pattern of boosted hadronic objects, and it can be used to effectively “count” the number of subjets in a given jet.

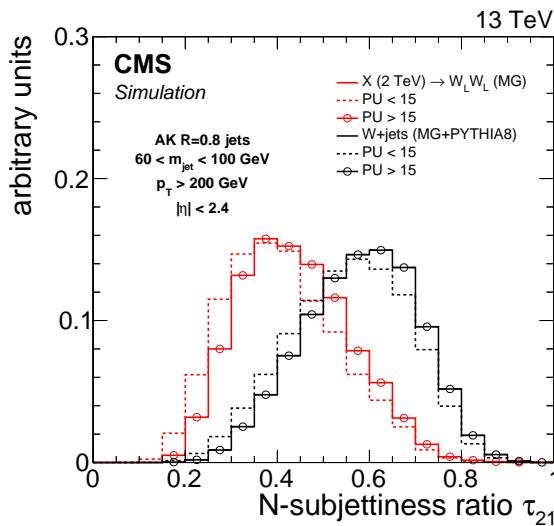
2051     The N-subjettiness is a generalized jet shape observable which defines a measure,  $\tau_N$ ,  
 2052    for a jet to have  $N$  subjets. The constituents of the jet before the pruning procedure are  
 2053    reclustered using the  $k_T$  algorithm (Section 6.4), until  $N$  joint objects (subjets) remain in  
 2054    the iterative combination procedure of the algorithm. The observable  $\tau_N$  is then defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\}, \quad (7.2)$$

2055    where  $k$  runs over the constituents of the jet, and the distances  $\Delta R_{n,k}$  are calculated  
 2056    relative to the axis of the  $n$ th subjet. The normalization factor  $d_0$  is taken as

$$d_0 = \sum_k p_{T,k} R_0, \quad (7.3)$$

2057    where  $R_0$  is the characteristic jet radius used in the original jet clustering algorithm. The  
 2058    subjet axes are obtained by running the exclusive  $k_T$  algorithm [64], and reversing the last  
 2059     $N$  clustering steps. The variable  $\tau_N$  quantifies the compatibility of the jet clustering with  
 2060    the hypothesis that exactly  $N$  subjets are present. Jets with  $\tau_N \approx 0$  have all their radiation  
 2061    aligned with the candidate subjet directions and therefore have  $N$  (or fewer) subjets. Jets  
 2062    with  $\tau_N \gg 0$  have a large fraction of their energy distributed away from the candidate subjet  
 2063    directions and therefore have at least  $N + 1$  subjets. The ratio between 2-subjettiness and  
 2064    1-subjettiness,  $\tau_{21} = \tau_2/\tau_1$ , is found to be a powerful discriminant between jets originating  
 2065    from hadronic V decays and from gluon and single-quark hadronization. Jets from  $V \rightarrow q\bar{q}^{(\prime)}$   
 2066    decays in signal events are characterized by lower values of  $\tau_{21}$  relative to QCD background.  
 2067    Figure 7.2 shows the  $N$ -subjettiness ratio  $\tau_{21}$  distribution for W jets and QCD jets after  
 2068    requiring  $60 < m_{\text{jet}} < 100$  GeV, demonstrating its discrimination power after the pruned jet  
 2069    mass selection.



**Figure 7.2:** Distribution in N-subjettiness ratio  $\tau_{21}$  for simulated events of highly boosted and longitudinally polarized W bosons and inclusive QCD jets expected in the W+jet process. The distributions are shown after a selection on the pruned jet mass requiring  $60 < m_{\text{jet}} < 100$  GeV. The histograms are the expected distributions after full CMS simulation with pileup corresponding to an average number above and below 15 interactions.

## 2070 7.2 The V tagging algorithm

2071 The jet substructure observables described in the previous section are employed for identifying,  
 2072 or “tagging”,  $W$  and  $Z$  jets (“V jets”). The V tagging of the jets is obtained combining  
 2073 selections on both the pruned jet mass  $m_{\text{jet}}$  and N-subjettiness ratio  $\tau_{21}$  observables.

2074 The selection criteria have been optimized in the context of searches for resonances  
 2075 decaying into diboson in the  $\ell\nu$ +jet and dijet final states [93–95]. The optimization, based  
 2076 on simulation, aims at maximizing the analysis sensitivity and it leads to slightly different  
 2077 working points for each analysis. In particular, the baseline selection values have been changed  
 2078 from Run 1 to Run 2. Typical signal efficiencies and mistagging rates of QCD jets obtained,  
 2079 respectively, from simulations and measurements with 8 and 13 TeV data are summarized in  
 2080 Table 7.1, for jets with  $p_T = 500$  GeV.

**Table 7.1:** Typical selection criteria for V tagging used in Run 1 and Run 2. The corresponding signal efficiency and mistagging rate of QCD jets are also reported for jets with  $p_T = 500$  GeV, obtained from 8 and 13 TeV data and from simulation.

Data sets	V tagging selections	signal efficiency	mistagging rate
8 TeV	$60 < m_{\text{jet}} < 100$ GeV $\tau_{21} < 0.5$	0.65	0.04
13 TeV	$65 < m_{\text{jet}} < 105$ GeV $\tau_{21} < 0.45$	0.55	0.03

2081 The  $\ell\nu q\bar{q}$  analysis described in this work makes use of a looser  $\tau_{21}$  working point of 0.6  
 2082 resulted from an optimization which takes into account signal efficiency and background  
 2083 rejection over a large jet  $p_T$  range. In fact, this channel is characterized by a low background  
 2084 rate and a  $\tau_{21}$  selection providing a higher signal efficiency over the whole jet  $p_T$  range is  
 2085 therefore preferred. This working point corresponds to a signal efficiency of 65% and a  
 2086 mistagging rate of 5%.

2087 The V tagging performance at 8 TeV has been studied in detail in Ref. [68]. From simulation  
 2088 studies it is observed that the efficiency of the  $m_{\text{jet}}$  selection increases with  $p_T$  up to about  
 2089 600 GeV since at higher  $p_T$  the showers from the  $W$  decay quarks are more likely to be  
 2090 reconstructed within a single fat jet. Above 600 GeV, the efficiency begins to decrease as  
 2091 a function of jet  $p_T$ , since at very large values the PF candidate reconstruction degrades  
 2092 in resolving the jet substructure, and the pruning algorithm therefore removes too large a  
 2093 fraction of the jet mass. For Run 2, the PF reconstruction has been optimized by exploiting  
 2094 the full potential of the CMS ECAL granularity to resolve jet substructure and a constant  
 2095 efficiency is maintained up to at least  $p_T = 2.5$  TeV [73, 96].

2096 The efficiency of the additional  $\tau_{21}$  selection also drops as a function of  $p_T$ , thus a fixed  
 2097 working point will degrade the efficiency with increasing  $p_T$ . However, the same efficiency  
 2098 at an equivalent background rejection rate can be reached by adjusting the  $\tau_{21}$  selection as  
 2099 a function of  $p_T$ . This possibility has not been explored yet in any of the searches which  
 2100 employ V tagging.

2101 The efficiency of the V tagging selection as a function of the number of reconstructed  
 2102 primary vertices (PV) has also been studied [73]. It is observed that the efficiency of the  
 2103  $m_{\text{jet}}$  selection is constant as a function of PV, whereas the additional  $\tau_{21}$  selection efficiency  
 2104 drops from 60% at 0 PV to 40% at 30 PV. However, the mistagging of the background also  
 2105 decreases with pileup for the same selection, yielding similar discrimination. Efficiency and  
 2106 mistagging rate are affected by pileup in the same way, since additional pileup shifts the  $\tau_{21}$   
 2107

2108 distribution towards higher values (towards background like) for both signal and background  
 2109 (Fig. 7.2). Therefore, the same signal efficiency can be reached at the same background  
 2110 rejection rate for up to 30 reconstructed vertices by merely adjusting the  $\tau_{21}$  selection.

2111 An important factor that influences the V tagging performance is the polarization of the  
 2112 reconstructed V bosons. In fact, the pruned jet mass selection is less efficient for transversely  
 2113 polarized ( $V_T$ ) V bosons. This can be explained by a higher asymmetry in the  $p_T$  of the two  
 2114 quarks from the  $V_T$  boson decay, such that the pruning algorithm in a considerable fraction  
 2115 of events rejects the particles from the lower  $p_T$  quark and yields a much lower jet mass.  
 2116 In addition, the  $\Delta R$  separation between the partons for pure longitudinally polarized ( $V_L$ )  
 2117 V bosons is smaller on average than for  $V_T$  bosons and is more likely to be accepted by a  
 2118 large-cone jet. In the analysis presented in this work only  $V_L$  bosons are considered.  
 2119

2120 This analysis relies on the modelling of the jet substructure variables  $m_{jet}$  and  $\tau_{21}$  in  
 2121 simulation. The data/simulation discrepancies in  $m_{jet}$  and  $\tau_{21}$  can bias the signal efficiency  
 2122 estimated from simulated samples. Therefore, the modelling of signal efficiency is cross-  
 2123 checked in a signal-free sample with jets having characteristics that are similar to those  
 2124 expected for a genuine signal [73]. A pure sample of high- $p_T$  W bosons, that decay to quarks  
 2125 and are reconstructed as a single jet, is obtained selecting  $t\bar{t}$  and single top quark events.  
 2126 Scale factors for the  $\tau_{21}$  selection efficiency are extracted by estimating the selection efficiency  
 2127 on both data and simulation for the pure W jet signal. This is achieved by subtracting the  
 2128 background contribution. The generated W boson in the  $t\bar{t}$  simulation provides a model  
 2129 of the contribution from the W jet peak in the pruned jet mass. The contribution from  
 2130 combinatorial background is derived from  $t\bar{t}$  simulation as well. This signal plus background  
 2131 model is fitted directly in the distributions of data and in their simulation.

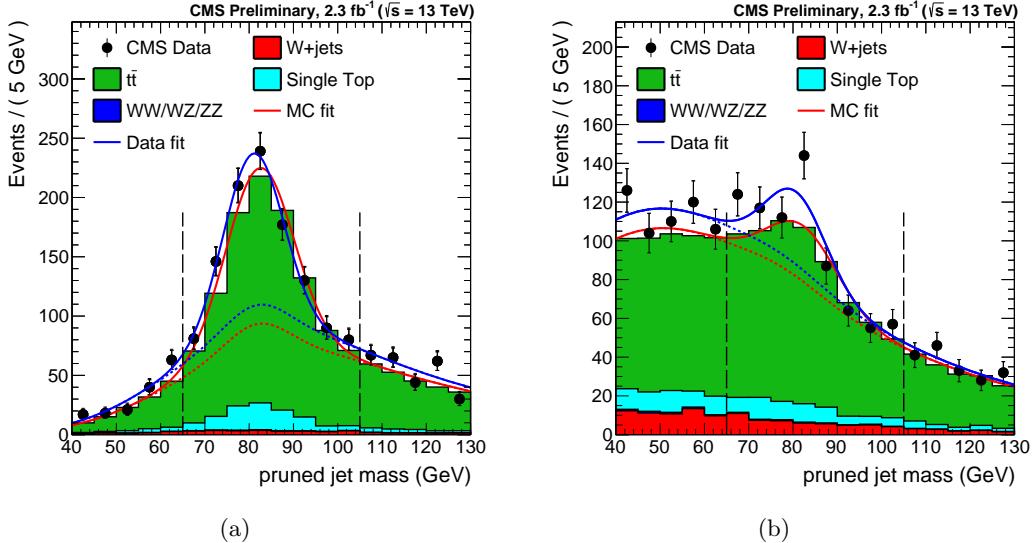
2132 The pruned jet mass distribution of events that pass and fail the  $\tau_{21}$  selection are fitted  
 2133 simultaneously to extract the selection efficiency on the pure W jet component. The ratio of  
 2134 data and simulation efficiencies are taken as the V tagging efficiency scale factor. Figure 7.3  
 2135 shows the fits obtained with 13 TeV data for the  $\tau_{21} < 0.45$  selection and similar results are  
 2136 obtained for the looser  $\tau_{21} < 0.6$  selection used in the  $\ell\nu q\bar{q}$  analysis presented in this work. The  
 2137 extracted scale factor for this selection is  $1.01 \pm 0.03$  and it is used to correct the total signal  
 2138 efficiency and the VV background normalization predicted by the simulation. The quoted  
 2139 uncertainty includes two systematic effects. One comes from the modelling of the nearby jets  
 2140 and  $p_T$  spectrum in  $t\bar{t}$  MC events, obtained by comparing the selection efficiency estimated  
 2141 from LO and NLO  $t\bar{t}$  simulation. The other is due to the choice of the models used to fit  
 2142 signal and background. The quadratic sum of these systematic uncertainties is found to be  
 2143 smaller than half of the statistical uncertainty on the scale factor. An additional uncertainty  
 2144 is calculated to account for the extrapolation of the scale factor from  $t\bar{t}$  events with an average  
 2145 jet  $p_T \sim 200$  GeV to higher momenta. This is estimated from the difference between PYTHIA8  
 2146 and HERWIG++ [97] showering models with a resulting factor of  $4.53\% \times \ln(p_T/200\text{ GeV})$ .

2147 The peak position in the W jet mass and its resolution are also extracted to obtain data-  
 2148 to-simulation corrections on the pruned jet mass listed in Table 7.2, as obtained from 13 TeV  
 2149 data and from simulation. The quoted uncertainties are statistical. The W jet mass scale in  
 2150 data is  $\approx 1\%$  smaller than in simulation while its resolution is found to be larger by about  
 2151 5%. In the simulation  $m_{jet}$  must therefore be shifted and enlarged by the above quantities to  
 2152 correct for the difference between data and simulation.

2153 The mass peak position is slightly shifted relative to the W boson mass. The shift is found  
 2154 to be primarily due to extra radiation in the W jet from the nearby b quark, and additional  
 2155 effects are due to the presence of the extra energy deposited in the jet cone from pileup, un-  
 2156 derlying event, and initial-state radiation not completely removed in the jet pruning procedure.

2157

2158 Because the kinematic properties of  $W$  jets and  $Z$  jets are very similar, the same corrections  
 2159 are also used when the  $V$  jet is assumed to arise from a  $Z$  boson.



**Figure 7.3:** Distribution in pruned jet mass for events that (a) pass and (b) fail the  $\tau_{21} < 0.45$  selection in the  $t\bar{t}$  control sample. The result of the fit to data and simulation are shown, respectively, by the solid and long-dashed line and the background components of the fit are shown as dashed-dotted and short-dashed line [73].

**Table 7.2:**  $W$  jet mass peak position and resolution, as extracted from top quark enriched sample in 8 TeV data and from simulation, after applying the  $\tau_{21} < 0.5$  selection [68].

$\tau_{21} < 0.45$	$m_{jet} [\text{GeV}]$	Standard deviation [GeV]
Data	$84.1 \pm 0.4$	$8.4 \pm 0.6$
Simulation	$82.7 \pm 0.3$	$7.6 \pm 0.4$

### 2160 7.3 The H tagging algorithm

2161 As discussed in the previous sections boosted  $V$  bosons are reconstructed using jet substructure  
 2162 methods through the  $V$  tagging algorithm, providing large discrimination against multijet  
 2163 backgrounds. However, if one or more of the decay products is a  $b$  quark, adding  $b$  jet  
 2164 identification (Section 6.4) along with jet substructure information can significantly improve  
 2165 the sensitivity of these methods.

2166 Two different approaches to identify boosted  $H \rightarrow b\bar{b}$  candidates have been explored and  
 2167 used at CMS [82]:

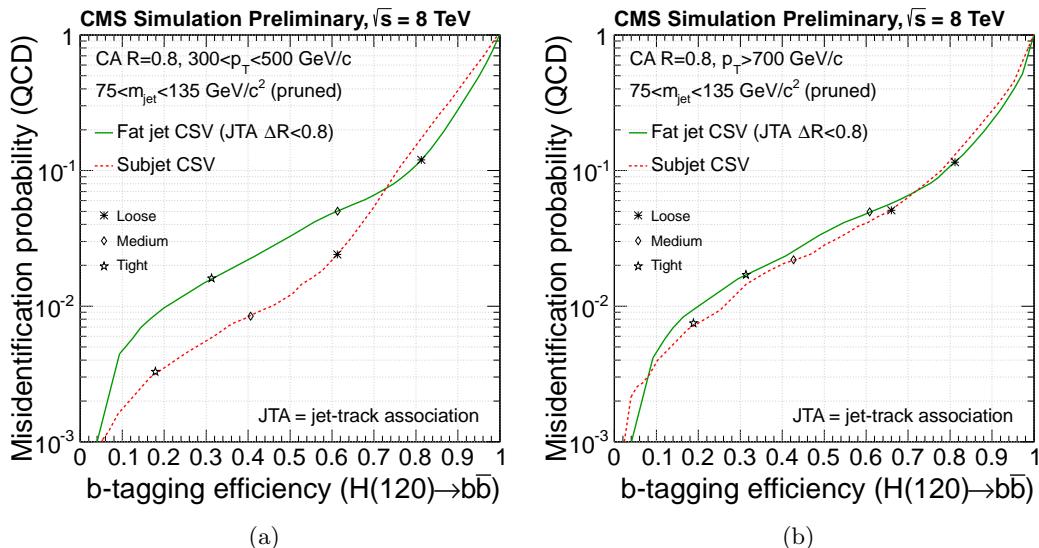
- 2168 application of  $b$  tagging to the fat jet (“fat jet  $b$  tagging”)
- 2169 application of  $b$  tagging to the subjets reconstructed within the fat jet (“subjets  $b$   
 2170 tagging”)

2171 Both approaches are based on the standard  $b$  tagging algorithms which take advantage of  
 2172 the tracking and vertexing information and are designed to identify jets from single  $b$  quarks.

As described in Section 6.4, the b tagging procedure starts with an association of tracks to jets, based on the angular distance between the tracks and the jet axis. The default b tagging algorithms use the selection  $\Delta R < 0.3$ . However, when applying this to a large-cone jet of size  $R = 0.8$ , the criteria is suboptimal. Hence, to apply b tagging to fat jets, this angular distance is enlarged to  $\Delta R < 0.8$ . For the application of b tagging to subjets, the angular distance remains at the default value of  $\Delta R < 0.3$ .

2179

The H tagging technique starts requiring that the pruned jet mass of the H jet candidate lies in a window around the Higgs boson mass (Fig. 7.1(b)), as this requirement rejects a large fraction of QCD background as demonstrated in the previous sections. The subjets are then obtained by reversing the last step of the pruning recombination algorithm described in Section 7.1.1. In addition to the jet mass requirement, the b tagging is applied either to the whole fat jet or to the two subjets, where both subjets are required to pass the same selection on the CSV discriminator. The b tagging efficiency and misidentification probability of QCD jets after applying the selection  $75 < m_{\text{jet}} < 135 \text{ GeV}$  are shown in Fig. 7.4. The subjet b tagging outperforms the fat jet tagging for most of the phase space.



**Figure 7.4:** Misidentification probability as a function of b tagging efficiency for boosted  $H \rightarrow b\bar{b}$  jets and inclusive QCD jets for the CSV algorithm applied to CA8 jets and pruned subjets for jets with (a)  $300 < p_T < 500 \text{ GeV}/c$  and (b)  $p_T > 700 \text{ GeV}/c$ . Loose, medium, and tight operating points of the CSV discriminator are indicated [82].

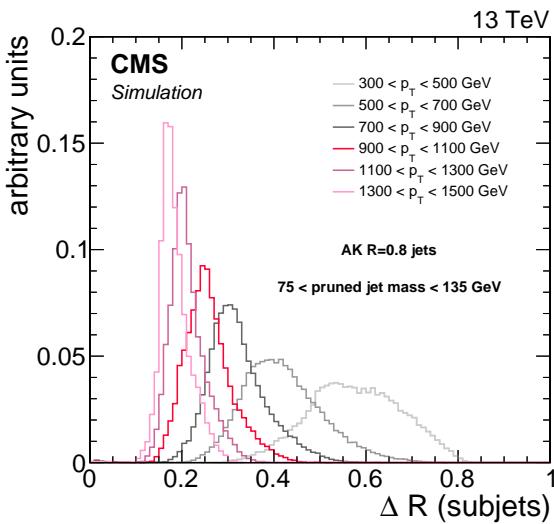
The H tagging efficiency obtained combining the requirement on the pruned jet mass ( $75 < m_{\text{jet}} < 135 \text{ GeV}$ ) and the subjet b tagging at the CSVL operating point is between 40 and 50% for a H jet  $p_T$  range spanning from 300 GeV to 1 TeV, with a suppression of QCD background to about 0.4%.

2193

The use of a fixed-size jet-track association cone inevitably leads to track sharing between the subjets of the jets once their angular separation becomes comparable or smaller than the size of the association cone. For boosted H jets the fraction of shared tracks, defined as the ratio of the number of tracks within  $\Delta R < 0.3$  from more than one subjet and the number of all tracks within  $\Delta R < 0.3$  from any of the subjets, ranges from a few percent at a jet  $p_T$  of 400 GeV and increases to 40% at a jet  $p_T$  of 700 GeV and to 80% at a jet  $p_T$  of

2200 1 TeV. Because of track sharing, the b tagging probabilities for individual subjets deteriorate  
 2201 at large jet  $p_T$  and the subjet b tagging performance approach the fat jet b tagging one as  
 2202 can be seen in Fig. 7.4. The lost in efficiency is then recovered applying the two approaches  
 2203 depending on the  $\Delta R$  between the two subjets. In particular, the analysis involving boosted  
 2204 Higgs bosons such as the one presented in this work apply subjet b tagging and fat jet b  
 2205 tagging if  $\Delta R > 0.3$  and  $< 0.3$ , respectively. The distribution of the angular separation  $\Delta R$   
 2206 of the two subjets reconstructed within the fat jet for different jet  $p_T$  ranges in simulated  
 2207 events of highly boosted Higgs bosons decaying to  $b\bar{b}$ , is shown in Fig. 7.5.

2208 In this analysis a requirement on the pruned jet mass of the reconstructed H jet candidate  
 2209 given by  $110 < m_{\text{jet}} < 135$  GeV is applied. The  $m_{\text{jet}}$  window is chosen such that a contamina-  
 2210 tion from possible signals with boosted V jets in the Higgs boson mass region is minimized.  
 2211 The b tagging is performed with the algorithm described above using the loose working point  
 2212 of the CSV discriminant. The total H tagging efficiency for these selections is about 35% for  
 2213 jet  $p_T$  of about 1 TeV with a mistagging probability below 1%.



**Figure 7.5:** Distributions of the angular separation  $\Delta R$  of the two subjets reconstructed within the fat jet for simulated events of highly boosted Higgs bosons decaying to  $b\bar{b}$ . The distributions are compared for different ranges of the H jet  $p_T$ .

2214 The validation of b tagging in boosted H jets is performed selecting events containing jets  
 2215 from gluon splitting to  $b\bar{b}$  (  
 2216  $mathrm{mg} \rightarrow b\bar{b}$ ) in which the b quarks hadronize inside the fat jet [82]. To enrich a sample  
 2217 of fat jets in  $g \rightarrow b\bar{b}$  component, used as an analogue of boosted  $H \rightarrow b\bar{b}$  jets, the fat jets are  
 2218 required to be double-muon-tagged with both subjets matched to distinct muon candidates  
 2219 within a cone of size  $\Delta R < 0.4$ . This sample is used to study the modelling of b tagging  
 2220 efficiencies in boosted  $H \rightarrow b\bar{b}$  topologies. The scale factors, given by the ratio between the  
 2221 efficiencies measured in data and simulation, are found to be in good agreement with those  
 2222 measured in the standard, non-boosted topologies, indicating that the simulation reproduces  
 2223 the b tagging performance in boosted and non-boosted environments equally well. These  
 2224 scale factors are used in the analysis to reweight the simulated events.

---

# Analysis strategy

2228 This chapter describes in details the strategy followed in this search, that starting from the  
 2229 physics objects and identification algorithms described in the previous chapters, leads to  
 2230 the final results of the analysis. Although preliminary selections on the objects expected  
 2231 in the final state have already been discussed, tighter requirements and a categorization of  
 2232 the events are applied as described in Section 8.1 to maximize the analysis sensitivity to  
 2233 the signals under study. The final discriminating observable used to search for the signal is  
 2234 represented by the invariant mass of the diboson system. In fact, a possible signal would  
 2235 appear as a localized excess of data in one of the bins on the top of a smooth background.  
 2236 An accurate description of the expected background and signal distributions is therefore  
 2237 fundamental. A background estimation method for the main W+jets component, which  
 2238 makes use of data in sideband regions is used and described in Section 8.2. Another important  
 2239 source of background is represented by top quark production, which is estimated from data in  
 2240 a dedicated control region as discussed in Section 8.3. The background model together with  
 2241 the signal model presented in Section 8.4 is used to perform a maximum likelihood fit of the  
 2242 data in the statistical analysis. The systematic uncertainties in the signal and background  
 2243 predictions discussed in Section 8.5 are treated as nuisance parameters in the statistical  
 2244 interpretation. Finally, Section 8.6 describes the standard procedure for the statistical test of  
 2245 the new signal hypothesis commonly used by LHC experiments and originally developed for  
 2246 the Higgs boson search. The final results are presented in the next chapters.

2247 **8.1 Final event selection and categorization**

2248 Events are selected online with triggers requiring either one muon or electron (Sections 6.2.2  
 2249 and 6.3.2). Several requirements are then applied offline to the selected events to enhance  
 2250 the analysis sensitivity as described in the following.

2252 The two analyses described in this work feature the same selection strategy on the leptonic  
 2253 part of the final state. Both analyses require exactly one muon or one electron satisfying  
 2254 certain  $p_T$  and  $\eta$  requirements and passing the high- $p_T$  lepton identification criteria described  
 2255 in Sections 6.3.3 and 6.2.3. As summarized in Tables 8.1 and 8.2, the only difference is in the  
 2256  $p_T$  threshold of the lepton which is higher for the 13 TeV data analysis to match the increase  
 2257 in the trigger threshold. The offline reconstructed  $p_T$  of the electron must be greater than  
 2258 90 (120) GeV for the 8 (13) TeV data analysis, where the trigger reaches the plateau. This  
 2259 is required in order to avoid any bias on the distributions due to the turn-on of the trigger  
 2260 efficiency curve and its description in simulation. Reconstructed electrons must have  $|\eta| < 2.5$   
 2261 and also be located outside of the overlap region between the ECAL barrel and endcaps,  
 2262 because the reconstruction of an electron object in this region is not optimal. In a similar  
 2263 way, the offline reconstructed  $p_T$  of the muon must be greater than 50 (53) GeV for the 8  
 2264 (13) TeV analysis, and within  $|\eta| < 2.1$  as a consequence of the trigger criteria. Events with  
 2265 additional well-identified muons and/or electrons are rejected to avoid contamination from  
 2266 events containing  $Z \rightarrow ll$  decays.

**Table 8.1:** Summary of the final selection for the 8 TeV data analysis in the  $\ell\nu b\bar{b}$  final state.

Selection	Value
Lepton selection	
Electron	$p_T > 90 \text{ GeV}$ $ \eta  < 2.5$ except [1.44, 1.57] range
Muon	$p_T > 50 \text{ GeV}$ $ \eta  < 2.1$
AK5 jet selections	
Jet $p_T$	$p_T > 30 \text{ GeV}$
Jet $\eta$	$ \eta  < 2.4$
$E_T^{\text{miss}}$ selections	
$E_T^{\text{miss}}$ (electron channel)	$E_T^{\text{miss}} > 80 \text{ GeV}$
$E_T^{\text{miss}}$ (muon channel)	$E_T^{\text{miss}} > 40 \text{ GeV}$
Boson selections	
$W \rightarrow \ell\nu$	$p_T > 200 \text{ GeV}$
$H \rightarrow b\bar{b}$ (CA8 jet)	$p_T > 200 \text{ GeV}$
Back-to-back topology	$ \eta  < 2.4$ except [1.0, 1.8] range $\Delta R(\ell, H_{b\bar{b}}) > \pi/2$ $\Delta\phi(H_{b\bar{b}}, \vec{p}_T^{\text{miss}}) > 2$ $\Delta\phi(H_{b\bar{b}}, W_{\ell\nu}) > 2$
Diboson invariant mass	$m_{WH} > 0.7 \text{ TeV}$
H tagging selections	
Pruned jet mass	$110 < m_{\text{jet}} < 135 \text{ GeV}$
Combined b-tagging cut	2 CSVL b-tagged subjets if $\Delta R(\text{subjets}) > 0.3$ 1 CSVL b-tagged CA8 jet if $\Delta R(\text{subjets})$
$t\bar{t}$ rejection	
B-tag veto	no CSVM b-tagged AK5 jet within $\Delta R(H_{b\bar{b}}, AK5) = 0.8$
Top quark mass veto	$m_{\text{top}}^l < 120 \parallel m_{\text{top}}^l > 240$ $m_{\text{top}}^h < 160 \parallel m_{\text{top}}^h > 280$

2268     The requirements  $E_T^{\text{miss}} > 40$  and  $> 80 \text{ GeV}$  are applied, respectively, in the muon and  
 2269     electron channels. The threshold is higher in the electron channel to further suppress the larger  
 2270     background from multijet processes expected at low values of  $E_T^{\text{miss}}$  due to jets misidentified  
 2271     as electrons. This background is expected to be negligible in the muon channel, for which a  
 2272     lower  $E_T^{\text{miss}}$  threshold can be used to preserve a higher efficiency for a low-mass signal. The  
 2273     identified lepton and the  $E_T^{\text{miss}}$  are used to reconstruct the  $W \rightarrow \ell\nu$  candidate as described in  
 2274     Section 6.6, which is required to have  $p_T > 200 \text{ GeV}$ .

2275

2276     A different strategy is instead used in the two analyses, for the hadronic part of the final  
 2277     state. As described in Section 6.4, the CA8 and AK8 algorithms are used to reconstruct  
 2278     the H and V jet candidates in the 8 and 13 TeV analysis, respectively. In both cases the jet  
 2279     is required to have  $p_T > 200 \text{ GeV}$  and  $|\eta| < 2.4$ . For CA8 jets, the pseudorapidity region  
 2280      $1.0 < |\eta| < 1.8$  is excluded corresponding to the barrel-endcap transition region of the silicon  
 2281     tracker where the reconstruction of tracks is not optimal (Section 6.4.2). The probability of  
 2282     signal events with jets outside this region is 80% (92%) for a resonance mass of 1.0 (2.5) TeV.

2283     The 8 TeV analysis aims at isolating events with a high- $p_T$  Higgs boson decaying to  $b\bar{b}$   
 2284     and the H tagging algorithm described in Section 7.3 is applied. The H tagging requires  
 2285     the selected CA8 jet to have pruned mass in the range  $110 < m_{\text{jet}} < 135 \text{ GeV}$ . Furthermore,  
 2286     the subjets are required to be b-tagged with the CSVL algorithm if their angular distance  
 2287      $\Delta R < 0.3$ . Otherwise, b tagging is applied to the whole CA8 jet using the same algorithm.

2288     The 13 TeV analysis is instead focused on events with a high- $p_T$  V boson decaying to  $q\bar{q}$

**Table 8.2:** Summary of the final selection for the 13 TeV data analysis in the  $\ell\nu q\bar{q}$  final state.

Selection	Value
Lepton selection	
Electron	$p_T > 120 \text{ GeV}$ $ \eta  < 2.5$ except [1.44, 1.57] range
Muon	$p_T > 53 \text{ GeV}$ $ \eta  < 2.1$
AK4 jet selections	
Jet $p_T$	$p_T > 30 \text{ GeV}$
Jet $\eta$	$ \eta  < 2.4$
$E_T^{\text{miss}}$ selections	
$E_T^{\text{miss}}$ (electron channel)	$E_T^{\text{miss}} > 80 \text{ GeV}$
$E_T^{\text{miss}}$ (muon channel)	$E_T^{\text{miss}} > 40 \text{ GeV}$
Boson selections	
$W \rightarrow \ell\nu$	$p_T > 200 \text{ GeV}$
$V \rightarrow q\bar{q}$ (AK8 jet)	$p_T > 200 \text{ GeV}$ $ \eta  < 2.4$
Back-to-back topology	$\Delta R(\ell, V_{q\bar{q}}) > \pi/2$ $\Delta\phi(V_{q\bar{q}}, \vec{p}_T^{\text{miss}}) > 2$ $\Delta\phi(V_{q\bar{q}}, W_{\ell\nu}) > 2$
Diboson invariant mass	$m_{VV} > 0.7 \text{ TeV}$
V tagging selections	
Pruned jet mass	$65 < m_{\text{jet}} < 105 \text{ GeV}$
2- to 1-subjettiness ratio	$\tau_{21} < 0.6$
$m_{\text{jet}}$ categories	
WW-enriched	$65 < m_{\text{jet}} < 85 \text{ GeV}$
WZ-enriched	$85 < m_{\text{jet}} < 105 \text{ GeV}$
$t\bar{t}$ rejection	
B-tag veto	no CSV b-tagged AK5 jet within $\Delta R(V_{q\bar{q}}, AK5) = 0.8$

and the V tagging algorithm described in Section 7.3 is applied in this case. The pruned jet mass window is shifted down to the V boson mass, requiring the selected AK8 jet to have pruned mass in the range  $65 < m_{\text{jet}} < 105 \text{ GeV}$ . Furthermore, the V jet is required to have  $\tau_{21} < 0.6$ . Finally, the V jet is deemed a W-boson candidate if its pruned mass falls in the range 65–85 GeV, while it is deemed a Z-boson candidate if it falls in the range 85–105 GeV instead. This categorization has been added on the top of the V tagging requirements on the  $m_{\text{jet}}$  to enhance discrimination between resonances with different charge and spin. Indeed, the first category, referred to as “WW-enriched”, has a higher sensitivity for resonances such as the neutral spin-2 graviton or the neutral spin-1  $Z'$  decaying to WW, where a W jet is expected. The second category, referred to as “WZ-enriched”, is instead optimized for resonances such as the charged spin-1  $W'$  decaying to WZ, where a Z jet is expected.

2300

In addition, there are specific topological selection criteria chosen for both the analyses. It is required that the two V bosons from the decay of a massive resonance are approximately back-to-back: the  $\Delta R$  between the lepton and the signal jet is greater than  $\pi/2$ ; the  $\Delta\phi$  between the vector  $\vec{p}_T^{\text{miss}}$  and the signal jet, as well as between the  $W \rightarrow \ell\nu$  and signal jet candidates, are both greater than 2 radians.

To reduce the level of the  $t\bar{t}$  background, events with one or more reconstructed AK5 (or AK4) jets, not overlapping with the signal jet candidate are analyzed: if one or more of these jets is b-tagged with the CSV algorithm, the event is rejected. For the 8 TeV analysis additional selections are applied to further reduce contamination from  $t\bar{t}$  background. In

2306

fact, the b tagging requirements in this analysis enhance the contribution from top quark production where real b jets are present. A leptonically decaying top quark candidate mass  $m_{\text{top}}^l$  is reconstructed from the lepton,  $E_T^{\text{miss}}$ , and the closest AK5 jet to the lepton using the method described in Section 6.6. A hadronically decaying top quark candidate mass  $m_{\text{top}}^h$  is also reconstructed, from the H jet candidate and the closest AK5 jet. Events with  $120 < m_{\text{top}}^l < 240 \text{ GeV}$  or  $160 < m_{\text{top}}^h < 280 \text{ GeV}$  are rejected. The chosen windows around the top quark mass are the result of an optimization carried out in this analysis, taking into account the asymmetric tails at larger values due to combinatorial background.

2318

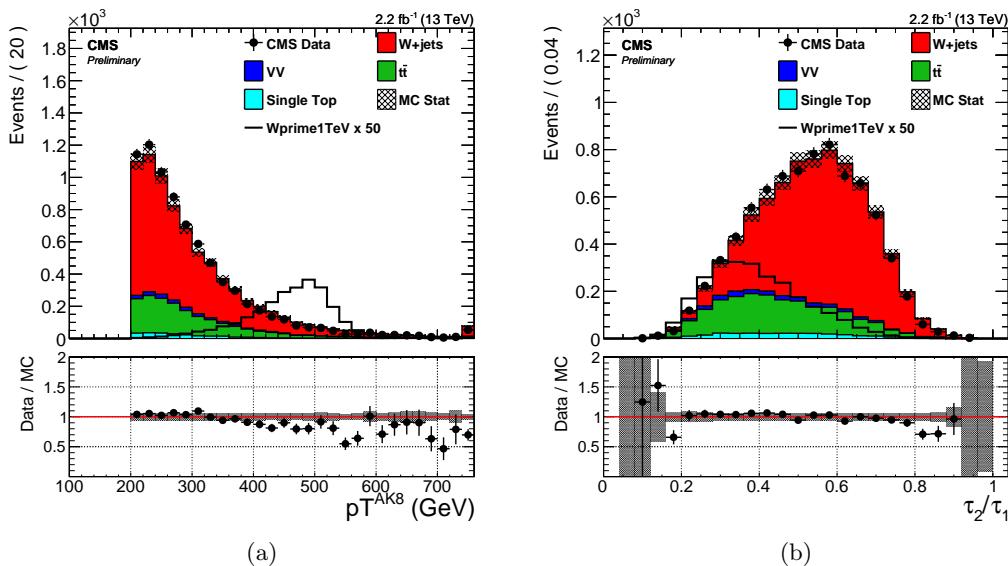
According to the above description of the final selections, the event categorization is based on 2 orthogonal classes of events for the 8 TeV data analysis in the  $\ell\nu bb$  final state, depending on the lepton flavour (muon or electron), and on 4 orthogonal classes of events for the 13 TeV data analysis in the  $\ell\nu q\bar{q}$  final state, depending on the lepton flavour and on the pruned jet mass category (WW or WZ).

The two boson candidates are combined into a diboson candidate, with presence of signal then inferred from the observation of localized excesses in the  $m_{\ell\nu+\text{jet}}$  distribution. When several diboson resonance candidates are present in the same event, only the one with the highest  $p_T$  V or H jet is kept for further analysis.

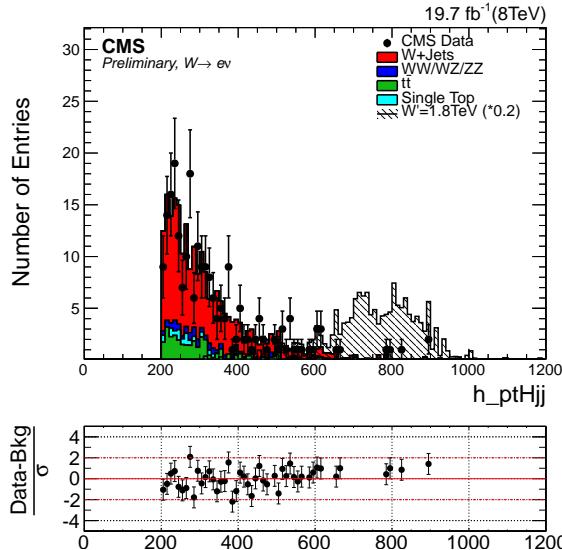
The reconstructed invariant mass of the resonance is required to be at least 0.7 TeV.

2329

The distributions in  $p_T$  and N-subjettiness ratio  $\tau_{21}$  distributions for the V jet candidate in the  $\ell\nu q\bar{q}$  channel is shown in Fig. 8.1, after requiring  $65 < m_{\text{jet}} < 105 \text{ GeV}$ , for both simulation and 13 TeV data. Figure 8.2 shows the distribution in  $p_T$  for the H jet candidate after requiring  $40 < m_{\text{jet}} < 110 \text{ GeV}$ , for both simulation and 8 TeV data. **FIXME: If I have time I will redo these plots.**



**Figure 8.1:** Distributions in  $p_T$  (a) and N-subjettiness ratio  $\tau_{21}$  (b) for the V jet candidate obtained requiring  $65 < m_{\text{jet}} < 105 \text{ GeV}$  after merging muon and electron channels. The SM diboson,  $t\bar{t}$ , and single top quark backgrounds are taken from simulation and are normalized to the integrated luminosity of the 13 TeV data sample. The  $W+jets$  background is rescaled to match the number of events.



**Figure 8.2:** Distributions in  $p_T$  for the H jet candidate obtained requiring  $40 < m_{\text{jet}} < 110 \text{ GeV}$  for events in the muon channel. The SM diboson,  $t\bar{t}$ , and single top quark backgrounds are taken from simulation and are normalized to the integrated luminosity of the 8 TeV data sample. The W+jets background is rescaled to match the number of events in data.

## 2335 8.2 W+jets background estimate with $\alpha$ ratio method

2336 The  $m_{\ell\nu+\text{jet}}$  distribution observed in data is dominated by SM background processes where  
 2337 single quark or gluon jets are falsely identified as signal jets. The dominant processes is  
 2338 inclusive W boson production. Since both normalization and shape discrepancies are visible  
 2339 between data and simulation [89], a data driven method has been developed to estimate this  
 2340 background component, as described in the following. Sub-dominant backgrounds include  $t\bar{t}$ ,  
 2341 single top quark, and non resonant diboson SM production, which are estimated from MC,  
 2342 after applying correction factors for residual data-to-simulation disagreement measured in  
 2343 control samples selected in data.

### 2344 8.2.1 Description

2345 The W+jets background is estimated through the so called  $\alpha$  *ratio* method. This method  
 2346 assumes that the correlation between  $m_{\text{jet}}$  and  $m_{\ell\nu+\text{jet}}$  for the dominant W+jets background  
 2347 can be adequately modelled by simulation. A signal-depleted control region (sideband) is  
 2348 defined by requiring the mass of the V or H jet to lie below or above the nominal selection;  
 2349 the  $m_{\ell\nu+\text{jet}}$  distribution observed in this region is then extrapolated to the nominal region  
 2350 through a transfer function estimated from simulation. Other minor sources of background,  
 2351 such as  $t\bar{t}$ , single top quark, and SM diboson production, are estimated using simulated events  
 2352 after applying correction factors based on control regions in data, as described in Sections 7.2  
 2353 and 8.3. The sideband region is defined around the jet mass window that represents the  
 2354 analysis signal region (Section 8.1). The lower and upper sidebands for the two analyses  
 2355 are summarized in Table 8.3. For the 13 TeV analysis a “gap” is introduced between the  
 2356 signal region and the upper sideband, since the range defined by  $105 < m_{\text{jet}} < 135$  might  
 2357 include contribution from signals with highly Lorentz-boosted Higgs bosons in the final state.  
 2358 Since these types of searches at 13 TeV [98] have been performed simultaneously with the  
 2359 one described in this work, this region has been discarded to avoid introducing a bias in  
 2360 the shape and normalization extrapolation due to a possible signal. On the other hand,

the lower sideband of the 8 TeV  $\ell\nu b\bar{b}$  analysis includes the region where signals from highly Lorentz-boosted V bosons might occur. In fact, this analysis has been performed after the search for WV resonances in the semi-leptonic final state at 8 TeV discovered the signal region, where no deviation from the predicted SM background have been observed [93].

**Table 8.3:** Sideband regions used in the two analyses to estimate the contribution from the main W+jets background.

$m_{\text{jet}}$ sideband	final state	
	$\ell\nu b\bar{b}$	$\ell\nu q\bar{q}$
Low sideband (LSB)	40–110 GeV	40–65 GeV
High sideband (HSB)	135–150 GeV	135–150 GeV

### 8.2.2 Extraction of the W+jets normalization

The overall normalization of the W+jets background in the signal region is determined from a fit to the  $m_{\text{jet}}$  distribution in the lower and upper sidebands of the data. The analytical form of the fitting function is chosen from simulation studies, as are the contributions from minor backgrounds. A summary of the empirical functional forms used to parametrize each background contribution are listed in Table 8.4, and defined as follows:

$$\begin{aligned}
 F_{\text{Exp}}(x) &= e^{cx} \\
 F_{\text{ErfExp}}(x) &= e^{cx} \cdot \frac{1 + \text{Erf}((x - a)/b)}{2} \\
 F_{\text{ExpGaus}}(x) &= c_0 \cdot e^{cx} + c_1 \cdot \text{Gaus}(x, x_1, \sigma_1) \\
 F_{4\text{Gaus}}(x) &= c_1 \cdot \text{Gaus}(x, x_1, \sigma_1) + c_2 \cdot \text{Gaus}(x, x_2, \sigma_2) + c_3 \cdot \text{Gaus}(x, x_3, \sigma_3) + c_4 \cdot \text{Gaus}(x, x_4, \sigma_4) \\
 F_{\text{ErfExp2Gaus}}(x) &= e^{cx} \cdot \frac{1 + \text{Erf}((x - a)/b)}{2} + c_1 \cdot \text{Gaus}(x, x_1, \sigma_1) + c_2 \cdot \text{Gaus}(x, x_2, \sigma_2)
 \end{aligned} \tag{8.1}$$

**Table 8.4:** Summary of the empirical functional forms used to fit the  $m_{\text{jet}}$  spectra of each background component in the two analyses.

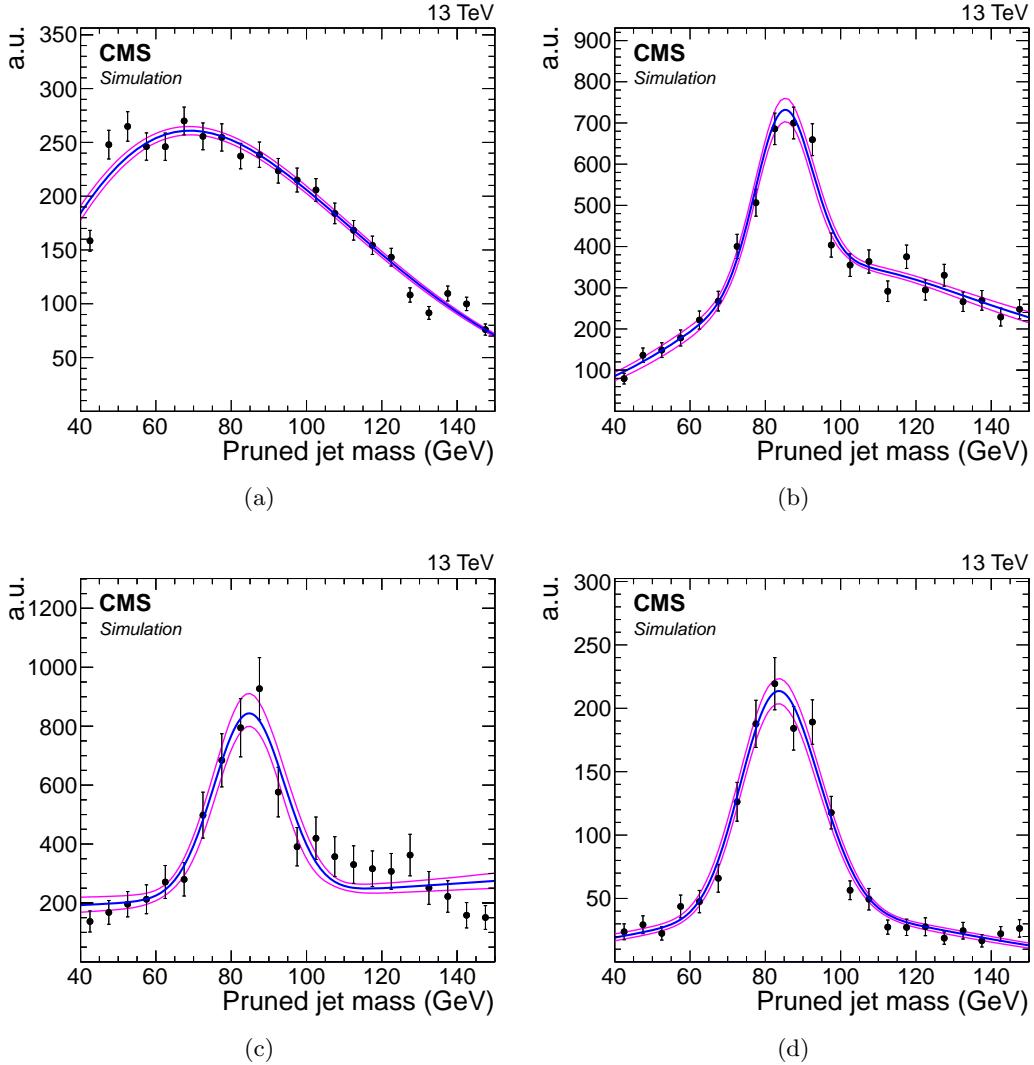
Final state	W+jets	$t\bar{t}$	single top quark	diboson
$\ell\nu b\bar{b}$	$F_{\text{ErfExp}}(x)$	$F_{\text{Exp}}(x)$	$F_{\text{Exp}}(x)$	$F_{\text{ExpGaus}}(x)$
$\ell\nu q\bar{q}$	$F_{\text{ErfExp}}(x)$	$F_{\text{ErfExp2Gaus}}(x)$	$F_{\text{ExpGaus}}(x)$	$F_{4\text{Gaus}}(x)$

Figure 8.3 shows the functional forms listed in Table 8.4 for the  $\ell\nu q\bar{q}$  channel, after fitting the simulation data of each background component, demonstrating that the chosen functions well reproduce the expected  $m_{\text{jet}}$  spectra.

The results of this fit procedure to extract the W+jets normalization are shown in Fig. 8.4 and 8.5 for the  $\ell\nu b\bar{b}$  and the  $\ell\nu q\bar{q}$  channel, respectively. The factors for correcting the simulated W-peak position and resolution to represent the observed data, taken from the top quark enriched control sample as described in Section 7.2, are included in the  $m_{\text{jet}}$  spectra of Fig. 8.5.

### 8.2.3 Extraction of the W+jets shape

The form of the  $m_{\ell\nu+\text{jet}}$  distribution for the W+jets background in the signal region (SR) is determined from the lower  $m_{\text{jet}}$  sideband, through the transfer function  $\alpha_{\text{MC}}(m_{\ell\nu+\text{jet}})$  obtained from the W+jets simulation, and defined as:



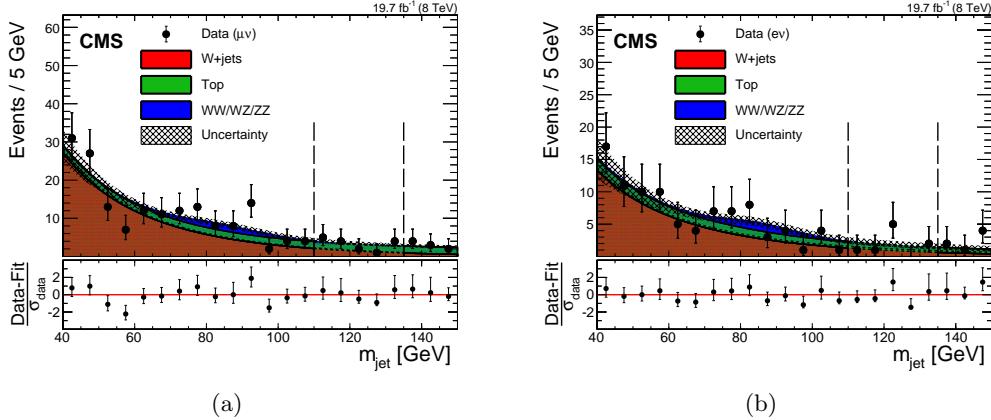
**Figure 8.3:** Functional forms describing the  $m_{\text{jet}}$  spectra for each background contribution after fitting the simulation data. (a) W+jets. (b)  $t\bar{t}$ . (c) Single top quark. (d) Diboson.

$$\alpha_{\text{MC}}(m_{\ell\nu+\text{jet}}) = \frac{F_{\text{MC,SR}}^{\text{W+jets}}(m_{\ell\nu+\text{jet}})}{F_{\text{MC,SB}}^{\text{W+jets}}(m_{\ell\nu+\text{jet}})}, \quad (8.2)$$

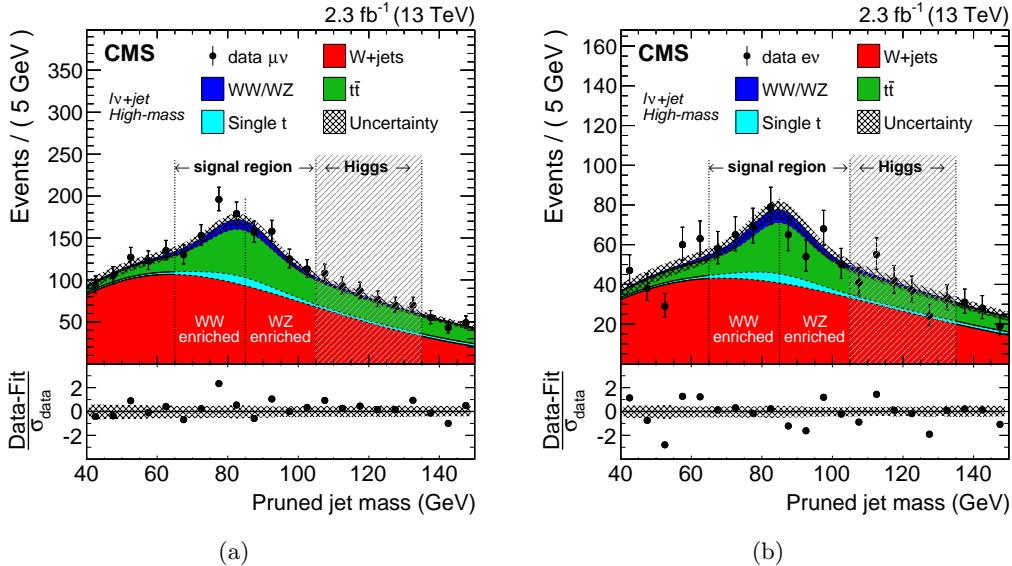
where  $F_{\text{MC,SB}}^{\text{W+jets}}$  and  $F_{\text{MC,SR}}^{\text{W+jets}}$  are the probability density functions used to describe the simulated  $m_{\ell\nu+\text{jet}}$  spectrum in the lower  $m_{\text{jet}}$  sideband and signal region, respectively. The upper  $m_{\text{jet}}$  sideband is not considered since the W+jets shape is different here compared to what expected in the lower sideband. Furthermore, the upper sideband suffers from a larger  $t\bar{t}$  background contamination.

Since the lower sideband region does not represent a perfectly pure sample of W+jets events in data, the presence of minor backgrounds is subtracted from the observed diboson invariant mass distribution to obtain an estimation of the W+jets contribution in the sideband control region of the data,  $F_{\text{data,SB}}^{\text{W+jets}}(m_{\ell\nu+\text{jet}})$ .

The  $m_{\ell\nu+\text{jet}}$  range used in the estimate of the background distribution determines the region of masses probed by these searches. This range is chosen to ensure a smoothly falling



**Figure 8.4:** Distributions in pruned jet mass  $m_{\text{jet}}$  in the muon (a) and electron (b) channels for the  $\ell\nu b\bar{b}$  analysis at 8 TeV. All selections are applied except the requirement on the  $m_{\text{jet}}$  signal window. The signal region lies between the dashed vertical lines. The hatched region indicates the statistical uncertainty of the fit. At the bottom of each plot, the bin-by-bin fit residuals,  $(N_{\text{data}} - N_{\text{fit}})/\sigma_{\text{data}}$ , are shown.



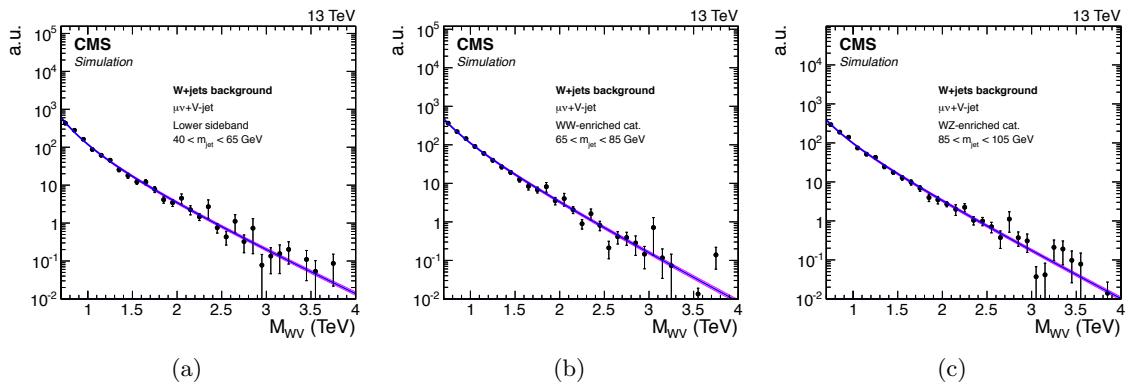
**Figure 8.5:** Distributions in pruned jet mass in the muon (a) and electron (b) channels for the  $\ell\nu q\bar{q}$  analysis at 13 TeV. All selections are applied except the requirement on the  $m_{\text{jet}}$  signal window. The signal regions and  $m_{\text{jet}}$  categories of the analysis are indicated by the vertical dotted lines. The shaded  $m_{\text{jet}}$  region 105–135 GeV is not used in the analysis. At the bottom of each plot, the bin-by-bin fit residuals,  $(N_{\text{data}} - N_{\text{fit}})/\sigma_{\text{data}}$ , are shown together with the uncertainty band of the fit normalized by the statistical uncertainty of data,  $\sigma_{\text{data}}$ .

background spectrum, and therefore far enough from the kinematic turn-on at low masses generated by the acceptance selections, allowing for a good stability and a robust control of the background estimation. For this reason the low edge of the range is chosen at 0.7 TeV while the high edge is chosen such that it is not too far from the last value where data are still present. Therefore, the fits are performed in the range  $0.7 < m_{\ell\nu+\text{jet}} < 4$  TeV for the 13 TeV analysis, while at 8 TeV no data are present above  $m_{\ell\nu+\text{jet}} \approx 3$  TeV and the chosen range is therefore  $0.7 < m_{\ell\nu+\text{jet}} < 3$  TeV.

To describe the smoothly falling W+jets background distribution, a parametrization of the form of a leveled exponential is adopted, defined as

$$F_{\text{ExpTail}}(x) = e^{-\frac{x}{a+bx}}. \quad (8.3)$$

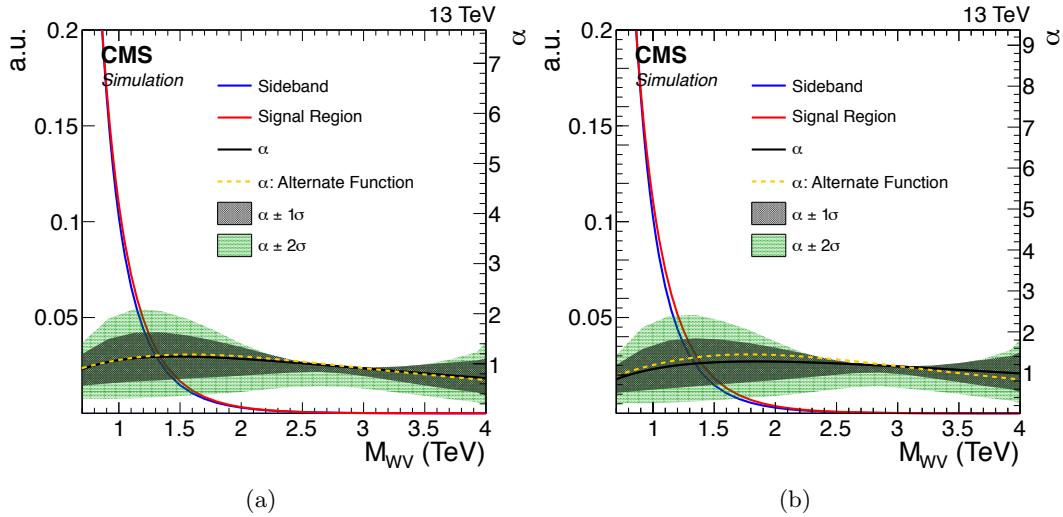
This functional form is found to adequately describe the simulation in both the signal region and the low sideband as demonstrated in Fig. 8.6. Tests are performed with alternative functional forms, and the background prediction is found to agree with the one of the default function within the uncertainties. The minor background contributions are parametrized with a simple exponential functional form, except for the diboson contribution for which the  $F_{\text{ExpTail}}(x)$  defined above is used.



**Figure 8.6:** Functional form describing the diboson invariant mass spectrum of the W+jets background after fitting the simulation data. The distributions for the lower  $m_{\text{jet}}$  sideband (a), and the WW-enriched (b) and WZ-enriched (c) signal regions of the  $\ell\nu q\bar{q}$  analysis are shown.

For the  $\ell\nu q\bar{q}$  analysis, the  $\alpha_{\text{MC}}$  is computed independently for the two WW- and WZ-enriched categories, which are therefore treated as two different signal regions. Figure 8.7 shows the  $\alpha_{\text{MC}}$  for the two categories, obtained from a simultaneous fit of W+jets simulated data in the lower sideband and in the signal region defined by the category using the parametrization in Eq. 8.3. The blue and the red lines represent the probability density functions describing the W+jets background with  $m_{\text{jet}}$  in the lower sideband and signal region, respectively, and given by the leveled-exponential function of Eq. 8.3. A simultaneous fit is performed of the two distributions, where the parameters used to model the distribution in the signal region are correlated with the ones used to model the distribution in the sideband. The transfer function  $\alpha_{\text{MC}}$  is shown as a solid black line, while the dark (light) shaded region corresponds to the  $1\sigma$  ( $2\sigma$ ) statistical uncertainty of the fit. These uncertainties only represent the uncertainty in the modelling of the W+jets distribution. The bands have a size of approximately zero around 2 TeV as the  $\alpha_{\text{MC}}$  is the ratio of two probability density functions which have to cross in order to conserve the total probability. Similar results are obtained for the  $\ell\nu b\bar{b}$  analysis.

In Fig. 8.8, the result of the fit to the  $m_{\ell\nu+\text{jet}}$  distribution of the data with  $m_{\text{jet}}$  in the lower sideband is shown for the electron and muon channels of the  $\ell\nu q\bar{q}$  analysis. From this fit, an estimation of  $F_{\text{data},\text{SB}}^{\text{W+jets}}(m_{\ell\nu+\text{jet}})$  is obtained. Finally, the W+jets background distribution in the signal region is then extrapolated by rescaling  $F_{\text{data},\text{SB}}^{\text{W+jets}}$  by  $\alpha_{\text{MC}}$ . The minor backgrounds are then added to the W+jets background to obtain the total SM prediction in the signal region, which is given by



**Figure 8.7:** The transfer functions  $\alpha_{MC}$  from the lower  $m_{jet}$  sideband to the signal region defined by the WW-enriched (a) and WZ-enriched (b) category of the  $\ell\nu q\bar{q}$  analysis. The dark and light shaded areas represent the statistical uncertainty of the fit. The blue and the red lines represents the probability density functions describing the W+jets background with  $m_{jet}$  in the lower sideband and signal region, respectively. The  $\alpha_{MC}$  obtained fitting the W+jets with and alternative function is shown as yellow line.

$$N_{SR}^{bkg}(m_{\ell\nu+jet}) = N_{SR}^{W+jets} \times \alpha_{MC}(m_{\ell\nu+jet}) \times F_{data,SB}^{W+jets}(m_{\ell\nu+jet}) + \sum_k N_{SR}^k \times F_{MC,SR}^k(m_{\ell\nu+jet}). \quad (8.4)$$

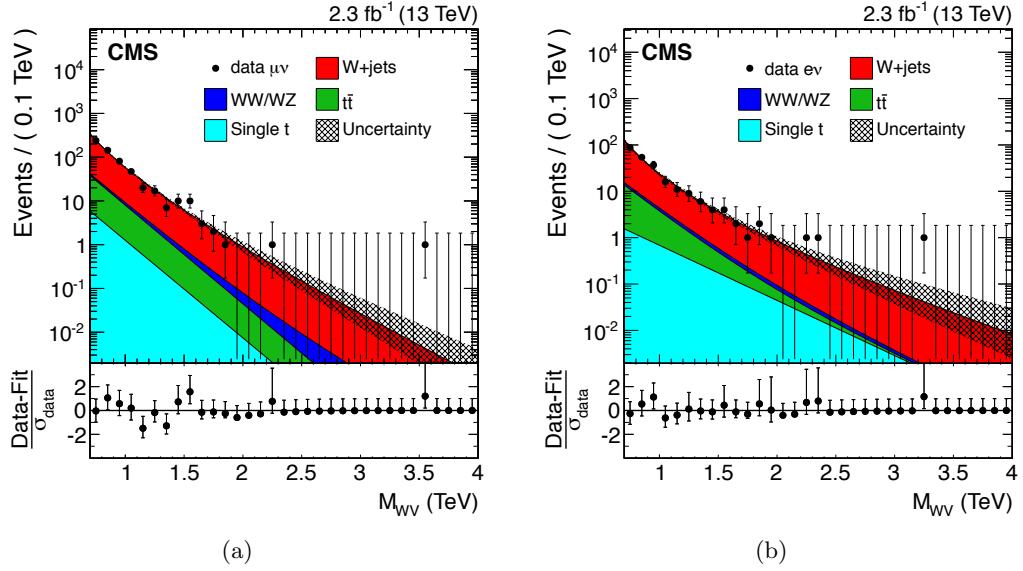
In the above equation, the sum runs over the products of the normalization  $N_{MC,SR}^k$  and probability density function  $F_{MC,SR}^k$  of each minor background contribution  $k$ , while  $N_{SR}^{W+jets}$  and  $F_{data,SB}^{W+jets}$  represent the normalization and probability density function of the W+jets background derived from data as described previously in this chapter. The transfer function  $\alpha_{MC}$  accounts for small kinematic differences between the signal and the sideband regions.

Results of the final background extraction in the signal region will be presented in Chapters 9 and 10 or the  $\ell\nu bb$  and  $\ell\nu q\bar{q}$  analysis, respectively.

#### 8.2.4 Validation of the $\alpha$ method

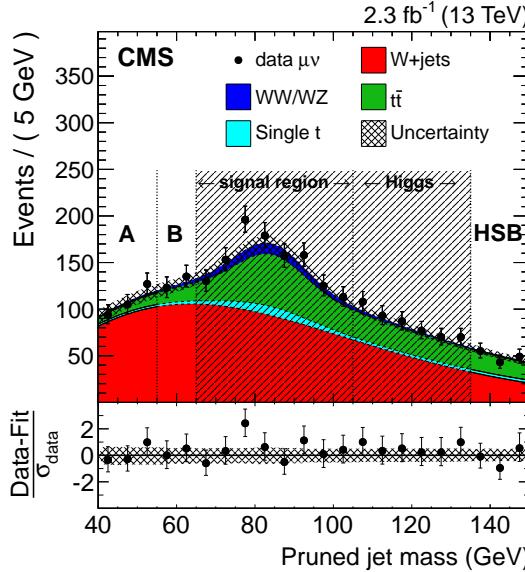
To test the validity and the robustness of the data driven method used to estimate the W+jets contribution and described previously in this section, a closure test is performed. In this test, the background is extracted to a signal free control region that allows to check the compatibility with data for both the distribution and normalization. In order to achieve this, the low mass sideband defined in Table 8.3 is divided into two regions:  $40 < m_{jet} < 55$  GeV, referred to as “region A”, is used as sideband, while  $55 < m_{jet} < 65$  GeV, referred to as “region B”, is used as signal region. The W+jets background normalization is then predicted in region B by performing a fit to the  $m_{jet}$  distribution of the data in region A and in the upper sideband (Table 8.3), while its distribution in  $m_{\ell\nu+jet}$  is extrapolated in region B with a fit of the data in region A and a suitable transfer function  $\alpha_{MC}$ . In this test, the  $\alpha_{MC}$  is defined as the ratio between the simulated W+jets background distributions in  $m_{\ell\nu+jet}$  in region B and A.

An example of the result of this test is presented in the following for the muon channel in



**Figure 8.8:** Results of the fit to the  $m_{WW}$  distribution of the data with  $m_{jet}$  in the lower sideband to estimate  $F_{data,SB}^{W+jets}$  for both muon (a) and electron (b) channels of the  $\ell\nu q\bar{q}$  analysis. Minor backgrounds are estimated from simulation, while the W+jets contribution is the result of the fit to the data.

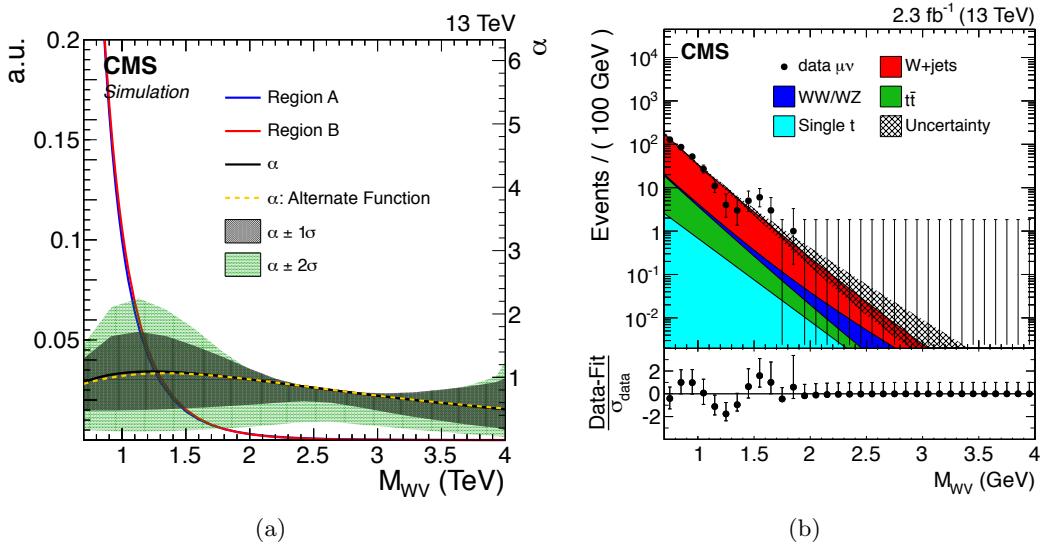
2452 the  $\ell\nu q\bar{q}$  analysis. Figure 8.9 shows the results of the fit to the  $m_{jet}$  distribution of the data  
 2453 inside the region A and the HSB, performed to extract the expected W+jets normalization  
 2454 inside the region B.



**Figure 8.9:** Result of the closure test for the muon channel in the  $\ell\nu q\bar{q}$  analysis. The plot shows the fit to the pruned jet mass distribution considering only events in data with  $m_{jet}$  in the ranges 40–55 GeV (A) and 135–150 GeV (HSB) performed to extract the W+jets normalization inside region B.

2455 Figure 8.10(a) shows the transfer function  $\alpha_{MC}$  obtained from a simultaneous fit of  
 2456 W+jets simulated events in the region A and in the region B, using the leveled-exponential  
 2457 parametrization defined in Eq. 8.3. In Fig. 8.10(b), the result of the fit to the  $m_{\ell\nu+jet}$

2458 distribution of the data with  $m_{\text{jet}}$  in the lower sideband is shown, where the W+jets shape is  
 2459 modelled through the same leveled-exponential function.



**Figure 8.10:** (a) The transfer function  $\alpha_{\text{MC}}$  obtained by simultaneously fitting the diboson invariant mass distributions of simulation data inside the sideband (A) and signal region (B). (b) Diboson invariant mass distribution for events with  $40 < m_{\text{jet}} < 55$  GeV (A). The W+jets shape is fitted, after subtracting contaminations from minor backgrounds, by means of a leveled-exponential function.

2460 Finally, Fig. 8.11 shows a comparison between the total predicted background, obtained  
 2461 through Eq. 8.4, and the data inside the signal free region B. A good agreement is found over  
 2462 the whole  $m_{\ell\nu+\text{jet}}$  range. The test has been performed for both lepton flavours for the  $\ell\nu q\bar{q}$   
 2463 analysis, as well as for the  $\ell\nu b\bar{b}$  analysis where slightly different definitions for region A and  
 2464 B are used. In all the cases, consistency between the predicted background and the data is  
 2465 observed, thus validating the proposed strategy for the W+jets background estimation.

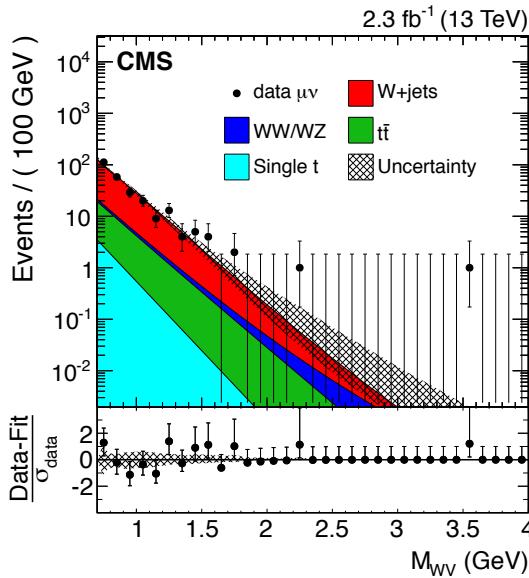
2466 **FIXME: consider adding results with dijet function here or in the appendix**

### 2467 8.3 Modelling of top quark production

2468 The backgrounds from  $t\bar{t}$  and single top quark production in both analysis channels are esti-  
 2469 mated from data-based correction factors in the normalization of the simulation. A top quark  
 2470 enriched control sample is selected by applying all the analysis requirements except that the  
 2471 b jet veto is inverted by requiring, instead, at least one b-tagged AK4 (or AK5) jet in the event.  
 2472

2473 For the  $\ell\nu q\bar{q}$  channel, the comparison between data and simulation yields normalization  
 2474 correction factors for  $t\bar{t}$  and single top quark background processes evaluated in the pruned  
 2475 jet mass signal region  $65 < m_{\text{jet}} < 105$  GeV. The measured correction factors are  $0.87 \pm$   
 2476  $0.04$  and  $0.83 \pm 0.07$  for the muon and electron channel, respectively, where the quoted  
 2477 uncertainty is only statistical. The disagreement is consistent with the difference between  
 2478 NLO and NNLO shape prediction for large top quark  $p_T$  [99].  
 2479

2480 For the  $\ell\nu b\bar{b}$  channel, a unique correction factor is calculated with a simultaneous fit  
 2481 to number of data events in the muon and electron channels in the pruned jet mass region  
 2482  $40 < m_{\text{jet}} < 150$  GeV. The difference in normalization between data and simulation is found



**Figure 8.11:** Distributions in diboson invariant mass for data and the expected backgrounds for events inside the pruned mass region defined by  $55 < m_{jet} < 65$  GeV (B). The W+jets background distribution is extracted using events within  $40 < m_{jet} < 55$  GeV (A).

2483 to be  $4.6 \pm 5.6\%$ , where the quoted uncertainty is only statistical.

2484

2485 These scale factors include both the W boson signal and the combinatorial components  
 2486 mainly due to events where the extra b jet from the top quark decay is in the proximity of  
 2487 the W, and are used to correct the normalization of the tt and single top quark simulated  
 2488 background predictions in the signal regions. The relative uncertainties are used to quantify  
 2489 the uncertainty in the tt and single top quark background normalization.

2490

2491 The  $m_{jet}$  distribution in the top quark enriched sample for the 13 TeV data  $\ell\nu q\bar{q}$  analysis  
 2492 and for simulation is shown in Fig. 8.12(a), while Fig. 8.12(b) shows the  $\tau_{21}$  distribution. The  
 2493 same distribution is also shown for the  $\ell\nu b\bar{b}$  analysis channel in Fig. 8.13, where 8 TeV data  
 2494 and simulation are compared. In all cases, the  $m_{jet}$  spectrum shows a clear peak for events with  
 2495 a W boson decaying to hadrons, including the combinatorial background, while a reasonable  
 2496 agreement between the shapes in data and simulation is observed. Comparisons of data  
 2497 and simulation are also shown in Fig. 8.14 for other distributions such as the reconstructed  
 2498  $m_{\ell\nu+jet}$ , as well as  $m_{top}^l$  and  $m_{top}^h$ . In the latter a clear peak at the top quark mass is visible.

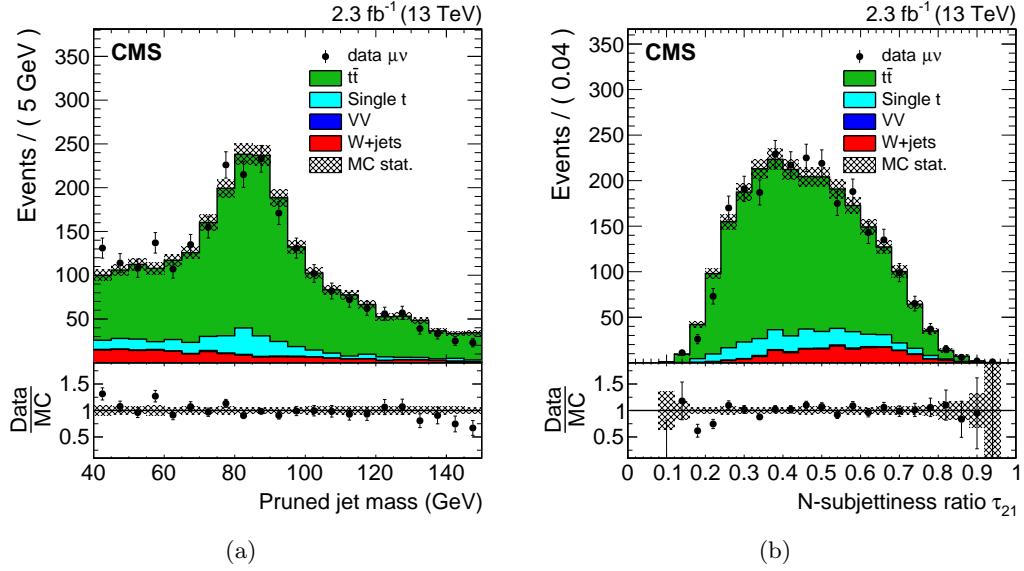
2498

## 8.4 Signal modeling

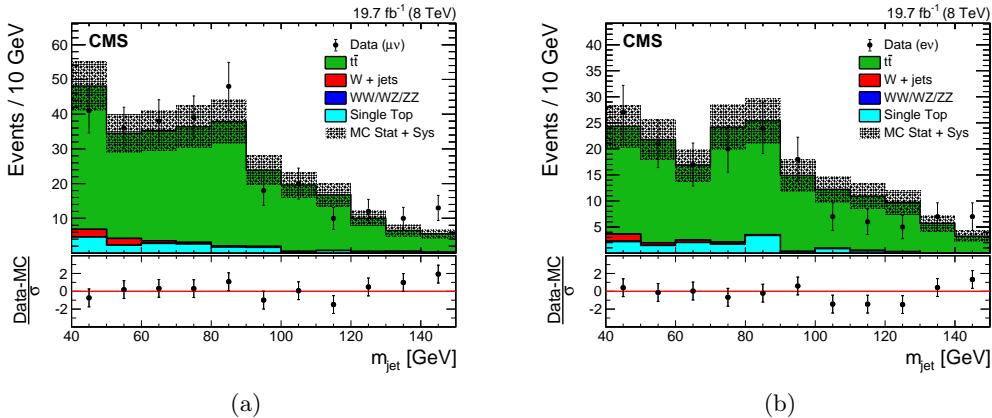
2499 The potential discovery and exclusion power of these analyses rely on the ability of finding a  
 2500 local enhancement on the top of a smoothly falling background. This is ultimately achieved  
 2501 through an unbinned likelihood fit of the signal + background model to the reconstructed  
 2502 diboson invariant mass, which depends on the accurate description of the signal shape.

2503

2504 An analytical parametrization of the signal shape is chosen such that it well reproduces  
 2505 the simulated resonance distributions. As stated in Section 5.2.1, simulated signal events are  
 2506 generated with a resonance natural width sufficiently small compared to the detector resolution.  
 2507 This makes the model used for generating the events independent from the detector effects on  
 2508 the signal shape, allowing a model independent search for narrow resonances where only the  
 2509 detector resolution has to be described. A double-sided Crystal-Ball (CB) function [100] (i.e.



**Figure 8.12:** Distributions in the N-subjettiness ratio  $\tau_{21}$  (a) and pruned jet mass  $m_{\text{jet}}$  (b) from the top quark enriched control sample in the muon channel of the  $\ell\nu q\bar{q}$  analysis. The  $t\bar{t}$  background is rescaled such that the total number of background events matches the number of events in 13 TeV data.

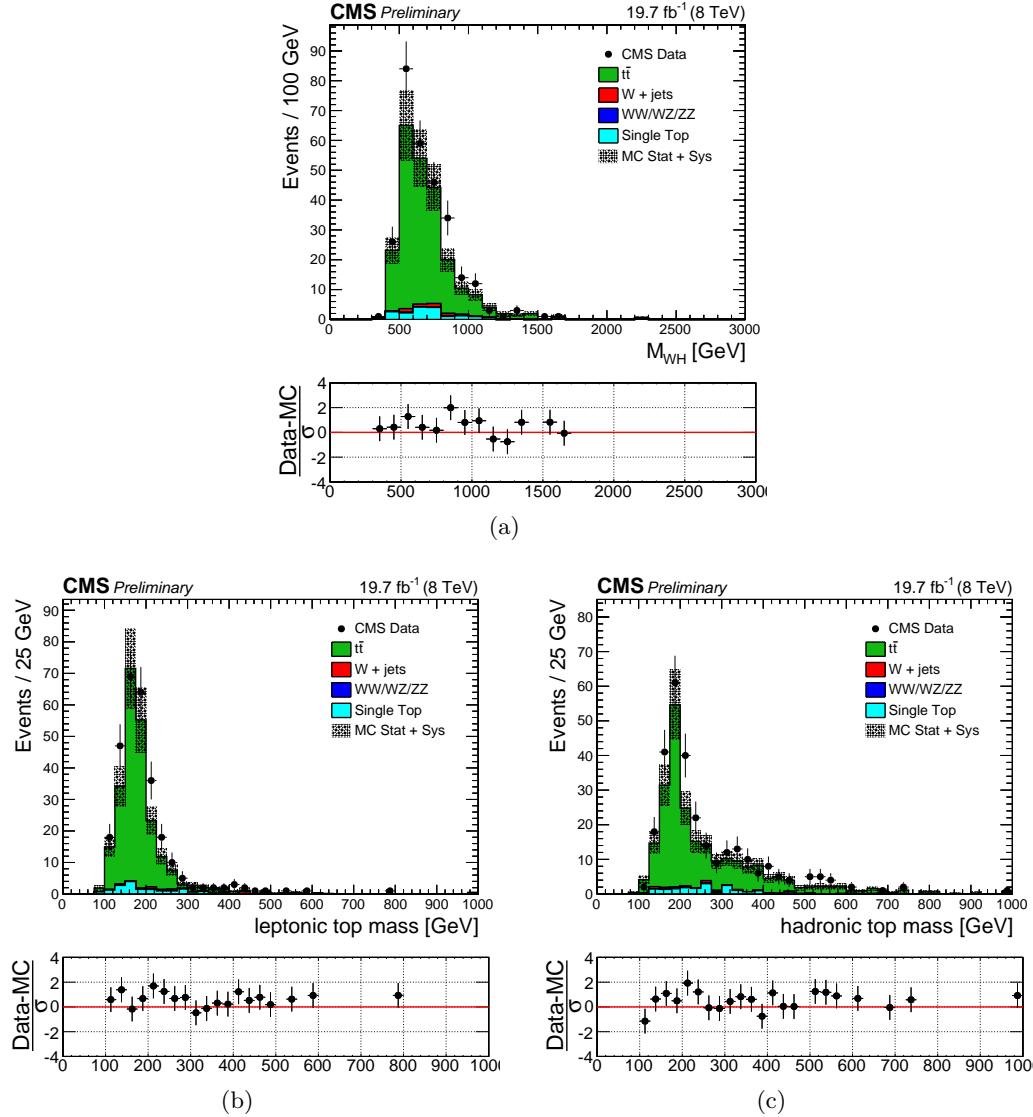


**Figure 8.13:** Distributions in pruned jet mass  $m_{\text{jet}}$  in the top quark enriched control sample in the electron (a) and muon (b) channels of the  $\ell\nu b\bar{b}$  analysis. The hatched region indicates the overall uncertainty in the background. In the lower panels, the bin-by-bin residuals,  $(\text{Data} - \text{MC})/\sigma$  are shown, where  $\sigma$  is the sum in quadrature of the statistical uncertainty of the 8 TeV data, the simulation, and the systematic uncertainty in the  $t\bar{t}$  background.

2509 a Gaussian core with power law tails on both sides) is found to well serve this purpose. To  
 2510 take into account differences between muon and electron momentum resolutions, the signal  
 2511 invariant mass distribution is parametrized separately in the two lepton flavor categories.

2512 Figure 8.15 shows examples of the fitted signal distribution through a CB function, for  
 2513 several signal benchmarks and in different  $m_{\text{jet}}$  categories of the  $\ell\nu q\bar{q}$  analysis. Similar results  
 2514 are obtained for the  $W'$  signal used in the  $\ell\nu b\bar{b}$  analysis.

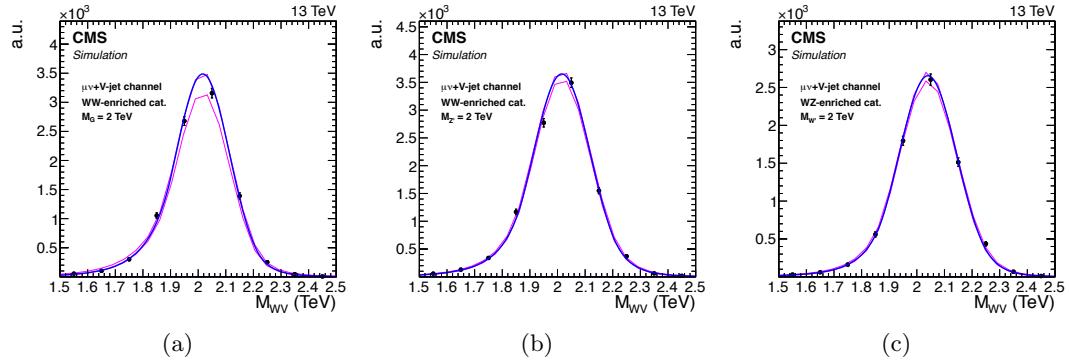
2515 Because of the limited number of available simulated samples, a liner interpolation is  
 2516 performed for each parameter of the CB function between the shapes obtained for some  
 2517 reference mass points, in order to extrapolate the distribution for intermediate values of the



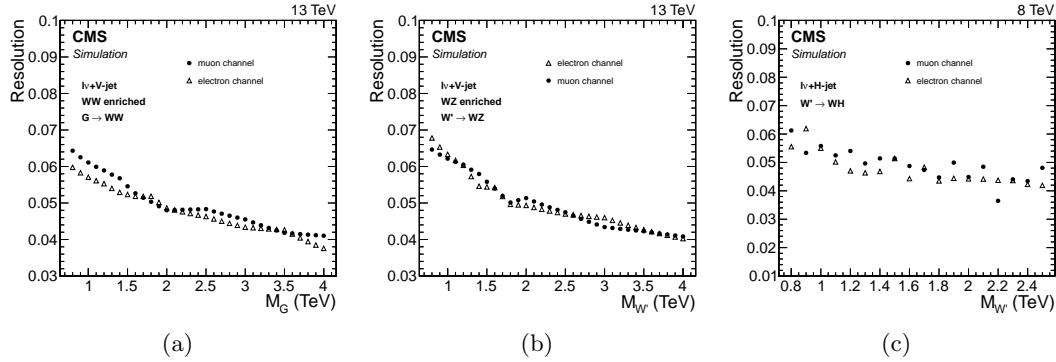
**Figure 8.14:** Distributions for 8 TeV data and for simulation in  $m_{\text{WH}}$  (a),  $m_{\text{top}}^l$  (b) and  $m_{\text{top}}^h$  (c) in the top quark enriched control sample for the muon channel of the  $\ell\nu b\bar{b}$  analysis.

resonance mass. The resolution of the reconstructed diboson invariant mass is given by the width of the Gaussian core and it ranges between 7 and 4% depending on the resonance mass, as summarized in Fig. 8.16. The resolution is dominated by the jet and  $E_T^{\text{miss}}$  contributions.

The signal selection efficiency, evaluated for each category, is defined as the number of selected signal events over the number of generated ones, which include all the possible lepton flavours ( $e$ ,  $\mu$  and  $\tau$ ). As shown in Fig. 8.17 the efficiency for a  $Z'$  or bulk graviton signal in the WW-enriched category is  $\approx 2$  times larger compared to a  $W'$  signal. On the other hand, the efficiency for a  $W'$  signal in the WZ-enriched category is  $\approx 4$  times larger compared to a  $Z'$  or bulk graviton signal. For both categories and for each signal hypothesis the efficiency is smaller compared to the large  $m_{\text{jet}}$  window used for V tagging. However, the resulting loss in sensitivity in each of the category is recovered with a combination of the two  $m_{\text{jet}}$  categories which allows the use of all the available data. With this solution the discrimination between the two type of signals is maximized together with a gain in sensitivity of 10–20% depending on the resonance mass.

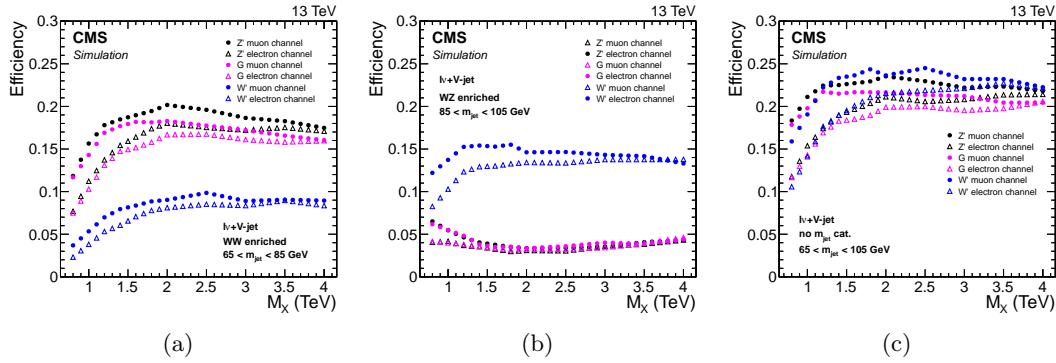


**Figure 8.15:** Modeling of the reconstructed signal distribution with a double-sided Crystal Ball function, for different signal benchmarks and in different  $m_{\text{jet}}$  categories of the  $\ell\nu q\bar{q}$  analysis: bulk graviton (a) and  $Z'$  (b) signals in the WW-enriched category; (c)  $W'$  signal in the WZ-enriched category. In all cases, a signal sample with a generated mass of 2 TeV is considered.

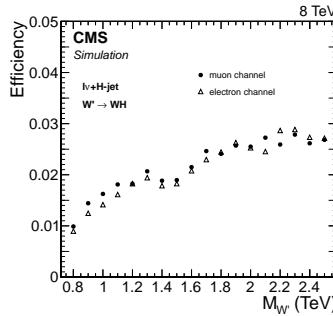


**Figure 8.16:** Relative resolution of the fitted signal distribution as given by the width of the Gaussian core, as a function of the generated resonance mass for different signal benchmarks and for the two analysis: bulk graviton (a) and  $W'$  (b) signals in the WW-enriched and WZ-enriched category, respectively, of the  $\ell\nu q\bar{q}$  analysis; (c)  $W'$  signal for the  $\ell\nu b\bar{b}$  analysis.

2532 A linear interpolation of the signal efficiency is performed between the values obtained for  
 2533 some reference mass points in order to extrapolate the efficiency for intermediate resonance  
 2534 masses for which a simulated sample is not available. The efficiency for the electron channel  
 2535 is lower compared to the muon channel over most of the phase space due to the tighter  
 2536 requirements on the electron  $p_T$  and  $E_T^{\text{miss}}$ . This effect is less visible in the  $\ell\nu b\bar{b}$  channel  
 2537 (Fig. 8.18) where the electron selections are less strict. For all cases, at low masses the  
 2538 efficiency increases with the resonance mass because of the increase in the acceptance of the  
 2539 lepton,  $E_T^{\text{miss}}$  and  $m_{WV}/WH$  selections together with the inefficiency of the jet algorithms in  
 2540 reconstructing the merged jet for a low boosted V boson (Fig. 6.13). At larger resonance  
 2541 masses the efficiency slightly decreases due to  $\tau_{21}$  selection inefficiency for very high  $p_T$  V  
 2542 jets, as described in Section 7.2. For the electron channel this effect is compensated by a  
 2543 larger increase in the lepton selection acceptance, resulting in a nearly flat efficiency at high  
 2544 resonance masses. Similar considerations hold for the efficiency in the  $\ell\nu b\bar{b}$  channel shown in  
 Fig. 8.18.



**Figure 8.17:** Signal efficiency in the  $\ell\nu q\bar{q}$  analysis channel as a function of the generated resonance mass for all signal benchmarks and for different  $m_{\text{jet}}$  selection: (a) WW-enriched category; (b) WZ-enriched category; (c)  $65 < m_{\text{jet}} < 105 \text{ GeV}$ .



**Figure 8.18:** Signal efficiency in the  $\ell\nu b\bar{b}$  analysis channel as a function of the generated  $W'$  mass.

## 2546 8.5 Systematic uncertainties

2547 This section describes the systematic uncertainties in the signal ad background predictions  
 2548 affecting both the normalizations and the  $m_{\ell\nu+\text{jet}}$  distributions. The uncertainties described  
 2549 below are include as nuisance parameters in the calculation of the limits on the cross section  
 2550 as well as of the p-values of potential excesses of events observed in the data.

### 2551 8.5.1 Systematic uncertainties in the background estimation

2552 The uncertainty in the  $W+\text{jets}$  background normalization is mainly due to the uncertainties  
 2553 in the parameters extracted from the fit of the data in the pruned jet mass sideband. This  
 2554 contribution is statistical in nature since it depends on the amount of data in the  $m_{\text{jet}}$  sideband  
 2555 regions, and it is evaluated varying the fit parameters from the final fit values by random  
 2556 amounts sampled from the covariance matrix. An additional effect due to the difference  
 2557 arising from alternative parametrization of the  $W+\text{jets}$   $m_{\text{jet}}$  distribution is taken into account  
 2558 and added in quadrature to the pure statistical contribution. This contribution is found to  
 2559 constitute up to 15% of the total uncertainty. The total uncertainty on the  $W+\text{jets}$  yields  
 2560 remains below 10% in the  $\ell\nu q\bar{q}$  channel, while uncertainties above 40% are obtained for the  
 2561  $\ell\nu b\bar{b}$  channel where the amount of data in the sidebands is largely reduced by the tight b  
 2562 tagging requirements.

2563 As described in Section 8.2.3 the extrapolated background shape in the signal region is  
 2564 computed from the product of  $F_{\text{data},\text{SB}}^{\text{W+jets}}$  and  $\alpha_{\text{MC}}$ . Thus, the shape uncertainty comes from  
 2565 both uncertainties in the  $W+\text{jets}$   $m_{\ell\nu+\text{jet}}$  shape obtained from the fit of the data in the lower

2566  $m_{\text{jet}}$  sideband region and in the modelling of the transfer function  $\alpha_{\text{MC}}$ . Both contributions  
 2567 are mainly statistical in nature, as they are driven by the available amount of data in the  
 2568 sideband and by the number of simulated W+jets events passing the analysis requirements,  
 2569 respectively. These effects are estimated from the covariance matrix of the fit and included in  
 2570 the final limit and p-value calculations after a procedure which diagonalizes the matrix to  
 2571 decorrelate the fitted parameters. In this procedure, the new parameters are defined in such a  
 2572 way to be centered at zero and with error equal to unity. The background fit parameterization  
 2573 is then redefined as a function of these new, uncorrelated parameters. This new fit function  
 2574 together with the uncertainties in the fitted parameter is used to describe the background  
 2575 distribution in the limit and p-value calculations explained in Section 8.6.

2576 Additionally, the  $\alpha_{\text{MC}}$  (Fig. 8.7) is affected by variations due to the choice of the para-  
 2577 metrization used to model the W+jets distribution. Previous studies showed that additional  
 2578 variations of about the same size are due to the use of different parton showering algorithms [93].  
 2579 This effect has been evaluated comparing the  $\alpha$  obtained with simulated samples with parton  
 2580 showering implemented through HERWIG++ and PYTHIA. All these variations are found  
 2581 to be equal or slightly smaller than the statistical uncertainties on the  $\alpha$ , and hence the  
 2582 associated systematic effect is taken into account by enlarging the errors on the decorrelated  
 2583 fit parameters by a factor  $\sqrt{2}$ . This is sufficiently conservative to cover all the shape variations.  
 2584 In a similar way, variations in the  $F_{\text{data},\text{SB}}^{\text{W+jets}}$  due to the same effects, are as well taken into  
 2585 account.

2586 The uncertainties in the W+jets normalization are treated as uncorrelated among the  
 2587 different lepton flavor channels and  $m_{\text{jet}}$  categories, while the uncertainties in the W+jets  
 2588 distribution are partially correlated according to the following scheme:

- 2589     • uncertainties in the  $F_{\text{data},\text{SB}}^{\text{W+jets}}$  parameters are correlated;
- 2590     • uncertainties in the  $\alpha_{\text{MC}}$  parameters are uncorrelated.

2591     This solution takes into account the fact that in the different  $m_{\text{jet}}$  categories the same  
 2592 data in the sideband are used to estimate the W+jets distribution, while the transfer function  
 2593 is used to predict the shape in the two orthogonal signal regions defined by the categories.

2594     The systematic uncertainty in the normalization of the  $t\bar{t}$ /single top quark backgrounds  
 2595 is driven by the uncertainties in the data-to-simulation scale factors estimated in the top  
 2596 quark enriched control sample (Section 8.3). In the  $\ell\nu q\bar{q}$  channel these uncertainties are  
 2597 measured to be 4.6% and 8.4% in the muon and electron channel, respectively. For the  $\ell\nu b\bar{b}$   
 2598 channel, this uncertainty amounts to 5.6%. For the single top quark background an additional  
 2599 systematic uncertainty related to the cross section calculations is assigned to be 15% and 5%,  
 2600 for the 8 and 13 TeV data analysis respectively [101, 102].

2601     The  $t\bar{t}$  background distribution in  $m_{\ell\nu+\text{jet}}$  is taken from simulation and this choice is  
 2602 found to be reasonable given the agreement between data and simulation in the top quark  
 2603 enriched control sample (Fig. 8.14(a)). However, previous studies [93] showed that variations  
 2604 in the shape occur due to the choices of regularization or factorization scales (varied up and  
 2605 down by a factor of 2), to the matching scales in the MADGRAPH simulation, and to different  
 2606 generators (MADGRAPH or POWHEG). In order to cover all these effects, the errors on the  
 2607 decorrellated fit parameters for the  $t\bar{t}$  distribution is enlarged by a factor of 2.

2608     The systematic uncertainties in the diboson background normalization is due to the  
 2609 uncertainty in the inclusive cross sections, which are assigned to be 10% [103] and 3% [104]  
 2610 for the 8 and 13 TeV data analysis, respectively. For the  $\ell\nu q\bar{q}$  channel, the uncertainty in the

2613 diboson background normalization is as well due to the uncertainty of 3% in the measured  
 2614 data-to-simulation scale factors for the V tagging efficiency derived in the top quark enriched  
 2615 control sample (Section 7.2).

2616 Additional sources of systematic uncertainties in the background normalization are due  
 2617 to the uncertainty in the integrated luminosity, and in the measured data-to-simulation scale  
 2618 factors for the efficiency of lepton trigger and identification, described in the following section.

2620 A summary of the systematic uncertainties in the normalization of the predicted back-  
 2621 ground is provided in Tables 8.5 and 8.6 for the  $\ell\nu q\bar{q}$  and  $\ell\nu b\bar{b}$  analysis channel, respectively.

Source	W+jets	t <bar>t</bar>	single top quark	diboson
Integrated luminosity	-	2.6%	2.6%	2.6%
Cross section	-	-	15%	10%
Data-driven prediction	42% ( $\mu$ ) / 59% (e)	5.6%	5.6%	-
Lepton trigger ( $\mu/e$ )	-	1% / 1%	1% / 1%	1% / 1%
Lepton identification ( $\mu/e$ )	-	1% / 3%	1% / 3%	1% / 3%

**Table 8.5:** Summary of the systematic uncertainties in the normalization of the predicted background in the  $\ell\nu bb$  analysis at 8 TeV.

Source	W+jets	t <bar>t</bar>	single top quark	diboson
Integrated luminosity	-	2.7%	2.7%	2.7%
Cross section	-	-	5%	3%
V-tagging efficiency	-	-	-	3%
Data-driven prediction	5–9%	5–8%	5–8%	-
Lepton trigger ( $\mu/e$ )	-	1% / 1%	1% / 1%	1% / 1%
Lepton identification ( $\mu/e$ )	-	1% / 3%	1% / 3%	1% / 3%

**Table 8.6:** Summary of the systematic uncertainties in the normalization of the predicted background in the  $\ell\nu q\bar{q}$  analysis at 13 TeV.

### 2622 8.5.2 Systematic uncertainties in the signal prediction

2623 Systematic uncertainties affecting the predicted signal efficiency (or normalization) and  
 2624  $m_{\ell\nu+jet}$  distribution arise from several sources as described in the following and summarized  
 2625 in Tables 8.7 and 8.8. The effect of each source is evaluated for each considered simulated  
 2626 signal hypothesis as a function of the resonance mass.

2627 One the primary sources affecting the signal normalization for the  $\ell\nu q\bar{q}$  channel is due to  
 2628 uncertainties in data-to-simulation scale factors for the V tagging efficiency, derived from  
 2629 top quark enriched control sample as described in Section 7.2. These uncertainties include  
 2630 separately the uncertainty of 3% on the scale factor measured in tt events with an average  $p_T$   
 2631  $\approx 200$  GeV, and the uncertainty due to the extrapolation of the scale factor to higher momenta,  
 2632 which is assigned to be 6–10% depending on the signal mass. Additional uncertainties are  
 2633 assigned due to the pruned jet mass scale and resolution measured in tt events (Table 7.2).  
 2634 These are computed by rescaling or smearing the  $m_{jet}$  value according to the uncertainties in  
 2635 the respective  $m_{jet}$  scale or resolution. The selection efficiencies are recalculated on these

modified events, with the resulting changes taken as systematic uncertainties that depend on the resonance mass.

In a similar way, systematic uncertainties are assigned in the  $\ell\nu b\bar{b}$  channel due to the uncertainty in the H tagging efficiency. This contribution arises from both uncertainties in the data-to-simulation scale factors for the pruned jet mass scale and resolution, derived from the top quark enriched control sample with 8 TeV data, and for b-tagged jet identification efficiencies (Section 6.4.3). These sources introduce a systematic uncertainty in the mass tagging and b tagging of the Higgs boson of 2–10% and 2–8%, respectively, depending on the signal mass.

2646

The accuracy on energy and momentum measurements for leptons and jets represents an important source of systematic uncertainties in the signal efficiency. In particular, the muon momentum scale and resolution, the electron energy scale and resolution, and the jet energy scale and resolution are considered. The event selection is applied to the signal samples after varying the lepton four-momenta within one standard deviation of the corresponding uncertainty in the muon momentum scale [57] or electron energy scale [105], or applying an appropriate Gaussian momentum/energy smearing in case of resolution uncertainties. The same procedure is also applied for the jet four-momenta using the corresponding energy scale and resolution uncertainties. In this process, variations in the lepton and jet four-momenta are propagated consistently to the  $\vec{p}_T^{\text{miss}}$  vector. The signal efficiency is then recalculated using modified lepton and jet four-momenta separately for each source of systematic uncertainties. The largest relative change in the signal efficiency compared to the default value is taken as the systematic uncertainty for that specific source. The induced relative migration among V jet mass categories is evaluated for the  $\ell\nu q\bar{q}$  channel, but do not affect the overall signal efficiency. The muon, electron, and jet uncertainties are assumed to be uncorrelated. Finally, the resulting changes on the reconstructed resonances are propagated on the reconstructed  $m_{\ell\nu+\text{jet}}$  signal distribution, resulting in a small effect on both peak position and width of the Gaussian core.

2665

The systematic uncertainties in the lepton trigger, identification, and isolation efficiencies are derived using a dedicated T&P analysis in  $Z \rightarrow \ell^+\ell^-$  events. For both analysis channels, an uncertainty of 1% is assigned to the trigger efficiency for both lepton flavors, while for lepton identification and isolation efficiency, the systematic uncertainty is estimated to be 1% for the muon and 3% for electron flavors.

2671

The 2.7% and 2.6% uncertainty in the integrated luminosity affects to the normalization of both signal and backgrounds in the  $\ell\nu q\bar{q}$  and  $\ell\nu b\bar{b}$  channel, respectively, as obtained in measurements performed for the 2015 and 2012 data taking periods [106, 107].

2675

For the  $\ell\nu q\bar{q}$  channel, uncertainties on the signal yield due to variations in the parton distribution function and the choice of factorization ( $\mu_f$ ) and renormalization ( $\mu_r$ ) scales are also taken into account. The PDF uncertainties are evaluated using the NNPDF 3.0 [22] PDF set. The uncertainty related to the choice of  $\mu_f$  and  $\mu_r$  scales is evaluated following the proposal in Refs. [108, 109] by varying the default choice of scales in the following 6 combinations of factors:  $(\mu_f, \mu_r) \times (1/2, 1/2), (1/2, 1), (1, 1/2), (2, 2), (2, 1)$ , and  $(1, 2)$ . The uncertainty in the signal cross section from the choice of PDFs and of factorization and renormalization scales ranges from 4 to 77%, and from 1 to 22%, respectively, depending on the resonance mass, particle type and its production mechanism. For the  $\ell\nu b\bar{b}$  channel, only the impact of the proton PDF uncertainties on the signal efficiency is evaluated with the PDF4LHC prescription [110, 111], using the MSTW2008 [112] and NNPDF 2.1 [113] PDF

sets. This effect is found to be < 0.5%.

Finlay, the systematic uncertainty due to the modelling of pileup is estimated by reweighting the signal simulation samples such that the distribution of the number of interactions per bunch crossing is shifted according to the uncertainty in the inelastic proton-proton cross section compared with that found in data. This contribution is found to be 0.5% in both channels.

**Table 8.7:** Summary of the systematic uncertainties in the signal prediction for the  $\ell\nu b\bar{b}$  analysis channel and their impact on the event yield in the signal region and on the reconstructed  $m_{WH}$  shape (mean and width) for both muon and electron channels.

Source	Relevant quantity	Uncertainty (%)
Lepton trigger ( $\mu/e$ )	Signal yield	1 / 1
Lepton identification ( $\mu/e$ )	Signal yield	1 / 3
Lepton $p_T$ scale ( $\mu/e$ )	Signal yield	1 / 0.5
Lepton $p_T$ resolution ( $\mu/e$ )	Signal yield	0.1 / 0.1
Jet energy scale	Signal yield	1–3
Jet energy resolution	Signal yield	0.5
Integrated luminosity	Signal yield	2.6
Pileup	Signal yield	0.5
PDFs	Signal yield	< 0.5
H jet mass tagging efficiency	Signal yield	2–10
H jet b tagging efficiency	Signal yield	2–8
Jet energy scale	Resonance shape (mean)	0.5
Jet energy scale	Resonance shape (width)	4
Jet energy resolution	Resonance shape (mean)	0.2
Jet energy resolution	Resonance shape (width)	4
Lepton $p_T$ resolution	Resonance shape (mean)	0.1
Lepton $p_T$ resolution	Resonance shape (width)	1.2
Lepton $p_T$ scale	Resonance shape (mean)	0.7
Lepton $p_T$ scale	Resonance shape (width)	2.5

## 8.6 Testing new resonance hypothesis

The purpose of this analysis is to infer a constraint on the existence of a new resonance decaying into diboson for a set of different signal mass hypotheses. The comparison between the diboson invariant mass distribution observed in data and the SM background prediction is used to check for the presence of the new resonance. A hypothesis test is built to decide between a null hypothesis given by the predicted SM background only, against an alternative hypothesis which includes both background as well as the sought after signal. In principle one can either test the background-only hypothesis and exclude it if there is a large deviation of the data from the SM background prediction, or test the signal hypothesis and exclude it if there is a large deviation of the data from the expected signal model. In particular, if no significant deviation from the SM background prediction is observed in data, compatible with the signal hypothesis, an upper limit on production cross section of such signal is usually set, up to a certain degree of belief. The CMS community has agreed upon a procedure for computing upper limits, which is based on the modified frequentist method, often referred to as  $CL_s$ . While a detailed description of such method can be found in Refs. [114, 115], the basic ingredients will be summarized Section 8.6.1. A description of the procedure followed

**Table 8.8:** Summary of the systematic uncertainties in the signal prediction for the  $\ell\nu q\bar{q}$  analysis and their impact on the event yield in the signal region and on the reconstructed  $m_{WV}$  shape (mean and width) for both muon and electron channels. The last uncertainty results in migrations between event categories, but does not affect the overall signal efficiency.

Source	Relevant quantity	Uncertainty (%)
Lepton trigger ( $\mu/e$ )	Signal yield	1 / 1
Lepton identification ( $\mu/e$ )	Signal yield	1 / 3
Lepton $p_T$ scale ( $\mu/e$ )	Signal yield	0.7 / 0.2
Lepton $p_T$ resolution ( $\mu/e$ )	Signal yield	0.1 / 0.1
Jet energy and $m_{jet}$ scale	Signal yield	0.2–4
Jet energy and $m_{jet}$ resolution	Signal yield	0.1–2
Integrated luminosity	Signal yield	2.7
Pileup	Signal yield	0.5
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
V tagging efficiency	Signal yield	3
V tagging $p_T$ -dependence	Signal yield	6–10
Jet energy scale	Resonance shape (mean)	1.3
Jet energy scale	Resonance shape (width)	3
Jet energy resolution	Resonance shape (mean)	0.1
Jet energy resolution	Resonance shape (width)	3
Lepton $p_T$ resolution	Resonance shape (mean)	0.1
Lepton $p_T$ resolution	Resonance shape (width)	0.1
Lepton $p_T$ scale	Resonance shape (mean)	0.1
Lepton $p_T$ scale	Resonance shape (width)	0.5
Jet energy and $m_{jet}$ scale	Migration	2–24

2708 to quantify an excess of events is provided in Section 8.6.3. A summary of the final results  
2709 will be given in the next chapter.

### 2710 8.6.1 Limit setting procedure

2711  
2712 The procedure to establish the exclusion of a given signal hypothesis is based on a  
2713 frequentist significance test which uses a log-likelihood ratio as a test statistic. In order to  
2714 construct the test statistic a likelihood function is defined as

$$\mathcal{L}(data|\mu, \theta) = \text{Poisson}(data|\mu \cdot s(\theta) + b(\theta)) \cdot p(\tilde{\theta}|\theta). \quad (8.5)$$

2715 In this definition,  $s$  and  $b$  denote the expected signal and background event yields,  
2716 respectively, which, before the scrutiny of the observed data entering the statistical analysis,  
2717 are subject to multiple uncertainties that are treated by introducing nuisance parameters  
2718  $\theta$ , so that signal and background expectations depend on these parameters as  $s(\theta)$  and  $b(\theta)$ .  
2719 The exclusion of a signal hypothesis is generally expressed as an upper limit on the *signal*  
2720 *strength modifier*  $\mu$  which scales the cross section used as input in the evaluation of the

expected signal yields. With this definition, the likelihood represents the Poisson probability of observing a certain amount of data when the expected yield is  $\mu \cdot s(\theta) + b(\theta)$  and given the probability  $p(\tilde{\theta}|\theta)$  of measuring a value  $\tilde{\theta}$  for the nominal nuisance parameter  $\theta$ . Note that, in this likelihood definition, “data” stands for a generic dataset, either experimental or a pseudo-data generated randomly.

The likelihood can be either binned or unbinned. In the first case the function  $\text{Poisson}(data|\mu \cdot s + b)$  in Eq. 8.5 is the product of Poisson probabilities for observing  $n_i$  events in each bin  $i$  of the signal+background model

$$\prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}. \quad (8.6)$$

For the unbinned case each event enters the calculation as follows

$$k^{-1} \prod_i (\mu S f_s(x_i) + B f_b(x_i)) e^{-(\mu S + B)}, \quad (8.7)$$

where  $f_s(x)$  and  $f_b(x)$  are the probability density functions of signal and background of the observable  $x$ , while  $S$  and  $B$  are the total event rates expected for signal and background. In this analysis the unbinned form for the likelihood is used, where the observable  $x$  coincides with the reconstructed diboson invariant mass.

To compare the compatibility of the data with the background-only and signal+background hypotheses, where the prediction for the signal is allowed to be scaled by some factor  $\mu$ , the test statistic  $\tilde{q}_\mu$  is constructed based on the profile likelihood ratio as

$$\tilde{q}_\mu = -2 \ln \frac{\mathcal{L}(data|\mu, \hat{\theta}_\mu)}{\mathcal{L}(data|\hat{\mu}, \hat{\theta})}, \quad \text{with} \quad 0 \leq \hat{\mu} \leq \mu. \quad (8.8)$$

Here  $\hat{\theta}_\mu$  denotes the value of  $\theta$  that maximizes the likelihood for the hypothesized  $\mu$ , i.e. it is the conditional maximum-likelihood (ML) estimator of  $\theta$  (and thus is a function of  $\mu$ ). The procedure of refitting the nuisance parameters to maximize the likelihood for each possible value of the parameter of interest  $\mu$ , is usually referred to as “profiling”. The denominator is the maximized (unconditional) likelihood function, i.e.,  $\hat{\mu}$  and  $\hat{\theta}$  are the global maximum of the likelihood. The presence of the nuisance parameters broadens the profile likelihood as a function of  $\mu$  relative to what one would have if their values were fixed. This reflects the loss of information about  $\mu$  due to the systematic uncertainties. Higher values of  $\tilde{q}_\mu$  correspond to increasing incompatibility between the data and the hypothesized signal of strength  $\mu$ . The lower constraint for  $\hat{\mu}$  in the denominator excludes the possibility of negative signal yields. The upper constraint is introduced to avoid that data with  $\hat{\mu} > \mu$  (upward fluctuations) are considered as representing less compatibility with  $\mu$  than what obtained with data.

The observed value of the test statistic,  $\tilde{q}_\mu^{\text{obs}}$  for the given signal strength modifier  $\mu$  under test is computed, as well as the nuisance parameters  $\hat{\theta}_0^{\text{obs}}$  and  $\hat{\theta}_\mu^{\text{obs}}$  maximizing the likelihood under the background-only and signal+background hypothesis, respectively. Furthermore, the probability density functions of the chosen test statistic  $\tilde{q}_\mu$  under the signal+background hypothesis,  $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$ , and the background-only hypothesis, and  $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$ , are constructed by means of ensembles of toy MC pseudo-experiments generated according to the same Poisson probabilities used to build the likelihood. In this process the nuisance parameters are fixed to the values  $\hat{\theta}_0^{\text{obs}}$  and  $\hat{\theta}_\mu^{\text{obs}}$  obtained by fitting the observed data.

Using the  $f(\tilde{q}_\mu|0, \hat{\theta}_0^{\text{obs}})$  and  $f(\tilde{q}_\mu|\mu, \hat{\theta}_\mu^{\text{obs}})$  distributions, two p-values are computed

$$\begin{aligned}
p_\mu \equiv \text{CL}_{s+b} &= P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | \mu s(\hat{\theta}_\mu^{\text{obs}}) + b(\hat{\theta}_\mu^{\text{obs}})) = \int_{\tilde{q}_\mu^{\text{obs}}}^{+\infty} f(\tilde{q}_\mu | \mu, \hat{\theta}_\mu^{\text{obs}}) d\tilde{q}_\mu \\
p_0 \equiv \text{CL}_b &= P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | b(\hat{\theta}_0^{\text{obs}})) = \int_{\tilde{q}_\mu^{\text{obs}}}^{+\infty} f(\tilde{q}_\mu | 0, \hat{\theta}_0^{\text{obs}}) d\tilde{q}_\mu.
\end{aligned} \tag{8.9}$$

2759 The two probabilities are shown in the example in Fig. 8.19(a). In the classical frequentist  
2760 approach, the level of agreement between the data and hypothesized  $\mu$  is evaluated by using  
2761 the  $\text{CL}_{s+b}$  probability only, and one says that the hypothesized signal  $\mu$  is excluded at 95%  
2762 CL if  $\text{CL}_{s+b} \leq 0.05$ .

2763 However, such a definition as a caveat. If the distributions of the test statistic for the  
2764 signal+background and background-only hypotheses have a not negligible overlap as in the  
2765 plot (c) of Fig. 8.19(b), the experiment would tend to exclude the hypothesized signal  $\mu$  even  
2766 if the experiment in this case has little sensitivity to discriminate it against the background.  
2767 In fact, in this case the experimental data are highly contaminated with background and a  
2768 statement about the signal would be a mistake of interpretation. To prevent the inference of  
2769 a signal in such cases, the so-called modified frequentist approach has been introduced at  
2770 the time of LEP [114, 115]. In this approach, the level of agreement between the data and  
2771 hypothesized  $\mu$  is evaluated by using instead the quantity

$$\text{CL}_s = \frac{\text{CL}_{s+b}}{\text{CL}_b}, \tag{8.10}$$

2772 and the hypothesized signal  $\mu$  is excluded at 95% confidence level (CL) if  $\text{CL}_s \leq 0.05$ .  
2773 It is straightforward to see from plot (a) of Fig. 8.19(b) that, if the distribution of the test  
2774 statistic for the signal+background hypothesis is well separated from the background-only  
2775 distribution, then  $\text{CL}_s \sim \text{CL}_{s+b}$  and there is no risk of misinterpretation.

2776 In order to quote, as conventionally done, 95% CL observed upper limits, the full procedure  
2777 is iterated for different values of  $\mu$ , until  $\text{CL}_s = 0.05$  is found. This value of  $\mu$  is denoted as  
2778  $\mu_{95\%}$ , and one can infer that the hypothesized resonance  $X \rightarrow WV/VH$  with a cross section  
2779  $\mu$ -times larger than the one predicted by some specific theoretical model  $\sigma_{th}$  used as input to  
2780 the statistical analysis, is excluded at 95% CL. In this analysis, model-independent limits on  
2781 the cross section are set by rescaling the  $\mu^{95\%} = \sigma_{95\%}/\sigma_{th}$  by the input cross section in order  
2782 to obtain  $\sigma_{95\%}$ .

2783 In addition to the observed upper limit derived from the actual data distribution, it is  
2784 important to study also the expected limit given the observed data. In fact, the expected  
2785 limit quantifies the sensitivity of the experiment independent from statistical fluctuations in  
2786 the data. In order to compute the median-expected upper limit, and the associated  $\pm 1\sigma$  and  
2787  $\pm 2\sigma$  bands, a large set of background-only pseudo-experiments is generated and, for each of  
2788 them, the  $\mu_{95\%}$  is calculated. From the cumulative distribution of  $\mu_{95\%}$ , the median value is  
2789 taken as the expected limit, while the  $\pm 1(2)\sigma$  uncertainty bands on the expected limits are  
2790 extracted from the values of the 16% (2.5%) and 84% (97.5%) quantiles.

### 2791 8.6.2 The asymptotic approximation

2792 In order to compute the  $\text{CL}_s$  the probability density functions of the test statistics are required.  
2793 In particular, one needs the probability density functions  $f(\tilde{q}_\mu | \mu')$ , where  $\mu' = 0$  or  $\mu' = \mu$ ,  
2794 which are obtained from MC toys requiring very expensive computational resources. An  
2795 approximation for the  $\text{CL}_s$  method, valid in the large sample limit, also referred to as “asym-  
2796 totic approximation” has been proposed in Ref. [116] and it is briefly described in the following.

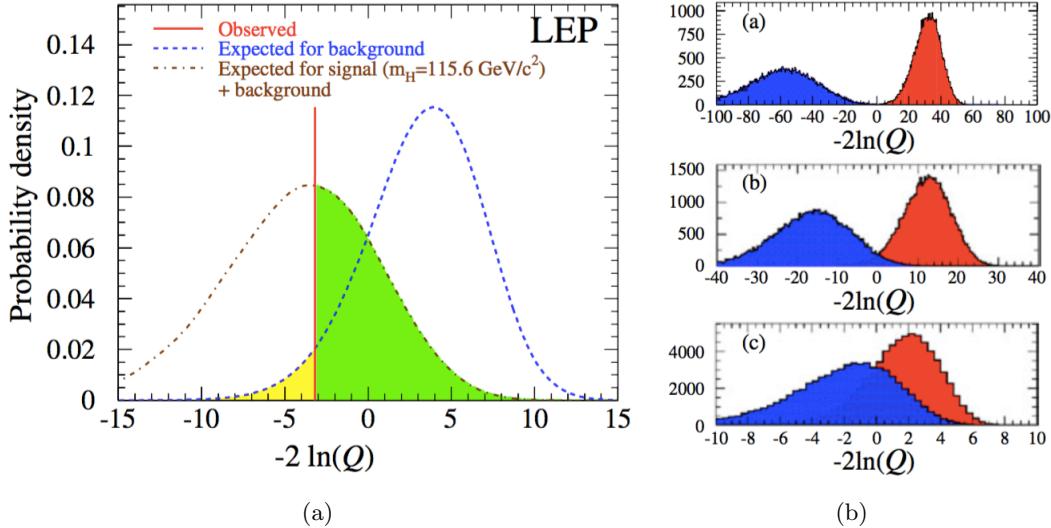


Figure 8.19: blabla

By using the Wald approximation [117] the desired distribution  $f(\tilde{q}_\mu | \mu')$  can be obtained by expressing the test statistic given by the log-likelihood ratio as

$$\tilde{q}_\mu = \frac{(\mu - \hat{\mu})^2}{\sigma^2} + \mathcal{O}(1/\sqrt{N}), \quad (8.11)$$

where  $\hat{\mu}$  follows a Gaussian distribution with a mean  $\mu'$  and standard deviation  $\sigma$ , and  $N$  represents the data sample size. For large data samples ( $N \rightarrow \infty$ ), the  $\mathcal{O}(1/\sqrt{N})$  can be neglected and it can be shown [118] that the distribution  $f(\tilde{q}_\mu | \mu')$  of the test statistic  $\tilde{q}_\mu$  follows a *noncentral chi-square* distribution for one degree of freedom with noncentrality parameter

$$\Lambda = \frac{(\mu - \mu')^2}{\sigma^2}. \quad (8.12)$$

For the special case  $\mu' = \mu$  one has  $\Lambda = 0$  and the test statistic is distributed as a chi-square for one degree of freedom. For the general case in which  $\mu' \neq \mu$ , the standard deviation  $\sigma$  of  $\hat{\mu}$  has to be evaluated, which depends on the MLE estimator of the nominal nuisance parameters. The evaluation of  $\sigma$  is greatly simplified considering a special, artificial data set, referred to as the “Asimov data set”, where all statistical fluctuations are suppressed and the estimators for all parameters are replaced by their expectation values as follows:

$$\hat{\mu} = \mu' \quad \text{and} \quad \hat{\theta} = \theta. \quad (8.13)$$

With these assumptions the test statistic  $\tilde{q}_{\mu,A}$  for the Asimov dataset is given by

$$\tilde{q}_{\mu,A} \approx \frac{(\mu - \mu')^2}{\sigma^2} = \Lambda. \quad (8.14)$$

From the Asimov data set one therefore obtains an estimate of the noncentrality parameter  $\Lambda$  that characterizes the distribution  $f(\tilde{q}_\mu | \mu')$ . Equivalently, the above equation can be used to obtain the variance  $\sigma^2$  which characterizes the distribution of  $\hat{\mu}$ , namely,

$$\sigma_A^2 = \frac{(\mu - \mu')^2}{\tilde{q}_{\mu,A}}, \quad (8.15)$$

so that the distribution obtained by using  $\sigma_A^2$  has a median given by the corresponding Asimov value  $\tilde{q}_{\mu,A}$ . Using these formulae, asymptotic relations are derived which are easily solved for the observed upper limits with the  $CL_s$  method, as well as for the expected median and error bands.

### 8.6.3 Quantifying an excess of events

The presence of the signal is quantified by the background-only p-value, i.e. the probability for the background to fluctuate and give an excess of events as large or larger than the observed one. As for the upper limits, this evaluation requires defining a test statistic and the construction of its probability density function. For a given resonance mass hypothesis  $M_X$ , the test statistic used in this case is  $\tilde{q}_0$ , defined as

$$\tilde{q}_0 = -2 \ln \frac{\mathcal{L}(\text{data}|0, \hat{\theta}_0)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}, \quad \text{with } \hat{\mu} \geq 0. \quad (8.16)$$

The probability density function  $f(\tilde{q}_0|0, \hat{\theta}_0^{\text{obs}})$  is built by generating toy MC pseudo-data under the assumption of the background-only hypothesis. From this distribution, the p-value corresponding to a given experimental observation  $\tilde{q}_0^{\text{obs}}$  is evaluated:

$$p_0 = P(\tilde{q}_0 \geq \tilde{q}_0^{\text{obs}} | b(\hat{\theta}_0^{\text{obs}})) = \int_{\tilde{q}_0^{\text{obs}}}^{+\infty} f(\tilde{q}_0|0, \hat{\theta}_0^{\text{obs}}) d\tilde{q}_0. \quad (8.17)$$

This probability is converted into a *significance*, also referred to as *Z value*, as follows

$$Z = \Phi^{-1}(1 - p_0). \quad (8.18)$$

A significance of  $5\sigma$ , corresponding to a p-value of  $2.87 \times 10^{-7}$ , is conventionally used in high energy physics to claim a discovery, and  $3\sigma$  for an evidence.

It can be demonstrated that in the asymptotic approximation (Section 8.6.2), the likelihood ratio test statistic  $\tilde{q}_0$  follows a chi-square distribution for one degree of freedom, and a fair estimate of the p-value and of the significance can be obtained from the observed value  $\tilde{q}_0^{\text{obs}}$  itself, without the need for generating pseudo-data, as follows

$$p_0 = \frac{1}{2} [1 - \text{erf}(\sqrt{\tilde{q}_0^{\text{obs}}}/2)] \quad (8.19)$$

$$Z = \sqrt{\tilde{q}_0^{\text{obs}}}.$$

The p-value discussed above is evaluated at a fixed resonance mass  $M_X$  and can be referred to as a *local p-value*. In this search, a scan is performed over a wide range of resonance mass hypotheses with the aim of finding the minimum local p-value, which describes the probability of a background fluctuation for that particular resonance mass hypothesis. However, it is important to distinguish the probability of finding a fluctuation in some particular location from the probability of finding such a fluctuation anywhere else. The former is associated to the so called *local significance*, whereas the latter is referred to as the *global significance*. The fact that the global significance is usually smaller than the largest local one is often referred to as the “look-elsewhere effect” (LEE). As demonstrated in Ref. [119], the global and local p-values are related to each other by a multiplicative factor, usually referred to as “trial factor”, proportional to the number of independent search regions. In the asymptotic approximation the trial factor grows linearly with the local significance, through a proportional

2848 constant that is related to the ratio between the mass range under consideration divided by  
2849 its resolution. In particular, it can be shown that

$$\text{trial\#} = \frac{p_{\text{global}}}{p_{\text{local}}} \approx \frac{1}{3} \frac{\text{mass range}}{\text{mass resolution}} Z_{\text{local}}. \quad (8.20)$$

2850 The trial factor is best estimated through MC methods as it will be shown in Section 9.2.  
2851 However, a good agreement with the equation above is obtained.

# Results with 8 TeV data

2854

2855 The final results of the analysis performed with 8 TeV data and focused on the search for  
 2856 a heavy charged resonance decaying into W and Higgs bosons in the  $\ell\nu b\bar{b}$  final state, are  
 2857 presented and discussed in this chapter. In particular, the final observed  $m_{\text{WH}}$  spectrum is  
 2858 used to check for the presence of a new resonance. Firstly, a search is conducted for local  
 2859 enhancement in the  $m_{\text{WH}}$  distribution, which might be due to a signal. As described in the  
 2860 following, since no significant excesses are found, upper limits are set on the production cross  
 2861 section of the new resonance.

## 2862 9.1 Final $m_{\text{WH}}$ distribution

2863 The predicted number of background events in the signal region after the inclusion of all  
 2864 backgrounds is summarized in Table 9.1 and compared with observations. The yields are  
 2865 quoted in the range  $0.7 < m_{\text{WH}} < 3 \text{ TeV}$ . The expected background is derived with the  
 2866 sideband procedure described in Section 8.2. The uncertainties in the background prediction  
 2867 from data are statistical in nature, as they depend on the number of events in the sideband  
 2868 region. The muon channel has more expected background events than the electron channel  
 2869 owing to the lower  $E_T^{\text{miss}}$  requirement and its worse mass resolution at high  $p_T$ .

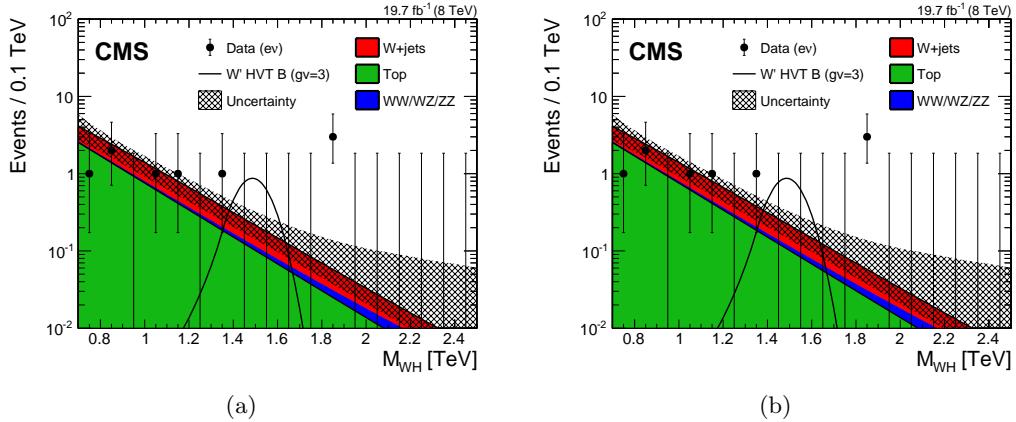
**Table 9.1:** Observed and expected yields in the signal region together with statistical uncertainties.

	$e\nu + \text{H-jet}$	$\mu\nu + \text{H-jet}$
Observed yield	9	16
Expected total background	$11.3 \pm 3.1$	$14.9 \pm 3.1$
W+jets	$4.7 \pm 2.9$	$7.0 \pm 3.1$
Top	$6.3 \pm 1.1$	$7.3 \pm 0.4$
VV	$0.4 \pm 0.1$	$0.6 \pm 0.2$

2870 Figure 9.1 shows the final observed  $m_{\text{WH}}$  spectra after all selection criteria have been  
 2871 applied. The highest mass event is in the electron category and has  $m_{\text{WH}} \approx 1.9 \text{ TeV}$ . The  
 2872 observed data and the predicted background in the muon channel agree. In the electron  
 2873 channel, an excess of three events is observed with  $m_{\text{WH}} > 1.8 \text{ TeV}$ , where about 0.3 events  
 2874 are expected, while in the muon channel no events with  $m_{\text{WH}} > 1.8 \text{ TeV}$  are observed, where  
 2875 about 0.3 events are expected.

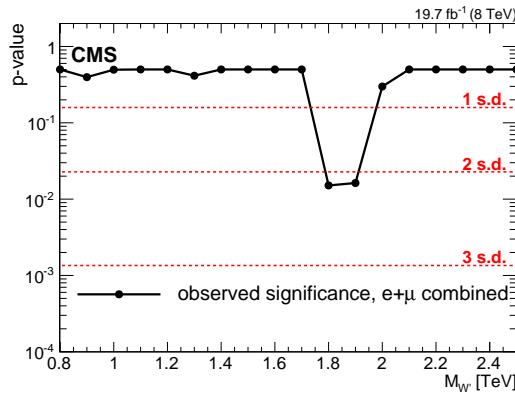
## 2876 9.2 Significance of the data

2877 A comparison between the  $m_{\text{WH}}$  distribution observed in data and the largely data-driven  
 2878 background prediction is used to test for the presence of a resonance decaying into WH.  
 2879 As described in Section 8.6, the statistical test is performed based on a profile likelihood  
 2880 discriminant for an unbinned shape analysis. Systematic uncertainties in the signal and back-  
 2881 ground yields are treated as nuisance parameters and profiled in the statistical interpretation



**Figure 9.1:** Final distributions in  $m_{\text{WH}}$  for data and expected backgrounds for electron (a) and muon (b) categories. The 68% error bars for Poisson event counts are obtained from the Neyman construction [120]. The hatched region indicates the statistical uncertainty of the fit combined with the systematical uncertainty in the shape. This figure also shows a hypothetical  $W'$  signal with mass of 1.5 TeV, normalized to the cross section predicted by the HVT model B with parameter  $g_V = 3$  as described in Section.

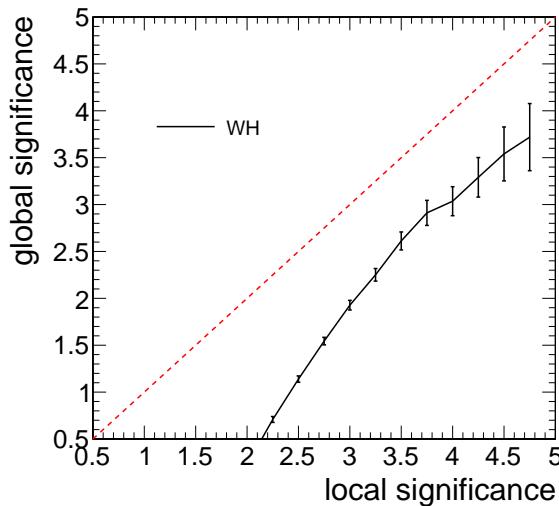
2882 using log-normal priors, while Gaussian priors are used for shape parameters only. The local  
 2883 significance of the observations is evaluated in the context of the described statistical test,  
 2884 under the assumptions of a narrow resonance decaying into the WH final state and lepton  
 2885 universality for the  $W$  boson decay, by combining the two event categories. Correlations  
 2886 arising from the uncertainties common to both channels are taken into account. The result  
 2887 is shown in Fig. 9.2. The highest local significance of 2.2 standard deviations is found for  
 2888 a resonance mass of 1.8 TeV, driven by the excess in the electron channel described in the  
 2889 previous section. The corresponding local significance for a resonance of 1.8 TeV in the  
 2890 electron channel is 2.9 standard deviations, while in the muon channel there is no significance.



**Figure 9.2:** Local p-value of the combined electron and muon data as a function of the  $W'$  boson mass, probing a narrow WH resonance.

2891 Taking into account the look-elsewhere effect (Section 8.6.3), the local significance of  
 2892 2.9 standard deviations can be translated into a global significance value by computing the  
 2893 trial factor as given by Eq. 8.20. Considering the mass range 0.8–2.5 TeV and the step of  
 2894 0.1 TeV used for the search, a trial factor of  $\approx 16.4$  is obtained. The factor, when multiplied  
 2895 by the local p-value, gives a global significance of 1.9 standard deviations when searching

for resonances over the full mass range and across two channels. In order to cross check this final value, the LEE is also estimated by means of background-only pseudo-experiments. The relation between the global and local significances obtained with this method is shown in Fig. 9.3, and it agrees with the calculation performed with the trial factor. It can be concluded that the results are thus statistically compatible with the SM expectation within 2 standard deviations.



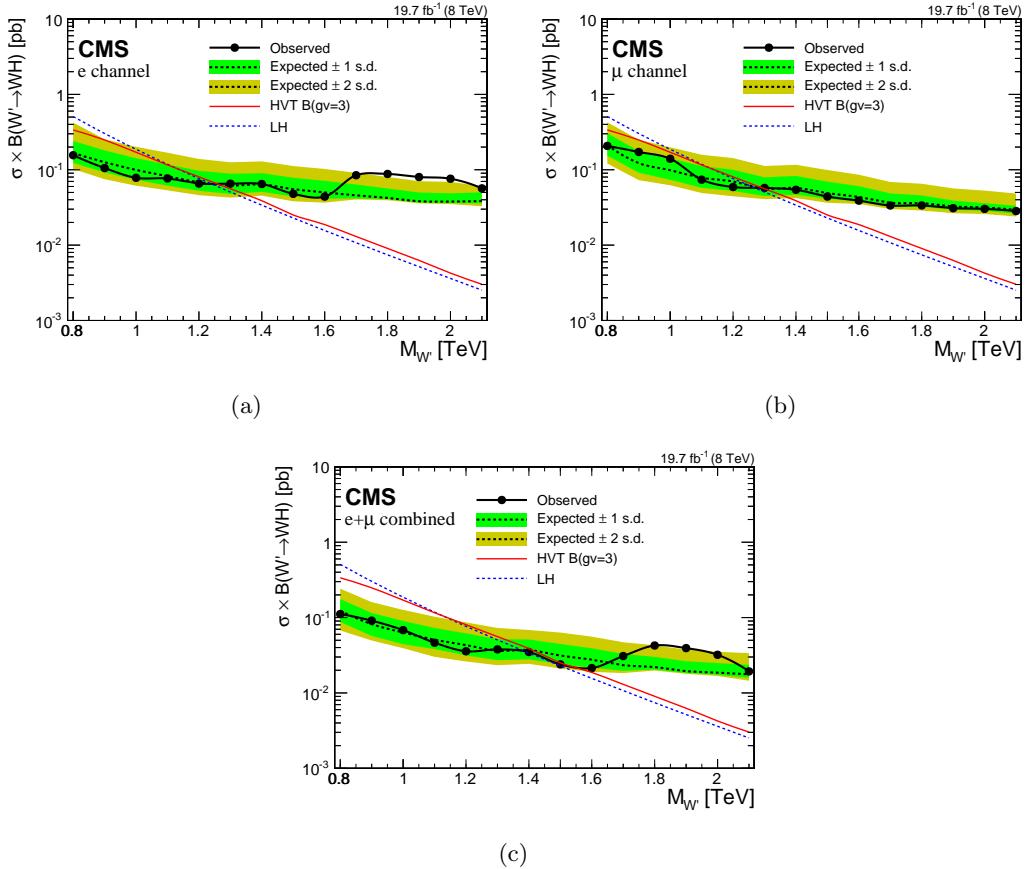
**Figure 9.3:** Global significance as a function of the local significance which corresponds to the maximal significance in the  $m_{\text{WH}}$  range 0.8–2.5 TeV in the two categories. The global significance is estimated with a frequentist approach using background-only pseudo-experiments and corresponds to the fraction of toys (translated from a p-value to significance) with at least a certain local significance in the  $m_{\text{WH}}$  range in the two categories.

### 9.3 Cross section limits

Since no excesses with significance larger than three standard deviations are observed, upper limits are set on the production cross section of the new resonance following the modified-frequentist  $\text{CL}_s$  method described in Section 8.6. Exclusion limits can be set as a function of the  $W'$  resonance mass, under the narrow-width approximation. The results are interpreted in the HVT model B and in the context of the little Higgs model.

Figure 9.4 shows the expected and observed exclusion limits at 95% CL on the product of the  $W'$  production cross section and the branching fraction of  $W' \rightarrow \text{WHWH}$  for the electron and muon channels separately, and for the combination of the two. The limits are compared with the prediction of the two theoretical models. For the combined channels, the observed and expected lower limits on the  $W'$  mass are 1.4 TeV in the LH model and 1.5 TeV in the HVT model B. For the electron (muon) channel, the observed and expected lower limits on the  $W'$  mass are 1.2 (1.3) TeV in the LH model and 1.3 (1.3) TeV in the HVT model B.

These results are finally combined with other searches for heavy resonances decaying into diboson performed with pp collisions at 8 and 13 TeV as described in Chapter 11.



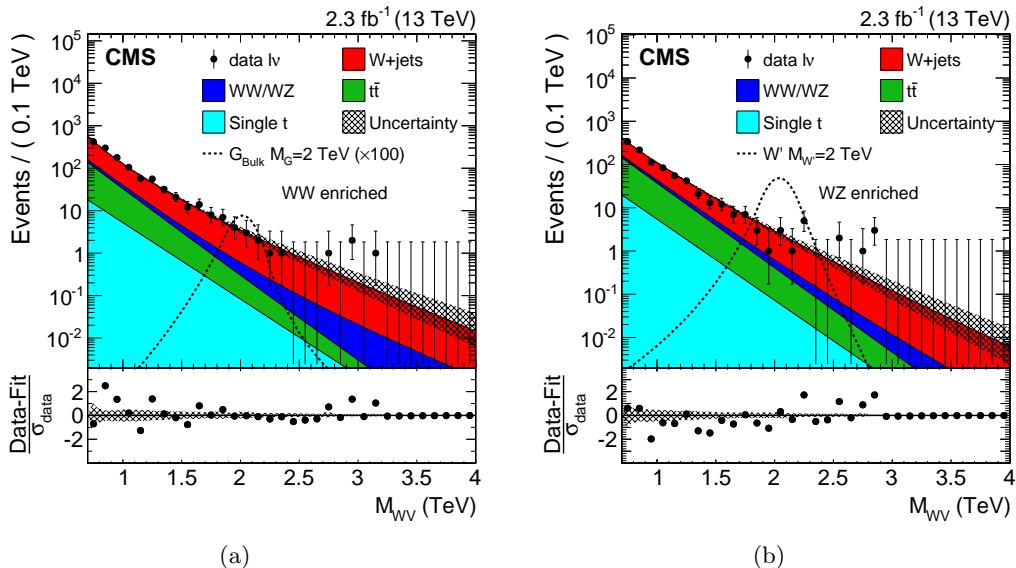
**Figure 9.4:** Observed (solid) and expected (dashed) upper limits at 95% CL on the product of the  $W'$  production cross section and the branching fraction of  $W' \rightarrow WH$  for electron (a) and muon (b) channels, and the combination of the two channels (c). The products of cross sections and branching fractions for  $W'$  production in the LH and HVT models are overlaid.

# Results with 13 TeV data

In this chapter, the final results of the analysis performed with 13 TeV data and focused on the search for a heavy resonances decaying into a pair of vector bosons (WW/WZ) in the  $\ell\nu q\bar{q}$  final state, are presented and discussed. As for the analysis conducted with 8 TeV data described in the previous chapter, the final  $m_{WV}$  spectrum observed in data is used to check for the presence of a new resonance. No bins with an excess with significance larger than three standard deviations are observed and upper limits are set on the production cross section of such resonances under a variety of signal benchmarks by combining all the event categories.

## 10.1 Final $m_{WV}$ distribution

The final  $m_{WV}$  spectra observed in data and for the background predicted with the  $\alpha$  ratio method (Section 8.2) for all event categories are shown in Fig. 10.1. The observed data and the predicted background are found to well agree. The highest mass events are at  $m_{WV} = 2.95$  and  $3.15$  TeV for the muon and electron category, respectively.



**Figure 10.1:** Final  $m_{WV}$  distributions for data and expected backgrounds obtained combining muon and electron channels in the WW-enriched (a) and WZ-enriched (b) signal regions. In each plot the solid curve represents the background estimation provided by the  $\alpha$  ratio method. The hatched band includes both statistical and systematic uncertainties. The data are shown as black points. At the bottom of each plot are the bin-by-bin fit residuals,  $(N_{\text{data}} - N_{\text{fit}})/\sigma_{\text{data}}$ , shown together with the uncertainty band of the fit normalized by the statistical uncertainty of data,  $\sigma_{\text{data}}$ . The distributions for a bulk graviton and for a  $W'$  signal are also shown with black dashed lines.

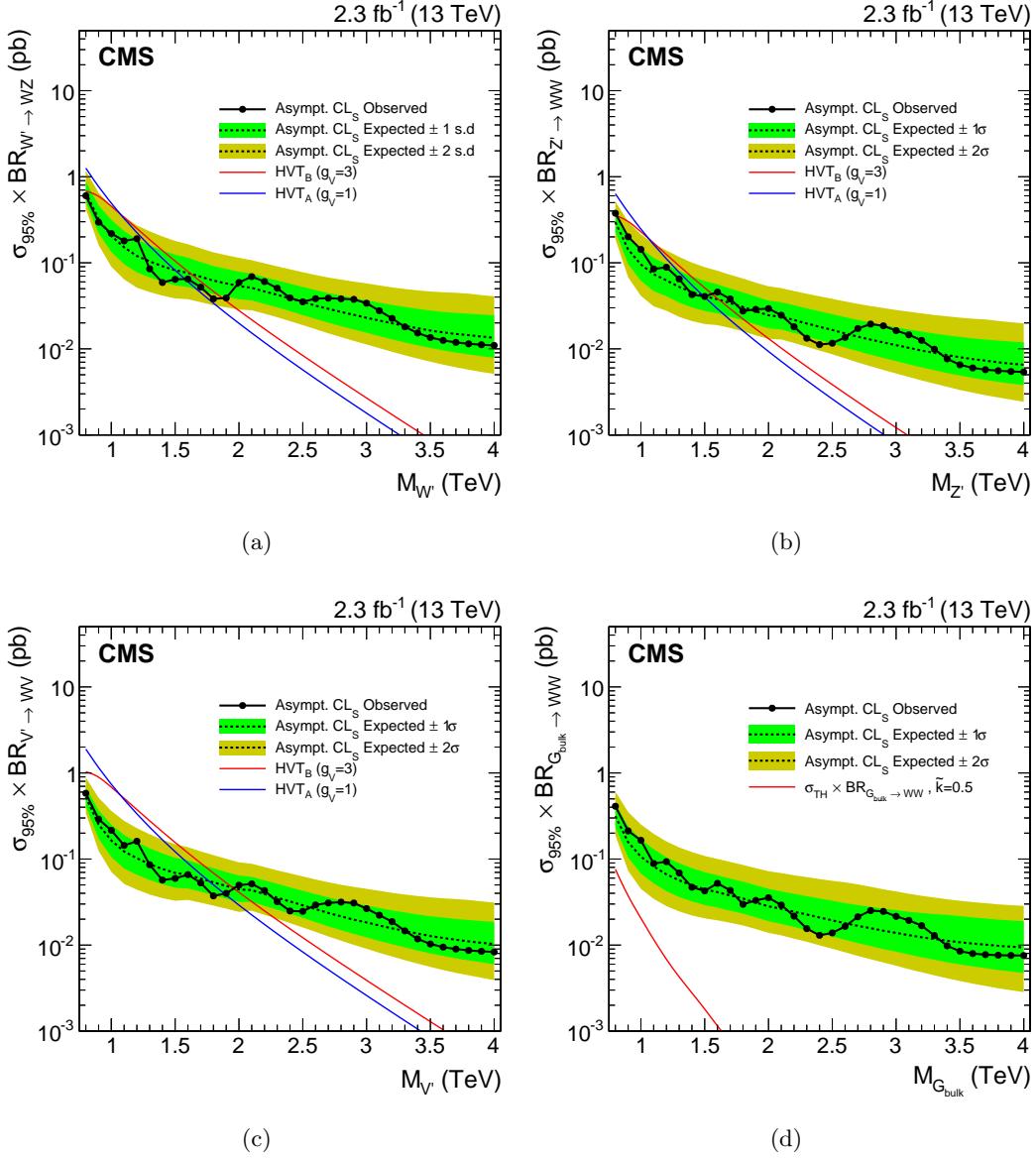
## 2934 10.2 Cross section limits

2935 Since no excesses with significance larger than three standard deviations are observed, upper  
 2936 limits are set on the production cross section of the new resonance by combining all event  
 2937 categories. The asymptotic approximation of the  $\text{CL}_s$  criterion described in Section 8.6 is  
 2938 followed. The exclusion limits computed with this approach are found to agree with the  
 2939 results obtained using the modified frequentist prescription. Systematic uncertainties are  
 2940 treated as nuisance parameters in the statistical interpretation using log-normal, and they  
 2941 are profiled following the frequentist convention as discussed in Section 8.6.

2942  
 2943 Exclusion limits are set in the context of the bulk graviton model and of the HVT Models  
 2944 A and B, under the assumption of a natural width negligible compared to the experimental  
 2945 resolution. Figure 10.2 shows the resulting 95% CL expected and observed exclusion limits  
 2946 on the signal cross section as a function of the resonance mass for all signal hypotheses. The  
 2947 limits are compared with the product of cross section and branching fraction ( $\sigma \times \mathcal{B}$ ) to WW  
 2948 for a bulk graviton with  $k/\bar{M}_{\text{Pl}} = 0.5$ , and with  $\sigma \times \mathcal{B}$  for WZ and WW for spin-1 particles  
 2949 predicted by the HVT Models A and B. In this context, a scenario is considered, where the  
 2950  $W'$  and  $Z'$  bosons are expected to be degenerate in mass (triplet hypothesis). In addition,  
 2951 the statistical interpretation is provided in a scenario where only a charged ( $W'$ ) or a neutral  
 2952 ( $Z'$ ) resonance is expected at a given mass (singlet hypothesis).

2953 In the narrow-width bulk graviton model, the sensitivity of the search is not large enough  
 2954 to set mass limits, however, cross sections are excluded in the range 0.007–0.4 pb. For HVT  
 2955 Model A (B), the data exclude singlet  $W'$  resonances with masses  $< 1.6$  (1.9) TeV and  $Z'$   
 2956 resonances with masses below  $< 1.5$  (1.6) TeV. Under the triplet hypothesis, spin-1 resonances  
 2957 with masses  $< 1.9$  and  $< 2$  TeV are excluded for HVT Models A and B, respectively.

2958 These results supersede the ones obtained analyzing 8 TeV data, where the lower mass  
 2959 limit of 1.5 TeV for a  $W'$  in the context of the HVT model B is reached (Fig. 9.4). However,  
 2960 the most stringent limits are obtained in the final combination of these results with other  
 2961 searches for heavy resonances decaying into diboson with 8 and 13 TeV data, as described in  
 2962 Chapter 11.



**Figure 10.2:** 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 charged  $W'$  (a) and neutral  $Z'$  (b) resonances, and compared with the prediction of the HVT Models A and B. (c) Limits are set in the same model under the triplet hypothesis ( $W'$  and  $Z'$ ). (d) Limits are set in the context of a bulk graviton with  $k/\bar{M}_{\text{Pl}} = 0.5$  and compared with the prediction.

# Combination of searches for diboson resonances at $\sqrt{s} = 8$ and 13 TeV

In addition to the analyses described in this work, several similar searches for narrow-width massive resonances decaying to pairs of W, Z, and Higgs bosons in various final states have been performed with the CMS experiments in both LHC Run 1 and Run 2 [93–95, 98, 121–124]. As these searches have individually very similar sensitivity to benchmark physics scenarios of interest, a statistical combination to maximize the overall sensitivity is performed and presented in this chapter. Furthermore, the combination of these analyses is fundamental to fully understand the compatibility of the excess observed in the  $\ell\nu b\bar{b}$  final state at  $m_{WH} = 1.8$  TeV as discussed in Chapter 9. The interest in this excess was further enhanced by the observation of an excess at the same diboson invariant mass values by the ATLAS experiment in the all-hadronic final state [125].

The analyses taken into account in the statistical combination are based on pp collision data collected by the CMS experiment during 2012 and 2015 at  $\sqrt{s} = 8$  TeV and 13 TeV, corresponding to an integrated luminosity of  $19.7 \text{ fb}^{-1}$  and  $2.3\text{--}2.7 \text{ fb}^{-1}$ , respectively. Analyses with all-leptonic, semi-leptonic, and all-jets final states are considered. This includes the decay into charged leptons ( $\ell$ ) and neutrinos ( $\nu$ ) of W and Z bosons, as well as reconstructed jets containing the decay products of hadronically decaying W or Z bosons. The latter are labeled as q $\bar{q}$  final states that include  $W \rightarrow q\bar{q}' \rightarrow \text{jet}$  and  $Z \rightarrow q\bar{q} \rightarrow \text{jet}$ . For Higgs bosons, hadronic decays labeled as b $\bar{b}$  or q $\bar{q}q\bar{q}$  final states referring to  $H \rightarrow b\bar{b}$  or  $H \rightarrow q\bar{q}'q\bar{q}'$  are considered.

Altogether, results are combined corresponding to the following final states:  $\ell\nu q\bar{q}$  (13 TeV, this work) [95], q $\bar{q}q\bar{q}$  (13 TeV) [95],  $\ell\ell b\bar{b}/\ell\nu b\bar{b}/\nu\nu b\bar{b}$  (13 TeV) [98],  $3\ell\nu$  (8 TeV) [122],  $\ell\nu q\bar{q}$  (8 TeV) [93],  $\ell\ell q\bar{q}$  (8 TeV) [93], q $\bar{q}q\bar{q}$  (8 TeV) [94],  $\ell\nu b\bar{b}$  (8 TeV, this work) [123], q $\bar{q}b\bar{b}/6q$  (8 TeV) [121], q $\bar{q}\tau\tau$  (8 TeV) [124]. As for the analyses described in this thesis, also the other searches feature a similar experimental signature given by highly boosted bosons in the final state. Therefore, all these analyses exploit same V tagging and H tagging algorithms to help resolve jet decays of massive bosons and achieve large suppression of SM backgrounds.

The results are interpreted in the context of the BSM models described in Section 2.3 and summarized in Table 4.1, namely, heavy vector triplet and singlet models predicting W' and Z' bosons, and the bulk graviton model. Combined cross section limits as a function of resonance mass are obtained. This work represents the first combined search for high mass resonances with both WW/WZ and WH/ZH signatures.

This chapter is organized as follows. A summary of the analyses entering the combination is given in Section 11.1. The combination procedure is described in Section 11.2, and finally the results are presented and discussed in Section 11.3.

## 3003 11.1 Inputs to the combination

3004 A statistical combination is carried out of searches for new heavy resonances that are performed  
 3005 on top of the steeply falling invariant mass distribution of two reconstructed W, Z or Higgs  
 3006 bosons. Various decay modes of these bosons are considered. The  $Z \rightarrow \ell\ell$  candidates are  
 3007 reconstructed from electron and muon candidates, while  $W \rightarrow \ell\nu$  candidates are reconstructed  
 3008 from identified muons or electrons with the method described in Section 6.6, which makes use  
 3009 of the missing transverse momentum under the constraint that the  $\ell\nu$  invariant mass is equal  
 3010 to the known W-boson mass. The  $H \rightarrow \tau\tau$  candidates are reconstructed from electron, muon  
 3011 and hadronically-decaying  $\tau$  candidates in combination with missing transverse momentum.  
 3012 The  $W \rightarrow q\bar{q}'$ ,  $Z \rightarrow q\bar{q}$ ,  $H \rightarrow b\bar{b}$  and  $H \rightarrow q\bar{q}'q\bar{q}'$  candidates are reconstructed with jet  
 3013 algorithms with a distance parameter of 0.8 (CA for the 8 TeV data analyses, AK for the  
 3014 13 TeV analyses).

3015 All analyses are focused on high mass resonances which decay in highly boosted W/Z/H  
 3016 bosons. Hence, their decay products are reconstructed close-by in angle, requiring the special  
 3017 reconstruction techniques already described previously in this thesis. For highly boosted  
 3018 W/Z/H bosons decaying to electron, muon and  $\tau$  candidates, identification and isolation  
 3019 requirements are adapted such that the nearby reconstructed leptons do not reduce the  
 3020 identification efficiency.

3021 For highly boosted V bosons decaying to quark anti-quark pairs, the V algorithm described  
 3022 in Section 7.2 is applied. In the 8 TeV data analyses, a V jet candidate is identified if its  
 3023 pruned mass,  $m_{jet}$ , falls in a range around the W or Z mass. In the 13 TeV data analyses,  
 3024 two distinct categories enriched in W or Z bosons are defined by two exclusive ranges in  $m_{jet}$   
 3025 as described in Section 8.1. In the 8 TeV data analyses the sensitivity is further enhanced  
 3026 by distinguishing two categories, a low purity (LP) and a high purity (HP) one based on  
 3027 the  $\tau_{21}$  variable. This same strategy is follow in the dijet 13 TeV analysis. Although the HP  
 3028 category dominates the total sensitivity of the analyses, the LP category is retained, since for  
 3029 large masses of a new resonance it provides improved signal efficiency with only moderate  
 3030 background contamination.

3031 Higgs-boson identification is similarly performed using a pruned jet mass window around  
 3032 the Higgs mass together with b-tagging algorithms applied to the H jet or to its subjets as  
 3033 described in Section 7.3. To distinguish  $H \rightarrow WW \rightarrow q\bar{q}'q\bar{q}'$  jets from background, a similar  
 3034 technique as V tagging is applied using the  $\tau_{42}$  ratio. The selection efficiencies for each signal  
 3035 and channel are summarized in Table 11.1.

3036 In all-jets final states, the background dominated by QCD multijets production is estimated  
 3037 with a fit of signal+background to the data, where the background is described by a smooth  
 3038 functional form. In semi-leptonic final states, the dominant backgrounds from V+jets  
 3039 production are estimated using data in  $m_{jet}$  sidebands with the method described in Section 8.2.  
 3040 In all-leptonic final states, the dominant background from standard model diboson production  
 3041 is estimated using simulated events.

3042 More details are given in the following for the analyses where not all signal models  
 3043 presented in the combination were originally considered.

### 3044 11.1.1 Reinterpretations

3045 In the searches for new heavy resonances decaying into a pair of vector bosons (WW, ZZ or  
 3046 WZ) in the semi-leptonic ( $\ell\nu q\bar{q}$  and  $\ell\ell q\bar{q}$ ) final states [93] with pp collision data collected at  
 3047  $\sqrt{s} = 8$  TeV, exclusion limits at 95% CL have been set on the production cross section of  
 3048 a bulk graviton. The results were published with a parametrization for the reconstruction  
 3049 efficiency as a function of W and Z boson kinematics, enabling a reinterpretation in the

**Table 11.1:** Summary of the signal efficiencies of all analysis channels for all signal models for a 2 TeV resonance. For analyses with categorization in high-purity (HP) and low-purity (LP) categories, both efficiencies are quoted in the form HP/LP. The signal efficiencies are in percent and include the SM branching ratios of the bosons to the final state of the analysis channel, effects from detector acceptance, as well as reconstruction and selection efficiencies.

Channel	HVT				RS bulk	
	$W'$		$Z'$		$G_{\text{bulk}}$	
	WZ	WH	WW	ZH	WW	ZZ
$3\ell\nu$ (8 TeV)	0.6	-	-	-	-	-
$\ell\ell q\bar{q}$ (8 TeV)	1.1/-	-	-	0.2/-	-	3.0/1.0
$\ell\nu q\bar{q}$ (8 TeV)	4.8/-	-	9.4/-	-	10.6/7.1	-
$q\bar{q}q\bar{q}$ (8 TeV)	5.9/5.5	0.8/0.7	5.7/5.3	0.8/0.7	3.8/3.1	5.7/4.2
$\ell\nu bb$ (8 TeV)	-	0.9	-	-	-	-
$q\bar{q}\tau\tau$ (8 TeV)	-	1.2	-	1.3	-	-
$q\bar{q}bb/6q$ (8 TeV)	-	3.0/1.8	-	1.7/1.1	-	-
$\ell\nu q\bar{q}$ (13 TeV)	10.2/-	1.7/-	19.4/-	-	18.1/-	-
$q\bar{q}q\bar{q}$ (13 TeV)	9.7/12.3	1.8/2.5	8.2/10.6	1.9/2.6	8.7/12.4	11.0/13.5
$\ell\ell bb$ (13 TeV)	-	-	-	1.5	-	-
$\ell\nu bb$ (13 TeV)	-	4.0	-	-	-	-
$\nu\nu bb$ (13 TeV)	-	-	-	4.2	-	-

context of neutral and charged the spin-1 resonances as predicted by HVT models. The reinterpretation in the context of this model is obtained by rescaling the bulk graviton signal efficiencies by scale factors taking into account the different kinematics of W and the Z bosons from  $W'$  and  $Z'$  production compared to the graviton production. The scale factors have been derived for each mass point by means of the tables published in Ref. [93]. Since the efficiency parametrization is restricted to the HP category of the analyses, the LP category is not used for the  $W'$  and  $Z'$  interpretations of these channels. The  $m_{\text{jet}}$  window that defines the signal regions of the analysis channels is such that the  $\ell\nu q\bar{q}$  channel is sensitive to both the charged and neutral resonance predicted by HVT models. This is taken into account in the statistical combination.

The searches for new heavy resonances decaying into a pair of vector bosons (WW, ZZ or WZ) in the semi-leptonic ( $\ell\nu q\bar{q}$  and  $\ell\ell q\bar{q}$ ) [93, 93, 95], and all-hadronic ( $q\bar{q}q\bar{q}$ ) final states [94, 95] at 8 and 13 TeV, are also sensitive to WH and ZH signatures, since a small fraction of jets initiated by Higgs bosons have a pruned jet mass in the range considered to identify W or Z bosons. These searches were therefore re-interpreted with WH and ZH signals to profit from this additional signal sensitivity. The additional signal efficiencies for those signals are indicated in Table 11.1.

The search for resonances in the  $q\bar{q}\tau\tau$  final state [121] was optimized for a resonance  $Z'$  decaying into Z and a Higgs boson. However, given the large  $m_{\text{jet}}$  window ( $65 < m_{\text{jet}} < 105$  GeV) used to identify the hadronically decaying Z boson, this analysis channel is also sensitive to the production of the charged spin-1  $W'$  resonance decaying into W and Higgs bosons as predicted in HVT models. This overlap is also taken into account in the statistical combination.

## 3073 11.2 Combination procedure

3074 In all the analysis channels a search is performed for a peak on top of the falling background  
 3075 distribution in the diboson invariant mass by means of a maximum likelihood fit to the data.  
 3076 As done for the main analyses described in this work (Section 8.6), the likelihood function  
 3077 is maximized to obtain the best fit of the signal strength modified  $\mu$  for each signal and  
 3078 resonance mass hypothesis. The function is constructed from the reconstructed diboson  
 3079 invariant mass distribution observed in data, the background prediction, and the signal  
 3080 resonance shape to test for the presence of a new resonance decaying to two bosons. For  
 3081 the  $3\ell\nu$ ,  $q\bar{q}q\bar{q}$ ,  $q\bar{q}b\bar{b}/6q$ , and  $q\bar{q}\tau\tau$  analyses, the likelihood function is computed using events  
 3082 binned as a function of reconstructed diboson invariant mass as in Equation 8.6. For the  
 3083 remaining analyses ( $\ell\nu q\bar{q}$ ,  $\ell\ell q\bar{q}$ ,  $\ell\nu b\bar{b}$ ), the functional form for an unbinned likelihood is  
 3084 similarly defined using functional forms that describe the shape of the reconstructed diboson  
 3085 invariant mass for background and signal resonance as given by Equation 8.7.

3086 The treatment of the background in the maximum likelihood fit depends on the analysis  
 3087 channel. In the  $q\bar{q}q\bar{q}$  and  $q\bar{q}b\bar{b}/6q$  analyses, the background fit function parameters are left  
 3088 floating in the maximum likelihood fit, such that the background prediction is simultaneously  
 3089 obtained with the signal  $\mu$  for every hypothesis. The remaining analyses ( $\ell\nu q\bar{q}$ ,  $\ell\ell q\bar{q}$ ,  $\ell\ell b\bar{b}$ ,  
 3090  $\ell\nu b\bar{b}$ ,  $\nu\nu b\bar{b}$ ) follow the same procedure as for the analyses described in this work: the  
 3091 background is estimated using data sidebands and uncertainties related to its parametrized  
 3092 shape are treated as nuisance parameters constrained with Gaussian probability density  
 3093 functions in the maximum likelihood fit. Except for the cases described in Section 11.1, which  
 3094 have been found to be negligible, selection are exclusive. The combined likelihood is then  
 3095 obtained from the product of the likelihoods of each individual analysis channel.

3096 The asymptotic approximation of the  $CL_s$  criterion (Section 8.6.2) is used with the test  
 3097 statistic given by Eq. 8.8 to set upper limits on the cross section for resonance production.  
 3098 When combining 8 and 13 TeV analyses, limits are set on the signal scale factor  $\mu$  taking into  
 3099 account the production cross section ratio evaluated from theory between 8 TeV and 13 TeV.

3100 The dominant sources of systematic uncertainties are treated as nuisance parameters  
 3101 constrained with a log-normal probability density function. All nuisance parameters are  
 3102 profiled following the frequentist convention discussed in Section 8.6. When the likelihoods  
 3103 of multiple analyses channels are combined, the correlation of systematic effects across  
 3104 analysis channels is taken into account by categorizing the uncertainties into fully correlated  
 3105 (associate to same nuisance parameter) and fully uncorrelated (associate to different nuisance  
 3106 parameters). Table 11.2 summarizes which uncertainties are treated as correlated among 8  
 3107 and 13 TeV analyses, electron and muon channels, HP and LP categories and W, Z and Higgs  
 3108 enriched categories in the combination. Further categorisation within individual analyses are  
 3109 described therein.

3110 The most important and only nuisance parameters treated as correlated between 8 and  
 3111 13 TeV analyses are those related to the PDFs and the choice of factorization ( $\mu_f$ ) and  
 3112 renormalization ( $\mu_r$ ) scales used to estimate the signal cross sections. They have been re-  
 3113 evaluated for this combination for both 8 and 13 TeV analyses, estimating the full impact on  
 3114 the expected signal yield rather than the impact on only the signal acceptance. The PDF  
 3115 uncertainties are evaluated using the NNPDF 3.0 [22] PDFs. The uncertainty related to the  
 3116 choice of  $\mu_f$  and  $\mu_r$  scales is evaluated following the proposal in Refs. [108, 109] by varying  
 3117 the default choice of scales in the following 6 combinations of factors:  $(\mu_f, \mu_r) \times (1/2, 1/2)$ ,  
 3118  $(1/2, 1)$ ,  $(1, 1/2)$ ,  $(2, 2)$ ,  $(2, 1)$ , and  $(1, 2)$ . The experimental uncertainties are all treated as  
 3119 uncorrelated between 8 and 13 TeV. At 13 TeV the systematic uncertainties are dominated by  
 3120 the statistical uncertainty of the datasets used to evaluate scale factors applied to the signal

**Table 11.2:** Correlation of systematic uncertainties.

Systematic uncertainty	Type	8+13 TeV	e+μ	HP+LP	W+Z+Higgs
Lepton trigger	yield	no	no	yes	yes
Lepton identification	yield	no	no	yes	yes
Lepton momentum scale	yield, shape	no	no	yes	yes
Jet energy scale	yield, shape	no	yes	yes	yes
Jet energy resolution	yield, shape	no	yes	yes	yes
Jet mass scale	yield	no	yes	yes	yes
Jet mass resolution	yield	no	yes	yes	yes
b tagging	yield	no	yes	yes	yes
W tagging $\tau_{21}$ (HP/LP)	yield	no	yes	yes	yes
Integrated luminosity	yield	no	yes	yes	yes
Pileup	yield	no	yes	yes	yes
PDF	yield	yes	yes	yes	yes
$\mu_f, \mu_r$ scales	yield	yes	yes	yes	yes

3121 simulation to reproduce data.

## 3122 11.3 Results

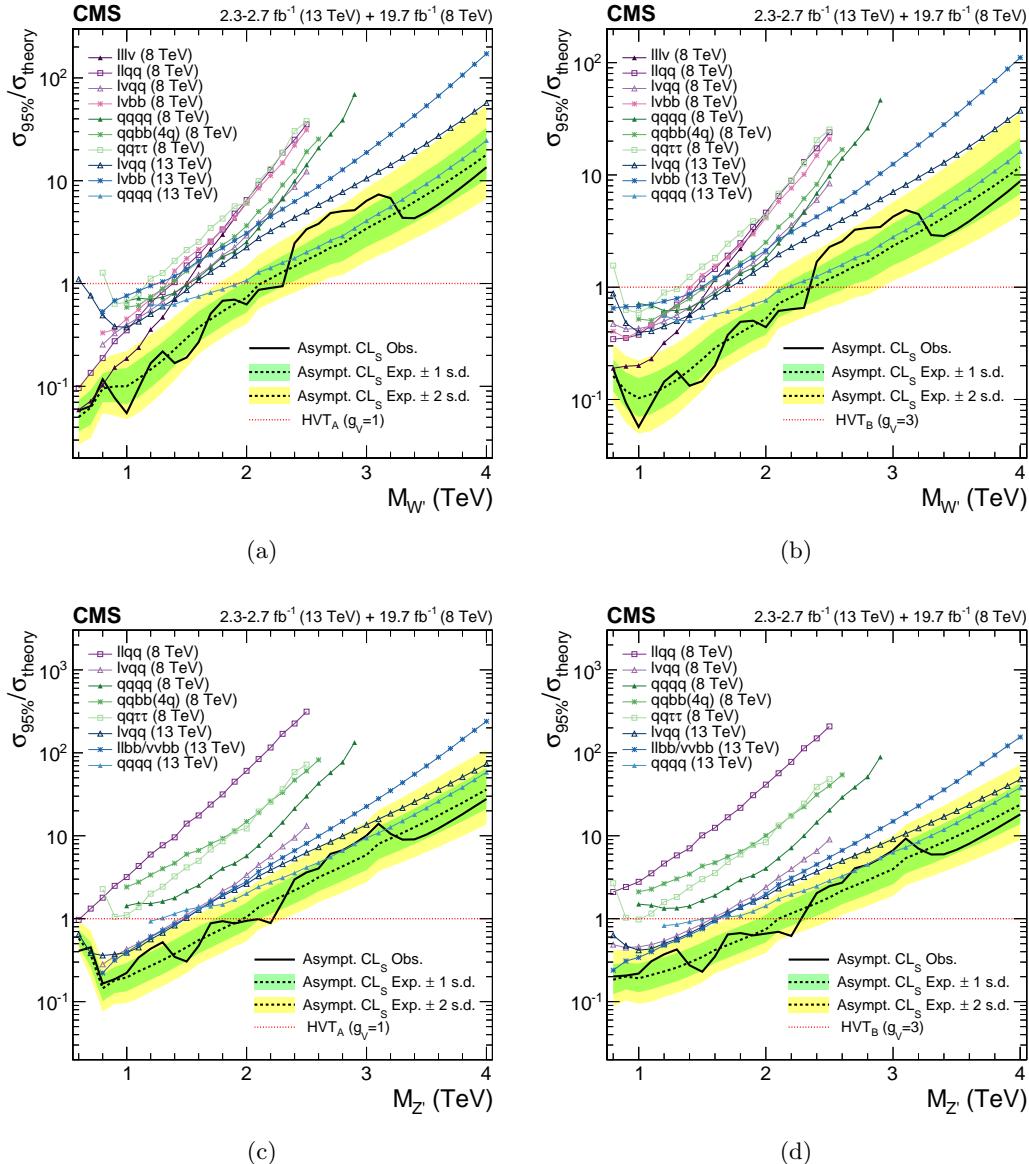
3123 In this section the combination of the individual analysis channels described in Section 11.1  
3124 is presented, for each of the signal hypothesis described in Section 2.3. For each channel the  
3125 95% CL exclusion limits on the signal strength modifier  $\mu = \sigma_{95\%}/\sigma_{\text{theory}}$  are presented.

### 3126 11.3.1 Limits on W' and Z' singlets

3127 Figures 11.1(a) and 11.1(b) show the comparison and combination of the results obtained in  
3128 the 8 and 13 TeV searches for a W' singlet resonance for model A and model B, respectively.  
3129 The 95% CL exclusion limits on the signal strength in the resonance mass range  $0.6 < m_{W'} <$   
3130 4 TeV for model A and  $0.8 < m_{W'} < 4$  TeV for model B are shown. Table 11.3 summarizes  
3131 the resulting resonance mass exclusion limits. Below resonance mass values of about 1.4  
3132 TeV, the  $3\ell\nu$  channel at 8 TeV is most sensitive. At higher masses, the  $q\bar{q}q\bar{q}$  search at 13  
3133 TeV dominates the sensitivity. The overall sensitivity benefits from the combination up to  
3134 resonance masses of about 2 TeV, lowering the cross section exclusion limit by up to a factor  
3135 1/3 when comparing to the most sensitive single channel. Above masses of 2 TeV the 8 TeV  
3136 channels do not add any significant contribution compared to the  $q\bar{q}q\bar{q}$  search at 13 TeV. The  
3137 observed mass limit is not affected by the combination compared to that obtained from the  
3138 13 TeV searches. However, the expected mass limit is slightly improved from 2.3 to 2.4 TeV.

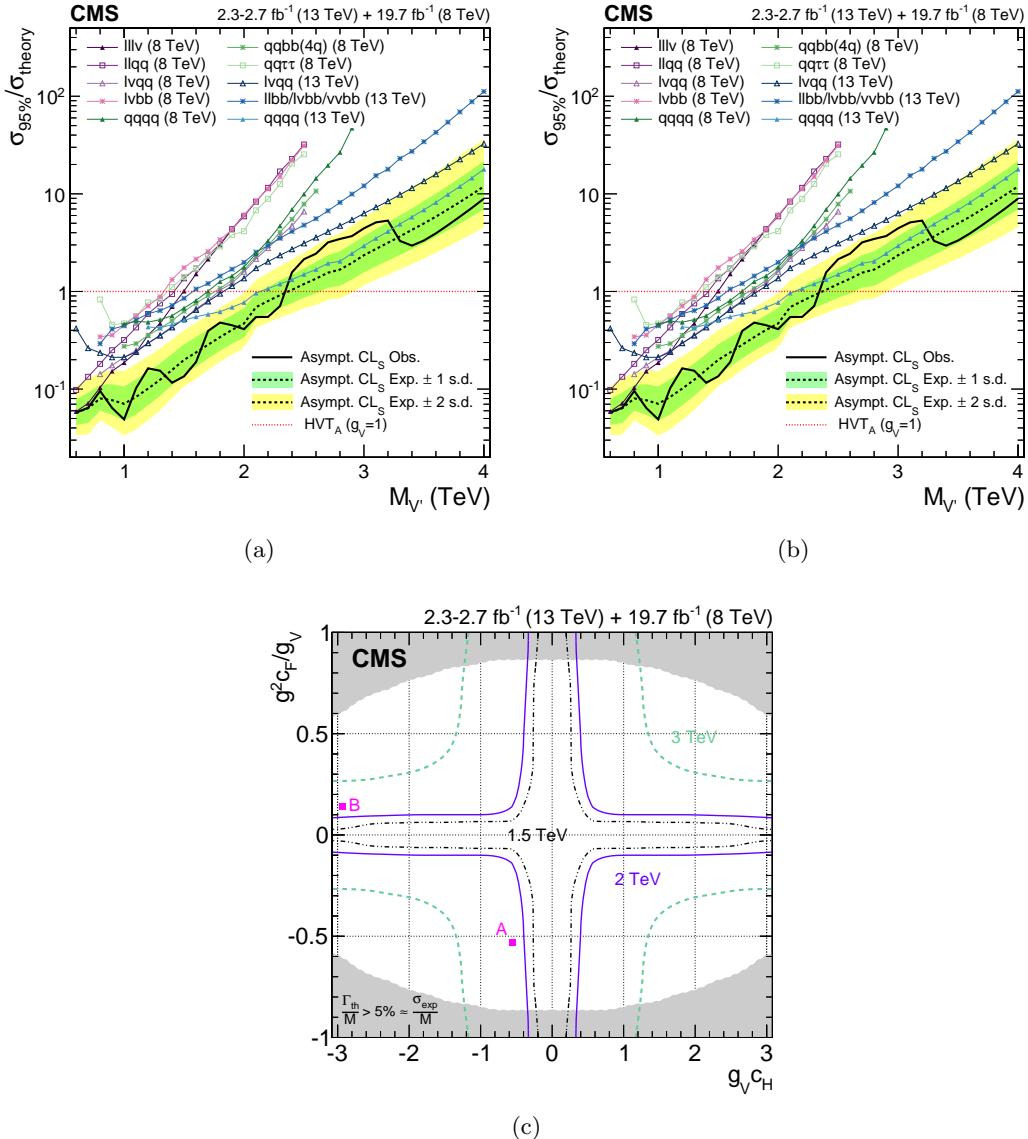
**Table 11.3:** Resonance mass 95% CL exclusion limits in HVT model scenarios.

Model	Observed limit (TeV)	Expected limit (TeV)
W' (model A)	2.3	2.1
Z' (model A)	2.2	2.0
HVT (W' +Z') (model A)	2.4	2.4
W' (model B)	2.3	2.4
Z' (model B)	2.3	2.1
HVT (W' +Z') (model B)	2.4	2.6



**Figure 11.1:** Exclusion limits at 95% CL on the signal strength for (top)  $W' \rightarrow WZ/WH$  and (bottom)  $Z' \rightarrow WW/ZH$  in (left) HVT model A and (right) HVT model B as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches. In each of the plots the different colored lines correspond to the searches entering the combination.

Figures 11.1(c) and 11.1(d) show the comparison and combination of the results obtained in the 8 and 13 TeV searches for a  $Z'$  singlet resonance for model A and model B, respectively. The  $\ell\nu q\bar{q}$  channel at 8 TeV and the  $q\bar{q}q\bar{q}$ ,  $\ell\nu q\bar{q}$ ,  $\ell\ell b\bar{b}/\nu\nu b\bar{b}$  channels at 13 TeV dominate the sensitivity over the whole range, with 8 and 13 TeV analyses giving almost equal contributions for masses below 2 TeV. Above this value, the sensitivity is mainly driven by the 13 TeV analyses. Under this signal hypothesis the sensitivities reached by the 8 and 13 TeV channels are similar at low resonance masses. As for the  $W'$  case, the mass limit is not affected by the combination compared to what is obtained from the 13 TeV searches.

3147 11.3.2 Limits on heavy vector triplet ( $W' + Z'$ )

**Figure 11.2:** Exclusion limits at 95% CL on the signal strength in (a) HVT model A and (b) HVT model B as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches. In both plots the different colored lines correspond to the searches entering the combination. (c) Exclusion regions in the plane of the HVT-model couplings ( $g_V c_H$ ,  $g^2 c_F/g_V$ ) for three resonance masses, 1.5, 2, and 3 TeV, where  $g$  denotes the weak gauge coupling. The points A and B of the benchmark models used in the analysis are also shown. The boundaries of the regions outside these lines that are excluded by this search are indicated by the solid and dashed lines. The areas indicated by the solid shading correspond to regions where the resonance width is predicted to be more than 7% of the resonance mass and the narrow-resonance assumption is not satisfied.

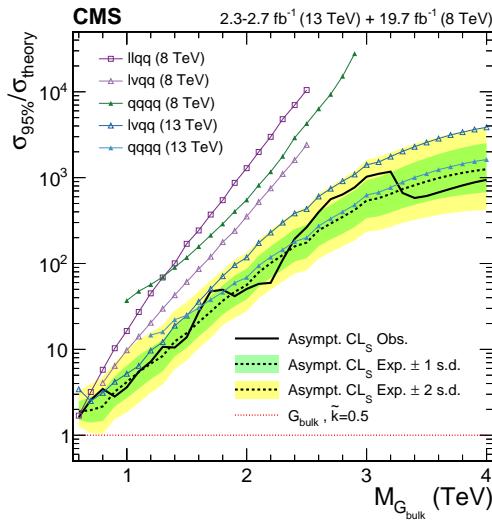
3148 Figures 11.2(a) and 11.2(b) shows the comparison and combination of the results obtained  
 3149 in the 8 and 13 TeV searches for a heavy vector triplet scenario. As for the  $W'$  and  $Z'$  cases,  
 3150 the observed mass limit of 2.4 TeV obtained combining 8 and 13 TeV searches is determined  
 3151 by the 13 TeV channels.

3152 In Fig. 11.2(c), a scan of the coupling parameters and the corresponding observed 95% CL

exclusion contours in the HVT model from the combination of the 8 and 13 TeV analyses are shown. The parameters are defined as  $g_{VH}$  and  $g^2 c_F/g_V$ , in terms of the coupling strengths (Section 2.3.3) of the new resonance to the Higgs boson and to fermions. The range of the scan is limited by the assumption that the new resonance is narrow. A contour is overlaid, representing the region where the theoretical width is larger than the experimental resolution of the searches, and hence where the narrow-resonance assumption is not satisfied. This contour is defined by a predicted resonance width of 5%, corresponding to the narrowest resonance mass resolution of the considered searches.

### 11.3.3 Limits on bulk graviton

Figure 11.3 shows the comparison and combination of the results obtained in the 8 and 13 TeV VV searches in the bulk graviton scenario with  $k/\bar{M}_{Pl} = 0.5$ . The sensitivity is mainly driven by the 13 TeV  $q\bar{q}q\bar{q}$  and  $\ell\nu q\bar{q}$  channels. Under this signal hypothesis, the sensitivity reached by the 13 TeV searches supersedes the 8 TeV combination down to very low resonance masses (0.7 TeV), since this signal is produced via gluon-fusion in contrast to the HVT resonances produced via  $q\bar{q}$  annihilation. Hence, the contribution given by 8 TeV channels is less significant with respect to the spin-1 resonance hypotheses. The combination yields the most stringent signal strength limits on narrow bulk graviton resonances ( $k/\bar{M}_{Pl} = 0.5$ ) to date in the mass range from 0.6 to 4 TeV.



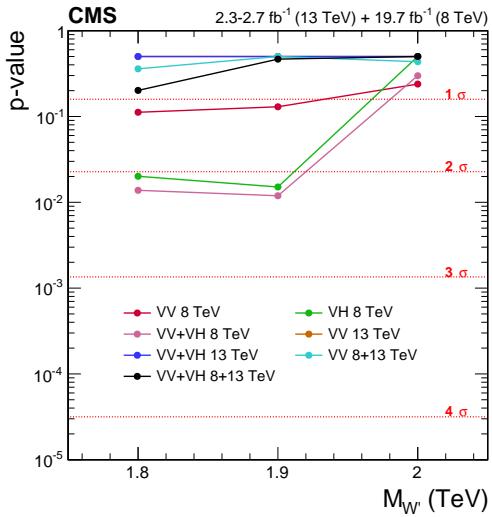
**Figure 11.3:** Exclusion limits at 95% CL on the signal strength as a function of the resonance mass obtained by combining the 8 and 13 TeV diboson searches in the bulk graviton scenario with  $k/\bar{M}_{Pl} = 0.5$ . The different colored lines correspond to the searches entering the combination.

### 11.3.4 Significance at 2 TeV

ATLAS reported an excess in the all-hadronic  $VV \rightarrow q\bar{q}q\bar{q}$  search corresponding to a local significance of  $3.4\sigma$  for a  $W'$  resonance with a mass of 2 TeV [125]. For CMS, the largest deviation of  $2.2\sigma$  has been observed in the semi-leptonic  $WH \rightarrow \ell\nu b\bar{b}$  search described in this work (Chapter 9). The combined significance of the 8 and 13 TeV CMS searches in the range 1.8–2.0 TeV is here evaluated and showed in Figure 11.4 for a  $W'$  hypothesis.

Combining all 8 TeV VH searches in the  $W'$  hypothesis, the local significance of the excess at 1.8 TeV is slightly reduced to  $2.1\sigma$ . Combining all 8 TeV VV and VH searches,

it is increased back to  $2.2\sigma$ , since the VV searches observed a small deviation in the same resonance mass range. However, in combination with the 13 TeV VV and VH searches, the overall significance at 1.8 TeV is reduced to  $0.8\sigma$ . This remains the largest significance for the overall combination of 8+13 TeV searches considering all signal hypothesis over the mass range 1.8–2.0 TeV, thus not supporting the excesses observed in the two individual channels in 8 TeV data.



**Figure 11.4:** Local p-values of the excesses observed in the resonance mass range 1.8–2 TeV in the various combinations of searches for a  $W'$  hypothesis.

# Conclusions

3188 A search for new massive resonances decaying into a pair of vector bosons (WW/WZ) or  
 3189 into a W boson and a Higgs boson (WH) in semi-leptonic final states has been presented. In  
 3190 particular two analyses have been described featuring a similar search strategy.

3191 The first analysis is performed with pp collision data at  $\sqrt{s} = 8$  TeV collected in 2012, and  
 3192 is focused on the final state given by the W boson decay to  $\ell\nu$ , with  $\ell = \mu$  or e, and the Higgs  
 3193 boson decay to a pair of bottom quarks. The second analysis is performed with pp collision  
 3194 data at  $\sqrt{s} = 13$  TeV collected in 2015, and also in this case a final state is considered given  
 3195 by the  $W \rightarrow \ell\nu$  decay together with the decay of the second boson into quarks, where the  
 3196 second boson (V) can be either a W or a Z.

3197 In both analyses, each event is reconstructed as a leptonic W boson candidate recoiling  
 3198 against a jet with mass compatible with the Higgs or V boson mass for the  $\ell\nu b\bar{b}$  or  $\ell\nu q\bar{q}$   
 3199 analysis channel, respectively. Specialized methods, referred to as V tagging and H tagging,  
 3200 are exploited to help resolve jet decays of massive bosons and achieve large suppression of  
 3201 background from multijet processes. In particular, the H tagging algorithm combines jet  
 3202 substructure information with identification techniques based on the peculiarities of b jets.

3203 In the  $\ell\nu b\bar{b}$  analysis channel, no excess of events above the standard model prediction  
 3204 is observed in the muon channel, while an excess with a local significance of 2.9 standard  
 3205 deviations is observed in the electron channel at  $m_{WH} \approx 1.8$  TeV. Taking into account the  
 3206 look-elsewhere effect, the results are statistically compatible with the standard model within  
 3207 2 standard deviations. In the context of the little Higgs and the heavy vector triplet models,  
 3208 upper limits at 95% CL are set on the  $W'$  production cross section in a range from 100 to  
 3209 10 fb for masses between 0.8 and 2.5 TeV, respectively. Within the little Higgs model, a lower  
 3210 limit on the  $W'$  mass of 1.4 TeV has been set. A heavy vector triplet model that mimics the  
 3211 properties of composite Higgs models has been excluded up to a  $W'$  mass of 1.5 TeV.

3212 These results are superseded by the limits set by the analysis in the  $\ell\nu q\bar{q}$  final state. No  
 3213 evidence for a signal is found in this search with new 2015 data, and the result is interpreted  
 3214 as an upper limit on the production cross section of a narrow-width resonance as a function  
 3215 its mass, in the context of several benchmark models for spin-1 and spin-2 resonances. In  
 3216 particular, for the same heavy vector triplet model as mentioned above the data exclude a  
 3217  $W'$  resonance with masses  $< 1.9$  TeV.

3218 However, the best results are provided by a statistical combination of all searches performed  
 3219 in CMS with 8 and 13 TeV data for massive resonances decaying to pairs of W, Z, and Higgs  
 3220 bosons in various final states. The results are interpreted in the context of heavy vector  
 3221 singlet and triplet models predicting a  $W'$  and a  $Z'$  decaying to WZ, WH, WW, and ZH and  
 3222 a model with a bulk graviton that decays into WW and ZZ. The combined significance of a  
 3223 potential resonances at 1.8–2.0 TeV has been evaluated and has been found to be 0.8 standard  
 3224 deviations for the hypothesis of a  $W'$ , thus the excesses observed in the  $\ell\nu b\bar{b}$  channel in  
 3225 8 TeV data is not supported. The combination yields mass limits at the 95% CL on spin-1  
 3226 resonance in the range 2.2–2.4 TeV, depending on the specific benchmark. The most stringent  
 3227 cross section limits on a narrow-width bulk graviton resonance with  $\tilde{k} = 0.5$  to date are set  
 3228 in the mass range from 0.6 to 4 TeV.

3229

## Part II

3230

# Calibration and upgrade of the CMS pixel barrel detector

3231

# 3234 Introduction

---

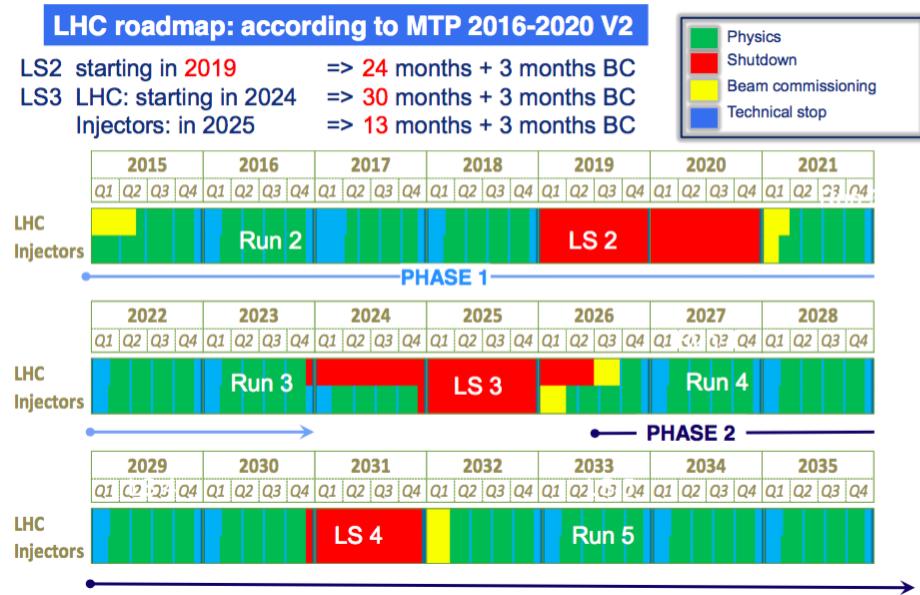
3235 The extremely high particle fluxes at small distances from the interaction point require the  
 3236 innermost tracking layers to be composed of pixel devices delivering spatial information  
 3237 with high resolution. Over the full acceptance of the CMS detector, the silicon pixel system  
 3238 provides two or more hits per track, which allow secondary vertices to be reconstructed  
 3239 for tagging long-lived objects, like b or c quarks and  $\tau$ -leptons, and to distinguish them  
 3240 from a large background of light quark and gluon jets. It is also an important detector for  
 3241 identifying the primary vertex, and separating it from dozens of additional pileup vertices.  
 3242 Hence, this detector plays a special role in the physics analyses described in this thesis. In  
 3243 fact, its performances have a large impact on the identification of b-quark jets as well as  
 3244 on jet substructure observables, being the latter highly dependent on pileup. The pixel  
 3245 detector consists of central barrel layers (BPix) and forward disks (FPix). This part of the  
 3246 thesis is dedicated to several aspects of the BPix system, including its calibration and upgrade.

3247     The pixel detector was installed in 2008 and showed an excellent performance during  
 3248 the first period of data taking at the LHC (2010–2012). After the first long shut-down of  
 3249 the machine (2012–2015), that allowed to increase the center-of-mass energy of the collisions  
 3250 to 13 TeV, the detector was re-installed into CMS and successfully continued taking data  
 3251 throughout the first two years of LHC Run 2 (2015–2016). The excellent performance of the  
 3252 BPix at the re-start of collisions have been guaranteed by the efforts spent during LS1 in  
 3253 recovering the broken channels as well as in re-calibrating the detector after the radiation  
 3254 damage suffered during Run 1.

3255     The current planning for the LHC and injector chain foresees other two long shut-downs,  
 3256 LS2 and LS3 (Fig. 13.1). In the period through LS2 (2019–2020), the injector chain will be  
 3257 improved and upgraded to deliver very bright bunches into the LHC. In LS3 (2024–2026),  
 3258 the LHC itself will be upgraded with new components to optimize the bunch overlap at the  
 3259 interaction region. Further upgrades will then be foreseen beyond the 2030.

3260     The present pixel detector was originally designed for a luminosity of  $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$   
 3261 and a pileup of 25 in LHC collisions with 25 ns bunch spacing. These parameters have already  
 3262 been exceeded in 2016, when collisions at 13 TeV happened at instantaneous luminosities  
 3263 up to  $1.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  with an average pileup of 25 [8]. Based on the excellent LHC  
 3264 performances to date, it can be anticipated that the peak luminosity will keep increasing  
 3265 until 2018 reaching values up to  $1.7\text{--}1.8 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , and beyond these after LS2. Thus,  
 3266 starting from 2017 the CMS experiment must be prepared to operate with an average pileup  
 3267 of 50 as a baseline, with the possibility that it may be significantly higher (up to 100) if  
 3268 collisions will happen at 50 ns bunch spacing after LS2. In order to maintain efficient and  
 3269 robust tracking at CMS, the pixel detector will be replaced with an upgraded pixel system,  
 3270 referred to as *Phase 1 pixel upgrade*, in the LHC winter shutdown 2016/2017. The design  
 3271 of the upgraded detector allows to cope with the aforementioned harsh conditions expected  
 3272 at the LHC in the upcoming years. A more complex upgrade step is planned for the LS3  
 3273 and referred to as *Phase 2 upgrade*. It will include deeper changes in the whole CMS, among  
 3274 which a complete substitution of the entire tracker detectors system.

3275     The Phase 1 pixel detector is expected to be operative up to the Phase 2 upgrade, around



**Figure 13.1:** The outline LHC schedule out to 2035 as officially approved in June 2015 [126].

3277 the 2024. During the planned 5 years of operation before LS3, the LHC is expected to deliver  
 3278 about  $500 \text{ fb}^{-1}$ . The proposed Phase 1 upgrade system has been designed and tested to be  
 3279 operative up to this target, with the only exception of the innermost layer. In fact, the  
 3280 estimated hadron fluence that will be accumulated in the innermost pixel layer is too high for  
 3281 the pixel sensor, thus a replacement of the innermost barrel layer is planned after  $250 \text{ fb}^{-1}$ .

3282 The Phase 1 pixel upgrade project is now at its last stage of assembly and testing of the  
 3283 entire system. A test stand have been setup at the University of Zurich (UZH) in 2013 and  
 3284 was operated until the end of 2016. It has been fundamental to test the performance of the  
 3285 complete upgraded BPix system and gain experience in its operations. This activity will  
 3286 guarantee the success of the installation and commissioning of the new detector planned for  
 3287 the beginning of March 2017, as well as excellent and stable performance during its first year  
 3288 of data-taking.

3289     This part of thesis is organized as follows. First, a description of the design and function-  
 3290 ality of the present BPix detector is given in Chapter 14. The work carried out during LS1  
 3291 aimed at optimizing the detector for LHC Run 2 is discussed in Chapter 15. The same chapter  
 3292 also includes details on the operations conducted during its re-installation into CMS and  
 3293 commissioning. Chapter 16 provides an overview of the design and features of the upgraded  
 3294 BPix system. In this chapter, a section is dedicated to the description of the test stand at  
 3295 UZH which I contributed to setup. The last section focuses on the development of new tests  
 3296 and procedures to be used during the upgraded detector assembly and commissioning.

# The CMS pixel barrel detector

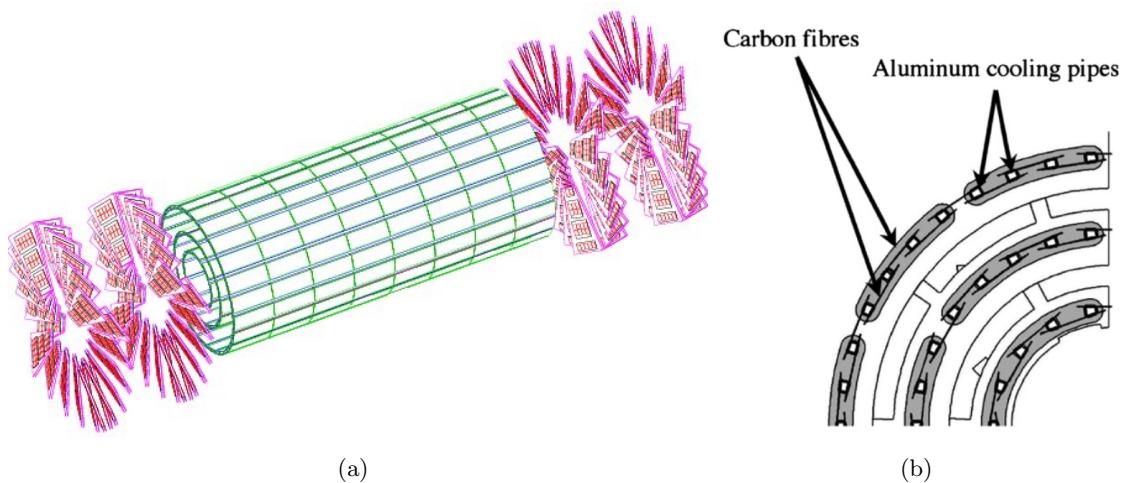
3301 This chapter provides an introduction to the design and functionality of the present CMS  
 3302 pixel barrel detector. It was developed, designed and built at the Paul Scherrer Institute (PSI)  
 3303 in cooperation with Eidgenössische Technische Hochschule Zurich (ETH) and the University  
 3304 of Zurich (UZH). The detector was installed in 2008 and it has been in operation for the  
 3305 entire first LHC run. During the first long shut-down it was temporarily installed in the  
 3306 clean room at the CMS experimental site, LHC-P5. The detector has been re-installed in  
 3307 December 2014 and commissioned for Run 2 in January 2015.

3308 Section 14.1 provides an overview of the detector design and mechanical structure, followed  
 3309 by a description of the detector module and its main building blocks (Section 14.2). In  
 3310 Section 14.3, the detector readout and control system are described. The last section provides  
 3311 an introduction to the structure and functionality of the pixel online software (POS) used for  
 3312 controlling and calibrating the detector. The detector calibration and its performance at the  
 3313 re-start of collisions in 2015 are discussed in the next chapter.

3314 

## 14.1 Design

3315 The CMS BPix detector [127] consists of three cylindrical layers at mean radii of 4.4, 7.3  
 3316 and 10.2 cm from the center of the detector and with a length of 53 cm. A three dimensional  
 3317 representation of the detector can be seen in Fig. 14.1(a). The layers are composed of 768  
 3318 modular detector units that consist of thin segmented silicon sensors, with a pixel size of  
 3319  $100 \times 150 \text{ m}^2$  providing about 48 million readout channels. The pixels are almost square  
 3320 shaped in order to achieve a similar track resolution in both the  $r\phi$  and  $z$  direction.



**Figure 14.1:** (a) Layout of the CMS pixel detector with three barrel layers (green) and four forward disks (red). (b) Detailed view in  $r\phi$  of the geometric layout.

3322 aluminum cooling pipes with 0.3 mm wall thickness. The cylindrical shape of the barrel is  
 3323 achieved by assembling several ladders on each layer half shell. The three half shells are then  
 3324 mounted together at the end flange building up half of the BPix detector. The total number  
 3325 of ladders per half shell is 10 for layer 1, 16 for layer 2 and 22 for layer 3. To guarantee full  
 3326 spatial coverage ladders are mounted with overlap on alternating sides of the cooling tubes.  
 3327 This is shown in Fig. 14.1(b). The resulting two detector parts are mechanically separated  
 3328 and the half shells are joined together with special ladders equipped with half modules. The  
 3329 overall layout results in 96 half modules and 672 full modules.

3330 The BPix detector is connected to four 2.2 m-long half cylinders (supply tubes) that carry  
 3331 the services along the beam pipe, accommodate the cooling lines and house the electronics for  
 3332 detector readout and control. The supply tubes are a complex system in design as well as in  
 3333 production due to the thin radial shell thickness (1–2 cm), the large number of circuits, plugs  
 3334 and sensors, the fine wires and thin printed circuit boards that are assembled. The detector  
 3335 and the supply tubes are connected via a six layer printed circuit board (PCB) which is  
 3336 mounted on the detector end flange and distributes the power and the control signals to the  
 3337 individual modules. The final BPix system consists of two independent half cylinders placed  
 3338 at  $+x$  (inner) and  $-x$  (outer) coordinates, with each half shell connected to two supply tubes  
 3339 placed at  $+z$  and  $-z$  coordinates.

## 3340 14.2 Detector modules

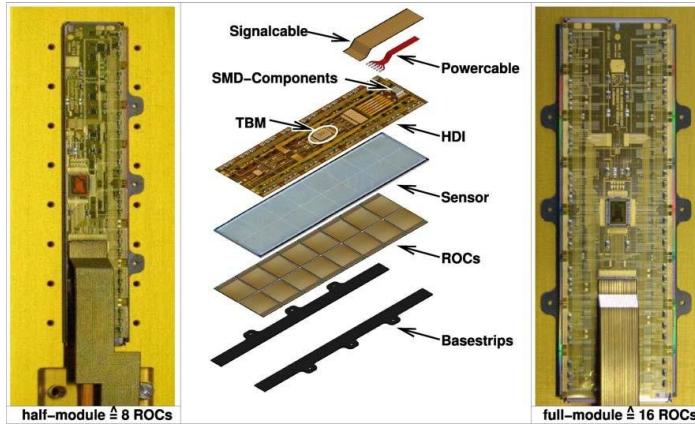
3341 The BPix modules are made of a thin, segmented silicon sensor [128] with a dimension of  
 3342  $66.6 \times 18.6 \text{ mm}^2$  that enables the detection of particles that pass through it by measuring  
 3343 the ionization charge that they produce. The charge measurement is performed by readout  
 3344 chips (ROCs) [129] that are connected to the sensor using the bump bonding technique with  
 3345 Indium solder. Full modules consist of two rows of 8 ROCs, while half modules use a smaller  
 3346 sensor with  $1 \times 8$  ROCs. Each ROC has a size of  $8 \times 8 \text{ mm}^2$  and reads a matrix of 4160 pixel  
 3347 readout channels. Table 14.1 summarizes the configuration of full and half modules on the  
 3348 three different BPix detector layers. The ROCs are wire-bonded to a three layer high density  
 3349 interconnect (HDI) flex printed circuit glued onto the backside of the sensor. Mounted on  
 3350 the top of the HDI, a token bit manager (TBM) chip [130] controls the readout of the ROCs,  
 3351 receives all external control signals and distributes them to the chips. Base-strips made of  
 3352  $250 \mu\text{m}$  thick silicon nitride ( $\text{Si}_3\text{N}_4$ ) are glued underneath the ROCs allowing the module  
 3353 to be mounted on the mechanical support structure. A power cable consisting of 6 copper  
 3354 coated aluminum wires is soldered to the HDI and brings analog, digital and high voltage  
 3355 to the module. The control and readout signals are sent through a two layer Kapton signal  
 3356 cable which is wire-bonded to the HDI. Additionally, the HDI distributes the signals and the  
 3357 voltages to the ROCs. The signal cables from the modules are plugged into the end flange  
 3358 that exists on both sides of the barrel and connects the three layers to the detector supply  
 3359 tubes. The modules are attached to cooling frames, with the cooling tubes being an integral  
 3360 part of the mechanical structure. The size of a full module is  $66.6 \times 26 \text{ mm}^2$  and the weight  
 3361 is up to 3.5 g depending on the length of the signal and power cables. The average power  
 3362 consumption of a full module is 2 W. An exploded view of the BPix module is shown in  
 3363 Fig. 14.2.

### 3364 14.2.1 Sensor

3365 The sensor is made from a  $n$ -type silicon wafer with a thickness of  $285 \mu\text{m}$ . Charged particles  
 3366 that travel through the sensor material leave electron-hole pairs as the result of multiple

**Table 14.1:** Number of modules, readout chips and pixel channels for the three detector layers.

radius (cm)	full modules #	half modules #	ROCs #	pixels ( $10^6$ )
4.4	128	32	2304	9.6
7.3	224	32	3840	16.0
10.2	320	32	5376	22.4
Total	672	96	11520	48

**Figure 14.2:** Picture of a BPix half module (left) and full module (right). In the center, the components of the module are shown. From top to bottom: the Kapton signal cable, the power cable, the HDI, the silicon sensor, the 16 ROCs and the base strips.

3367 interactions with the atoms in the material. For charged particles at intermediate energies  
 3368 ( $0.1 \leq \beta\gamma \leq 1000$ ), the average energy loss  $dE$  in a thickness  $dx$  of material is described by  
 3369 the *Bethe-Bloch formula*

$$-\left\langle \frac{dE}{dx} \right\rangle = 4\pi N_A r_e^2 m_e c^2 z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[ \frac{1}{2} \ln \frac{2m_e c^2 \beta^2 \gamma^2 W_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right]. \quad (14.1)$$

3370 In the above equation,  $N_A$  is the Avogadro's number,  $r_e$  the classical electron radius,  
 3371  $m_e$  the electron mass,  $z$  the charge of the particle,  $Z$  ( $A$ ) the atomic number (mass) of the  
 3372 material ( $Z = 14$  and  $A = 28.1$  u for silicon),  $W_{\max}$  the maximum energy transfer to an  
 3373 electron in a single collision,  $I$  the mean excitation energy, and  $\delta$  a density effect correction.  
 3374 At a particle velocity  $\beta \approx 0.96$  ( $\beta\gamma \approx 3$ ) a broad minimum is reached. At higher energies  
 3375 the logarithmic term leads to a slow rise again, which is eventually canceled by the density  
 3376 correction. A particle with an energy loss in the minimum is called a minimum ionizing  
 3377 particle (MIP).

3378 The energy loss in a finite medium is subject to statistical fluctuations well described by  
 3379 a *Landau distribution* (Fig. 14.3). If a particle is not stopped in the medium, the energy loss  
 3380 (and therefore the number of charge carriers) varies around the peak of the distribution (most  
 3381 probable value or MPV). In rare but measurable cases ( $\delta$ -rays or  $\delta$ -electrons), the transferred  
 3382 energy is large, so that these cases are responsible for the asymmetric long tail towards high  
 3383 charge deposits. Due to this tail the most probable value of energy transfer is about 30%  
 3384 lower than the average value. For a MIP crossing the sensor at an angle of  $90^\circ$  the most  
 3385 probable number of electron-hole pairs generated in  $1 \mu\text{m}$  of silicon is 76. Therefore, a MIP

3386 generates a signal of about 22,500 electron-hole pairs (MPV).

3387

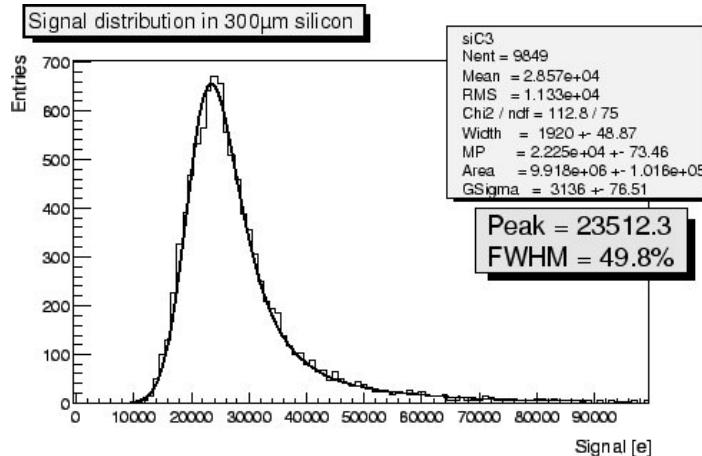
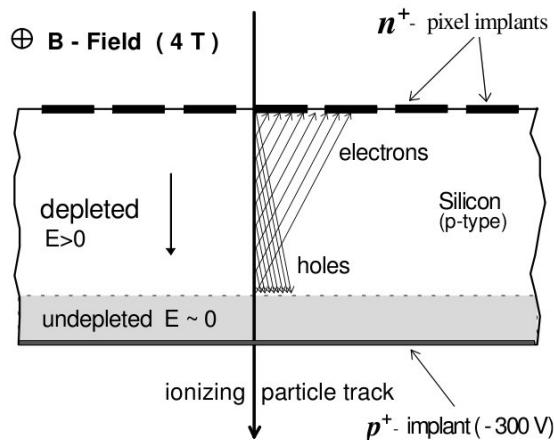


Figure 14.3: Measured MIP signal distribution in a Silicon detector of  $300\text{ }\mu\text{m}$  thickness.

3388 The silicon sensor used for the CMS pixel detector adopts a double sided  $n^+$ -in- $n$  design:  
 3389 pixels consist of high dose  $n^+$  implants on a high resistance  $n$  substrate. The backside of  
 3390 the substrate is p-doped, therefore the p-n junction is placed on this side of the sensor. A  
 3391 cross-section of the sensor is shown in Fig. 14.4. If the junction is reverse biased, a depletion  
 3392 zone forms that extends towards the pixel implants. In this zone, an electric field is established  
 3393 that allows ionization charge to drift. Electrons drift toward  $n^+$  implants while holes drift  
 3394 toward the back of the sensor. In Fig. 14.4, the bulk of the silicon is p-type because of the type  
 3395 inversion occurring in the bulk after prolonged exposure to high fluences of radiation. In fact,  
 3396 the effective concentration of impurities gradually decreases with exposure, until a transition  
 3397 to the other type material behavior occurs. At this stage, the depletion zone grows  
 3398 from the pixel implants toward the back of the sensor, enabling the collection of electrons  
 3399 even when the sensor is only partially depleted. Extremely high operating voltages can  
 3400 therefore be avoided, reducing the problems of leakage currents and high voltage breakdowns.  
 3401 Furthermore, the double-sided processing of  $n^+$ -in- $n$  detectors allows a guard ring concept  
 3402 which keeps all sensor edges at ground potential and avoids the risk of disruptive discharges  
 3403 to the very closely spaced front-end chip.

3404 Additional processing is needed on the readout side to electrically isolate the  $n^+$  implants  
 3405 from each other. The electron accumulation layer induced by ionizing radiation otherwise  
 3406 tends to short-circuit the pixel implants. A moderated p-spray technique is used, which  
 3407 consists of a medium dose p-type Boron implants.

3408  
 3409 The position resolution of single-pixel hits is given by the pixel pitch divided by the  $\sqrt{12}$ .  
 3410 However, the spatial resolution can be improved exploiting charge sharing among adjacent  
 3411 pixels. A group of pixels collecting a signal from the same particle is usually called “cluster”.  
 3412 Significant charge sharing is a consequence of the Lorentz drift in the strong magnetic field  
 3413 of 4 T inside CMS. In fact, charge carriers released by the ionizing particle in the silicon  
 3414 sensor do not follow the electric field lines to the collection electrodes, but are deflected by  
 3415 the Lorentz force (Fig. 14.4). Furthermore, the analog readout of the CMS pixel detector  
 3416 allows for an interpolation of the amounts of collected charge for each of the pixels in the  
 3417 cluster. This effect influenced the choice of the barrel pixels size which have been optimized  
 3418 to achieve the optimal spatial resolution. Two-pixel clusters and the interpolation allow a



**Figure 14.4:** Illustration of a charged particle crossing a sensor of the CMS pixel detector. The charge carriers produced by the passage of the ionizing particle are collected at the high dose  $n^+$  implants.

much better resolution, limited only by fluctuations of the charge deposition. Since division of the signal charge among more than two pixels increases the data rate and reduces the signal charge per pixel without an improvement of the resolution, an ideal choice of the pixel size in the direction perpendicular to the magnetic field ( $r\phi$ ) is therefore given by the length over which charges are spread when they reach the surface of the sensor. For the usual  $\sim 300\text{ }\mu\text{m}$  sensor thickness and a Lorentz angle of  $28^\circ$  this amounts to  $\sim 150\text{ }\mu\text{m}$ . A slightly smaller size of  $100\text{ }\mu\text{m}$  was chosen to maintain charge sharing, and hence resolution, after irradiation. The area of a pixel must also be large enough to accommodate the readout electronics. With one dimension fixed by the Lorentz drift, this leads to a more or less quadratic shape of  $100\text{ }\mu\text{m}(r\phi) \times 150\text{ }\mu\text{m}(z)$ , resulting in comparable resolution in both directions.

### 14.2.2 Readout chip

The readout chip is responsible for measuring the charge deposited by a particle in the sensor's pixel. It amplifies and samples the signal with a time resolution of 25 ns, which is the time between two LHC bunch crossings. The pixel hit information have to be stored on-chip during the CMS Level-1 trigger latency of  $3.2\text{ }\mu\text{s}$  after which they are either read out or discarded. Each pixel sensor is connected via a bump bond to its own readout circuit on the ROC, referred to as *Pixel Unit Cell* (PUC). The PUCs are arranged in  $26 \times 80$  double columns. Each double column represents an independent readout unit controlled by a circuit sitting in the column periphery from where the PUC is controlled, data are buffered and global functions common to all pixels are located.

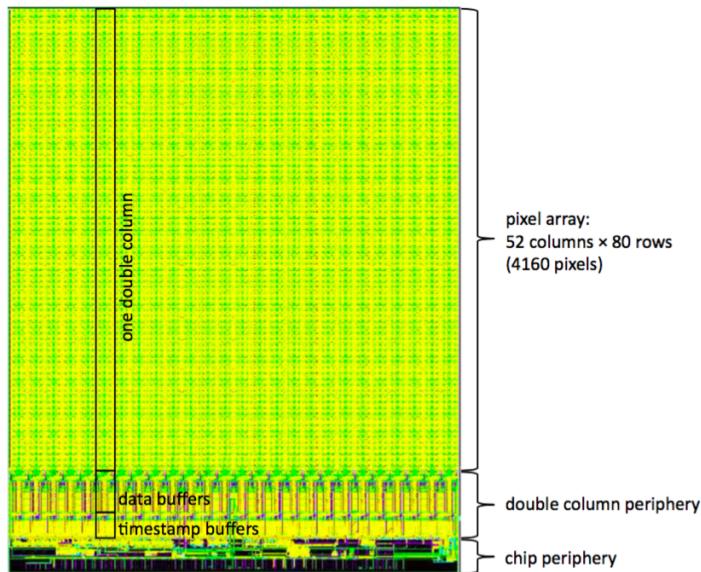
To control and optimize the readout, 26 digital-to-analog converters (DAC) can be programmed using a serial protocol similar to I<sup>2</sup>C modified to operate at 40 MHz.

The PUC can receive a signal either through a charge deposition in the sensor or by injecting a calibration signal. Within the PUC, the signal is first passed through a two stage pre-amplifier/shaper system to a comparator where zero-suppression is performed. It compares the shaper output to a threshold value which is programmed by a DAC distributed globally to all pixels. Since variations of the threshold of the individual pixels caused by transistor mismatch, voltage drops or preamplifier gain variations can lead to an increased noise rate, each pixel has a 4-bit DAC to trim the threshold. Furthermore, a mask bit allows

3448 to disable noisy pixels. When the rising edge of the signal has passed the threshold, the  
 3449 signal height is sampled after some delay and stored in the sample-and-hold capacitance until  
 3450 the readout mechanism is started from the periphery. During this time the pixel becomes  
 3451 insensitive.

3452 Since the L1 trigger latency time in CMS is  $3.2\ \mu\text{s}$  (128 bunch crossings), the information  
 3453 of a hit pixel, including the associated bunch crossing number and the analog signal, can not  
 3454 be kept on the pixel itself during this time without introducing a significant inefficiency. In  
 3455 the *Column Drain Architecture* chosen for the CMS pixel readout, all pixel hits occurring in  
 3456 a pixel double column are immediately and quickly copied into the column periphery in order  
 3457 to free the pixels for the next hit. In this case the probability of having a second hit in the  
 3458 pixel during the latency is significantly reduced. Each double column informs the column  
 3459 periphery immediately of any hits that have been occurred by sending to the periphery  
 3460 a current with adjustable intensity. The column periphery writes the value of the bunch  
 3461 crossing counter into a time stamp buffer and initiates a token scan of the double column  
 3462 passing a readout token from cell to cell. Once the hit pixel is found, in the readout block  
 3463 of the PUC the token signal initiates the transfer of pixel address and analog pulse-height  
 3464 information, which are stored in a data buffer located in the periphery waiting for the L1  
 3465 trigger. The hit pixels remain inactive until their hit information has been transferred. The  
 3466 double column periphery verifies the trigger by comparing the time stamp with a counter  
 3467 running behind the bunch crossing counter by the trigger delay. In case of agreement the  
 3468 column is set into readout mode and the data acquisition is stopped, otherwise the data are  
 3469 discarded. When the readout token arrives at the double column periphery the validated  
 3470 data are sent to the chip periphery and the double column is reset. The ROCs are read out  
 3471 serially via a 40 MHz analog link.

3472 A picture of the BPix readout chip is shown in Fig. 14.5.



**Figure 14.5:** Picture of the CMS pixel readout chip showing the three main building blocks: double column, double column periphery and chip periphery [4].

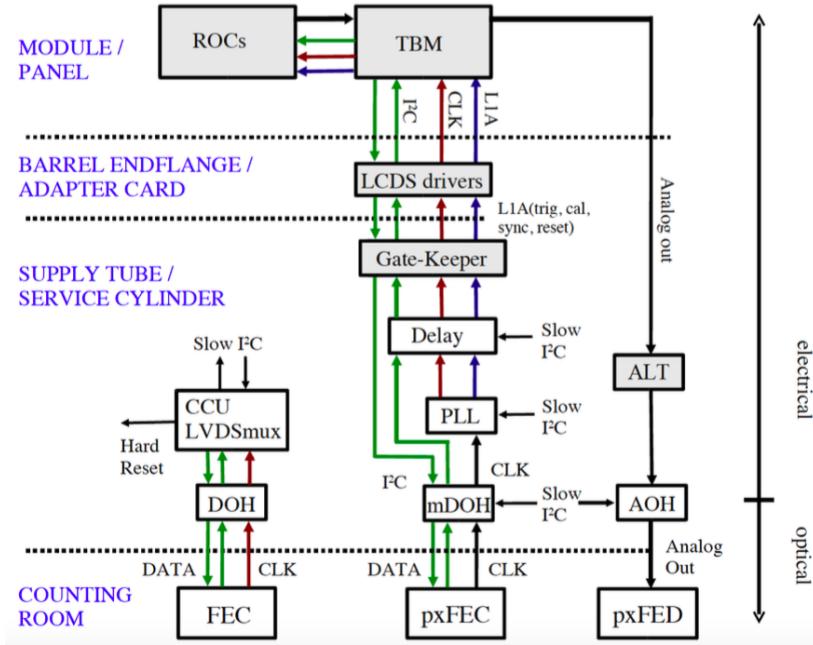
### 3473 14.2.3 Token bit manager

3474 A token bit manager chip is wire-bonded to the HDI and controls the readout of the ROCs. It  
 3475 serves as an interface for data acquisition and programming and is responsible to synchronize  
 3476 the readout of the ROCs on the module. For each incoming L1 trigger, the TBM sends a  
 3477 token in a fixed order from chip to chip, and waits until the token returns from the last chip  
 3478 in the chain. The chip that has the token transmits all hits for a given trigger and then  
 3479 passes the token to the next chip. Upon the arrival of the token, each ROC sends a three  
 3480 clock cycle header. While the header is transmitted, the token is passed through the chip  
 3481 looking for a double column with validated hits belonging to that token. The length of the  
 3482 header is sufficient for the token to skip all 26 double columns if no triggered hits were present  
 3483 and to be passed on to the next chip with the right timing. Triggers and readout tokens  
 3484 are both counted and hits are only readout when the token number matches the readout  
 3485 number. It must be ensured that exactly one token for every trigger is issued and that there  
 3486 is never more than one. The ROCs in the module are either serviced by a single token that  
 3487 sequentially passes through all the 16 chips, or a second channel in a dual TBM chip is used  
 3488 such that the ROCs are divided into two groups of eight. This method is employed for the  
 3489 two innermost layers of the detector which experiences higher hit rates per module than the  
 3490 others. It requires two separate buses for the readout of the two groups of ROCs, and the  
 3491 data streams are also individually transmitted through two separate readout links for the  
 3492 data acquisition. The TBM multiplexes the signal from the ROCs, adds a header and a  
 3493 trailer to the data stream and drives the signal through the readout link. In addition, the  
 3494 TBM distributes the L1 trigger and the clock to the ROCs. The header contains an event  
 3495 number and the trailer a status information, such as the stack overflow warning.

3496 The TBM keeps track of triggers arriving while the token is still under way with a trigger  
 3497 stack of 32 entries that is filled each time a trigger arrives and reduced every time a token  
 3498 returns. In case of a stack overflow, the TBM withholds the incoming triggers from the ROCs  
 3499 until the stack is reduced. It notifies the data acquisition that events have become lost in  
 3500 this case. The TBM chip also includes a communication component called the HUB which  
 3501 serves as a port for programming commands sent from the DAQ.

## 3502 14.3 Detector readout and control

3503 A schematic drawing of the pixel readout and control systems is shown in Fig. 14.6. The path  
 3504 on the right shows the readout part of the system. Signals from a group of ROCs are amplified  
 3505 and converted into a 40 MHz analog optical signal in the analog opto-hybrid circuits mounted  
 3506 on the supply tube. Optical fibers allow the data to be transferred over approximately 60 m  
 3507 distance to the underground counting room, where a VME front-end driver unit (FED) [131]  
 3508 digitizes the signal, builds event fragments and sends them to the DAQ. The signal path in  
 3509 the middle shows the fast detector control link. In the counting room, the control signals  
 3510 are driven by front-end control (FEC) units [132] which are used to program the detector  
 3511 modules. The signals enter the supply tubes through optical fibers to be converted in digital  
 3512 opto-hybrid circuits. Several other electronic devices are needed by the system and are placed  
 3513 on the supply tubes. Some of these components need to be programmed. This happens  
 3514 through a dedicated slow control link corresponding to the signal path on the left. Also  
 3515 shown in Fig. 14.6 is the Timing Trigger and Control (TTC) [133] system which distributes  
 3516 the clock and trigger signals to all detector components. The individual electronic devices of  
 3517 the detector readout and control system are described in more detail in the following.



**Figure 14.6:** Overview of the BPix readout and control system.

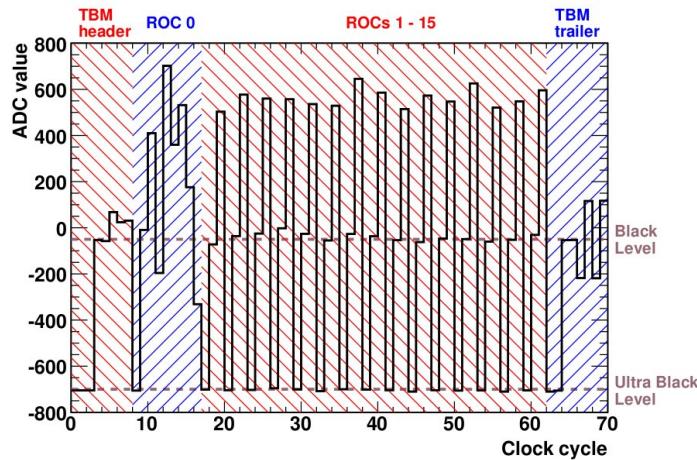
### 14.3.1 Readout of the analog signal

An example of an analog readout signal from a module with a single pixel hit is shown in Fig. 14.7. The TBM header uses eight clock cycles and starts with three ultra black levels (UBL). An UBL is simply a large negative signal level well outside of the range of pixel data. The three UBLs are followed by a black level (BL), which defines the zero level of the differential analog signal. The four remaining clock cycles encode an 8-bit event number. The minimal readout of each ROC starts with an UBL, a BL and a level called last DAC which represents the value of the most recently programmed DAC. A pixel hit adds a block of six clock cycles to the ROC minimal readout: two for the double-column address, three for the row address, and one for the pulse height. In order to speed up the transmission of digital pixel hit information while maintaining the global 40 MHz clock, the pixel addresses are not sent in a common binary fashion, but the available signal amplitude is divided into six possible analog levels (2.5 bits/clock). The readout is terminated by the TBM trailer, containing two UBLs, two BLs, and four clock cycles with the TBM status information.

The data stream which contains all hit information belonging to a single trigger is sent out by the TBM through the module Kapton cable. Kapton cables bring the analog signals to the PCB on which the analog optical hybrids (AOHs) [134] are plugged. The electric analog signals are amplified in an Analog Level Translator (ALT) chip and converted into 40 MHz analog optical signals in the AOHs. Each AOH is equipped with 6 lasers with adjustable gain and threshold, which drive the signal through optical fibers to the front-end drivers.

A total of 32 FED modules, located in the underground counting room, receive the data packets. They convert the signals from optical to electrical, perform the digitization at the LHC frequency, and decode the pixel hit information. The pixel FED also builds event fragments and sends them to the CMS central DAQ system. It is a 9 U VME module designed at HEPHY Vienna that includes optical receivers, ADCs and several FPGAs for signal processing.

A FED has three opto-receiver devices each of which has twelve input channels where



**Figure 14.7:** Analogue readout of a pixel module with one hit in ROC 0.

3545 the fibers terminate. Each input channel is equipped with a 10-bit ADC. The ADC has a  
 3546 clock with adjustable phase w.r.t. the global clock in steps of 1.6 ns to select the optimum  
 3547 digitization sampling point for each input. A programmable offset voltage can be set for each  
 3548 optical input in order to compensate for bias shifts in the analog signal. An additional, single  
 3549 optical input receives TTC signals such as clock, trigger and reset. Each channel has a 1k  
 3550 words 32-bit data buffer (FIFO1) which stores the digitized module signal (Fig. 14.7). The  
 3551 data package from four or five (depending on the location) FIFO1 channels are collected in a  
 3552 FIFO2 with a size of 8k words and a width of 64 bits plus 4 control bits. During the data  
 3553 transfer to FIFO2 the input event number is compared with that of the CMS TTC system.  
 3554 The data of all FIFO2 memories are collected by two final memories (FIFO3) of 8k words  
 3555 each over two buses of 64+4 bits at 40 MHz. Four front FPGAs, each handling 9 inputs,  
 3556 house FIFO1 and FIFO2 buffers, while a center FPGA houses the FIFO3 where the whole  
 3557 event fragment is built, together with the S-link connection to the central DAQ. The FED  
 3558 can also be operated in a transparent mode making unprocessed ADC output data available  
 3559 for calibration and testing purpose.

### 3560 14.3.2 Detector control and programming

3561 The detector control and programming is performed through front-end control modules (pixel  
 3562 FECs or pxFEC) located in the underground counting room. The function of the pixel  
 3563 control system is to send the 40 MHz clock, the trigger and control signals (e.g. resets) to the  
 3564 front-end electronics, and to program all front-end devices (TBMs and ROCs). All the signals  
 3565 are sent through optical fibers and converted to electrical signals by digital optical hybrids  
 3566 (DOHs) [134] mounted on the supply tube before forwarding them to the pixel modules. A  
 3567 DOH is connected to four optical fibers, two for receiving and two for sending signals. The  
 3568 LHC 40 MHz clock and trigger information is encoded in one signal which is sent over a  
 3569 single fiber to the DOH. A modified version of the common I<sup>2</sup>C protocol has been developed  
 3570 to cope with the required volume of the data that have to be downloaded to configure the  
 3571 detector modules. The main modifications include the increase of the clock speed to 40 MHz  
 3572 and dropping the requirement of an acknowledge signal. Each DOH contains two laser drivers  
 3573 and two PIN diodes. The DOH is mounted on the digital opto-board together with other  
 3574 electronic devices needed by the system. In particular, a phase locked loop (PLL) chip [135] is  
 3575 used to split the clock from the trigger, and the Delay25 chips [136] adjust the relative phases

3576 of all control signals. The Gate-Keeper chip converts the LVDS signals used by PLLs and  
3577 Delay25s to low current differential signals (LCDS) used by the pixel front-end chips. Finally,  
3578 the LCDS-driver chips mounted on the end flange PCB are used to drive the signals on the  
3579 Kapton cables to each detector module. In addition, these chips are used to compensate the  
3580 signal phases for the different Kapton cable lengths.

3581

3582 The electronics on the supply tubes (DOHs, PLLs, Delay25s, AOHS, and so on) have to be  
3583 controlled and programmed. This is achieved through a system of four CCU (Communication  
3584 and Control Unit) boards equipped with 9 CCU chips [137]. This is indicated in Fig. 14.6  
3585 as “slow I<sup>2</sup>C”, since the standard I<sup>2</sup>C protocol is implemented for this task. The boards are  
3586 mounted on the supply tubes and each of them supervises one quarter of the detector.

3587

3588 The slow control links are implemented as a ring architecture. A ring consists of 9 CCUs,  
3589 two optical drivers and receivers that bring clock, trigger and control data to the CCUs,  
3590 and a front-end controller (tracker FEC or tkFEC) which provides the communication with  
3591 the CCUs and the programming signals. Each CCU distributes the digital control and  
3592 programming signals to a set of individual boards forming one readout sector of the detector.  
3593 A CCU chip supports two I<sup>2</sup>C channels to communicate with the front-end electronics, and  
3594 three PIA channels to generate the necessary signals to reset the circuits and the ROCs of  
3595 one sector. Eight CCUs are used for the control of the eight sectors in which each supply tube  
3596 is divided, the ninth CCU is a dummy chip used for redundancy. Since a large number of  
3597 front-end channels depend on the same control link, a very high reliability of the system is of  
3598 utmost importance. A CCU failure leads to a loss of communication to all electronics attached  
3599 to it. A redundancy scheme based on doubling signal paths and bypassing of interconnection  
3600 lines, between the CCUs and between the CCUs and the FEC, is supported. The dummy  
3601 CCU allows to mitigate a single DOH failure. The CCU is equipped with two DOHs which  
3602 form separated control rings and thus ensure a high operational reliability. The DOHs on the  
3602 CCU board are programmed by the first two CCU chips.

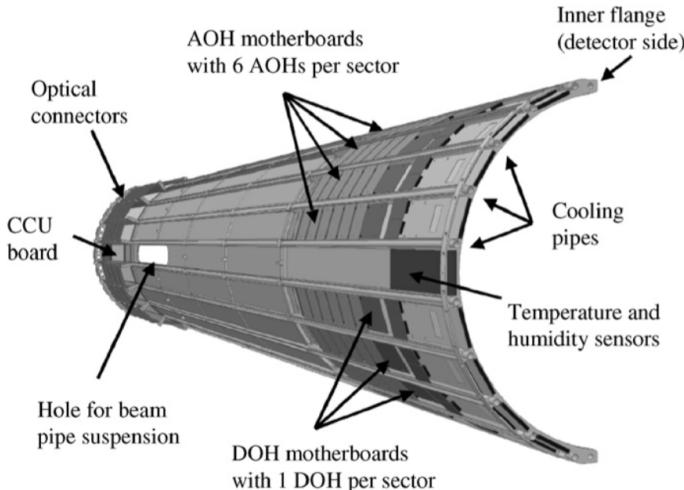
3603

### 14.3.3 Supply tubes

3604 As mentioned previously in this chapter, the readout and control circuits of the pixel detector  
3605 are integrated on four supply tube half cylinders. In addition, the supply tubes bring the  
3606 power and cooling lines to the detector. A schematic view of a supply tube is shown in  
3607 Fig. 14.8. A supply tube is divided into 8 sectors which contain the power lines and electronics  
3608 of two readout groups, one serving the modules of the first two layers, the other serving the  
3609 modules of the third layer. One sector includes an analog opto-board with six AOHS, a digital  
3610 opto-board with two DOHs, two PLL chips, two Delay25 chips and two Gate-Keeper chips.  
3611 A total of 192 AOHS and 72 DOHs are used for the pixel barrel detector. For each sector, 44  
3612 optical fibers drive the communication with the front-end modules, 36 for the analog readout  
3613 and 8 for the digital control. The CCU board is placed in the central sector of the supply  
3614 tube.

3615

3616 The stability of the analog signal is strongly affected by the temperature dependence  
3617 of the AOHS. A shift of 50 ADC counts is observed in the level of the analog signal for a  
3618 temperature variation of the AOH of 1° C. The FED is able to internally correct for a drift  
3619 within a temperature range of ±2° C. Consequently, the temperature of the AOHS has to be  
3620 controlled within a very narrow range in order to assure a stable operation of the detector.  
3621 The barrel pixel supply tubes are equipped with a total of 124 temperature sensors and  
3622 8 humidity sensors. The temperature sensors are placed on the CCU boards, the analog  
3622 opto-board and on the supply tube cooling lines.



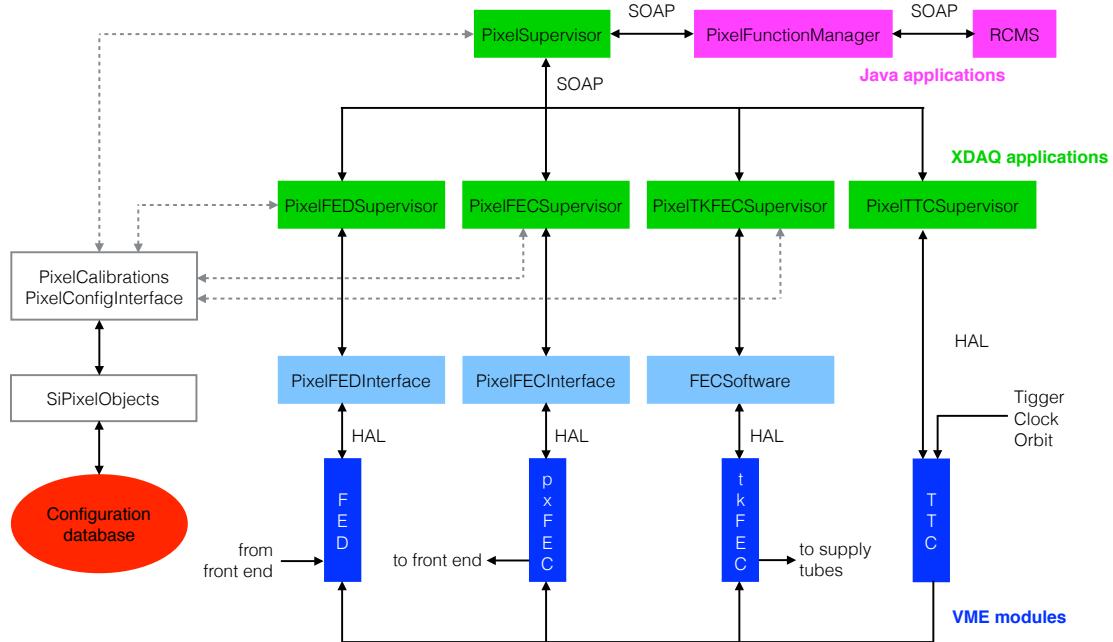
**Figure 14.8:** Layout of a BPix supply tube [127].

## 3623 14.4 Pixel online software

3624 The pixel online software is the framework used for controlling and calibrating the CMS pixel  
 3625 detector and that runs on the PCs of the CMS control room at LHC-P5. Its main functional-  
 3626 ties are to configure the detector, perform calibrations, analyze calibration data and monitor  
 3627 the detector during data taking. The pixel online software is based on the XDAQ toolkit [138]  
 3628 and is written in C++. It has a very complex structure built from a large number of different  
 3629 applications and packages. The dependencies among the main applications and packages  
 3630 is presented in Fig. 14.9. The top level application is represented by the **PixelSupervisor**  
 3631 which is responsible for the overall coordination of the pixel DAQ. Its main function is  
 3632 to coordinate the activities of the other supervisors, particularly during configuration and  
 3633 calibration. It is also responsible for updating the configuration database with new settings  
 3634 obtained by calibrations. Among the other supervisors there is the **PixelFECSupervisor** that  
 3635 controls the pxFECs and is responsible for loading the configuration parameters for the ROCs  
 3636 and TBMs from the configuration database and programming those parameters into the  
 3637 detector. Similarly, the **PixelTKFECSupervisor** controls the tkFECs and the initialization of  
 3638 all the electronics placed on the supply tubes. The **PixelFEDSupervisor** controls the FEDs.  
 3639 An additional supervisor is included in the software to control the pixel TTC module used  
 3640 for trigger and timing. Among other things the TTC module is used during calibrations  
 3641 to generate triggers. The various supervisors run as independent processes, or even on  
 3642 different computers. Therefore, in order to communicate with each other they must exchange  
 3643 messages on the network. This is done using the XML-based SOAP (Simple Object Access  
 3644 Protocol) protocol. A set of classes such as **PixelFEDInterface**, **PixelFECInterface**, and  
 3645 **FECSoftware**, provides the direct communication between the supervisors and the VME  
 3646 hardware via Hardware Access Library (HAL) [139].

3647 A function manager acts as an interface between the global run control (Run Control and  
 3648 Monitoring System or RCMS) and POS. It is a JAVA application that basically passes the  
 3649 state machine of CMS (Halted, Running, Configured, and so on) to the **PixelSupervisor**  
 3650 which then forwards state requests to the underlying supervisors to carry out the different  
 3651 tasks needed in state transitions of the run control.

3652 Another key element of the software is represented by the **PixelConfigInterface** package



**Figure 14.9:** Illustration of the dependencies among the main applications and packages implemented in the pixel online software.

which provides access methods for retrieving and storing configuration data. Several different classes are available in the `SiPixelObjects` package, each responsible for storing a specific set of detector settings as well as the configuration needed by the calibration code (e.g., which detector parameter to scan and its range). For instance, the `PixelNameTranslation` class translates from the naming scheme used to label each individual ROC to the hardware addresses used by both the FEC and the FED to identify a specific ROC. Similarly, the `PixelDACSettings` and `PixelTBMSettings` classes store, respectively, the DAC settings for all ROCs on one module and the settings for one TBM. The `PixelSupervisor` features a web GUI that can be accessed as an html page. It displays information about the current configuration, or if it is not configured it allows the user to select a possible configuration from a list and configure the detector using that choice. The GUI is also used to run and monitor the detector calibrations. The calibration routines are implemented in independent and separate classes contained in the `PixelCalibrations` package. The description of the detector calibration procedures is presented in the following chapter.

# Optimization and commissioning for LHC Run 2

3670

3671 The CMS pixel detector was designed to cope with the high radiation environment of LHC and  
 3672 to operate with the highest performance even after the accumulation of significant radiation  
 3673 doses. Nevertheless, radiation damage affects hit efficiency and resolution and hence, it is  
 3674 necessary to monitor its effects during operations. As described in this chapter, throughout  
 3675 Run 1, re-calibrations of the detector have been performed to compensate for these effects  
 3676 and recover full performance.

3677 During LS1, both BPix and FPix were extracted from CMS for maintenance with the  
 3678 purpose to recover broken channels. In this period, the calibration procedure has been  
 3679 exercised and improved in view of commissioning and operations for Run 2.

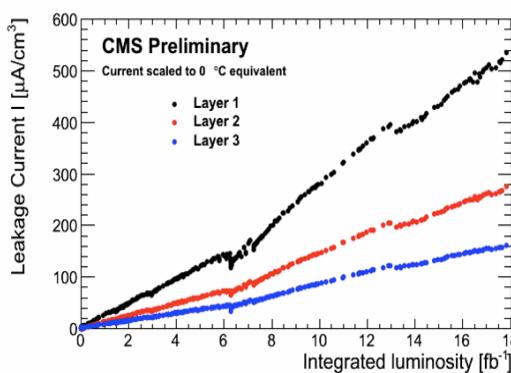
3680 The pixel detector has been operated with a coolant temperature of 7.4 °C in 2008–2011  
 3681 and 0 °C in 2012, which for the pixel sensors translates to values  $\approx 10$  °C higher. In order to  
 3682 limit the impact of radiation damage, during Run 2 the detector has been operated at much  
 3683 lower temperature, down to -10 °C. This has been made possible thanks to a major effort  
 3684 during the long shut-down to implement a tracker wide sealing that ensures minimal humidity  
 3685 levels. The flow of dry gas into the tracker volume was increased and a new safety system  
 3686 was developed that shuts down the detector safely in case a sudden increase of temperature,  
 3687 electric current or humidity is detected. During LS1, the pixel detector functionalities at very  
 3688 low temperature have been checked and its (temperature dependent) settings re-calibrated to  
 3689 allow for optimal operations under such conditions. This activity, described in the following,  
 3690 have been crucial to achieve a quick and reliable re-installation and commissioning for Run 2,  
 3691 as well as for stable and excellent operations during 2015 and 2016.

## 3692 15.1 Effects of radiation damage in LHC Run 1

3693 One of the first visible effect of radiation is the increase of the sensor leakage current with  
 3694 integrated luminosity, due to damages in the silicon bulk. The most fundamental type of bulk  
 3695 radiation damage is a defect, produced by the displacement of an atom of the semiconductor  
 3696 material from its normal lattice site. The vacancy left behind, together with the original  
 3697 atom now at an interstitial position, constitutes a trapping site for normal charge carriers.  
 3698 The formation of mid-gap states facilitates the transition of electrons from the valence to the  
 3699 conduction band leading to an increase of the leakage current in the depletion region. The  
 3700 primary defects caused by irradiation are not stable but able to move through the crystal. As  
 3701 result of this diffusion process, there is the possibility of combination of more complex defects.  
 3702 This process is called *annealing*, with a beneficial part reducing the damage and a reverse  
 3703 one degrading macroscopic sensor properties, called *reverse annealing*. During beneficial  
 3704 annealing, with a time constant of a few days at room temperature, the leakage current  
 3705 decreases, while later it rises due to reverse annealing process until it finally saturates at a  
 3706 value which is significantly above the initial level. At temperatures below 0 °C however, both  
 3707 effects can be frozen and the detector current remains constant. Thus, irradiated detectors

3708 should be operated and stored at low temperature, while it is favorable to shortly expose  
 3709 them to room temperature to take advantage of the beneficial annealing.

3710 Figure 15.1 shows the increase of the leakage current during Run 1 for the pixel barrel  
 3711 layers measured from readings of the high voltage power supplies as a function of the integ-  
 3712 rated luminosity. The damage was only partially recovered by beneficial annealing that took  
 3713 place during a longer shut-down after about  $6 \text{ fb}^{-1}$  and a shorter technical stop after about  
 3714  $13 \text{ fb}^{-1}$  delivered integrated luminosity. Between the end of 2011 and the beginning of 2012  
 3715 the operating temperature was decreased from  $7.4^\circ\text{C}$  to  $0^\circ\text{C}$  achieving a reduction in leakage  
 3716 current by a factor two and preventing reverse annealing which would eventually require too  
 3717 high depletion voltages.  
 3718

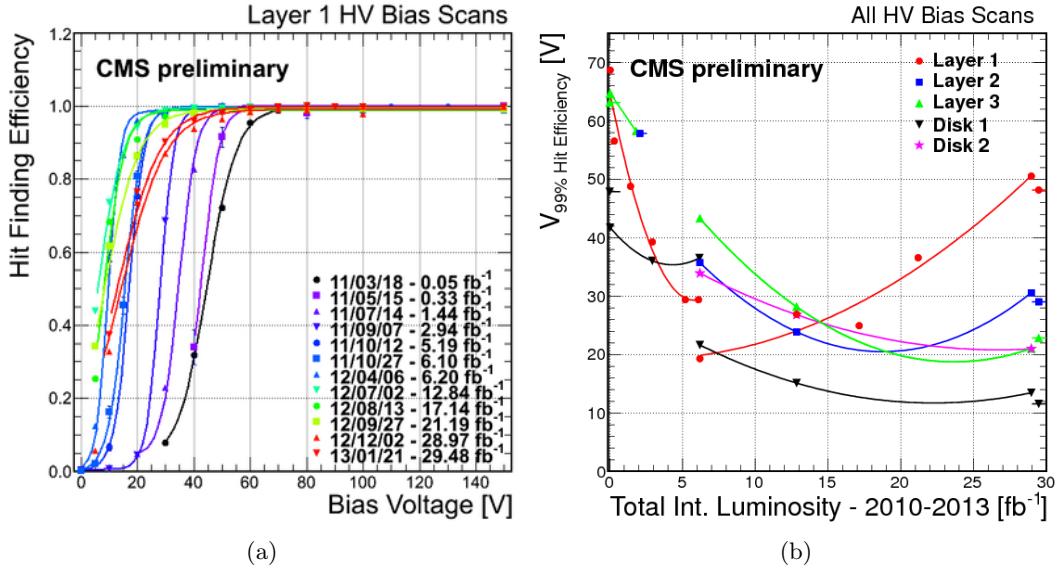


**Figure 15.1:** Leakage current scaled to  $0^\circ\text{C}$  operational temperature for the barrel layers as a function of the integrated luminosity delivered in Run 1 that represents the accumulated irradiation [11].

3719 The depletion voltage was also monitored during operations. With irradiation, defects  
 3720 with a negative space charge are generated throughout the bulk leading to variations in  
 3721 the effective doping concentration. When starting with n-type bulk, the effective doping  
 3722 concentration decreases because of the negatively charged defects until the bulk is transformed  
 3723 into an effective p-type. This process, called *type inversion*, happens at a relatively low dose  
 3724 of several  $10^{12} n_{eq}/cm^2$  (neutron equivalent fluence) [140]. As a consequence of this space  
 3725 charge sign inversion, the depletion zone now expands from the  $n^+$  pixel implants towards  
 3726 the p-type back. The depletion voltage scales with the bulk doping concentration: it initially  
 3727 decreases reaching a minimum at the inversion point, and then rises with the effective bulk  
 3728 doping concentration.

3729 A dedicated scan of the bias voltage was performed several times per year, by varying the  
 3730 detector bias voltage from 0 V to the normal operating value of 150 V, and measuring  
 3731 the single hit efficiency. The results of the hit efficiency measurements for the innermost  
 3732 barrel layer between 2011 and the beginning of 2013 are shown in Fig. 15.2(a). The bias  
 3733 voltage that is needed to reach a depletion depth corresponding to full hit efficiency decreases  
 3734 with irradiation at first, then increases as expected due to the aforementioned changes in  
 3735 the effective doping. The dependence of the voltage needed to achieve full hit efficiency on  
 3736 the integrated luminosity is shown in Fig. 15.2(b) for the barrel layers and endcap disks.  
 3737 The presence of a minimum for the layer 1 and layer 2 is evidence for type inversion occurrence.  
 3738

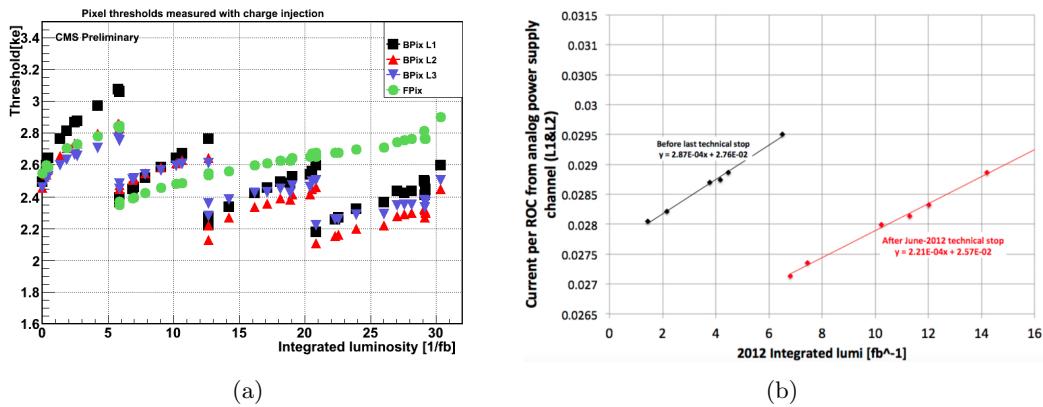
3739 The evolution of the pixel threshold (Fig. 15.3(a)) and the analog current (Fig. 15.3(b))  
 3740 was also frequently monitored in Run 1, and an increase of both parameters with integrated  
 3741 luminosity was observed. The possible explanation for these changes is the radiation damage



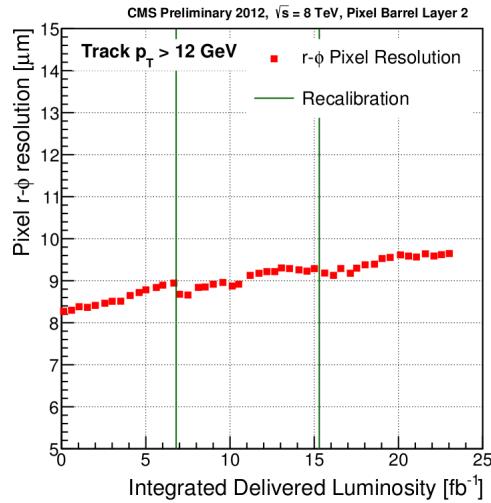
**Figure 15.2:** (a) Scans of the bias voltage performed on the innermost barrel layer. (b) Bias voltage corresponding to full single hit efficiency for all barrel layers and forward disks as a function of the integrated luminosity delivered in Run 1 [11].

in the bad-gap reference voltage circuit, which would shift all voltage settings inside the ROC. Because of the described effect, a re-calibration of the analog voltage and the pixel threshold during technical stops was necessary to recover the optimal ROC performance.

The pixel hit resolution also exhibits a slow degradation with integrated luminosity as shown in Fig. 15.4. The two points of improvement correspond to re-calibrations of the pixel threshold.



**Figure 15.3:** (a) Average pixel threshold in units of 1 ke for the barrel layers and forward disks, and (b) average analog current per ROC drawn by the power supply for BPix layers 1 and 2, as a function of the integrated luminosity delivered in Run 1 [11].



**Figure 15.4:** Single hit resolution for barrel layer 2 in the  $r\phi$  plane as a function of the integrated luminosity delivered in Run 1 [11].

## 3748 15.2 Optimization for LHC Run 2

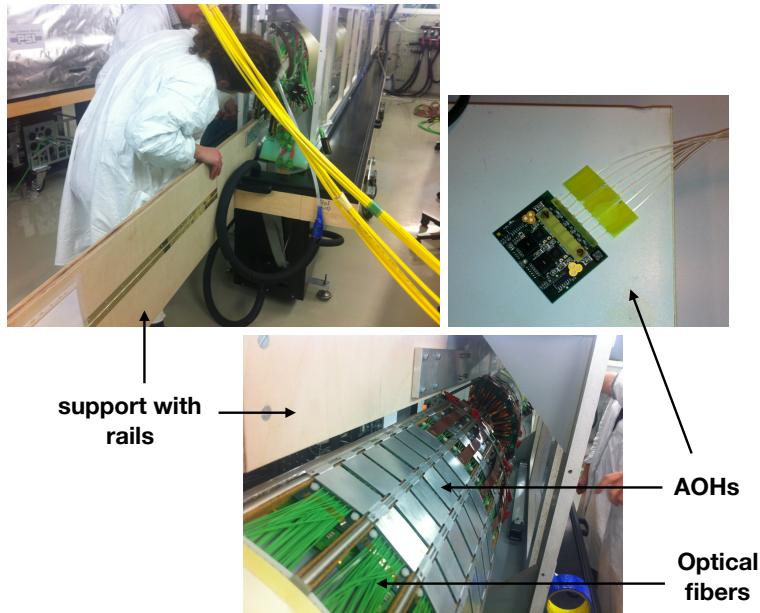
3749 In Summer 2013, after the first LHC run, the BPix and FPix detectors were extracted  
 3750 from CMS, and throughout LS1 they were kept in a refrigerated, climate-controlled room  
 3751 environment (Fig. 15.5) located at the CMS experimental site, LHC-P5. The BPix was  
 3752 maintained in two cold boxes in a laboratory with repair workbenches, and all the electronics  
 3753 and computers necessary to control and readout the detector for maintenance and tests.



**Figure 15.5:** Barrel pixel detector temporarily installed in the clean room at LHC P5.

3754 At the end of Run 1, the fraction of operational channels in the barrel pixel detector was  
 3755 97.7% and the long shut-down was used to recover the faulty channels. The main reasons

were broken wire-bond connections between the ROC and the HDI as well as issues with the lasers on few AOHs. Furthermore, some modules had an old ROC design, which caused operational problems, and were therefore disabled. Replacements were attempted only for the barrel layer 3 outer shell, which made up 52% of the faulty channels, since the other layers and the inner shell of layer 3 were considered too risky to touch without breaking further parts. Two AOHs were found with disconnected wire bonds between the laser and the AOH PCB, and they were also replaced. Figure 15.6 shows pictures from the laboratory in LHC-P5 during this operation. In order to proceed with the replacement one of the two cold boxes was opened and the half shell of interest extracted using a support equipped with rails. The shields covering the AOHs were unscrewed and all the AOHs of the sector in the outside direction had to be unplugged in order to replace the two malfunctioning ones. Before restoring the detector in its original position inside the box, the two new AOHs were tested by checking with the oscilloscope the variations in the optical output when changing the laser bias settings with commands sent through the tkFEC. The same tests were performed for the other functioning AOHs that had to be unplugged to perform the replacement. It was found that during the operation, two additional AOHs were damaged and they had to be replaced as well.



**Figure 15.6:** Pictures of the operations conducted in the clean room at LHC-P5 to replace the broken AOHs. The support with rails used to extract one half shell from the box is visible. The AOHs are mounted on the supply tube and covered by metal shields. A picture of an AOH is shown on the top right.

There was, however, a serious incident in mid-August 2014. After having replaced a BPix module, tests of the corresponding quadrant showed severe damage: 55 new unresponsive modules were found. It was decided to take that part of the detector to PSI for further tests and repairs. Shorts were discovered at the ROCs and in several modules between the TBM and cable pads. Eventually, the detector was repaired using 40 new modules and 19 repaired ones within three months. The shorts were suspected to be caused by humidity due to unobserved condensation in the cold box. After being repaired, the functionalities of the new modules were successfully confirmed and at the end of LS1 the good detector fraction was 99%.

3782 Part of the time available during LS1 have been employed to exercise and improve  
3783 calibration procedures in view of commissioning and operations for Run 2. An overview of  
3784 the calibration procedure is given in the following.

3785 **15.3 Overview of pixel calibrations**

3786 Detector functioning and performance depend on proper calibrations of readout chain para-  
3787 meters. Most of these parameters are quite stable unless major changes occur, such as  
3788 the detector operating temperature. Other parameters are more sensitive to environmental  
3789 variations. For these parameters a re-calibration on a regular basis was necessary during  
3790 Run 1 operations.

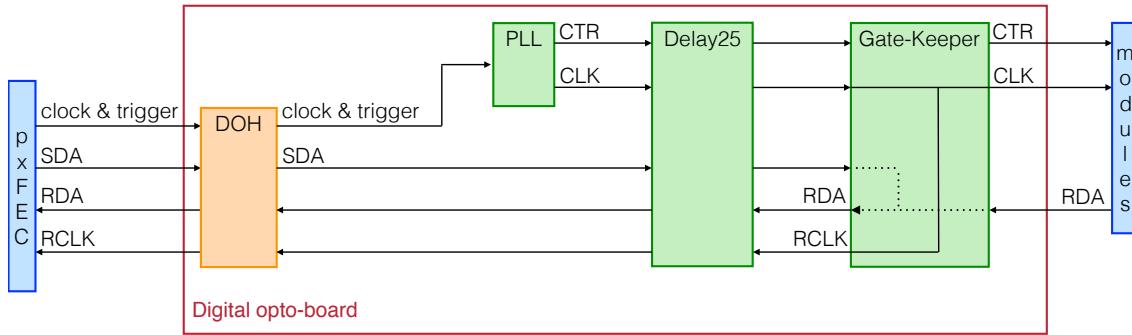
3791 Further expertise in the calibration procedure was achieved during LS1 and used for the  
3792 re-commissioning of the detector to prepare it for a successful data-taking in 2015–2016. In  
3793 addition, the detector was fully re-calibrated at low temperature after re-installation. As for  
3794 Run 1, in these two years, re-calibrations have been performed during technical stops, and in  
3795 particular in mid-2016 when the analog current drawn by the ROCs of the innermost layer  
3796 reached critical values ( $\approx 6\text{ A}$ ) that led to the trip of the power supplies in several occasions.

3797 The calibrations are performed with POS which was as well installed and run on the  
3798 computers available in the clean room. As described in the following, there are a large number  
3799 of different calibration tasks that need to be executed sequentially and sometimes iterated.  
3800 The calibration process consists first of a part where the readout chain settings are adjusted.  
3801 It is meant to put the detector in a state in which it can correctly reconstruct hits and involves  
3802 tuning of the settings of the FED, of the electronic components placed on the supply tubes, as  
3803 well as the threshold and timing settings of the ROCs, which are controlled by programmable  
3804 DACs (Section 14.2.2). In the second part of the process the signal performance are optimized.  
3805 The steps involved here are lengthy and require several iterations to reach the target signal  
3806 rise speed as well as the lowest practical value for the threshold of each ROC. In the final step,  
3807 an optimization of the analog pulse height is performed. Most of the calibrations produce  
3808 directly new optimal settings which can then be used in subsequent runs. Other calibrations  
3809 write binary data files which have to be analyzed offline, these include the pixel alive test,  
3810 the threshold and noise measurement and the gain calibration.

3811 **15.3.1 Adjustment of readout chain settings**

3812 **1) Delay25 chip**

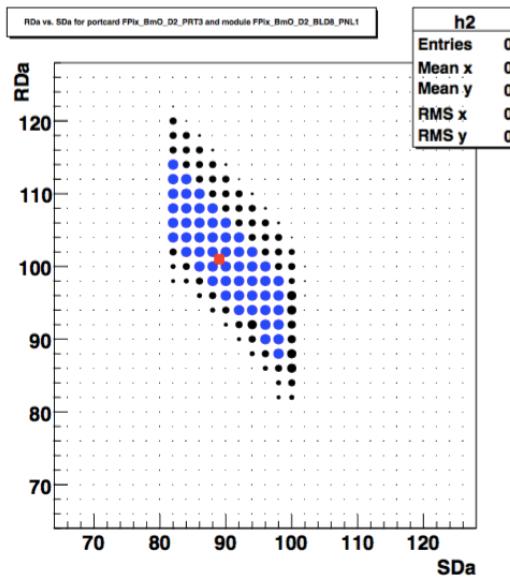
3813 As described in Section 14.3 (Fig. 14.6), the LHC clock, L1 trigger and programming signals  
3814 are transmitted from the pxFEC placed in the underground counting room to the detector  
3815 through fibers. The optical signals are first converted into digital signals by the DOH to be  
3816 then sent to the detector through the Delay25, PLL and Gate-Keeper chips integrated in the  
3817 same digital circuit. The clock and trigger are encoded as a single signal transmitted using  
3818 one single fiber. As schematically illustrated in Fig. 15.7, this signal is decoded by the PLL  
3819 chip and sent via two separate lines, LHC clock (CLK) and Calibrate/Trigger/Reset (CTR),  
3820 through the Delay25 chip to the BPix modules. In addition, the CLK signal is split in the  
3821 Gate-Keeper chip and one line (RCK) is returned and sent back to the pxFEC through the  
3822 Delay25 chip. The digital programming and control data (SDA) also goes through the Delay25  
3823 and Gate-Keeper chips. If the gate is open the SDA is transmitted to the BPix modules  
3824 which sends the acknowledge signal (RDA) back, otherwise the data packet is returned in the  
3825 Gate-Keeper.



**Figure 15.7:** Diagram illustrating the functionality of the BPix digital circuit.

3826     The SDA signal can only be decoded by the TBM if it is in phase with the CLK signal.  
 3827     The purpose of the Delay25 chip is to adjust the timing between the two lines to make this  
 3828     communication work. Hence, a calibration is performed where the delays for the SDA and  
 3829     RDA lines are scanned and for each set of values commands are sent to the TBM and the  
 3830     return status in the pxFEC is checked. The main output from this calibration is new SDA and  
 3831     RDA delay settings. If the calibration converges the old settings stored in the configuration  
 3832     database are updated with the new ones. An example of the scan is shown in Fig. 15.8 for  
 3833     one module. The set of values chosen by the algorithm is indicated with a red point and  
 3834     corresponds to a region where the communication between the TBM and the pxFEC has  
 3835     been established for each trial.

3836     This calibration is fundamental to ensure correct communication with the pxFEC, but  
 3837     once the settings are found they do not need to be readjusted often.



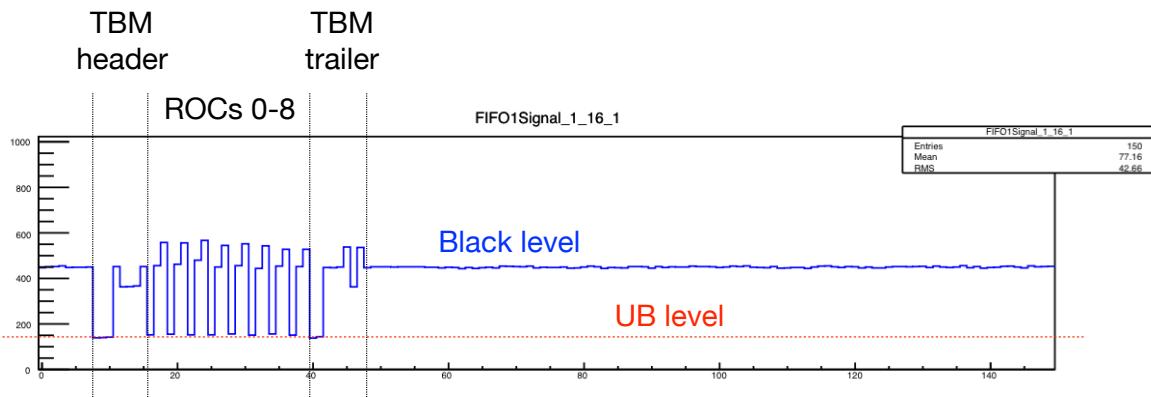
**Figure 15.8:** Example of output of the Delay25 calibration for one module. For each set of SDA and RDA delay settings the communication with the TBM is checked. The blue dots indicates areas with 100% communication efficiency. The black dots indicated partial efficiency where larger dots have higher efficiency. The red square indicates the point chosen by the algorithm.

## 3838 2) FED receiver offset

3839 This calibration adjusts the individual offsets included in each input channel of the FED such  
 3840 that the baseline of the analog signal (black level) is tuned to be near a given target value,  
 3841 normally 450 ADC counts, which is near the midpoint of the dynamic range of the ADC. The  
 3842 main output consists of new FED parameters that, if satisfactory, are used to update the  
 3843 previous settings. This calibration is performed often because the AOH is very temperature  
 3844 sensitive. Already 1 °C temperature change shifts the signal by 50 ADC counts (out of 1024).  
 3845 The pixel FED automatically corrects for baseline shifts during a run but it is important to  
 3846 start with a uniform baseline distribution. The baseline calibration adjusts each optical input  
 3847 to be  $\pm 5$  ADC counts from the target value. During normal LHC operations it is performed  
 3848 at least once a day during the LHC fills.

3849 The calibration also produces an output file with the analog signal for each module where  
 3850 its several components are visible, namely the TBM header and trailer, and each ROC header  
 3851 (Fig. 15.9). It runs quickly and provides an information on the data buffer for each FED  
 3852 channel. It is therefore very useful as a debugging tool since it provides a feedback on the  
 3853 basic functionalities of optical links, AOHs, TBMs and ROCs, needed to assess the status of  
 3854 the detector.

3855 If this step fails to converge, for instance when part of the analog signal is not visible,  
 3856 a calibration can be run that adjusts the timing of the signal digitization in the FED by  
 3857 changing the phase of the ADC clock. This calibration is usually very stable, and needs to  
 3858 be repeated only when the FEDs, fibers, or other parts of the detector are touched, or if  
 3859 modifications of the fine phase of the global clock occurs.



**Figure 15.9:** Example of analog signal from the TBM displayed at the end of the FED baseline calibration. The FED parameters are adjusted to center the baseline (or black level) in the middle of the FED ADC range.

## 3860 3) AOH bias and gain

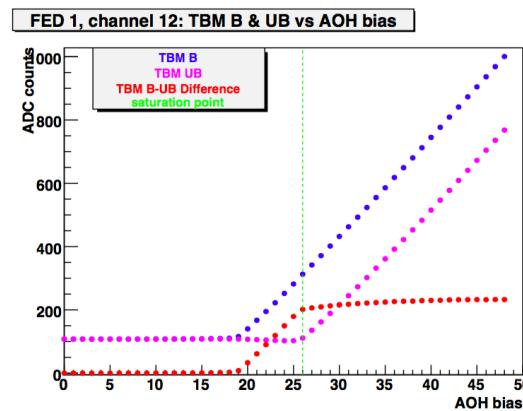
3861 Each AOH is equipped with 6 lasers for which the bias and the gain can be adjusted  
 3862 individually. The optical fibers connected to the lasers are combined in groups of 12 and each  
 3863 of them is connect to one FED channel. The AOH bias is a setting that controls the laser  
 3864 bias current and hence, the amount of light (optical power) sent to the FED. The optical  
 3865 power, and hence the ADC counts in the FED, increases with the laser bias setting. As

3866 shown in Fig. 15.10, at low values of the bias, the black (BL) and ultra black (UBL) levels  
 3867 are unaffected, so that there is no separation between the two. At some threshold, the BL  
 3868 begins to increase approximately linearly, followed by the UBL at a higher laser bias value  
 3869 and with about the same slope.

3870 The maximum BL-UBL separation depends on the TBM settings discussed in the next  
 3871 section, and it is low if these parameters are configured to low values. In fact, the BL is  
 3872 independent on these settings, whereas the linear rise of the UBL begins at a later point  
 3873 if the configured values in the TBM are higher. As a consequence, the BL-UBL difference  
 3874 saturates at a higher laser bias value. The goal of the AOH bias calibration is to determine  
 3875 a laser bias setting for each FED channel that is just high enough to saturate the BL-UBL  
 3876 difference. The calibration measures this difference, using the levels from the TBM header  
 3877 and trailer, as a function of the laser bias. It is important, though, that during this scan  
 3878 the TBM settings are set to reasonable values, at least as high as they will be set in later  
 3879 calibrations and physics runs. Otherwise, the laser bias value determined from the saturation  
 3880 point will be too low.

3881 Temperature variations alter the response of the AOH, essentially shifting the curves in  
 3882 Fig. 15.10 to the left or right by 4 bias counts for 5 °C variation. In order to provide a margin  
 3883 of error for these variations, the optimal laser bias setting is chosen by the calibration to be  
 3884 4 counts higher than the saturation value. This offset can be externally configured before  
 3885 running the calibration.

3886 It is also important that the laser bias is not too high to avoid that the signal moves  
 3887 out of the dynamic range of the FED. In the last part of the AOH bias calibration a coarse  
 3888 baseline adjustment is performed to bring the black level into the target range by re-adjusting  
 3889 the FED optical receiver offsets and laser bias settings. In this step the AOH bias is not  
 3890 decreased below the saturation value unless it is absolutely necessary. The main output of  
 3891 the calibration is a new configuration for the AOH bias and FED offset values that puts all  
 3892 FED baselines near the center of the dynamic range, with laser bias values that allow for  
 3893 a large BL-UBL separation. After the AOH bias calibration, the FED baseline calibration  
 3894 should be run to obtain a finer adjustment of the baseline (using the freedom to move each  
 3895 channel offset).



**Figure 15.10:** Black and ultra black levels as a function of the AOH laser bias. An optimal value for this parameter is found at the saturation value of the BL-UBL separation.

3896 The gain is a setting for each AOH laser that can accept just 4 possible values (0, 1, 2,  
 3897 3). This setting does not affect the black level, whereas it scales the size of deviations from  
 3898 the BL expanding or shrinking the signal. Larger settings correspond to larger deviations

and hence, to a larger BL-UBL separation. Although the adjustment of TBM and ROC parameters will be the primary method for tuning the UBL to the optimal value, the aim of this calibration is to set the laser gain at the lowest level that will allow the TBM UBL to be sufficiently low for an optimal readout of the signal. In fact, too high laser gain values will increase the power drawn and they are intended to be used to compensate for radiation damage over time. The optimal laser gain setting is chosen as the lowest value that provides an UBL below a user-defined threshold.

#### 3906 4) TBM and ROC ultra black levels

With the black level set at about 450 ADC counts by the FED baseline calibration and automatic correction, the next step consists in a fine adjustment of the TBM and ROC ultra black levels. First, the TBM settings are calibrated to set the TBM header and trailer UBL to a value of about 250 ADC counts. There are three registers on the TBM affecting the UBL, where higher configured values correspond to lower UBL and hence, larger BL-UBL separation. Furthermore, two of them affect also the signal from the ROCs. A simultaneous scan of all three registers is usually performed. Although higher settings generally provide lower ultra black levels, at very high values the UBL may actually increase. Thus, if the calibration finds multiple settings that give the target UBL, it will choose the lower ones.

Dual TBMs represent a special case. The two channels on a dual TBM share the same registers, so that they cannot be adjusted independently to tune both ultra black levels at the target value. In this case, the settings are optimized such that one channel is at the target UBL, and the other is below.

A second calibration is run that sets the ultra black level for each ROC equal to the corresponding TBM's UBL. There are two DAC settings on the ROC which affect the UBL, and higher configured values correspond to lower UBL (and larger BL-UBL separation).

These calibrations have to be repeated every time the previous steps 2 and 3 are run and modify the correlated parameters.

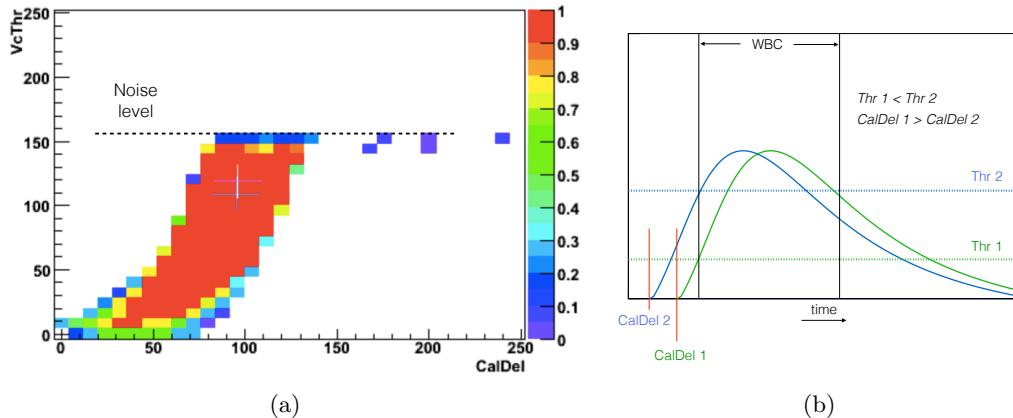
#### 3926 5) Threshold and charge injection delay

The rest of the calibrations require the use of the charge injection feature of the ROC. For the injected charge to be readout as a hit, it has to cross the comparator threshold and be validated by the trigger (which involves the timing of the injection). Thus, a calibration is first run that aims at finding the settings for each ROC for the comparator threshold and for the delay at which the charge is injected into the pixels. It is meant at quickly finding a working point in which the injected test charge can be read out. The amplitude of the injected signal is set by programming the corresponding DAC register ( $V_{cal}$ ). Since these settings are common to all pixels in a ROC, only few cells can be enabled for this calibration. A 2D scan of the threshold and delay settings is performed: for each pair of values, a defined number of triggers are sent and for each of them the event is readout from FIFO1 or FIFO3 to verify that the hits have been collected for each ROC. The settings are changed by programming the corresponding DAC registers,  $VcThr$  and  $CalDel$ .

It has to be mentioned that the  $CalDel$  setting is only relevant for calibration data taken with charge injection. For real data, only the trigger delay has to be known and programmed into the so called *WBC* register of each ROC. The trigger delay basically sets the bunch crossing in which data is read out and is estimated from the known cable/fiber lengths and delays introduced by the electronics.

3944 An example of  $VcThr$  versus  $CalDel$  scan is shown in Fig. 15.11(a). For large  $VcThr$   
 3945 values, which correspond to low thresholds, a large number of pixel fire due to noise such  
 3946 that to block a double column (cross-talk). For lower values, hits are collected in the  $CalDel$   
 3947 range that corresponds to the  $WBC$  used. The curve bends to smaller delay values as the  
 3948  $VcThr$  decreases. The explanation for this behavior is illustrated in Fig. 15.11(b). Since a  
 3949 low  $VcThr$  value corresponds to a higher threshold, the signal reaches the threshold later  
 3950 and hence a smaller delay is needed for the signal, i.e. the signal is injected earlier.

3951 At the end of the calibration an optimal set of values is chosen in the region where the  
 3952 efficiency for detecting a pixel hit is 100%. The optimal working point is also chosen such  
 3953 that it is sufficiently far away from the noise level.



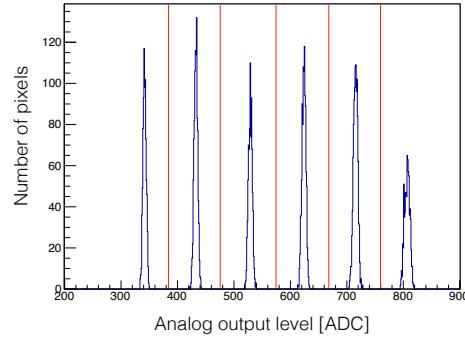
**Figure 15.11:** (a) Efficiency for detecting a hit as a function of the comparator threshold ( $VcThr$ ) and delay ( $CalDel$ ) settings for one ROC. Large values of  $VcThr$ , corresponding to a low threshold, generates much noise that saturates the digital circuit and no hits are detected. The optimal point is indicated in black while the blue point indicates the old settings. For small values of  $VcThr$  the signal reaches the threshold later and hence a smaller delay is needed for the signal. This behavior is illustrated in (b).

## 3954 6) Address levels

3955 The row and column address of the hit pixel is encoded in 6 discrete analog levels (Section  
 3956 14.3.1) which have to be well separated for being correctly decoded by the FED. The  
 3957 position of the address levels is determined by measuring the levels of all pixels in a ROC  
 3958 and overlaying them in a histogram. Pixels are scanned to make sure that combinations of  
 3959 address levels that could potentially cause problems are probed, such as transitions from  
 3960 high to low levels and vice versa. An example of the results is shown in Fig. 15.12, where the  
 3961 six peaks corresponding to the six address levels can be seen. The separation between the  
 3962 levels is good and the decoding limits are chosen in the center between to neighboring peaks  
 3963 to be then downloaded to the FED. The separation can mainly be affected by dirty optical  
 3964 connectors and poor light transmission, or by large temperature changes not compensated by  
 3965 the automatic baseline correction. Hence, during stable running conditions this calibration is  
 3966 run once every few days.

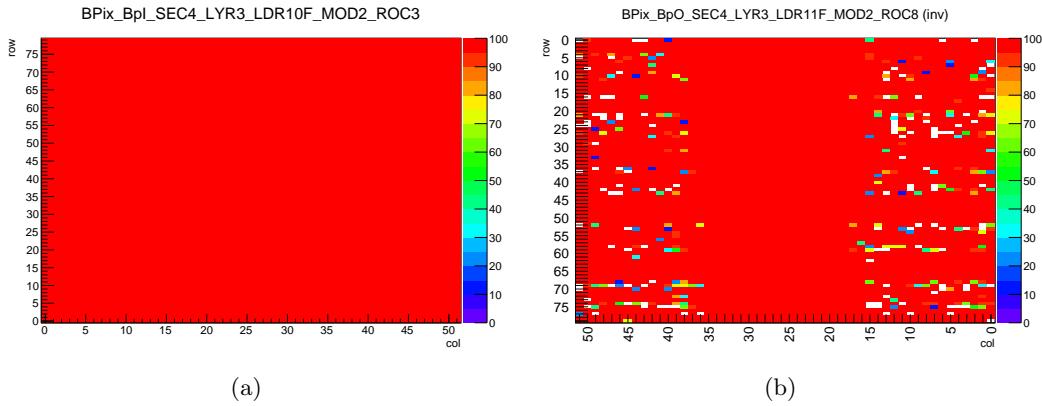
## 3967 7) Pixel alive test

3968 In this test, the functionality of each pixel in a ROC is checked by verifying that it responds  
 3969 to an injected calibration signal above threshold. Charge is injected in each pixel several



**Figure 15.12:** Address levels of all pixels in a ROC as received by the FED. The red lines are the separation limits used for the decoding of the pixel addresses in the FED.

times and the number of output signals is recorded. The pixel is fully working if all signals are registered; the pixel is defective, if no output signal is registered at all. The data are then analyzed offline to produce an efficiency map that displays the efficiency for each pixel. Examples of the results for two ROCs are shown in Fig. 15.13. For the case on the left all cells are functioning, whereas on the right an example is shown of a ROC with a large number faulty pixels.

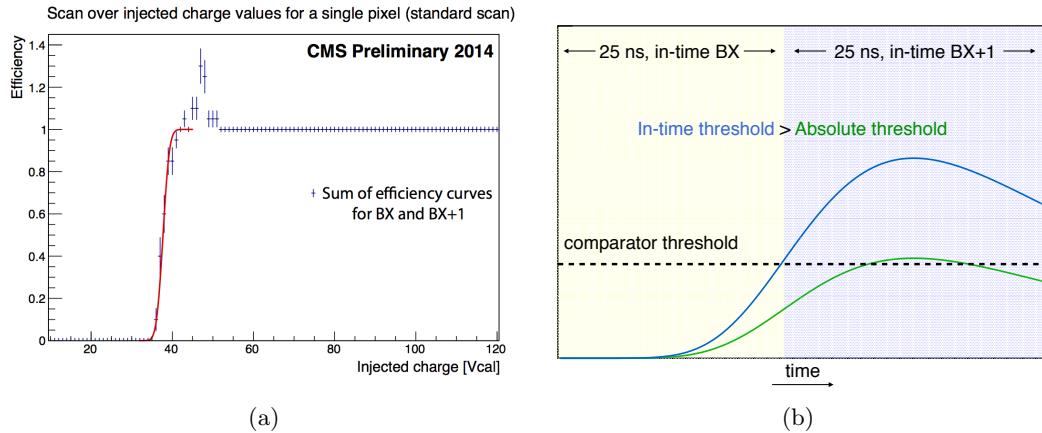


**Figure 15.13:** Examples of pixel alive test results for two ROCs: (a) all cells are functioning and (b) a large number of pixels are broken.

## 8) Measurement of threshold and noise

This is the last step of the calibration chain aimed at verifying and adjusting the basic functionalities of the detector. At this stage it is important to perform a measurement of the threshold and noise of each pixel, which will be afterwards optimized in the second part of the procedure as described in the next section. In fact, the detection thresholds are an important parameter of the pixel detector since they influence the hit position resolution (Fig. 15.4).

The thresholds are measured through the so called “S-curve” scan, which provides the pixel response efficiency as a function of the amplitude of the injected test charge ( $V_{cal}$ ), varied from 0 to its maximum. The  $V_{cal}$  value where the signal shows 50% efficiency is



**Figure 15.14:** (a) Single pixel efficiency curve determined performing a scan over injected signal amplitudes ( $V_{cal}$ ). The curve is the result of the sum of efficiency curves for the in-time bunch crossing (BX) and the following one (BX+1). Points exceeding the 100% efficiency are due to statistical fluctuations of the two curves in the turn-on region. The effect on the fit is negligible. (b) Diagram illustrating the difference between in-time and absolute thresholds due to the finite rise-time of the signals.

3985 defined as the threshold. As for the pixel alive test, the data are analyzed offline to produce  
 3986 the final results. An example of such scan is shown in Fig. 15.14(a) for a test conducted in  
 3987 the clean room at a temperature of -15 °C.

3988

3989 Due to the finite rise-time of the signals the threshold measurement is somewhat more  
 3990 complicated. One should distinguish the *absolute threshold* defined as the comparator level  
 3991 above which the signal is accepted and the pixel hit is available for readout, and the *in-time*  
 3992 *threshold* where the signal is fast enough to be correctly labeled by the right bunch crossing.  
 3993 The difference between the two is due to *time-walk* and is related to the speed of the pixel  
 3994 amplifier. The absolute threshold is relevant when discussing noise and cross-talk, that is the  
 3995 optimum conditions under which the ROC still works. It also determines the pixel detector  
 3996 hit occupancy. The in-time threshold determines the lowest amplitude signals useful for hit  
 3997 reconstruction and affects the position resolution. Both thresholds can be measured using  
 3998 the S-curve method: for the in-time measurement the *WBC* (or trigger delay) is set to the  
 3999 nominal value; for the absolute threshold measurement the *WBC* is shifted down by one unit  
 4000 making the readout of the lowest amplitude (i.e. slowest) signal possible. By definition the  
 4001 in-time threshold is higher than the absolute. This behavior is illustrated in Fig. 15.14(b).

4002

4003 The noise can also be measured with the S-curve method since it is proportional to the  
 4004 width of the region where the signal efficiency switches from 0 to 100%. Both noise and  
 4005 threshold are measured in  $V_{cal}$  units, representing the parameter which determines the  
 4006 magnitude of the injected charge. The calibration of the  $V_{cal}$  unit itself was done during  
 4007 module testing using data from X-ray sources of known energies, and it varies from pixel to  
 4008 pixel and from ROC to ROC. On the average, one  $V_{cal}$  unit corresponds to 65.5 electrons,  
 4009 representing the slope of the calibration curve, whereas the average offset is -414 electrons.  
 4010 However, the spreads of the two distributions are rather large, the slope parameter has an  
 4011 RMS of 9 and the RMS of the offset is about 570 [141].

4012 Running this method for the whole detector is very time consuming. Instead, for each ROC

4013 the thresholds and noise are measured using only 81 pixels, which was found to be sufficient to  
 4014 determine the average values. The results of the noise and threshold measurements performed  
 4015 in 2015 during commissioning for Run 2 will be discussed in Section 15.4.

### 4016 15.3.2 Optimization of signal performance

#### 4017 1) Signal rise speed

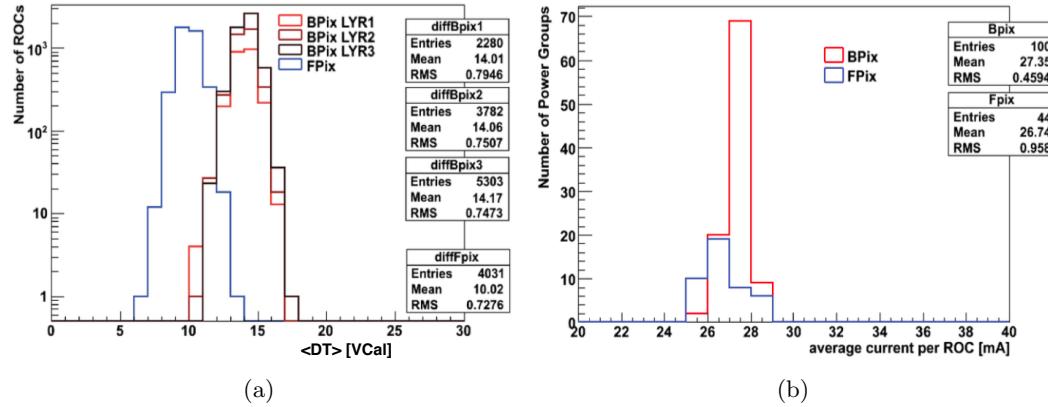
4018 The in-time threshold depends on the amount of time-walk introduced in the amplification  
 4019 and shaping that occur before the signal reaches the comparator. The speed of the pixel  
 4020 amplifier is controlled by  $V_{ana}$ , a 8-bit DAC that regulates the voltage applied to the analog  
 4021 part of the ROC, which can be varied in the range from 800 to 1,300 mV. The  $V_{ana}$  has to be  
 4022 optimized such that a compromise is obtained between the desire to minimize the time-walk  
 4023 and the need to keep the current drawn by the analog part of the ROC, or analog current,  
 4024 at a reasonable level. During module testing the optimal  $V_{ana}$  setting was determined for  
 4025 each BPix ROC by measuring directly the analog current drawn by the ROC, and then  
 4026 choosing the value that corresponds to 26–28 mA. In fact, this value for the current has been  
 4027 found optimal to avoid exceeding the limit of the power supply. Nevertheless, the radiation  
 4028 damage affects the ROC analog current and a re-calibration is necessary during operations  
 4029 (Fig. 15.3(b)).

4030 Once the detector is fully assembled, it is no longer possible to access the value of the  
 4031 analog current for each ROC, since at this stage the only available information is the total  
 4032 current drawn from a single power supply, which services more than one-hundred ROCs.  
 4033 Thus, a procedure has been developed in the past to optimize  $V_{ana}$  that does not make use of  
 4034 this information. The analog current can indeed be directly related to the time-walk, whose  
 4035 value DT can be obtained by the difference between the in-time and absolute thresholds. The  
 4036 higher  $V_{ana}$ , the faster is the detector (smaller DT), but also the higher the current drawn  
 4037 by the ROC. The target value of DT is then chosen such that the average analog current per  
 4038 ROC in each power group is near the optimal value of 26–28 mA. However, the correct target  
 4039 for DT depends on radiation damage and temperature, so that a fixed number to target  
 4040 cannot be given. Instead, one should tune the target based on the average analog current  
 4041 per ROC as read from the power supply. Figure 15.15 shows the DT distributions for both  
 4042 BPix and FPix measured in 2012, as well as the corresponding average analog current per  
 4043 ROC. For BPix, a target DT value of 14  $V_{cal}$  was found to be sufficient to reach the optimal  
 4044 current.

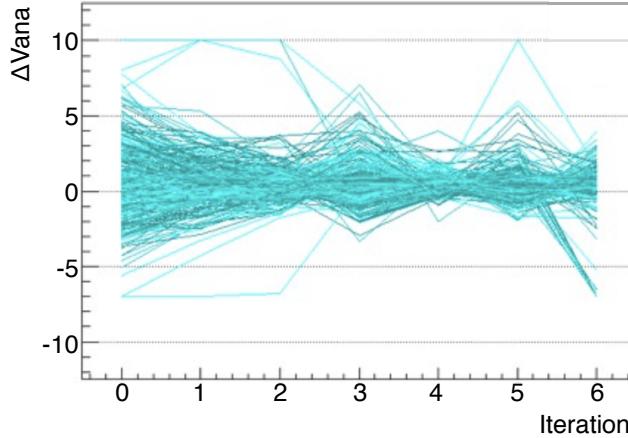
4045 In order to optimize the DT, the calibration is implemented as an iterative procedure,  
 4046 which makes use of the in-time and absolute threshold measurements given by the S-curve  
 4047 method. For each iteration, using the relation  $\Delta(DT) = DT - DT_{target} \simeq \Delta V_{ana}$ , the new  
 4048  $V_{ana}$  settings are computed to be downloaded to the ROCs in the next iteration. In each  
 4049 iteration, the absolute threshold and charge injection timing has to be re-calibrated (step  
 4050 5 in Section 15.3.1) because of their dependence from  $V_{ana}$ . Figure 15.16 illustrates the  
 4051 evolution of the  $V_{ana}$  settings with the iterations showing how these converge to the value  
 4052 corresponding to the target DT.

#### 4053 2) Threshold minimization

4054 This step is meant to set the threshold of each ROC at the lowest practical value, so that  
 4055 the threshold is low enough to detect low amplitude signals and ensure high hit resolution,  
 4056 but above the noise level at which cross-talk occurs. The procedure for minimizing the  
 4057 threshold starts setting a large value of the comparator threshold (for instance 50  $V_{cal}$ ) in



**Figure 15.15:** (a) Final distributions in DT obtained at the end of the optimization of the signal rise speed in 2012 for each barrel pixel layer and for FPix . The target DT value is chosen to reach an average analog current per ROC of 26–28 mA (b).

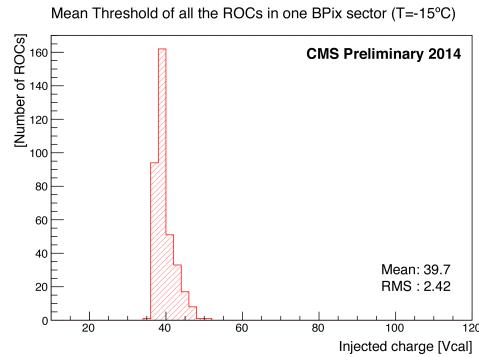


**Figure 15.16:** Example of optimization of the signal rise speed obtained from tests conducted in the clean room during LS1. The evolution of the  $V_{ana}$  settings for some ROCs with the iterations is shown and illustrates how these converge to the values corresponding to the target DT.

each ROC such that it is above the level of cross-talk. The threshold is then lowered by 2 units and a pixel alive test or S-curve is run to check whether a ROC is failing because the threshold is too low and cross-talk occur. The procedure is iterated until all ROCs reach the minimum achievable value. For each iteration the charge injection timing has to be re-optimized as well. Several scripts have been implemented during LS1 to automatize this time consuming procedure. An example of the results from tests conducted in the clean room during LS1 is shown in Fig. 15.17. The final threshold and noise distributions for the whole detector obtained with this method before the start of data-taking in 2015 are discussed in Section 15.4.

### 3) Analog signal response calibration

The final part of the calibration procedure is aimed at maximizing the range and linearity of the detector. In fact, the hit position is interpolated from the charge information of all



**Figure 15.17:** Distribution of the minimized thresholds for all the ROCs in one BPix sector. The measurement was performed at  $-15^{\circ}\text{C}$  coolant temperature in the clean room during LS1.

pixels in a cluster. For a precise position resolution it is therefore crucial to know for each pixel the exact response curve (or pixel gain) that converts the analog pulse height (in ADC counts) into the corresponding charge. The response curve is measured by injecting signals with increasing amplitudes to each pixel and measuring the analog pulse height. Before this calibration, the linearity of the response curve is optimized by adjusting few DAC registers in the ROC. The linearity is required for two reasons. On one hand the non linear behavior in the low range does not allow to reconstruct the charge of the signal, on the other hand fewer parameters have to be stored in the data base. The *VhldDel* register controls the delay that is applied to each pulse before its height is sampled and stored in the sample and hold capacitance until the readout mechanism is started from the periphery. The supply voltage of the sample and hold circuit is regulated by the *Vsf* register.

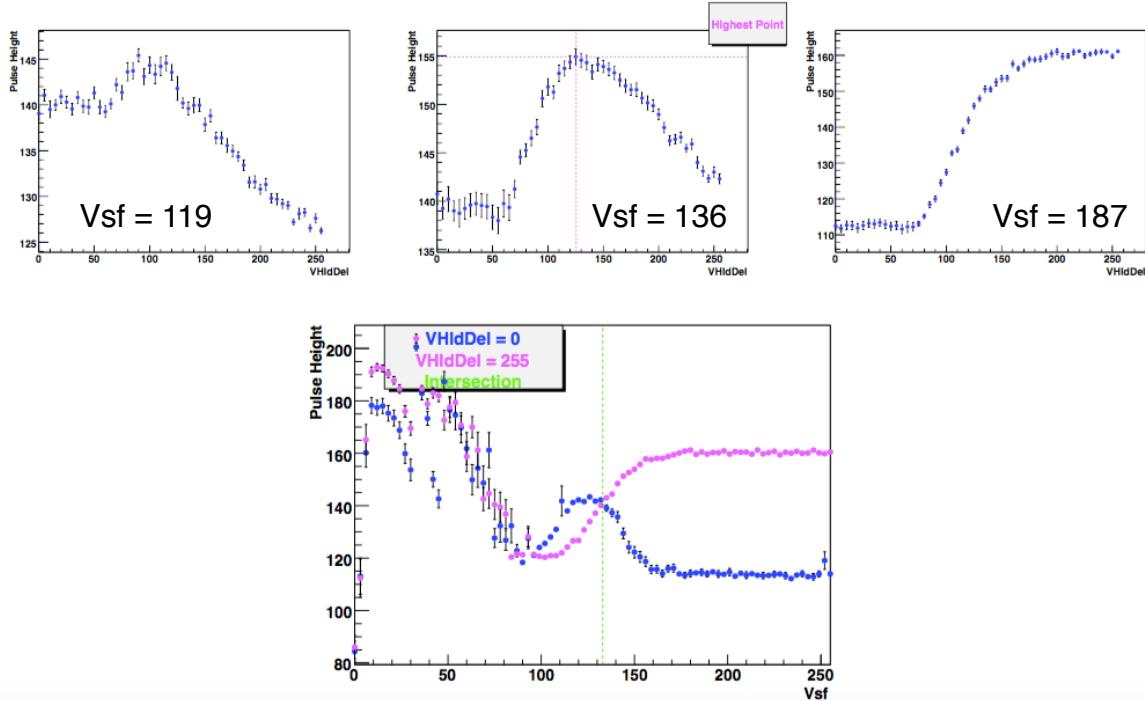
Figure 15.18 shows the pulse height as a function of *VhldDel* settings at low, medium, and high values of *Vsf*, for a fixed injected signal amplitude. A good *Vsf* value is one for which this curve rises and then falls so that the pulse heights at the two endpoints (lowest and highest *VhldDel*) are equal. The figure also includes a plot of these endpoints as a function of *Vsf*; the rightmost intersection point is the *Vsf* value chosen. Low values of *Vsf*, below  $\sim 90$  are discarded because found to be not optimal for a correct readout. After choosing the *Vsf* value, *VhldDel* is set to the value that maximizes the pulse height.

A number of ROC DAC settings also affect the scaling of the pulse height signal that is sent out to the FED. The difference in recorded pulse height between a small and large amounts of collected charge should be preferably large. However, the pulse height signal should not go low enough to be confused with the UB level, nor high enough to exceed the FED's dynamic range. Hence, a calibration is run to optimize these settings as well.

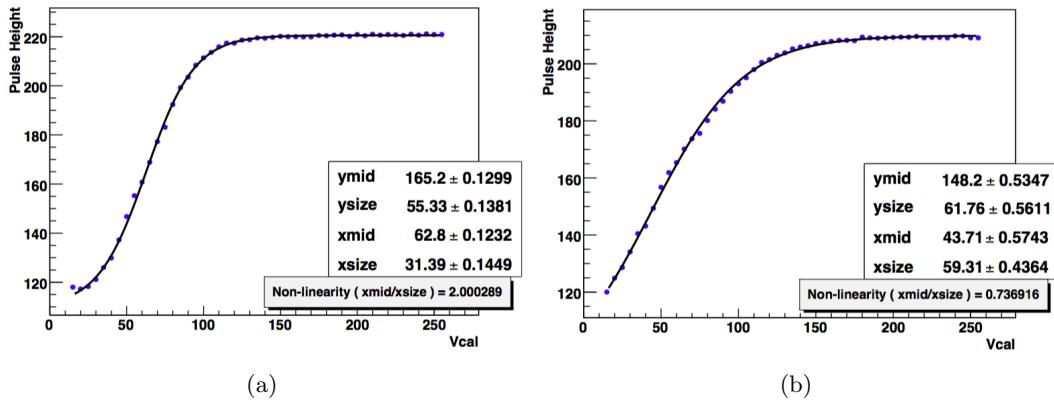
After these fine adjustments, the measurement of the response curve for each pixel is performed. For each pixel about 30 charge values are injected. During the scan, the acquired data is stored in binary files and is later analyzed offline. All pixels have to be calibrated, therefore, the procedure is time consuming and takes about 8 hours for the whole detector. An example of such measurement for one pixel is shown in Fig. 15.19. For comparison, an example exhibiting a non-linear behavior for a not optimal *Vsf* setting is also shown. The saturation in the high range is less important since it occurs for charges of more than 30-40 ke.

The pixel response curves are parametrized with the following function:

$$PH = f(Vcal) = y_{mid} + y_{size} \cdot \tanh\left(\frac{Vcal - x_{mid}}{x_{size}}\right), \quad (15.1)$$



**Figure 15.18:** Top row: pulse height as a function of  $VhldDel$  for low, medium, and high values of  $Vsf$ . As  $Vsf$  increases, the right endpoint shifts to the right. The best  $Vsf$  value is the one for which the pulse heights measured at the extremes of the  $VhldDel$  range are equal. Bottom plot: pulse height at the extremes, as a function of  $Vsf$ . Low values of  $Vsf$  are discarded because not optimal.



**Figure 15.19:** Examples of pixel response curve (gain calibration) representing the scan of the pulse height as a function of  $Vcal$ . The scan on the left presents poor linearity as performed for a not optimal value of  $Vsf$ .

where  $PH$  is the recorded pulse height,  $(x_{mid}, y_{mid})$  is the point at the center of the quasi-linear rise region of the hyperbolic tangent,  $x_{size}$  and  $y_{size}$  are the horizontal and vertical scales of the quasi-linear region, respectively. If  $x_{mid}/x_{size} \approx 1$ , the response curve is linear in the whole region of interest. Thus, the linear region below the saturation is parametrized by only the slope (gain) and offset (pedestal) of a linear fit. These parameters are then used in the data reconstruction. The results of the gain calibration performed for the whole detector

4108 before the start of data-taking in 2015 are discussed in Section 15.4.

## 4109 15.4 Re-commissioning for LHC Run 2

4110 The barrel pixel detector was installed back into CMS on 8th December 2014. The operations,  
4111 described in Section 15.4.1, were coordinated by the PSI and UZH teams (Fig. 15.20), and  
4112 were completed in only 5 days. After that, the FPix detector was also installed following a  
4113 similar check out procedure as for BPix so that most work was already completed before  
4114 Christmas. The full pixel detector was re-commissioned in January 2015 within about a  
4115 fortnight using the procedure described in the previous section. Section 15.4.2 presents the  
4116 results of the detector calibrations performed for the whole detector at low temperature after  
4117 the installation. Finally, in Section 15.5, the detector performance at the beginning of the  
4118 new LHC run are discussed.

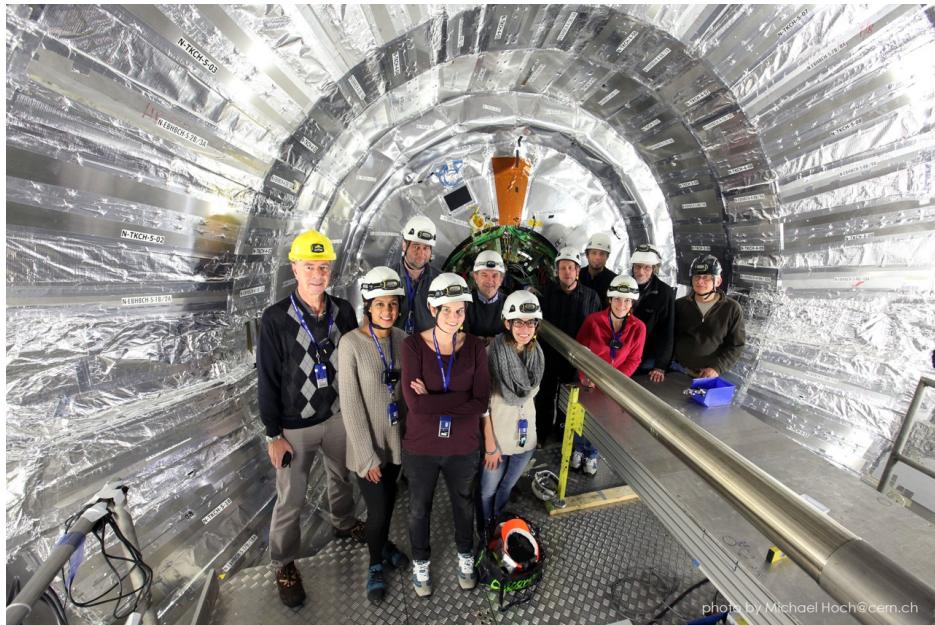


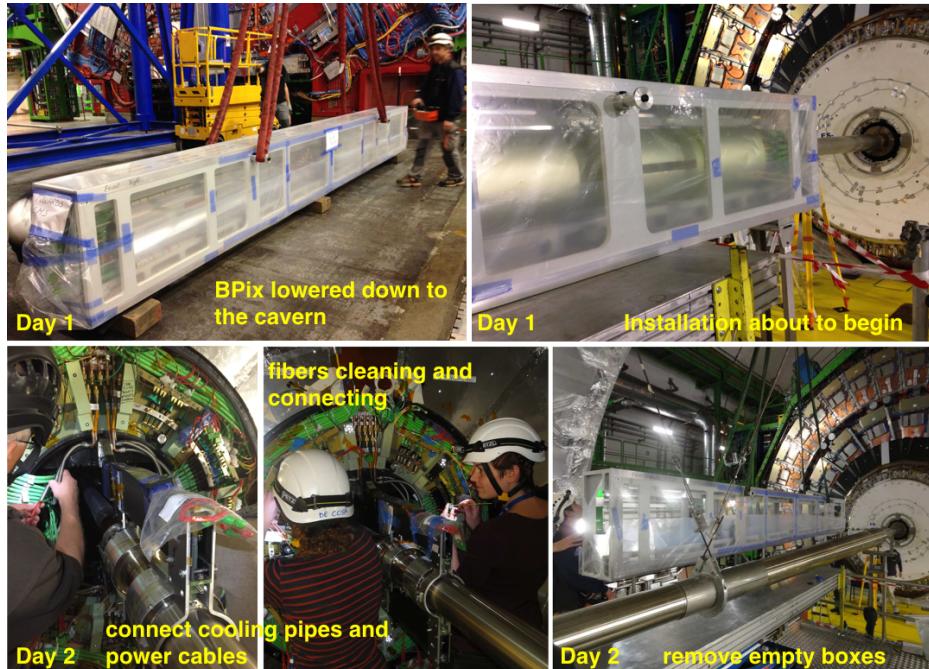
photo by Michael Hoch@cern.ch

**Figure 15.20:** Pictures taken on the CMS underground platform after the re-installation of the barrel pixel detector in December 2014. The operations were coordinated by the PSI and UZH teams.

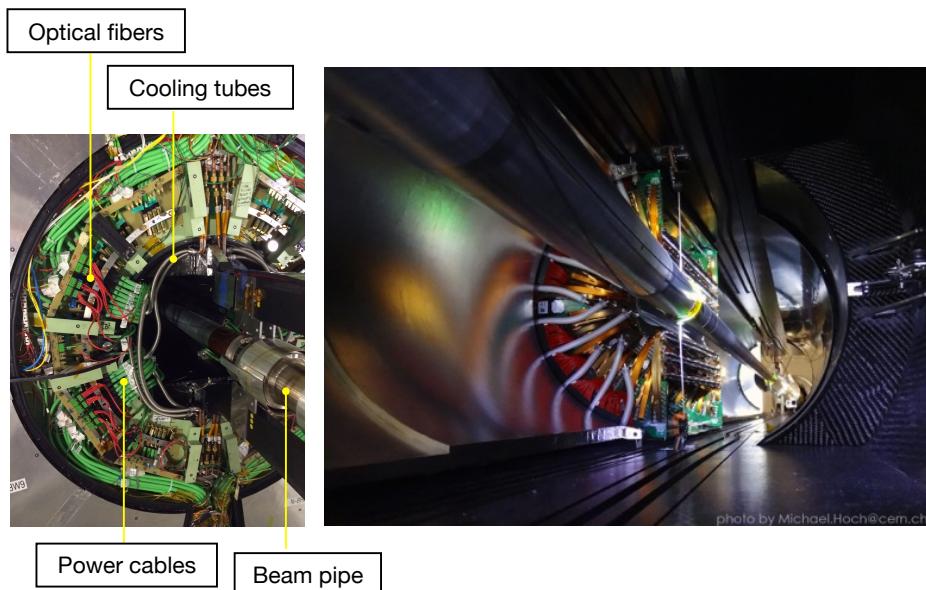
### 4119 15.4.1 Installation into CMS

4120 The barrel pixel detector re-installation into CMS took place only three months later than  
4121 originally planned due to the incident mentioned in the previous section. Particular care  
4122 was given to the centering of the detector with respect to the new beam pipe, required for  
4123 the upgraded pixel detector planned for Spring 2017 (Chapter 16), since before it had been  
4124 slightly shifted irradiating one side stronger than the other. Figure 15.21 shows pictures  
4125 taken on the CMS underground platform illustrating the operations conducted in the first  
4126 two days. The first day, each half of the detector was moved inside a transport box from the  
4127 clean room and lowered down to the cavern through the main shaft. A system with rails on  
4128 top and bottom inside CMS had been designed to insert the pixel detector and the supply  
4129 tubes along the beam pipe. The transport box with the detector was lifted to the insertion  
4130 table and the rail system inside the box was joint with the rail system inside CMS using

4131 temporary extension rails. In this way, the detector could slide out of the transport box into  
 4132 its final position. The following day, all cooling loops, power cables and fibers were connected,  
 4133 and first attempts to power the detector and to test a sector were made. A picture of the  
 4134 detector in the final position with all power and control cables and optical fibers connected is  
 4135 shown in Fig. 15.22.



**Figure 15.21:** Pictures illustrating the steps of the BPix re-installation in December 2014. The operations have been completed in 2 days.



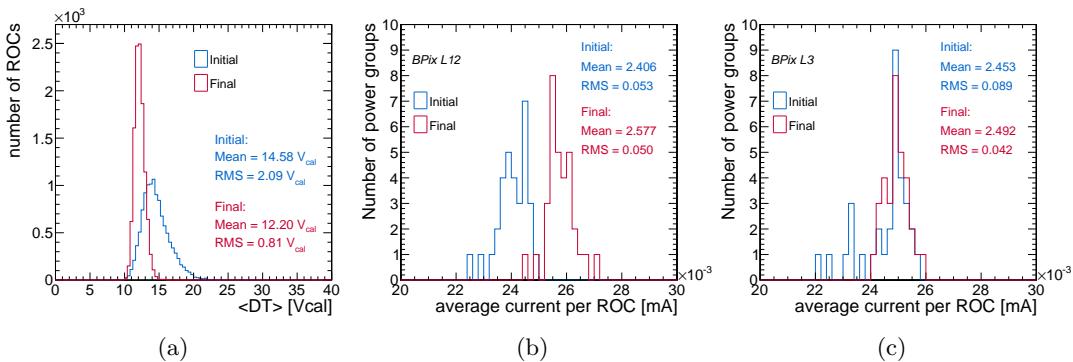
**Figure 15.22:** View of the barrel pixel detector in its final position inside CMS. The central beam pipe and the detector end-flanges with cooling lines and power and signal cables can be seen.

After the installation, the detector was then checked out at room temperature of about 16 °C. The basic set of calibrations was run from the CMS control room aimed at assessing the detector status and setting the basic operating parameters. These include calibrations of the Delay25 chip, FED baseline, AOH bias and gain, TBM and ROC UB, and address levels. The absence of a good quality TBM signal in the FED (Fig. 15.9) or bad address levels indicate poor optical connections. This kind of problems were immediately established and solved underground on the platform by re-cleaning the optical connectors with special tools. Few iterations were needed. These operations were completed in about 3 days establishing the functionality of the whole BPix detector. Only 1% dead or disabled channels were found and most of them were acknowledged during LS1. The check out procedure was repeated after the insertion of the FPix.

#### 15.4.2 Calibrations at -10 °C

As discussed at the beginning of this chapter, it was planned to operate the detector at -10 °C since low temperatures are favorable to mitigate the effects of radiation damage and guarantee excellent performance. Since the detector settings largely depend on the temperature a full calibration of the detector under the new conditions have to be performed. Because of the limited amount of time, it was not possible to achieve this during LS1, and tests were conducted for only few sectors aimed at verifying some basic functionalities at such low temperatures. The full calibration procedure was instead run in January 2015 and completed in only 8 days, including the final optimization of the signal performance (Section 15.3.2). The improvements added to the procedure during LS1 as well as the time spent in practicing it were crucial to make these time consuming operations much faster and smoother with respect to 2011 and 2012.

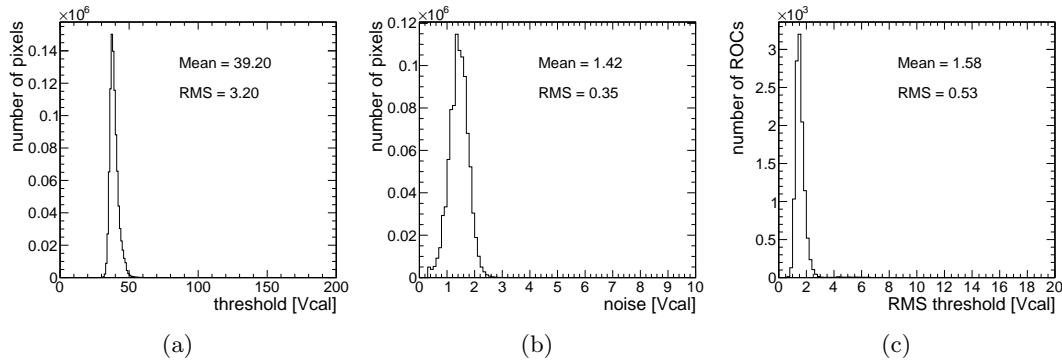
The results of the optimization of the signal rise speed are presented in Fig. 15.23, which shows the distributions in DT for all the ROCs at the beginning and at the end of the procedure. A value for the DT of 12  $V_{cal}$  was chosen as a target. The corresponding average current per ROC measured from the power supplies for each power group are also shown separately for layers 1 and 2 (Fig. 15.23(b)), and layer 3 ((Fig. 15.23(c))).



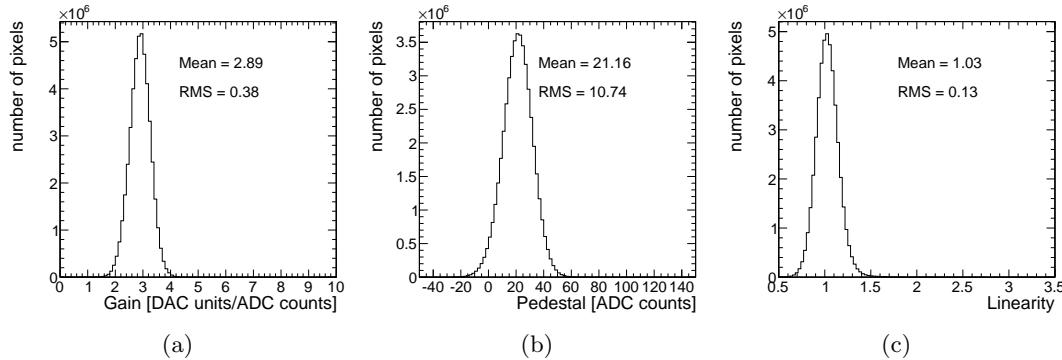
**Figure 15.23:** (a) Distributions in DT at the beginning and end of the optimization of the signal rise speed performed in 2015 for BPix Run 2 commissioning. (b-c) Corresponding average analog current per ROC after reaching the target DT value of 12  $V_{cal}$ .

Figure 15.24 shows the final threshold and noise distributions for all pixels obtained after the procedure of minimization described in Section 15.3.2. The spread of the thresholds in

4167 each ROC is also shown, quantified by the RMS of the individual ROC distributions. A  
 4168 final average threshold of  $\approx 40$  Vcal (2,200 electrons) was obtained showing agreement with  
 4169 the results of the tests performed in the clean room (Fig. 15.17) and with the Run 1 values  
 4170 (Fig. 15.3). Finally, the measured distributions of the gain and pedestal for each pixel used  
 4171 for the offline reconstruction of clusters are presented in Fig. 15.25. The distribution of the  
 4172 linearity parameter of the response curve as extracted from the fits is also shown.



**Figure 15.24:** (a) Threshold and (b) noise distribution for pixels after the calibrations performed in 2015 for Run 2 commissioning. (c) The RMS of the threshold distributions within single ROCs quantifying its spread among cells. All distributions are in units of Vcal (1 Vcal = 65.5 electrons).

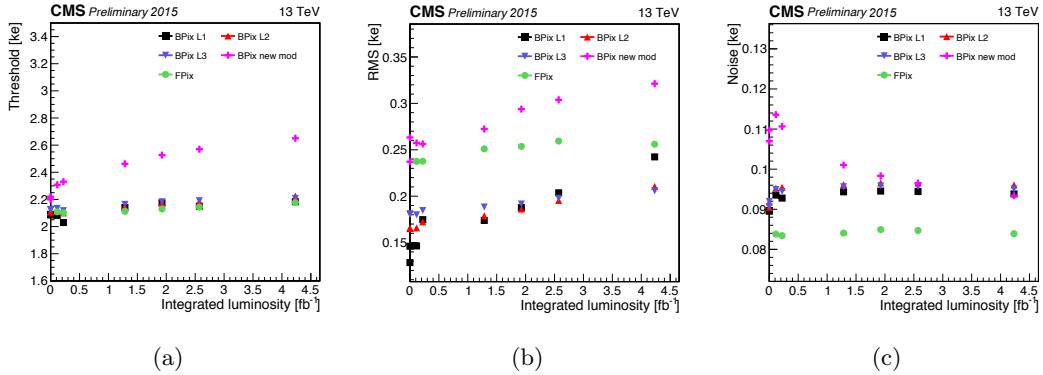


**Figure 15.25:** Gain (a) and pedestal (b) distributions extracted from the linear fits to the gain response curves of all pixels. These parameters are used for the offline reconstruction of clusters. (c) Distribution of the linearity parameter of the response curve as extracted from the fit.

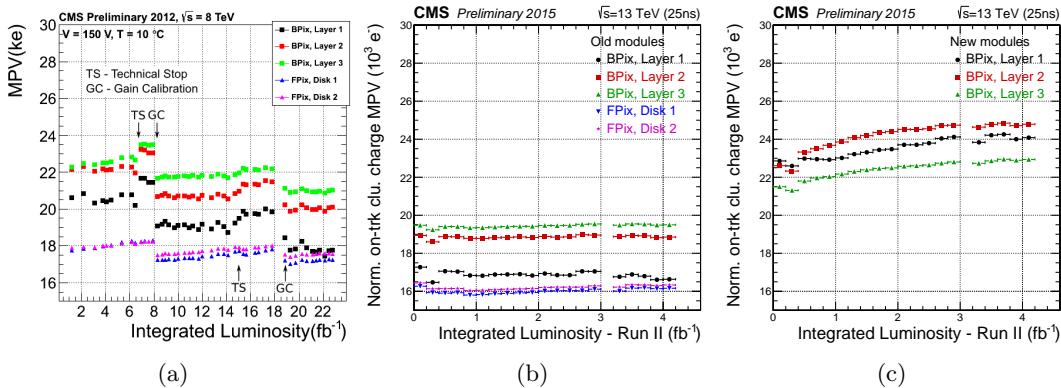
## 4173 15.5 Performance at the start of Run 2

4174 The detector re-calibration discussed in the previous section have been crucial to ensure  
 4175 excellent performance during the start-up of data-taking in 2015. Figure 15.26 shows the  
 4176 average threshold, RMS and noise in units of 1 ke for each barrel pixel layer as a function  
 4177 of the integrated luminosity delivered in 2015. Due to the different levels of irradiation, the  
 4178 old and new modules have been monitored separately. The threshold of the new modules  
 4179 rapidly increased with irradiation as was observed also in Run 1 (Fig. 15.3). The noise quickly

reached similar values as that of the old modules, which no longer experience such large changes due to irradiation.



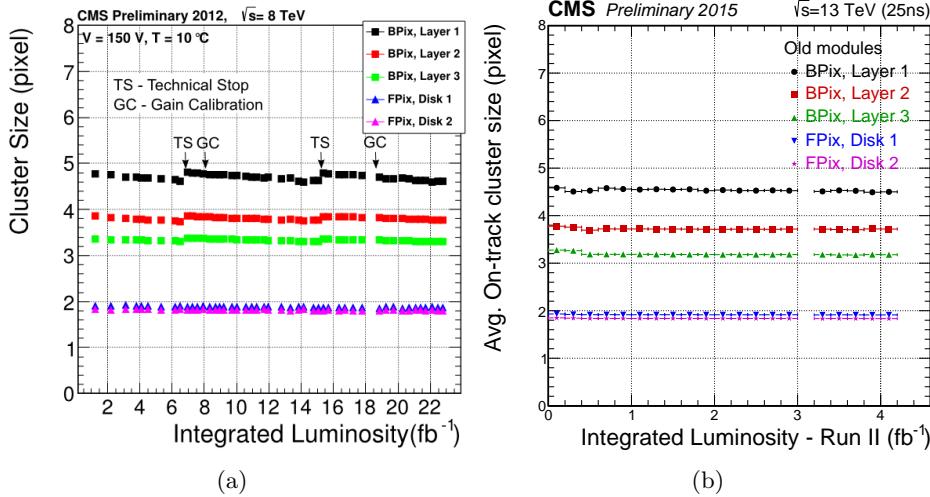
**Figure 15.26:** Average pixel thresholds (a), RMS (b), and noise (c) measured with charge injection, using the S-curve method. The BPix modules substituted during LS1 are considered separately [11].



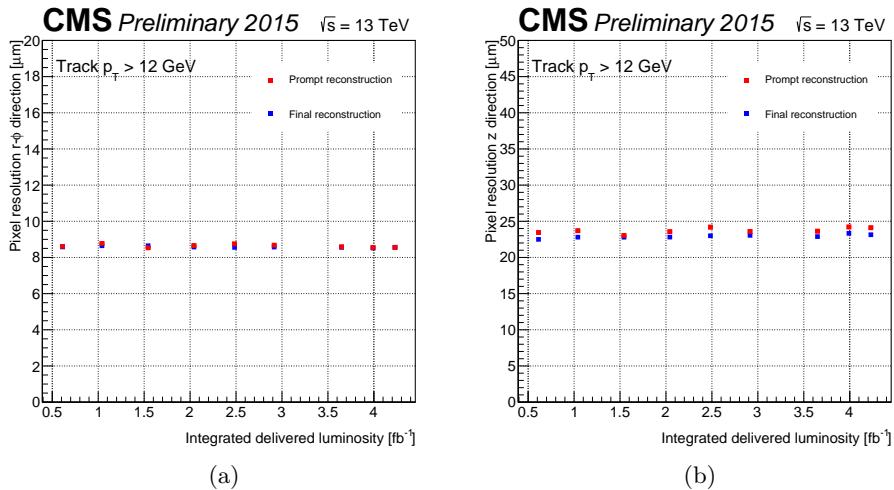
**Figure 15.27:** The MPV of the on-track cluster charge as a function of integrated luminosity (a) in 2012, and in 2015 separately for (b) old and (c) new modules [11].

Cluster properties like charge and size are important indicators of detector conditions and they have been monitored throughout the year by the pixel group. The cluster charge is determined by fitting the Landau distribution (Fig. 14.3) arising from the hits of tracks with  $p_T > 1 \text{ GeV}$  and extracting the MPV parameter. In the Run 1 measurements, the MPV changed significantly throughout the year and also after calibrations during technical stop periods (Fig. 15.27(a)). While the MPV of old modules did not change much in 2015 (Fig. 15.27(b)), the new modules showed a rapid increase (Fig. 15.27(c)). This behavior was also observed for old modules in Run 1 in the beginning of their lifetime. No significant change in the cluster size (Fig. 15.28) was observed.

Finally, the hit resolution have also been measured by the pixel group for layer 2 with tracks that have hits on layer 1 and layer 3. The tracks are then re-fitted without the hit in the middle, and the residual between the original and the interpolated hit positions are measured. The residual distribution is then fitted with a student-t function. Figure 15.29 shows the hit resolution as a function of the delivered luminosity in 2015. A large improvement was observed with respect to the measurements performed at the end of Run 1 (Fig. 15.4).



**Figure 15.28:** Average on-track cluster size as a function of integrated luminosity in (a) Run 1 and (b) Run 2. Both old and new modules showed very similar behavior [11].



**Figure 15.29:** Hit resolution of barrel pixel modules in (a) the  $r\phi$  and (b) the beam direction as a function of integrated luminosity in 2015 [11].

# Phase I upgrade of the CMS pixel barrel detector

4201 The present pixel detector will be replaced with a new pixel system in order to maintain the  
 4202 excellent tracking performance of CMS with the upcoming higher luminosity conditions at  
 4203 the LHC. This project is referred to as *Phase I pixel upgrade* and it was defined in 2012 [142].  
 4204 The new upgrade detector comprises four barrel layers and three forward disks to provide  
 4205 on average one more spatial point measurement per track compared to the present system,  
 4206 in the whole detector acceptance range. It also provides improved track impact parameter  
 4207 resolution reducing the radius of the innermost layer and increasing radial acceptance. Fur-  
 4208 ther improvement is obtained thanks to optimized engineering of the mechanical design and  
 4209 services of the detector, that provide a substantial reduction of the passive material in the  
 4210 tracking volume despite the addition of one barrel layer. Since the innermost sensitive layer is  
 4211 closer to the interaction point compared to the present detector, faster front-end electronics  
 4212 has been developed to operate with high hit efficiency and low dead-time. In this chapter,  
 4213 the main features of the new barrel pixel system are introduced.

4215 At the end of LS1 eight prototype Phase-1 pixel modules were installed in the CMS  
 4216 detector, on the third unpopulated forward disk. This so-called pilot system was commissioned  
 4217 and integrated into the central DAQ and control system, and took data in 2015–2016, with  
 4218 the aim of gaining operation experience under realistic conditions [143].

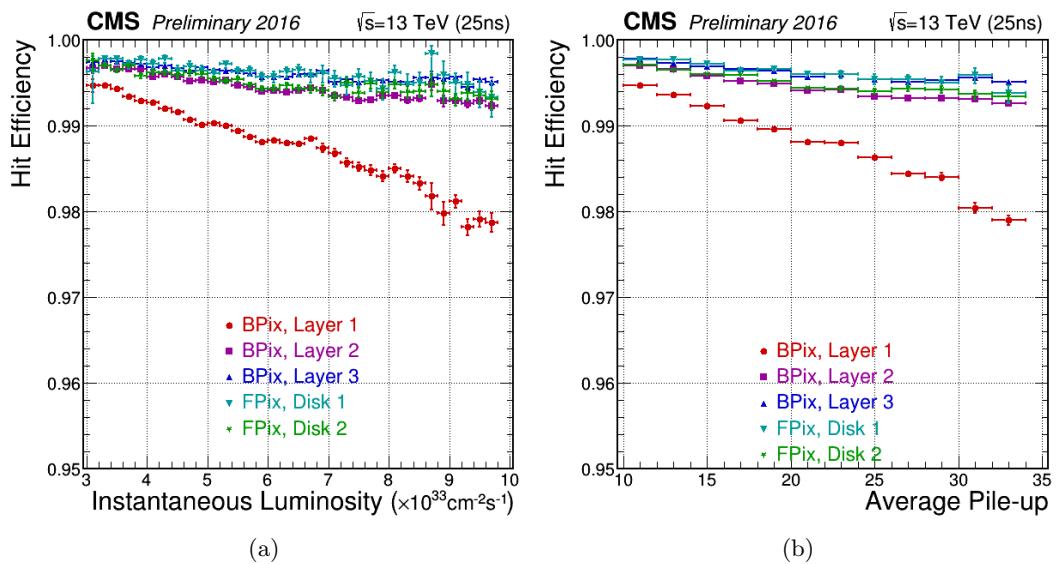
As for the present barrel pixel detector, the supply tubes have been assembled and tested  
 at the University of Zurich, while the modules have been mounted on the detector mechanical  
 structure at PSI. The integration of the supply tubes with the detector is currently ongoing  
 and the installation into CMS and commissioning of the complete system is planned for  
 March 2017.

Several procedures for testing the new system have been developed over the last three  
 years, thanks to a test stand assembled at the University of Zurich. The test stand, described  
 in this chapter, includes a slice of the CMS pixel data-acquisition system and all components  
 of the upgrade readout chain, together with a number of detector modules. It allowed for  
 detailed evaluation and verification of the components placed on the supply tubes before  
 their integration. I have contributed to the assembly of the test system and I implemented  
 some of its functionalities. Furthermore, I employed the system to test new calibration  
 procedures that I developed to be included in the main pixel online software. The aim of the  
 procedure is to guarantee a quick verification of the detector functionality during assembly  
 and commissioning. This work, detailed in the following, has been crucial to gain experience  
 with the new barrel pixel system and to acknowledge and implement several modifications to  
 the pixel software to be able to operate with the detector.

## 4236 16.1 Motivations

4237 The proposed upgrade of the CMS pixel detector aims at maintaining the excellent perform-  
 4238 ance of the present detector up to and beyond an instantaneous luminosity of  $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$   
 4239 and a pileup of 50. The limitations of the present detector for increasing luminosity and  
 4240 pileup can be seen in Fig. 16.1, which shows the hit efficiency for the various layers of the  
 4241 present pixel detector in collisions during 2016. The leading effect is a dynamic data loss  
 4242 in the readout chip which increases with instantaneous luminosity and trigger rate. This  
 4243 loss of data depends on both the occupancy and trigger rates and comes primarily from two  
 4244 sources, buffer size and readout speed. Between L1 triggers, pixel hits are stored in a finite  
 4245 sized buffer before being readout at the next L1 trigger. Therefore, if this buffer is full the  
 4246 ROC cannot record any more hits and subsequent hits are lost. Furthermore, when a L1  
 4247 triggers the readout, the involved double columns are blocked from having hits recorded and  
 4248 the buffer is cleared only after data have been sent. Thus, data can be lost if the readout is  
 4249 slow or the L1 trigger rate is very high. Simulation studies showed that for an instantaneous  
 4250 luminosity of  $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  and a bunch crossing time of 25 ns (50 ns), the expected  
 4251 dynamic inefficiency for the present pixel detector increases up to 15% (50%) for ROCs in  
 4252 the first barrel layer. As a consequence, the track reconstruction efficiency is affected. This  
 4253 can be seen in Fig. 16.2, which shows the track reconstruction efficiency for muons coming  
 4254 from the Z boson decay as a function of the number of primary vertices, as measured in 2016  
 4255 data with a T&P method. The efficiency is high and well described in the simulation, but  
 4256 slowly degrades as the number of pileup events increases. A new ROC for the upgrade pixel  
 4257 detector will largely reduce these effects.

4258



**Figure 16.1:** Hit efficiency for the various layers of the present pixel detector for 2016 collisions as a function of (a) the instantaneous luminosity and (b) the average number of inelastic pp collisions [11].

4259 Further effects contributing to inefficiencies in the track reconstruction arise from failures in  
 4260 the tracking algorithms for events with a large number of hits. In fact, with more interactions  
 4261 per crossing giving rise to additional hits in the tracking detectors, the pattern recognition  
 4262 becomes more challenging. Under these conditions, the CPU time required for tracking largely

increases in both the HLT and offline processing. In addition, keeping the same level of tracking efficiency results in a higher level of fake tracks; alternatively, the tracking can be tuned for lower fake rate at the expense of reduced efficiency. In order to keep both the CPU time and fake rate under control for luminosities of  $2 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , the tracking has to be tuned to have generally lower efficiency than at lower luminosities. This is obtained requiring hits in 3 pixel barrel layers. With an extra pixel layer negative effects of pileup can be partly mitigated.

4269

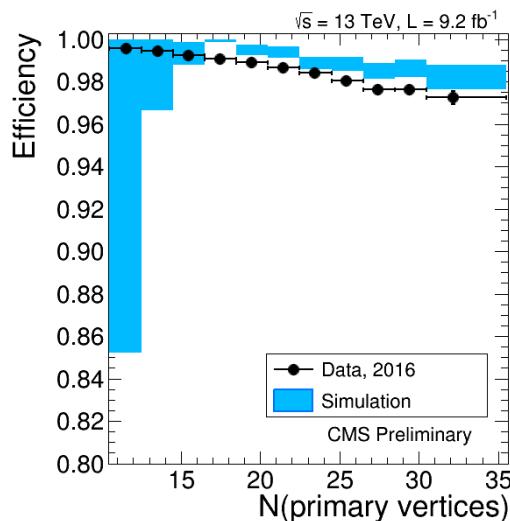
Degradation in the performance of the present detector are further due to radiation damage resulting in reduced charge collection and hence, in degradation of hit detection efficiency and resolution. Although the degradation can initially be mitigated mostly by increasing the voltage, and modification of the pixel cluster hit templates, eventually the reduced collected charge cannot be compensated. The hit efficiency is expected to be less affected but the reduced charge sharing and eventual breaking up of clusters will degrade the hit resolution. The upgrade pixel sensor would suffer similar radiation damage, however, such effects can be compensated by a much lower comparator threshold for the new readout chip. This improvement would largely mitigate the effects of reduced collected charge, so degradation in hit resolution should be much reduced comparing to the same radiation fluence.

4280

The passive material in the tracking volume is known to lead to tracking inefficiencies. In particular, a significant portion of material is present in the region near  $|\eta| = 1.5$  where the bulkhead with services from BPix meets the FPix. This material also contribute to additional challenges for track pattern recognition in a high pileup environment. The upgrade pixel detector, even with an extra layer features less passive material in the tracking volume, due to a new lightweight construction, cooling, and relocation of passive material out of the tracking region.

4281

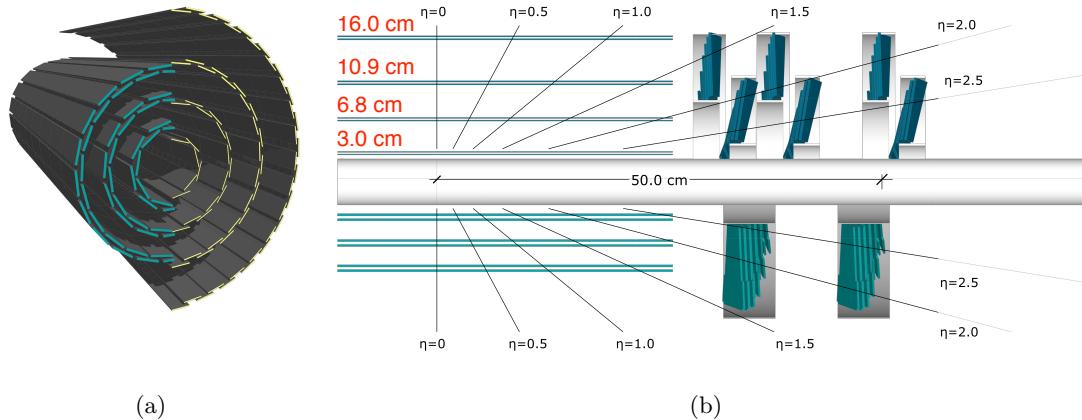
Details on the new detector layout and front-end electronics are given in the next chapter.



**Figure 16.2:** Track reconstruction efficiency for 2016 data and simulation for muons coming from the Z decay as a function of the number of primary vertices [48].

## 4289 16.2 Detector layout

4290 The proposed upgrade pixel detector consists of four barrel layers and three disks on either  
 4291 side of the interaction point. The layouts of the present and upgrade pixel systems are  
 4292 compared in Fig. 16.3. The barrel layers have a length of 548.8 mm and are placed at radii  
 4293 of 30, 68, 109, and 160 mm. Compared to the present BPix, there is one new layer at high  
 4294 radius. The radius of the innermost layer is reduced by 10 mm while layers 2 and 3 are almost  
 4295 unchanged.

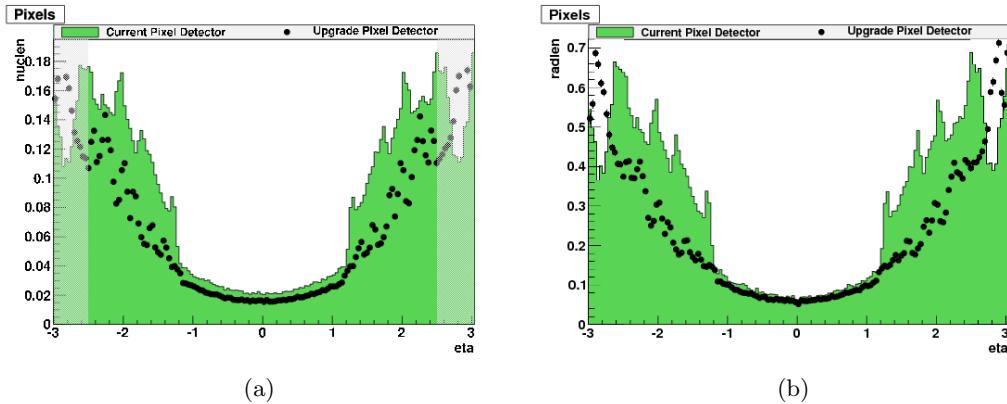


**Figure 16.3:** (a) Layout of the proposed upgrade pixel detector compared to the present detector in longitudinal view. (b) Three-dimensional view of the upgrade and present BPix detectors.

4296 The total number of BPix modules will increase to 1,184 compared to 768 modules in  
 4297 the present detector, with an increase in the number of pixel channels from 48 million to 79  
 4298 million. The modules are mounted on lightweight mechanical structures built from carbon  
 4299 fiber. The design and composition of the modules are similar in the whole pixel detector,  
 4300 except for the innermost layer where a considerable higher data rate is expected. Furthermore,  
 4301 half modules are no longer used to join the two halves, whereas a slightly more complex  
 4302 design of the mechanical support structure enables the use of full modules throughout. The  
 4303 pixel detector modules will be described in more details in the next section.

4304 The cooling pipes diameter is significantly reduced with respect to the present detector  
 4305 thanks to the usage of a two-phase CO<sub>2</sub> cooling system, which requires a much smaller mass  
 4306 flow than C<sub>6</sub>F<sub>14</sub>. This reduces substantially the amount of material in the tracking region.  
 4307 A further, significant reduction is achieved by moving the module connector area from the  
 4308 detector bulkheads to higher  $z$ , outside of the tracker acceptance, by using longer and more  
 4309 flexible module cables. As a replacement, micro twisted pair cables made of copper and with  
 4310 a diameter of only 127  $\mu\text{m}$  are used. Multiple twisted pairs are used to transmit the different  
 4311 signals, including clock, I<sup>2</sup>C, trigger, data, and so on. Power is transmitted in parallel through  
 4312 multiple copper clad aluminium wires with a diameter of 90  $\mu\text{m}$ . Signal and power cables  
 4313 are braided into a single strand. They are about 95 cm in length for all modules. Each wire  
 4314 of the strand is soldered onto a custom made board that fits into a commercial connector.  
 4315 The connector on the module side is soldered to the HDI. The obtained reduction in the  
 4316 material budget can be seen in Fig. 16.4, which shows a comparison of the radiation length  
 4317 and nuclear interaction length of the present and upgrade pixel detectors as a function of  $\eta$ .

4318 The overall layout of the system is unchanged. The detector barrel is complemented with



**Figure 16.4:** Material budget in the pixel detector shown in units of radiation length (a), and in units of nuclear interaction length (b) as a function of  $\eta$ ; this is given for the present (green histogram) and upgrade (black points) pixel detector. The shaded region at high  $\eta$  is outside the region for track reconstruction [142].

4319 four supply tubes on the  $+z$  and  $-z$  sides. The supply tubes carry electrical connections and  
 4320 cooling lines from the patch panels to the barrel bulkheads, and house auxiliary front-end  
 4321 electronics. However, the upgrade system has to fit in the same mechanical envelope as the  
 4322 present system and reuse existing services, power cables and optical fibers. This puts strong  
 4323 constraints on the design of the new system. In particular, higher bandwidth electronics is  
 4324 need. Since the upgrade detector has 1.9 times more channels than the present detector,  
 4325 the power consumption increases accordingly. The upgrade system uses DC-DC power  
 4326 converters [144] to supply the necessary current to the modules while reusing the existing  
 4327 infrastructure.

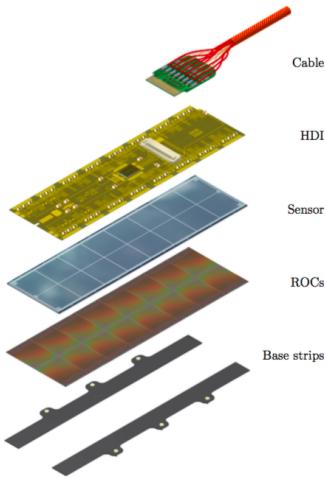
### 4328 16.3 Pixel modules

4329 The pixel modules for the upgrade are of similar design compared to the ones employed in  
 4330 the present detector. The main changes concern the design of the ROCs and the TBMs, as  
 4331 described in the following. Figure 16.5 shows a drawing of the pixel module employed for the  
 4332 outer barrel layers. The innermost barrel layer features a different ROC that allows to cope  
 4333 with even more extreme conditions at such small radii, while its modules differ mostly by  
 4334 the way they are mounted and by the cables used. From top to bottom, the figure shows  
 4335 the cables with a connector print, the HDI with the TBM mounted in the center, the silicon  
 4336 pixel sensor,  $2 \times 8$  ROCs and base strips for mounting.

4337 The sensor used in the upgrade is built with the same technology as the one used in the  
 4338 present detector. For the innermost layer, where the close proximity to the interaction point  
 4339 leads to the highest radiation damage, the sensor is expected to operate up to an integrated  
 4340 luminosity of  $250 \text{ fb}^{-1}$ . For this reason it is planned to exchange this layer once during the  
 4341 detector's expected lifetime of  $500 \text{ fb}^{-1}$ . The sensors in the rest of the detector can sustain  
 4342 for the entire duration because of the greater distance from the interaction point.

#### 4343 16.3.1 The digital ROC

4344 The ROC for the upgrade detector [145] is not a completely new development but rather an  
 4345 evolution of the well-proven ROC operating in CMS since its commissioning. It is designed



**Figure 16.5:** Exploded view of the new pixel module employed for the outer barrel layers of the upgrade BPix detector.

in the same 250 nm CMOS technology and the well understood core of its double-column architecture is mostly unaltered. However, to cope with the higher data bandwidth the readout protocol has been changed from a 40 MHz analog to a 160 Mbit/s digital readout. An ADC digitizes the analog pulse height information in the ROC periphery. The key additional elements are an 8-bit successive approximation current ADC running at 80 MHz with a programmable range, and a PLL which generates the 160 and 80 MHz clocks for the serial readout links and the ADC, respectively, from the 40 MHz LHC master clock. To reduce data losses, the number of hit buffer cells in each double column has been increased from 32 to 80 and the time stamp buffers have been increased from 12 to 24. To limit the increase of the area used by the buffers the layout has been entirely redesigned. An additional readout buffer stage has been introduced in the ROC periphery to reduce dead time during the column readout: the data is transferred (after being digitized) into the new readout buffer immediately after the trigger arrives so that the double columns are again operative. Improved performance of the analog amplifier and the discriminator in the pixel unit cell allow for operations at lower threshold, which is reduced from about 3,500 electrons in the present detector to under 2,000 electrons after the upgrade. This guarantees higher radiation tolerance and therefore, a longer lifetime of the detector.

The chip just described is suitable for the whole upgrade pixel detector except for the innermost barrel layer, where the data rates up to 600 MHz/cm<sup>2</sup> are expected, four times higher compared to the second layer. In order to cope with such extreme conditions the newly developed PROC600 [146] readout chip is used in the innermost layer. The new chip features a new 40 MHz dynamic cluster column drain mechanism based on dynamic cluster ( $2 \times 2$  pixels) finding in the double column.

### 16.3.2 The TBM and readout

Differently to the present detector, for the upgrade all barrel modules use at least two data channels in order to improve the bandwidth of the readout. Because of the limited number of fibers available for this purpose, two channels are always multiplexed into one data stream through a Data-Keeper multiplexer and encoder. For this purpose few modifications have been applied to the TBM. In particular, it now combines the digital 160 Mbit/s readout from

4375 the ROCs from two buses into a 320 Mbit/s signal to which it then applies 4-to-5 bit encoding.  
 4376 This results in a 400 Mbit/s data stream. The readout scheme is adapted to the different  
 4377 barrel layers (Fig. 16.6). Layers 3 and 4 employ a dual core TBM, referred to as TBM08,  
 4378 that passes two tokens simultaneously to achieve the parallel readout of two groups of 8  
 4379 ROCs, called Port 0 (or Channel  $\alpha$ ) and Port 1 (or Channel  $\beta$ ). The data are then combined  
 4380 into one data stream as described above, requiring one fiber. Except for the multiplexing  
 4381 step, this is very similar to the method used for the first two layers of the present detector  
 4382 (Section 14.2.3). Layer 2 employs a different TBM, called TBM09, which is the equivalent of  
 4383 two TBM08s (TBM A and TBM B in Fig. 16.6), each equipped with its own Data-Keeper.  
 4384 This TBM is capable of issuing four tokens simultaneously, so that the 16 ROCs are therefore  
 4385 divided into four groups that are readout in parallel. The two Data-Keeper then produce one  
 4386 400 Mbit/s data stream each and two fibers are required for the readout. For the innermost  
 4387 layer two identical TBM09 chips per module are employed and identified by two different  
 4388 HUB addresses. In this case eight tokens are passed in parallel on these modules, resulting in  
 4389 four 400 Mbit/s data streams and hence, four fibers for transmission.

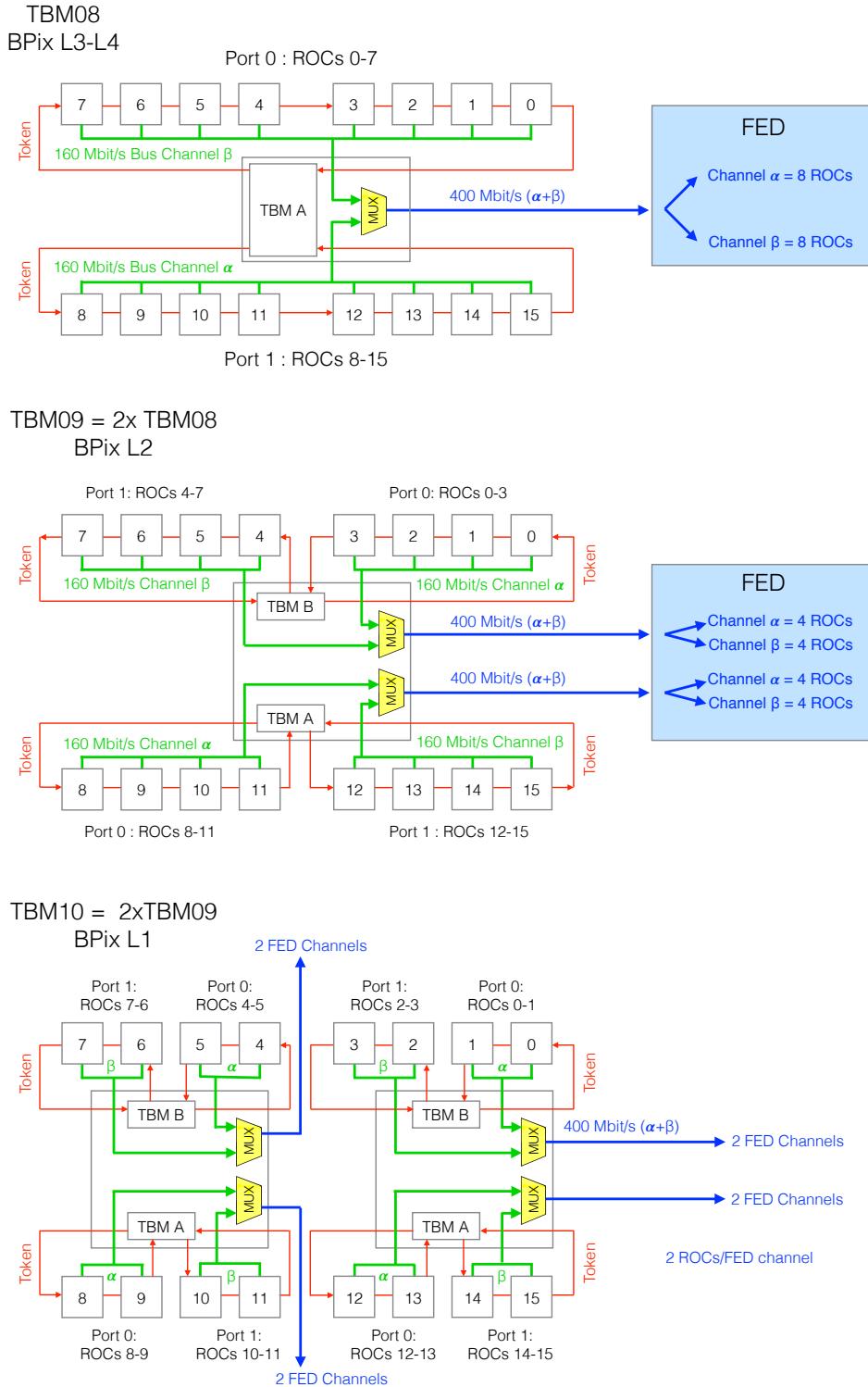
4390 An extensive set of control registers have been built into the TBM, which allow various  
 4391 functions and operating modes of the TBM to be controlled by issuing commands to the  
 4392 TBM through the communication control HUB. For the TBM09 the commands have to be  
 4393 issued to both TBM08s controlled by one unique HUB.

4394 As for the present detector, the module output signal is characterized by TBM header  
 4395 and trailer, ROC headers and pixel hit information, which are now encoded in binary data as  
 4396 shown in Fig. 16.7. A TBM readout begins by transmitting a twelve clock cycle (160 MHz)  
 4397 header sequence. The next sixteen clock cycles of the header are used to transmit the 8-bit  
 4398 event counter, 2 bits of error information, and 8 bits encoding the data contained in the  
 4399 last 8-bit TBM register accessed. Coincident with the next to last clock cycle, the token  
 4400 is transmitted to the ROCs. The TBM now goes into standby mode, waiting for the last  
 4401 ROC in the chain to return the token to the TBM. At this stage, the TBM transmits a  
 4402 twelve clock cycle trailer sequence. The next sixteen clock cycles of the trailer include 10  
 4403 bits with the values programmed in the TBM registers that control its mode of operations,  
 4404 and a 6-bit stack count value. The counting is used to monitor the timeout on the token  
 4405 returning. If the token fails to return, before the timer expires, the TBM will automatically  
 4406 issue a ROC reset, ending the token pass. The data contained in the ROCs are deleted, and  
 4407 error bits are returned in the TBM trailer 8 clock cycles later. The ROC data consist of 12  
 4408 bits for the header, 16 bits used for the pixel hit address and the final 8 bits for the pulse height.  
 4409

4410 In order to readout the new fully digital pixel system a VME-digital FED has been firstly  
 4411 designed. It is a hybrid solution featuring new daughter boards on the existing FED, and it  
 4412 has been used at the beginning of the operations with the pilot system and with test stands.  
 4413 This solution will be replaced by a  $\mu$ TCA system with high-speed signal links providing data  
 4414 rates up to 10 Gbits/sec. Since the results presented in this work are based on the VME  
 4415 system, only this is described in the following.

4416 The ADC daughter boards of the analog FED are not needed anymore for digital trans-  
 4417 mission. For the purpose of system developments, a special add-on board was produced,  
 4418 which is mounted on the current VME module for data readout, receives the 400 Mbit/s  
 4419 digitized data, and passes the data to the FPGAs in the same format as the present system.  
 4420 Thanks to this modular approach, the other parts of the FED did not require any hardware  
 4421 modification, allowing for a quick start of the tests with the new upgrade pixel system.

4422 As shown in Fig. 16.6, the signal from each fiber is split at the FED into two channels,  
 4423 whose content is buffered and processed in the FIFOs. Each channel will then correspond to



**Figure 16.6:** Readout scheme of the different TBMs used in the BPix layers.

4424 the data from half of the initial number of ROCs originally present in one fiber.

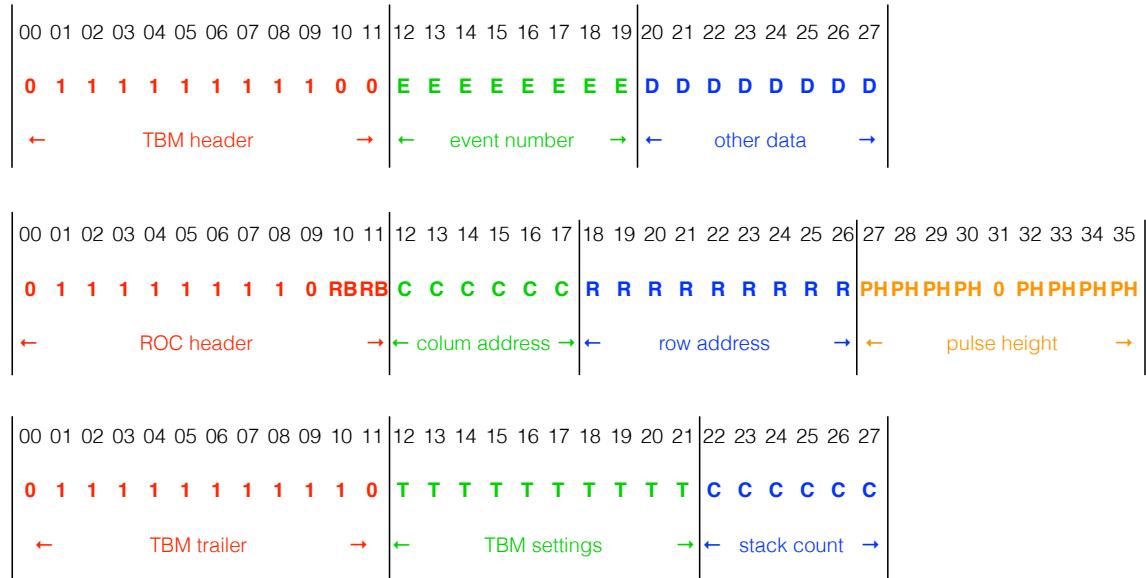


Figure 16.7: Decoding of the module digital output data.

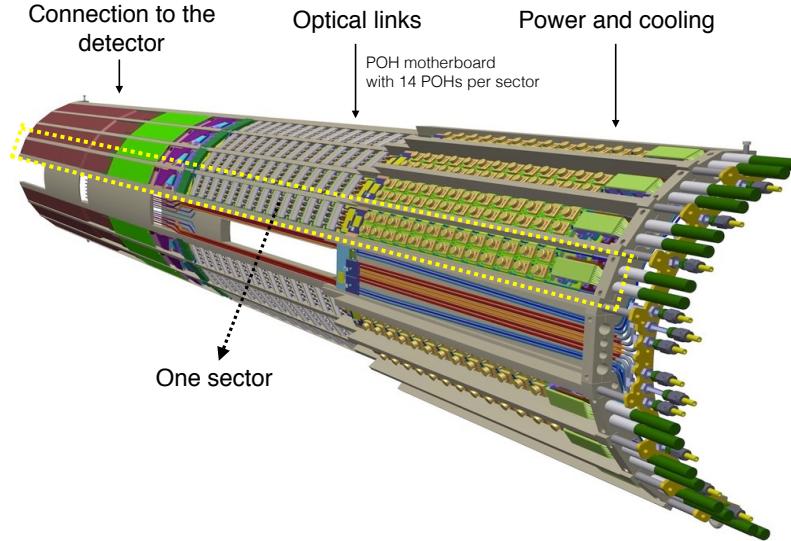
## 16.4 Supply tubes

As for the present detector, the power, readout and control circuits, as well as the cooling lines are housed by four half-cylinder supply tubes. Figure 16.8 shows the layout of one supply tube together with some of the new electronic components. The mechanical structure of the supply tubes is made from layers of carbon fiber composites. Each supply tube is divided in 8 sectors, which hold the electronics for one readout group of detector modules. Each sector includes DOHs as well as the auxiliary chips (PLL, Delay25, Gate-Keeper) for the transmission of control, clock and trigger signals. So-called pixel opto-hybrids (POHs) [147] are used for the transmission of the module readout data as a replacement of the AOIs used for the present detector. The change from analog to digital module readout in the upgrade system also requires the adoption of new optical links. POHs are built from four transmitter optical subassemblies (TOSA), linear laser-driver and level-translator chips and have been designed specifically for their use in the pixel upgrade system. All other components used in the control and readout chain are identical to the ones used in the present system. CCU chips are used for slow control, monitoring and timing distribution. Furthermore, pairs of DC-DC converters are mounted on the service cylinders. Each sector consists of a stack of boards, DC-DC converters, optical links and cooling loops, resulting in tight space constraints and a non-trivial assembly procedure.

The complete supply tube system has been integrated and tested sector by sector at the University of Zurich. **FIXME:** add nice picture from UZH

## 16.5 The test stand

In order to test the performance of the complete upgrade pixel system and gain experience in its operations, a test stand for BPix has been set up at the University of Zurich in 2013. The setup includes a slice of the full CMS pixel DAQ system together with prototypes of all components of the upgrade power system, control and readout chain as well as a number of



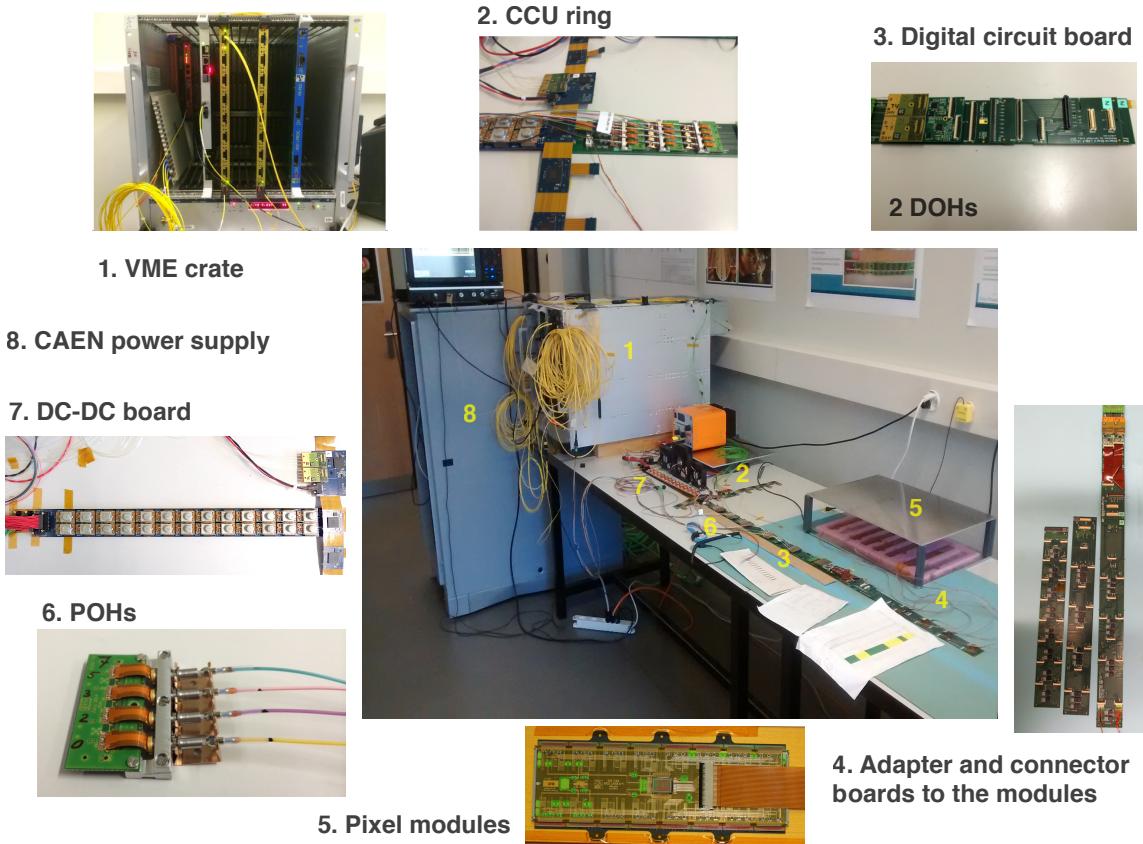
**Figure 16.8:** Layout of one of the four supply tubes for BPix equipped with the new electronic components. Each supply tube is divided into 8 sectors.

4450 detector modules. The system test was operated until November 2016 when the integration  
 4451 of the supply tubes started. Its main goals included:

- 4452 • test all components of the detector system prior to full production;
- 4453 • test all final components before and after integration on the supply tubes;
- 4454 • establish test and calibration procedures for the assembly and commissioning (Sec-  
4455 tion 16.6);
- 4456 • exercise the transition from the VME- to the  $\mu$ TCA-based DAQ system.

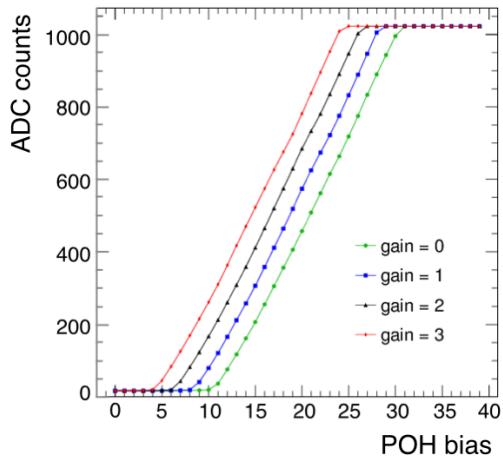
4457 Figure 16.9 shows a picture of the test stand, which consists of several sensor modules,  
 4458 electronics for their operation, a CAEN power supply and a VME back-end system to control  
 4459 and readout the modules. The test stand is also equipped with a linux PC connected to the  
 4460 VME and used to run and control the system through the installed XDAQ applications.

4461 Most of the tests that have been conducted with this system make use of a standalone  
 4462 software based on the socket technology provided by the PYTHON framework, which allows  
 4463 for a direct communication with the hardware. These tools have been firstly developed  
 4464 for the testing of the present pixel detector prior to assembly [148], and necessitated many  
 4465 fundamental changes to be able to operate with the upgrade system. I implemented part  
 4466 of this transition and I was able to perform the first tests with the POHs. In the example  
 4467 shown in Fig. 16.10, the POH laser bias setting is varied from 0 to 40 and for each value  
 4468 the ADC counts (or optical power) at the VME-analog FED are checked. This scan is  
 4469 performed for different values of POH laser gain setting. The optical power sharply increases  
 4470 until the saturation of the receiver at the FED is reached. This test ensures the slow I<sup>2</sup>C  
 4471 communication with the auxiliary electronic components of the system. Other tests have been  
 4472 performed after pixel modules have been added to the system and aimed at establishing the  
 4473 functionality of the fast I<sup>2</sup>C communication to/from the TBM. This was checked by verifying  
 4474 for instance the presence of the TBM signal at the corresponding POH output through an

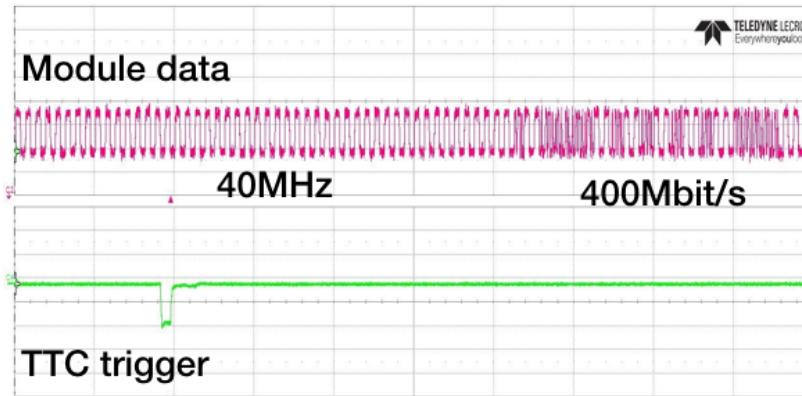


**Figure 16.9:** Test stand setup at UZH including all components of one BPix sector together with a few pixel modules. The CAEN power supply and the VME back-end system are also indicated.

4475 optical probe plugged into the oscilloscope. If the fast I<sup>2</sup>C communication through the digital  
 4476 circuit is functional, the TBM sends the 400 Mbit/s data stream upon the arrival of a trigger  
 4477 as shown in Fig. 16.11.



**Figure 16.10:** Typical result of the scan of a POH laser bias and gain. For each value of the laser settings, the optical power is readout at the VME-analog FED.



**Figure 16.11:** Module output signal initialized by a trigger sent by the TTC. The signal is acquired at the oscilloscope with an optical probe connected to the POH laser output. The 400 Mbit/s signal contains TBM trailer and header, and the header from 8 ROCs.

These kinds of test are not suitable when dealing with a large number of channels. For detector assembly and commissioning the full functionality of each module has to be checked and calibrations have to be performed over the whole detector. In order to achieve this in a reasonable amount of time, the full pixel online software (Section 14.4) has to be used and upgraded to be able to operate with the new system. Many fundamental changes to the software have been firstly applied for the pilot system, as well as for operations with the FPix test stand at CERN. These changes mainly concerned the `FEDInterface` class, where the new features of the digital VME-FED had to be implemented. In order to test and debug the new developments I installed POS and applied the modifications required to run it with the system test at UZH. After establishing the basic functionalities, I developed new calibrations for the upgrade detector as described in the next section. This work represents the first attempt to operate the upgrade BPix system with POS and hence, has been crucial to understand, implement and test new BPix specific developments.

In mid-2016 a separate table has been setup in the same laboratory to operate the upgrade electronics with the  $\mu$ TCA DAQ system as this became available only at a later time. A new  $\mu$ TCA version of POS was developed in the meantime by the FPix group, and the second test stand has been very useful to gain expertise with the new developments. Several of the new features for the upgrade system that I included in the VME-POS have been easily exported to the  $\mu$ TCA-POS. The transition has been straightforward since I developed the new code such that it is transparent with respect to the parts of the software handling the  $\mu$ TCA or VME communication.

## 16.6 Testing and calibration

Since the upgrade system features a new digital readout, some of the steps in the calibration procedures developed for the present detector become obsolete, whereas new kind of tests are needed. In particular, the adjustment of the FED receiver offset, of the UB and address levels are dropped for obvious reasons. Furthermore, since the digital FED is not able to register the optical power, it is not possible to perform a laser bias scan (Fig. 15.10). In this section, a calibration procedure for the Phase-1 detector is described with main focus on the new developments. This procedure is suitable for testing each sector of the detector after assembly at PSI as well as for the commissioning of the whole detector at CERN. It is implemented in several steps, each aimed at testing one particular functionality. Each step

4509 produces new detector settings, such as POH bias and gain or TBM and ROC DACs, after a  
 4510 dedicated scan of a set of parameters. The procedure has been fully tested with the table top  
 4511 system at UZH and few pixel modules, and results will be presented and discussed in the  
 4512 following. Only modules equipped with TBM08 and TBM09 have been used for these tests,  
 4513 since the TBM10 was not available at the time of this work. The TBM10 features slightly  
 4514 different readout and programming mechanisms, and additional modifications to the software  
 4515 are required. Further improvements to this preliminary version of the procedure have been  
 4516 implemented later on by the BPix group in order to have a finalized version for detector  
 4517 commissioning that also includes specific developments to operate with the TBM10.

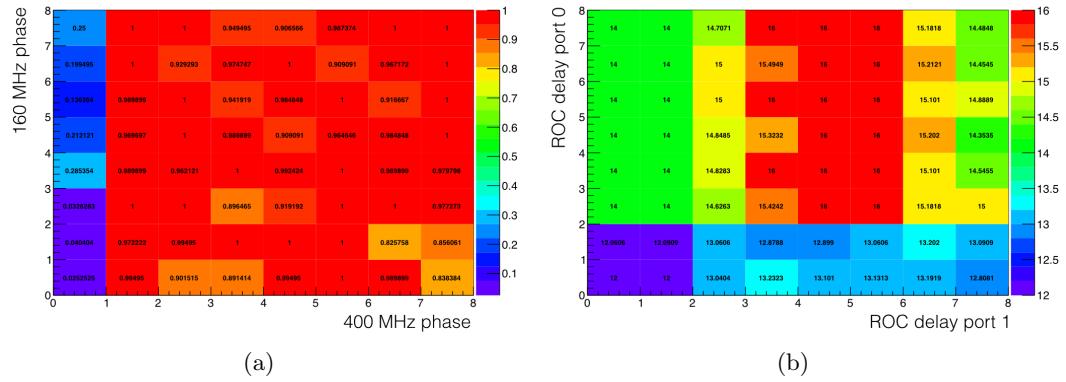
#### 4518 16.6.1 Delay adjustments

4519 The synchronization between the 160 Mbit/s ROC data and the 400 Mbit/s final output data  
 4520 stream has to be adjusted in order to be able to correctly readout the module output signal.  
 4521 The data alignment can be adjusted by programming two 8-bit registers internal to the TBM.  
 4522 The first register controls the phases of the 160 MHz and 400 MHz PLLs integrated in the  
 4523 TBM. For the TBM09 only one of the two cores (TBM A and TBM B in Fig. 16.6) has to be  
 4524 programmed since the PLLs are common to both. The second register controls the delay in  
 4525 the data stream of the two readout groups of ROCs (Port 0 and Port 1 in Fig. 16.6), and  
 4526 for TBM09 it has to be programmed for each core. These 4 parameters can be varied in the  
 4527 range 0–7 ns in steps of 1 ns so that 3 bits are sufficient for each of them.

4528 After a calibration of the Delay25 chip that ensures communication between the pxFEC  
 4529 and the modules (Fig. 15.8), the first step is to perform a scan of the TBM PLLs' phases.  
 4530 For each set of values the digital data are readout at the FED, and it is checked that the  
 4531 TBM had sent header and trailer. This is done for a number of triggers defined by the user  
 4532 and for each FED channel corresponding to the same module (Fig. 16.6). The TBM header  
 4533 and trailer are only available in the FIFO1, which can be accessed only if the normal data  
 4534 flow to the other FIFOs is interrupted. This is achieved by programming a FED register.  
 4535 An example of such scan for one module is shown in Fig. 16.12(a), where the efficiency for  
 4536 TBM header and trailer to be recorded is measured as a function of the two phases. The  
 4537 efficiency is averaged over all the FED channels. At the end of the test new settings for the  
 4538 TBM are produced, corresponding to a pair of phase values corresponding to 100% efficiency.  
 4539 Since there might be more than one bin corresponding to 100% efficiency only the first one  
 4540 is chosen. However, this algorithm can be improved by picking a bin from a region that  
 4541 presents small variations. If no 100% efficiency bins are found, the bin corresponding to  
 4542 the maximum is picked and a error flag is saved in the output file. The histogram with the  
 4543 resulting scan is also saved for each module. This test ensures that the TBM settings are  
 4544 correct to be able to read out the signal. It is also very useful to have a feedback on the  
 4545 status of the detector, for instance after installation, similarly to the FED baseline calibration  
 4546 for the present detector (Fig. 15.9). In fact, with this test it is verified that clock, trigger  
 4547 and programming signals are correctly arriving at the pixel modules and that the TBM can  
 4548 be programmed. In addition, low efficiency might indicate poor optical connections and the  
 4549 problem can be immediately solved by cleaning the fibers. Issues in the mapping between mod-  
 4550 ules and FED channels, as well as possible broken channels, can also be identified at this stage.

4551 The second step consists of scanning the ROC delays, and for each set of values it is  
 4552 checked that each ROC had sent the header. As for the scan of the TBM PLLs' phases,  
 4553 this is done for a number of triggers defined by the user, and the efficiency is measured for  
 4554 each ROC. For the TBM09 the same value is programmed for the two cores. A histogram is  
 4555 produced with the efficiency averaged over all ROCs and FED channels for each pair of delay

values, as shown in Fig. 16.12(b). New settings for the TBM are also produced, corresponding to the delay values giving 16 ROC headers recorded at the FED. The choice of the best settings follows the same strategy described above for the TBM PLLs' phases. This test ensures that the TBM settings are correct to be able to read out the signal and it also verifies the functionality of the token passage.



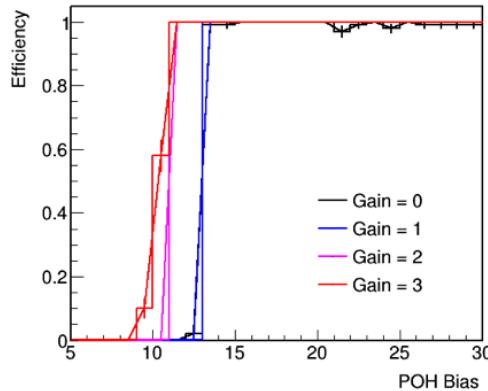
**Figure 16.12:** (a) Example of scan of the 2 TBM PLLs' phases for one module. For each pair of values the module signal is read out several times at the digital FED, and the presence of TBM header and trailer is verified. An efficiency is obtained by dividing by the number of times the signal has been read out. (b) Example of scan of the delays in the data from the two ROC readout groups. For each pair of values the presence of each ROC header in the readout data at digital FED is verified. The average number of ROCs that sent the header is obtained by dividing by the number of times the signal has been read out.

### 16.6.2 POH bias and gain

The tests described in the previous section assume that the POH laser settings are good enough to allow for a correct readout of the signal. In fact, a too low laser bias and gain result in a small difference between the 0 and 1 levels of the digital signal and consequently to a large error bit rate. It has been found with the test stand that a bias of 40 and a gain of 3 are sufficient to be able to correctly read out the signal so that the tests of the previous section can be safely run with such values. However, once the functionality of the TBM has been verified a scan of the POH setting should be run to obtain a finer adjustment aimed at minimizing the power consumption. As already mentioned, measuring the optical power as a function of the laser settings is not possible with the digital FED, whereas an approximate indication of the error bit rate can be obtained by measuring the known TBM signal. In this calibration, a scan of the POH bias and gain is performed, and for each value it is checked that the TBM had sent header and trailer. An example of such scan is shown in Fig. 16.13. At the end of the calibration new settings are produced corresponding to a safe area, i.e. 5 units above the beginning of the saturation.

### 16.6.3 Read back test

As discussed in Section 16.3.2, the ROC header is followed by 3 bits where the first one is always a 0 and the other 2 contain the so called read back data. These data contain different information depending on the value programmed in the ROC read back register. For instance, for a value of 12 the ROC analog current is returned. It has been found after several tests



**Figure 16.13:** Example of scan of a POH laser bias and gain. For each value of the laser settings the module signal is read out several times at the digital FED and the presence of TBM header and trailer is verified.

4582 that not all delay settings allow for a correct read out of the information contained in the  
 4583 ROC read back data. Hence, another scan of the TBM PLLs' phases and ROC delays is  
 4584 performed and for each value the information in the read back data is verified as explained in  
 4585 the following.

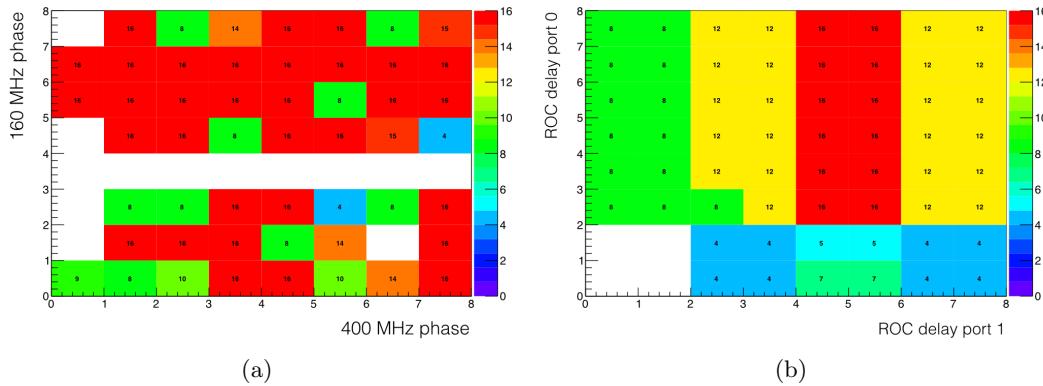
4586 The read back word is encoded in 16 bits (Table 16.1), which cannot be sent at once given  
 4587 the limited size of the ROC data stream. As mentioned above there are 2 bits available for  
 4588 the read back data which are sent every time the readout is initialized by a trigger. Of these  
 4589 two, only one belongs to the actual information, so that 16 triggers have to be sent before  
 4590 the entire word can be decoded. The second bit is used to indicate the start of the word. Of  
 4591 the 16 bits collected, the first 8-bits are used for the required information (for instance the  
 4592 analog current), 4 bits contain the value written to the read back register (for instance 12 for  
 4593 the analog current), and the last 4 bits encode the ROC I<sup>2</sup>C address (from 0 to 15).

4594 In order to ensure that the TBM settings chosen by the previous algorithms allow for a  
 4595 correct read back, a scan is performed over the TBM PLLs phases and ROC delays, and for  
 4596 each value it is checked that the ROC address in the read back word is the expected one.  
 4597 This is indeed known from the mapping of FED channels and ROCs specified in the detector  
 4598 configuration database. At the end of the scan, it is verified that the current settings from  
 4599 previous calibrations give a correct read back of the ROC address for each ROC, otherwise  
 4600 new settings are produced. Examples of such scans are shown in Fig. 16.14. In order to  
 4601 acquire statistics the 16-bit word is read several times. The algorithm to chose the optimal  
 4602 phases or delays is the same as for the tests described in Section 16.6.1.

4603 It has to be noted that this test gives better results if charge is injected in at least one  
 4604 pixel in each of the ROCs. In fact, without a pixel hit, the read back bit is the last bit being  
 4605 sent before the TBM trailer. If the timing of the data stream is not optimal, this bit might  
 4606 get lost and the read back word would not be meaningful. However, if a pixel hit is added  
 4607 after the ROC header, the last bit being sent before the TBM trailer belongs to the pulse  
 4608 height information and the read back word would not be affected. On the other hand, in  
 4609 order to be able to readout the pixel hit the settings for the comparator threshold (*VcThr*)  
 4610 and the delay (*CalDel*) in the charge injection have to be correct. It is then fundamental to  
 4611 run a *VcThrCalDel* scan (Fig. 15.11) before performing the read back test. This calibration  
 4612 is also useful to verify the functionality of the charge injection mechanism.

**Table 16.1:** Mechanism used by the ROC to send read back data. Sixteen triggers are needed to read out the full 16-bit word. One bit is used to indicate the start of the message. In this example, the ROC 2 is readout, and the value written in the read back register is 12, which allows for the readout of the analog current. The latter is equal to 88 ADC counts.

Trigger #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Message start	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Message data	0	0	1	0	1	1	0	0	0	1	0	1	1	0	0	0
Comment	ROC I <sup>2</sup> C address				last value written to the read back register											



**Figure 16.14:** (a) Example of scan of the 2 TBM PLLs' phases (a) and ROC delays (b) for one module. For each pair of values it is verified that the message data obtained with the read back mechanism contain the expected ROC I<sup>2</sup>C address. If the timing is correct, 16 ROCs per module have to be correctly identified.

#### 4613 16.6.4 ROC analog current and digital voltage

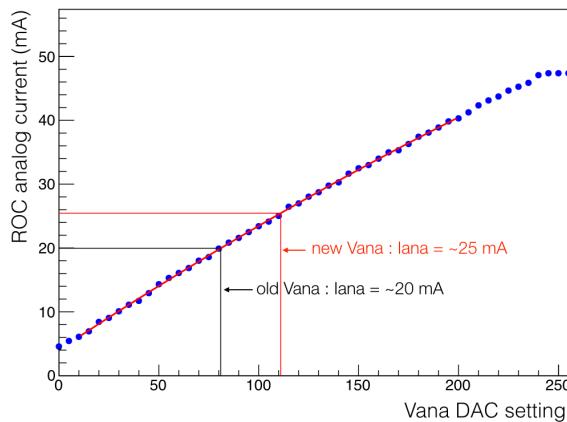
4614 The read back mechanism described in the previous section allows to access the ROC analog  
 4615 current information (*Iana*). For the present detector it is possible to access only the total  
 4616 current of a large number of ROCs provided by a reading of the power supply, so that  
 4617 only an average value per ROC can be obtained. As discussed in Chapter 15, the analog  
 4618 current increases with irradiation and a re-calibration of the ROC analog voltage setting  
 4619 *Vana* is necessary to bring back the current in the safe range and recover the optimal  
 4620 ROC performance (Fig. 15.3(b)). The *Vana* calibration for each ROC is a time consuming  
 4621 procedure that converges after a large number of iterations of the S-curve test, which provides  
 4622 measurements of the threshold every time the settings are changed (Section 15.3.2). A much  
 4623 more simple procedure has been developed for the upgrade ROCs, since the measurement of  
 4624 the analog current is directly provided in the read back data. The value is returned in ADC  
 4625 units and a conversion curve measured per each ROC during module testing is applied to  
 4626 obtain the corresponding value in mA. The curve is of the form

$$I_{ana}(\text{ADC}) = a + b \times I_{ana}(\text{mA}) + c \times I_{ana}^2(\text{mA}), \quad (16.1)$$

4627 and its parameters are saved for each ROC in the detector configuration database. As  
 4628 shown in Fig. 16.15, for each value of the *Vana* setting the analog current is read out from  
 4629 the ROC read back bits, and the conversion curve above applied to the original value in ADC.

4630 At the end of the scan, a  $V_{ana}$  setting corresponding to the target  $I_{ana}$  value of 25 mA is  
4631 chosen.

4632 The same calibration code can be also configured to perform a measurement of the digital  
4633 voltage  $Vd$  per each ROC. This parameter is controlled by a corresponding DAC and it is  
4634 useful to monitor it as well during operations. Its value can be returned by using the same  
4635 read back mechanism but by changing the value written in the read back register. As for  
4636 the analog current, a ADC/mV conversion curve has been measured per each ROC during  
4637 module testing and it is applied in the final result of this measurement.



**Figure 16.15:** ROC analog current as readout with the ROC read back mechanism as a function of the analog voltage regulator setting.

# Conclusions

4641 The contributions to the calibration and upgrade of the CMS pixel barrel detector have been  
4642 presented in this part of thesis. First, a major effort was made during the long shut-down to  
4643 recover the full detector performance after the first LHC physics run. This effort included  
4644 the replacement of broken channels and the re-calibration of detector parameters at low  
4645 temperature needed to compensate for the effects of radiation damage. The detector was  
4646 re-installed into CMS in December 2014 and commissioned in January 2015 for the second  
4647 LHC physics run at the highest center-of-mass energy of 13 TeV. The entire operation was  
4648 completed in only a few days thanks to the expertise gained during the long shut-down. The  
4649 first pp collision at 13 TeV occurred on March, 21st 2015 and since then the detector was  
4650 running stable with a high data-taking efficiency.

4651 A second contribution has been made during this thesis to the Phase 1 upgrade pixel  
4652 project. The upgrade detector will allow to maintain the excellent tracking performance of  
4653 CMS at the upcoming higher luminosity conditions at the LHC. A test stand at the University  
4654 of Zurich has been set up and run with the aim of testing the performance of the complete  
4655 upgrade pixel system and gain experience in its operations. The setup includes a slice of the  
4656 full CMS pixel DAQ together with all the upgrade electronics for the power, readout and  
4657 control systems as well as newly developed pixel modules. The test system has been employed  
4658 to implement and test new developments in the pixel online software used to operate with  
4659 the detector. Although the software architecture remains unchanged, several calibration  
4660 procedures for the present detector become obsolete with the novel digital readout of the  
4661 upgrade system, whereas brand-new tests have been developed. Additional fundamental  
4662 modifications had to be understood and implemented to be able to operate the software with  
4663 the upgrade detector. The Phase 1 pixel upgrade project is now at its last stage of assembly  
4664 and testing of the entire system. The novel calibration procedures are aimed at guaranteeing  
4665 the success of these operations, as well as of the installation and commissioning of the new  
4666 system planned for March 2017.

4667

## Part III

4668

## Summary

# Studies on track reconstruction problems

4669

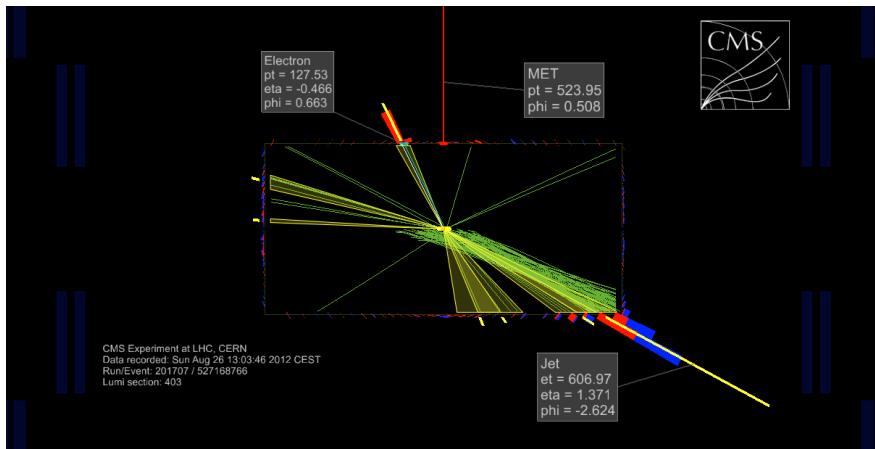
4670

4671

4672

A scan of the displays of all the events in 8 TeV data with  $m_{\text{WH}} > 1.6 \text{ TeV}$  passing all the selection criteria for the  $\ell\nu b\bar{b}$  final state (Table 8.1), reported that presence of two events characterized by a rare specific kind of noise. This noise arises from an anomalous behaviour of the tracking algorithm in the transition region between the barrel (TOB) and endcap (TEC) regions of the silicon tracker, namely, in the pseudorapidity range  $1 < |\eta| < 1.5$  (Fig. 3.5). As a consequence, many fake (displaced) tracks are associated to the selected H jet candidate. Figure A.1 shows the event display of one of the two events affected by this problem, while Figure A.2 shows the same feature in simulation.

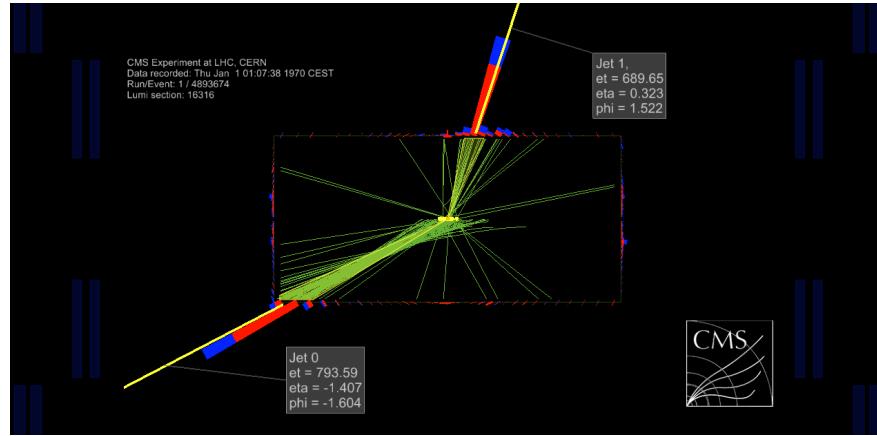
In order to reject this type of noise, it is common in CMS analyses to apply a standard filtering algorithm that discards the event if there is an anomalous amount of tracks that have been seeded in the TOB-TEC transition region. The efficiency of this filter on signal events is about 97% independently on the H jet  $p_T$ .



**Figure A.1:** Display of one typical anomalous event found in data recorded by the CMS experiment. Many fake and displaced tracks are reconstructed creating a bias in the jet reconstruction. Only tracks with  $p_T$  larger than 2 GeV are shown.

However, further checks performed on the anomalous events showed that after applying the standard filter residual noise can still be identified in the problematic  $\eta$  region. Therefore, it has been decided for the analysis described in this work to apply an additional requirement on the  $\eta$  of the selected H jet candidate (Section 6.4.2). In particular, CA8 jets are rejected if their pseudorapidity falls in the problematic region  $1 < |\eta| < 1.8$ . As described in the following, the choice for this fiducial cut is motivated by the disagreement between data and simulation in the rate at which the noise occur.

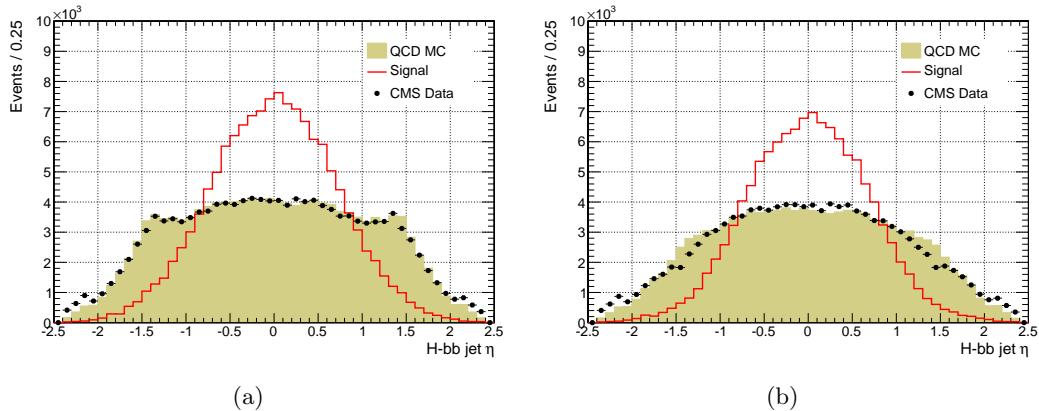
The efficiency of the standard filter is studied as a function of the H jet  $p_T$  and  $\eta$  in a dijet sample with high statistics in both data and simulation. The sample is selected requiring at least two jets, with  $p_T > 400 \text{ GeV}$  for the leading jet and  $p_T > 80 \text{ GeV}$  for the sub-leading



**Figure A.2:** Display of one typical anomalous event in simulation. Only tracks with  $p_T$  larger than 2 GeV are shown.

one. At least one of the jet has to be b-tagged using the same combined b tagging algorithm as for the main analysis selection, representing thus the H jet candidate. The jet that fails the b tagging is required to have low pruned mass ( $m_{jet} < 40$  GeV).

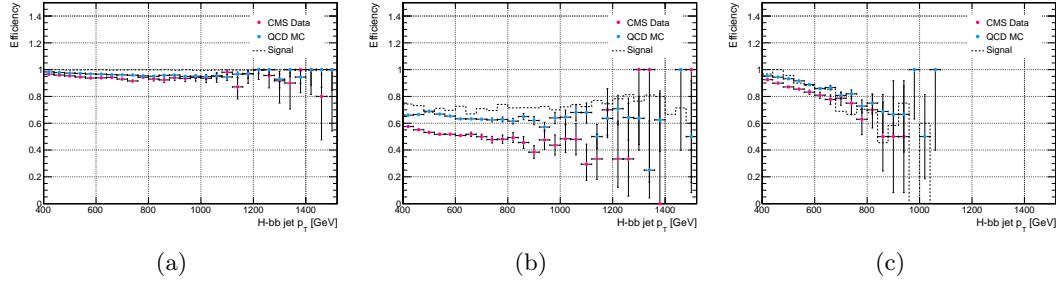
Figure A.3 shows the effect of the filter on the jet  $\eta$  distribution comparing data, simulated signal and QCD background: the signal distribution is rather unaffected while data and QCD background distributions show a reduction of events in the problematic  $\eta$  region after applying the filter, as expected.



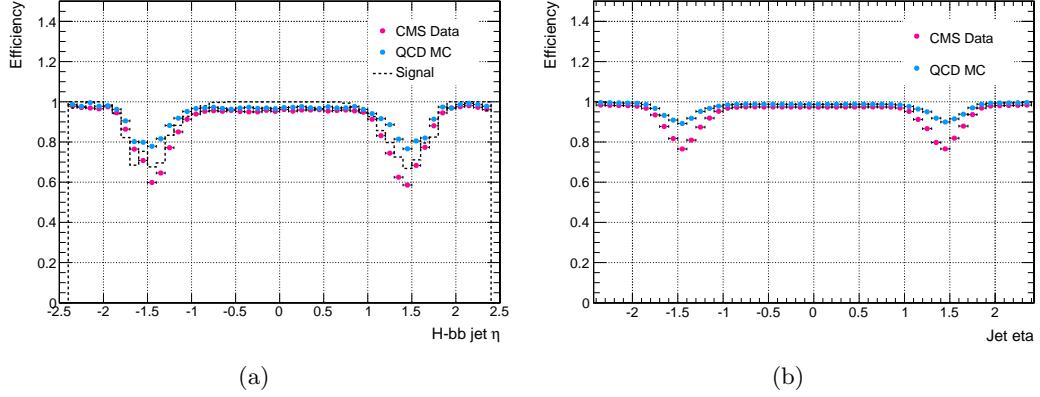
**Figure A.3:** Comparison of the H jet  $\eta$  distributions for data, and simulated signal and QCD background before (a) and after (b) applying the tracking noise filter. Signal jets are mostly central in the detector.

Figure A.4 shows the filter efficiency on data and simulated signal and QCD background as a function of the H jet candidate  $p_T$  for different jet  $\eta$  regions. A little dependence of the filter efficiency with the jet  $p_T$  is observed in the regions  $0 < |\eta| < 1$  and  $1.0 < |\eta| < 1.5$ , while in the forward region  $1.5 < |\eta| < 2.4$  the efficiency decreases with the jet  $p_T$ . The performance of the filter in the different  $\eta$  regions is summarized in Figure A.5(a). A large discrepancy between data and simulation is found in the pseudorapidity region  $1.0 < |\eta| < 1.8$ , where the simulation does not sufficiently well describe the full material budget of the tracking detector. The same studies are also performed removing the b-tagging requirement. The filter efficiency as a function of the leading jet  $\eta$  for this case is shown in Fig. A.5(b), for both data and

simulation. The increase in efficiency compared to what obtained in b-tagging shows that the b-tagging requirement enriches the samples with events characterized by this noise up to 30%, making this analysis systematically prone to it.



**Figure A.4:** Efficiency of the tracking noise filter as a function of the H jet  $p_T$  for data, and simulated signal and QCD background for jets reconstructed in the pseudorapidity regions  $0 < |\eta| < 1$  (a),  $1.0 < |\eta| < 1.5$  (b), and  $1.5 < |\eta| < 2.4$  (c).



**Figure A.5:** Efficiency of the tracking noise filter as a function of the leading jet  $p_T$  for data, and simulated signal and QCD background. (a) The leading jet is required to be b-tagged with the combined b tagging algorithm used in the main analysis. (b) The b tagging requirements for the leading jet are removed.

4714

# Bibliography

- [1] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001.
- [2] W.J. Stirling, “Parton luminosity and cross section plots”.  
<http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>. *Private communication*.
- [3] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.
- [4] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [5] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.
- [6] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.
- [7] F. Marcastel, “CERN’s Accelerator Complex. La chane des accrateurs du CERN”, General Photo.
- [8] CMS Collaboration, “CMS Public Luminosity Results”.  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [9] N. Cartiglia, “Measurement of the proton-proton total, elastic, inelastic and diffractive cross sections at 2, 7, 8 and 57 TeV”, arXiv:1305.6131.
- [10] CMS Collaboration, “The CMS tracker system project: Technical Design Report”. Technical Design Report CMS. CERN, 1997.
- [11] CMS Collaboration, “Pixel performance plots”.  
<https://twiki.cern.ch/twiki/bin/view/CMS/PixelOfflinePlots>.
- [12] CMS Collaboration Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [13] CMS Collaboration Collaboration, “The CMS hadron calorimeter project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [14] P. Adzic, “Energy resolution of the barrel of the CMS Electromagnetic Calorimeter”, *Journal of Instrumentation* **2** (2007), no. 04, P04004.
- [15] M. A. Dobbs et al., “Les Houches guidebook to Monte Carlo generators for hadron collider physics”, in *Physics at TeV colliders. Proceedings, Workshop, Les Houches, France, May 26-June 3, 2003*, pp. 411–459. 2004. arXiv:hep-ph/0403045.
- [16] J. C. Collins and D. E. Soper, “The Theorems of Perturbative QCD”, *Ann. Rev. Nucl. Part. Sci.* **37** (1987) 383–409, doi:10.1146/annurev.ns.37.120187.002123.
- [17] ZEUS Collaboration, “The ZEUS detector: Status report 1993”,.

- [18] H1 Collaboration, “The H1 detector at HERA”, *Nucl. Instrum. Meth. A* **386** (1997), no. 2, 310 – 347, doi:[http://dx.doi.org/10.1016/S0168-9002\(96\)00893-5](http://dx.doi.org/10.1016/S0168-9002(96)00893-5).
- [19] G. Altarelli and G. Parisi, “Asymptotic freedom in parton language”, *Nuclear Physics B* **126** (1977), no. 2, 298 – 318, doi:[http://dx.doi.org/10.1016/0550-3213\(77\)90384-4](http://dx.doi.org/10.1016/0550-3213(77)90384-4).
- [20] J. Pumplin et al., “New generation of parton distributions with uncertainties from global QCD analysis”, *JHEP* **07** (2002) 012, doi:[10.1088/1126-6708/2002/07/012](https://doi.org/10.1088/1126-6708/2002/07/012), arXiv:hep-ph/0201195.
- [21] H.-L. Lai et al., “New parton distributions for collider physics”, *Phys. Rev. D* **82** (2010) 074024, doi:[10.1103/PhysRevD.82.074024](https://doi.org/10.1103/PhysRevD.82.074024), arXiv:1007.2241.
- [22] R. D. Ball et al., “Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology”, *Nucl. Phys. B* **849** (2011) 296, doi:[10.1016/j.nuclphysb.2011.03.021](https://doi.org/10.1016/j.nuclphysb.2011.03.021), arXiv:1101.1300.
- [23] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1”, *Comput. Phys. Commun.* **178** (2008) 852, doi:[10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036), arXiv:0710.3820.
- [24] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05** (2006) 026, doi:[10.1088/1126-6708/2006/05/026](https://doi.org/10.1088/1126-6708/2006/05/026), arXiv:hep-ph/0603175.
- [25] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011) 128, doi:[10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128), arXiv:1106.0522.
- [26] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:[10.1088/1126-6708/2007/11/070](https://doi.org/10.1088/1126-6708/2007/11/070), arXiv:0709.2092.
- [27] S. Frixione, P. Nason, and B. R. Webber, “Matching NLO QCD and parton showers in heavy flavor production”, *JHEP* **08** (2003) 007, doi:[10.1088/1126-6708/2003/08/007](https://doi.org/10.1088/1126-6708/2003/08/007), arXiv:hep-ph/0305252.
- [28] GEANT4 Collaboration, “GEANT4 — a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:[10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [29] CMS Collaboration, “Measurement of the Underlying Event Activity at the LHC with  $\sqrt{s} = 7$  TeV and Comparison with  $\sqrt{s} = 0.9$  TeV”, *JHEP* **09** (2011) 109, doi:[10.1007/JHEP09\(2011\)109](https://doi.org/10.1007/JHEP09(2011)109), arXiv:1107.0330.
- [30] CMS Collaboration, “Study of the underlying event at forward rapidity in pp collisions at  $\sqrt{s} = 0.9$ , 2.76, and 7 TeV”, *JHEP* **04** (2013) 072, doi:[10.1007/JHEP04\(2013\)072](https://doi.org/10.1007/JHEP04(2013)072), arXiv:1302.2394.
- [31] D. Pappadopulo, A. Thamm, R. Torre, and A. Wulzer, “Heavy Vector Triplets: Bridging Theory and Data”, *JHEP* **09** (2014) 060, doi:[10.1007/JHEP09\(2014\)060](https://doi.org/10.1007/JHEP09(2014)060), arXiv:1402.4431.
- [32] G. Burdman, M. Perelstein, and A. Pierce, “Large Hadron Collider tests of a little Higgs model”, *Phys. Rev. Lett.* **90** (2003) 241802, doi:[10.1103/PhysRevLett.90.241802](https://doi.org/10.1103/PhysRevLett.90.241802), arXiv:hep-ph/0212228. [Erratum: *Phys. Rev. Lett.* 92, 049903(2004)].

- [33] J. M. Campbell, R. K. Ellis, and D. L. Rainwater, “Next-to-leading order QCD predictions for  $W + 2$  jet and  $Z + 2$  jet production at the CERN LHC”, *Phys. Rev.* **D68** (2003) 094021, doi:10.1103/PhysRevD.68.094021, arXiv:hep-ph/0308195.
- [34] J. M. Campbell, R. K. Ellis, and C. Williams, “Vector boson pair production at the LHC”, *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.
- [35] J. M. Campbell and R. K. Ellis, “Top-quark processes at NLO in production and decay”, *J. Phys.* **G42** (2015), no. 1, 015005, doi:10.1088/0954-3899/42/1/015005, arXiv:1204.1513.
- [36] J. M. Campbell, R. K. Ellis, and F. Tramontano, “Single top production and decay at next-to-leading order”, *Phys. Rev.* **D70** (2004) 094012, doi:10.1103/PhysRevD.70.094012, arXiv:hep-ph/0408158.
- [37] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in FEWZ”, *Phys. Rev.* **D86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.
- [38] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”, *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.
- [39] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.
- [40] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [41] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [42] W. Adam, B. Mangano, T. Speer, and T. Todorov, “Track Reconstruction in the CMS tracker”, Technical Report CMS-NOTE-2006-041, CERN, Dec, 2006.
- [43] P. Billoir, “Progressive track recognition with a Kalman like fitting procedure”, *Comput. Phys. Commun.* **57** (1989) 390–394, doi:10.1016/0010-4655(89)90249-X.
- [44] R. Frhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth.* **A262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.
- [45] T. Speer et al., “Vertex Fitting in the CMS Tracker”, Technical Report CMS-NOTE-2006-032, CERN, Feb, 2006.
- [46] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* **86** (Nov, 1998) 2210–2239, doi:10.1109/5.726788.
- [47] W. Waltenberger, R. Frhwirth, and P. Vanlaer, “Adaptive vertex fitting”, *Journal of Physics G: Nuclear and Particle Physics* **34** (2007), no. 12, N343.

- 4830 [48] CMS Collaboration, “CMS Tracking Performance Results”.  
4831 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsTRK>.
- 4832 [49] S. Baffioni et al., “Electron reconstruction in CMS”, *Eur. Phys. J.* **C49** (2007)  
4833 1099–1116, doi:10.1140/epjc/s10052-006-0175-5.
- 4834 [50] CMS Collaboration, “Commissioning of the Particle-flow Event Reconstruction with  
4835 the first LHC collisions recorded in the CMS detector”, CMS Physics Analysis  
4836 Summary CMS-PAS-PFT-10-001, 2010.
- 4837 [51] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for  
4838 Jets, Taus, and MET”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009.
- 4839 [52] W. Adam, R. Frhwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons  
4840 with the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005), no. 9, N9.
- 4841 [53] CMS Collaboration, “Measurements of Inclusive  $W$  and  $Z$  Cross Sections in  $pp$   
4842 Collisions at  $\sqrt{s} = 7$  TeV”, *JHEP* **01** (2011) 080, doi:10.1007/JHEP01(2011)080,  
4843 arXiv:1012.2466.
- 4844 [54] CMS Collaboration, “Performance of Electron Reconstruction and Selection with the  
4845 CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV”, *JINST* **10** (2015),  
4846 no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- 4847 [55] CMS Collaboration, “CMS ECAL Performance Results”.  
4848 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EcalDPGResults>.
- 4849 [56] CMS Collaboration, “Search for physics beyond the standard model in dilepton mass  
4850 spectra in proton-proton collisions at  $\sqrt{s} = 8$  TeV”, *JHEP* **04** (2015) 025,  
4851 doi:10.1007/JHEP04(2015)025, arXiv:1412.6302.
- 4852 [57] CMS Collaboration, “Performance of CMS muon reconstruction in  $pp$  collision events  
4853 at  $\sqrt{s} = 7$  TeV”, *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002,  
4854 arXiv:1206.4071.
- 4855 [58] CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector  
4856 Performance and Software”. Technical Design Report CMS. CERN, 2006.
- 4857 [59] CMS Collaboration, “Additional Performance Results for EPS 2015”,.
- 4858 [60] R. Radogna, A. Colaleo, and S. Nuzzo, “Search for high-mass resonances decaying into  
4859 muon pairs with the CMS experiment at LHC”. PhD thesis, Bari U., 2016. Presented  
4860 14 Mar 2016.
- 4861 [61] CMS Collaboration, “Performance of the CMS Level-1 Trigger”, Technical Report  
4862 CMS-CR-2012-322, CERN, 2012. Comments: Presented at 36th International  
4863 Conference on High Energy Physics, July 4-11, 2012, Melbourne, Australia.
- 4864 [62] G. P. Salam and G. Soyez, “A Practical Seedless Infrared-Safe Cone jet algorithm”,  
4865 *JHEP* **05** (2007) 086, doi:10.1088/1126-6708/2007/05/086, arXiv:0704.0292.
- 4866 [63] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, “Longitudinally  
4867 invariant  $k_t$  clustering algorithms for hadron hadron collisions”, *Nucl. Phys. B* **406**  
4868 (1993) 187, doi:10.1016/0550-3213(93)90166-M.
- 4869

- [64] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev. D* **48** (1993) 3160, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.
- [65] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering algorithms”, *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.
- [66] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross-sections in deep inelastic scattering”, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*. 1998. arXiv:hep-ph/9907280.
- [67] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_t$  jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [68] CMS Collaboration, “Identification techniques for highly boosted W bosons that decay into hadrons”, *JHEP* **12** (2014) 017, doi:10.1007/JHEP12(2014)017, arXiv:1410.4227.
- [69] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *Submitted to: JINST* (2016) arXiv:1607.03663.
- [70] T. C. collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *Journal of Instrumentation* **6** (2011), no. 11, P11002.
- [71] CMS Collaboration, “Pileup Jet Identification”, Technical Report CMS-PAS-JME-13-005, CERN, 2013.
- [72] CMS Collaboration, “Pileup Removal Algorithms”, Technical Report CMS-PAS-JME-14-001, CERN, 2014.
- [73] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, Technical Report CMS-PAS-JME-16-003, CERN, 2014.
- [74] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [75] M. Cacciari and G. P. Salam, “Dispelling the  $N^3$  myth for the  $k_t$  jet-finder”, *Phys. Lett.* **B641** (2006) 57–61, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.
- [76] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett.* **B659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.
- [77] “Jet energy scale and resolution performances with 13TeV data”,
- [78] CMS Collaboration, “Search for resonant pair production of Higgs bosons decaying to  $b\bar{b}$  and  $\tau^+\tau^-$  in proton-proton collisions at  $\sqrt{s} = 8$  TeV”, Technical Report CMS-PAS-EXO-15-008, 2015.
- [79] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.
- [80] W. Waltenberger, “Adaptive Vertex Reconstruction”, Technical Report CMS-NOTE-2008-033, CERN, Geneva, Jul, 2008.

- 4909 [81] CMS Collaboration, “Measurement of  $B\bar{B}$  Angular Correlations based on Secondary  
4910 Vertex Reconstruction at  $\sqrt{s} = 7$  TeV”, *JHEP* **03** (2011) 136,  
4911 doi:10.1007/JHEP03(2011)136, arXiv:1102.3194.
- 4912 [82] CMS Collaboration, “Performance of b tagging at  $\text{sqrt}(s)=8$  TeV in  
4913 multijet,  $t\bar{t}$  and boosted topology events”, Technical Report CMS-PAS-BTV-13-001,  
4914 CERN, Geneva, 2013.
- 4915 [83] CMS Collaboration, “Identification of b quark jets at the CMS Experiment in the LHC  
4916 Run 2”, Technical Report CMS-PAS-BTV-15-001, CERN, 2016.
- 4917 [84] T. C. collaboration, “Missing transverse energy performance of the CMS detector”,  
4918 *Journal of Instrumentation* **6** (2011), no. 09, P09001.
- 4919 [85] CMS Collaboration, “Performance of the CMS missing transverse momentum  
4920 reconstruction in pp data at  $\sqrt{s} = 8$  TeV”, *JINST* **10** (2015), no. 02, P02006,  
4921 doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.
- 4922 [86] CMS Collaboration, “Performance of missing energy reconstruction in 13 TeV pp  
4923 collision data using the CMS detector”, Technical Report CMS-PAS-JME-16-004,  
4924 CERN, 2016.
- 4925 [87] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys. C* **38**  
4926 (2014) 090001, doi:10.1088/1674-1137/38/9/090001.
- 4927 [88] J. Bauer, “Prospects for the Observation of Electroweak Top-Quark Production with  
4928 the CMS Experiment”. PhD thesis, Karlsruher Institut für Technologie (KIT), 2010.
- 4929 [89] CMS Collaboration, “Studies of jet mass in dijet and W/Z + jet events”, *JHEP* **05**  
4930 (2013) 090, doi:10.1007/JHEP05(2013)090, arXiv:1303.4811.
- 4931 [90] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Techniques for improved heavy particle  
4932 searches with jet substructure”, *Phys. Rev. D* **80** (2009) 051501,  
4933 doi:10.1103/PhysRevD.80.051501, arXiv:0903.5081.
- 4934 [91] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Recombination Algorithms and Jet  
4935 Substructure: Pruning as a Tool for Heavy Particle Searches”, *Phys. Rev. D* **81**  
4936 (2010) 094023, doi:10.1103/PhysRevD.81.094023, arXiv:0912.0033.
- 4937 [92] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”,  
4938 *JHEP* **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.
- 4939 [93] CMS Collaboration, “Search for massive resonances decaying into pairs of boosted  
4940 bosons in semi-leptonic final states at  $\sqrt{s} = 8$  TeV”, *JHEP* **08** (2014) 174,  
4941 doi:10.1007/JHEP08(2014)174, arXiv:1405.3447.
- 4942 [94] CMS Collaboration, “Search for massive resonances in dijet systems containing jets  
4943 tagged as W or Z boson decays in pp collisions at  $\sqrt{s} = 8$  TeV”, *JHEP* **08** (2014)  
4944 173, doi:10.1007/JHEP08(2014)173, arXiv:1405.1994.
- 4945 [95] CMS Collaboration, “Search for massive resonances decaying into pairs of boosted W  
4946 and Z bosons at  $\sqrt{s} = 13$  TeV”, Technical Report CMS-PAS-EXO-15-002, CERN,  
4947 2015.

- [96] CMS Collaboration, “V Tagging Observables and Correlations”, Technical Report CMS-PAS-JME-14-002, CERN, 2014.
- [97] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J.* **C58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.
- [98] CMS Collaboration, “Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos, and b quarks”, arXiv:1610.08066. Submitted to PLB.
- [99] M. Czakon, D. Heymes, and A. Mitov, “High-precision differential predictions for top-quark pairs at the LHC”, *Phys. Rev. Lett.* **116** (2016), no. 8, 082003, doi:10.1103/PhysRevLett.116.082003, arXiv:1511.00549.
- [100] M. J. Oreglia, “A study of the reactions  $\psi' \rightarrow \gamma\gamma\psi'$ ”. PhD thesis, Stanford University, 1980. SLAC Report SLAC-R-236.
- [101] CMS Collaboration, “Measurement of the single-top-quark  $t$ -channel cross section in  $pp$  collisions at  $\sqrt{s} = 7$  TeV”, *JHEP* **12** (2012) 035, doi:10.1007/JHEP12(2012)035, arXiv:1209.4533.
- [102] P. Kant et al., “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”, *Comput. Phys. Commun.* **191** (2015) 74–89, doi:10.1016/j.cpc.2015.02.001, arXiv:1406.4403.
- [103] CMS Collaboration, “Measurement of W+W- and ZZ production cross sections in pp collisions at  $\text{sqrt}(s) = 8$  TeV”, *Phys. Lett.* **B721** (2013) 190–211, doi:10.1016/j.physletb.2013.03.027, arXiv:1301.4698.
- [104] T. Gehrmann et al., “ $W^+W^-$  Production at Hadron Colliders in Next to Next to Leading Order QCD”, *Phys. Rev. Lett.* **113** (2014), no. 21, 212001, doi:10.1103/PhysRevLett.113.212001, arXiv:1408.5243.
- [105] CMS Collaboration, “Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV”, *JINST* **8** (2013) P09009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016. [JINST8,9009(2013)].
- [106] CMS Collaboration, “CMS Luminosity Measurement for the 2015 Data Taking Period”, CMS Physics Analysis Summary CMS-PAS-LUM-15-001, CERN, 2015.
- [107] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update”, CMS Physics Analysis Summary CMS-PAS-LUM-13-001, 2013.
- [108] M. Cacciari et al., “The  $t\bar{t}$  cross-section at 1.8-TeV and 1.96-TeV: A Study of the systematics due to parton densities and scale dependence”, *JHEP* **04** (2004) 068, doi:10.1088/1126-6708/2004/04/068, arXiv:hep-ph/0303085.
- [109] S. Catani, D. de Florian, M. Grazzini, and P. Nason, “Soft gluon resummation for Higgs boson production at hadron colliders”, *JHEP* **07** (2003) 028, doi:10.1088/1126-6708/2003/07/028, arXiv:hep-ph/0306211.
- [110] M. Botje et al., “The PDF4LHC Working Group Interim Recommendations”, (2011). arXiv:1101.0538.

- 4988 [111] S. Alekhin et al., “The PDF4LHC Working Group Interim Report”, (2011).  
4989 [arXiv:1101.0536](https://arxiv.org/abs/1101.0536).
- 4990 [112] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, “Parton distributions for the  
4991 LHC”, *Eur. Phys. J. C* **63** (2009) 189, doi:[10.1140/epjc/s10052-009-1072-5](https://doi.org/10.1140/epjc/s10052-009-1072-5),  
4992 [arXiv:0901.0002](https://arxiv.org/abs/0901.0002).
- 4993 [113] R. D. Ball et al., “Impact of heavy quark masses on parton distributions and LHC  
4994 phenomenology”, *Nucl. Phys. B* **849** (2011) 296,  
4995 doi:[10.1016/j.nuclphysb.2011.03.021](https://doi.org/10.1016/j.nuclphysb.2011.03.021), [arXiv:1101.1300](https://arxiv.org/abs/1101.1300).
- 4996 [114] A. L. Read, “Presentation of search results: The CL<sub>S</sub> technique”, *J. Phys. G* **28**  
4997 (2002) 2693, doi:[10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- 4998 [115] T. Junk, “Confidence level computation for combining searches with small statistics”,  
4999 *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:[10.1016/S0168-9002\(99\)00498-2](https://doi.org/10.1016/S0168-9002(99)00498-2),  
5000 [arXiv:hep-ex/9902006](https://arxiv.org/abs/hep-ex/9902006).
- 5001 [116] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “Asymptotic formulae for  
5002 likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1554,  
5003 doi:[10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0), [10.1140/epjc/s10052-013-2501-z](https://doi.org/10.1140/epjc/s10052-013-2501-z),  
5004 [arXiv:1007.1727](https://arxiv.org/abs/1007.1727). [Erratum: *Eur. Phys. J. C* **73** (2013) 2501].
- 5005 [117] A. Wald, “Tests of Statistical Hypotheses Concerning Several Parameters When the  
5006 Number of Observations is Large”, *Transactions of the American Mathematical  
5007 Society* **54** (1943), no. 3, 426–482.
- 5008 [118] S. S. Wilks, “The Large-Sample Distribution of the Likelihood Ratio for Testing  
5009 Composite Hypotheses”, *Ann. Math. Statist.* **9** (03, 1938) 60–62,  
5010 doi:[10.1214/aoms/1177732360](https://doi.org/10.1214/aoms/1177732360).
- 5011 [119] E. Gross and O. Vitells, “Trial factors or the look elsewhere effect in high energy  
5012 physics”, *Eur. Phys. J. C* **70** (2010) 525–530,  
5013 doi:[10.1140/epjc/s10052-010-1470-8](https://doi.org/10.1140/epjc/s10052-010-1470-8), [arXiv:1005.1891](https://arxiv.org/abs/1005.1891).
- 5014 [120] F. Garwood, “Fiducial Limits for the Poisson Distribution”, *Biometrika* **28** (1936)  
5015 437, doi:[10.1093/biomet/28.3-4.437](https://doi.org/10.1093/biomet/28.3-4.437).
- 5016 [121] CMS Collaboration, “Search for a massive resonance decaying into a Higgs boson and  
5017 a W or Z boson in hadronic final states in proton-proton collisions at  $\sqrt{s} = 8$  TeV”,  
5018 *JHEP* **02** (2016) 145, doi:[10.1007/JHEP02\(2016\)145](https://doi.org/10.1007/JHEP02(2016)145), [arXiv:1506.01443](https://arxiv.org/abs/1506.01443).
- 5019 [122] CMS Collaboration, “Search for new resonances decaying via WZ to leptons in  
5020 proton-proton collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Lett. B* **740** (2015) 83,  
5021 doi:[10.1016/j.physletb.2014.11.026](https://doi.org/10.1016/j.physletb.2014.11.026), [arXiv:1407.3476](https://arxiv.org/abs/1407.3476).
- 5022 [123] CMS Collaboration, “Search for massive WH resonances decaying into the  $\ell\nu b\bar{b}$  final  
5023 state at  $\sqrt{s} = 8$  TeV”, *Eur. Phys. J. C* **76** (2016) 237,  
5024 doi:[10.1140/epjc/s10052-016-4067-z](https://doi.org/10.1140/epjc/s10052-016-4067-z), [arXiv:1601.06431](https://arxiv.org/abs/1601.06431).
- 5025 [124] CMS Collaboration, “Search for Narrow High-Mass Resonances in Proton-Proton  
5026 Collisions at  $\sqrt{s} = 8$  TeV Decaying to a Z and a Higgs Boson”, *Phys. Lett. B* **748**  
5027 (2015) 255, doi:[10.1016/j.physletb.2015.07.011](https://doi.org/10.1016/j.physletb.2015.07.011), [arXiv:1502.04994](https://arxiv.org/abs/1502.04994).

- 5028 [125] ATLAS Collaboration, “Search for high-mass diboson resonances with boson-tagged  
 5029 jets in proton-proton collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, *JHEP* **12**  
 5030 (2015) 055, doi:10.1007/JHEP12(2015)055, arXiv:1506.00962.
- 5031 [126] CERN, “LHC commissioning schedule”. <https://lhcb-commissioning.web.cern.ch/lhc-commissioning/schedule/LHC-long-term.htm>.
- 5033 [127] H. Kästli et al., “CMS barrel pixel detector overview”, *Nucl. Instrum. Meth. A* **582**  
 5034 (2007), no. 3, 724 – 727, doi:<http://dx.doi.org/10.1016/j.nima.2007.07.058>.  
 5035 Proceedings of the 15th International Workshop on Vertex Detectors.
- 5036 [128] Y. Allkofer et al., “Design and performance of the silicon sensors for the {CMS} barrel  
 5037 pixel detector”, *Nucl. Instrum. Meth. A* **584** (2008), no. 1, 25 – 41,  
 5038 doi:<http://dx.doi.org/10.1016/j.nima.2007.08.151>.
- 5039 [129] H. C. Kästli et al., “Design and performance of the CMS pixel detector readout chip”,  
 5040 *Nucl. Instrum. Meth. A* **565** (2006) 188–194, doi:10.1016/j.nima.2006.05.038,  
 5041 arXiv:physics/0511166.
- 5042 [130] E. Bartz, “The 0.25um Token Bit Manager Chip for the CMS Pixel Readout”,  
 5043 Proceedings of the 11th Workshop on Electronics for LHC and Future Experiments  
 5044 CERN-2005-011, CERN-LHCC-2005-038, LHCC-G-101.
- 5045 [131] M. Pernicka et al., “The CMS Pixel FED”, Proceedings of the Topical Workshop on  
 5046 Electronics for Particle Physics CERN-2007-007.
- 5047 [132] K. A. Gill et al., “Progress on the CMS Tracker control system”, Proceedings of the  
 5048 11th Workshop on Electronics for LHC and Future Experiments CERN-2005-011,  
 5049 CERN-LHCC-2005-038, LHCC-G-101.
- 5050 [133] B. G. Taylor, “Timing distribution at the LHC”, Proceedings of the 8th Workshop on  
 5051 Electronics for LHC Experiments CERN-2002-003, CERN-LHCC-2002-034,  
 5052 CERN-LHCC-G-014.
- 5053 [134] J. Troska et al., “Optical readout and control systems for the CMS tracker”, *IEEE  
 5054 Transactions on Nuclear Science* **50** (Aug, 2003) 1067–1072,  
 5055 doi:10.1109/TNS.2003.815124.
- 5056 [135] P. Placidi, A. Marchioro, and P. Moreira, “CMS Tracker PLL Reference Manual”.  
 5057 <http://cmstrackercontrol.web.cern.ch/CMSTrackerControl/documents/PauloMoreira/TrackerPLLManual.pdf>, 2000. CERN EP Microelectronics Group.
- 5059 [136] P. M. H. Correia, A. Marchioro and J. Schrader, “Delay25 - A 4 channel 1/2 ns  
 5060 programmable delay line”.  
 5061 <http://proj-delay25.web.cern.ch/proj-delay25/manuals/manual.pdf>, 2005.  
 5062 CERN EP Microelectronics Group.
- 5063 [137] C. Paillard, C. Ljuslin, and A. Marchioro, “The CCU25: a network oriented  
 5064 communication and control unit integrated circuit in a  $0.25\mu\text{m}$  CMOS technology”,  
 5065 Proceedings of the 8th Workshop on Electronics for LHC Experiments  
 5066 CERN-2002-003, CERN-LHCC-2002-034, CERN-LHCC-G-014.
- 5067 [138] V. Brigljevic et al., “Using XDAQ in application scenarios of the CMS experiment”,  
 5068 *eConf* **C0303241** (2003) MOGT008, arXiv:hep-ex/0305076.

- 5069 [139] CMS Collaboration, “A library developed for the CMS data acquisition system at  
5070 CERN”. <http://cmsdoc.cern.ch/~cschwick/software/documentation/HAL/>.
- 5071 [140] N. Wermes, L. Rossi, P. Fischer, and T. Rohe, “Pixel Detectors, From Fundamentals  
5072 to Applications”. Springer-Verlag, 2006.
- 5073 [141] D. Kotliński, “Status of the CMS Pixel detector”, *Journal of Instrumentation* **4**  
5074 (2009), no. 03, P03019.
- 5075 [142] A. Dominguez et al., “CMS Technical Design Report for the Pixel Detector Upgrade”,  
5076 Technical Report CERN-LHCC-2012-016. CMS-TDR-11, Sep, 2012.
- 5077 [143] CMS Collaboration, “Pilot system for the Phase 1 pixel upgrade of CMS”, *PoS*  
5078 **VERTEX2015** (2015) 018.
- 5079 [144] L. Feld et al., “The DC-DC conversion power system of the CMS Phase-1 pixel  
5080 upgrade”, *Journal of Instrumentation* **10** (2015), no. 01, C01052.
- 5081 [145] H. K'astli, “Frontend electronics development for the CMS pixel detector upgrade”,  
5082 *Nucl. Instrum. Meth. A* **731** (2013) 88 – 91,  
5083 doi:<http://dx.doi.org/10.1016/j.nima.2013.05.056>. Proceedings of the 8th  
5084 International Workshop on Semiconductor Pixel Detectors for Particles and Imaging.
- 5085 [146] CMS Collaboration Collaboration, “High rate capability and radiation tolerance of the  
5086 new CMS pixel detector readout chip PROC600”, Technical Report CMS-CR-2016-269,  
5087 CERN, Geneva, Oct, 2016.
- 5088 [147] J. Troska et al., “Prototype pixel optohybrid for the CMS phase 1 upgraded pixel  
5089 detector”, *Journal of Instrumentation* **7** (2012), no. 01, C01113.
- 5090 [148] L. M. Caminada, “Study of the Inclusive Beauty Production at CMS and Construction  
5091 and Commissioning of the CMS Pixel Barrel Detector”. PhD thesis, Zurich, ETH,  
5092 2010.