



VRIJE  
UNIVERSITEIT  
BRUSSEL



1

Graduation thesis submitted in partial fulfilment of blah

# SOMETHING ABOUT HIGGS+CHARM

my thesis subtitle

2

Felix Heyen

March 2026

3

Promotors: prof. dr. Michael Tytgat prof. dr. Gerrit Van Onsem

sciences and bioengineering sciences



# **Abstract**

<sup>4</sup> My abstract



# Contents

6	<b>Abstract</b>	iii
8	<b>1 Introduction</b>	1
9	1.1 The Standard Model of particle physics . . . . .	1
10	1.2 The Higgs-charm Yukawa coupling . . . . .	3
11	1.2.1 The Brout-Englert-Higgs mechanism . . . . .	6
12	1.2.2 The Yukawa couplings . . . . .	7
13	1.3 Measuring the charm quark Yukawa coupling . . . . .	8
14	1.3.1 The cH process . . . . .	8
15	1.3.2 The $\kappa$ -framework . . . . .	10
16	1.4 An EFT interpretation of the cH process . . . . .	11
17	1.4.1 The chromomagnetic dipole operator . . . . .	13
18	1.4.2 Validity of an EFT . . . . .	13
19	<b>2 The CMS experiment at the LHC</b>	17
20	2.1 The CMS detector . . . . .	17
21	2.1.1 The CMS coordinate system . . . . .	17
22	2.1.2 The silicon tracker . . . . .	19
23	2.1.3 The electromagnetic calorimeter . . . . .	20
24	2.1.4 The hadronic calorimeter . . . . .	21
25	2.1.5 The superconducting solenoid magnet . . . . .	21
26	2.1.6 The muon chambers . . . . .	22
27	2.1.7 The triggering system . . . . .	23
28	2.1.8 The 2018 dataset of the CMS detector . . . . .	24
29	2.2 Event reconstruction with the CMS detector . . . . .	24
30	2.2.1 Track and vertex reconstruction . . . . .	24
31	2.2.2 The Particle Flow algorithm . . . . .	24
32	2.2.3 Reconstruction and identification of muons . . . . .	25
33	2.2.4 Reconstruction and identification of jets . . . . .	27
34	2.2.5 Missing transverse momentum . . . . .	28
35	2.3 Identification of charm quark-induced jets . . . . .	29
36	2.3.1 Properties of heavy-flavour jets . . . . .	29
37	2.3.2 The DeepJet algorithm . . . . .	30
38	<b>3 Search for the cH(<math>ZZ \rightarrow 4\mu</math>) process</b>	33
39	3.1 cH event selection . . . . .	33
40	3.1.1 Higgs candidate selection . . . . .	34
41	3.1.2 Jet candidate selection . . . . .	34

42	3.1.3	Reconstruction efficiency . . . . .	36
43	3.2	Signal and background estimation . . . . .	36
44	3.2.1	Monte Carlo simulation of proton-proton collisions . . . . .	36
45	3.2.2	Estimation of cH( $ZZ \rightarrow 4\mu$ ) process . . . . .	38
46	3.2.3	Estimation of irreducible backgrounds . . . . .	42
47	3.2.4	Estimation of reducible backgrounds . . . . .	43
48	3.3	Validation of cH( $ZZ \rightarrow 4\mu$ ) event selection in sideband regions . . . . .	48
49	3.4	Statistical inference . . . . .	48
50	3.4.1	Statistical model . . . . .	48
51	3.4.2	Uncertainty model . . . . .	49
52	3.5	Results . . . . .	51
53	<b>4</b>	<b>An EFT interpretation of the cH(<math>ZZ \rightarrow 4\mu</math>) process</b>	<b>53</b>
54	<b>Conclusion</b>		<b>55</b>
55	<b>Bibliography</b>		<b>57</b>
56	<b>Appendix</b>		<b>63</b>
57	.1	Columnar analysis framework . . . . .	63
58	.2	Jet selection likelihood templates . . . . .	63
59	.3	Jet-parton matching efficiency . . . . .	63

60 **Chapter 1**

61 **Introduction**

62 The Standard Model (SM) of particle physics is the theory that best describes our current understanding of fundamental particles and their interactions. It describes a broad range of phenomena and makes a plethora of predictions, many of which have been confirmed via measurement to great degrees of accuracy [1]. A notable feature of the SM is the Brout-Englert-Higgs (BEH) mechanism [2][3], which predicts the existence of a Brout-Englert-Higgs (or often simply Higgs) boson. The BEH mechanism is considered a central part of the SM as it provides a unique mechanism by which SM particles may acquire mass through their interaction with the Higgs boson. As such, the experimental discovery of a Higgs-like scalar boson in 2012 [4][5] was a major milestone in particle physics. Since this discovery, a significant open question that remains is whether this particle indeed behaves entirely in an SM-like way. Measuring the exact properties of the discovered scalar particle has thus been a major feature of LHC experiments such as the CMS collaboration [6]. A significant subset of these properties are the so-called Yukawa interactions between the Higgs boson and massive fermions. As can be seen in Figure 1.1, a number of these have previously been measured and indeed align with the values expected from the SM. However, the measurement of the Yukawa couplings of several of the lighter fermions still remain an open challenge as these couplings decrease in strength with smaller fermion masses.

78 The next lightest fermion candidate for such a measurement is the charm quark. Consequentially, the study of the Yukawa-coupling between the Higgs boson and the charm quark is of significant interest [7]. Apart from a brief discussion of the SM, this section introduces the charm-Yukawa coupling. Additionally, LHC processes that may be targeted to exploit their sensitivity to the Higgs-charm Yukawa coupling with an experiment such as the CMS detector are discussed.

84 **1.1 The Standard Model of particle physics**

85 The SM is formulated through the formalism of Quantum Field Theory (QFT). This is a formalism that combines concepts of classical field theory, quantum mechanics as well as special relativity into a single, coherent description of fundamental particles as excitations of underlying fields that pervade space-time. In this description, SM particles fall into two categories: fermions and bosons. The former are the massive particles which make up the matter of the universe while the latter are the force-carrying particles of the strong and electro-weak forces. The distinction between these categories is made based on the spin of the particle, which may be of either half-integer or integer respectively.

93

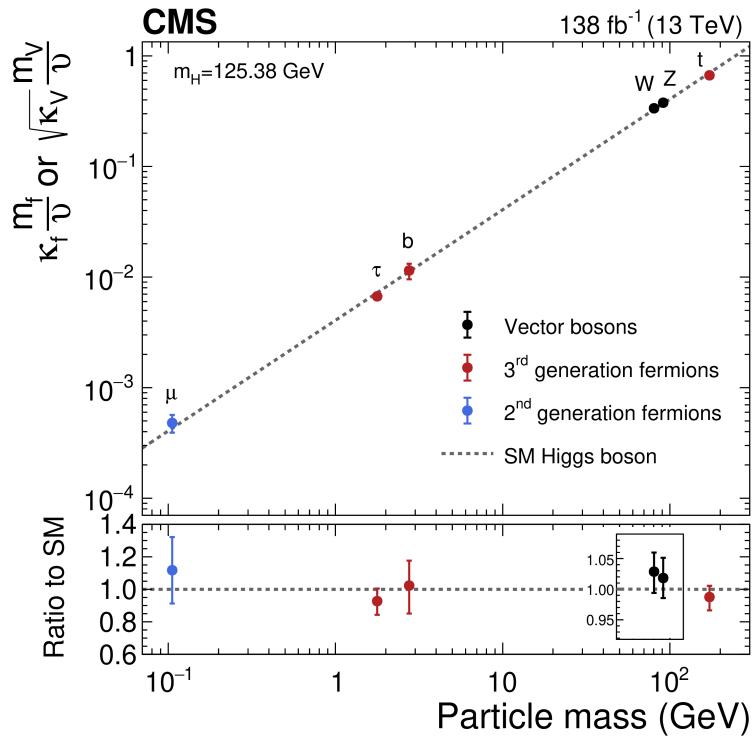


Figure 1.1: The measured coupling modifiers  $\kappa_f$  and  $\kappa_V$  of the coupling between the Higgs boson and fermions as well as heavy gauge bosons as functions of fermion or gauge boson mass  $m_f$  and  $m_V$ , where  $\nu$  is the vacuum expectation value of the Higgs field. [6]

94 The fermion content of the SM consists of 12 unique particles. These include six leptons, namely  
 95 the electron, muon and tau as well as their respective neutrinos as well as six different quarks  
 96 that are distinguished by their so-called flavour. The different quark flavours include up, down,  
 97 charm, strange, bottom and top and specifies a quark's mass eigenstate as well as electric charge.  
 98 These fermions are typically arranged into three generations typically depicted as

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}, \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}, \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}, \begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}. \quad (1.1)$$

99  
 100 However, there are distinct differences between the leptons and quarks. Leptons carry integer  
 101 (or no) charge while quarks carry fractional charges. More importantly, while both quarks and  
 102 leptons may interact via the electro-weak force, only the quarks interact via the strong force.  
 103 Due to the nature of the strong force, quarks almost exclusively form composite states called  
 104 hadrons. Lastly, the existence of anti-fermions must be mentioned. These carry the exact op-  
 105 posite quantum numbers (e.g. charge) as their fermion counterparts, though otherwise behave  
 106 similarly (take the electron and positron for instance). For simplicity, references to a fermion  
 107 in this work may be understood as referencing both the fermion and anti-fermion counterpart,  
 108 unless otherwise explicitly indicated. Examples of the latter are e.g. referring explicitly to elec-  
 109 trons  $e^-$  and positrons  $e^+$  or charm quark  $c$  and anti-charm quark  $\bar{c}$  pairs.

110  
 111 There exist 13 unique bosons in the SM. These include the photon  $\gamma$ ,  $W^\pm$  and  $Z$  which me-  
 112 diate the electro-weak force as well as 8 gluons  $g$  that mediate the strong force. The final piece  
 113 is the Higgs boson. Contrary to the force carriers, which all are spin 1, the Higgs boson is spin  
 114 0. By interacting with the Higgs boson, the massive particles of the SM acquire their mass and  
 115 is thus a central element of the SM.

116  
 117 Considering the introduced particles and forces, the SM has a rich and detailed phenome-  
 118 ology. A great example of a mathematically rigorous delineation of this can be found for example  
 119 in [8]. Given the focus of this work on the Yukawa coupling between the Higgs boson and charm  
 120 quark, only this aspect of the SM is discussed in further detail.

## 121 1.2 The Higgs-charm Yukawa coupling

122 The coupling that defines the strength of the interaction between massive fermions and the Higgs  
 123 boson is the so-called Yukawa coupling. To better understand this and associated concepts, some  
 124 knowledge of the electro-weak sector of the SM is required. These are discussed in this section  
 125 while a comprehensive overview may be found in [9].

126  
 127 To understand the origin of the Yukawa-couplings, a brief discussion of Lagrangian densities,  
 128 gauge transformations and the role of symmetries in the SM is warranted. The Lagrangian  
 129 density  $\mathcal{L}(\phi_i; a_i)$  is a quantity dependent on a set of fields  $\phi_i$  and constants  $a_i$  from which the  
 130 equations of motions for the particles associated with these fields may be derived. Commonly,  
 131 theories of particles and their behaviour in a QFT are thus defined through the formulation of a  
 132 Lagrangian density. The form of this expression determines the nature of the particles that are  
 133 included as well as their interactions.

134  
 135 A central component to the way in which particle interactions are introduced in the SM is  
 136 the concept of gauge symmetries. These originate from the fact that the quantum fields in a

- <sup>137</sup> QFT carry phase information, which may depend on the space-time coordinate of the field. This  
<sup>138</sup> phase information describes (local) degrees of freedom of the field and should have no effect on  
<sup>139</sup> the physical observables of the system. Thus,  $\mathcal{L}$  should remain invariant under arbitrary phase  
<sup>140</sup> transformations. Such transformations are typically referred to as a choice of gauge and such an  
<sup>141</sup> invariance is accordingly referred to as a *local gauge symmetry*.
- <sup>142</sup>
- <sup>143</sup> In the Lagrangian of the SM, invariance in the presence of local gauge symmetries is insured  
<sup>144</sup> through the addition of additional fields. These gauge fields couple to the previously existing  
<sup>145</sup> fields and effectively serve as mediators of phase information between space-time points of the  
<sup>146</sup> original fields. It is exactly these gauge fields which we identify as the fields force-mediating  
<sup>147</sup> bosons introduced previously and which are required to maintain local gauge symmetry. A very  
<sup>148</sup> interesting conclusion from this is that the dynamics of the bosons and the corresponding force  
<sup>149</sup> are determined entirely by the structure of the local gauge symmetry that must be preserved.  
<sup>150</sup> For the electro-weak force, the corresponding symmetry is referred to as  $SU(2)_L \times U(1)_Y$ . Here,  
<sup>151</sup> the  $L$  denotes that the associated force only acts on left-handed chiral particles while the  $Y$   
<sup>152</sup> denotes the charge that is carried by the corresponding bosons and is referred to as the weak  
<sup>153</sup> hypercharge. There are a total of four boson associated with the electro-weak force. These are  
<sup>154</sup> the photon  $\gamma$  that mediates the electromagnetic force as well as the electromagnetically charged  
<sup>155</sup>  $W^\pm$  and electromagnetically neutral Z boson that mediate the weak force.
- <sup>156</sup>
- <sup>157</sup> With these concepts in mind the nature of the electro-weak sector's Lagrangian in the SM  
<sup>158</sup> may be discussed. Naively, the form of this would be given by

$$\mathcal{L}_{EW} = i\bar{\psi}_L \gamma^\mu D_\mu^L \psi_L + i\bar{\psi}_R \gamma^\mu D_\mu^R \psi_R - \frac{1}{2} \text{Tr} (W_{\mu\nu}^a W^{a\mu\nu}) - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}. \quad (1.2)$$

for a generic combination of a left-handed isospin doublet  $\psi_L$  and right-handed isospin singlet  $\psi_R$ . The individual elements of  $\mathcal{L}_{EW}$  are briefly summarised below

$g'$ :	coupling constant of $U(1)_Y$
$g$ :	coupling constant of $SU(2)_L$
$\psi_L$ ,	left-handed isospin doublet
$\psi_R$ ,	right-handed isospin doublet
$B_\mu$ :	gauge field of $U(1)_Y$
$W_\mu^a$ :	gauge fields of $SU(2)_L$ , $a = 1, 2, 3$
$W_{\mu\nu}$ :	field strength tensor
$B_{\mu\nu}$ :	field strength tensor
$t^a = \frac{\sigma^a}{2}$ ,	$SU(2)$ generators
$Y_L = -1$ ,	left chiral hypercharge
$Y_R = -2$ ,	right chiral hypercharge
$D_\mu^L = \partial_\mu + ig' \frac{Y_L}{2} B_\mu + igt^a W_\mu^a$	
$D_\mu^R = \partial_\mu + ig' \frac{Y_R}{2} B_\mu$	

The terms  $D_\mu^{L/R}$  are so-called covariant derivatives that ensure the local  $SU(2)_L \times U(1)_Y$  gauge symmetry is upheld for  $\mathcal{L}_{EW}$ . In this formulation, the observed charged gauge bosons  $W^\pm$  arise from linear combinations of the  $W_1$  and  $W_2$  gauge fields

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2), \quad (1.3)$$

while the Z boson and photon  $\gamma$  arise from linear combinations of the  $W_3$  and  $B$  gauge fields achieved via a rotation

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix}. \quad (1.4)$$

with the weak mixing angle  $\theta_W$ .

The massive natures of the  $W^\pm$  and Z bosons, as first reported in [10], are however incompatible with such a formulation. This is as naive mass term such as

$$m_W^2 W_\mu^+ W^{-,\mu} + \frac{1}{2} m_Z^2 Z_\mu Z^\mu. \quad (1.5)$$

do not remain invariant under arbitrary  $SU(2)_L$  gauge transformations. This is as gauge fields  $A_\mu$  generically transform as

$$A_\mu \rightarrow A'_\mu = A_\mu - \frac{1}{g} \partial_\mu \mathcal{V}(x) \quad (1.6)$$

where  $\mathcal{V}(x)$  is some arbitrary phase. Substituting Equation 1.6 into Equation 1.5 thus introduces additional terms that do not cancel. The same is true for fermion mass terms in the form of

$$m_f \bar{\psi} \psi. \quad (1.7)$$

<sup>166</sup>  
<sup>167</sup> There is however a subtle distinction in this case, as the invariance breaking terms in Equation 1.7  
<sup>168</sup> arise from the different transformation behaviour of the  $\psi_L$  and  $\psi_R$  components of  $\psi$  under  
<sup>169</sup>  $SU(2)_L \times U(1)_Y$  gauge transformations.

### <sup>170</sup> 1.2.1 The Brout-Englert-Higgs mechanism

The BEH mechanism provides a way to circumvent the gauge symmetry breaking nature of the aforementioned generic mass terms. This is achieved through a process referred to as spontaneous symmetry breaking. A spontaneously broken symmetry refers to a symmetry that is upheld in a global view of the system (i.e. the overall Lagrangian density  $\mathcal{L}_{EW}$  remains invariant under a relevant gauge transformation) while the energetic ground state of the system explicitly breaks this symmetry. This is a process formally described by the Goldstone theorem [11] that states that each broken symmetry in a relativistic QFT generates an additional massless boson. These introduce additional degrees of freedom into the theory and are coined Goldstone bosons. The BEH mechanism exploits this by adding an additional term

$$\mathcal{L}_{\text{Higgs}} = D_\mu \phi^\dagger D^\mu \phi - V(\phi) \quad (1.8)$$

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (1.9)$$

to  $\mathcal{L}_{EW}$  with the complex field  $\phi$ . This is a  $SU(2)_L$  doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (1.10)$$

with the scalar components  $\phi^+$  and  $\phi^0$ . Here,  $V(\phi)$  corresponds to the potential energy term of the field. Again, the covariant derivative

$$D_\mu = \partial_\mu + ig' \frac{Y_\phi}{2} B_\mu + ig t^a W_\mu^a \quad (1.11)$$

ensures  $\mathcal{L}_{\text{Higgs}}$  remains locally gauge invariant under  $SU(2)_L \times U(1)_Y$  transformations. The constants of the potential term Equation 1.9 are chosen in such a way that the ground state of  $V(\phi)$  is non-zero. This can be achieved by choosing them such that  $\lambda > 0$  and  $\mu^2 > 0$ . The result is a ground state of  $V$  that is identified as the vacuum expectation value

$$v = \sqrt{\frac{\mu^2}{2\lambda}}. \quad (1.12)$$

<sup>171</sup>  
<sup>172</sup> The center of the potential is now an unstable local maximum and the only stable configuration

<sup>173</sup> can be found in the non-zero ground state. Through this, the symmetry of the potential is  
<sup>174</sup> effectively broken. A popular choice of gauge for  $\phi$  is

$$\phi = \begin{pmatrix} 0 \\ v + \frac{h}{\sqrt{2}} \end{pmatrix} \quad (1.13)$$

<sup>175</sup>  
<sup>176</sup> where  $h$  is a new scalar field that is used to parametrise radial perturbations of the potential's  
<sup>177</sup> ground state. This choice is referred to as the unitary gauge and  $h$  is identified as the field  
<sup>178</sup> corresponding to the physical Higgs boson. By expanding Equation 1.8 with this choice of  $\phi$ , a  
<sup>179</sup> range of terms are introduced to  $\mathcal{L}_{\text{EW}}$ . These contain a variety of interaction terms between the  
<sup>180</sup> gauge fields and the Higgs field, as well as newly generated mass terms for the Z and W bosons

$$\left(\frac{g}{2}\right)^2 v^2 W_\mu^+ W^\mu_- = m_W^2 W_\mu^+ W^\mu_- \quad (1.14)$$

$$\left(\frac{\sqrt{g^2 + g'}}{2}\right)^2 v^2 Z_\mu Z^\mu = m_Z^2 Z_\mu Z^\mu. \quad (1.15)$$

<sup>181</sup>  
<sup>182</sup> This can be understood to mean that the electro-weak coupling constants  $g$  and  $g'$  along with  $v$   
<sup>183</sup> effectively determine the mass of the Z and  $W^\pm$  bosons. A full description and compilation of  
<sup>184</sup> all the terms of the electro-weak Lagrangian density of the SM can be found in [9].

### <sup>185</sup> 1.2.2 The Yukawa couplings

<sup>186</sup> By including the Higgs contribution in our theory, mass terms for fermions may now be generated  
<sup>187</sup> by including a term of the form

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi} \phi \psi, \quad (1.16)$$

$$= -y_f v \bar{\psi} \psi \left(1 + \frac{1}{v} \frac{h}{\sqrt{2}}\right) \quad (1.17)$$

which is invariant under  $SU(2)_L \times U(1)_Y$  gauge transformations due to the addition of  $\phi$ . Similarly to the W and Z mass terms, the relation

$$m_f = y_f v. \quad (1.18)$$

<sup>188</sup>  
<sup>189</sup> is obtained. A curious feature of the SM is that the Yukawa-couplings  $y_f$  are free parameters  
<sup>190</sup> of the theory with no a priori values. As a result these must be measured experimentally, with  
<sup>191</sup> the measurement of the charm quark Yukawa coupling  $y_c$  being the goal of this work. Since the  
<sup>192</sup> charm quark mass has previously been determined from experiment to be  $m_c = 1.27$  GeV [1],  
<sup>193</sup> a measurement of  $y_c$  thus represents an important consistency test of the SM. To this end, one  
<sup>194</sup> can exploited that an interaction between fermions and the Higgs field is introduced as can be  
<sup>195</sup> seen in Equation 1.17, with an interaction strength proportional to  $y_c$ . It is exactly this feature  
<sup>196</sup> that may be exploited by experiments at the LHC to measure  $y_c$ .

<sup>197</sup> **1.3 Measuring the charm quark Yukawa coupling**

<sup>198</sup> By measuring the frequency of occurrence of physics processes in which the coupling between the  
<sup>199</sup> Higgs boson and charm quark appears,  $y_c$  may be determined. As such, a suitable process must  
<sup>200</sup> be found that can be detected by an experiment such as CMS. These fall into two categories. The  
<sup>201</sup> first consists of processes in which a Higgs boson decays into a charm and anti-charm quark pair  
<sup>202</sup> ( $H \rightarrow c\bar{c}$ ). Previous analysis of e.g. top quark pair and vector boson associated Higgs production  
<sup>203</sup> has been able to observe a 95% CL upper limit on the charm quark Yukawa coupling modifier  
<sup>204</sup>  $\kappa_c$  (see subsection 1.3.2 for a detailed discussion) of  $|\kappa_c| < |3.5|$  [12], the most stringent limit to  
<sup>205</sup> date. The second category consists of processes in which a Higgs boson is produced in association  
<sup>206</sup> with a charm quark. This latter category of processes is the focus of this work and is henceforth  
<sup>207</sup> referred to as the cH process.

<sup>208</sup> **1.3.1 The cH process**

<sup>209</sup> The cH process encompasses processes in proton-proton collisions in which a charm-quark is  
<sup>210</sup> produced alongside a Higgs boson. At leading order, this consists of 2 processes sensitive to  $y_c$ ,  
<sup>211</sup> represented by the Feynman diagrams shown in Figure 1.3. The first two diagrams, namely the s  
<sup>212</sup> and t-channel diagrams, constitute the  $y_c$  sensitive contribution. There exist also additional cH  
<sup>213</sup> processes, mediated through the effective Higgs boson to gluon coupling, which are not sensitive  
<sup>214</sup> to  $y_c$ . These account for approximately 80% of the inclusive cH cross section and thus represents  
<sup>215</sup> a significant background to the cH process sensitive to the charm quark Yukawa coupling.

<sup>216</sup> Targeting the cH process to measure  $y_c$  is a relatively novel strategy in comparison to targeting  
<sup>217</sup>  $H \rightarrow c\bar{c}$ . A key advantage of this approach is that contributions from the abundant QCD  
<sup>218</sup> background at the LHC are greatly reduced due to only needing to identify the flavour of single  
<sup>219</sup> jet resulting from a charm quark, as opposed to two. Additionally, since the sensitivity to  $y_c$  does  
<sup>220</sup> not originate from the decay of the Higgs boson, the Higgs boson decay mode to target can be  
<sup>221</sup> chosen freely. Especially signatures such as  $H \rightarrow ZZ \rightarrow 4\mu$ , which may be resolved cleanly by an  
<sup>222</sup> experiment such as CMS, can be targeted. However, an analysis of the cH process also comes  
<sup>223</sup> with drawbacks. A significant experimental difficulty results from the fact that the associated  
<sup>224</sup> charm flavour jets are typically produced at lower transverse momenta  $p_T$ , as seen in Figure 1.2.  
<sup>225</sup> These can be experimentally difficult to reconstruct and thus a significant portion of this signal  
<sup>226</sup> may be lost due to detector acceptance effects. Another drawback is that Higgs boson decay  
<sup>227</sup> channels such as  $H \rightarrow ZZ \rightarrow 4\mu$  have very small branching ratios (e.g.  $BR(H \rightarrow ZZ \rightarrow 4\mu) = 0.3\%$  [1])  
<sup>228</sup> and thus the overall cross section of the cH process may be very small. As a result of these effects,  
<sup>229</sup> a key challenge of a search for the cH process is expected to lie in the statistical uncertainty of  
<sup>230</sup> the analysis.

<sup>232</sup> As a novel strategy, targeting the cH process is of recent interest and results in the cH(WW)  
<sup>233</sup> and cH( $\gamma\gamma$ ) channels using Run 2 data of the CMS experiment are published. Upper limits on  
<sup>234</sup>  $\kappa_c$  at 95% CL are reported with  $|\kappa_c^{\text{cH}(WW)}| < 47$  [13] and  $|\kappa_c^{\text{cH}(\gamma\gamma)}| < 38.1$  [14]. While not  
<sup>235</sup> as sensitive as the limit observed in the  $H \rightarrow c\bar{c}$  channels, these nonetheless provide important  
<sup>236</sup> complementary results and can contribute significantly in combination. This is especially impor-  
<sup>237</sup> tant given that even at the High-Luminosity LHC, the projected sensitivity to the charm quark  
<sup>238</sup> Yukawa coupling in individual channels is only starting to approach one [15].

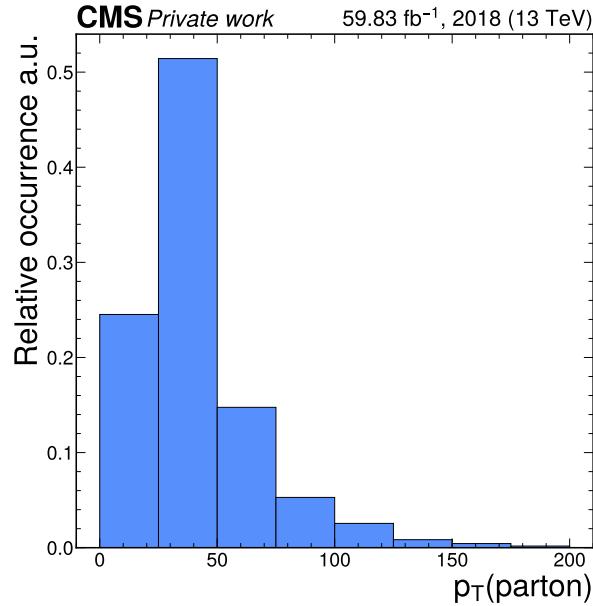


Figure 1.2: Transverse momentum of the parton produced alongside a Higgs boson in a simulation of the cH process, which typically takes on relatively small values.

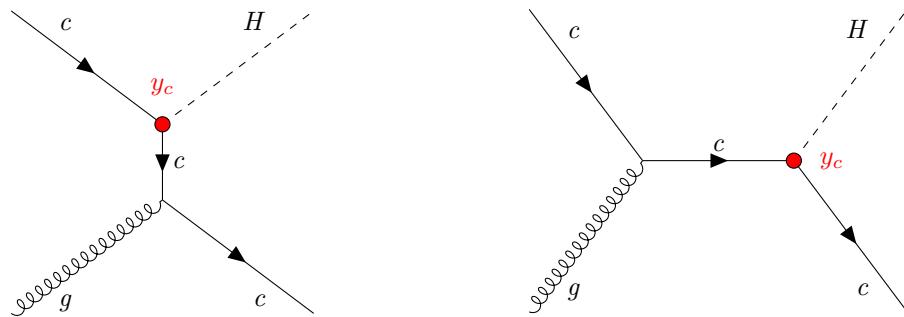


Figure 1.3: The leading order cH processes through which  $y_c$  may be probed as each diagram contains a vertex with a charm-quark and Higgs boson, here denoted in red. The corresponding diagrams with an anti-charm quark  $\bar{c}$  are implied.

<sup>240</sup> **1.3.2 The  $\kappa$ -framework**

The  $\kappa$ -framework [16] is a tool to parametrise modifications to couplings between the Higgs boson and other particles with respect to the expected SM values of the couplings. For example, the coupling modifiers for the charm quark Yukawa coupling is introduced as

$$\kappa_f = \frac{y_f}{y_f^{\text{SM}}}. \quad (1.19)$$

where  $y_f$  is the measured Yukawa-coupling and  $y_f^{\text{SM}}$  is the expected Yukawa-coupling of the SM, calculated from the known charm quark mass. Thus modifications to the Yukawa-coupling of the charm quark are parametrised in this way as deviations from  $\kappa_c = 1$ . However,  $y_c$  is not a quantity that can be measured directly. Instead a signal strength measurement  $\mu_{if}$ , where  $i$  represents the production process and  $f$  represents the decay process, relative to the SM expectation is made. Thus a measurement of  $\mu_{if}$  must be converted into an interpretation of  $\kappa_c$ . This is a step that contains some finer subtleties.

The rate of a Higgs production and decay process in relation to the expected SM signal (i.e. a signal strength) may be written as

$$\mu_{if} = \frac{\sigma_i \cdot \text{BR}_f}{(\sigma_i \cdot \text{BR}_f)^{\text{SM}}}, \quad (1.20)$$

where  $\sigma_i$  is the production cross section in a given channel  $i$  and  $\text{BR}_f$  is the decay branching ratio in a given channel  $f$ . This can be rewritten as

$$\sigma_i \cdot \text{BR}_f = \kappa_{r,i} \sigma_i^{\text{SM}} \cdot \frac{\kappa_f \Gamma_f^{\text{SM}}}{\Gamma_H} \quad (1.21)$$

to give a general expression in which modifications to the production cross section and partial SM decay width  $\Gamma_f^{\text{SF}}$  are introduced via  $\kappa_{r,i}$  and  $\kappa_f$  respectively. The denominator  $\Gamma_H$  represents the total decay width which can be written as

$$\begin{aligned} \Gamma_H &= \Gamma_H^{\text{SM}} (\kappa_b^2 \text{BR}_{bb}^{\text{SM}} + \kappa_W^2 \text{BR}_{WW}^{\text{SM}} + \kappa_g^2 \text{BR}_{gg}^{\text{SM}} + \kappa_\tau^2 \text{BR}_{\tau\tau}^{\text{SM}} + \kappa_Z^2 \text{BR}_{ZZ}^{\text{SM}} + \kappa_c^2 \text{BR}_{cc}^{\text{SM}} \\ &\quad + \kappa_\gamma^2 \text{BR}_{\gamma\gamma}^{\text{SM}} + \kappa_{Z\gamma}^2 \text{BR}_{Z\gamma}^{\text{SM}} + \kappa_s^2 \text{BR}_{ss}^{\text{SM}} + \kappa_\mu^2 \text{BR}_{\mu\mu}^{\text{SM}}) \end{aligned} \quad (1.22)$$

$$:= \Gamma_H^{\text{SM}} \kappa_H^2 \quad (1.23)$$

Here,  $\Gamma_H^{\text{SM}}$  is the SM total decay width of the Higgs boson and  $\text{BR}_f^{\text{SM}}$  are the branching ratios of the possible decay modes (the loop induced coupling of the Higgs boson to gluons and photons are included as independent quantities) where  $\kappa_f$  parametrises modifications thereof. Substituting Equation 1.23 into Equation 1.20, the rate modifier may be written as

$$\mu_{if} = \frac{\kappa_{r,i} \kappa_f^2}{\kappa_H^2}. \quad (1.24)$$

Now, assuming in the production of the Higgs boson only modifications to the charm quark Yukawa coupling plays a role as well as that the decay mode (e.g.  $H \rightarrow ZZ \rightarrow 4\mu$ ) is unmodified, Equation 1.24 becomes

$$\mu_{if} = \frac{\kappa_c^2}{\kappa_H^2} \quad (1.25)$$

Using the flat direction approach discussed in [7] and [17], a simplification of  $\kappa_H$  can be introduced. This approach is based on the finding that, when performing fits to existing Higgs boson production and decay rates, increases in the Yukawa couplings of light quarks (including the charm quark) can be compensated by increases in the couplings of the gauge bosons and heavy fermions. This is referred to as a “flat direction” in the fit, where observed Higgs boson production and decay rates can be modeled equally well for any value of  $\kappa_c$  by a respective scaling of all other processes. The authors thus replace the individual modifiers in the sum of Equation 1.22 with a single modifier  $\kappa$ . This allows Equation 1.24 to be rewritten as

$$\mu_{if} = \frac{\kappa^4}{\kappa^2(1 - BR_{cc}^{SM}) + \kappa_c^2 BR_{cc}^{SM}} \quad (1.26)$$

which has a solution for  $\kappa$  given by

$$\kappa = \frac{(1 - BR_{cc}^{SM})\mu}{2} + \frac{\sqrt{(1 - BR_{cc}^{SM})^2\mu^2 + 4\mu BR_{cc}^{SM}\kappa_c^2}}{2}. \quad (1.27)$$

241  
Here, the expected SM decay width  $BR_{cc}^{SM} = 0.3$  can be substituted. Additionally, the fact that  
243 observed Higgs boson rates have been well measured to be close to their expected values (see e.g.  
244 [18]) can be reflected by setting  $\mu \approx 1$ , so that only a dependence on  $\kappa_c$  remains in the expression.  
245 Thus by replacing  $\kappa_H$  in Equation 1.25 with Equation 1.27, a final expression relating a measured  
246 signal strength of the cH process to  $\kappa_c$  is obtained, given by

$$\mu_{\sigma_{cH} BR(H \rightarrow ZZ)} = \frac{2\kappa_c^2}{0.97 + \sqrt{(0.97)^2 + 4 \cdot 0.97\kappa_c^2}}. \quad (1.28)$$

Rearranging for  $\kappa_c$  gives

$$\kappa_c = \pm \frac{\sqrt{4 \cdot 0.97 \cdot \mu_{\sigma_{cH} BR(H \rightarrow ZZ)} \cdot (1 + \mu_{\sigma_{cH} BR(H \rightarrow ZZ)})}}{2}. \quad (1.29)$$

247  
Effectively, this approach in interpreting  $\kappa_c$  from a signal strength measurement  $\mu_{\sigma_{cH} BR(H \rightarrow ZZ)}$   
248 thus ensures compatibility with existing Higgs boson rate measurements, given a non-unity value  
249 of  $\kappa_c$  leads to modifications of the Higgs boson partial decay widths. It should be noted that  
250 this already indirectly implies bounds on  $\kappa_c$ , as discussed in [7].  
251

## 1.4 An EFT interpretation of the cH process

The cH process may also be interpreted in terms of Standard Model Effective Field Theory (SMEFT). In SMEFT theory, potential effects from physics processes not described by the SM (commonly referred to as beyond-the-SM or BSM physics) are parametrised in a mostly model-independent way. Specifically, the SMEFT framework can be used at colliders with a characteristic energy scale  $E$  to describe the effects of processes with a characteristic energy scale above  $E$ . This concept is illustrated in Figure 1.4.

Formally, SMEFT is a collection of all possible combinations of field interactions that obey the gauge invariance conditions of the SM. Generically, this can be expressed as an expansion in the

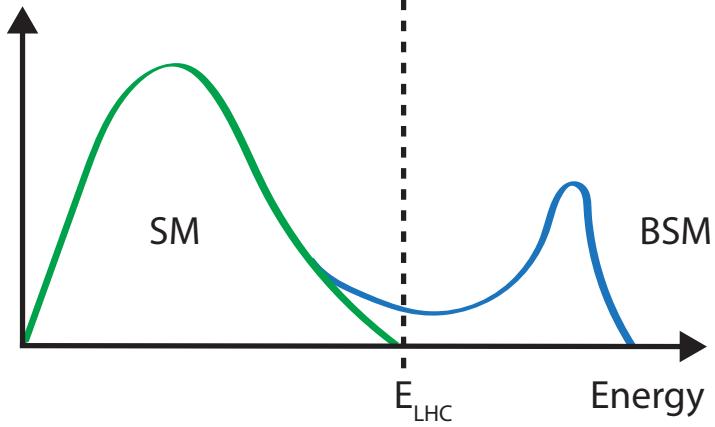


Figure 1.4: Illustration of how the presence of BSM physics, which is primarily visible beyond the reach of current collider energies (e.g.  $E_{\text{LHC}}$ ), can lead to subtle modifications of SM observables. These effects can be parametrised by SMEFT.

energy scale of the new physics scale  $\Lambda$

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{d>4} \sum_i \frac{C_i}{\Lambda^{d-4}} \hat{O}_i^d \quad (1.30)$$

where  $\mathcal{L}_{\text{SM}}$  is the SM lagrangian,  $O_i$  denotes a particular operator (i.e. a particular combination of fields) with a dimensionless coupling coefficient  $C_i$  and  $d$  denotes the dimension of the operator. The dimensionality is derived through a dimensional analysis of a lagrangian and its fields, where energy dimensions of terms may be deduced from the requirement that the action

$$S = \int \mathcal{L} d^4x \quad (1.31)$$

<sup>253</sup>  
<sup>254</sup> remains dimensionless. Accordingly,  $\mathcal{L}_{\text{SM}}$  is of energy dimension four. Since the SMEFT operators  $O_i^d$  all have energy dimensions higher than four and  $\Lambda$  comes with energy dimension one, the terms in the sum of Equation 1.30 are scaled with  $1/\Lambda^{d-4}$  to ensure the combination also has an energy dimension of four.

<sup>255</sup>  
<sup>256</sup>  
<sup>257</sup>  
<sup>258</sup> Typically, operators in SMEFT are grouped by their energy dimension. In  $d=5$ , only one operator possible operator exists that violates lepton number [19] and is not relevant in this work.  
<sup>259</sup> In  $d=6$  however, a plethora of valid operators exist. In total, these amount to 59 different dimension six operators (not counting all possible flavour combinations), commonly represented in the Warsaw basis [20]. Since  $d=7$  operators again violate lepton number and each additional dimension adds a suppressive factor of  $\Lambda^{-1}$ , a simplified SMEFT schema is commonly used in which only the contribution of  $d=6$  operators is considered in the expansion. Thus Equation 1.30 simplifies to

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{C_i}{\Lambda^2} \hat{O}_i^{(6)} \quad (1.32)$$

267

268 A good overview of SMEFT can be found in [21].

269

270 **1.4.1 The chromomagnetic dipole operator**271 A particular operator relevant to this work is referred to as the chromomagnetic dipole (CMD)  
272 operator  $\hat{O}_{qG}$ . For the charm quark, the CMD operator is written as

$$\hat{O}_{cG} = (\bar{q}_{2,L} \sigma^{\mu\nu} T^a c) \tilde{\phi} G_{\mu\nu}^a. \quad (1.33)$$

273

274 Here,  $\bar{q}_{2,L}$  is the second generation, left-handed quark doublet,  $\sigma^{\mu\nu} = i[\gamma_\mu, \gamma_\nu]/2$  with the Dirac  
275 matrices  $\gamma_\mu$ ,  $T^a c$  are the generators of the SU(3),  $\tilde{\phi}$  is the adjoint Higgs doublet and  $G_{\mu\nu}^a$  is the  
276 field strength tensor of the strong interaction. This operator may be uniquely bounded with the  
277 cH process due to the unique chiral structure of the operator, which mixes left and right-handed  
278 spinors, a structure otherwise only found in the Yukawa and quark-Higgs boson interaction terms  
279 of the SM.

280

281 To better understand this, it is worth considering other processes such as inclusive Higgs boson  
282 production, which have been successfully leveraged to set strong constraints on the top quark  
283 CMD operator  $\hat{O}_{tG}$  [22]. Typically, the strategy that is used to probe even small wilson coefficients e.g.  $C_{tG}$  is to exploit interference of the relevant (small) SMEFT contribution with a larger  
284 SM contribution. Though the pure SMEFT contribution itself may be small and experimentally  
285 negligible due to limited analysis sensitivity, the much larger contribution of the SM process it  
286 interferes with can result in a non-negligeable interference effect with respect to the SM process.  
287 However, the chiral structure of the CMD operator influences the effectiveness of this strategy.  
288 Since the  $\hat{O}_{qG}$  operator effectively flips the chirality of the ingoing and outgoing quarks, a second  
289 *chirality flip* must be inserted for the SMEFT contribution to interfere with the SM process. This  
290 is visualised in Figure 1.5. Such a chirality flip is proportional to the mass  $m_q$  of the respective  
291 quark. As a result the interference contribution for a much lighter quark is significantly suppressed  
292 in comparison to the top quark, as also argued for the bottom quark in [23]. Effectively, the  
293 processes that prove effective in targeting  $\hat{O}_{tG}$  due to the large mass of the top quark are thus  
294 much less sensitive to  $\hat{O}_{cG}$ . However, since the cH process itself contains the chirality flipping  
295 quark-Higgs boson vertex, interference terms between the EFT and SM contributions do not  
296 suffer from the above described effect. Furthermore, due to the very low expected cross section  
297 of the cH process, quadratic contributions from  $\hat{O}_{cG}$  may be comparatively large even at small  
298 values of  $C_{cG}$ . Accordingly, the cH process may be an excellent target in constraining  $\hat{O}_{cG}$ .300 **1.4.2 Validity of an EFT**

In addition to EFT terms needing to satisfy the gauge invariance conditions of the SM, two additional key validity conditions are typically required of an EFT. The first is related to the fact that in an EFT, the particle nature of e.g. new, heavy mediator particles is simplified into the introduction of a new effective vertex. For example, a 2→2 particle resonant scattering via

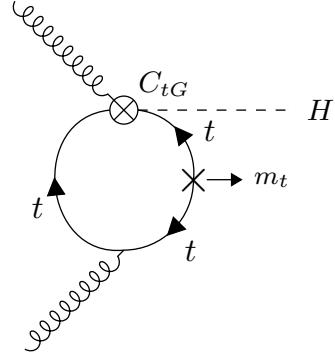


Figure 1.5: A modification of the gluon fusion process with a top quark loop by including the vertex introduced by the top quark CMD operator. Note that the arrows indicate chirality and not momentum flow. A chirality flip, denoted by the cross, proportional to the top quark mass  $m_t$  is required for the inclusion of the top quark CMD vertex.

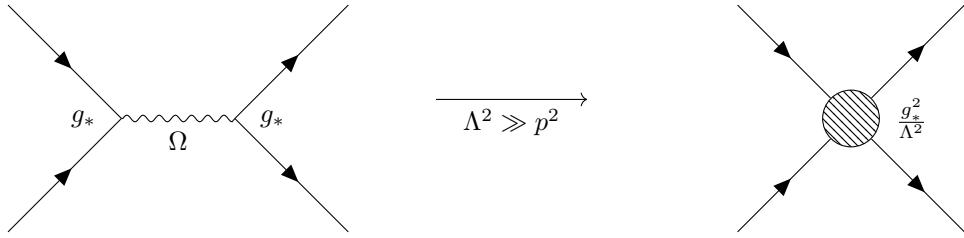


Figure 1.6: Feynman diagrams depicting a resonant process in which the new mediator particle  $\Omega$  is created (left) and the approximate description of this in an EFT, where the diagram is reduced to a four-point interaction.

a new heavy mediator particle  $\Omega$  with a newly introduced coupling constant  $g_*$  is simplified via the introduction of a four-point interaction, as visualised in Figure 1.6. This corresponds to a first order approximation of the new particle's mediator as

$$\frac{g_*}{p^2 - m_\Omega} \xrightarrow{m_\Omega^2 \gg p^2} -\frac{g_*}{m_\Omega} \left( 1 + \frac{p^2}{m_\Omega^2} + \frac{p^4}{m_\Omega^4} + \dots \right) \approx -\frac{g_*}{m_\Omega} \quad (1.34)$$

For the EFT description of this simplification to be valid, the energy involved in processes containing the effective vertex introduced by the relevant operator must thus lie well below  $m_\Omega$ , which represents the previously introduced new physics scale  $\Lambda$ . Practically, this can be achieved by placing an upper limit  $M_{\text{cut}}$  on the total energy that is considered in measurements of such processes. The requirement can be expressed as

$$M_{\text{cut}} < \Lambda. \quad (1.35)$$

A good estimator of  $M_{\text{cut}}$  is the invariant mass of the final state particles of a process. In case of the  $cH$  process, the invariant mass of the Higgs boson and jet system is a natural choice.

The second condition that must be met is related to the perturbativity of the theory. Concretely,

this means that higher dimensional operators should contribute increasingly smaller corrections so that the sum of operator contributions converges. In the case of this work where only d=6 operators are considered, this means ensuring contributions from d=8 operators and higher are sufficiently small. While this cannot be determined with certainty without explicit knowledge of the underlying theory the EFT is estimating, a popular choice is to require that at most  $g_* \sim 4\pi$  [24].

These two conditions may be combined into a single, simultaneous requirement. In [24] an effective lagrangian (ignoring relevant indices for simplicity) of the general form

$$\mathcal{L}_{\text{eff}} = \frac{\Lambda^4}{g_*^2} \mathcal{L} \left( \frac{D_\mu}{\Lambda}, \frac{g_h \phi}{\Lambda}, \frac{g_{\psi_{L,R}} \psi_{L,R}}{\Lambda^{3/2}}, \frac{g F_{\mu\nu}}{\Lambda^2} \right) \quad (1.36)$$

is obtained when a single BSM coupling  $g_*$  is introduced. This provides a prescription for the powers of the couplings and  $\Lambda$  that are associated with the SM fields  $\phi, \psi$  and  $F_{\mu\nu}$ , and the covariant derivate  $D_\mu$ . Here,  $g$  represents the unaltered gauge field couplings of the SM, while  $g_{\psi_{L,R}}$  and  $g_h$  represent the coupling of SM fermion and the Higgs doublet to the BSM theory. In a single BSM coupling scenario, this simplifies to  $g_{\psi_{L,R}} = g_h = g_*$ . Applying this prescription to the CMD operator gives

$$\hat{O}_{cG} \longrightarrow \frac{\Lambda^4}{g_*^2} \left[ \left( \frac{g_* \psi_{L,R}}{\Lambda^{3/2}} \right) \cdot \left( \frac{g_* \psi_{L,R}}{\Lambda^{3/2}} \right) \cdot \left( \frac{g_* \phi}{\Lambda} \right) \cdot \left( \frac{g_s G}{\Lambda^2} \right) \right] \quad (1.37)$$

$$= \frac{g_* g_s}{\Lambda^2} (\psi_{L,R} \cdot \psi_{L,R} \cdot \phi \cdot G) . \quad (1.38)$$

Reading off from Equation 1.38, one can see that the coupling of the CMD operator is given by  $g_* g_s / \Lambda^2$ . Comparing to Equation 1.30 thus reveals that the CMD Wilson coefficient is given by  $C_{cG} = g_* g_s$ . By requiring the first validity condition, the relation

$$\frac{C_{cG}}{\Lambda^2} < \frac{g_* g_s}{M_{\text{cut}}^2} \quad (1.39)$$

is obtained. Since both  $C_{cG}$  and  $\Lambda$  are a priori unknown, we can redefine  $\tilde{C}_{cG} = \frac{C_{cG}}{\Lambda^2}$ . With this redefinition and by setting  $g_* \sim 4\pi$ , the expression

$$\frac{|\tilde{C}_{cG}| M_{\text{cut}}^2}{4\pi g_s} < 1 . \quad (1.40)$$

can be used to define a plane in  $\tilde{C}_{cG}$  and  $M_{\text{cut}}$  that satisfies the previously discussed conditions.



317 **Chapter 2**

318 **The CMS experiment at the LHC**

319 The Compact Muon Solenoid (CMS) detector [25] is large, general purpose particle detector  
320 located at the Large Hadron Collider (LHC)[26] accelerator in Geneva, Switzerland. Run by the  
321 European Organisation for Nuclear Research (CERN), the LHC's largest ring spans a circumfer-  
322 ence of 27km, making it the largest particle accelerator in the world. In their circular trajectory  
323 through the beam pipe, collimated bunches of  $\sim 10^{11}$  protons are accelerated in both directions  
324 of the ring. At each of the four collision points, of which CMS is built around one, the trajec-  
325 tories of these proton bunches are crossed such that highly energetic proton-proton collisions are  
326 produced. A sketch of the LHC accelerator complex can be seen in Figure 2.1. A detector such  
327 as CMS effectively acts as a camera taking very complex snapshot of each collision. During Run  
328 2 of the LHC, approximately 30 protons collide on average per bunch crossing with a centre of  
329 mass energy of  $\sqrt{s} = 13$  TeV. These collisions produce a plethora of particles, many of which  
330 decay to sets of particles of varying multiplicities themselves. As such, these collision produce a  
331 complex and varied phenomenology that require a complex machine such as the CMS detector  
332 to fully capture. By recording the information from many millions of collisions, a multitude of  
333 different statistical analyses may be performed. This includes analyses of the Higgs boson and its  
334 properties, such as the Yukawa coupling of the charm quark. To this end, this chapter gives an  
335 overview of the CMS detector and its subsystems as well as the techniques used to reconstruct  
336 individual proton-proton collisions.

337 **2.1 The CMS detector**

338 The CMS detector is designed to be able to detect a wide range of signatures and is built from a  
339 set of complementary sub-detectors. An overview of the detector may be seen in Figure 2.2. By  
340 combining data from these sub-detectors, a comprehensive reconstruction of individual proton-  
341 proton collisions, commonly referred to as an *event*, may be made. The role and functioning of  
342 the individual sub-detectors is covered in this section. While several of the detector components  
343 have undergone changes for the current Run-3 of the LHC[29], the configuration relevant to this  
344 work is that of Run-2.

345 **2.1.1 The CMS coordinate system**

346 Due to the cylindrical nature of the CMS detector, using cylindrical coordinates to describe  
347 positions within the detector is a natural choice. Thus, the z coordinate describes the position  
348 along the beam pipe,  $r$  the radius and  $\phi$  the azimuthal angle, where the proton-proton collision

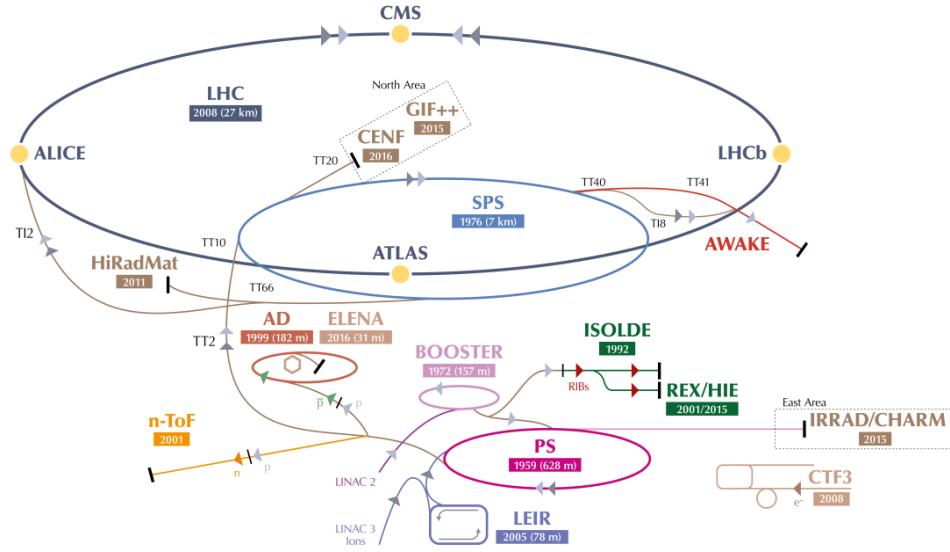


Figure 2.1: An overview of the LHC accelerator complex [27]. Before entering the large LHC ring, particles must pass through a number of increasingly powerful set of accelerators.

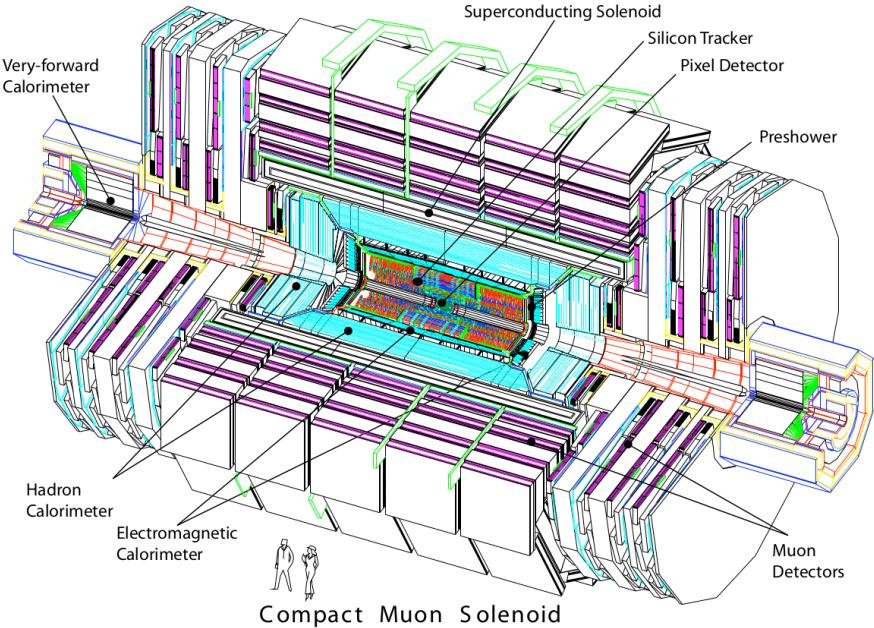


Figure 2.2: An overview of the CMS detector [28].

349 point is taken as the coordinate system's centre. Trajectories of particles with energy  $E$  within  
 350 the detector into the plane perpendicular to  $z$  may be described by the rapidity

$$y = \ln \sqrt{\frac{E + p_z c}{E - p_z c}}. \quad (2.1)$$

351  
 352 Small momenta in the  $z$ -direction  $p_z$  give a rapidity of zero, while the rapidity tends to  $\pm\infty$  for  
 353 large  $p_z$ . However, this requires knowledge of  $E$  and  $p_z$ , which can be difficult to measure. By  
 354 assuming the particle is ultra-relativistic, as is typically the case at the LHC, it is possible to  
 355 simply this description and introduce the pseudorapidity

$$\eta = \ln \left( \tan \left( \frac{\theta}{2} \right) \right) \quad (2.2)$$

which is dependent solely on  $\theta$ , the polar angle. A convenient feature of the (pseudo)rapidity is that differences of (pseudo)rapidity are Lorentz invariant and thus not dependant on the initial longitudinal boost of the proton-proton system, which is a priori not known due to the varying momenta fractions of its constituents. Together with the particle's transverse (to the beam axis) momentum  $p_T$  and mass  $m$ , a particle's four-vector may be described by

$$p = \begin{pmatrix} m \\ p_T \\ \eta \\ \phi \end{pmatrix}. \quad (2.3)$$

356  
 357 The CMS detector may be broadly split into two distinct regions inward and outward of the  
 358 boundary  $|\eta| = 1.479$ . The inner region or *barrel* consists of concentric layers around the beam  
 359 pipe. The outer *endcap* region consists of two caps that close off the detector at either end.  
 360 In this way, the CMS detector is designed for the best possible hermetic coverage around the  
 361 collision point.

### 362 2.1.2 The silicon tracker

363 The silicon tracker [30] is the innermost system of the CMS detector, situated closest to the  
 364 beampipe. It is designed to track the trajectories of charged particles as they emerge from the  
 365 collision point while producing minimal energy losses of the particles themselves. This subde-  
 366 tector is split into two main components, the pixel detector and silicon strip detector. A sketch  
 367 of these components may be seen in Figure 2.3.

368  
 369 The pixel detector is situated right around the beampipe and as of 2017 consists of four cir-  
 370 cular layers of individual silicon pixels in the barrel region and three disk layers in the endcap  
 371 region. These consist of rectangular silicon chips with a size of  $100 \times 150 \mu\text{m}^2$ . When a charged  
 372 particle traverses through the active material of these chips, an electrical signal is induced that  
 373 is recorded. This is typically referred to as a *hit*. The small pixel size allows for position mea-  
 374 surements with a very high resolution, namely  $\sim 10\mu\text{m}$  in the  $r\phi$  direction and  $\sim 20\mu\text{m}$  in the  $z$   
 375 direction [31]. An important feature of the pixel detector is its high radiation tolerance due to the  
 376 close proximity of these modules to the beam pipe.

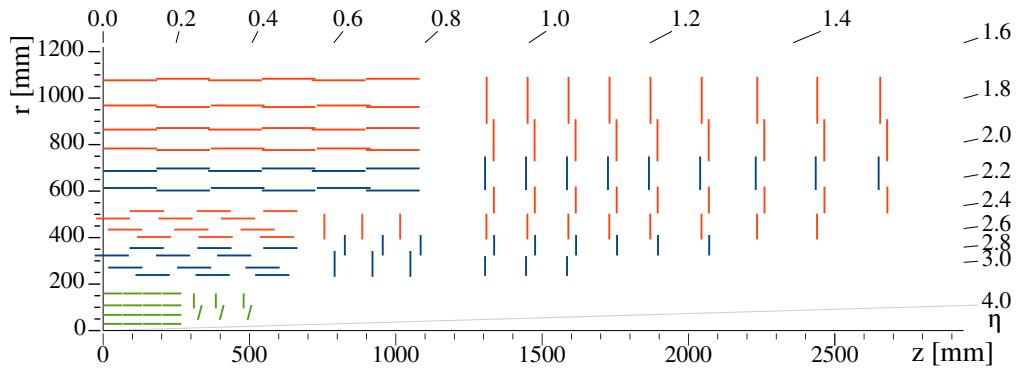


Figure 2.3: An overview of the CMS silicon tracker [30], shown in the  $r$ - $z$  plane after its upgrade during Run-2. The pixel detector is denoted in green while the silicon strip detector is denoted in blue and orange.

377  
 378 Following the pixel detector is the silicon strip detector. It is composed of silicon strips of  
 379 varying sizes, with increases in size at greater distances to the beam pipe due to the reduced  
 380 overall particle flux they must contend with. In the barrel region, this consists of 10 layers of  
 381 silicon strips, while in the endcap regions this consists of nine layers. The latter extend the  
 382 coverage of the detector to  $|\eta|=2.5$ .

383  
 384 The tracking system provides key information that is essential to the reconstruction of events. As  
 385 charged particle fly through the CMS detector, their trajectories are curved due to the magnetic  
 386 field generated by the solenoid magnet (see subsection 2.1.5). By measuring the curvature of  
 387 these trajectories with this system, the transverse momentum  $p_T$  of particles can be constructed.  
 388 Additionally, the tracker plays a key role in methods used to determine the nature of hadronic  
 389 particle cascades and the progenitor particles (quarks or gluons) from which these originate.

390

391

### 392 2.1.3 The electromagnetic calorimeter

393 The second innermost subsystem is the electromagnetic calorimeter (ECAL) [32][33], a sketch  
 394 of which may be seen in Figure 2.4. It is designed to measure the energies of electromagnetic  
 395 showers initiated by photons and electrons. The ECAL is a homogenous calorimeter, consisting  
 396 of over 75,000 lead tungstate crystals. These crystals scintillate as charged particles pass through  
 397 them and the produced photons can be collected via photodiodes, producing an electrical signal.  
 398 This signal may be evaluated to infer the energy that is deposited. Not only do the crystals  
 399 scintillate but they are also extremely dense and thus are very effective in absorbing the energy  
 400 of incoming electrons and photons. This allows a very compact thickness of 23cm (22cm) in  
 401 the barrel (endcap) region, which corresponds to  $\sim 26$  ( $\sim 25$ ) radiation lengths. An additional  
 402 component of the ECAL is the preshower detector. This consists of lead absorbers interlaced  
 403 with scintillating layers and help to distinguish high energy photons from neutral pions. The  
 404 latter decays into photon pairs which may mimic high energy photons in this part of the detector  
 405 with an increased likelihood. The increased granularity of the preshower detector helps mitigate  
 406 this effect. The energy resolution of the ECAL is  $\sim 1\text{-}4\%$ .

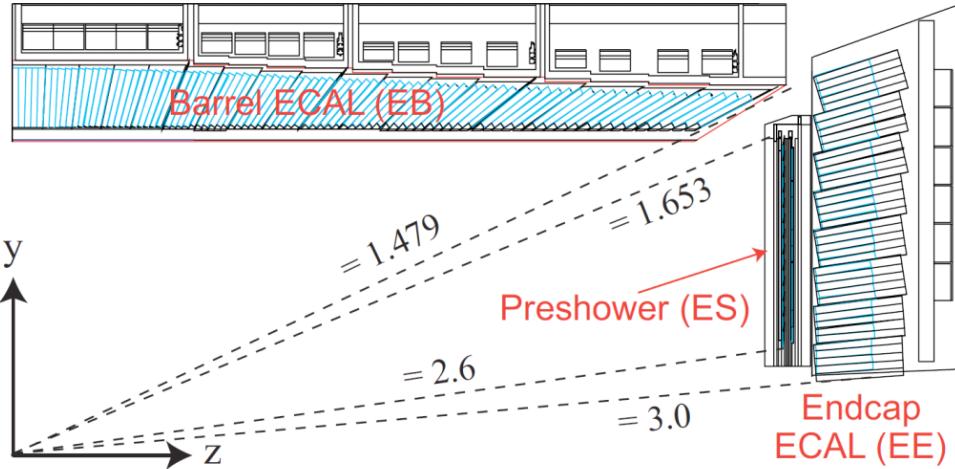


Figure 2.4: An overview of the CMS ECAL [34], shown in the  $r(y)$ - $z$  plane. The dashed lines denote the coverage of the barrel and endcap ECAL region as well as the preshower detector.

#### 2.1.4 The hadronic calorimeter

Following the ECAL is the hadronic calorimeter (HCAL) [35]. It is designed to measure the presence and energy of hadrons, which typically traverse the ECAL with minor energy losses. It is the most hermetic part of the CMS detector, with a coverage out to  $|\eta| = 5.0$ , in order to absorb almost all particle produced in the proton-proton collision. The only exceptions to this are muons which are particles that minimally deposit their energy and neutrinos, which have an interaction probability that is so low that they cannot be measured with the CMS detector at all.

In contrast to the ECAL, the HCAL is a sampling calorimeter. This means layers of absorber are interleaved with layers of a scintillator. Different materials are used in different parts of the calorimeter, which is split into the barrel ( $|\eta| < 1.5$ ), endcap ( $1.5 < |\eta| < 3.0$ ) and forward ( $3.0 < |\eta| < 5.0$ ) regions. Since the HCAL component inside the magnet system does not sufficiently absorb all hadronic showers, the system also extends past the magnet. Due to the sampling nature of the calorimeter, a lower number of respective interaction lengths and larger energy fluctuations in hadronic particle showers, the energy resolution of the HCAL is significantly worse than the ECAL. It lies in the order of 10-30% and with a strong dependence on the energy and pseudorapidity of the initiating particles.

#### 2.1.5 The superconducting solenoid magnet

A key component of the CMS detector is the superconducting solenoid magnet [36]. It is responsible for maintaining a strong 3.8 T magnetic field that homogeneously permeates the barrel of the detector. A measurement of the field strength can be seen in Figure 2.5. With its toroidal shape, the field is orientated along the  $z$ -axis and covers the 12.9m long barrel region of the detector, curving the trajectories of charged particles emerging from the interaction point in the  $\phi$ -direction. This allows for a measurement of the particles transverse momentum  $p_T$ , which together with the  $\phi$  and  $\eta$  directions fully characterise the particle's momentum vector. The magnet itself is composed of superconducting niobium-titanium coils that are cooled to a

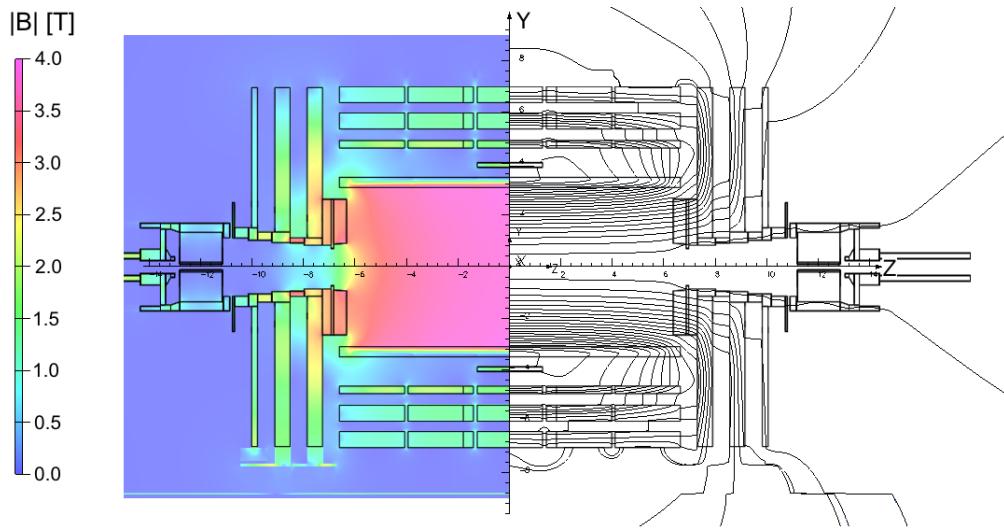


Figure 2.5: An overview of the magnetic flux (left) and magnetic field lines(right) inside the CMS detector, shown in the r-z plane [37].

433 temperature of 4.65K, at which these are superconducting. The magnet is encased by a 12,000t  
434 steel yoke that captures the magentic field that is produced outside of the solenoid.

### 435 2.1.6 The muon chambers

436 The muon subdetector consists of a dedicated system of gaseous detectors [38][39], which are  
437 placed outside of the solenoid magnet. As suggested by the CMS name, a strong focus is placed  
438 on the performance of this subdetector. This is as muons may often be produced in collisions that  
439 are of physics interest (such as in this work) and thus an emphasis is laid on detecting these with  
440 great efficiency. Due to muons being minimally ionising particles, they easily pass through the  
441 inner subdetector layers to reach the muon chambers and information from the moun chambers  
442 as well as the tracker and calorimeters may be used to identify and reconstruct them.

443

444 Like the other subdetectors, the muon chambers are separated in a barrel ( $|\eta| < 1.2$ ) and endcap  
445 ( $1.2 < |\eta| < 2.4$ ) region, which are composed of drift tubes and cathode strip chambers respec-  
446 tively. The drift tubes each consists of a gas volume containing a mixture of Argon and CO<sub>2</sub>  
447 in which a positively charged wire is stretched through the center. When charged particles  
448 such as muons traverse these tubes, the gas is ionised. Due to the positive charge of the wire,  
449 the resulting electrons drift towards the wire producing an electrical signal. Thus the presence  
450 of muons may be determined by activation of the drift tubes. The cathode strip chambers on  
451 the other hand consist of layers of positively charged (anode) wires, which are arranged in a  
452 perpendicular fashion to a set of negatively charged (cathode) strips. Combining signals from  
453 both the wires and strips allows for a position measurement in both the R and  $\phi$  direction. Both  
454 types of detector are supplemented by resistive plate chambers, which act as a trigger providing  
455 a precise timing resolution of  $\sim$ 1ns. This makes it possible to unambiguously assign muons to  
456 individual collisions. These consist of parallel, oppositely charged plastic plates that are coated  
457 with a conductive graphite layer and are contained in a gas volume. Ionisation of the gas due to  
458 the traversal of a charged particle thus leads to an electrical signal. An overview of the spatial

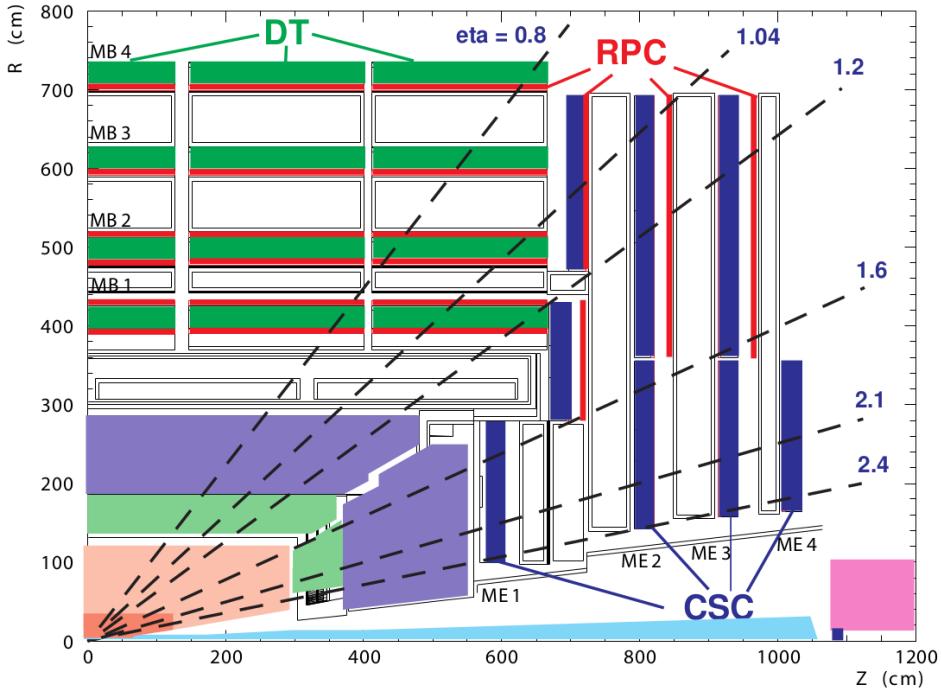


Figure 2.6: An overview of the CMS muon system, shown in the  $r$ - $z$  plane [28]. Shown are the drift tube (DT), the cathode strip chambers (CSC) and resistive plate chambers (RPC).

459 arrangement of these systems can be see in Figure 2.6. With this system, the bulk of muons may  
 460 be measured with a precise momentum resolution of  $\sim 1\text{-}2\%$ .

### 461 2.1.7 The triggering system

462 The triggering system is an essential component in manging the data output of the CMS detector [40]. With a nominal collision rate of  $\sim 40$  MHz, the data rate the CMS detector provides is  
 463 close to 40 TB/s. Not only is the storage of such a quantity of data unfeasible but a significant  
 464 portion consists of low-energy scattering events which are not of interest to analysts. As such,  
 465 the triggering system is implemented to extract a subset of events that are of physics interest.

467  
 468 The trigger systems is composed of two subsystems. The first the so-called level one (L1) trigger.  
 469 This is a very fast hardware-based system which reduces the event rate to  $\sim 100$  kHz by evaluating  
 470 the presence of e.g. energetic muons or other interesting signatures such as large energy  
 471 deposits in the calorimeters in an event. The total time allocated to decide whether an event  
 472 should be kept is  $3.2\mu\text{s}$ . Subtracting for signal propagation in the detector, the L1 system must  
 473 make a decision within  $1\mu\text{s}$ . From the L1, the events are passed to a software based high-level  
 474 trigger (HLT) system. This is composed of several thousand CPU cores, performing a simple  
 475 reconstruction of the event signatures to make a decision whether an event should be stored.  
 476 Since different analyses are interested in different signatures, a set of trigger paths are defined  
 477 so that only one such path must be satisfied for an event to pass the HLT. Since the HLT is  
 478 software-based, the trigger paths may be continuously updated. After the HLT, the event rate

479 is reduced to  $\sim$ 100 Hz and the passing events are permanently stored.

### 480 2.1.8 The 2018 dataset of the CMS detector

481 The dataset used in this work is comprised of the 2018 dataset, recorded by the CMS experiment  
482 during Run-2 of the LHC. This dataset is comprised of an integrated luminosity of  $58.83\text{fb}^{-1}$   
483 worth of data at a center-of-mass energy of 13 TeV. It is used in the estimation of reducible  
484 backgrounds described in subsection 3.2.4 as well as in section 3.3, where it is used to verify the  
485 background estimation in sidebands of the primary signal region.

## 486 2.2 Event reconstruction with the CMS detector

487 Events that pass the triggering system are stored and reconstructed using a more complicated  
488 set of reconstruction algorithms. An overview of the reconstruction techniques for the objects  
489 relevant to this work, namely muons and jets, is given in this section.

### 490 2.2.1 Track and vertex reconstruction

491 Particle tracks, describing the trajectories of particles through the detector, can be obtained by  
492 leveraging information from the pixel and strip detectors of the tracker [41]. By determining  
493 the track of a charged particle and thus the curvature of its trajectory in the detectors magnetic field,  
494 the particle's transverse momentum  $p_T$  may be implicitly determined. Since track  
495 reconstruction is a computationally intensive procedure given the large number of permutations  
496 in which individual pixel or strip hits may be combined, this procedure is performed iteratively.  
497 Initially, tracks which are easily identifiable due to e.g. their relatively high  $p_T$  or proximity to  
498 the interaction point are identified by matching hits in the pixel and silicon strip subdetectors  
499 and performing a fitting procedure. The hits associated with these tracks are then removed from  
500 the collection of unassociated hits. This procedure is repeated anew with looser fitting criteria  
501 so that hits that may originate from low  $p_T$  tracks or those with an origin displaced from the  
502 collision point, may also be associated to tracks.

503 From the reconstructed tracks, common track origins or *vertices* may be identified. Since several  
504 proton-proton collisions may occur in a single bunch crossing, this amounts to identifying  
505 the location of the individual collisions in an event. Tracks with a low perpendicular distance  
506 or low *impact parameter* to the center of the bunch crossing and that satisfy requirements on  
507 the number of pixel and strip detector hits as well as the quality of the track fit are chosen for  
508 this purpose. These tracks are clustered using a deterministic annealing algorithm [42], thus  
509 producing a set of candidate vertices with some location along the z-axis. The vertex candidate  
510 which is associated with the highest  $\sum p_T^2$  is assigned as the primary vertex of the collision. The  
511 remaining vertex candidates are referred to as pile-up vertices.

### 513 2.2.2 The Particle Flow algorithm

514 The Particle Flow (PF) algorithm [43] is used to combine information from many of the different  
515 CMS subsystems to give an improved and holistic description of an event. This includes  
516 reconstructed tracks, the energy deposits in the ECAL and HCAL as well as hits in the muon  
517 chamber system. Since different types of particles will interact with the CMS subdetector systems  
518 in unique ways, the properties of individual particles can be extrapolated from this information.  
519 These are briefly summarised in Table 2.1.

520  
521

Table 2.1: Overview of particle signatures in the CMS detector

Particle	Signature
Muons	Muons produce tracks in the tracker as well as the muon system with minimal energy deposits in the calorimeters.
Electrons	Electrons produce tracks in the tracker as well as energy deposits in the ECAL with minimal deposits in the HCAL.
Photons	Photons do not produce tracks in the tracker due to being uncharged and deposit their energy in the ECAL.
Charged hadrons	Charged hadrons produces tracks in the tracker, primarily depositing their energy in the HCAL.
Neutral hadrons	Neutral hadrons produce no tracks in the tracker, primarily depositing their energy in the HCAL.

522 A visual overview of these signatures and the particle type they correspond to can be found in  
 523 Figure 2.7. The PF algorithm leverages exactly these properties. Initially, matched tracks in  
 524 the tracker and muon systems are identified as muons and the corresponding components are  
 525 removed from the event. Subsequently, matched tracks and energy deposit clusters in the ECAL  
 526 are identified as electrons and the corresponding components are removed. An isolated cluster in  
 527 the ECAL with no associated track is reconstructed as a photon candidate and the corrsponding  
 528 cluster is removed. This is expected to leave only charged and neutral hadrons. Clusters of  
 529 energy deposits in the HCAL associated with a track are thus identified as charged hadrons.  
 530 However, it frequently occurs that photons are produced in the decay of neutral hadrons. Thus,  
 531 if the energy estimated from a track is considerably less than the associated cluster in the HCAL  
 532 and there is a corresponding energy deposit in the ECAL, an additional photon candidate is  
 533 reconstructed that is associated with the hadron. Finally, HCAL clusters with no associated  
 534 track are reconstructed as neutral hadrons.

535  
 536 This of course is a greatly simplified description, a more comprehensive version of which can  
 537 be found in [43]. The following section describe in greater detail the reconstruction of objects  
 538 relevant to this work. This includes muons, *jets*, which are collimated particle showers that  
 539 typically consist of a collection of reconstructed objects and missing transverse energy.

### 540 2.2.3 Reconstruction and identification of muons

541 Since muons are used to reconstruct the Higgs candidate of the cH process, they represent an  
 542 important element of the analysis described in this work. Using the available information from  
 543 the tracker and muon system, three different approaches may be used to intially reconstruct  
 544 muon tracks.

- 545 • **Standalone muon tracks:** A standalone muon track refers to a fit of individual hits  
 546 present in the muon detector.
- 547 • **Tracker muon tracks:** Tracker muon tracks are reconstructed by extrapolating tracks  
 548 from the tracker to the muon detector, referred to as an *inside-out* approach. If a hit in

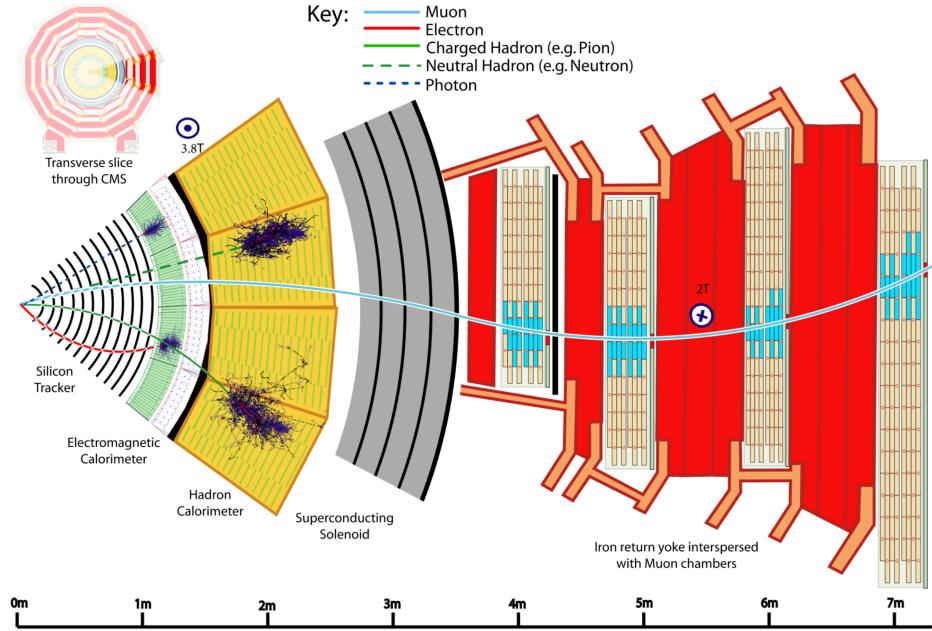


Figure 2.7: A transverse slice of the CMS detector, visualising the signatures that different particles produce in the different detector subsystems. [43].

549           the muon detector can be matched to the extrapolated track, then these matched tracks  
 550           are identified as a tracker muon track. This reduces the impact from atmospheric muons  
 551           traversing the detector, which may be falsely interpreted as standalone muon tracks.

- 552           • **Global muon tracks:** Global muon tracks are obtained through an *outside-in* approach,  
 553           matching standalone muon tracks with tracker muon tracks through a comparison of the  
 554           respective fitted track parameters. If the tracks are found to match, a combined fit of these  
 555           tracks is performed. This approach reduces the impact from remnants of hadronic showers  
 556           that reach the muon chambers, which may be incorrectly reconstructed as a tracker muon  
 557           track.

558           Naturally, there is a large overlap between global and tracker muon tracks. If two muon tracks  
 559           share the same track in the tracker, then they are merged into a single object. The collection  
 560           of standalone, tracker and global muons is passed to the previously introduced PF algorithm  
 561           which, by imposing additional quality requirements (see [43]) produces a set of reconstructed  
 562           muon candidates.

563           A useful criterium in identifying muons that originate directly from the proton-proton inter-  
 564           action is the relative isolation  $\mathcal{I}_{\text{rel}}^{\mu}$ . This is defined as

$$\mathcal{I}_{\text{rel}}^{\mu} = \left( \sum p_{\text{T}}^{\text{charged}} + \max(\sum p_{\text{T}}^{\text{neutral}} + \sum p_{\text{T}\gamma} - p_{\text{T}}^{\mu,\text{PU}}) \right) / p_{\text{T}}^{\mu}. \quad (2.4)$$

566 Here,  $\sum p_T^{\text{charged}}$  represents the scalar sum of the transverse momenta of charged hadron originating from the primary vertex of the event. The quantities  $\sum p_T^{\text{neutral}}$  and  $\sum p_T^\gamma$  represent the  
 567 respective transverse momenta sums for neutral hadrons and photons. These sums are calculated  
 568 by accounting from contributions within a conical volume around the muon direction. The size  
 569 of a cone between two positions  $i$  and  $j$  is defined as  $\Delta R(i, j) = \sqrt{\Delta\eta(i, j)^2 + \Delta\phi(i, j)^2}$  and in  
 570 this case the cone boundary around the muon direction is set at  $\Delta R = 0.4$ . The contribution to  
 571 the relative isolation from pile-up is estimated by subtracting  $p_T^{\mu, \text{PU}} = 0.5 \sum_k p_T^{k, \text{charged}}$  in Equa-  
 572 tion 2.4, where the sum over  $k$  represents charged hadron contributions not originating from the  
 573 PV. The factor 0.5 corrects for different fractions of charged and neutral particles in the cone  
 574 [44]. Lastly,  $p_T^\mu$  represents the transverse momentum of the muon. The relative isolation is thus  
 575 a variable that quantifies the presence of energy deposits in the ECAL and HCAL around the  
 576 trajectory of the muon, relatively to the  $p_T$  of the muon. Since muons are expected to produce  
 577 such deposits only minimally, good muon candidates are expected to be associated with small  
 578 values of  $\mathcal{I}_{\text{rel}}^\mu$ .

580  
 581 Two sets of muon identification criteria are defined for this work:

- 582 • **Loose muons:** Loose muons are PF muons reconstructed from either a global or tracker  
 583 muon track where the perpendicular distance of the extrapolated track to the event's pri-  
 584 mary vertex is less than 5mm in the  $z$  direction and less than 2mm in the  $r$  direction.
- 585 • **Tight muons:** Tight muons are loose muons which are reconstructed exclusively from a  
 586 global muon track. A number of additional criteria are applied. This includes that the fit  
 587 quality of the global muon track must be  $\chi^2/\text{ndf} < 10$  as well that the significance of the  
 588 track's 3D impact parameter  $\text{SIP}_{3\text{D}} = \text{IP}/\sigma_{\text{IP}}$  satisfies  $\text{SIP}_{3\text{D}} < 4$ . Here  $\text{IP}$  is the impact  
 589 parameter or point of closest approach to the primary vertex and  $\sigma_{\text{IP}}$  is the associated  
 590 uncertainty. Additionally, it is required that at least six layers with at least one pixel hit  
 591 are registered in the tracker in the associated track as well as two segments hit in the muon  
 592 detector. Lastly, a relative isolation requirement of  $\mathcal{I}_{\text{rel}}^\mu < 0.25$  is imposed.

593 The tight muon definition is used to select muons for reconstructing Higgs candidates while the  
 594 loose definition is used in the estimation of reducible backgrounds.

#### 595 2.2.4 Reconstruction and identification of jets

596 The quarks and gluons that are produced in proton-proton collisions rapidly hadronise, typically  
 597 producing collimated cones of particles referred to as *jets*. Details on the concept of hadronisa-  
 598 tion, which results from the nature of the strong interaction, can be found in [45]. Since the  $c$   
 599 quark of the  $c\bar{H}$  process too will produce a jet, jet objects also represent an important aspect of  
 600 the analysis presented in this work.

601 To produce jet objects, the hadrons reconstructed by the PF algorithm must be clustered. To  
 602 ensure a minimal impact of pile-up on this clustering, the contributions of pile-up are mitigated  
 603 through *charged hadron subtraction*. This involves the removal of charged hadron contributions  
 604 in the HCAL and ECAL if these may be associated with any of the pile-up vertices produced  
 605 in the collision, as described in subsection 2.2.1. Once this subtraction has been performed, the  
 606 remaining PF hadrons are passed to the anti- $k_T$  algorithm [46]. The anti- $k_T$  algorithm is an it-  
 607 erative clustering algorithm that is based on a principle of minimal distances between particles.  
 608 The distance  $d_{ij}$  between the particles  $i$  and  $j$  is defined as well as the distance  $d_{iB}$  between  
 609 particle  $i$  and the beam. These are given by

$$d_{ij} = \min\left(\frac{1}{p_{T,i}}, \frac{1}{p_{T,j}}\right) \frac{\Delta_{ij}^2}{R^2} \quad (2.5)$$

$$d_{iB} = \frac{1}{p_{T,i}} \quad (2.6)$$

$$\Delta_{ij} = \sqrt{\Delta y(i,j)^2 + \Delta\phi(i,j)^2}. \quad (2.7)$$

611  
612 Here,  $y$  is the rapidity of a particle and  $R$  is a constant parameter that determines the cone size  
613 of the clustered jets. The default choice used in CMS is  $R=0.4$ , which is also used in this work.  
614 Starting with the highest  $p_T$  object in the initial iteration, the distance  $d_{ij}$  with the closest PF  
615 candidate  $j$  is calculated. The two objects are clustered together and this process is repeated un-  
616 til a stopping condition  $d_{ih} > d_{iB}$  is met. At this point, the jet is considered fully reconstructed  
617 and the PF candidates used in its clustering are removed for the reconstruction of subsequent jets.

618  
619 Due to the presence of detector noise, unphysical low  $p_T$  jets can be erroneously reconstructed.  
620 This effect can be mitigated by applying additional criteria on reconstructed jets. This includes  
621 requiring that at least two PF candidates are clustered in the jet and that the jet's energy is  
622 not solely attributed to neutral hadrons or photons. These requirements remove almost all such  
623 unphysical jets while over 99% of physical jets fulfill them [47]. Additionally, a pile-up discrim-  
624 ination algorithm is described in [47], of which the loose working point is applied to jets with  
625  $p_T < 50$  GeV in this work.

626  
627 A calibration of jet energies is performed after reconstruction [48] in both simulation and data.  
628 This calibration accounts for pile-up contributions in the clustering, the non-linearity of the de-  
629 tector response and improper reconstruction of hadrons. A number of methods are used to derive  
630 sets of correction factors. An example is the use of events with a Z boson, the  $p_T$  of which may  
631 be precisely reconstructed via the  $Z \rightarrow \mu\mu$  decay, that a single jet recoils against. Additionally,  
632 significant discrepancies in the resolution of jets in simulation and data are observed, with the  
633 resolution being worse in the latter than the former. This is accounted for by a smearing method,  
634 in which the resolution of jets is artificially smeared in simulation so that a better comparison to  
635 data is achieved.

### 636 2.2.5 Missing transverse momentum

637 Due to the conservation of momentum, it is expected that the vectorial sum of momenta of all  
638 particles produced in a collision adds up to zero. However, this may not be the case when particles  
639 such as neutrinos are produced in a collision as these cannot be measured by the detector. As a  
640 result, it can be useful to define the missing transverse momentum as

$$p_T^{\text{miss}} = \sum_i^{\text{PF}} p_T^{(i)}. \quad (2.8)$$

641  
642 The presence of significant quantities of  $p_T^{\text{miss}}$  may thus be used to identify the presence of  
643 neutrinos in an event.

### 644 2.3 Identification of charm quark-induced jets

645 To identify the charm quark-induced jet of the cH process, one must be able to discriminate  
 646 against both bottom quark as well as light quark or gluon-induced jets. This is a task colloquially  
 647 referred to as *flavour tagging*, with a jet's *flavour* being determined by the type of particle that  
 648 initiated it. Modern flavour tagging techniques typically use machine learning to leverage key  
 649 jet properties that may differentiate jets of different flavours, though this remains a challenging  
 650 task. To discuss these properties, a definition of jet flavour is useful. In the context of CMS, a  
 651 ghost matching procedure [49] is applied to obtain such a definition for simulated events. This  
 652 involves adding information from the event simulation to the reconstructed event. Specifically,  
 653 hadrons containing bottom and charm quarks are identified in the simulation and added to the  
 654 list of reconstructed PF candidates, albeit with negligible momenta. With this addition of so-  
 655 called *ghost hadrons* the jet clustering is once again performed. Due to the negligible momenta  
 656 of the ghost hadrons, the clustering procedure itself is unaffected. However, the inclusion of the  
 657 ghost hadrons can be used for the following definitions:

- 658 • **c jets:** If at least one charm (*c*) ghost hadron and no bottom (*b*) hadrons are clustered  
   659 inside the jet, the jet is labelled as a *c* jet.
- 660 • **b jets:** If at least one *b* ghost hadron is clustered inside the jet, the jet is labelled as a *b*  
   661 jet.
- 662 • **light jets:** If no ghost hadrons are clustered inside the jet, the jet is labelled as a light  
   663 jet. Light jets may be initiated by quarks such as the up, down, or strange quark or by  
   664 gluons. An additional, technical category of *pile-up jets* exists depending on whether so-  
   665 called matching criteria between reconstructed and simulated jets are fulfilled, though they  
   666 are subsumed into the light jets category for the purpose of this work.

667 The task of identifying *c* jets is thus twofold and broken down into two tasks:

- 668 1. Discriminating *heavy-flavour* (HF) jets consisting of *b* jets and *c* jets against light jets.
- 669 2. Discriminating between *b* jets and *c* jets.

#### 670 2.3.1 Properties of heavy-flavour jets

671 The term heavy-flavour originates from the mass of the bottom and charm quarks, which is  
 672 an order of magnitude greater than the next heaviest quark, the strange quark. The *c* and *b*  
 673 hadrons have relatively long lifetimes that allow them to travel an observable distance from the  
 674 PV before decaying. The typical lifetime of a *b* hadron of the order of  $\sim 1.5$  ps while that of *c*  
 675 quarks ranges down to approximately an order of magnitude less [1]. This typically results in  
 676 the presence of a secondary vertex (SV) that is measurably displaced from the collision point  
 677 up to a distance of 1cm in the case of energetic hadrons and is thus a key signature of HF jets.  
 678 Tracks originating from the decay of a HF induced jet thus typically originate from a SV. This  
 679 effect can for example be seen when looking at the significance of 2D impact parameters of *b*, *c*  
 680 and light jets, as seen in Figure 2.8.

681 Another feature of heavy flavour jets is the presence of leptons in the jet. This results from  
 682 the relatively large branching fractions of HF hadrons into states containing leptons. These are  
 683 typically low-energy and are present in about 20% (10%) of *b*(*c*) jets, meaning the identification  
 684 of a low-energy electron or muon inside a jet serves as a good indicator that a jet originates from

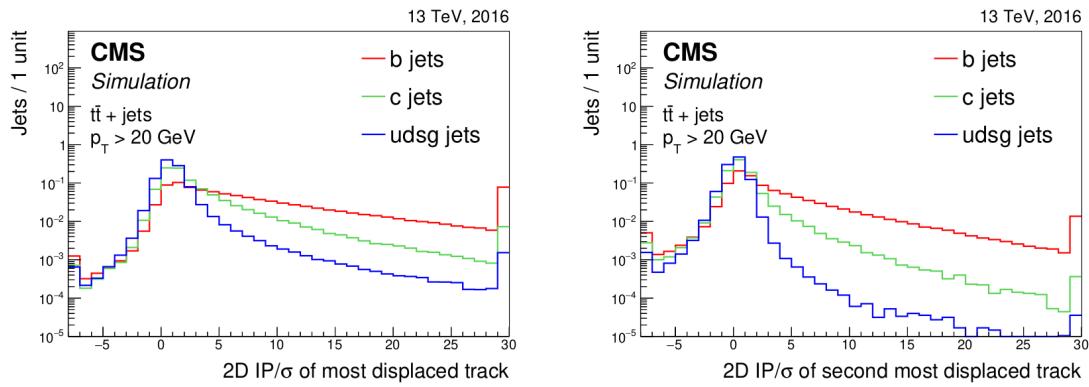


Figure 2.8: Plots showing the significance of the 2D impact parameter of the most and second most displaced tracks in a jet [50]. As can be seen, these variables can differentiate  $b$  and  $c$  jets from light jets to a significant degree.

686 a HF hadron. Also of significance are the relatively high masses HF hadrons exhibit in  
 687 comparison to their lighter counterparts. This results in HF induced jets having a broader energy  
 688 flux compared to their lighter counterparts, due to higher diffusion of momenta perpendicular  
 689 to the flight direction as well as a higher hadron multiplicity resulting from the decay of the HF  
 690 hadron. These features are illustrated in Figure 2.9.

### 691 2.3.2 The DeepJet algorithm

692 The DeepJet algorithm [51] is a machine learning algorithm used for jet-flavour identification in  
 693 this work. It improves on previous neural network based algorithms [50] used by CMS in the  
 694 Run-2 period of the LHC. A notable feature compared to earlier algorithms is its use of lower  
 695 level information such as use of track, PV and SV information, as well as PF candidate and  
 696 event kinematics information. An overview of the architecture employed by DeepJet can be see  
 697 in Figure 2.10. The network is comprised of three branches that individually process neutral and  
 698 charged hadrons as well as secondary vertices before this information is combined with global  
 699 variables in a set of fully connected layers. The network ouput consists of six output nodes  
 700 representing six individual output classes. The output value of the nodes  $\mathcal{P}(b/bb/lepb/c/l/g)$  for  
 701 a given jet are interpreted as the likelihood that a jet belongs to the respective class. These are  
 702 defined as

- 703 •  **$b/bb/lepb$  ( $b$  jets):** These three classes represent subclasses of jets originating from a  
 704  $b$  hadron. The  $b$  class represents a jet originating from a single  $b$  hadron, the  $bb$  class  
 705 originating from two  $b$  hadrons and  $lepb$  representing a jet originating from a  $b$  hadron  
 706 with the presence of a soft lepton.
- 707 •  **$c$  ( $c$  jets):** This class represents a jet originating from a  $c$  hadron.
- 708 •  **$l, g$  (light jets):** These two classes represent light jets originating from a light quark or  
 709 gluon respectively.

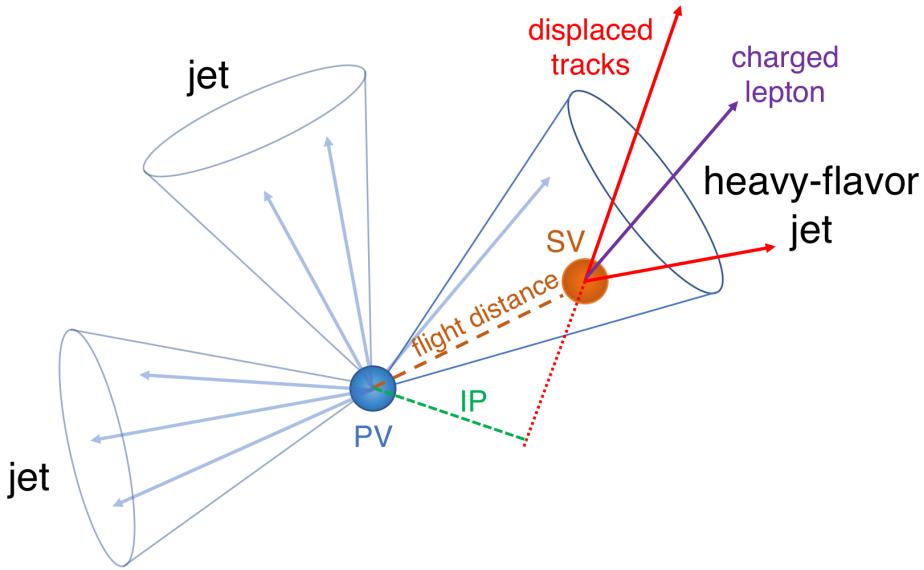


Figure 2.9: An illustration highlighting the properties of HF jets [50]. The presence of a secondary vertex (SV), characterised by the impact parameter (IP) in green, as well as the presence of a lepton is highlighted.

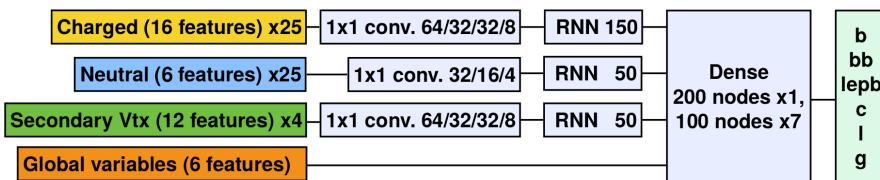


Figure 2.10: An illustration depicting the architecture of the DeepJet neural network [51]. Three individual branches separately process the charged hadrons, neutral hadrons and secondary vertex information before being passed onto a combining, fully connected layer together with global variables.

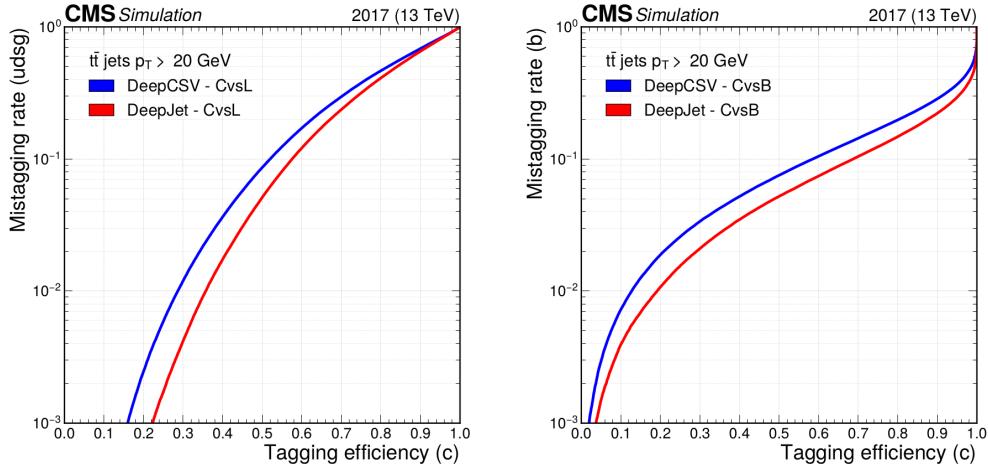


Figure 2.11: Performance of DeepJet algorithm in identifying  $c$  jets against  $b$  jets and light jets in simulated samples of top quark pair production, in which both top quarks decay hadronically [52]. The x-axis represents the efficiency with which  $c$  jets are identified, while the y-axis represents mis-identification rate with respect to either  $b$  jets or light jets.

From these output classes, two useful discriminators to identify  $c$  jets can be constructed. These are

$$\text{CvsB} = \frac{\mathcal{P}(c)}{\mathcal{P}(c) + \mathcal{P}(b) + \mathcal{P}(bb) + \mathcal{P}(lepb)}, \text{CvsL} = \frac{\mathcal{P}(c)}{\mathcal{P}(c) + \mathcal{P}(l) + \mathcal{P}(g)} \quad (2.9)$$

(2.10)

representing a discrimination of  $c$  jets against  $b$  jets and light jets respectively. The performance of DeepJet with the CvsL and CvsB discriminators in simulated samples of top quark pair production can be seen in Figure 2.11. A comparison to the DeepCSV jet-flavour identification algorithm is included, highlighting the performance gain that the DeepJet algorithm achieves.

Since neural network based algorithms are trained on simulated samples that do not perfectly describe their data counterpart, the neural network output must be calibrated with respect to data. To calibrate the entire shape of the algorithm's output distributions the approach described in [52] is used. This involves targeting phase spaces enriched in  $b$  jets (top quark pair production),  $c$  jets (charm associated  $W^\pm$  production) and light jets (jet associated Drell-Yan production). Using simulation, the fractions of  $b$ ,  $c$  and light-flavour jets are determined in each phase space and an iterative fitting procedure, minimising differences between simulation and data is performed. This allows for the derivation of correction factors which depend on the discriminators CvsL and CvsB as well as the true flavour of a simulated jet.

724 **Chapter 3**

725 **Search for the cH( $ZZ \rightarrow 4\mu$ )  
726 process**

727 To probe the charm Yukawa coupling through the cH process, a methodology must be devised to  
728 select and reconstruct cH candidate events. This is described in section 3.1, specifically targetting  
729 cH( $ZZ \rightarrow 4\mu$ ) final states. Additionally, a model describing the expected contributions from the  
730 cH( $ZZ \rightarrow 4\mu$ ) process as well as a number of background processes in the event selection must  
731 be constructed and is described in section 3.2. Finally, a statistical evaluation method using  
732 flavour-tagging discriminators to set 95% CL upper limits on  $\kappa_c$ , assuming the absence of signal,  
733 is presented in section 3.4.

734 **3.1 cH event selection**

735 To reconstruct a cH( $ZZ \rightarrow 4\mu$ ) candidate event, a Higgs boson candidate needs to be reconstructed  
736 and a corresponding jet candidate needs to be identified. These two procedures are described in  
737 this section. Distributions of cH( $ZZ \rightarrow 4\mu$ ) candidate events are shown using a simulation of the  
738 cH( $ZZ \rightarrow 4\mu$ ) process, which is discussed in subsection 3.2.2.

739 To reconstruct a Higgs (jet) candidate, an initial selection of muon (jet) objects must be made.  
740 These are summarised in Table 3.1 along with the HLT trigger path requirement used in this anal-  
741 ysis. The objective of this selection is to identify events with well-reconstructed, isolated muons  
742 as well as a least one well-reconstructed jet. Following this initial selection, the corresponding  
743 objects are passed onto the respective algorithms to select a final Higgs and jet candidate.

Table 3.1: Muon, jet object and HLT path selection requirements.

Object	Selection criteria
Muons	$p_T > 5 \text{ GeV}$ $ \eta  < 2.4$ Tight muon identification criteria
Jets	$p_T > 25 \text{ GeV}$ $ \eta  < 2.5$ Jet ID Pile-up ID, loose working point $\Delta R(\text{jet, selected muons}) > 0.4$ Jets in veto regions of detector are excluded
HLT	HLT IsoMu24 is triggered

### 3.1.1 Higgs candidate selection

A Higgs boson reconstruction algorithm (and muon object selection) very similar to those presented and validated in [53] is implemented. This reconstruction is performed for events in which exactly four selected muons are present to avoid introducing a potential bias when reconstructing non-Higgs (background) events. Then the following reconstruction steps are applied:

1. Of the four selected muons, the  $p_T$ -leading muon is required to satisfy  $p_T > 20 \text{ GeV}$  and the sub-leading muon is required to satisfy  $p_T > 10 \text{ GeV}$ . Additionally, the HLT IsoMu24 trigger requirement must be met. Lastly, to ensure two muons are not spuriously reconstructed from shared tracks, it is required that each muon candidate is separated from the others by  $\Delta R > 0.02$ .
2. Opposite-sign muon pairs are merged into Z boson candidates. At least two Z boson candidates must be reconstructed to proceed. Additionally, the invariant mass of any combination of opposite-sign muons must satisfy  $m_{\mu\mu} > 4 \text{ GeV}$ , to remove any contributions from low mass resonances such as  $J/\psi$ .
3. The Z candidate with a mass closest to the known Z boson mass of  $Z = 91.19 \text{ GeV}$  [1] is interpreted as an on-shell  $Z_1$  candidate. The  $Z_1$  candidate should satisfy  $40 \text{ GeV} < m_{Z_1} < 120 \text{ GeV}$ . The other candidate is taken as the  $Z_2$  candidate, which is typically more off-shell and thus the invariant di-muon mass requirement is relaxed to  $12 \text{ GeV} < m_{Z_1} < 120 \text{ GeV}$ .
4. The  $Z_1$  and  $Z_2$  candidates are combined to form a Higgs boson candidate. The four-muon invariant mass of the Higgs boson candidate must satisfy  $m_H > 70 \text{ GeV}$ .

The reconstructed Higgs boson candidate mass distribution in simulated  $cH(ZZ \rightarrow 4\mu)$  events can be seen in ???. As expected, a peak around the known Higgs mass  $m_H = 125.3 \text{ GeV}$  can be observed, with an elongated tail towards lower masses that originate from increasingly off-shell Z candidate contributions.

### 3.1.2 Jet candidate selection

Once a Higgs boson candidate is reconstructed, a likelihood ratio algorithm is applied to best identify and select the jet that is associated with (i.e. recoils off) the reconstructed Higgs

boson. This algorithm does not use jet-flavour identification methods and is based solely on kinematic properties of the jets so as to minimise the introduction of any flavour bias in the selection. Specifically, two variables related to momentum conservation in the transverse plane are exploited:

- 773 1. The difference in azimuthal angle  $\Delta\phi(H, \text{jet})$  between the Higgs boson candidate  $H$  and  
774 the jet is used. Due to an initial zero net momentum in the direction of the azimuthal  
775 angle, the Higgs boson and associated jet are expected to recoil off each other *back-to-back*  
776 and thus  $\Delta\phi(H, \text{jet})$  is expected to be  $\sim \pm\pi$ .
- 781 2. Since the Higgs boson and associated jet recoil off each other, their  $p_T$  is expected to be  
782 approximately balanced. This information can be captured by transverse momentum ratio  
783  $p_T(H)/p_T(\text{jet})$ .

To derive the relevant distributions to be used in a likelihood ratio, a parton-to-jet matching is performed in simulated  $cH(ZZ \rightarrow 4\mu)$  events. This is achieved by, in a simulated event, taking the directional information of the simulated parton and matching it to a reconstructed jet with the matching requirement  $\Delta R(\text{jet}, \text{parton}) < 0.3$ . All jets which match the initial jet selection are considered in this process. A jet which is matched in this way is labelled as the associated jet, while the remaining non-matched jets are labelled non-associated jets. The efficiency with which this matching is performed can be seen in ???. Once this labelling is performed, the distributions of  $\Delta\phi(H, \text{jet})$  and  $p_T(H)/p_T(\text{jet})$  for associated and non-associated jets are extracted as templates and treated as probability density functions. To capture kinematic differences associated with higher and lower  $p_T$  Higgs candidates, this procedure is repeated in different bins of  $p_T(H)$  listed in Table 3.2.

Using the extracted templates, a per-jet likelihood evaluation can be made in each event. For this, the per-variable likelihood ratio

$$\mathcal{L}(x) = \frac{\mathcal{L}_{\text{associated}}(x)}{\mathcal{L}_{\text{non-associated}}(x)}, \text{ with } x \in \left\{ \Delta\phi(H, \text{jet}), \frac{p_T(H)}{p_T(\text{jet})} \right\} \quad (3.1)$$

is defined. From this follows the per-jet likelihood

$$\mathcal{L}(\text{jet}) = \mathcal{L}\left(\Delta\phi(H, \text{jet})\right) \cdot \mathcal{L}\left(\frac{p_T(H)}{p_T(\text{jet})}\right) \quad (3.2)$$

784 that is evaluated. The jet with the highest associated likelihood in an event is selected as the jet  
785 candidate.

Table 3.2: The  $p_T(H)$  bins in which the jet selection procedure is performed.

Bin number	$p_T(H)$ range
1	0 - 15 GeV
2	15 - 30 GeV
3	30 - 50 GeV
4	50 - 100 GeV
5	100 - 200 GeV
6	>200 GeV

786 With this, the individual components of the  $cH(ZZ \rightarrow 4\mu)$  process are thus reconstructed and  
787 events satisfying the described requirements are selected for evaluation.

788    **3.1.3 Reconstruction efficiency**

789    A total selection efficiency of  $\sim 10\%$  is achieved on the simulated  $cH(ZZ \rightarrow 4\mu)$  sample using the  
 790    described methods. The losses in efficiency can be attributed to the following:

- 791    • Approximately 50% of generated events fall outside of the geometrical detector acceptance.
- 792    • Of the 50% of generated events within the geometrical detector acceptance, approximately  
 793    50% fulfill the described muon selection requirements.
- 794    • Approximately 91% of these events pass the Higgs reconstruction, leading to an additional  
 795    event yield loss of 9%
- 796    • Of the remaining events, the event yield is reduced by close to 60% by the jet selection  
 797    requirements that are imposed.

798    Also of interest is the efficiency with which the ‘correct’ jet is selected in the  $cH(ZZ \rightarrow 4\mu)$  sample  
 799    with the described jet selection method. This can be seen in ??.

800    **3.2 Signal and background estimation**

801    The  $cH(ZZ \rightarrow 4\mu)$  process as well as background processes which may mimic its signature must  
 802    be estimated to accurately reflect the underlying processes as well as their interaction with the  
 803    detector. This is done primarily using Monte Carlo simulations, which are described in subsec-  
 804    tion 3.2.1. The simulation of the  $cH(ZZ \rightarrow 4\mu)$  process is specifically discussed in subsection 3.2.2.  
 805    The estimation of processes that make up the irreducible and reducible backgrounds to the  
 806     $cH(ZZ \rightarrow 4\mu)$  process is discussed in subsection 3.2.3 and subsection 3.2.4 respectively. From  
 807    these estimations, a comprehensive model of the expected yields and distributions resulting from  
 808    applying the described selection on the 2018 dataset of the CMS detector, which is discussed  
 809    in ??, can be constructed. The statistical evaluation procedure that is ultimately applied is  
 810    discussed in in ??.

811    **3.2.1 Monte Carlo simulation of proton-proton collisions**

812    Since the complexity of a proton-proton collisions in a detector cannot realistically be captured by  
 813    analytic calculations, Monte Carlo methods [54] can be used as an approximation. The concept  
 814    of such a simulation relies on a phenomenological approach, sampling the known distributions  
 815    of process and detector quantities and properties to construct a comprehensive simulation of a  
 816    process and its interaction with the detector. The simulation process occurs in discrete steps,  
 817    each dealing with different aspects of the simulated process. These can be summarised as:

- 818    1. **The hard scattering process:** The hard scattering process refers to the immediate,  
 819    high energy transfer scattering of two protons resulting in the production of additional  
 820    particles. To calculate this, two main ingredients are required. The first is a calculation of  
 821    the matrix elements that describe the simulated process in which proton constituents collide  
 822    to produce additional particles. These matrix elements allow for the calculation of a cross  
 823    section for the process. However, the proton itself is a complex object consisting not only of  
 824    its valence quarks (two up-type quarks and one down-type quark) but also of a constantly  
 825    changing ensemble of additional quarks and gluons that are created and annihilated. This  
 826    behaviour must thus be captured for an accurate process description and is parametrised  
 827    via so-called *Parton Distribution Functions*. These describe the likelihood with which a

parton, that carries some fraction  $x$  of the proton's total momentum, may be found in a proton at some energy scale  $Q^2$ . The evolution of the PDF with changing  $Q^2$  is described by the DGLAP equations [55]. Software used to simulate the hard scattering are referred to as *event generators*. Commonly used event generators include `Madgraph5_aMC@NLO` [56] and `POWHEG` [57].

- 828     2. **Parton showering:** Particles such as quarks and gluons that are produced in the hard  
       829     scattering carry the colour charge of the strong interaction. As a result, these may produce  
       830     soft radiation or branch into other particles. While a most physically accurate description  
       831     would be given by including these contributions in the calculation of the hard scattering  
       832     process, this greatly increases the complexity of the calculation. As such a *parton shower*  
       833     model, such as in the `Pythia` software package [58], is used instead to describe the splitting  
       834     of a single mother particle into two daughter particles. In QCD, this describes to  
       835     gluon radiation ( $q \rightarrow qg$ ) and gluon splitting ( $g \rightarrow gg$  and  $g \rightarrow q\bar{q}$ ) and in QED describes  
       836     Bremsstrahlung ( $f \rightarrow f\gamma$ ) and pair creation ( $\gamma \rightarrow f\bar{f}$ ). In case this the parton showering  
       837     originates from initial state partons it is referred to as initial state radiation (ISR). Ac-  
       838     cordingly, parton showering originating from final state partons is referred to as final state  
       839     radiation (FSR). In cases with final states containing multiple partons, there can be some  
       840     ambiguity in the combination of matrix elements and parton showering since both can  
       841     describe the same processes. For this merging schemes are applied that resolve potential  
       842     double counting of events. A prescription used for this work is the FxFx scheme [59].
  - 843     3. **Hadronisation:** At an energy around the QCD scale  $\Lambda_{\text{QCD}}$ , the perturbative parton  
       844     shower prescription loses its validity as the running coupling of the strong force  $\alpha_s$  be-  
       845     comes too strong. Here the individual, colour-charged partons *hadronise* into colour-neutral  
       846     states. Since this process currently cannot be described from first principles, a phenomeno-  
       847     logical description must be applied. In `Pythia`, the *Lund string* model is used [60]. It  
       848     describes the interaction between two partons as a coloured field, the lines of which pass  
       849     through a tube that is extended between the partons. The potential energy of the tube  
       850     (or string) is described by a term linear in the distance between the partons. Thus if the  
       851     partons are separated at a large enough distance and the potential energy is sufficiently  
       852     large, the string may 'break' and new colourless quark-antiquark pairs are formed. This  
       853     procedure may be repeated with these new parton pairs if they posses an invariant mass  
       854     above some threshold.
  - 855     4. **The underlying event:** A description of a variety of effects secondary to the hard scatter-  
       856     ing must be included in the simulation. These can have several origins such as secondary,  
       857     *soft* interactions of the proton-proton collision or remnants of the collided protons, which  
       858     will hadronise themselves. These effects are modeled from data [61].
  - 859     5. **Detector simulation:** Finally, the detector response to the particles emerging from the  
       860     previously described steps must be simulated. This is performed with the `GEANT4` package  
       861     [62], which is configured to model the CMS detector. This includes modelling the curving  
       862     of particle trajectories due to the detector's magnet, the interaction of particles with the  
       863     materials of the detector, as well as the digitisation of the signals in the electronic modules  
       864     of the subdetectors.
- 870     A diagrammatic overview of what an event simulation looks like can be found in Figure 3.1. The  
       871     output of this simulation is passed to the reconstruction algorithms described in section 2.2.

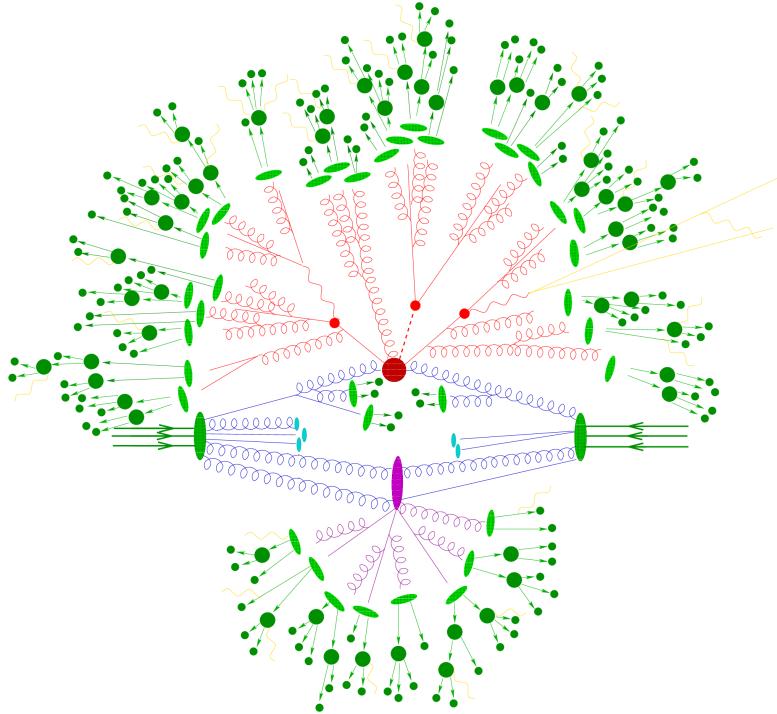


Figure 3.1: An overview of what an event simulation may look like (adapted from [63]).

### 3.2.2 Estimation of cH(ZZ $\rightarrow$ 4 $\mu$ ) process

The cH(ZZ $\rightarrow$ 4 $\mu$ ) process is estimated using a simulation generated by MadGraph5\_aMC@NLO. The following MadGraph5\_aMC@NLO syntax is used, which illustrates some important concepts related to the simulation of the cH process:

```

876 import model loop_sm_MSbar_yb_yc-yc4FS
877 define p = g u u~ d d~ s s~ c c~
878 define j = g u u~ d d~ s s~ c c~
879 generate p p >h [QCD] @ 0
880 generate p p >h j [QCD] @ 1

```

In the first line, it can be read off that the `loop_sm` model is used, a model allowing NLO calculations of the SM. Only the Yukawa-couplings of the bottom and charm quarks are included to ensure orthogonality of the cH simulation to simulations of other Higgs production processes such as gluon fusion. Additionally, a so-called *four flavour scheme* (4FS) version of the model is used [6]. The flavour scheme denotes which quarks are included as constituents of the proton, in which they are approximated as massless. The 4FS includes the up, down, strange and charm quarks as proton constituents. In contrast to the 4FS, a three flavour scheme 3FS could also be used. Here, the charm quark is not included in the proton but instead must be produced via gluon splitting, i.e.  $g \rightarrow c\bar{c}$ .

In the following two lines, the proton and jet constituents are defined. Finally in the last two

892 lines, the processes included in the simulation are defined. These are, calculated to next to leading  
 893 order in QCD, the  $pp \rightarrow H$  and  $pp \rightarrow H + j$  processes. Both are included to give the most  
 894 accurate possible kinematic description of the cH process. The reasoning for this is related to the  
 895 modelling of final state partons and can be better understood by considering what is included  
 896 in the leading order (LO) and next-to-leading (NLO) contributions to  $pp \rightarrow H$  and  $pp \rightarrow H + j$   
 897 respectively. At leading order, an additional jet in  $pp \rightarrow H$  can only be generated via the parton  
 898 shower. Thus, this contribution is expected to best model the lower momentum behaviour of  
 899 final state partons. The NLO contributions to  $pp \rightarrow H$ , which correspond to LO contributions  
 900 of  $pp \rightarrow H + j$ , in turn are expected to better model higher momentum behaviour of the final  
 901 state parton. The same logic is applied to the LO contributions of  $pp \rightarrow H + j$  and the NLO  
 902 contributions of  $pp \rightarrow H + j$ , where two final state partons explicitly appear in the calculation of  
 903 the latter. However, this approach introduces double counting of processes. These are accounted  
 904 for along with double counting between parton shower and matrix element contributions using  
 905 the FxFx merging scheme.

906  
 907 The types of event topologies generated by these commands can be best understood by cat-  
 908 egorising them according to the partons that initiate them. In this way, four categories can be  
 909 identified:

- 910 •  **$c\bar{c}$ :** At leading order, this topology arises from an initial state c and anti-c quark that  
 911 interact to produce a Higgs boson. Final state partons must thus be produced via next-  
 912 to-leading order contributions to the matrix element (i.e. gluon radiation) or via parton  
 913 showering. Notably, the charm quark(s) associated with the Higgs boson are not present  
 914 in the final state of this type of topology. This means that for the described event selec-  
 915 tion strategy, this analysis can only be sensitive to these topologies via an ‘incorrect’ jet  
 916 association. A visualisation of this type of event topology can be seen in Figure 3.3a. It is  
 917 the most prominent in the generated sample.
- 918 •  **$cg$ :** At leading order, this topology arises from an initial state (anti-)c quark that emits  
 919 a Higgs boson and absorbs a gluon. This topology corresponds to the leading order cH  
 920 process that is discussed in subsection 1.3.1. Additional final state partons may be produced  
 921 via next-to-leading order contributions to the matrix element or via parton showering. A  
 922 visualisation of this type of event topology can be seen in Figure 3.3b. It is the second  
 923 most prominent in the generated sample.
- 924 •  **$gg$ :** This topology arises from two initial state gluon which split to form  $c\bar{c}$  pairs that  
 925 interact to produce a Higgs boson. Since this is already considered a next-to-leading order  
 926 contribution to the matrix element by the generator, additional final state partons can  
 927 only be generated via parton showering. A visualisation of this type of event topology can  
 928 be seen in Figure 3.3c. It contributes to only a small fraction of events in the generated  
 929 sample.
- 930 •  **$cl$ :** This topology arises from an initial state (anti-)c quark that interacts with an addi-  
 931 tional initial state light quark via gluon exchange, which produces a Higgs boson. Since this  
 932 is already considered a next-to-leading order contribution to the matrix element by the  
 933 generator, additional final state partons can only be generated via parton showering. A  
 934 visualisation of this type of event topology can be seen in Figure 3.3d. Like the  $gg$  topology,  
 935 it contributes to only a small fraction of events in the generated sample.

936 An overview of the relative occurrence of these topologies in the generated sample can be seen in  
 937 Figure 3.2a. Clearly, the  $c\bar{c}$  topology dominates with a relative proportion > 60%, while the  $cg$

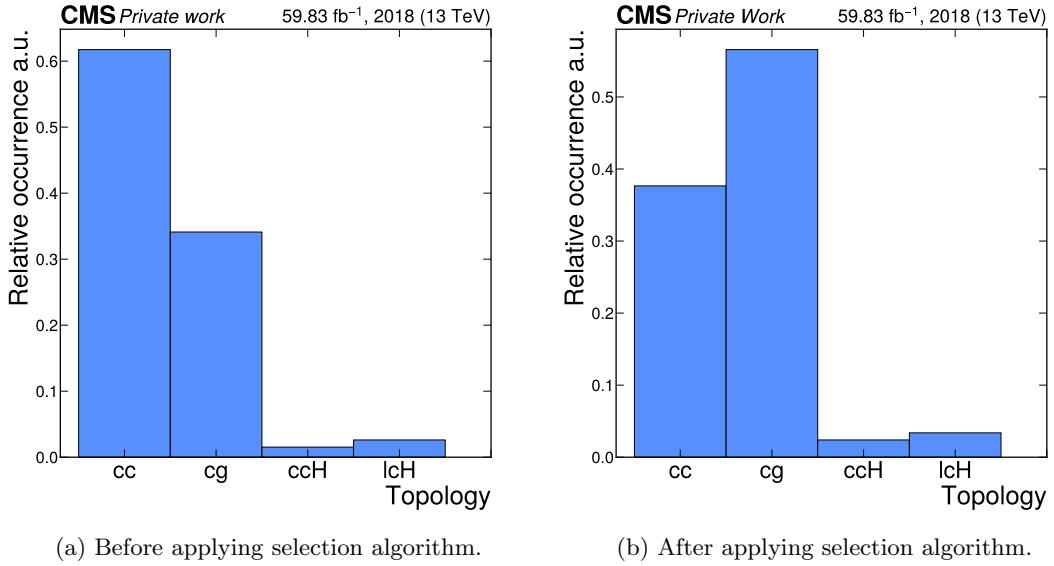


Figure 3.2: The relative occurrence of the  $c\bar{c}$ ,  $cg$ ,  $gg$  and  $cl$  event topologies in the simulated  $cH(ZZ \rightarrow 4\mu)$  sample before and after applying the selection algorithm.

938 topology only comprises  $\sim 35\%$  of events. Since the jet identification plays no role in the former  
 939 topology, it is worth considering how the selection algorithm described in section 3.1 affects the  
 940 relative occurrence of each topology. This can be seen in Figure 3.2b. Clearly, the selection  
 941 algorithm (specifically the jet selection) introduces a significant bias towards  $cg$  topologies as  
 942 expected and desired. However, a relative proportion of close to 40% of  $c\bar{c}$  topologies remain,  
 943 which may still contribute towards the sensitivity of the analysis. These are typically associated  
 944 with light flavour jets. This motivates the use of an inclusive statistical evaluation strategy,  
 945 exemplified by for example the use of the full CvsB and CvsL DeepJet distributions of the selected  
 946 jet, in contrast to the cut-based approach presented in [14].

947 To capture uncertainties associated with the choice of a particular flavour scheme in the simulation  
 948 of  $cH(ZZ \rightarrow 4\mu)$ , additional  $cH(ZZ \rightarrow 4\mu)$  samples are used. These specifically simulate the  
 949  $cH(ZZ \rightarrow 4\mu)$  process in the 3FS and 4FS, without the use of FxFx merging, effectively capturing  
 950 cases where a charm *must* originate from gluon splitting or directly from the proton respectively.  
 951 From these samples, an uncertainty envelope is constructed for the discriminators distribution  
 952 in the statistical evaluation presented in section 3.4. This envelope is determined by taking the  
 953 relative discrepancy between the 3FS and 4FS samples and interpreting it as an uncertainty band  
 954 around the the nominal 4FS FxFx sample distributions. An overview of all used signal samples  
 955 can be seen in Table 3.3.

956  
 957

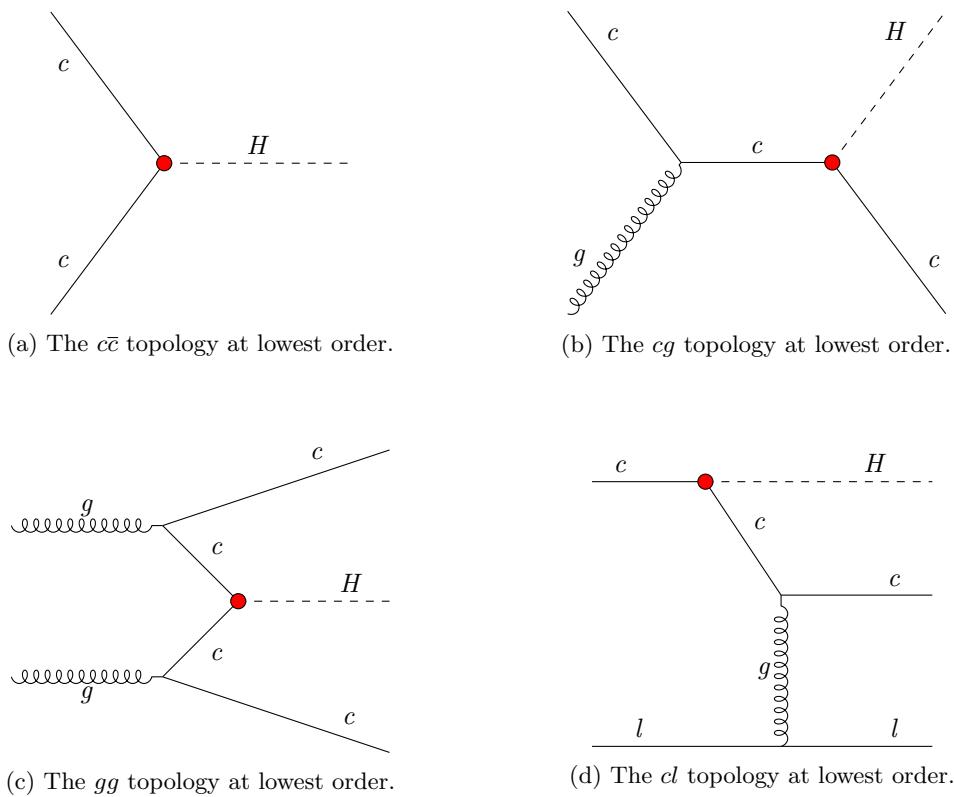


Figure 3.3: Examples of feynman diagrams that illustrate the four different event topologies generated in the  $cH(ZZ \rightarrow 4\mu)$  sample at the lowest order. The red dots represent the  $y_c$  coupling.

Table 3.3: cH(ZZ $\rightarrow$ 4 $\mu$ ) samples used in this work

Process	$\sigma$	# of simulated events
cH(ZZ $\rightarrow$ 4 $\mu$ ) 4FS FxFx	xx	
cH(ZZ $\rightarrow$ 4 $\mu$ ) 3FS	xx	
cH(ZZ $\rightarrow$ 4 $\mu$ ) 4FS	xx	

### 3.2.3 Estimation of irreducible backgrounds

Irreducible background processes are background processes that produce the same final state particles as the signal process in question. Thus, processes which produce four final state muons along with the presence of a, or several, jet(s) constitute the irreducible background. These again fall into two categories, namely those where the four muons originate from a Higgs boson and those in which they do not. When analysing e.g. the mass spectrum of Higgs candidates, the processes of the former category is thus clearly resonant around the Higgs mass of  $\sim$ 125 GeV, while those of the latter category may take on a more continuous shape. The irreducible backgrounds considered in this work are:

967

- **Gluon fusion (ggH):** In these processes, a pair of gluons produce an intermediary (top) quark pair loop that produces a Higgs boson. The Higgs boson decay may produce four muons while parton radiation may produce additional jets. An example Feynman diagram of the ggH process can be seen in Figure 3.4a.
- **Top quark pair associated Higgs production(ttH):** In ttH processes, a pair of top quarks is produced alongside a Higgs boson. The Higgs boson decay may produce four muons while the decay of the top quark pairs produces additional b jets. An example Feynman diagram of the ttH process can be seen in Figure 3.4b.
- **Vector boson associated Higgs production (VH):** The vector boson produced alongside a Higgs boson may refer to  $W^\pm$  and Z bosons. Thus, three different process types contribute to this background, namely  $W^+H$ ,  $W^-H$  and  $ZH$ . The Higgs boson decay may produce four muons while a  $W^\pm$  boson decay as well as a Z boson decay may produce additional jets. An example Feynman diagram of these processes can be seen in Figure 3.4c.
- **Vector boson fusion (qqH):** In vector boson fusion processes, two vector bosons emitted by initial state quarks fuse to produce a Higgs boson. The Higgs boson decay may produce four muons while additional final state quarks that produce jets are also present. An example Feynman diagram of the qqH process can be seen in Figure 3.4d.
- **Top quark plus quark associated Higgs production (tqH):** In tqH processes, a top quark as well as a quark of a different flavour are produced alongside a Higgs boson. The Higgs boson decay may produce four muons while the additional (top) quark will produce a (b) jet. An example Feynman diagram of the tqH process can be seen in Figure 3.4e.
- **ZZ production from gluons (gg $\rightarrow$ ZZ):** A pair of Z bosons can be produced from a pair of initial state gluons via an intermediate quark loop. The Z boson decays may produce four muons while additional parton radiation may produce jets. An example Feynman diagram of the gg $\rightarrow$  ZZ process can be seen in Figure 3.4f

- 993 • **ZZ production from quarks ( $qq \rightarrow ZZ$ ):** A pair of Z bosons can be produced from two  
 994 initial state quarks. The Z boson decays may produce four muons while additional parton  
 995 radiation may produce jets. An example Feynman diagram of the  $qq \rightarrow ZZ$  process can be  
 996 seen in Figure 3.4g.

997 The irreducible backgrounds are estimated using simulation. An overview of the samples that  
 998 are used can be found in Table 3.4. Like with the  $cH(ZZ \rightarrow 4\mu)$  process, additional samples are  
 999 used for the  $bH(ZZ \rightarrow 4\mu)$  background to account for an uncertainty related to the choice of  
 1000 flavour scheme. However, here the five flavour scheme (5FS) is now the nominal FS to include  
 1001 the bottom quark in the proton.

Table 3.4: Simulated processes used for background estimation in this work. The listed processes represent the irreducible backgrounds to the  $cH(ZZ \rightarrow 4\mu)$  analysis with the exception of the  $WZ \rightarrow 3\ell\nu$  process, which is used solely in the estimation of the irreducible backgrounds.

Process	$\sigma$	# of simulated events
$ggH(ZZ \rightarrow 4L)$	xx	
$ttH(ZZ \rightarrow 4L)$	xx	
$W^-H(ZZ \rightarrow 4L)$	xx	
$W^+H(ZZ \rightarrow 4L)$	xx	
$ZH(ZZ \rightarrow 4L)$	xx	
$qqH(ZZ \rightarrow 4L)$	xx	
$tqH(ZZ \rightarrow 4L)$	xx	
$gg \rightarrow ZZ(4\mu)$	xx	
$gg \rightarrow ZZ(4\tau)$	xx	
$gg \rightarrow ZZ(2\mu 2\tau)$	xx	
$qq \rightarrow ZZ/Z\gamma^* \rightarrow 4L$	xx	
<hr/>		
$bH$ 5FS $FxFx$	xx	
$bH$ 4FS	xx	
$bH$ 5FS	xx	
<hr/>		
$WZ \rightarrow 3\ell\nu$	xx	

### 1002 3.2.4 Estimation of reducible backgrounds

1003 Reducible background processes are background processes that do not produce the same final  
 1004 state particles as the signal process but where mis-identification of physics objects can still falsify  
 1005 the sought-after signature. Since jets are typically abundant in most collisions, this amounts to  
 1006 the mis-identification of additional muons for this analysis. Major contributions to this back-  
 1007 ground are expected to come from the Drell-Yan process while other processes which may produce  
 1008 at least two leptons, such as  $t\bar{t}$ , may also contribute. Example Feynman diagrams of these two  
 1009 processes can be seen in Figure 3.5. Since the simulation of mis-identified muons is subject to  
 1010 significant modelling uncertainties, a data-driven approach may be used. This involves deter-  
 1011 mining the mis-identification rate of muons in data and applying it to a side-band region from

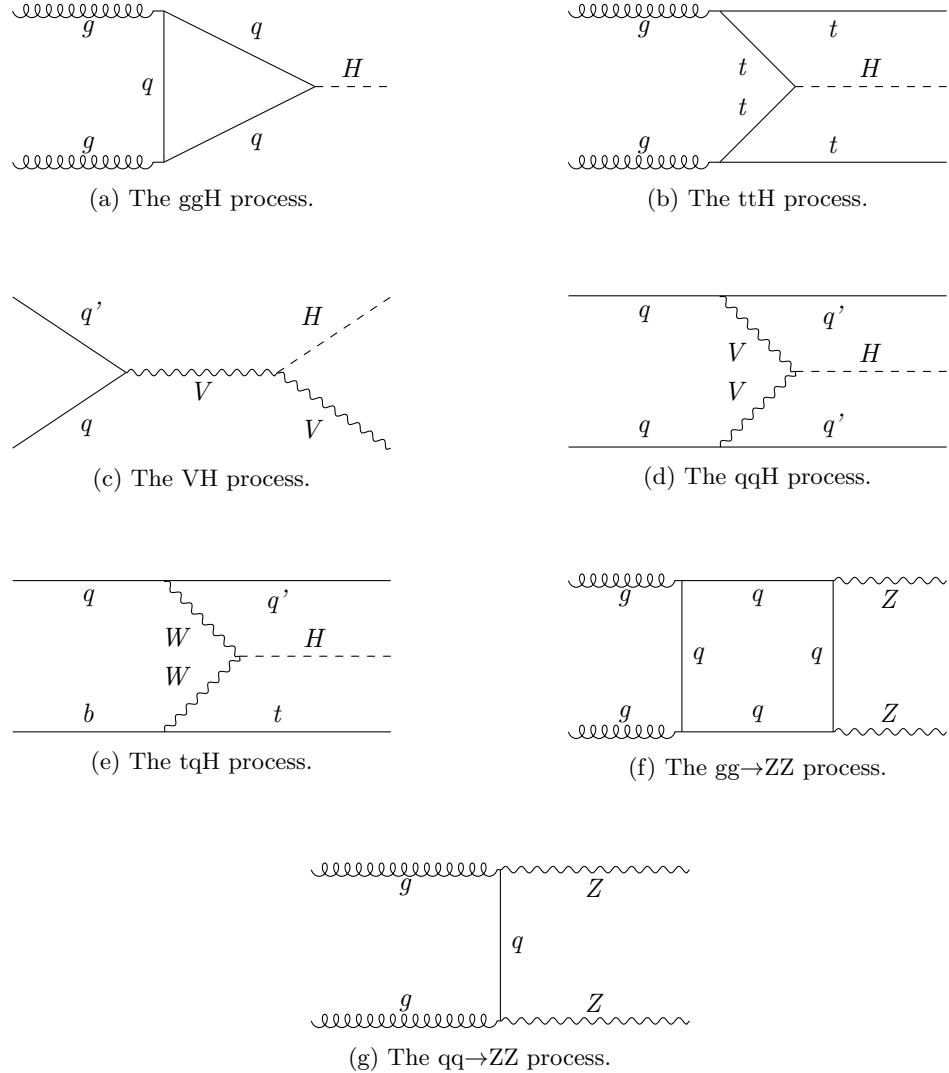


Figure 3.4: Representative Feynman diagrams of the irreducible background processes relevant to the  $cH(ZZ \rightarrow 4\mu)$  analysis. This includes the ggH, ttH, VH, qqH, tqH, gg $\rightarrow$ ZZ and qq $\rightarrow$ ZZ processes. When no final state parton is explicitly present in the diagram it is implied via, for example, parton radiation.



Figure 3.5: Representative Feynman diagrams of the Drell-Yan and  $t\bar{t}$  processes, which are expected to contribute significantly to the reducible background of the  $cH(ZZ \rightarrow 4\mu)$  analysis. Jets are produced via for example parton radiation for the Drell-Yan process, while the decay of the top quarks in the  $t\bar{t}$  process automatically produces b jets from the hadronisation of b quarks.

which the contributions of reducible backgrounds are extrapolated into the signal region. This methodology is presented in this section and follows that which is used in [53].

#### Determination of muon mis-identification rate

To determine the mis-identification of muons with respect to the tight muon requirement outlined in subsection 2.2.3, a three-muon selection is applied to data. Specifically, events with a  $Z \rightarrow \mu^+ \mu^-$  decay that also contain a third muon are targeted. Since hard-scattering processes that produces a Z boson are not expected to produce any additional muons, the third reconstructed muon (typically referred to as the *probe muon*) is assumed to be one that is mis-identified as such. To determine a mis-identification rate, the ratio of probe muons that pass the tight muon requirement with respect to those that pass the loose muon requirement is calculated. This procedure is performed in bins of the probe muon  $p_T$  for the barrel ( $|\eta| \leq 1.2$ ) and endcap ( $|\eta| > 1.2$ ) regions respectively. The exact reconstruction algorithm that is applied is the following:

- Events that contain at least two muons that pass the tight identification requirement and where the third passes at least the loose identification requirement are chosen. The  $p_T$ -leading muon is required to satisfy  $p_T > 20$  GeV and the sub-leading muon is required to satisfy  $p_T > 10$  GeV. Additionally, the HLT\_IsoMu24 trigger requirement must also be met. Lastly, to ensure two muons are not spuriously reconstructed from shared tracks, it is required that each muon candidate is separated from the others by  $\Delta R > 0.02$ .

- Opposite-sign muon pairs are merged into Z boson candidates and the candidate closest to the nominal Z mass is taken as the final Z candidate. Additionally, the invariant mass of any combination of opposite-sign muons must satisfy  $m_{\mu\mu} > 4$  GeV, to remove any contributions from low mass resonances such as  $J/\psi$ .

- The remaining, third muon not selected as part of the Z candidate is taken as the probe muon

The mis-identification rate of the probe muon that is determined in this way in bins of the probe muon  $p_T$ , can be seen in ???. However, the contribution from processes that indeed produce three muons in the hard-scattering must be subtracted. This consists primarily of  $WZ \rightarrow 3\ell\nu$  processes that artificially inflate the calculated mis-identification rate at higher probe muon  $p_T$ . This contribution is subtracted using simulation. It is this corrected version of the mis-identification rate that is used in the following section.

<sup>1042</sup> **Application of muon mis-identification rate**

The muon mis-identification rate is applied to a control region to estimate the contribution of reducible backgrounds to the previously described  $CH(ZZ \rightarrow 4\mu)$  selection. It is useful to introduce some of the related terminology at this point. The four pass (4P) region henceforth refers to the inclusive signal region that is defined via the  $CH(ZZ \rightarrow 4\mu)$  selection. The three-pass-one-fail (3P1F) and two-pass-two-fail (2P2F) regions respectively refer to regions in which the  $CH(ZZ \rightarrow 4\mu)$  reconstruction is performed as previously described but where only three (two) of the muons satisfy the tight identification criteria and the remaining one (two) muon(s) satisfy only the loose identification criteria. The 3P1F and 2P2F are collectively referred to as the application region (AR).

The extrapolation of the AR to the 4P is performed using the previously determined mis-identification rate. The prescription for this application can be obtained from the mis-identification rate  $f_i$  which is defined as

$$f = \frac{N_{\text{tight}}}{N_{\text{loose}}}. \quad (3.3)$$

Here  $N_{\text{loose}}$  and  $N_{\text{tight}}$  are the number of probe muons in a given bin that pass the loose and tight identification criteria respectively. From this, the relation

$$N_{\text{tight}} = N_{\text{loose}} f \quad (3.4)$$

follows. Since one is interested in the contributions of muons which pass the loose but not the tight identification requirement in the AR, Equation 3.4 can be reinterpreted as

$$N_{\text{loose-not-tight}} = N_{\text{loose}}(1 - f). \quad (3.5)$$

By substituting this back into Equation 3.4, the desired prescription is found:

$$N_{\text{tight}} = N_{\text{loose-not-tight}} \frac{f}{(1 - f)}. \quad (3.6)$$

<sup>1043</sup> Thus, for each muon that fails the tight identification requirements but passes the loose ones  
<sup>1044</sup> in the 3P1F and 2P2F regions, the weight  $f/(1 - f)$  is applied, where  $f$  is the  $p_T$  and  $\eta$   
<sup>1045</sup> dependant muon misidentification rate. This leads to the following expressions for the individual  
<sup>1046</sup> contributions of the AR to the 4P region:

<sup>1047</sup> **1. 2P2F:** Since this region contains two muons that pass the loose identification criteria but  
<sup>1048</sup> not the tight, the weight  $f/(1 - f)$  must be applied twice. The total contribution of this  
<sup>1049</sup> region in the 4P region can thus be written as

$$N_{4P}^{(2P2F)} = \sum_k^{N_{2P2F}} \frac{f_k^{(3)}}{(1 - f_k^{(3)})} \frac{f_k^{(4)}}{(1 - f_k^{(4)})}, \quad (3.7)$$

<sup>1050</sup> where the  $f_k^{(3/4)}$  is the misidentification rate associated with each of the non-passing muon  
<sup>1051</sup> for the  $k$ -th event. Major contributors to the 2P2F region are expected to be Drell-Yan  
<sup>1052</sup> and  $t\bar{t}$  processes, which produce only two prompt muons.

- 1053 2. **3P1F:** Since this region contains only one muon that passes the loose identification criteria  
 1054 but not the tight, the weight  $f/(1-f)$  is only applied. The total contribution of this region  
 1055 in the 4P region can thus be written as

$$N_{4P}^{(3P1F)} = \sum_k^{N_{3P1F}} \frac{f_k^{(4)}}{(1 - f_k^{(4)})}, \quad (3.8)$$

1056 where the  $f_k^{(4)}$  is the misidentification rate associated with the non-passing muon for the  
 1057  $k$ -th event. The major contributor to the 3P1F region is expected to consist of  $WZ \rightarrow 3\ell\nu$   
 1058 due to the presence of three prompt leptons.

To obtain the total contribution of the AR to the 4P region, potential contaminations of the 2P2F and 3P1F regions must be accounted for. The first source of contamination is the potential overlap of the 3P1F region with contributions from the 2P2F region, where an additional muon has been mis-identified in the former and erroneously passes the tight identification criteria. This may lead to an overestimation of the 3P1F region. Such contributions can be estimated via the term

$$N_{\text{exp.}}^{(3P1F)} = \sum_k^{N_{2P2F}} \left( \frac{f_k^{(3)}}{(1 - f_k^{(3)})} + \frac{f_k^{(4)}}{(1 - f_k^{(4)})} \right). \quad (3.9)$$

Effectively, contributions from the 2P2F region are weighed with the mis-identification rate for both muons that fail the tight identification criteria. To then extrapolate this contribution to the 4P region, the fake rate must once again be applied, this time to the respective complementary muon. This produces the final term

$$N_{4P\text{exp.}}^{(3P1F)} = \sum_k^{N_{2P2F}} \left( \frac{f_k^{(3)}}{(1 - f_k^{(3)})} \frac{f_k^{(4)}}{(1 - f_k^{(4)})} + \frac{f_k^{(4)}}{(1 - f_k^{(4)})} \frac{f_k^{(3)}}{(1 - f_k^{(3)})} \right) \quad (3.10)$$

$$= 2 \sum_k^{N_{2P2F}} \left( \frac{f_k^{(3)}}{(1 - f_k^{(3)})} \frac{f_k^{(4)}}{(1 - f_k^{(4)})} \right) \quad (3.11)$$

An additional source of potential contamination is the contribution of four muon processes in which muons either fail the tight identification criteria or fall outside the detector acceptance. This is primarily relevant in the 3P1F region with contributions from  $qq \rightarrow ZZ/Z\gamma^* \rightarrow 4L$  processes, which again lead to an overestimation of the 3P1F region. Contributions to the 4P region from this, denoted with  $N_{4P}^{(ZZ,3P1F)}$  are estimated via simulation with the same prescription as data events in the 3P1F region:

$$N_{4P}^{(ZZ,3P1F)} = \sum_k^{N_{3P1F}} \frac{f_k^{(4)}}{(1 - f_k^{(4)})}, \quad (3.12)$$

The total contribution to the 4P region  $N_{4P}$  may finally be estimated by taking the sum of the

2P2F and 3P1F contributions and subtracting the discussed contamination terms. This leads to

$$N_{4P} = N_{4P}^{(2P2F)} + N_{4P}^{(3P1F)} - N_{4P}^{(3P1F)} - N_{4P}^{(ZZ,3P1F)} \quad (3.13)$$

$$= \sum_k^{N_{3P1F}} \frac{f_k^{(4)}}{(1 - f_k^{(4)})} - \sum_k^{N_{3P1F}^{ZZ}} \frac{f_k^{(4)}}{(1 - f_k^{(4)})} - \sum_k^{N_{2P2F}} \left( \frac{f_k^{(3)}}{(1 - f_k^{(3)})} \frac{f_k^{(4)}}{(1 - f_k^{(4)})} \right) \quad (3.14)$$

1059

1060 However, it is found that due to the strong muon identification criteria an extremely small  
 1061 yield of reducible backgrounds are observed. Since this leads to non-continuous distributions of  
 1062 observables, a simplification must be made for the final statistical evaluation.

1063

...  
  

### 1064 3.3 Validation of cH(ZZ→ 4μ) event selection in sideband 1065 regions

1066 Here, a validation of the described event selection is presented. Specifically, the 2018 dataset of  
 1067 the CMS detector is compared to simulation in sideband regions of the signal for a variety of  
 1068 observables. These side-band regions include all selected events outside of the 120 GeV < m(H)  
 1069 >130 GeV region, which is where the cH(ZZ→ 4μ) is expected to be found. The ..

### 1070 3.4 Statistical inference

1071 For the statistical inference process, both the DeepJet CvsB and CvsL discriminators are used  
 1072 in a pseuo-two-dimensional fit. Specifically, an *unrolling* process is applied to events to project  
 1073 the two-dimensional information given by the CvsB and CvsL discriminators into a single dis-  
 1074 criminator. This involves constructing a histogram of the CvsB discriminator in bins of the  
 1075 corresponding CvsL value. The resulting histogram can be seen in ???. It is on these distribu-  
 1076 tions that the statistical inference process described in the following is performed. This consists  
 1077 of constructing a statistical model with which a fit may be performed as described in subsec-  
 1078 tion 3.4.1 as well as an uncertainty model, as described in section 3.5

1079

1080

#### 1081 3.4.1 Statistical model

Assuming the absence of a measurable signal, a upper limit on  $\sigma\text{BR}(\text{cH}(Z Z \rightarrow 4\mu))$  may be set. This in turn may be interpreted as a limit on  $\kappa_c$  using the prescription derived in subsection 1.3.2. To do this, a statistical model is required to predict the distribution of the number entries in each bin of the discriminator histogram according to a given hypothesis. Since each of the bins are filled in a discrete couting process, a Poissonian Ansatz with

$$\mathcal{P}(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (3.15)$$

is appropriate. Here,  $\lambda$  denotes the expectation value with  $k$  observed events. In the  $i$ -th bin of the discriminator distribution, the expecation value  $\lambda_i$  is calculated using the models of the signal and background processes discussed in section 3.2. This amounts to  $\lambda_i = \mu s_i + b_i$  where  $s_i$  and  $b_i$  are the signal and background estimations respectively. The signal strength modifier  $\mu$

allows for an arbitrary scaling of the signal contribution and is used as a floating parameter in the inference process. From this, a combined binned likelihood function

$$\mathcal{L}(d | \mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})) = \prod_{i \in \text{bins}} \mathcal{P}(d_i | \mu \cdot s(\boldsymbol{\theta}) + b(\boldsymbol{\theta})) \times \prod_{j \in \text{nuis.}} \mathcal{C}(\hat{\theta}_j | \theta_j). \quad (3.16)$$

may be derived for some set of measured data  $d_i$ . Here, an uncertainty model is introduced via the nuisance parameters  $\boldsymbol{\theta}$  that account for uncertainties related to the signal and background estimation. These follow a distribution  $\mathcal{C}$  and may alter the scaling of the signal and background contributions as well as their histogram shape with respect to the discriminator that is us. The estimate of  $\boldsymbol{\theta}$  that is used to obtain estimations of  $s_i$  and  $b_i$  in the inference process is denoted by  $\hat{\boldsymbol{\theta}}$  and is found by finding the global maximum of the likelihood. The specifics of the uncertainty model employed in this analysis are discussed in section 3.5. Since the statistical evaluation is not applied to data in this analysis, an Asimov dataset [**AsimovMaybe?**] is used. To compare different hypotheses, the test statistic that is used is the *profile likelihood ratio*

$$q_\mu = -2\ln \left( \frac{\mathcal{L}(d | \mu \cdot s(\hat{\boldsymbol{\theta}}_\mu) + b(\hat{\boldsymbol{\theta}}_\mu))}{\mathcal{L}(d | \hat{\mu} \cdot s(\hat{\boldsymbol{\theta}}) + b(\hat{\boldsymbol{\theta}}))} \right), \quad (3.17)$$

where  $(\hat{\mu}, \hat{\boldsymbol{\theta}})$  are the parameter values that globally maximise the likelihood while  $\hat{\boldsymbol{\theta}}_\mu$  maximises the likelihood for a given  $\mu$ . A significant advantage of this test statistic is that in the large sample limit, the distribution  $f(q | \mu)$  approaches the  $\chi^2_k$  distribution with  $k = 1$  degrees of freedom, according to Wilk's theorem [**wilksTheorem**]. This is very useful as knowing the distribution of the test statistic allows one to calculate a p-value

$$p_\mu = \int_{q_{\text{obs.}}}^{\infty} f(q | \mu) dq. \quad (3.18)$$

With this p-value, the CL<sub>s</sub> method may be applied [**CLs**]. For this, a CL<sub>s</sub> value

$$\text{CL}_s = \frac{p_\mu}{1 - p_0} = \frac{\int_{q_{\text{obs.}}}^{\infty} f(q | \mu) dq}{\int_{q_{\text{obs.}}}^{\infty} f(q | 0) dq}. \quad (3.19)$$

is computed. A feature of this method is that large overlaps of test statistic distributions between the  $H_\mu$  and  $H_0$  hypotheses, due to the very small signal cross section associated with  $H_\mu$ , are accounted for. This is achieved by weighting  $p_\mu$  with  $(1 - p_0)^{-1}$ , thus decreasing the CL<sub>s</sub> value with larger overlaps.

### 3.4.2 Uncertainty model

The sources of uncertainty in the model of the signal and background processes are described in this section. These consist of two types. The first are shape uncertainties, which may change the normalisation as well as the shape of the distributions in question. These uncertainties are captured by creating variations of these distributions that are interpreted as  $\pm 1\sigma$  variations. The second type consists of normalisation uncertainties. These only affect the normalisation of distributions without affecting the shape and can thus be captured as a single, real paarameter. Both types of uncertainties are associated with respective nuisance parameters in the statistical model introduced in subsection 3.4.1.

1100 **Theoretical uncertainties common to simulation**

1101 A number of theoretical uncertainties that are common to all simulated sample are included in  
1102 the uncertainty model. This includes:

- 1103 • Shape uncertainties related to the choice of the normalisation and factorisation scales  $\mu_R$  and  $\mu_F$ . These are varied independently by factors of 2 and 0.5 using a reweighting technique that is applied to the simulation, thus introducing four shape-changing nuisance parameters.
- 1107 • Shape uncertainties related to the modelling of the Parton Distribution Function (PDF). These uncertainties are obtained from the NNPDF3.1 PDF set that is used in simulation  
1108 [NNPDF3.1] and includes 100 PDF variations associated with variations of individual  
1109 parameters. These are combined into a single nuisance parameter with the prescription

$$\sigma^+ = \sigma^- = \sqrt{\sum_{i=1}^{N_{\text{par}}} (F_i - F_0)^2}, \quad (3.20)$$

1112 where  $F_0$  is the nominal value of the observable in question and  $F_i$  is the varied value  
1113 associated with one of the  $N_{\text{par}}$  parameter variations.

- 1114 • Normalisation uncertainties related to the cross sections of the individual processes. These  
1115 are:
  - 1116 – Higgs production and branching ratio uncertainties taken from [16]
  - 1117 – A 50% normalisation uncertainty on the modelling of the gluon fusion process in  
1118 association with heavy quarks ?? Ask gerrit.
  - 1119 – Other process uncertainties...
- 1120 • Shape uncertainties related to the tuning of the parton showering used in the generator. These are implemented by varying parameters related to initial-state radiation and final-state radiation independently, introducing two nuisance parameters.
- 1123 • A shape uncertainty related to the application of a pile-up reweighting procedure, with which the pile-up profile of simulation is matched to that of data. These uncertainties are captured via a single nuisance parameter.

1126 **Experimental uncertainties common to simulation**

1127 The following experimental uncertainties are common to all simulated sample:

- 1128 • Shape uncertainties related to the jet energy scale. For this, a simplified schema is used in  
1129 which closely correlated uncertainty sources are grouped and captured across 11 individual  
1130 nuisance parameters. Sources of uncertainty include for example the limited set of data  
1131 available for the calibration procedure or differences in calibration response to different jet  
1132 flavours.
- 1133 • A shape uncertainty related to the smearing method used to correct the jet energy resolution. This is captured via a single nuisance parameter.

- 1135 • Shape uncertainties related to the calibration of the jet flavour-tagging algorithm that is  
 1136 used. These are captured via 13 nuisance parameters and include for example uncertainty  
 1137 sources related to the individual phase spaces targeted in the calibration or the limited set  
 1138 of available data.
- 1139 • A shape uncertainty related to the pile-up identification method that is used. This is  
 1140 captured via a single nuisance parameter.
- 1141 • Shape uncertainties related to the muons used in the analysis. The response of the muon  
 1142 identification criteria, the isolation as well as the muon trigger path in simulation is cal-  
 1143 ibrated to match data. Uncertainties associated with this calibration are captured via a  
 1144 nuisance parameter.
- 1145 • A normalisation uncertainty related to the luminosity with which the simulation is scaled,  
 1146 which is known with a limited precision. This amounts 2.5% for the 2018 dataset and is  
 1147 captured as a single nuisance parameter.

### 1148 **Uncertainties related to $cH(ZZ \rightarrow 4\mu)$ and $bH(ZZ \rightarrow 4\mu)$ modelling**

1149 A systematic uncertainty that is unique to  $cH(ZZ \rightarrow 4\mu)$  and  $bH(ZZ \rightarrow 4\mu)$  is that associated with  
 1150 the use of a specific flavour scheme. As discussed in subsection 3.2.2, an envelope representing  
 1151 a  $\pm 1\sigma$  variation is extracted from the respective 3FS (4FS) and 4FS (5FS) samples and applied  
 1152 to the nominal  $cH(ZZ \rightarrow 4\mu)$  4FS FxFx ( $bH(ZZ \rightarrow 4\mu)$  5FS FxFx) sample.

### 1153 **Uncertainties related to reducible background modelling**

1154 Due to the almost negligible nature of the reducible background contribution, a simple normal-  
 1155 isation uncertainty is introduced to account for uncertainties associated with this estimation  
 1156 procedure...

### 1157 **Uncertainties related to statistical precision of background estimation**

1158 Simulation estimation methods themselves include a statistical uncertainty due to the limited  
 1159 number of events that are generated to estimate each process. This results in the introduction  
 1160 of nuisance parameter per process per bin of the final discriminator. This can be simplified to  
 1161 introducing only a single, per bin nuisance parameter by using the Barlow-Beeston approach  
 1162 [Barlow-Beeston].

## 1163 **3.5 Results**

1164 In this section, the result of the presented methods is discussed.



<sub>1165</sub> Chapter 4

<sub>1166</sub> An EFT interpretation of the  
<sub>1167</sub>  $cH(ZZ \rightarrow 4\mu)$  process



# **Conclusion**

1168



# <sup>1169</sup> Bibliography

- <sup>1170</sup> [1] S. Navas et al. “Review of Particle Physics”. In: *Phys. Rev. D* 110 (3 Aug. 2024), p. 030001.  
<sup>1171</sup> DOI: 10.1103/PhysRevD.110.030001. URL: <https://link.aps.org/doi/10.1103/PhysRevD.110.030001>.
- <sup>1173</sup> [2] P.W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- <sup>1176</sup> [3] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- <sup>1179</sup> [4] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716 (1 2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021.. URL: <https://www.sciencedirect.com/science/article/pii/S0370269312008581>.
- <sup>1183</sup> [5] G. Aad et al. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020. URL: <https://www.sciencedirect.com/science/article/pii/S037026931200857X>.
- <sup>1187</sup> [6] CMS Collaboration. “A portrait of the Higgs boson by the CMS experiment ten years after the discovery”. en. In: *Nature* 607.7917 (July 2022), pp. 60–68. DOI: 10.1038/s41586-022-04892-x. URL: <http://dx.doi.org/10.1038/s41586-022-04892-x>.
- <sup>1190</sup> [7] Nina M. Coyle, Carlos E. M. Wagner, and Viska Wei. “Bounding the charm Yukawa coupling”. In: *Phys. Rev. D* 100 (7 Oct. 2019), p. 073013. DOI: 10.1103/PhysRevD.100.073013. URL: <https://link.aps.org/doi/10.1103/PhysRevD.100.073013>.
- <sup>1193</sup> [8] Michael Edward Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley (1995) 842 p. Westview Press, 1995.
- <sup>1195</sup> [9] R. Wolf. *The Higgs Boson Discovery at the Large Hadron Collider*. Springer, 2015. ISBN: 978-3-319-18512-5, 978-3-319-18511-8. DOI: 10.1007/978-3-319-18512-5.
- <sup>1197</sup> [10] G. Arnison et al. “Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV”. In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)91177-2. URL: <https://www.sciencedirect.com/science/article/pii/0370269383911772>.
- <sup>1201</sup> [11] J. Goldstone, A. Salam, and S. Weinberg. “Broken Symmetries”. In: *Phys. Rev.* 127 (3 Aug. 1962), pp. 965–970. DOI: 10.1103/PhysRev.127.965. URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965>.

- 1204 [12] CMS Collaboration. *Simultaneous probe of the charm and bottom quark Yukawa couplings*  
 1205 *using ttH events*. 2025. DOI: 10.48550/arXiv.2509.22535. arXiv: 2509.22535 [hep-ex].  
 1206 URL: <https://arxiv.org/abs/2509.22535>.
- 1207 [13] CMS Collaboration. *Search for a Higgs boson produced in association with a charm quark*  
 1208 *and decaying to a W boson pair in proton-proton collisions at  $\sqrt{s} = 13$  TeV*. 2025. arXiv:  
 1209 2508.14988 [hep-ex]. URL: <https://arxiv.org/abs/2508.14988>.
- 1210 [14] CMS Collaboration. *Search for the associated production of a Higgs boson with a charm*  
 1211 *quark in the diphoton decay channel in pp collisions at  $\sqrt{s} = 13$  TeV*. 2025. arXiv: 2503.  
 1212 08797 [hep-ex]. URL: <https://arxiv.org/abs/2503.08797>.
- 1213 [15] Nuoyu Dong et al. “Probing charm Yukawa coupling through ch associated production  
 1214 at the hadron colliders”. In: *Phys. Rev. D* 111 (5 Mar. 2025), p. 053003. DOI: 10.1103/  
 1215 PhysRevD.111.053003. URL: <https://link.aps.org/doi/10.1103/PhysRevD.111.053003>.
- 1217 [16] C T Potter et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of*  
 1218 *the LHC Higgs Cross Section Working Group*. en. 2013. DOI: 10.5170/CERN-2013-004.  
 1219 URL: <http://cds.cern.ch/record/1559921>.
- 1220 [17] J. A. Aguilar-Saavedra, J. M. Cano, and J. M. No. “More light on Higgs flavor at the  
 1221 LHC: Higgs boson couplings to light quarks through  $h + \gamma$  production”. In: *Phys. Rev.*  
 1222 *D* 103 (9 May 2021), p. 095023. DOI: 10.1103/PhysRevD.103.095023. URL: <https://link.aps.org/doi/10.1103/PhysRevD.103.095023>.
- 1224 [18] The CMS Collaboration. “Combined measurements of Higgs boson couplings in proton-  
 1225 proton collisions at  $\sqrt{s} = 13$  TeV”. In: *The European Physical Journal C* 79.5 (May 2019).  
 1226 ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6909-y. URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6909-y>.
- 1228 [19] Steven Weinberg. “Baryon and Lepton Nonconserving Processes”. In: *Phys. Rev. Lett.* 43  
 1229 (1979), pp. 1566–1570. DOI: 10.1103/PhysRevLett.43.1566.
- 1230 [20] B. Grzadkowski et al. “Dimension-six terms in the Standard Model Lagrangian”. In:  
 1231 *Journal of High Energy Physics* 2010.10 (Oct. 2010). ISSN: 1029-8479. DOI: 10.1007/  
 1232 jhep10(2010)085. URL: [http://dx.doi.org/10.1007/JHEP10\(2010\)085](http://dx.doi.org/10.1007/JHEP10(2010)085).
- 1233 [21] Gino Isidori, Felix Wilsch, and Daniel Wyler. “The standard model effective field theory at  
 1234 work”. In: *Reviews of Modern Physics* 96.1 (Mar. 2024). ISSN: 1539-0756. DOI: 10.1103/  
 1235 revmodphys.96.015006. URL: <http://dx.doi.org/10.1103/RevModPhys.96.015006>.
- 1236 [22] CMS Collaboration. *Combined effective field theory interpretation of Higgs boson, elec-*  
 1237 *troweak vector boson, top quark, and multi-jet measurements*. 2025. arXiv: 2504.02958  
 1238 [hep-ex]. URL: <https://arxiv.org/abs/2504.02958>.
- 1239 [23] Joseph Bramante et al. “Boosted Higgs bosons from chromomagnetic b’s:  $b\bar{b}h$  at high  
 1240 luminosity”. In: *Phys. Rev. D* 93 (5 Mar. 2016), p. 053001. DOI: 10.1103/PhysRevD.93.  
 1241 053001. URL: <https://link.aps.org/doi/10.1103/PhysRevD.93.053001>.
- 1242 [24] J. Elias-Miró et al. “Higgs windows to new physics through d=6 operators: constraints and  
 1243 one-loop anomalous dimensions”. In: *Journal of High Energy Physics* 2013.11 (Nov. 2013).  
 1244 ISSN: 1029-8479. DOI: 10.1007/jhep11(2013)066. URL: [http://dx.doi.org/10.1007/JHEP11\(2013\)066](http://dx.doi.org/10.1007/JHEP11(2013)066).
- 1246 [25] The CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of In-*  
 1247 *strumentation* 3.08 (Aug. 2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004. URL:  
 1248 <https://doi.org/10.1088/1748-0221/3/08/S08004>.

- [26] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. doi: 10.5170/CERN-2004-003-V-1. URL: <http://cds.cern.ch/record/782076>.
- [27] E. A. Mobs. “The CERN accelerator complex. Complexe des accélérateurs du CERN”. In: (Oct. 2016). General Photo. URL: <http://cds.cern.ch/record/2225847>.
- [28] The CMS Collaboration. *CMS Physics : Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. Geneva: CERN, 2006. URL: <https://cds.cern.ch/record/922757>.
- [29] CMS Collaboration. “Development of the CMS detector for the CERN LHC Run 3”. In: *Journal of Instrumentation* 19.05 (May 2024), P05064. ISSN: 1748-0221. doi: 10.1088/1748-0221/19/05/p05064. URL: <http://dx.doi.org/10.1088/1748-0221/19/05/P05064>.
- [30] *CMS Tracker Detector Performance Public Results*. Last accessed 24.11.25. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.
- [31] The CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009–P10009. ISSN: 1748-0221. doi: 10.1088/1748-0221/9/10/p10009. URL: <http://dx.doi.org/10.1088/1748-0221/9/10/P10009>.
- [32] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Tech. rep. Geneva, 1997. URL: <https://cds.cern.ch/record/349375>.
- [33] CMS Collaboration. *The CMS ECAL performance with examples*. Tech. rep. Geneva: CERN, 2014. doi: 10.1088/1748-0221/9/02/C02008. URL: <https://cds.cern.ch/record/1632384>.
- [34] CMS Collaboration. *The CMS ECAL performance with examples*. Tech. rep. Geneva: CERN, 2014. doi: 10.1088/1748-0221/9/02/C02008. URL: <https://cds.cern.ch/record/1632384>.
- [35] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Tech. rep. Geneva, 1997. URL: <https://cds.cern.ch/record/357153>.
- [36] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Tech. rep. Geneva, 1997. URL: <https://cds.cern.ch/record/331056>.
- [37] CMS Collaboration. “Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays”. In: *Journal of Instrumentation* 5.03 (Mar. 2010), T03021–T03021. ISSN: 1748-0221. doi: 10.1088/1748-0221/5/03/t03021. URL: <http://dx.doi.org/10.1088/1748-0221/5/03/T03021>.
- [38] CMS Collaboration. *The CMS muon project: Technical Design Report*. Tech. rep. Geneva, 1997. URL: <https://cds.cern.ch/record/343814>.
- [39] The CMS collaboration. “Performance of CMS muon reconstruction in pp collision events at  $\sqrt{s} = 7\text{TeV}$ ”. In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10002–P10002. ISSN: 1748-0221. doi: 10.1088/1748-0221/7/10/p10002. URL: <http://dx.doi.org/10.1088/1748-0221/7/10/P10002>.
- [40] Mia Tosi. *The CMS trigger in Run 2*. Tech. rep. Geneva: CERN, Oct. 2017. doi: 10.22323/1.314.0523. URL: <https://cds.cern.ch/record/2290106>.
- [41] CMS collaboration et al. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (2014), P10009.

- 1293 [42] K. Rose. “Deterministic annealing for clustering, compression, classification, regression,  
1294 and related optimization problems”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–  
1295 2239. DOI: 10.1109/5.726788.
- 1296 [43] A. M. Sirunyan et al. “Particle-flow reconstruction and global event description with the  
1297 CMS detector.” In: *JINST* 12 (June 2017). Replaced with the published version. Added  
1298 the journal reference and DOI. All the figures and tables can be found at <http://cms->  
1299 results.web.cern.ch/cms-results/public-results/publications/PRF-14-001(CMS Public Pages),  
1300 P10003. 82 p. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965. URL: <https://cds.cern.ch/record/2270046>.
- 1302 [44] CMS Collaboration. “Pileup mitigation at CMS in 13 TeV data”. In: *Journal of Instrumentation*  
1303 15.09 (Sept. 2020), P09018–P09018. ISSN: 1748-0221. DOI: 10.1088/1748-  
1304 0221/15/09/p09018. URL: <http://dx.doi.org/10.1088/1748-0221/15/09/P09018>.
- 1305 [45] B. Andersson et al. “Parton fragmentation and string dynamics”. In: *Physics Reports*  
1306 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: [https://doi.org/10.1016/0370-1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7). URL: <https://www.sciencedirect.com/science/article/pii/0370157383900807>.
- 1309 [46] M. Cacciari et al. “FastJet user manual”. In: *The European Physical Journal C* 72.3 (Mar.  
1310 2012), p. 1896. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-1896-2. URL: <https://doi.org/10.1140/epjc/s10052-012-1896-2>.
- 1312 [47] CMS Collaboration. *Jet algorithms performance in 13 TeV data*. Tech. rep. Geneva: CERN,  
1313 2017. URL: <https://cds.cern.ch/record/2256875>.
- 1314 [48] CMS Collaboration. “Jet energy scale and resolution in the CMS experiment in pp collisions  
1315 at 8 TeV”. In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. ISSN: 1748-  
1316 0221. DOI: 10.1088/1748-0221/12/02/p02014. URL: <http://dx.doi.org/10.1088/1748-0221/12/02/P02014>.
- 1318 [49] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas”. In: *Physics  
1319 Letters B* 659.1–2 (Jan. 2008), pp. 119–126. ISSN: 0370-2693. DOI: 10.1016/j.physletb.  
1320 2007.09.077. URL: <http://dx.doi.org/10.1016/j.physletb.2007.09.077>.
- 1321 [50] CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp  
1322 collisions at 13 TeV”. In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011.  
1323 ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/05/p05011. URL: <http://dx.doi.org/10.1088/1748-0221/13/05/P05011>.
- 1325 [51] E. Bols et al. “Jet flavour classification using DeepJet”. In: *Journal of Instrumentation*  
1326 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/  
1327 p12012. URL: <http://dx.doi.org/10.1088/1748-0221/15/12/P12012>.
- 1328 [52] CMS Collaboration. “A new calibration method for charm jet identification validated with  
1329 proton-proton collision events at  $\sqrt{s} = 13$  TeV”. In: *Journal of Instrumentation* 17.03  
1330 (Mar. 2022), P03014. ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/03/p03014. URL:  
1331 <http://dx.doi.org/10.1088/1748-0221/17/03/P03014>.
- 1332 [53] CMS Collaboration. “Measurements of production cross sections of the Higgs boson in the  
1333 four-lepton final state in proton–proton collisions at  
1334  $\sqrt{s} = 13$  TeV  
1335 ”. In: *The European Physical Journal C* 81.6 (June 2021). ISSN: 1434-6052. DOI: 10.1140/  
1336 epjc/s10052-021-09200-x. URL: <http://dx.doi.org/10.1140/epjc/s10052-021-09200-x>.

- [54] W. K. Hastings. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications”. In: *Biometrika* 57.1 (1970), pp. 97–109. ISSN: 00063444, 14643510. URL: <http://www.jstor.org/stable/2334940> (visited on 12/22/2025).
- [55] G. Altarelli and G. Parisi. “Asymptotic freedom in parton language”. In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. ISSN: 0550-3213. DOI: [https://doi.org/10.1016/0550-3213\(77\)90384-4](https://doi.org/10.1016/0550-3213(77)90384-4). URL: <https://www.sciencedirect.com/science/article/pii/0550321377903844>.
- [56] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: [10.1007/jhep07\(2014\)079](10.1007/jhep07(2014)079). URL: [http://dx.doi.org/10.1007/JHEP07\(2014\)079](http://dx.doi.org/10.1007/JHEP07(2014)079).
- [57] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with parton shower simulations: the POWHEG method”. In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), p. 070. DOI: <10.1088/1126-6708/2007/11/070>. URL: <https://doi.org/10.1088/1126-6708/2007/11/070>.
- [58] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: <https://doi.org/10.1016/j.cpc.2015.01.024>. URL: <https://www.sciencedirect.com/science/article/pii/S0010465515000442>.
- [59] Rikkert Frederix and Stefano Frixione. “Merging meets matching in MC@NLO”. In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). ISSN: 1029-8479. DOI: [10.1007/jhep12\(2012\)061](10.1007/jhep12(2012)061). URL: [http://dx.doi.org/10.1007/JHEP12\(2012\)061](http://dx.doi.org/10.1007/JHEP12(2012)061).
- [60] Bo Andersson. *The Lund Model*. Vol. 7. Cambridge University Press, 1998. ISBN: 978-1-009-40129-6, 978-1-009-40125-8, 978-1-009-40128-9, 978-0-521-01734-3, 978-0-521-42094-5, 978-0-511-88149-7. DOI: <10.1017/9781009401296>.
- [61] CMS Collaboration. “Event generator tunes obtained from underlying event and multi-parton scattering measurements”. In: *The European Physical Journal C* 76.3 (Mar. 2016). ISSN: 1434-6052. DOI: <10.1140/epjc/s10052-016-3988-x>. URL: <http://dx.doi.org/10.1140/epjc/s10052-016-3988-x>.
- [62] “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8). URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [63] Stefan Höche. “Introduction to parton-shower event generators”. In: *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*. 2015, pp. 235–295. DOI: [10.1142/9789814678766\\_0005](10.1142/9789814678766_0005). arXiv: [1411.4085 \[hep-ph\]](1411.4085).



# <sup>1372</sup> Appendix

- <sup>1373</sup> **A Columnar analysis framework**
- <sup>1374</sup> **B Jet selection likelihood templates**
- <sup>1375</sup> **C Jet-parton matching efficiency**