



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	2.2 Event reconstruction with the CMS detector . . . . .	24
29	2.2.1 Track and vertex reconstruction . . . . .	24
30	2.2.2 The Particle Flow algorithm . . . . .	24
31	2.2.3 Reconstruction and identification of muons . . . . .	25
32	2.2.4 Reconstruction and identification of jets . . . . .	27
33	2.2.5 Missing transverse momentum . . . . .	28
34	2.3 Identification of charm quark-induced jets . . . . .	29
35	2.3.1 Properties of heavy-flavour jets . . . . .	29
36	2.3.2 The DeepJet algorithm . . . . .	30
37	<b>3 Search for the cH(<math>ZZ \rightarrow 4\mu</math>) process</b>	33
38	3.1 cH event selection . . . . .	33
39	3.1.1 Higgs candidate selection . . . . .	34
40	3.1.2 Jet candidate selection . . . . .	34
41	3.2 Signal and background estimation . . . . .	36

42	3.2.1	Estimation of cH process . . . . .	36
43	3.2.2	Estimation of irreducible backgrounds . . . . .	36
44	3.2.3	Estimation of reducible backgrounds . . . . .	36
45	3.3	Statistical evaluation . . . . .	36
46	<b>4</b>	<b>An EFT interpretation of the cH(<math>ZZ \rightarrow 4\mu</math>) process</b>	<b>37</b>
47	<b>Conclusion</b>		<b>39</b>
48	<b>Bibliography</b>		<b>41</b>

49 **Chapter 1**

50 **Introduction**

51 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.

67 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.

73 **1.1 The Standard Model of particle physics**

74 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.

82

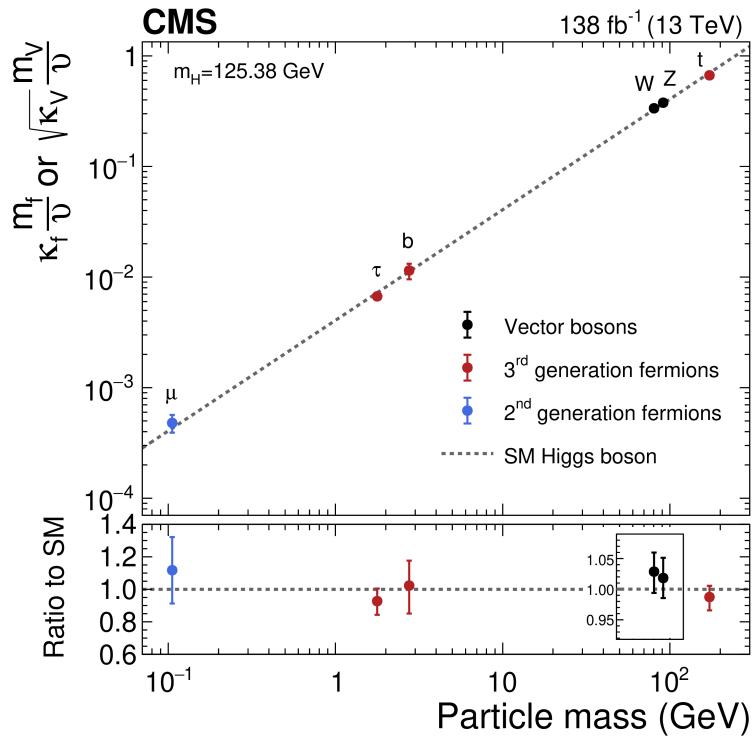


Figure 1.1: The measured coupling modifiers  $\kappa_f$  and  $\kappa_V$  of the coupling between the Higgs boson and fermions as well as heavy gauge bosons as functions of fermion or gauge boson mass  $m_f$  and  $m_V$ , where  $\nu$  is the vacuum expectation value of the Higgs field. [6]

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

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

88  
 89 However, there are distinct differences between the leptons and quarks. Leptons carry integer  
 90 (or no) charge while quarks carry fractional charges. More importantly, while both quarks and  
 91 leptons may interact via the electro-weak force, only the quarks interact via the strong force.  
 92 Due to the nature of the strong force, quarks almost exclusively form composite states called  
 93 hadrons. Lastly, the existence of anti-fermions must be mentioned. These carry the exact op-  
 94 posite quantum numbers (e.g. charge) as their fermion counterparts, though otherwise behave  
 95 similarly (take the electron and positron for instance). For simplicity, references to a fermion  
 96 in this work may be understood as referencing both the fermion and anti-fermion counterpart,  
 97 unless otherwise explicitly indicated. Examples of the latter are e.g. referring explicitly to elec-  
 98 trons  $e^-$  and positrons  $e^+$  or charm quark  $c$  and anti-charm quark  $\bar{c}$  pairs.

99  
 100 There exist 13 unique bosons in the SM. These include the photon  $\gamma$ ,  $W^\pm$  and  $Z$  which me-  
 101 diate the electro-weak force as well as 8 gluons  $g$  that mediate the strong force. The final piece  
 102 is the Higgs boson. Contrary to the force carriers, which all are spin 1, the Higgs boson is spin  
 103 0. By interacting with the Higgs boson, the massive particles of the SM acquire their mass and  
 104 is thus a central element of the SM.

105  
 106 Considering the introduced particles and forces, the SM has a rich and detailed phenome-  
 107 ology. A great example of a mathematically rigorous delineation of this can be found for example  
 108 in [8]. Given the focus of this work on the Yukawa coupling between the Higgs boson and charm  
 109 quark, only this aspect of the SM is discussed in further detail.

## 110 1.2 The Higgs-charm Yukawa coupling

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

115  
 116 To understand the origin of the Yukawa-couplings, a brief discussion of Lagrangian densities,  
 117 gauge transformations and the role of symmetries in the SM is warranted. The Lagrangian  
 118 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  
 119 equations of motions for the particles associated with these fields may be derived. Commonly,  
 120 theories of particles and their behaviour in a QFT are thus defined through the formulation of a  
 121 Lagrangian density. The form of this expression determines the nature of the particles that are  
 122 included as well as their interactions.

123  
 124 A central component to the way in which particle interactions are introduced in the SM is  
 125 the concept of gauge symmetries. These originate from the fact that the quantum fields in a

<sup>126</sup> QFT carry phase information, which may depend on the space-time coordinate of the field. This  
<sup>127</sup> phase information describes (local) degrees of freedom of the field and should have no effect on  
<sup>128</sup> the physical observables of the system. Thus,  $\mathcal{L}$  should remain invariant under arbitrary phase  
<sup>129</sup> transformations. Such transformations are typically referred to as a choice of gauge and such an  
<sup>130</sup> invariance is accordingly referred to as a *local gauge symmetry*.

<sup>131</sup>

<sup>132</sup> In the Lagrangian of the SM, invariance in the presence of local gauge symmetries is insured  
<sup>133</sup> through the addition of additional fields. These gauge fields couple to the previously existing  
<sup>134</sup> fields and effectively serve as mediators of phase information between space-time points of the  
<sup>135</sup> original fields. It is exactly these gauge fields which we identify as the fields force-mediating  
<sup>136</sup> bosons introduced previously and which are required to maintain local gauge symmetry. A very  
<sup>137</sup> interesting conclusion from this is that the dynamics of the bosons and the corresponding force  
<sup>138</sup> are determined entirely by the structure of the local gauge symmetry that must be preserved.  
<sup>139</sup> For the electro-weak force, the corresponding symmetry is referred to as  $SU(2)_L \times U(1)_Y$ . Here,  
<sup>140</sup> the  $L$  denotes that the associated force only acts on left-handed chiral particles while the  $Y$   
<sup>141</sup> denotes the charge that is carried by the corresponding bosons and is referred to as the weak  
<sup>142</sup> hypercharge. There are a total of four boson associated with the electro-weak force. These are  
<sup>143</sup> the photon  $\gamma$  that mediates the electromagnetic force as well as the electromagnetically charged  
<sup>144</sup>  $W^\pm$  and electromagnetically neutral Z boson that mediate the weak force.

<sup>145</sup>

<sup>146</sup> With these concepts in mind the nature of the electro-weak sector's Lagrangian in the SM  
<sup>147</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

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

<sup>148</sup>  
<sup>149</sup> while the Z boson and photon  $\gamma$  arise from linear combinations of the  $W_3$  and  $B$  gauge fields  
<sup>150</sup> 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

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

<sup>151</sup>  
<sup>152</sup> do not remain invariant under arbitrary  $SU(2)_L$  gauge transformations. This is as gauge fields  
<sup>153</sup>  $A_\mu$  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>154</sup>

<sup>155</sup> There is however a subtle distinction in this case, as the invariance breaking terms in Equation 1.7 arise from the different transformation behaviour of the  $\psi_L$  and  $\psi_R$  components of  $\psi$  under  $SU(2)_L \times U(1)_Y$  gauge transformations.

<sup>156</sup>

<sup>157</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>159</sup>

<sup>160</sup> The center of the potential is now an unstable local maximum and the only stable configuration

<sup>161</sup> can be found in the non-zero ground state. Through this, the symmetry of the potential is  
<sup>162</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>163</sup>  
<sup>164</sup> where  $h$  is a new scalar field that is used to parametrise radial perturbations of the potential's  
<sup>165</sup> ground state. This choice is referred to as the unitary gauge and  $h$  is identified as the field  
<sup>166</sup> corresponding to the physical Higgs boson. By expanding Equation 1.8 with this choice of  $\phi$ , a  
<sup>167</sup> range of terms are introduced to  $\mathcal{L}_{\text{EW}}$ . These contain a variety of interaction terms between the  
<sup>168</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>169</sup>  
<sup>170</sup> This can be understood to mean that the electro-weak coupling constants  $g$  and  $g'$  along with  $v$   
<sup>171</sup> effectively determine the mass of the Z and  $W^\pm$  bosons. A full description and compilation of  
<sup>172</sup> all the terms of the electro-weak Lagrangian density of the SM can be found in [9].

### <sup>173</sup> 1.2.2 The Yukawa couplings

<sup>174</sup> By including the Higgs contribution in our theory, mass terms for fermions may now be generated  
<sup>175</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>176</sup>  
<sup>177</sup> is obtained. A curious feature of the SM is that the Yukawa-couplings  $y_f$  are free parameters  
<sup>178</sup> of the theory with no a priori values. As a result these must be measured experimentally, with  
<sup>179</sup> the measurement of the charm quark Yukawa coupling  $y_c$  being the goal of this work. Since the  
<sup>180</sup> charm quark mass has previously been determined from experiment to be  $m_c = 1.27$  GeV [1],  
<sup>181</sup> a measurement of  $y_c$  thus represents an important consistency test of the SM. To this end, one  
<sup>182</sup> can exploited that an interaction between fermions and the Higgs field is introduced as can be  
<sup>183</sup> seen in Equation 1.17, with an interaction strength proportional to  $y_c$ . It is exactly this feature  
<sup>184</sup> that may be exploited by experiments at the LHC to measure  $y_c$ .

<sup>185</sup> **1.3 Measuring the charm quark Yukawa coupling**

<sup>186</sup> By measuring the frequency of occurrence of physics processes in which the coupling between the  
<sup>187</sup> Higgs boson and charm quark appears,  $y_c$  may be determined. As such, a suitable process must  
<sup>188</sup> be found that can be detected by an experiment such as CMS. These fall into two categories. The  
<sup>189</sup> first consists of processes in which a Higgs boson decays into a charm and anti-charm quark pair  
<sup>190</sup> ( $H \rightarrow c\bar{c}$ ). Previous analysis of e.g. top quark pair and vector boson associated Higgs production  
<sup>191</sup> has been able to observe a 95% CL upper limit on the charm quark Yukawa coupling modifier  
<sup>192</sup>  $\kappa_c$  (see subsection 1.3.2 for a detailed discussion) of  $|\kappa_c| < |3.5|$  [12], the most stringent limit to  
<sup>193</sup> date. The second category consists of processes in which a Higgs boson is produced in association  
<sup>194</sup> with a charm quark. This latter category of processes is the focus of this work and is henceforth  
<sup>195</sup> referred to as the cH process.

<sup>196</sup> **1.3.1 The cH process**

<sup>197</sup> The cH process encompasses processes in proton-proton collisions in which a charm-quark is  
<sup>198</sup> produced alongside a Higgs boson. At leading order, this consists of 2 processes sensitive to  $y_c$ ,  
<sup>199</sup> represented by the Feynman diagrams shown in Figure 1.3. The first two diagrams, namely the s  
<sup>200</sup> and t-channel diagrams, constitute the  $y_c$  sensitive contribution. There exist also additional cH  
<sup>201</sup> processes, mediated through the effective Higgs boson to gluon coupling, which are not sensitive  
<sup>202</sup> to  $y_c$ . These account for approximately 80% of the inclusive cH cross section and thus represents  
<sup>203</sup> a significant background to the cH process sensitive to the charm quark Yukawa coupling.

<sup>204</sup> Targeting the cH process to measure  $y_c$  is a relatively novel strategy in comparison to targeting  
<sup>205</sup>  $H \rightarrow c\bar{c}$ . A key advantage of this approach is that contributions from the abundant QCD  
<sup>206</sup> background at the LHC are greatly reduced due to only needing to identify the flavour of single  
<sup>207</sup> jet resulting from a charm quark, as opposed to two. Additionally, since the sensitivity to  $y_c$  does  
<sup>208</sup> not originate from the decay of the Higgs boson, the Higgs boson decay mode to target can be  
<sup>209</sup> chosen freely. Especially signatures such as  $H \rightarrow ZZ \rightarrow 4\mu$ , which may be resolved cleanly by an  
<sup>210</sup> experiment such as CMS, can be targeted. However, an analysis of the cH process also comes  
<sup>211</sup> with drawbacks. A significant experimental difficulty results from the fact that the associated  
<sup>212</sup> charm flavour jets are typically produced at very low transverse momenta  $p_T$ , as seen in Figure  
<sup>213</sup> 1.2. These can be experimentally difficult to reconstruct and thus a significant portion of  
<sup>214</sup> this signal may be lost due to detector acceptance effects. Another drawback is that Higgs boson  
<sup>215</sup> decay channels such as  $H \rightarrow ZZ \rightarrow 4\mu$  have very small branching ratios (e.g.  $BR(H \rightarrow ZZ \rightarrow 4\mu) =$   
<sup>216</sup> 0.3% [1]) and thus the overall cross section of the cH process may be very small. As a result of  
<sup>217</sup> these effects, a key challenge of a search for the cH process is expected to lie in the statistical  
<sup>218</sup> uncertainty of the analysis.

<sup>219</sup> As a novel strategy, targeting the cH process is of recent interest and results in the cH(WW)  
<sup>220</sup> and cH( $\gamma\gamma$ ) channels using Run 2 data of the CMS experiment are published. Upper limits on  
<sup>221</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>222</sup> as sensitive as the limit observed in the  $H \rightarrow c\bar{c}$  channels, these nonetheless provide important  
<sup>223</sup> complementary results and can contribute significantly in combination. This is especially impor-  
<sup>224</sup> tant given that even at the High-Luminosity LHC, the projected sensitivity on the charm quark  
<sup>225</sup> Yukawa coupling in individual channels is only starting to approach one [15].

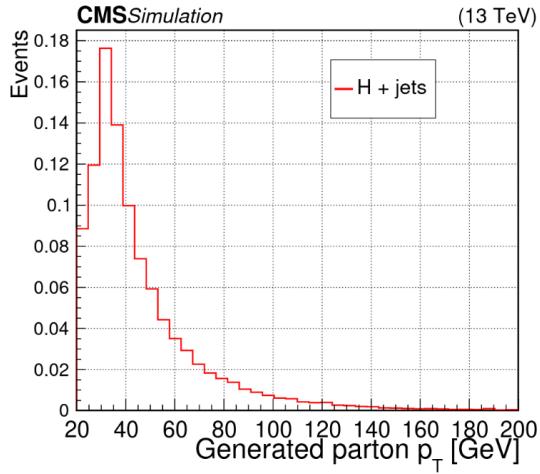


Figure 1.2: Transverse momentum of the charm quark produced alongside a Higgs boson in a simulation of the  $cH$  process, which typically takes on relatively small values. THIS IS A PLACEHOLDER PLOT.

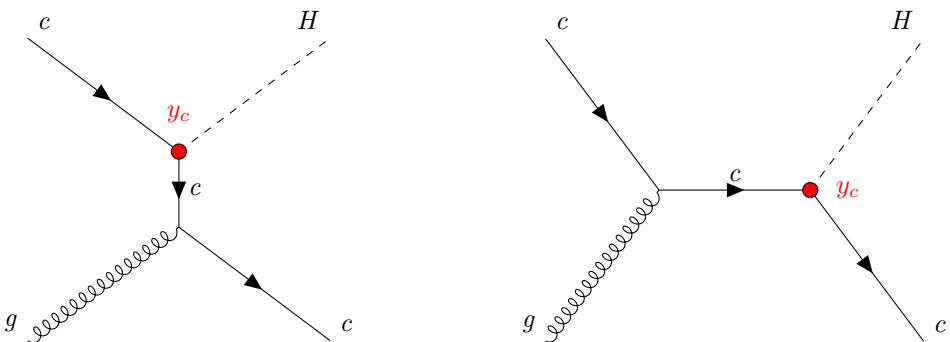


Figure 1.3: The leading order  $cH$  processes through which  $y_c$  may be probed as each diagram contains a vertex with a charm-quark and Higgs boson, here denoted in red. The corresponding diagrams with an anti-charm quark  $\bar{c}$  are implied.

<sup>228</sup> **1.3.2 The  $\kappa$ -framework**

<sup>229</sup> The  $\kappa$ -framework [16] is a tool to parametrise modifications to couplings between the Higgs boson  
<sup>230</sup> and other particles with respect to the expected SM values of the couplings. For example, the  
<sup>231</sup> 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

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

232  
Here, the expected SM decay width  $BR_{cc}^{SM} = 0.3$  can be substituted. Additionally, the fact that  
233 observed Higgs boson rates have been well measured to be close to their expected values (see e.g.  
234 [18]) can be reflected by setting  $\mu \approx 1$ , so that only a dependence on  $\kappa_c$  remains in the expression.  
235 Thus by replacing  $\kappa_H$  in Equation 1.25 with Equation 1.27, a final expression relating a measured  
236 signal strength of the cH process to  $\kappa_c$  is obtained, given by  
237

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

Rearranging for  $\kappa_c$  gives

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

238  
Effectively, this approach in interpreting  $\kappa_c$  from a signal strength measurement  $\mu_{\sigma_{cH} BR(H \rightarrow ZZ)}$   
239 thus ensures compatibility with existing Higgs boson rate measurements, given a non-unity value  
240 of  $\kappa_c$  leads to modifications of the Higgs boson partial decay widths. It should be noted that  
241 this already indirectly implies bounds on  $\kappa_c$ , as discussed in [7].  
242

## 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

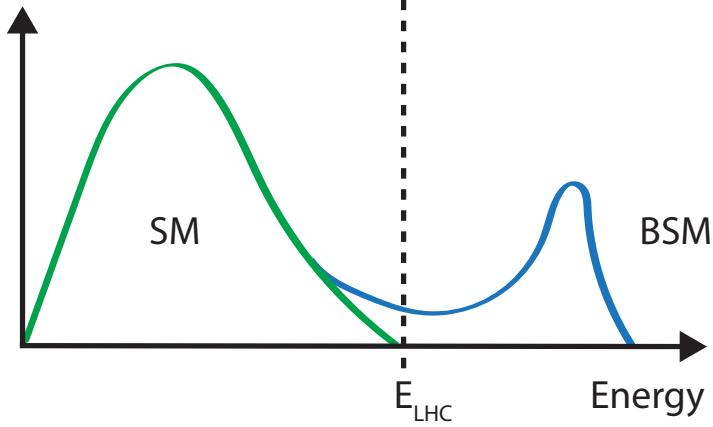


Figure 1.4: Illustration of how the presence of BSM physics, which is primarily visible beyond the reach of current collider energies (e.g.  $E_{\text{LHC}}$ ), can lead to subtle modifications of SM observables. These effects can be parametrised by SMEFT.

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>244</sup>  
<sup>245</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>249</sup>  
<sup>250</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>251</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

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

258

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

260

261 **1.4.1 The chromomagnetic dipole operator**262 A particular operator relevant to this work is referred to as the chromomagnetic dipole (CMD)  
263 operator  $\hat{O}_{qG}$ . For the charm quark, the CMD operator is written as

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

264

265 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  
266 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  
267 field strength tensor of the strong interaction. This operator may be uniquely bounded with the  
268 cH process due to the unique chiral structure of the operator, which mixes left and right-handed  
269 spinors, a structure otherwise only found in the Yukawa and quark-Higgs boson interaction terms  
270 of the SM.

271

272 To better understand this, it is worth considering other processes such as inclusive Higgs boson  
273 production, which have been successfully leveraged to set strong constraints on the top quark  
274 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  
275 SM contribution. Though the pure SMEFT contribution itself may be small and experimentally  
276 negligible due to limited analysis sensitivity, the much larger contribution of the SM process it  
277 interferes with can result in a non-negligeable interference effect with respect to the SM process.  
278 However, the chiral structure of the CMD operator influences the effectiveness of this strategy.  
279 Since the  $\hat{O}_{qG}$  operator effectively flips the chirality of the ingoing and outgoing quarks, a second  
280 *chirality flip* must be inserted for the SMEFT contribution to interfere with the SM process. This  
281 is visualised in Figure 1.5. Such a chirality flip is proportional to the mass  $m_q$  of the respective  
282 quark. As a result the interference contribution for a much lighter quark is significantly suppressed  
283 in comparison to the top quark, as also argued for the bottom quark in [23]. Effectively, the  
284 processes that prove effective in targeting  $\hat{O}_{tG}$  due to the large mass of the top quark are thus  
285 much less sensitive to  $\hat{O}_{cG}$ . However, since the cH process itself contains the chirality flipping  
286 quark-Higgs boson vertex, interference terms between the EFT and SM contributions do not  
287 suffer from the above described effect. Furthermore, due to the very low expected cross section  
288 of the cH process, quadratic contributions from  $\hat{O}_{cG}$  may be comparatively large even at small  
289 values of  $C_{cG}$ . Accordingly, the cH process may be an excellent target in constraining  $\hat{O}_{cG}$ .290 **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

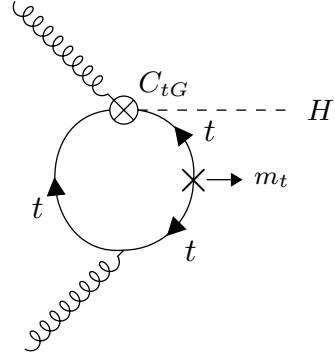


Figure 1.5: A modification of the gluon fusion process with a top quark loop by including the vertex introduced by the top quark CMD operator. Note that the arrows indicate chirality and not momentum flow. A chirality flip, denoted by the cross, proportional to the top quark mass  $m_t$  is required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

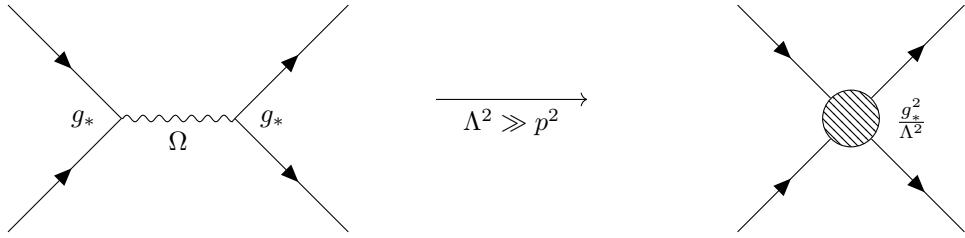


Figure 1.6: Feynman diagrams depicting a resonant process in which the new mediator particle  $\Omega$  is created (left) and the approximate description of this in an EFT, where the diagram is reduced to a four-point interaction.

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.



308 **Chapter 2**

309 **The CMS experiment at the LHC**

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

328 **2.1 The CMS detector**

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

336 **2.1.1 The CMS coordinate system**

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

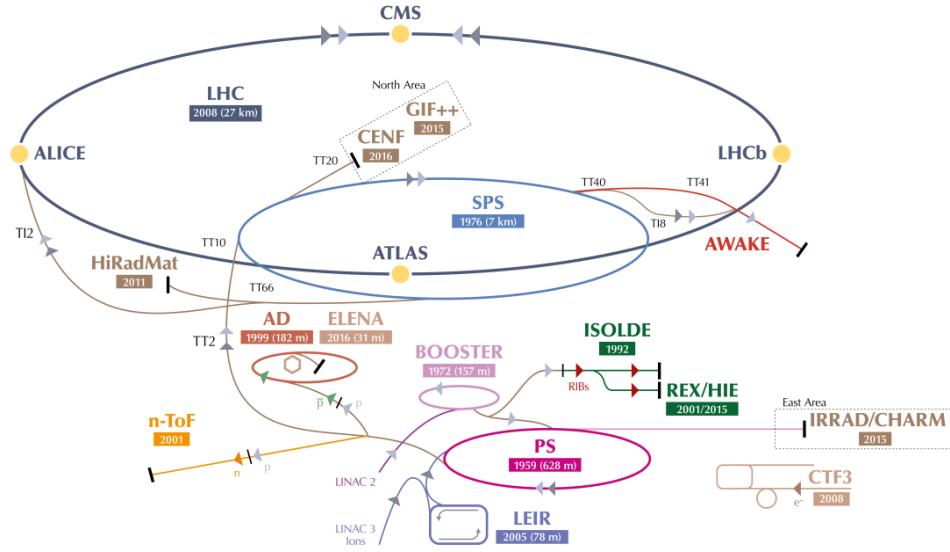


Figure 2.1: An overview of the LHC accelerator complex [27]. Before entering the large LHC ring, particles must pass through a number of increasingly powerful set of accelerators.

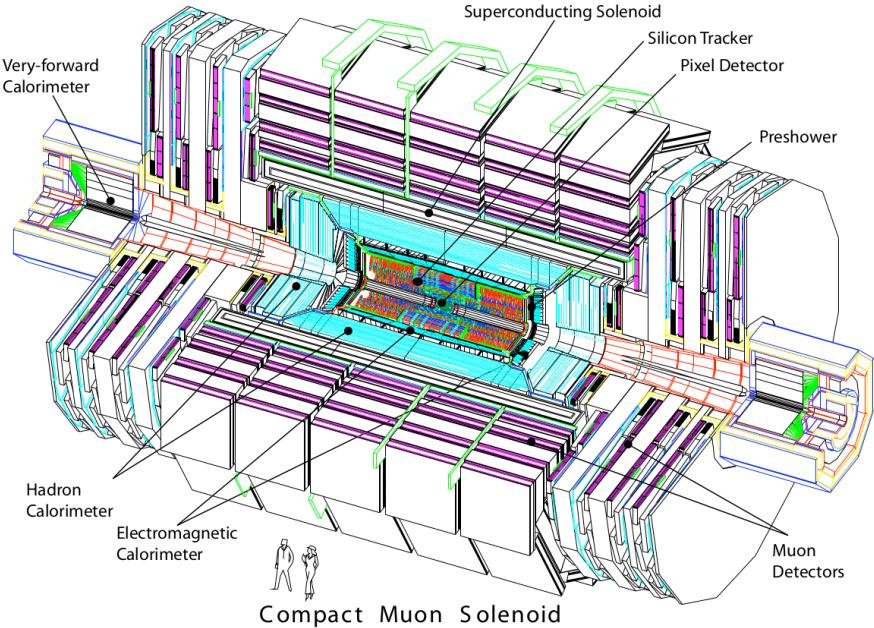


Figure 2.2: An overview of the CMS detector [28].

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

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

342  
 343 Small momenta in the  $z$ -direction  $p_z$  give a rapidity of zero, while the rapidity tends to  $\pm\infty$  for  
 344 large  $p_z$ . However, this requires knowledge of  $E$  and  $p_z$ , which can be difficult to measure. By  
 345 assuming the particle is ultra-relativistic, as is typically the case at the LHC, it is possible to  
 346 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

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

347  
 348 The CMS detector may be broadly split into two distinct regions inward and outward of the  
 349 boundary  $|\eta| = 1.479$ . The inner region or *barrel* consists of concentric layers around the beam  
 350 pipe. The outer *endcap* region consists of two caps that close off the detector at either end.  
 351 In this way, the CMS detector is designed for the best possible hermetic coverage around the  
 352 collision point.

### 353 2.1.2 The silicon tracker

354 The silicon tracker [30] is the innermost system of the CMS detector, situated closest to the  
 355 beampipe. It is designed to track the trajectories of charged particles as they emerge from the  
 356 collision point while producing minimal energy losses of the particles themselves. This subde-  
 357 tector is split into two main components, the pixel detector and silicon strip detector. A sketch  
 358 of these components may be seen in Figure 2.4.

359  
 360 The pixel detector is situated right around the beampipe and as of 2017 consists of four cir-  
 361 cular layers of individual silicon pixels in the barrel region and three disk layers in the endcap  
 362 region. These consist of rectangular silicon chips with a size of  $100 \times 150 \mu\text{m}^2$ . When a charged  
 363 particle traverses through the active material of these chips, an electrical signal is induced that  
 364 is recorded. This is typically referred to as a *hit*. The small pixel size allows for position mea-  
 365 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  $z$   
 366 direction [31]. An important feature of the pixel detector is its high radiation tolerance due to the  
 367 close proximity of these modules to the beam pipe.

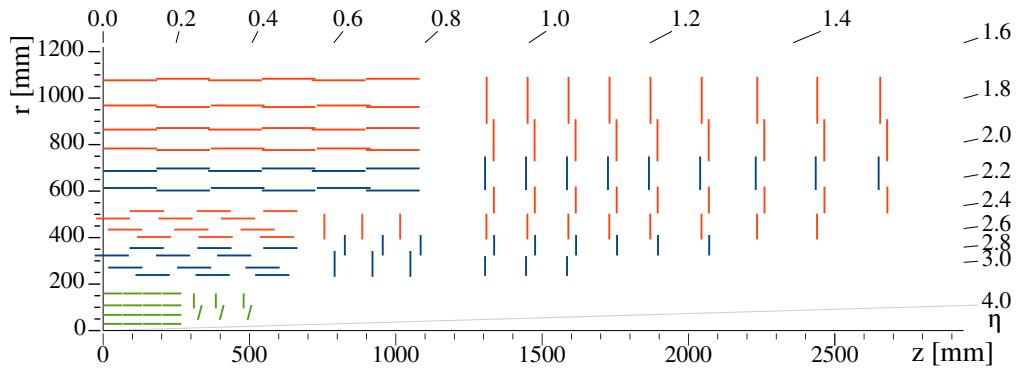


Figure 2.3: An overview of the CMS silicon tracker [30], shown in the  $r$ - $z$  plane after its upgrade during Run-2. The pixel detector is denoted in green while the silicon strip detector is denoted in blue and orange.

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

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

381  
382

### 2.1.3 The electromagnetic calorimeter

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

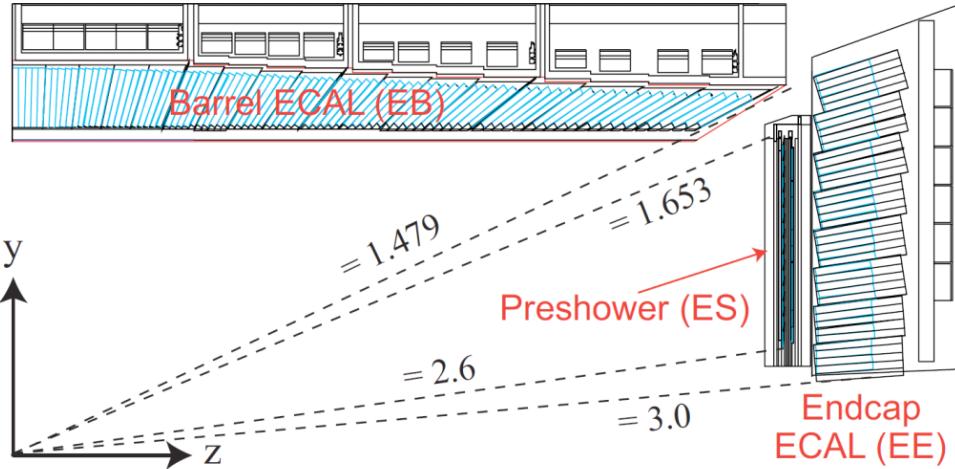


Figure 2.4: An overview of the CMS ECAL [34], shown in the  $r(y)$ - $z$  plane. The dashed lines denote the coverage of the barrel and endcap ECAL region as well as the preshower detector.

#### **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

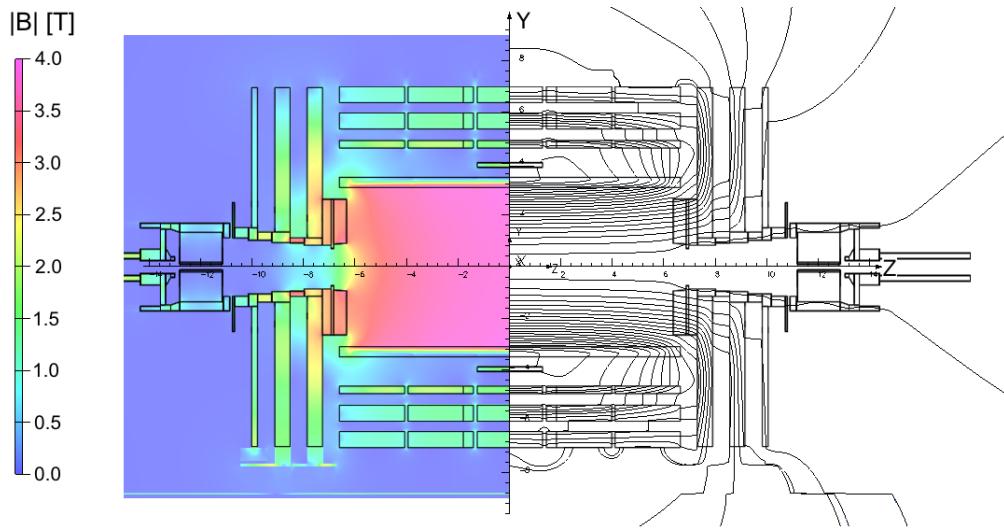


Figure 2.5: An overview of the magnetic flux (left) and magnetic field lines(right) inside the CMS detector, shown in the r-z plane [37].

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

#### 426 2.1.6 The muon chambers

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

434 Like the other subdetectors, the muon chambers are separated in a barrel ( $|\eta| < 1.2$ ) and endcap  
 435 ( $1.2 < |\eta| < 2.4$ ) region, which are composed of drift tubes and cathode strip chambers respec-  
 436 tively. The drift tubes each consists of a gas volume containing a mixture of Argon and CO<sub>2</sub>  
 437 in which a positively charged wire is stretched through the center. When charged particles  
 438 such as muons traverse these tubes, the gas is ionised. Due to the positive charge of the wire,  
 439 the resulting electrons drift towards the wire producing an electrical signal. Thus the presence  
 440 of muons may be determined by activation of the drift tubes. The cathode strip chambers on  
 441 the other hand consist of layers of positively charged (anode) wires, which are arranged in a  
 442 perpendicular fashion to a set of negatively charged (cathode) strips. Combining signals from  
 443 both the wires and strips allows for a position measurement in both the R and  $\phi$  direction. Both  
 444 types of detector are supplemented by resistive plate chambers, which act as a trigger providing  
 445 a precise timing resolution of  $\sim 1\text{ns}$ . This makes it possible to unambiguously assign muons to  
 446 individual collisions. These consist of parallel, oppositely charged plastic plates that are coated  
 447 with a conductive graphite layer and are contained in a gas volume. Ionisation of the gas due to  
 448 the traversal of a charged particle thus leads to an electrical signal. An overview of the spatial  
 449

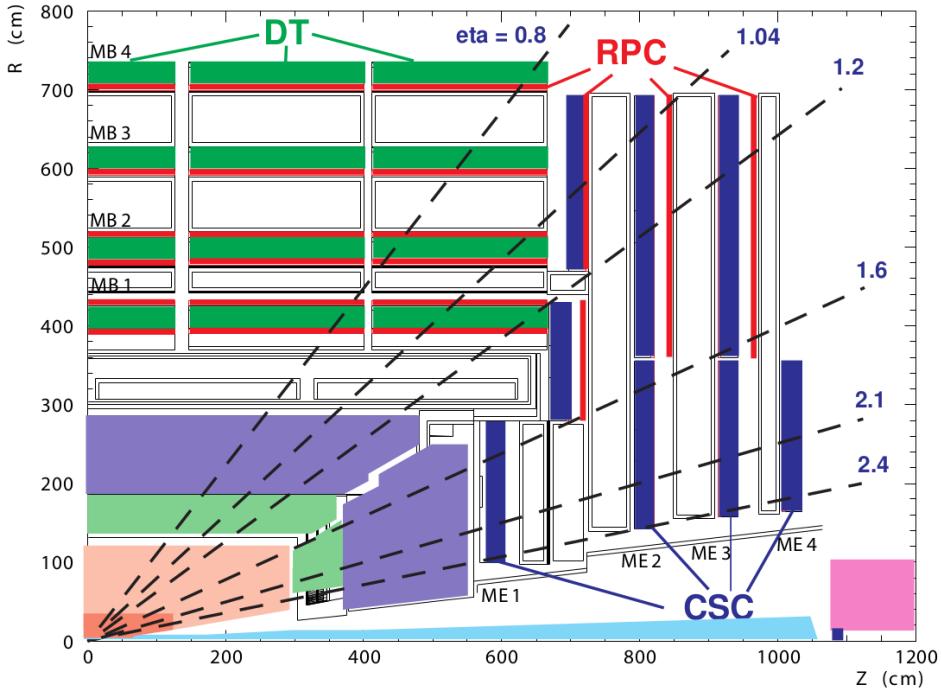


Figure 2.6: An overview of the CMS muon system, shown in the  $r$ - $z$  plane [28]. Shown are the drift tube (DT), the cathode strip chambers (CSC) and resistive plate chambers (RPC).

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

### 452 2.1.7 The triggering system

453 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  
454 close to 40 TB/s. Not only is the storage of such a quantity of data unfeasible but a significant  
455 portion consists of low-energy scattering events which are not of interest to analysts. As such,  
456 the triggering system is implemented to extract a subset of events that are of physics interest.  
457

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

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

## 471 2.2 Event reconstruction with the CMS detector

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

### 475 2.2.1 Track and vertex reconstruction

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

488

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

### 498 2.2.2 The Particle Flow algorithm

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

505  
506

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.

507 A visual overview of these signatures and the particle type they correspond to can be found in  
 508 Figure 2.7. The PF algorithm leverages exactly these properties. Initially, matched tracks in  
 509 the tracker and muon systems are identified as muons and the corresponding components are  
 510 removed from the event. Subsequently, matched tracks and energy deposit clusters in the ECAL  
 511 are identified as electrons and the corresponding components are removed. An isolated cluster in  
 512 the ECAL with no associated track is reconstructed as a photon candidate and the corrsponding  
 513 cluster is removed. This is expected to leave only charged and neutral hadrons. Clusters of  
 514 energy deposits in the HCAL associated with a track are thus identified as charged hadrons.  
 515 However, it frequently occurs that photons are produced in the decay of neutral hadrons. Thus,  
 516 if the energy estimated from a track is considerably less than the associated cluster in the HCAL  
 517 and there is a corresponding energy deposit in the ECAL, an additional photon candidate is  
 518 reconstructed that is associated with the hadron. Finally, HCAL clusters with no associated  
 519 track are reconstructed as neutral hadrons.

520  
 521 This of course is a greatly simplified description, a more comprehensive version of which can  
 522 be found in [43]. The following section describe in greater detail the reconstruction of objects  
 523 relevant to this work. This includes muons, *jets*, which are collimated particle showers that  
 524 typically consist of a collection of reconstructed objects and missing transverse energy.

### 525 2.2.3 Reconstruction and identification of muons

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

- 530 • **Standalone muon tracks:** A standalone muon track refers to a fit of individual hits  
 531 present in the muon detector.
- 532 • **Tracker muon tracks:** Tracker muon tracks are reconstructed by extrapolating tracks  
 533 from the tracker to the muon detector, referred to as an *inside-out* approach. If a hit in  
 534 the muon detector can be matched to the extrapolated track, then these matched tracks

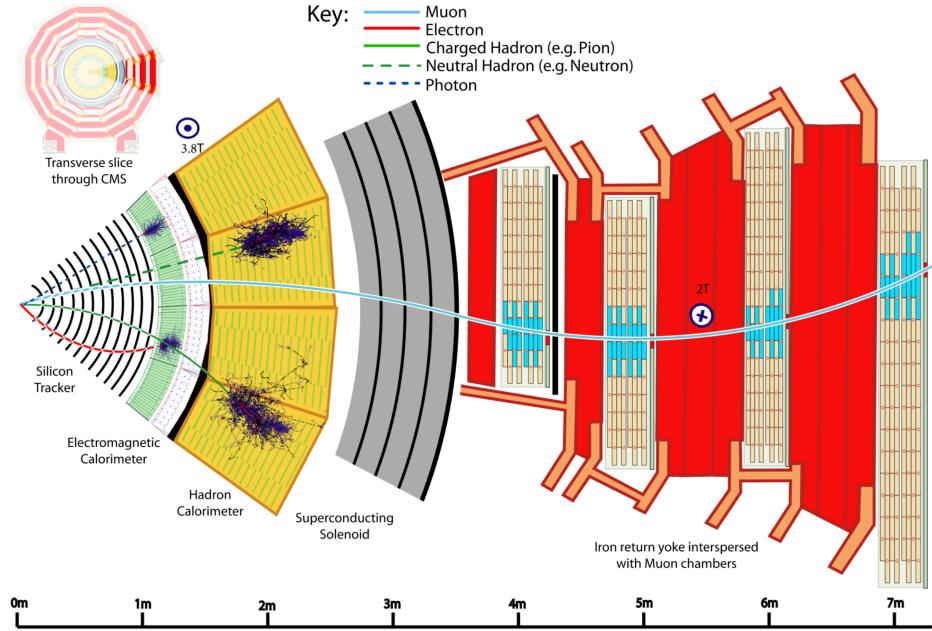


Figure 2.7: A transverse slice of the CMS detector, visualising the signatures that different particles produce in the different detector subsystems. [43].

535        are identified as a tracker muon track. This reduces the impact from atmospheric muons  
536        traversing the detector, which may be falsely interpreted as standalone muon tracks.

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

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

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

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

551        Here,  $\sum p_{\text{T}}^{\text{charged}}$  represents the scalar sum of the transverse momenta of charged hadron origi-  
552        nating from the primary vertex of the event. The quantities  $\sum p_{\text{T}}^{\text{neutral}}$  and  $\sum p_{\text{T}}^{\gamma}$  represent the

553 respective transverse momenta sums for neutral hadrons and photons. These sums are calculated  
 554 by accounting from contributions within a conical volume around the muon direction. The size  
 555 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  
 556 this case the cone boundary around the muon direction is set at  $\Delta R = 0.4$ . The contribution to  
 557 the relative isolation from pile-up is estimated by subtracting  $p_T^{\mu, PU} = 0.5 \sum_k p_T^{k, \text{charged}}$  in Equa-  
 558 tion 2.4, where the sum over  $k$  represents charged hadron contributions not originating from the  
 559 PV. The factor 0.5 corrects for different fractions of charged and neutral particles in the cone  
 560 [44]. Lastly,  $p_T^\mu$  represents the transverse momentum of the muon. The relative isolation is thus  
 561 a variable that quantifies the presence of energy deposits in the ECAL and HCAL around the  
 562 trajectory of the muon, relatively to the  $p_T$  of the muon. Since muons are expected to produce  
 563 such deposits only minimally, good muon candidates are expected to be associated with small  
 564 values of  $\mathcal{I}_{\text{rel}}^\mu$ .

565  
 566 Two sets of muon identification criteria are defined for this work:

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

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

#### 580 2.2.4 Reconstruction and identification of jets

581 The quarks and gluons that are produced in proton-proton collisions rapidly hadronise, typically  
 582 producing collimated cones of particles referred to as *jets*. Details on the concept of hadronisa-  
 583 tion, which results from the nature of the strong interaction, can be found in [45]. Since the c  
 584 quark of the cH process too will produce a jet, jet objects also represent an important aspect of  
 585 the analysis presented in this work.

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

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

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

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

596  
 597 Here,  $y$  is the rapidity of a particle and  $R$  is a constant parameter that determines the cone size  
 598 of the clustered jets. The default choice used in CMS is  $R=0.4$ , which is also used in this work.  
 599 Starting with the highest  $p_T$  object in the initial iteration, the distance  $d_{ij}$  with the closest PF  
 600 candidate  $j$  is calculated. The two objects are clustered together and this process is repeated un-  
 601 til a stopping condition  $d_{ih} > d_{iB}$  is met. At this point, the jet is considered fully reconstructed  
 602 and the PF candidates used in its clustering are removed for the reconstruction of subsequent jets.

603  
 604 Due to the presence of detector noise, unphysical low  $p_T$  jets can be erroneously reconstructed.  
 605 This effect can be mitigated by applying additional criteria on reconstructed jets. This includes  
 606 requiring that at least two PF candidates are clustered in the jet and that the jet's energy is  
 607 not solely attributed to neutral hadrons or photons. These requirements remove almost all such  
 608 unphysical jets while over 99% of physical jets fulfill them [47]. Additionally, a pile-up discrim-  
 609 ination algorithm is described in [47], of which the loose working point is applied to jets with  
 610  $p_T < 50$  GeV in this work.

611  
 612 A calibration of jet energies is performed after reconstruction [48] in both simulation and data.  
 613 This calibration accounts for pile-up contributions in the clustering, the non-linearity of the de-  
 614 tector response and improper reconstruction of hadrons. A number of methods are used to derive  
 615 sets of correction factors. An example is the use of events with a Z boson, the  $p_T$  of which may  
 616 be precisely reconstructed via the  $Z \rightarrow \mu\mu$  decay, that a single jet recoils against. Additionally,  
 617 significant discrepancies in the resolution of jets in simulation and data are observed, with the  
 618 resolution being worse in the latter than the former. This is accounted for by a smearing method,  
 619 in which the resolution of jets is artificially smeared in simulation so that a better comparison to  
 620 data is achieved.

### 621 2.2.5 Missing transverse momentum

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

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

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

## 628 2.3 Identification of charm quark-induced jets

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

- 642 • ***c* jets:** If at least one charm (*c*) ghost hadron and no bottom (*b*) hadrons are clustered  
   643 inside the jet, the jet is labelled as a *c* jet.
- 644 • ***b* jets:** If at least one *b* ghost hadron is clustered inside the jet, the jet is labelled as a *b*  
   645 jet.
- 646 • **light jets:** If no ghost hadrons are clustered inside the jet, the jet is labelled as a light  
   647 jet. Light jets may be initiated by quarks such as the up, down, or strange quark or by  
   648 gluons. An additional, technical category of *pile-up jets* exists depending on whether so-  
   649 called matching criteria between reconstructed and simulated jets are fulfilled, though they  
   650 are subsumed into the light jets category for the purpose of this work.

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

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

### 654 2.3.1 Properties of heavy-flavour jets

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

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

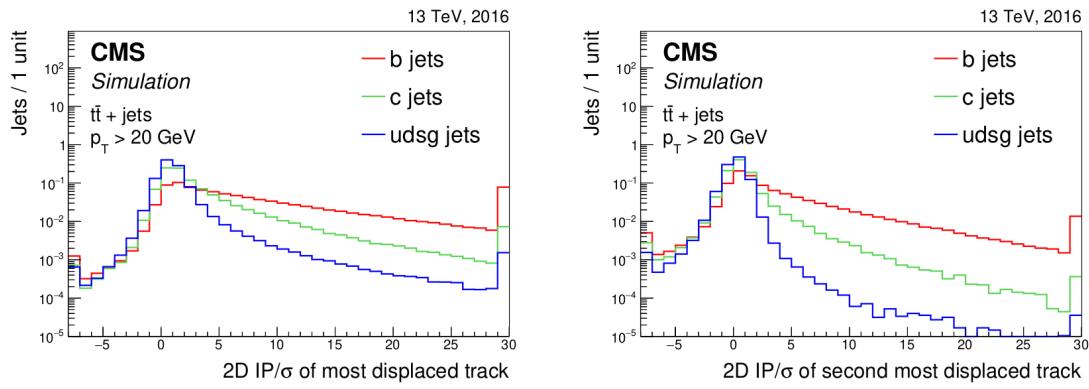


Figure 2.8: Plots showing the significance of the 2D impact parameter of the most and second most displaced tracks in a jet [50]. As can be seen, these variables can differentiate  $b$  and  $c$  jets from light jets to a significant degree.

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

### 675 2.3.2 The DeepJet algorithm

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

- 687 •  **$b/bb/lepb$  ( $b$  jets):** These three classes represent subclasses of jets originating from a  
 688  $b$  hadron. The  $b$  class represents a jet originating from a single  $b$  hadron, the  $bb$  class  
 689 originating from two  $b$  hadrons and  $lepb$  representing a jet originating from a  $b$  hadron  
 690 with the presence of a soft lepton.
- 691 •  **$c$  ( $c$  jets):** This class represents a jet originating from a  $c$  hadron.
- 692 •  **$l, g$  (light jets):** These two classes represent light jets originating from a light quark or  
 693 gluon respectively.

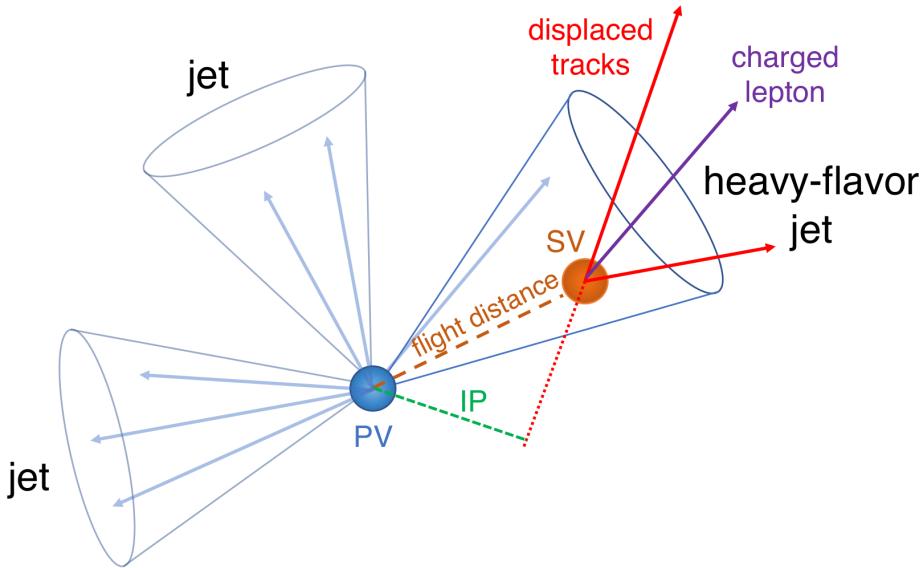


Figure 2.9: An illustration highlighting the properties of HF jets [50]. The presence of a secondary vertex (SV), characterised by the impact parameter (IP) in green, as well as the presence of a lepton is highlighted.

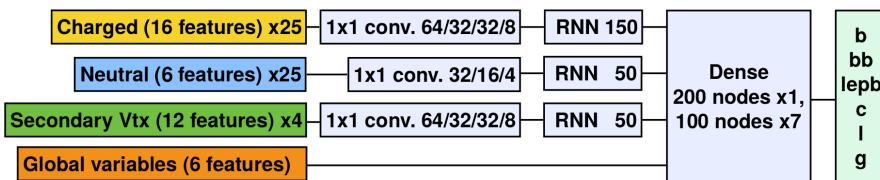


Figure 2.10: An illustration depicting the architecture of the DeepJet neural network [51]. Three individual branches separately process the charged hadrons, neutral hadrons and secondary vertex information before being passed onto a combining, fully connected layer together with global variables.

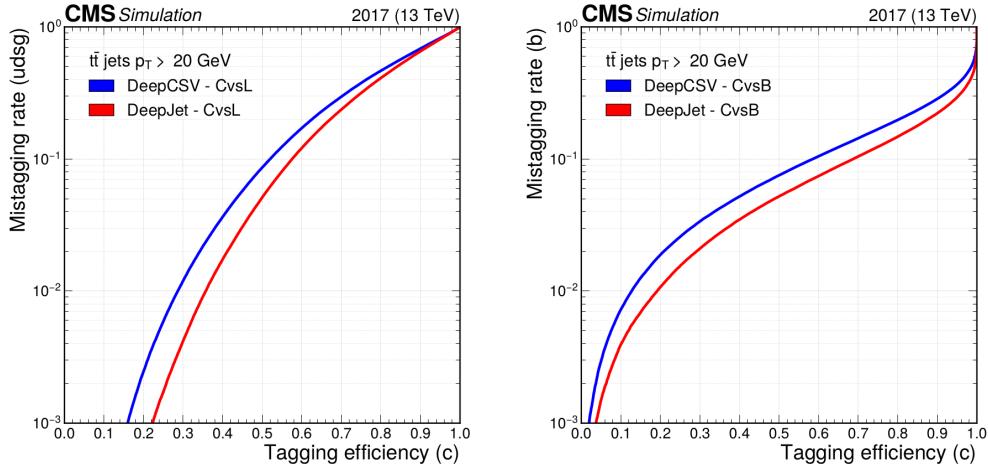


Figure 2.11: Performance of DeepJet algorithm in identifying  $c$  jets against  $b$  jets and light jets in simulated samples of top quark pair production, in which both top quarks decay hadronically [52]. The x-axis represents the efficiency with which  $c$  jets are identified, while the y-axis represents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

708 **Chapter 3**

709 **Search for the cH( $ZZ \rightarrow 4\mu$ )  
710 process**

711 To probe the charm Yukawa coupling through the cH process, a methodology must be devised  
712 to select and reconstruct cH candidate events. This is described in ??, specifically targetting  
713 cH( $ZZ \rightarrow 4\mu$ ) final states. Additionally, a model describing the expected contributions from  
714 the cH( $ZZ \rightarrow 4\mu$ ) process as well as a number of background processes in the event selection  
715 must be constructed, as described in ???. Finally, a statistical evaluation using flavour-tagging  
716 discriminators to set 95% CL upper limits on  $\kappa_c$ , assuming the absence of signal, is presented in  
717 ??.

718 **3.1 cH event selection**

719 To reconstruct a cH( $ZZ \rightarrow 4\mu$ ) candidate event, a Higgs boson candidate needs to be reconstructed  
720 and a corresponding jet candidate needs to be identified. These two procedures are described in  
721 this section and distributions of cH( $ZZ \rightarrow 4\mu$ ) candidate events are shown using a simulation of  
722 the cH( $ZZ \rightarrow 4\mu$ ) process as described in ??.

723  
724 To reconstruct a Higgs (jet) candidate, an initial selection of muon (jet) objects must be made.  
725 These are summarised in Table 3.1 along with the HLT trigger path requirement used in this anal-  
726 ysis. The objective of this selection is to identify events with well-reconstructed, isolated muons  
727 as well as a least one well-reconstructed jet. Following this initial selection, the corresponding  
728 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
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 ?? 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, 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. A selection efficiency of xxx% is achieved on the simulated cH(ZZ $\rightarrow$ 4 $\mu$ ) sample. The majority of loss in acceptance can be attributed to ....

### 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:

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

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

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

Table 3.2: Muon and jet object selection requirements.

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

783 With this, the individual components of the  $cH(ZZ \rightarrow 4\mu)$  are thus reconstructed and selected.  
784 Something about c jet vs non-c jet in yc sensitive events..

785 **3.2 Signal and background estimation**

786 **3.2.1 Estimation of  $cH$  process**

787 **3.2.2 Estimation of irreducible backgrounds**

788 **3.2.3 Estimation of reducible backgrounds**

789 **3.3 Statistical evaluation**

790 Chapter 4

791 An EFT interpretation of the  
792  $cH(ZZ \rightarrow 4\mu)$  process



# Conclusion

793



# Bibliography

- [1] S. Navas et al. “Review of Particle Physics”. In: *Phys. Rev. D* 110 (3 Aug. 2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001. URL: <https://link.aps.org/doi/10.1103/PhysRevD.110.030001>.
- [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>.
- [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>.
- [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>.
- [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>.
- [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>.
- [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>.
- [8] Michael Edward Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley (1995) 842 p. Westview Press, 1995.
- [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.
- [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>.
- [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>.

- 829 [12] CMS Collaboration. *Simultaneous probe of the charm and bottom quark Yukawa couplings*  
 830 *using ttH events*. 2025. DOI: 10.48550/arXiv.2509.22535. arXiv: 2509.22535 [hep-ex].  
 831 URL: <https://arxiv.org/abs/2509.22535>.
- 832 [13] CMS Collaboration. *Search for a Higgs boson produced in association with a charm quark*  
 833 *and decaying to a W boson pair in proton-proton collisions at  $\sqrt{s} = 13$  TeV*. 2025. arXiv:  
 834 2508.14988 [hep-ex]. URL: <https://arxiv.org/abs/2508.14988>.
- 835 [14] CMS Collaboration. *Search for the associated production of a Higgs boson with a charm*  
 836 *quark in the diphoton decay channel in pp collisions at  $\sqrt{s} = 13$  TeV*. 2025. arXiv: 2503.  
 837 08797 [hep-ex]. URL: <https://arxiv.org/abs/2503.08797>.
- 838 [15] Nuoyu Dong et al. “Probing charm Yukawa coupling through ch associated production  
 839 at the hadron colliders”. In: *Phys. Rev. D* 111 (5 Mar. 2025), p. 053003. DOI: 10.1103/  
 840 PhysRevD.111.053003. URL: <https://link.aps.org/doi/10.1103/PhysRevD.111.053003>.
- 842 [16] C T Potter et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of*  
 843 *the LHC Higgs Cross Section Working Group*. en. 2013. DOI: 10.5170/CERN-2013-004.  
 844 URL: <http://cds.cern.ch/record/1559921>.
- 845 [17] J. A. Aguilar-Saavedra, J. M. Cano, and J. M. No. “More light on Higgs flavor at the  
 846 LHC: Higgs boson couplings to light quarks through  $h + \gamma$  production”. In: *Phys. Rev.*  
 847 *D* 103 (9 May 2021), p. 095023. DOI: 10.1103/PhysRevD.103.095023. URL: <https://link.aps.org/doi/10.1103/PhysRevD.103.095023>.
- 849 [18] The CMS Collaboration. “Combined measurements of Higgs boson couplings in proton-  
 850 proton collisions at  $\sqrt{s} = 13$  TeV”. In: *The European Physical Journal C* 79.5 (May 2019).  
 851 ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6909-y. URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6909-y>.
- 853 [19] Steven Weinberg. “Baryon and Lepton Nonconserving Processes”. In: *Phys. Rev. Lett.* 43  
 854 (1979), pp. 1566–1570. DOI: 10.1103/PhysRevLett.43.1566.
- 855 [20] B. Grzadkowski et al. “Dimension-six terms in the Standard Model Lagrangian”. In:  
 856 *Journal of High Energy Physics* 2010.10 (Oct. 2010). ISSN: 1029-8479. DOI: 10.1007/jhep10(2010)085. URL: [http://dx.doi.org/10.1007/JHEP10\(2010\)085](http://dx.doi.org/10.1007/JHEP10(2010)085).
- 858 [21] Gino Isidori, Felix Wilsch, and Daniel Wyler. “The standard model effective field theory at  
 859 work”. In: *Reviews of Modern Physics* 96.1 (Mar. 2024). ISSN: 1539-0756. DOI: 10.1103/  
 860 revmodphys.96.015006. URL: <http://dx.doi.org/10.1103/RevModPhys.96.015006>.
- 861 [22] CMS Collaboration. *Combined effective field theory interpretation of Higgs boson, elec-*  
 862 *troweak vector boson, top quark, and multi-jet measurements*. 2025. arXiv: 2504.02958  
 863 [hep-ex]. URL: <https://arxiv.org/abs/2504.02958>.
- 864 [23] Joseph Bramante et al. “Boosted Higgs bosons from chromomagnetic b’s:  $b\bar{b}h$  at high  
 865 luminosity”. In: *Phys. Rev. D* 93 (5 Mar. 2016), p. 053001. DOI: 10.1103/PhysRevD.93.  
 866 053001. URL: <https://link.aps.org/doi/10.1103/PhysRevD.93.053001>.
- 867 [24] J. Elias-Miró et al. “Higgs windows to new physics through d=6 operators: constraints and  
 868 one-loop anomalous dimensions”. In: *Journal of High Energy Physics* 2013.11 (Nov. 2013).  
 869 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).
- 871 [25] The CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of In-*  
 872 *strumentation* 3.08 (Aug. 2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004. URL:  
 873 <https://doi.org/10.1088/1748-0221/3/08/S08004>.

- 874 [26] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva:  
875 CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: <http://cds.cern.ch/record/782076>.
- 877 [27] E. A. Mobs. “The CERN accelerator complex. Complexe des accélérateurs du CERN”. In:  
878 (Oct. 2016). General Photo. URL: <http://cds.cern.ch/record/2225847>.
- 879 [28] The CMS Collaboration. *CMS Physics : Technical Design Report Volume 1: Detector Per-*  
880 *formance and Software*. Technical design report. CMS. Geneva: CERN, 2006. URL: <https://cds.cern.ch/record/922757>.
- 882 [29] CMS Collaboration. “Development of the CMS detector for the CERN LHC Run 3”. In:  
883 *Journal of Instrumentation* 19.05 (May 2024), P05064. ISSN: 1748-0221. DOI: 10.1088/  
884 1748-0221/19/05/p05064. URL: <http://dx.doi.org/10.1088/1748-0221/19/05/P05064>.
- 886 [30] *CMS Tracker Detector Performance Public Results*. Last accessed 24.11.25. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK>.
- 888 [31] The CMS Collaboration. “Description and performance of track and primary-vertex re-  
889 construction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014),  
890 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>.
- 892 [32] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Re-*  
893 *port*. Tech. rep. Geneva, 1997. URL: <https://cds.cern.ch/record/349375>.
- 894 [33] CMS Collaboration. *The CMS ECAL performance with examples*. Tech. rep. Geneva:  
895 CERN, 2014. DOI: 10.1088/1748-0221/9/02/C02008. URL: <https://cds.cern.ch/record/1632384>.
- 897 [34] CMS Collaboration. *The CMS ECAL performance with examples*. Tech. rep. Geneva:  
898 CERN, 2014. DOI: 10.1088/1748-0221/9/02/C02008. URL: <https://cds.cern.ch/record/1632384>.
- 900 [35] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Tech.  
901 rep. Geneva, 1997. URL: <https://cds.cern.ch/record/357153>.
- 902 [36] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Tech. rep. Geneva,  
903 1997. URL: <https://cds.cern.ch/record/331056>.
- 904 [37] CMS Collaboration. “Precise mapping of the magnetic field in the CMS barrel yoke using  
905 cosmic rays”. In: *Journal of Instrumentation* 5.03 (Mar. 2010), T03021–T03021. ISSN: 1748-  
906 0221. DOI: 10.1088/1748-0221/5/03/t03021. URL: <http://dx.doi.org/10.1088/1748-0221/5/03/T03021>.
- 908 [38] CMS Collaboration. *The CMS muon project: Technical Design Report*. Tech. rep. Geneva,  
909 1997. URL: <https://cds.cern.ch/record/343814>.
- 910 [39] The CMS collaboration. “Performance of CMS muon reconstruction in pp collision events at  
911  $\sqrt{s} = 7\text{TeV}$ ”. In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10002–P10002. ISSN: 1748-  
912 0221. DOI: 10.1088/1748-0221/7/10/p10002. URL: <http://dx.doi.org/10.1088/1748-0221/7/10/P10002>.
- 914 [40] Mia Tosi. *The CMS trigger in Run 2*. Tech. rep. Geneva: CERN, Oct. 2017. DOI: 10.22323/  
915 1.314.0523. URL: <https://cds.cern.ch/record/2290106>.
- 916 [41] CMS collaboration et al. “Description and performance of track and primary-vertex recon-  
917 struction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (2014), P10009.

- 918 [42] K. Rose. “Deterministic annealing for clustering, compression, classification, regression,  
919 and related optimization problems”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–  
920 2239. DOI: 10.1109/5.726788.
- 921 [43] A. M. Sirunyan et al. “Particle-flow reconstruction and global event description with the  
922 CMS detector.” In: *JINST* 12 (June 2017). Replaced with the published version. Added  
923 the journal reference and DOI. All the figures and tables can be found at [http://cms-  
925 results.web.cern.ch/cms-results/public-results/publications/PRF-14-001](http://cms-<br/>924 results.web.cern.ch/cms-results/public-results/publications/PRF-14-001)(CMS Public Pages),  
P10003. 82 p. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965. URL: <https://cds.cern.ch/record/2270046>.
- 927 [44] CMS Collaboration. “Pileup mitigation at CMS in 13 TeV data”. In: *Journal of Instrumentation*  
928 15.09 (Sept. 2020), P09018–P09018. ISSN: 1748-0221. DOI: 10.1088/1748-  
929 0221/15/09/p09018. URL: <http://dx.doi.org/10.1088/1748-0221/15/09/P09018>.
- 930 [45] B. Andersson et al. “Parton fragmentation and string dynamics”. In: *Physics Reports*  
931 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: [https://doi.org/10.1016/0370-  
932 1573\(83\)90080-7](https://doi.org/10.1016/0370-1573(83)90080-7). URL: <https://www.sciencedirect.com/science/article/pii/0370157383900807>.
- 934 [46] M. Cacciari et al. “FastJet user manual”. In: *The European Physical Journal C* 72.3 (Mar.  
935 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>.
- 937 [47] CMS Collaboration. *Jet algorithms performance in 13 TeV data*. Tech. rep. Geneva: CERN,  
938 2017. URL: <https://cds.cern.ch/record/2256875>.
- 939 [48] CMS Collaboration. “Jet energy scale and resolution in the CMS experiment in pp collisions  
940 at 8 TeV”. In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. ISSN: 1748-  
941 0221. DOI: 10.1088/1748-0221/12/02/p02014. URL: <http://dx.doi.org/10.1088/1748-0221/12/02/P02014>.
- 943 [49] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas”. In: *Physics  
944 Letters B* 659.1–2 (Jan. 2008), pp. 119–126. ISSN: 0370-2693. DOI: 10.1016/j.physletb.  
945 2007.09.077. URL: <http://dx.doi.org/10.1016/j.physletb.2007.09.077>.
- 946 [50] CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp  
947 collisions at 13 TeV”. In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011.  
948 ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/05/p05011. URL: <http://dx.doi.org/10.1088/1748-0221/13/05/P05011>.
- 950 [51] E. Bols et al. “Jet flavour classification using DeepJet”. In: *Journal of Instrumentation*  
951 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/  
952 p12012. URL: <http://dx.doi.org/10.1088/1748-0221/15/12/P12012>.
- 953 [52] CMS Collaboration. “A new calibration method for charm jet identification validated with  
954 proton-proton collision events at  $\sqrt{s} = 13$  TeV”. In: *Journal of Instrumentation* 17.03  
955 (Mar. 2022), P03014. ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/03/p03014. URL:  
956 <http://dx.doi.org/10.1088/1748-0221/17/03/P03014>.