VRIJE UNIVERSITEIT BRUSSEL

Graduation thesis submitted in partial fulfilment of blah

SOMETHING ABOUT HIGGS+CHARM

my thesis subtitle

Felix Heyen

March 2026

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

sciences and bioengineering sciences

Abstract

My abstract

Contents

Chapter 1

Introduction

The Standard Model (SM) of particle physics is the theory that best decribes our current understanding of fundamental particles and their interactions. It describes a broad range of phenomena and makes a plethora predictions, many of which have been confirmed via measurement to great degrees of accuracy [1]. A noteable feature of the SM is the Brout-Englert-Higgs (BEH) mechanism [2][3], which predicts the existance of a Brout-Englert-Higgs (or often simply Higgs) boson. The EBH 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.

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.

1.1 The Standard Model of particle physics

The SM is formulated through the formalism of Quantum Field Theory (QFT). This is a formalism that combines concepts of classical field theory, quantum machanics 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.

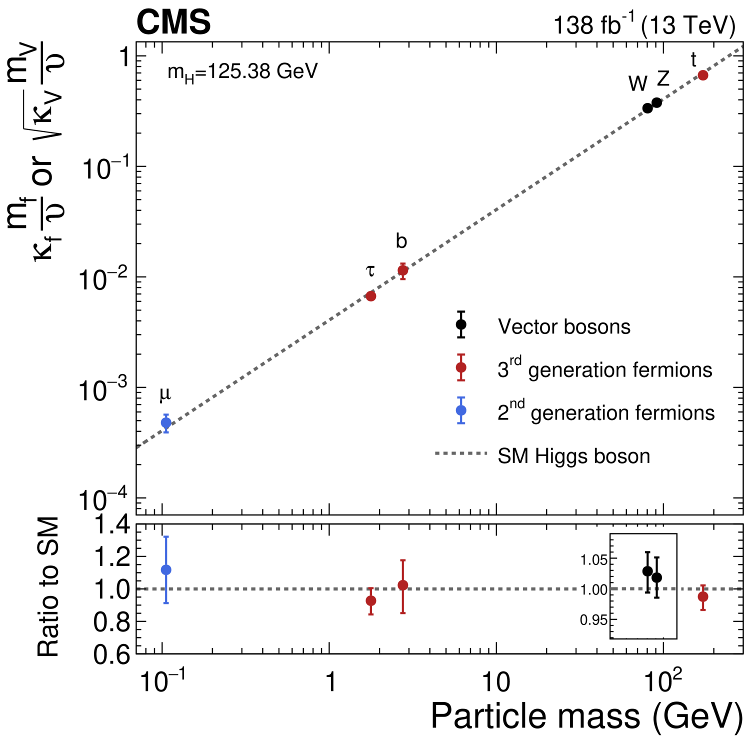

Figure 1.1: The measured coupling modifiers κ_f and κ_V of the coupling between the Higgs boson and fermions as well as heavy gauge bosons as functions of fermion or gauge boson mass m_f and m_V, where ν is the vacuum expectation value of the Higgs field. [6]

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

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

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

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

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

1.2 The Higgs-charm Yukawa coupling

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

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

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

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

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

With these concepts in mind the nature of the electro-weak sector's Lagrangian in the SM may be discussed. Naively, the form of this would be given by

$$\mathcal{L}_{\mathrm{EW}} = i\overline{\psi}_L \gamma^\mu D^L_\mu \psi_L + i\overline{\psi}_R \gamma^\mu D^R_\mu \psi_R - \frac{1}{2}\mathrm{Tr}\left(W^a_{\mu\nu} W^{a\mu\nu}\right) - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \,. \tag{1.2}$$

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

g' :	coupling constant of $\mathrm{U}(1)_Y$
g :	coupling constant of $\mathrm{SU}(2)_L$
ψ_L,	left-handed isospin doublet
ψ_R,	right-handed isospin doublet
B_μ :	gauge field of $\mathrm{U}(1)_Y$
W^a_μ :	gauge fields of $\mathrm{SU}(2)_L$, $a = 1, 2, 3$
$W_{\mu\nu}$:	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
$D^L_\mu = \partial_\mu + ig'\frac{Y_L}{2}B_\mu + igt^a W^a_\mu$	
$D^R_\mu = \partial_\mu + ig'\frac{Y_R}{2}B_\mu$	

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

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

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

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos\theta_\mathrm{W} & \sin\theta_\mathrm{W} \\ -\sin\theta_\mathrm{W} & \cos\theta_\mathrm{W} \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} . \tag{1.4}$$

with the weak mixing angle θ_W.

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

$$m^2_W W^+_\mu W^{-,\mu} + \frac{1}{2}m^2_Z Z_\mu Z^\mu. \tag{1.5}$$

do not remain invariant under arbitrary $\mathrm{SU(2)_L}$ gauge transformations. This is as gauge fields A_μ generically transform as

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

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

$$m_{\mathrm{f}}\overline{\psi}\psi. \tag{1.7}$$

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 ψ_L and ψ_R components of ψ under $\mathrm{SU(2)_L \, x \, U(1)}_Y$ gauge transformations.

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 $\mathcal{L}_{\mathrm{EW}}$ 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

$$\mathcal{L}_{\mathrm{Higgs}} = D_\mu \phi^\dagger D^\mu \phi - V(\phi) \tag{1.8}$$
$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2. \tag{1.9}$$

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

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.10}$$

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

$$D_\mu = \partial_\mu + ig' \frac{Y_\phi}{2} B_\mu + igt^a W_\mu^a \tag{1.11}$$

ensures $\mathcal{L}_{\mathrm{Higgs}}$ remains locally gauge invariant under $\mathrm{SU(2)_L \, x \, U(1)}_Y$ 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

$$v = \sqrt{\frac{\mu^2}{2\lambda}}. \tag{1.12}$$

The center of the potential is now an unstable local maximum and the only stable configuration can be found in the non-zero ground state. Through this, the symmetry of the potential is effectively broken. A popular choice of gauge for ϕ is

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

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

$$\left(\frac{g}{2}\right)^2 v^2 W_\mu^+ W^{\mu -} = m_{\mathrm{W}}^2 W_\mu^+ W^{\mu -} \tag{1.14}$$

$$\left(\frac{\sqrt{g^2 + g'}}{2}\right)^2 v^2 Z_\mu Z^\mu = m_{\mathrm{Z}}^2 Z_\mu Z^\mu \,. \tag{1.15}$$

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

1.2.2 The Yukawa couplings

By including the Higgs contribution in our theory, mass terms for fermions may now be generated by including a term of the form

$$\mathcal{L}_{\mathrm{Yukawa}} = -y_{\mathrm{f}} \overline{\psi} \phi \psi, \tag{1.16}$$

$$= -y_{\mathrm{f}} v \overline{\psi} \psi (1 + \frac{1}{v} \frac{h}{\sqrt{2}}) \tag{1.17}$$

which is invariant under $\mathrm{SU}(2)_{\mathrm{L}} \,\mathrm{x}\, \mathrm{U}(1)_Y$ gauge transformations due to the addition of ϕ. Similarly to the W and Z mass terms, the relation

$$m_{\mathrm{f}} = y_{\mathrm{f}} v. \tag{1.18}$$

is obtained. A curious feature of the SM is that the Yukawa-couplings y_{f} are free parameters of the theory with no a priori values. As a result these must be measured experimentally, with the measurement of the charm quark Yukawa coupling y_{c} being the goal of this work. Since the charm quark mass has previously been determined from experiment to be $m_{\mathrm{c}} = 1.27$ GeV [1], a measurement of y_{c} thus represents an important consistency test of the SM. To this end, one can exploited that an interaction between fermions and the Higgs field is introduced as can be seen in Equation 1.17, with an interaction strength proportional to y_{c}. It is exactly this feature that may be exploited by experiments at the LHC to measure y_{c}.

1.3 Measuring the charm quark Yukawa coupling

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

1.3.1 The cH process

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

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

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

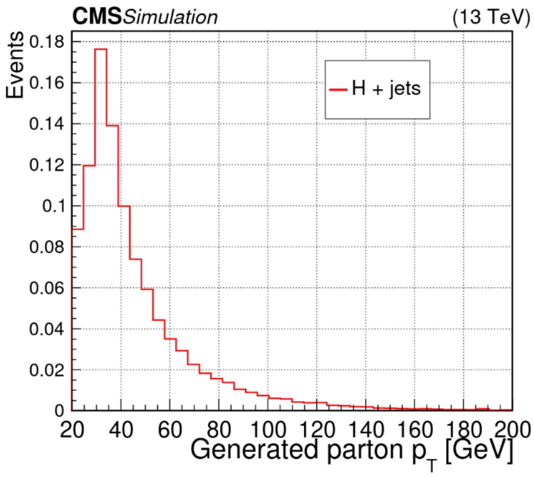

Figure 1.2: Transverse momentum of the charm quark produced alongside a Higgs boson in a simulation of the cH process, which typically takes on relatively small values. THIS IS A PLACEHOLDER PLOT.

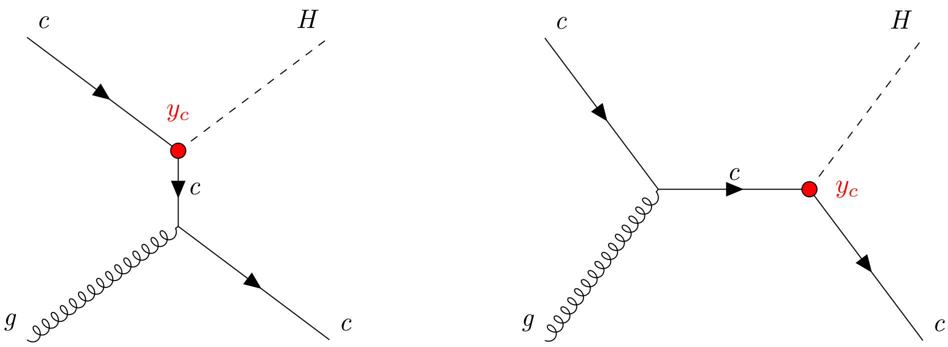

Figure 1.3: The leading order cH processes through which y_c may be probed as each diagramm contains a vertex with a charm-quark and Higgs boson, here denoted in red. The corresponding diagrams with an anti-charm quark $\overline{c}$ are implied.

1.3.2 The κ-framework

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

$$\kappa_{\mathrm{f}} = \frac{y_{\mathrm{f}}}{y_{\mathrm{f}}^{\mathrm{SM}}}. \tag{1.19}$$

where y_{f} is the measured Yukawa-coupling and $y_{\mathrm{f}}^{\mathrm{SM}}$ is the expected Yukawa-coulpling 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 $\kappa_{\mathrm{c}} = 1$. However, y_{c} is not a quantity that can be measured directly. Instead a signal strength measurement μ_{if}, where i represents the production process and f represents the decay process, relative to the SM expectation is made. Thus a measurement of μ_{if} must be converted into an interpretation of κ_{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

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

where σ_i is the production cross section in a given channel i and BR_f is the decay branching ratio in a given channel f. This can be rewritten as

$$\sigma_i \cdot \mathrm{BR}_f = \kappa_{r,i}\sigma_i^{\mathrm{SM}} \cdot \frac{\kappa_f \Gamma_f^{\mathrm{SM}}}{\Gamma_{\mathrm{H}}} \tag{1.21}$$

to give a general expression in which modifications to the production cross section and partial SM decay width Γ_f^{SF} are introduced via $\kappa_{r,i}$ and κ_f respectively. The denominator Γ_{H} represesents the total decay width which can be written as

$$\begin{aligned}\Gamma_{\mathrm{H}} =& \Gamma_{\mathrm{H}}^{\mathrm{SM}}(\kappa_{\mathrm{b}}^2\mathrm{BR}_{\mathrm{bb}}^{\mathrm{SM}} + \kappa_{\mathrm{W}}^2\mathrm{BR}_{\mathrm{WW}}^{\mathrm{SM}} + \kappa_{\mathrm{g}}^2\mathrm{BR}_{\mathrm{gg}}^{\mathrm{SM}} + \kappa_{\tau}^2\mathrm{BR}_{\tau\tau}^{\mathrm{SM}} + \kappa_{\mathrm{Z}}^2\mathrm{BR}_{\mathrm{ZZ}}^{\mathrm{SM}} + \kappa_{\mathrm{c}}^2\mathrm{BR}_{\mathrm{cc}}^{\mathrm{SM}} \\ &+ \kappa_{\gamma}^2\mathrm{BR}_{\gamma\gamma}^{\mathrm{SM}} + \kappa_{\mathrm{Z}\gamma}^2\mathrm{BR}_{\mathrm{Z}\gamma}^{\mathrm{SM}} + \kappa_{\mathrm{s}}^2\mathrm{BR}_{\mathrm{ss}}^{\mathrm{SM}} + \kappa_{\mu}^2\mathrm{BR}_{\mu\mu}^{\mathrm{SM}}) \qquad (1.22) \\ :=& \Gamma_{\mathrm{H}}^{\mathrm{SM}}\kappa_{\mathrm{H}}^2 \qquad (1.23)\end{aligned}$$

Here, $\Gamma_{\mathrm{H}}^{\mathrm{SM}}$ is the SM total decay width of the Higgs boson and $\mathrm{BR}_f^{\mathrm{SM}}$ 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 κ_f parametrises modifications thereof. Substituting Equation 1.23 into Equation 1.20, the rate modifier may be written as

$$\mu_{if} = \frac{\kappa_{r,i}^2\kappa_f^2}{\kappa_{\mathrm{H}}^2}. \tag{1.24}$$

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. $\mathrm{H} \to \mathrm{ZZ} \to 4\mu$) is unmodified, Equation 1.24 becomes

$$\mu_{if} = \frac{\kappa_{\mathrm{c}}^2}{\kappa_{\mathrm{H}}^2} \tag{1.25}$$

Using the flat direction approach discussed in [7] and [17], a simplification of κ_{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 κ_{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 κ. This allows Equation 1.24 to be rewritten as

$$\mu_{if} = \frac{\kappa^4}{\kappa^2(1 - \mathrm{BR}_{\mathrm{cc}}^{\mathrm{SM}}) + \kappa_{\mathrm{c}}^2 \mathrm{BR}_{\mathrm{cc}}^{\mathrm{SM}}} \tag{1.26}$$

which has a solution for κ given by

$$\kappa = \frac{(1 - \mathrm{BR}_{\mathrm{cc}}^{\mathrm{SM}})\mu}{2} + \frac{\sqrt{(1 - \mathrm{BR}_{\mathrm{cc}}^{\mathrm{SM}})^2 \mu^2 + 4\mu \mathrm{BR}_{\mathrm{cc}}^{\mathrm{SM}} \kappa_{\mathrm{c}}^2.}}{2}. \tag{1.27}$$

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

$$\mu_{\sigma_{\mathrm{cH}} \mathrm{BR(H \to ZZ)}} = \frac{2\kappa_{\mathrm{c}}^2}{0.97 + \sqrt{(0.97)^2 + 4 \cdot 0.97 \kappa_{\mathrm{c}}^2}}. \tag{1.28}$$

Rearranging for κ_{c} gives

$$\kappa_{\mathrm{c}} = \pm \frac{\sqrt{4 \cdot 0.97 \cdot \mu_{\sigma_{\mathrm{cH}} \mathrm{BR(H \to ZZ)}} \cdot (1 + \mu_{\sigma_{\mathrm{cH}} \mathrm{BR(H \to ZZ)}})}}{2}. \tag{1.29}$$

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

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

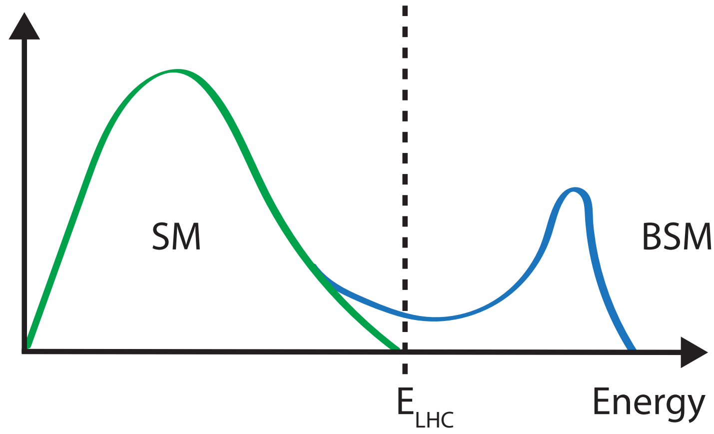

Figure 1.4: Illustration of how the presence of BSM physics, which is primarily visible beyond the reach of current collider energies (e.g. $\mathrm{E_{LHC}}$), can lead to subtle modifications of SM observables. These effects can be parametrised by SMEFT.

energy scale of the new physics scale Λ

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_{d>4} \sum_{i} \frac{C_i}{\Lambda^{d-4}} \hat{O}_i^d \tag{1.30}$$

where $\mathcal{L}_{\mathrm{SM}}$ 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

$$\mathrm{S} = \int \mathcal{L}\, d^4x \tag{1.31}$$

remains dimensionless. Accordingly, $\mathcal{L}_{\mathrm{SM}}$ is of energy dimension four. Since the SMEFT operators O_i^d all have energy dimensions higher than four and Λ 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.

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. 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 suppresive factor of Λ^{-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

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

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

1.4.1 The chromomagnetic dipole operator

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

$$\hat{O}_{cG} = (\overline{q}_{2,L}\sigma^{\mu\nu}T^a c)\tilde{\phi}G^a_{\mu\nu}. \tag{1.33}$$

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

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

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

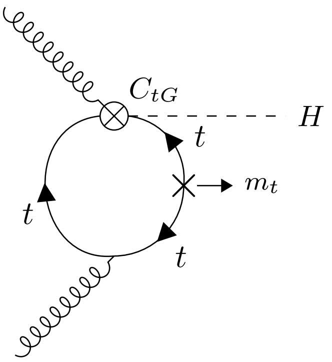

Figure 1.5: A modification of the gluon fusion process with a top quark loop by including the vertex introduced by the top quark CMD operator. Note that the arrows indicate chirality and not momentum flow. A chriality flip, denoted by the cross, proportional to the top quark mass m_t is required for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

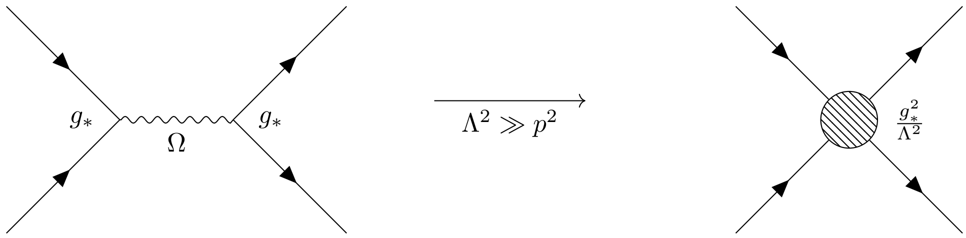

Figure 1.6: Feynman diagrams depicting a resonant process in which the new mediator particle Ω is created (left) and the approximate description of this in an EFT, where the diagram is reduced to a four-point interaction.

a new heavy mediator particle Ω 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 a first order approximation of the new particle's mediator as

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

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_Ω, which represents the previously introduced new physics scale Λ. Practically, this can be achieved by placing an upper limit $\mathrm{M_{cut}}$ on the total energy that is considered in measurements of such processes. The requirement can be expressed as

$$\mathrm{M_{cut}} < \Lambda. \tag{1.35}$$

A good estimator of $\mathrm{M_{cut}}$ is the inviariant 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

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

is obtained when a single BSM coupling g_* is introduced. This provides a prescription for the powers of the couplings and Λ that are associated with the SM fields ϕ, ψ and $F_{\mu\nu}$, and the covariant derivate D_μ. 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

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

$$= \frac{g_\star g_s}{\Lambda^2}\left(\psi_{L,R}\cdot\psi_{L,R}\cdot\phi\cdot G\right). \tag{1.38}$$

Reading off from Equation 1.38, one can see that the coupling of the CMD operator is given by $g_\star g_s/\Lambda^2$. 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

$$\frac{C_{cG}}{\Lambda^2} < \frac{g_* g_s}{\mathrm{M}_{\mathrm{cut}}^2} \tag{1.39}$$

is obtained. Since both C_{cG} and Λ 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

$$\frac{\left|\tilde{C}_{cG}\right| M_{\mathrm{cut}}^2}{4\pi g_s} < 1\,. \tag{1.40}$$

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

Chapter 2

The CMS experiment at the LHC

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

2.1 The CMS detector

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

2.1.1 The CMS coordinate system

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

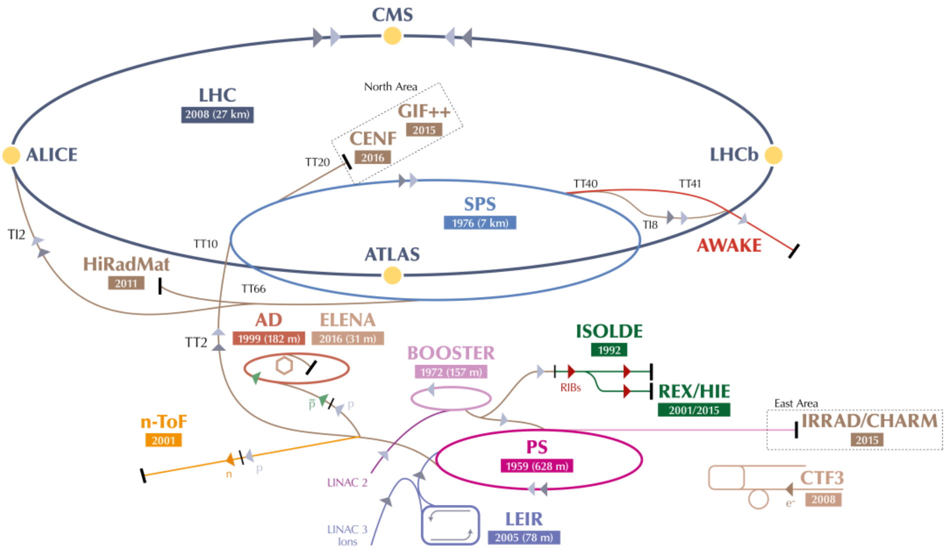

Figure 2.1: An overview of the LHC accelerator complex [27]. Before enetering the large LHC ring, particles must pass through a number of increasingly powerful set of accelerators.

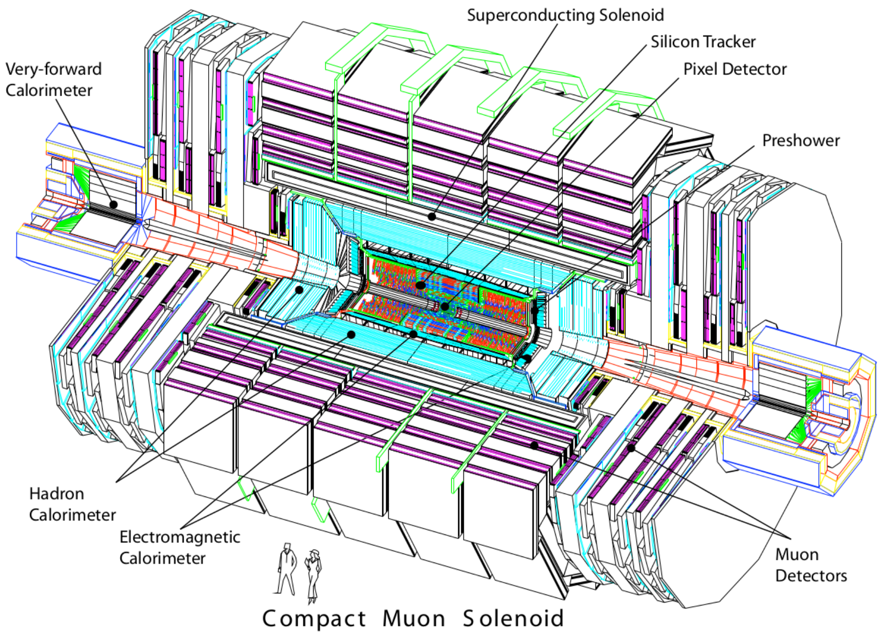

Figure 2.2: An overview of the CMS detector [28].

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

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

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

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

which is dependent solely on θ, 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 constitutents. Together with the particle's transverse (to the beam axis) momentum $\mathrm{p_T}$ and mass m, a particles four-vector may be described by

$$p = \begin{pmatrix} m \\ p_T \\ \eta \\ \phi \end{pmatrix}. \tag{2.3}$$

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

2.1.2 The silicon tracker

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

The pixel detector is situated right around the beampipe and as of 2017 consists of four circular layers of individual silicon pxiels in the barrel region and three disk layers in the endcap region. These consist of rectangular silicon chips with a size of 100 x 150 μm^2. When a charged particle traverses through the active material of these chips, an electrical signal is induced that is recorded. This is typically referred to as a *hit*. The small pixel size allows for position measurements with a very high resolution, namely $\sim 10\mu$m in the $r\phi$ direction and $\sim 20\mu$m in the z direction [31]. An important feature of the pixel detector its high radiation tolerance due to the close proximity of these modules to the beam pipe.

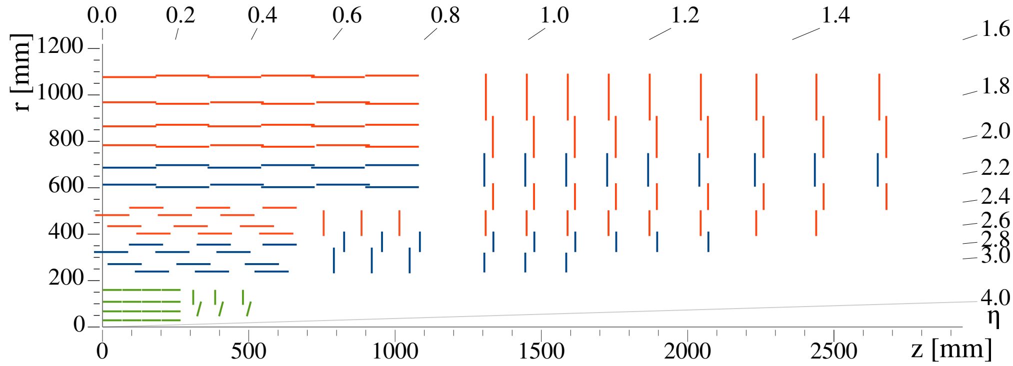

Figure 2.3: An overview of the CMS silicon tracker [30], shown in the r-z plane after its upgrade during Run-2. The pixel detector is denoted in green while the silicon strip detector is denoted in blue and orange.

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.

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 $\sim$ 1-4%.

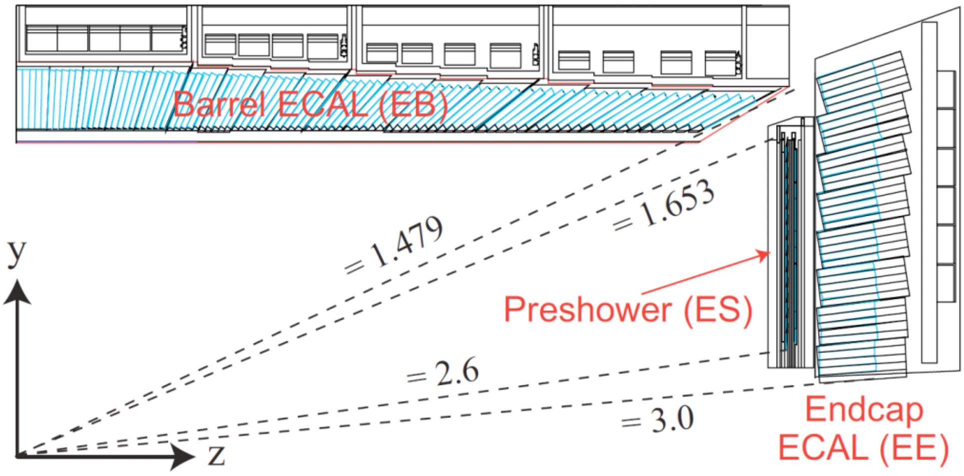

Figure 2.4: An overview of the CMS ECAL [34], shown in the r(y)-z plane. The dashed lines denote the coverage of the barrel and endcap ECAL region as well as the preshower detector.

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 homogenously 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 the trajectories of charged particles emerging from the interaction point in the ϕ-direction. This allows for a measurement of the particles transverse momentum p_T, which together with the ϕ and η directions fully characterise the particle's momentum vector. The magnet itself is composed of superconducting noibium-titanum coils that are cooled to a

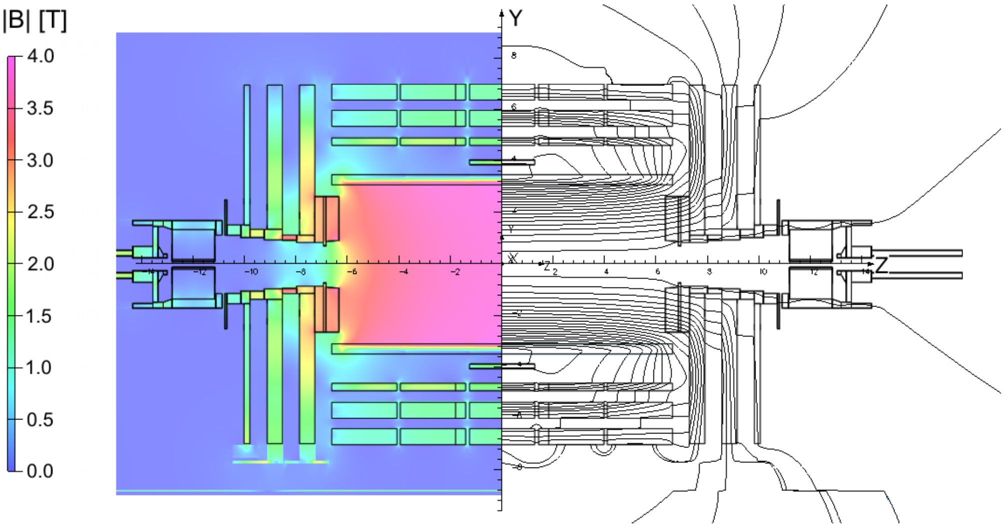

Figure 2.5: An overview of the magnetic flux (left) and magnetic field lines(right) inside the CMS detector, shown in the r-z plane [37].

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

2.1.6 The muon chambers

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

Like the other subdetectors, the muon chambers are separated in a barrel ($|\eta| < 1.2$) and endcap ($1.2 < |\eta| < 2.4$) region, which are composed of drift tubes and cathode strip chambers respectively. The drift tubes each consists of a gas volume containing a mixture of Argon and CO_2 in which a posititively charged wire is stretched through the center. When charged particles such as muons traverse these tubes, the gas is ionised. Due to the positive charge of the wire, the resulting electrons drift towards the wire producing an electrical signal. Thus the presence of muons may be determined by activation of the drift tubes. The cathode strip chambers on the other hand consist of layers of positively charged (anode) wires, which are arranged in a perpendicular fashion to a set of negatively charged (cathode) strips. Combining signals from both the wires and strips allows for a position measurement in both the R and ϕ direction. Both types of detector are supplemented by resistive plate chambers, which act as a trigger providing a precise timing resolution of $\sim$1ns. This makes it possible to unambiguously assign muons to individual collisions. These consist of parallel, oppositely charged plastic plates that are coated with a conductive graphite layer and are contained in a gas volume. Ionisation of the gas due to the traversal of a charged particle thus leads to an electrical signal. An overview of the spatial

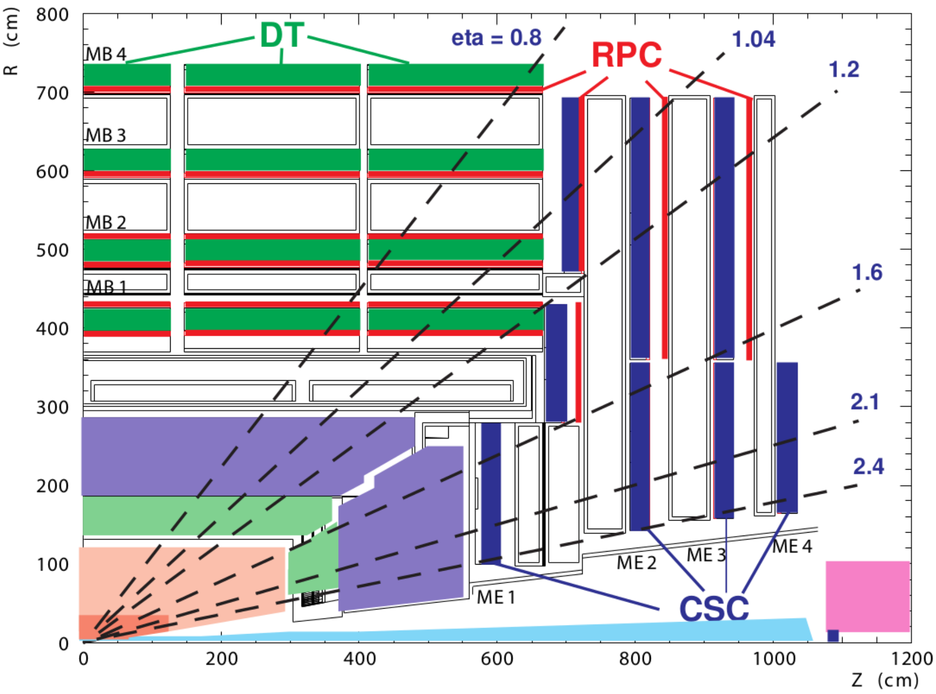

Figure 2.6: An overview of the CMS muon system, shown in the r-z plane [28]. Shown are the drift tube (DT), the cathode strip chambers (CSC) and resistive plate chambers (RPC).

arrangement of these systems can be see in Figure 2.6. With this system, the bulk of muons may be measured with a precise momentum resolution of ∼1-2%.

2.1.7 The triggering system

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

The trigger systems is composed of two subsystems. The first the so-called level one (L1) trigger. This is a very fast hardware-based system which reduces the event rate to ∼100 kHz by evaluating the presence of e.g. energetic muons or other interesting signatures such as large energy deposits in the calorimeters in an event. The total time allocated to decide whether an event should be kept is 3.2μs. Subtracting for signal propagation in the detector, the L1 system must make a decision within 1μs. From the L1, the events are passed to a software based high-level trigger (HLT) system. This is composed of several thousand CPU cores, performing a simple reconstruction of the event signatures to make a decision whether an event should be stored. Since different analyses are interested in different signatures, a set of trigger paths are defined so that only one such path must be satisfied for an event to pass the HLT. Since the HLT is software-based, the trigger paths may be continuously updated. After the HLT, the event rate is reduced to ∼100 Hz and the passing events are permanently stored.

2.1.8 The 2018 dataset of the CMS detector

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

2.2 Event reconstruction with the CMS detector

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

2.2.1 Track and vertex reconstruction

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

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

2.2.2 The Particle Flow algorithm

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

These are briefly summarised in Table 2.1.

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.

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

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

2.2.3 Reconstruction and identification of muons

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

- **Standalone muon tracks**: A standalone muon track refers to a fit of individual hits present in the muon detector.

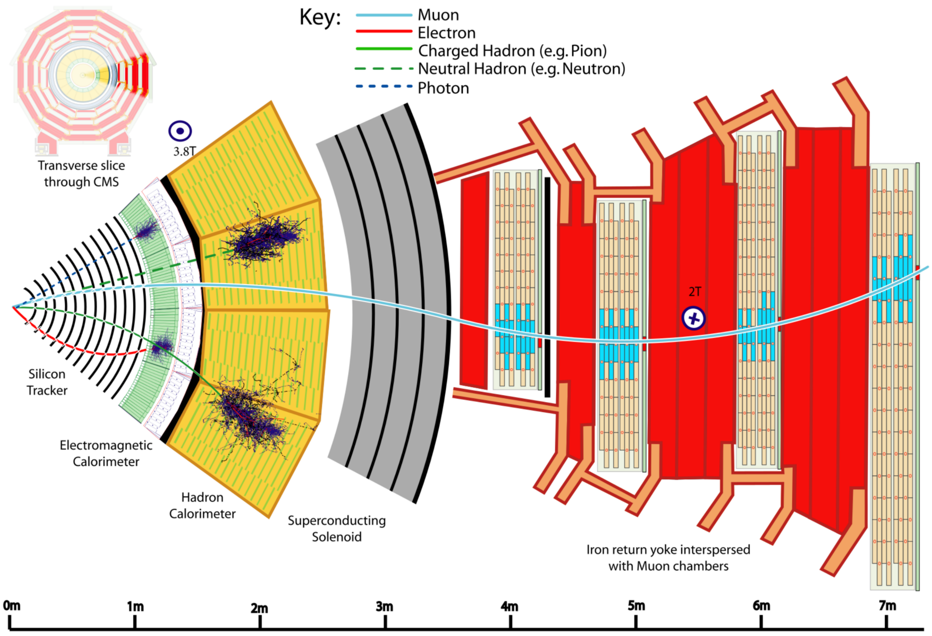

Figure 2.7: An transverse slice of the CMS detector, visualising the signatures that different particles produce in the different detector subsystems. [43].

- **Tracker muon tracks**: Tracker muon tracks are reconstructed by extrapolating tracks from the tracker to the muon detector, referred to as an *inside-out* approach. If a hit in the muon detector can be matched to the extrapolated track, then these matched tracks are identified as a tracker muon track. This reduces the impact from atmospheric muons traversing the detector, which may be falsely interpreted as standalone muon tracks.

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

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

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

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

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

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

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

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

2.2.4 Reconstruction and identification of jets

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

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

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

$$d_{iB} = \frac{1}{p_{\mathrm{T}}, i} \tag{2.6}$$

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

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

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

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

2.2.5 Missing transverse momentum

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

$$p_{\mathrm{T}}^{\mathrm{miss}} = \sum_{i}^{\mathrm{PF}} p_{\mathrm{T}}^{(i)}. \tag{2.8}$$

The presence of significant quantities of $p_{\mathrm{T}}^{\mathrm{miss}}$ may thus be used to identify the presence of neutrinos in an event.

2.3 Identification of charm quark-induced jets

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

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

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

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

2.3.1 Properties of heavy-flavour jets

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

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

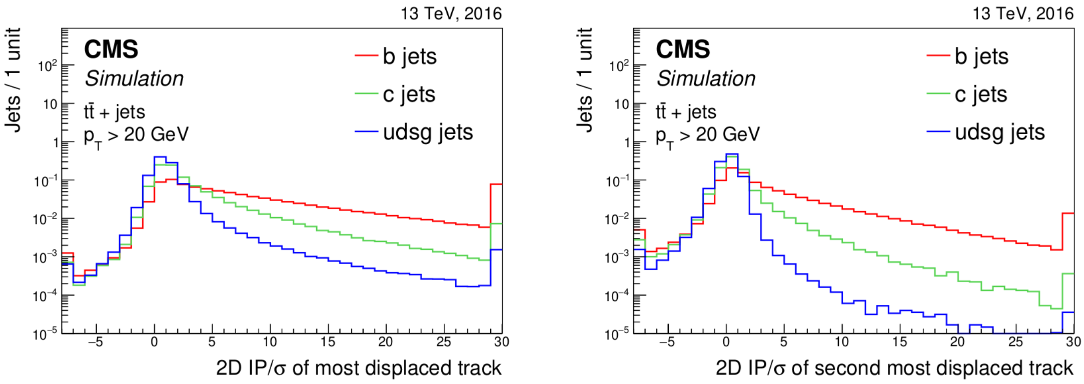

Figure 2.8: Plots showing the significance of the 2D impact parameter of the most and second most displaced tracks in a jet [50]. As can be seen, these variables can differentiate b and c jets from light jets to a significant degree.

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

2.3.2 The DeepJet algorithm

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

- ***b/bb/lepb* (*b* jets):** These three classes represent subclasses of jets originating from a b hadron. The b class represents a jet originating from a single b hadron, the bb class originating from two b hadrons and $lepb$ representing a jet originating from a b hadron with the presence of a soft lepton.

- ***c* (*c* jets)**: This class represents a jet originating from a c hadron.

- ***l, g* (light jets)**: These two classes represent light jets originating from a light quark or gluon respectively.

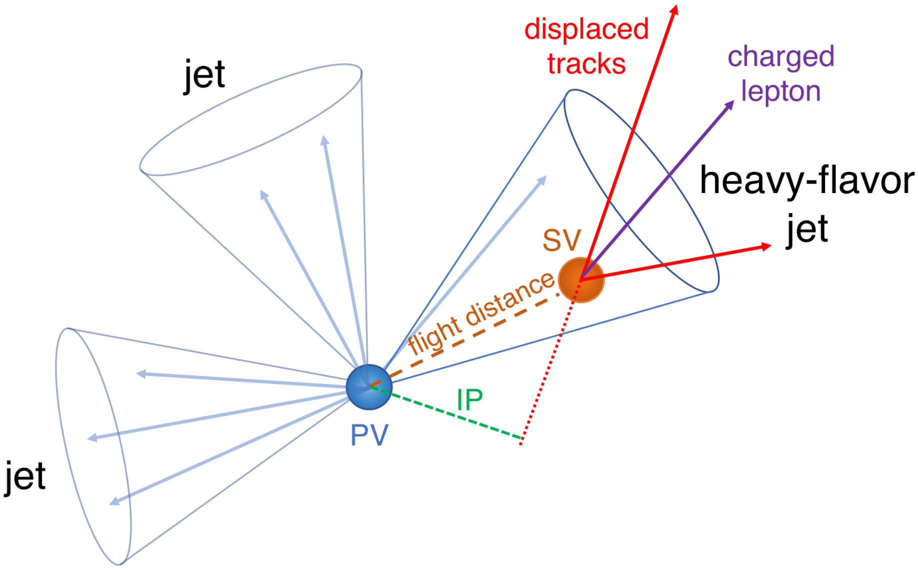

Figure 2.9: An illustration highlighting the properties of HF jets [50]. The presence of a secondary vertex (SV), characterised by the impact parameter (IP) in green, as well as the presence of a lepton is highlighted.

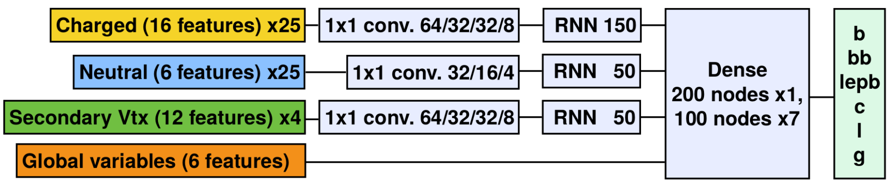

Figure 2.10: An illustration depicting the architecture of the DeepJet neural network [51]. Three individual branches separately process the charged hadrons, neutral hadrons and secondary vertex information before being passed onto a combining, fully connected layer together with global variables.

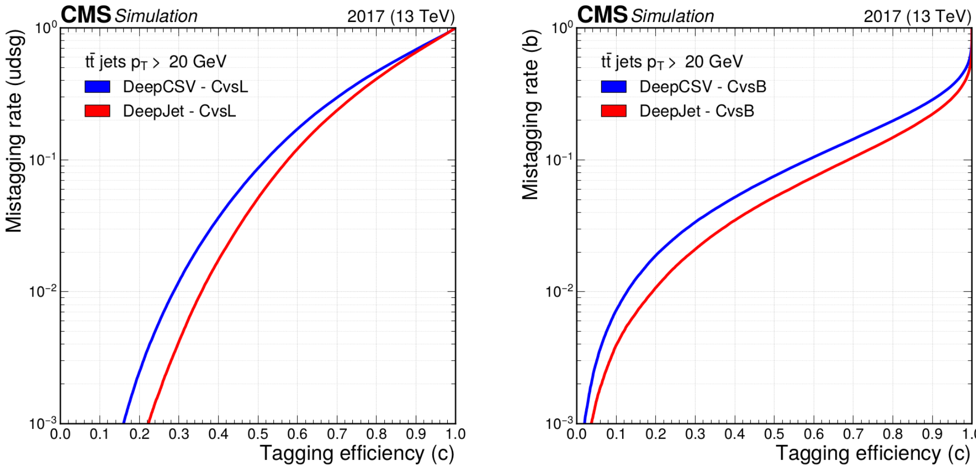

Figure 2.11: Performance of DeepJet algorithm in identifying c jets against b jets and light jets in simulated samples of top quark pair production, in which both top quarks decay hadronically [52]. The x-axis represents the efficiency with which c jets are identified, while the y-axis represents mis-identification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

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

(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 see 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 $\mathrm{W}^{\pm}$ 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 interative 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.

Chapter 3

Search for the cH(ZZ→4μ) process

To probe the charm Yukawa coupling through the cH process, a methodology must be devised to select and reconstruct cH candidate events. This is described in section 3.1, specifically targetting cH(ZZ→ 4μ) final states. Additionally, a model describing the expected contributions from the cH(ZZ→ 4μ) process as well as a number of background processes in the event selection must be constructed and is described in section 3.2. Finally, a statistical evaluation using flavour-tagging discriminators to set 95% CL upper limits on κ_c, assuming the absence of signal, is presented in ??.

3.1 cH event selection

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

To reconstruct a Higgs (jet) candidate, an initial selection of muon (jet) objects must be made. These are summarised in Table 3.1 along with the HLT trigger path requirement used in this analysis. The objective of this selection is to identify events with well-reconstructed, isolated muons as well as a least on well-reconstructed jet. Following this initial selection, the corresponding 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	$\mathrm{p_T} > 5$ GeV $\mid \eta \mid < 2.4$ Tight muon identification criteria
Jets	$\mathrm{p_T} > 25$ GeV $\mid \eta \mid < 2.5$ Jet ID Pile-up ID, loose working point ΔR(jet, selected muons) < 0.4
HLT	HLT_IsoMu24 is triggered

3.1.1 Higgs candidate selection

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

1. Of the four selected muons, the $\mathrm{p_T}$-leading muon is required to satisfy $\mathrm{p_T}$ >20 GeV and the sub-leading muon is required to satisfy $\mathrm{p_T}$ >10 GeV. Additionally, the HLT_IsoMu24 trigger requirement must be met. Lastly, to ensure two muons are not spuriously reconstructed from shared tracks, it is required that each muon candidate is separated from the others by ΔR >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 oppsite-sign muons must satisfy $m_{\mu\mu}>$ 4 GeV, to remove any contributions from low mass resonances such as J/ψ.

3. The Z candidate with a mass closest to the known Z boson mass of Z = 91.19 GeV [1] is interpreted as an on-shell Z_1 candidate. The Z_1 candidate should satisfy 40 GeV $< m_{Z_1}$ $<$ 120 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 GeV $< m_{Z_1} <$ 120 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 GeV.

The reconstructed Higgs boson candidate mass distribution in simulated cH(ZZ→ 4μ) events can be seen in **??**. As expected, a peak around the known Higgs mass m_H = 125.3 GeV can be oberserved, 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→ 4μ) sample. The majority off 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 $\Delta\phi(\mathrm{H}, \mathrm{jet})$ 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 eachother *back-to-back* and thus $\Delta\phi(\mathrm{H}, \mathrm{jet})$ is expected to be $\sim \pm\pi$.

2. Since the Higgs boson and associated jet recoil off eachother, their $\mathrm{p_T}$ is expected to be approximately balanced. This information can be captured by transverse momentum ratio $p_\mathrm{T}(H)/p_\mathrm{T}(\mathrm{jet})$.

To derive the relevant distributions to be used in a likelihood ratio, a parton-to-jet matching is performed in simulated cH(ZZ$\to 4\mu$) 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 $\Delta R(\mathrm{jet}, \mathrm{parton}) < 0.3$. All jets which match the initial jet selection are considered in this process. A jet which is matched in this way is labelled as the associated jet, while the remaining non-matched jets are labelled non-associated jets. The efficiency with which this matching is performed can be seen in **??**. Once this labelling is performed, the distributions of $\Delta\phi(\mathrm{H}, \mathrm{jet})$ and $p_\mathrm{T}(H)/p_\mathrm{T}(\mathrm{jet})$ 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 $\mathrm{p_T}$ Higgs candidates, this procedure is repeated in different bins of $\mathrm{p_T(H)}$ 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

$$\mathcal{L}(x) = \frac{\mathcal{L}_\mathrm{associated}(x)}{\mathcal{L}_\mathrm{non-associated}(x)},\ \mathrm{with}\, x \in \Big\{\Delta\phi(\mathrm{H}, \mathrm{jet}), \frac{p_\mathrm{T}(H)}{p_\mathrm{T}(\mathrm{jet})}\Big\} \tag{3.1}$$

is defined. From this follows the per-jet likelihood

$$\mathcal{L}(\mathrm{jet}) = \mathcal{L}\Big(\Delta\phi(\mathrm{H}, \mathrm{jet})\Big) \cdot \mathcal{L}\Big(\frac{p_\mathrm{T}(H)}{p_\mathrm{T}(\mathrm{jet})}\Big) \tag{3.2}$$

that is evaluated. The jet with the highest associated likelihood in an event is selected as the jet candidate. The efficiency with which the "correct" associated jet can be seen in **??**.

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

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

With this, the individual components of the cH(ZZ→ 4μ) are thus reconstructed and events satisfying the described requirements are selected for evaluation.

3.2 Signal and background estimation

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

3.2.1 Monte Carlo simulation of proton-proton collisions

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

1. **The hard scattering process**: The hard scattering process refers to the immediate, high energy transfer scattering of two protons resulting in the production of additional particles. To calculate this, two main ingredients are required. The first is a calculation of the matrix elements that describe the simulated process in which proton constituents collide to produce additional particles. These matrix elements allow for the calculation of a cross section for the process. However, the proton itself is a complex object consisting not only of its valence quarks (two up-type quarks and one down-type quark) but also of a constantly changing ensemble of additional quarks and gluons that are created and annihilated. This behaviour must thus be captured for an accurate process description and is parametrised via so-called *Parton Distribution Functions*. These describe the likelihood with which a parton, that carries some fraction x of the protons total momentum, may be found in a proton at some energy scale Q^2. The evolution of the PDF with changing Q^2 is described

by the DGLAP quations [55]. Software used to simulate the hard scattering are referred to as *event generators*. Commonly used event generators include Madgraph5_aMC@NLO [56] and POWHEG [57].

2. **Parton showering**: Particles such as quarks and gluons that are produced in the hard scattering carry the colour charge of the strong interaction. As a result, these may produce soft radiation or branch into other particles. While a most physically accurate description would be given by including these contributions in the calculation of the hard scattering process, this greatly increases the complexity of the calculation. As such a *parton shower* model, such as in the Pythia software package [58], is used instead to describe the splitting of a single mother particle into two daughter particles. In QCD, this describes to gluon radiation ($q \to qg$) and gluon splitting ($g \to gg$ and $g \to q\bar{q}$) and in QED describes Bremsstrahlung ($f \to f\gamma$) and pair creation ($\gamma \to f\bar{f}$). In case this the parton showering originates from initial state partons it is referred to as initial state radiation (ISR). Accordingly, parton showering originating from final state partons is referred to as final state radiation (FSR). In cases with final states containing multiple partons, there can be some ambiguity in the combination of matrix elements and parton showering since both can describe the same processes. For this merging schemes are applied that resolve potential double counting of events. A prescription used for this work is the FxFx scheme [59].

3. **Hadronisation**: At an energy around the QCD scale Λ_{QCD}, the perturbative parton shower prescription loses its validty as the running coupling of the strong force α_s becomes too strong. Here the individual, colour-charged partons *hadronise* into colour-neutral states. Since this process currently cannot be described from first principles, a phenomenological description must be applied. In Pythia, the *Lund string* model is used [60]. It describes the interaction between two partons as a coloured field, the lines of which pass through a tube that is extended between the partons. The potential energy of the tube (or string) is described by a term linear in the distance between the partons. Thus if the partons are separated at a large enough distance and the potential energy is sufficiently large, the string may 'break' and new colourless quark-antiquark pairs are formed. This procedure may be repeated with these new parton pairs if they posses an invariant mass above some threshold.

4. **The underlying event**: A description of a variety of effects secondary to the hard scattering must be included in the simulation. These can have several origins such as secondary, *soft* interactions of the proton-proton collision or remnants of the collided protons, which will hadronise themselves. These effects are modeled from data [61].

5. **Detector simulation**: Finally, the detector response to the particles emerging from the previously describeds steps must be simulated. This is performed with the GEANT4 package [62], which is configured to model the CMS detector. This includes modelling the curving of particle trajectories due to the detector's magnet, the interaction of particles with the materials of the detector, as well as the digitisation of the signals in the electronic modules of the subdetectors.

A diagrammatic overview of what an event simulation looks like can be found in Figure 3.1. The output of this simulation is passed to the reconstruction algorithms described in section 2.2.

3.2.2 Estimation of cH(ZZ→ 4μ) process

The cH(ZZ→ 4μ) process is estimated using a simulation generated by MadGraph5_aMC@NLO. The following MadGraph5_aMC@NLO syntax is used, which illustrates some important concepts

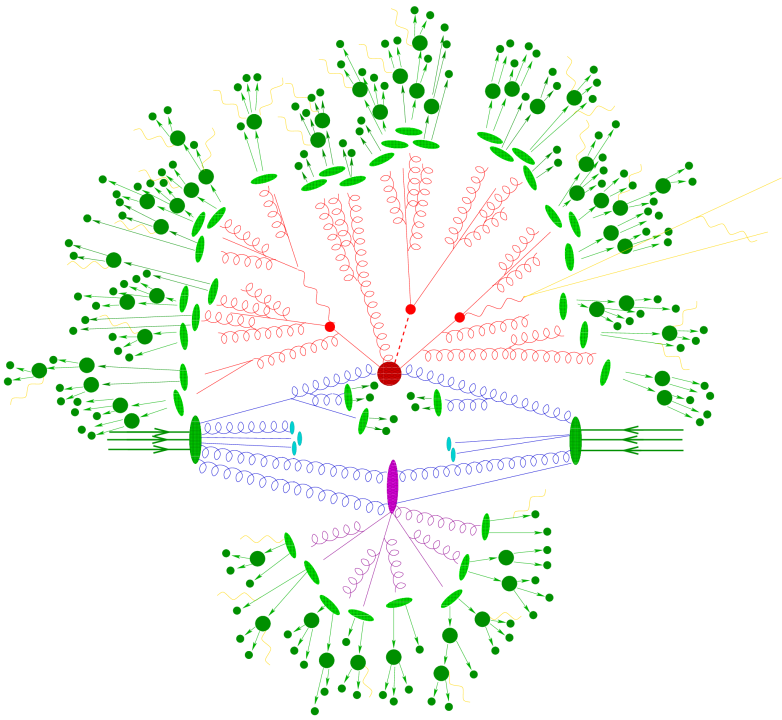

Figure 3.1: An overview of what an event simulation may look like (adapted from [63]).

related to the simulation of the cH process:

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

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

In the following two lines, the proton and jet constituents are defined. Finally in the last two lines, the processes included in the simulation are defined. These are, calculated to next to leading order in QCD, the $pp \to H$ and $pp \to H + j$ processes. Both are included to give the most accurate possible kinematic description of the cH process. The reasoning for this is related to the modelling of final state partons and can be better understood by considering what is included in the leading order (LO) and next-to-leading (NLO) contributions to $pp \to H$ and $pp \to H + j$ respectively. At leading order, an additional jet in $pp \to H$ can only be generated via the parton shower. Thus, this contribution is expected to best model the lower momentum behaviour of final state partons. The NLO contributions to $pp \to H$, which correspond to LO contritubtions of $pp \to H + j$, in turn are expected to beeter model higher momentum behaviour of the final state parton. The same logic is applied to the LO contributions of $pp \to H + j$ and the NLO contributions of $pp \to H + j$, where two final state partons explicitly appear in the calculation of the latter. This approach clearly introduces double counting of processes, however these are automatically accounted for by the event generator. Similarly, the FxFx merging scheme is used to remove any double counting between parton shower and matrix element contributions.

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

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

Process	**Tag**	σ
cH(ZZ$\to 4\mu$) 4FS FxFx	HPlusCharm_4FS_MuRFScaleDynX0p50_HToZZTo4L_M125_TuneCP5_13TeV_amcatnloFXFX_JHUGenV7011_pythia8	xx
cH(ZZ$\to 4\mu$) 3FS	HPlusCharm_3FS_MuRFScaleDynX0p50_HToZZTo4L_M125_TuneCP5_13TeV_amcatnlo_JHUGenV7011_pythia8	xx
cH(ZZ$\to 4\mu$) 4FS	HPlusCharm_4FS_MuRFScaleDynX0p50_HToZZTo4L_M125_TuneCP5_13TeV_amcatnlo_JHUGenV7011_pythia8	xx

Something about c jet vs non-c jet in yc sensitive events.. check cH sample definition

3.2.3 Estimation of irreducible backgrounds

Irreducible background processes are background processes that produce the same final state particles as the signal process in question. Thus, processes which produce four final state muons along with the presence of a, or several, jet(s) constitute the irreducible background. These again fall into two categories, namely those where the four muons originate from a Higgs boson and those in which they do not. When analysing e.g. the mass spectrum of Higgs candidates, th processes of the former category is thus clearly resonant around the Higgs mass of $\sim$125 GeV, while those of the latter category may take on a more continuous shape. The irreducible backgrounds of this analysis are estimated using simulation. An overview of the samples that are used can be found in Table 3.4. Like with the cH(ZZ$\to 4\mu$) process, additional samples are used for the bH($ZZ \to 4\mu$) background to account for an uncertainty related to the choice of flavour scheme. However, here the five flavour scheme (5FS) is now the nominal FS to include the bottom quark in the proton.

Table 3.4: Simulated processes used for background estimation in this work.

Process	**Tag**	σ
ggH(ZZ→4L)	GluGluHToZZTo4L_M125_TuneCP5_13TeV_powheg2_JHUGenV7011_pythia8	xx
ttH(ZZ→4L)	ttH_HToZZ_4LFilter_M125_TuneCP5_13TeV_powheg2_JHUGenV7011_pythia8	xx
W^-H(ZZ→4L)	WminusH_HToZZTo4L_M125_TuneCP5_13TeV_powheg2-minlo-HWJ_JHUGenV7011_pythia8	xx
W^+H(ZZ→4L)	WplusH_HToZZTo4L_M125_TuneCP5_13TeV_powheg2-minlo-HWJ_JHUGenV7011_pythia8	xx
ZH(ZZ→4L)	ZH_HToZZ_4LFilter_M125_TuneCP5_13TeV_powheg2-minlo-HZJ_JHUGenV7011_pythia8	xx
qqH(ZZ→4L)	VBF_HToZZTo4L_M125_TuneCP5_13TeV_powheg2_JHUGenV7011_pythia8	xx
tqH(ZZ→4L)	tqH_HToZZTo4L_M125_TuneCP5_13TeV-jhugenv7011-pythia8	xx
gg→ZZ(4μ)	GluGluToContinToZZTo4mu_TuneCP5_13TeV-mcfm701-pythia8	xx
gg→ZZ(4τ)	GluGluToContinToZZTo4tau_TuneCP5_13TeV-mcfm701-pythia8	xx
gg→ZZ(2μ2τ)	GluGluToContinToZZTo2mu2tau_TuneCP5_13TeV-mcfm701-pythia8	xx
qq→ZZ/Zγ^* →4L	ZZTo4L_TuneCP5_13TeV_powheg_pythia8	xx
bH 5FS FxFx	HPlusBottom_5FS_MuRFScaleDynX0p50_HToZZTo4L_M125_TuneCP5_13TeV_amcatnloFXFX_JHUGenV7011_pythia8	xx
bH 4FS	HPlusBottom_4FS_MuRFScaleDynX0p50_HToZZTo4L_M125_TuneCP5_13TeV_amcatnlo_JHUGenV7011_pythia8	xx
bH 5FS	HPlusBottom_5FS_MuRFScaleDynX0p50_HToZZTo4L_M125_TuneCP5_13TeV_amcatnlo_JHUGenV7011_pythia8	xx
WZ→ 3$\ell\nu$	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	xx

3.2.4 Estimation of reducible backgrounds

Reducible background processes are background processes that do not produce the same final state particles as the signal process but where mis-identification of physics objects can still falsify the sought-after signature. Since jets are typically abundant in most collisions, this amounts to the mis-identification of additional muons for this analysis. A major expected contribution in this background is expected from the Drell-Yan process. Since the simulation of mis-identified muons is subject to significant modelling uncertainties, a data-driven approach may be used. This involves determining the mis-identificaiton rate of muons in data and applying it to a side-band region from which the contributions of reducible backgrounds are extrapolated into the signal region. This methodology is presented in this section and follows the methods used in [53].

Determination of muon mis-identification rate

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

1. Events that contain at least two muons that pass the tight identification requirement and where the third passes at least the loose identification requirement are chosen. The p_T-leading muon is required to satisfy p_T >20 GeV and the sub-leading muon is required to satisfy p_T >10 GeV. Additionally, the HLT_IsoMu24 trigger requirement must also be met. Lastly, to ensure two muons are not spuriously reconstructed from shared tracks, it is required that each muon candidate is separated from the others by ΔR >0.02.

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

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

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

Application of muon mis-identification rate

The muon mis-identification rate is applied to a control region to estimate the contribution of reducible backgrounds to the previously described cH(ZZ$\rightarrow 4\mu$) selection. It is useful to introduce some of the related terminology at this point. The four pass (4P) region henceforth refers to the inclusive signal region that is defined via the cH(ZZ$\rightarrow 4\mu$) selection. The three-pass-one-fail (3P1F) and two-pass-two-fail (2P2F) regions respectively refer to regions in which the cH(ZZ$\rightarrow 4\mu$) reconstruction is performed as previously described but where only three (two) of the muons satisfy the tight identification criteria and the remaining one (two) muon(s) satisfy only the loose identification criteria. The 3P1F and 2P2P are collectively referred to as the application region (AR).

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

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

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

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

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

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

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

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

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

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

$$N_{\text{4P}}^{(\text{2P2F})} = \sum_{k}^{N_{\text{2P2F}}} \frac{f_k^{(3)}}{(1-f_k^{(3)})} \frac{f_k^{(4)}}{(1-f_k^{(4)})}, \tag{3.7}$$

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

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

$$N_{\text{4P}}^{(\text{3P1F})} = \sum_{k}^{N_{\text{3P1F}}} \frac{f_k^{(4)}}{(1-f_k^{(4)})}, \tag{3.8}$$

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

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

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

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

$$N_{\text{4Pexp.}}^{(\text{3P1F})} = \sum_{k}^{N_{\text{2P2F}}} \left(\frac{f_k^{(3)}}{(1-f_k^{(3)})} \frac{f_k^{(4)}}{(1-f_k^{(4)})} + \frac{f_k^{(4)}}{(1-f_k^{(4)})} \frac{f_k^{(3)}}{(1-f_k^{(3)})}\right) \tag{3.10}$$

$$= 2 \sum_{k}^{N_{\text{2P2F}}} \left(\frac{f_k^{(3)}}{(1-f_k^{(3)})} \frac{f_k^{(4)}}{(1-f_k^{(4)})}\right) \tag{3.11}$$

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

$$N_{\mathrm{4P}}^{(\mathrm{ZZ,3P1F})} = \sum_{k}^{N_{\mathrm{3P1F}}^{ZZ}} \frac{f_k^{(4)}}{(1-f_k^{(4)})}, \tag{3.12}$$

The total contribution to the 4P region N_{4P} may finally be estimated by taking the sum of the 2P2F and 3P1F contributions and subtracting the discussed contamination terms. This leads to

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

$$= \sum_{k}^{N_{\mathrm{3P1F}}} \frac{f_k^{(4)}}{(1-f_k^{(4)})} - \sum_{k}^{N_{\mathrm{3P1F}}^{ZZ}} \frac{f_k^{(4)}}{(1-f_k^{(4)})} - \sum_{k}^{N_{\mathrm{2P2F}}} \left(\frac{f_k^{(3)}}{(1-f_k^{(3)})}\frac{f_k^{(4)}}{(1-f_k^{(4)})}\right) \tag{3.14}$$

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

...

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

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

3.4 Statistical inference

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

3.4.1 Statistical model

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

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

is appropriate. Here, λ denotes the expectation value with k observed events. In the i-th bin of the discriminator distribution, the expecation value λ_i is calculated using the models of the signal and background processes discussed in section 3.2. This amounts to $\lambda_i = \mu s_i + b_i$ where s_i and b_i are the signal and background estimations respectively. The signal strength modifier μ allows for an arbitrary scaling of the signal contribution and is used as a floating parameter in the inference process. From this, a combined binned likelihood function

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

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

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

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

$$p_\mu = \int_{q_{\text{obs.}}}^{\infty} f(q \,|\, \mu) \mathrm{d}q. \tag{3.18}$$

With this p-value, the $\mathrm{CL_s}$ method may be applied [**CLs**]. For this, a $\mathrm{CL_s}$ value

$$\mathrm{CL_s} = \frac{p_\mu}{1 - p_0} = \frac{\int_{q_{\text{obs.}}}^{\infty} f(q \,|\, \mu)\, \mathrm{d}q}{\int_{q_{\text{obs.}}}^{\infty} f(q \,|\, 0)\, \mathrm{d}q}. \tag{3.19}$$

is computed. A feature of this method is that large overlaps of test statistic distributions between the H_μ and H_0 hypotheses, due to the very small signal cross section associated with H_μ, are accounted for. This is achieved by weighting p_μ with $(1 - p_0)^{-1}$, thus decreasing the $\mathrm{CL_s}$ value with larger overlaps.

3.4.2 Uncertainty model

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

Theoretical uncertainties common to simulation

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

- Shape uncertainties related to the choice of the normalisation and factorisation scales μ_R and μ_F. These are varied independently by factors of 2 and 0.5 using a reweighting technique that is applied to the simulation, thus introducing four shape-changing nuisance parameters.

- Shape uncertainties related to the modelling of the Parton Distribution Function (PDF). These uncertainties are obtained from the NNPDF3.1 PDF set that is used in simulation [**NNPDF3.1**] and includes 100 PDF variations associated with variations of indiviudal parameters. These are combined into a single nuisance parameter with the prescription

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

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

- Normalisation uncertainties related to the cross sections of the individual processes. These are:

  - Higgs production and branching ratio uncertainties taken from [16]
  - A 50% normalisation uncertainty on the modelling of the gluon fusion process in association with heavy quarks ?? Ask gerrit.
  - Other process uncertainties...

- Shape uncertainties related to the tuning of the parton showering used in the generator. These are implemented by varying parameters related to initial-state radiation and final-state radiation independently, introducing two nuisance parameters.

- A shape uncertainty related to the application of a pile-up reweighting procedure, with which the pile-up profile of simulation is matched to that of data. These uncertainties are captured via a single nuisance parameter.

Experimental uncertainties common to simulation

The following experimental uncertainties are common to all simulated sample:

- Shape uncertainties related to the jet energy scale. For this, a simplified schema is used in which closely correlated uncertainty sources are grouped and captured across 11 individual nuisance parameters. Sources of uncertainty include for example the limited set of data available for the calibration procedure or differences in calibration response to different jet flavours.

- A shape uncertainty related to the smearing method used to correct the jet energy resolution. This is captured via a single nuisance parameter.

- Shape uncertainties related to the calibration of the jet flavour-tagging algorithm that is used. These are captured via 13 nuisance parameters and include for example uncertainty sources related to the individual phase spaces targeted in the calibration or the limited set of avaialble data.

- A shape uncertainty related to the pile-up identification method that is used. This is captured via a single nuisance parameter.

- Shape uncertainties related to the muons used in the analysis. The response of the muon identification criteria, the isolation as well as the muon trigger path in simulation is calibrated to match data. Uncertainties associated with this calibration are captured via a nuisance parameter.

- A normalisation uncertainty related to the luminosity with which the simulation is scaled, which is known with a limited precision. This amounts 2.5% for the 2018 dataset and is caputured as a single nuisance parameter.

Uncertainties related to cH(ZZ→ 4μ) and bH(ZZ→ 4μ) modelling

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

Uncertainties related to reducible background modelling

Due to the almost neglible nature of the reducible background contribution, a simple normalisation uncertainty is introduced to account for uncertainties associated with this estimation procedure...

Uncertainties related to statistical precision of background estimation

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

3.5 Results

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

Chapter 4

An EFT interpretation of the cH(ZZ$\rightarrow$4μ) process

Conclusion

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.

[12] CMS Collaboration. *Simultaneous probe of the charm and bottom quark Yukawa couplings using ttH events*. 2025. DOI: 10.48550/arXiv.2509.22535. arXiv: 2509.22535 [hep-ex]. URL: https://arxiv.org/abs/2509.22535.

[13] CMS Collaboration. *Search for a Higgs boson produced in association with a charm quark and decaying to a W boson pair in proton-proton collisions at $\sqrt{s} = 13$ TeV*. 2025. arXiv: 2508.14988 [hep-ex]. URL: https://arxiv.org/abs/2508.14988.

[14] CMS Collaboration. *Search for the associated production of a Higgs boson with a charm quark in the diphoton decay channel in pp collisions at $\sqrt{s} = 13$ TeV*. 2025. arXiv: 2503.08797 [hep-ex]. URL: https://arxiv.org/abs/2503.08797.

[15] Nuoyu Dong et al. "Probing charm Yukawa coupling through *ch* associated production at the hadron colliders". In: *Phys. Rev. D* 111 (5 Mar. 2025), p. 053003. DOI: 10.1103/PhysRevD.111.053003. URL: https://link.aps.org/doi/10.1103/PhysRevD.111.053003.

[16] C T Potter et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. en. 2013. DOI: 10.5170/CERN-2013-004. URL: http://cds.cern.ch/record/1559921.

[17] J. A. Aguilar-Saavedra, J. M. Cano, and J. M. No. "More light on Higgs flavor at the LHC: Higgs boson couplings to light quarks through $h + \gamma$ production". In: *Phys. Rev. D* 103 (9 May 2021), p. 095023. DOI: 10.1103/PhysRevD.103.095023. URL: https://link.aps.org/doi/10.1103/PhysRevD.103.095023.

[18] The CMS Collaboration. "Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *The European Physical Journal C* 79.5 (May 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6909-y. URL: http://dx.doi.org/10.1140/epjc/s10052-019-6909-y.

[19] Steven Weinberg. "Baryon and Lepton Nonconserving Processes". In: *Phys. Rev. Lett.* 43 (1979), pp. 1566–1570. DOI: 10.1103/PhysRevLett.43.1566.

[20] B. Grzadkowski et al. "Dimension-six terms in the Standard Model Lagrangian". In: *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.

[21] Gino Isidori, Felix Wilsch, and Daniel Wyler. "The standard model effective field theory at work". In: *Reviews of Modern Physics* 96.1 (Mar. 2024). ISSN: 1539-0756. DOI: 10.1103/revmodphys.96.015006. URL: http://dx.doi.org/10.1103/RevModPhys.96.015006.

[22] CMS Collaboration. *Combined effective field theory interpretation of Higgs boson, electroweak vector boson, top quark, and multi-jet measurements*. 2025. arXiv: 2504.02958 [hep-ex]. URL: https://arxiv.org/abs/2504.02958.

[23] Joseph Bramante et al. "Boosted Higgs bosons from chromomagnetic b's: $b\bar{b}h$ at high luminosity". In: *Phys. Rev. D* 93 (5 Mar. 2016), p. 053001. DOI: 10.1103/PhysRevD.93.053001. URL: https://link.aps.org/doi/10.1103/PhysRevD.93.053001.

[24] J. Elias-Miró et al. "Higgs windows to new physics through d=6 operators: constraints and one-loop anomalous dimensions". In: *Journal of High Energy Physics* 2013.11 (Nov. 2013). ISSN: 1029-8479. DOI: 10.1007/jhep11(2013)066. URL: http://dx.doi.org/10.1007/JHEP11(2013)066.

[25] The CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://doi.org/10.1088/1748-0221/3/08/S08004.

[26] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: `10.5170/CERN-2004-003-V-1`. URL: `http://cds.cern.ch/record/782076`.

[27] E. A. Mobs. "The CERN accelerator complex. Complexe des accélérateurs du CERN". In: (Oct. 2016). General Photo. URL: `http://cds.cern.ch/record/2225847`.

[28] The CMS Collaboration. *CMS Physics : Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. Geneva: CERN, 2006. URL: `https://cds.cern.ch/record/922757`.

[29] CMS Collaboration. "Development of the CMS detector for the CERN LHC Run 3". In: *Journal of Instrumentation* 19.05 (May 2024), P05064. ISSN: 1748-0221. DOI: `10.1088/1748-0221/19/05/p05064`. URL: `http://dx.doi.org/10.1088/1748-0221/19/05/P05064`.

[30] *CMS Tracker Detector Performance Public Results*. Last accessed 24.11.25. URL: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK`.

[31] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009–P10009. ISSN: 1748-0221. DOI: `10.1088/1748-0221/9/10/p10009`. URL: `http://dx.doi.org/10.1088/1748-0221/9/10/P10009`.

[32] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Tech. rep. Geneva, 1997. URL: `https://cds.cern.ch/record/349375`.

[33] CMS Collaboration. *The CMS ECAL performance with examples*. Tech. rep. Geneva: CERN, 2014. DOI: `10.1088/1748-0221/9/02/C02008`. URL: `https://cds.cern.ch/record/1632384`.

[34] CMS Collaboration. *The CMS ECAL performance with examples*. Tech. rep. Geneva: CERN, 2014. DOI: `10.1088/1748-0221/9/02/C02008`. URL: `https://cds.cern.ch/record/1632384`.

[35] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Tech. rep. Geneva, 1997. URL: `https://cds.cern.ch/record/357153`.

[36] CMS Collaboration. *The CMS magnet project: Technical Design Report*. Tech. rep. Geneva, 1997. URL: `https://cds.cern.ch/record/331056`.

[37] CMS Collaboration. "Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays". In: *Journal of Instrumentation* 5.03 (Mar. 2010), T03021–T03021. ISSN: 1748-0221. DOI: `10.1088/1748-0221/5/03/t03021`. URL: `http://dx.doi.org/10.1088/1748-0221/5/03/T03021`.

[38] CMS Collaboration. *The CMS muon project: Technical Design Report*. Tech. rep. Geneva, 1997. URL: `https://cds.cern.ch/record/343814`.

[39] The CMS collaboration. "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7\text{TeV}$". In: *Journal of Instrumentation* 7.10 (Oct. 2012), P10002–P10002. ISSN: 1748-0221. DOI: `10.1088/1748-0221/7/10/p10002`. URL: `http://dx.doi.org/10.1088/1748-0221/7/10/P10002`.

[40] Mia Tosi. *The CMS trigger in Run 2*. Tech. rep. Geneva: CERN, Oct. 2017. DOI: `10.22323/1.314.0523`. URL: `https://cds.cern.ch/record/2290106`.

[41] CMS collaboration et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (2014), P10009.

[42] K. Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–2239. DOI: 10.1109/5.726788.

[43] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector." In: *JINST* 12 (June 2017). Replaced with the published version. Added the journal reference and DOI. All the figures and tables can be found at http://cms-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.

[44] CMS Collaboration. "Pileup mitigation at CMS in 13 TeV data". In: *Journal of Instrumentation* 15.09 (Sept. 2020), P09018–P09018. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/09/p09018. URL: http://dx.doi.org/10.1088/1748-0221/15/09/P09018.

[45] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7. URL: https://www.sciencedirect.com/science/article/pii/0370157383900807.

[46] M. Cacciari et al. "FastJet user manual". In: *The European Physical Journal C* 72.3 (Mar. 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.

[47] CMS Collaboration. *Jet algorithms performance in 13 TeV data*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256875.

[48] CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014–P02014. ISSN: 1748-0221. DOI: 10.1088/1748-0221/12/02/p02014. URL: http://dx.doi.org/10.1088/1748-0221/12/02/P02014.

[49] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Physics Letters B* 659.1–2 (Jan. 2008), pp. 119–126. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2007.09.077. URL: http://dx.doi.org/10.1016/j.physletb.2007.09.077.

[50] CMS Collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/05/p05011. URL: http://dx.doi.org/10.1088/1748-0221/13/05/P05011.

[51] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/p12012. URL: http://dx.doi.org/10.1088/1748-0221/15/12/P12012.

[52] CMS Collaboration. "A new calibration method for charm jet identification validated with proton-proton collision events at $\sqrt{s} = 13$ TeV". In: *Journal of Instrumentation* 17.03 (Mar. 2022), P03014. ISSN: 1748-0221. DOI: 10.1088/1748-0221/17/03/p03014. URL: http://dx.doi.org/10.1088/1748-0221/17/03/P03014.

[53] CMS Collaboration. "Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton–proton collisions at

$$\sqrt{s} = 13\,\mathrm{TeV}$$

". In: *The European Physical Journal C* 81.6 (June 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09200-x. URL: http://dx.doi.org/10.1140/epjc/s10052-021-09200-x.

[54] W. K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1 (1970), pp. 97–109. ISSN: 00063444, 14643510. URL: http://www.jstor.org/stable/2334940 (visited on 12/22/2025).

[55] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language". In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(77)90384-4. URL: https://www.sciencedirect.com/science/article/pii/0550321377903844.

[56] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: http://dx.doi.org/10.1007/JHEP07(2014)079.

[57] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. URL: https://doi.org/10.1088/1126-6708/2007/11/070.

[58] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024. URL: https://www.sciencedirect.com/science/article/pii/S0010465515000442.

[59] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *Journal of High Energy Physics* 2012.12 (Dec. 2012). ISSN: 1029-8479. DOI: 10.1007/jhep12(2012)061. URL: http://dx.doi.org/10.1007/JHEP12(2012)061.

[60] Bo Andersson. *The Lund Model*. Vol. 7. Cambridge University Press, 1998. ISBN: 978-1-009-40129-6, 978-1-009-40125-8, 978-1-009-40128-9, 978-0-521-01734-3, 978-0-521-42094-5, 978-0-511-88149-7. DOI: 10.1017/9781009401296.

[61] CMS Collaboration. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *The European Physical Journal C* 76.3 (Mar. 2016). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-3988-x. URL: http://dx.doi.org/10.1140/epjc/s10052-016-3988-x.

[62] "Geant4—a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[63] Stefan Höche. "Introduction to parton-shower event generators". In: *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*. 2015, pp. 235–295. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph].