UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"
Ph.D. course in Physics
XXXVI series

Search for heavy flavour Higgs boson decays at hadron and future Muon Collider

Thesis written with the financial contribution of INFN

Coordinator

Prof. Giulio Monaco

Supervisor

Prof. Donatella Lucchesi

Co-supervisor

Dr. Lorenzo Sestini

Candidate

Laura Buonincontri

Contents

Abstract

The goal of this thesis is to study the heavy flavour decays of the Higgs boson at LHCb, one of the experiments at the Large Hadron Collider, and at a Future Collider proposal, the Muon Collider, where muons will be collided at multi-TeV center of mass energy. At LHCb a set of Machine Learning algorithms have been used to set upper limits on the inclusive $H \to b\bar{b}$ and $H \to c\bar{c}$ production, using data from 2016 dataset at 13 TeV center of mass energy. At the Muon Collider, the performance in the heavy quarks flavour tagging in the presence of the beam-induced background have been evaluated and the uncertainties on the measurement of the double Higgs cross section and the trilinear Higgs self-coupling at 3 TeV center of mass energy have been calculated.

1 Introduction

The study of the Higgs boson physics provides a potential portal to New Physics. The precise measurement of its properties is fundamental to reveal possible deviation from the Standard Model (SM) theory of particle physics. Some of them have been measured at the Large Hadron Collider (LHC), like the Higgs mass and the main Higgs couplings to SM particles, while others, like the Higgs boson self-couplings, that determine the shape of the Higgs potential and the Electroweak Simmetry Breaking, have never been measured directly at colliders due to the low production cross section of double and triple Higgs. For this purpose, Future Collider projects are under development. The scientific program for Higgs Physics at Future Colliders foresees the improvement of the precision on the fermions and bosons coupling measurements and the measurements of Higgs boson self-couplings. So far, ATLAS and CMS experiments have performed the most precise measurements of the Higgs boson properties and their results were found in agreement with the SM.

In this thesis, a study of heavy flavour Higgs boson decays $H \to b\bar{b}$ and $H \to c\bar{c}$ at the LHCb experiment, and an evaluation of the performance on the measurement of the double Higgs production at the Muon Collider is performed. The search of the H at LHCb would allow to complement the H measurements performed by other CERN experiments like ATLAS and CMS, measuring the Higgs in a different phase space. The LHCb detector, differently from ATLAS and CMS, whose detectors are cylinders that surround the collision point, is a forward spectrometer that covers the very forward region of the pp collision. It has been designed to perform b- and c-hadron physics measurements, whose cross-sections are enhanced in the forward region of the pp collisions and offers excellent performance for heavy quarks flavour identification. Upper limits on the inclusive $H \to b\bar{b}$ and $H \to c\bar{c}$ production cross sections have been calculated in this thesis at a center-of-mass energy of 13 TeV, using data collected at the LHCb experiment during the 2016 stage of the Run 2, corresponding to an integrated luminosity of $\mathcal{L}$=1.6 fb^{-1}. These upper limits have been used to determine prospects on the future upgrades foreseen for LHC. The upper limit is set to the inclusive H production, since all possible H production processes are included the analysis. The former is the first upper limit calculation at LHCb on the inclusive $H \to b\bar{b}$, the latter is the first upper limit ever set to the inclusive $H \to c\bar{c}$ in the final state with two resolved jets.

The muon collider is a possible future machine where a large number of H and HH events could be produced by $\mu^+\mu^-$ collisions thanks to the high center of mass energies, in the regime of multi-TeV and the high luminosity conditions that are foreseen for this machine. The experimental condition are very different from the one at the LHC experiments. In this thesis the evaluation of the sensitivity on the SM production cross section measurement of the process $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$ and of the sensitivity on the determination of the Higgs boson trilinear self-coupling that can be reached at a 3 TeV center of mass energy Muon Collider is performed. This HH analysis at Muon Collider has been performed via detailed simulation of the detector effects on the events reconstruction and the effects of the background that arise from the muons decay along the beam pipe, the beam-induced background (BIB), have been evaluated. For both the analyses, similar techniques have

been employed to reconstruct jets, and analysis tagging techniques have been employed to identify the final state heavy quark flavour. The work have been structured in this thesis as follows:

1. Chapter 2: Status of the art of Higgs Physics
   In this chapter an overview of the Brout-Englert-Higgs mechanism of the symmetry breaking will be given and most recent measurements of the Higgs boson branching ratios and its couplings to bosons and heavy quarks, performed so far at LHC, will be also presented. The single, double and triple Higgs production processes at hadrons and lepton colliders will be described. An overview of the expected uncertainties on Higgs couplings at the Future Colliders will be also given. A particular focus will be given for the prospects on the Higgs measurements at Muon Collider.

2. Chapter 3: LHCb detector description
   After a brief introduction of the LHC accelerator facility, the sub-systems forming the LHCb detector and the LHCb trigger system are presented. The software framework used to simulate events and handle real data is also presented.

3. Chapter 4: Physics Objects reconstruction at LHCb
   The algorithms used for the particle identification and tracks, jets and primary and secondary vertex reconstruction algorithms at LHCb are presented together with their performances. The heavy flavour identification Neural Network algorithm used in this thesis to identify the heavy flavour final state jets is also described.

4. Chapter 5: Search for $H \to b\bar{b}$ and $H \to c\bar{c}$ al LHCb
   The data collected by LHCb during the year 2016 of the Run 2 data taking are analyzed. First, the Monte Carlo and data samples used for the analysis and the analysis strategy are presented. A data-driven method to model the multi-jet QCD heavy flavor di-jets background, that is the major source of background in the analysis is presented. A fit is performed on events of di-jet to determine the $H \to b\bar{b}$ and $H \to c\bar{c}$ yields and the multi-jet QCD background shape. Expected upper limits have been calculated by using the CLs method.

5. Chapter 6: Upper limits on the H $\to b\bar{b}$ and H $\to c\bar{c}$ search and future prospects
   The systematic uncertainties on the upper limit calculation have been studied in order to determine which impacts the most the results. The final observed limit are determined. In the conclusive part of the chapter, the results are discussed and scaled to calculate the prospects at the HL-LHC on $H \to b\bar{b}$ and $H \to c\bar{c}$ measurements at LHCb.

6. Chapter 7: The Muon Collider machine and detector
   The muon accelerator scheme developed by the Muon Accelerator Program (MAP) will be presented, starting from the muon production up to the acceleration section. The main technical issues regarding the muon collider and the detector are also described. In particular, the beam-induced background composition and mitigation strategies used in the HH analysis will be discussed. In this chapter the simulated detector provided by the Muon Collider framework used to reconstruct HH events is also described.

7. Chapter 8: Event reconstruction at Muon Collider
   In this chapter the algorithms used for the particle identification and tracks and jets reconstruction, implemented in the software framework, will be presented. Their performances have been evaluated in the presence of the BIB, and optimization requirements to reduce the BIB effects are discussed.

8. Chapter 9: Double Higgs production and Higgs self-coupling at muon collider
   In this chapter the methods used for the generation, simulation and reconstruction of the

process $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$ and the physics background will be discussed. Both the techniques to calculate the uncertainty on the cross section and the trilinear coupling exploit Machine learning techniques to separate signal from background. The strategy followed to perform both measurements is presented. A particular focus will be given in the study of the how the Beam Induced Background affects the measurements and how these effects can be limited.

2 Higgs boson physics

In this chapter an overview of the state of the art of the Higgs physics measurements performed so far at LHC and the prospects on Higgs physics at Future Colliders is given. The chapter is organized as follows:

- section 2.1 an overview of the Brout- mechanism of the symmetry breaking is given;
- then, the Higgs production mechanism at hadron and lepton colliders and its decay channels are described (sec. 2.2);
- the state of the art on Higgs physics measurements performed at LHC are shown in section 2.3;
- at last, the Future Colliders proposal are discussed together with their goal on Higgs physics measurement (sec. 2.4). The main advantages and challenges in the development of a future Muon Collider are also described.

2.1 Status of the art of Higgs Physics

The Standard Model provides the mathematical framework which gives a quantum-mechanical description of electromagnetic, weak and strong interactions. So far, there are seventeen fundamental particles in the Standard Model: twelve are the fundamental fermions, leptons and quarks. They interact via boson particles, that are force carriers: photons are exchanged in electromagnetic interactions, W^+, W^- and Z bosons are exchanged in weak interactions, and gluons are force carrier particles of the strong interaction.

The Standard Model is based on the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$, with the color group $SU(3)_C$ for the strong interaction and with $SU(2)_L \times U(1)_Y$ for the electroweak interaction spontaneously broken by the Higgs mechanism. The Higgs field is introduced in the electroweak theory [1] to break the gauge invariance spontaneously and to explain how fermions and bosons to acquire mass.

The Higgs field is a spin-zero field, a doublet under $SU(2)$, it has $Y = \frac{1}{2}$ hypercharge, and it is a singlet in color space.The dynamics of the Higgs field is described by the Lagrangian density $\mathcal{L}_{Higgs}$:

$$\mathcal{L}_{Higgs,bosons} = (D^\mu \Phi)^\dagger D_\mu \Phi - V(\Phi) \tag{2.1}$$

D_μ is the covariant derivative, defined as:

$$D_\mu = \partial_\mu + \frac{i}{2} g \tau_j W_\mu^j + \frac{i}{2} g' B_\mu$$

where $j = 1, 2, 3$, τ_j are Pauli matrices, W_μ^j and B_μ are the gauge fields, g is the weak coupling constant and g' is the coupling constant related to $U_Y(1)$ gauge symmetry. Φ is the scalar doublet of the Higgs field, defined as:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{2.2}$$

where ϕ_1, ϕ_2, ϕ_3, ϕ_4 are real scalar fields.

$V(\Phi)$ is the Higgs potential

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{2.3}$$

λ and μ are real constants related to the shape of the Higgs potential. $\lambda > 0$ is required to ensure the potential is bounded from below, guaranteeing the presence of a ground state. Since the Higgs

potential presents a continuum of infinite configurations of minimum energy, $\mu^2 < 0$ is also required. The choice of a particular minimum of the Higgs potential as ground state leads to spontaneous symmetry breaking. The expression for the Higgs field expansion around the vacuum (the chosen minimum), in the unitary gauge is:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.4}$$

where $v = \sqrt{\frac{-\mu^2}{\lambda}}$ is the vacuum expectation value and $h(x)$ is the real scalar field of the Higgs boson. By choosing this ground state the $SU(2)_L \times U(1)_Y$ symmetry gets spontaneously broken to the electromagnetic symmetry $U(1)_{QED}$. 2.1, when expression 2.4 is introduced, becomes:

$$\mathcal{L}_{Higgs} = \frac{1}{2}\partial_\mu h \partial^\mu h + (v+h)^2 \left[\frac{g^2}{4} W_\mu^\dagger W^\mu + \frac{1}{8}(g^2 + g'^2) Z_\mu^\dagger Z_\mu \right] - \lambda v^2 h^2 - \lambda v h^3 - \frac{\lambda}{4} h^4 \tag{2.5}$$

where W_μ^+, W_μ^-, Z_μ are the gauge fields are expressed in terms of the physical fields as:

$$\begin{aligned} W_\mu^+ &= \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2) \\ W_\mu^- &= \frac{1}{\sqrt{2}}(W_\mu^3 + iW_\mu^4) \\ Z_\mu &= \cos(\theta_W) W_\mu^3 - \sin(\theta_W) B_\mu \\ A_\mu &= \sin(\theta_W) W_\mu^3 + \cos(\theta_W) B_\mu \end{aligned} \tag{2.6}$$

$\sin(\theta_W)$ is the Weinberg angle. From equation 2.5 the expressions of the mass and of the Higgs boson and the Z and $W^\pm$ bosons can be found:

$$m_H = \sqrt{2\lambda v^2} \tag{2.7}$$

is the Higgs boson mass (the coefficient of the h^2 term);

$$\begin{aligned} m_{W+} = m_{W-} &= \frac{gv}{2} \\ m_Z = \frac{2\sqrt{g^2 + g'^2}}{2} &= \frac{m_W}{cos(\theta_W)} \end{aligned} \tag{2.8}$$

are the second term coefficient, and are the Z and $W^\pm$ bosons masses.

The coefficients of the h^3 and h^4 terms give the trilinear and quadrilinear Higgs self-couplings.

$$\begin{aligned} \lambda v h^3 &= \frac{m_H^2}{2v^2} v h^3 = \lambda_3 v h^3 \\ \frac{\lambda}{4} h^4 &= \frac{1}{4}\frac{m_H^2}{2v^2} h^4 = \frac{1}{4}\lambda_4 h^4 \end{aligned} \tag{2.9}$$

In the SM notation λ_3 and λ_4 are equal and expressed as:

$$.\lambda_3 = \lambda_4 = \frac{m_H^2}{2v^2} \tag{2.10}$$

Within the Standard Model the Higgs potential is then fully characterized by two parameters, m_h and v, that are free parameters of the theory. These parameters have been determined experimentally:

- the vacuum expectation value v is [2]

$$v = 1/\sqrt{\sqrt{2}G_F} \sim 246GeV \tag{2.11}$$

  and has been determined from the measurement of the Fermi constant, that is extremely well measured through the muon lifetime (0.6 ppm);

- The most updated measurement of the Higgs boson mass is $m_H = 125.11 \pm 0.11$ GeV [3]. This corresponds to a 0.09% precision and has been obtained by ATLAS.

Trilinear and quadrilinear Higgs self-coupling given by equation 2.10 are then:

$$\lambda_3 = \lambda_4 \sim 0.13$$

New physics can modify the Higgs potential giving different values of λ_3 and λ_4 [4]. Other terms of $\mathcal{L}_{Higgs}$ describe the couplings between the single and double Higgs with vector bosons.

Yukawa couplings introduce the mass terms for quarks and charged leptons after the spontaneous symmetry breaking of the Higgs field. For leptons, the Yukawa lagrangian is:

$$\mathcal{L}_{Yukawa} = -g_k \bar{l}_L^k \Phi e_R^k - Y_{ij}^u \bar{q}_L^i \tilde{\Phi} u_R^j - Y_{ij}^d \bar{q}_L^i \Phi d_R^j + h.c. \tag{2.12}$$

where g_k are the coupling constants to the flavour k and Y_{ij}^u and Y_{ij}^d are the matrices that contain the Yukawa coupling constants to quarks. The index k in the first term runs on the three flavours of leptons, while i, j are the quarks flavour. L and R refer, respectively to the left and right chiral projections of the spinors:

$$\begin{aligned} \psi_L &= \frac{1+\gamma^5}{2}\psi \\ \psi_R &= \frac{1-\gamma^5}{2}\psi \end{aligned} \tag{2.13}$$

$$l_L^i = \begin{pmatrix} e_L \\ \nu_e, L \end{pmatrix}, \begin{pmatrix} \mu_L \\ \nu_\mu, L \end{pmatrix}, \begin{pmatrix} \tau_L \\ \nu_\tau, L \end{pmatrix} \tag{2.14}$$

$$l_R^i = e_R, \mu_R, \tau_R \tag{2.15}$$

$$q_L^i = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \tag{2.16}$$

$$\begin{aligned} u_R^i &= u_R, c_R, t_R \\ d_R^i &= d_R, s_R, b_R \end{aligned} \tag{2.17}$$

By introducing in the Lagrangian of equation 2.12, the Higgs field of equation 2.4 the lepton masses are:

$$m^l = \frac{g^l v}{2} \tag{2.18}$$

In order to obtain the quark masses it is necessary to diagonalize the Y^u and Y^d matrices. A unitary transformation V allows to write:

$$M^u = \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} = \frac{v}{\sqrt{2}} V_L^u Y^u V_R^u$$

$$M^d = \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix} = \frac{v}{\sqrt{2}} V_L^d Y^d V_R^d$$

2.2 Higgs production cross section and decay channels at colliders

2.2.1 Higgs production mechanism at hadron collider

The theoretical calculation of single Higgs production cross section at hadron colliders for the different Higgs production processes as a function of the collision center of mass energy ($\sqrt{s}$) is shown in Figure 1. As can be seen, the cross section of Higgs production grows with the increasing of the center of mass energy up to order of magnitude.

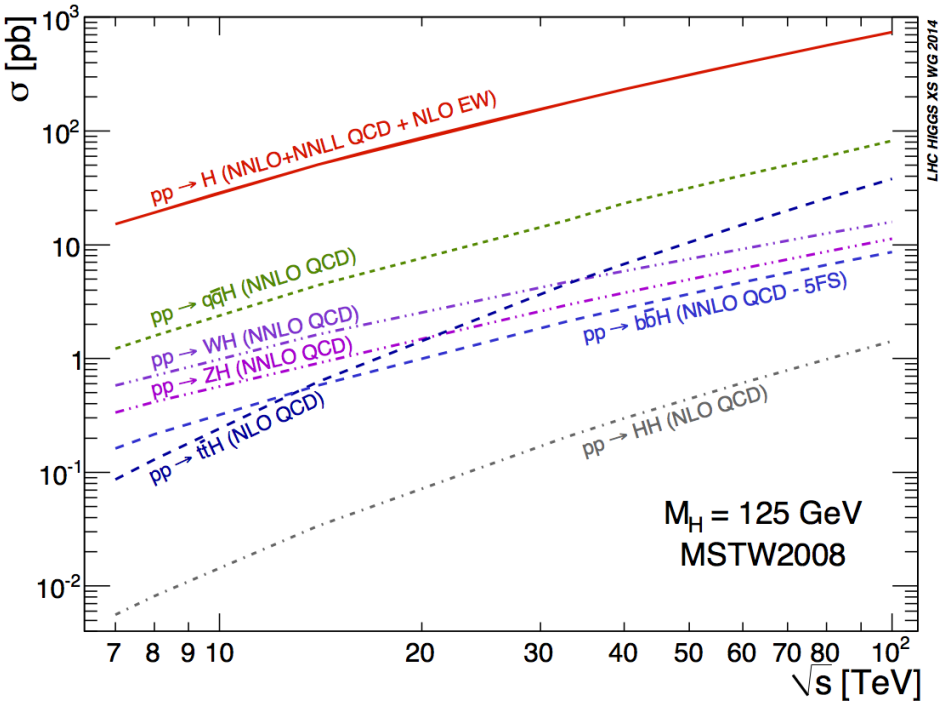

Figure 1: Higgs boson production cross section in a pp collider as a function of the center of mass energy up to 100 TeV[5]

The Higgs boson can be produced by pp collisions through five main processes, whose Feynman diagrams at leading order are shown in Figures 2 and 3.

- Gluon-gluon fusion ($gg \to H$, Figure 2 left), as can be seen in Figure 1 is the dominant. Higgs from ggF is produced by a loop-induced coupling, mediated by a massive fermion (the top is the main contribution) since no direct gluon-Higgs coupling exists within the Standard Model;
- Vector Boson Fusion (VBF, Figure 2 right) $qq \to VV \to H$, is the second-largest production mechanism, and it is dominated by t-channel W and Z exchange. The Higgs is produced from a $W^+ W^-$ or ZZ pair originated by initial state quarks with a change of flavour in the case of W boson;

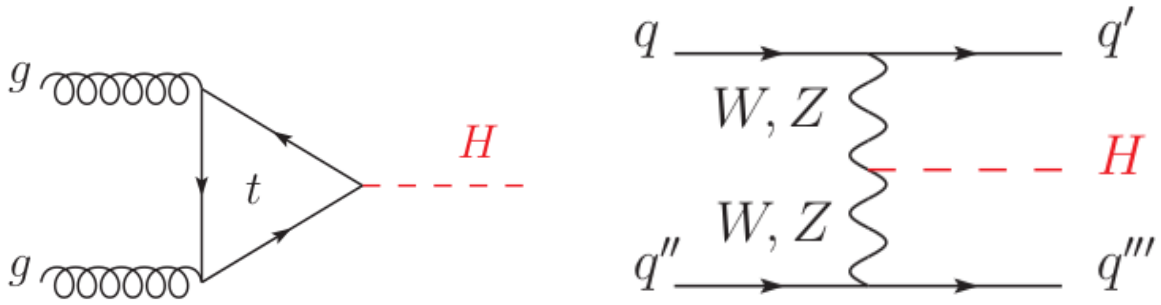

Figure 2: Feynman diagrams of leading Higgs production processes in proton-proton collisions: gluon-gluon fusion (left), vector boson fusion (right) [2]

- Associated production or Higgs-strahlung ($qq \to Z/WH$,Figure 3 left). Here a off-shell gauge boson produced at tree level from a quark-quark interaction, generates a W/Z and Higgs;
- Associated $t\bar{t}H$ higgs production ($gg \to t\bar{t}H$, Figure 3 right), where two gluons split in t$\bar{t}$ pair.

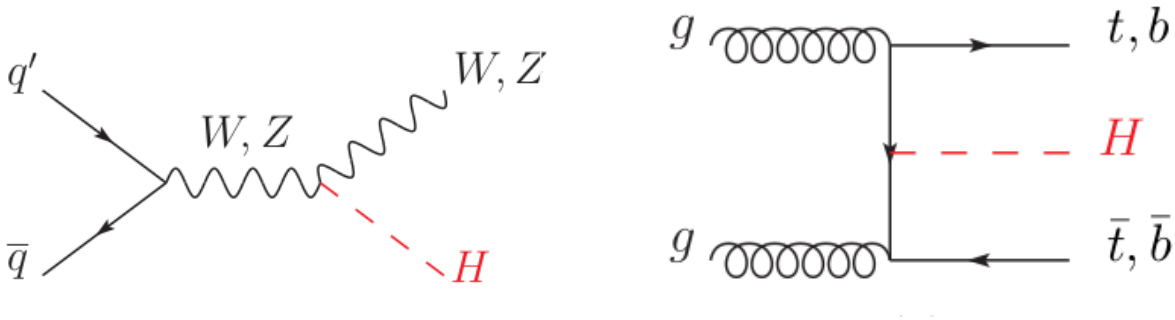

Figure 3: Feynman diagrams of leading Higgs production processes in proton-proton collisions: Higgs-strahlung (left) associated top production (right) [2]

2.2.2 Double Higgs production at hadron collider

Information on trilinear and quadrilinear couplings are experimentally accessible through double and triple Higgs production processes. At pp colliders, Higgs pairs production processes are similar to the single Higgs, and main representative diagrams can be seen in Figure 4. Processes containing the triple Higgs coupling are highlighted with a red circle. Destructive interference exists between the box and the triangle diagrams (Figure 4 up left and Figure 4 up right respectively). As can be seen in Figure 5, where the cross sections of these single processes are calculated as a function of the $\sqrt{s}$, the gluon-gluon fusion remains the dominant process over the entire range of different center of mass energies, also in Higgs pair production.

2.2.3 Higgs production mechanism at lepton colliders

At lepton colliders, the main Higgs production mechanism depends on the center of mass energy. There are two main proposal for lepton colliders, whose differences will be discussed more in detail in section 2.4, e^+e^- colliders and $\mu^+\mu^-$ colliders.

Figure 6 shows the cross sections for unpolarized muon beams for single Higgs production at center of mass energies up to 50 TeV.

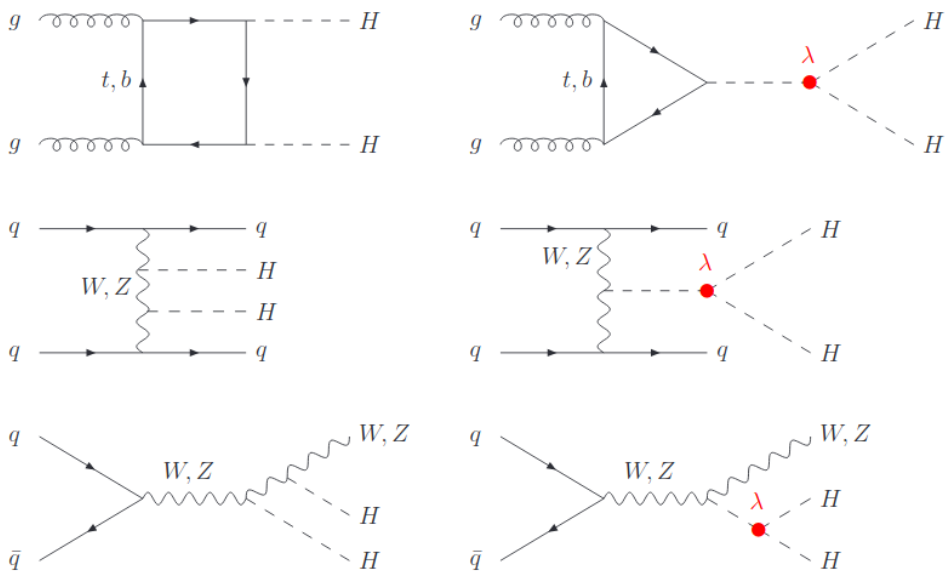

Figure 4: Representative diagrams for double Higgs production at a pp collider. Those containing the triple Higgs vertex are highlighted with red circle [6].

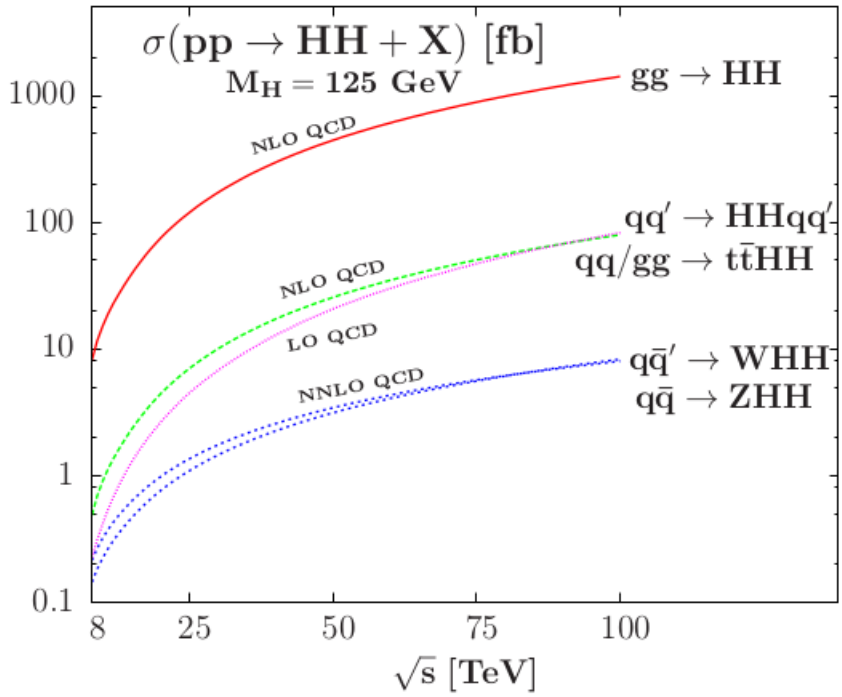

Figure 5: Cross section of double Higgs production at a pp collider as a function of the center of mass energy [6].

- at center of mass energy $\sqrt{s} < 1$ TeV the higgs-strahlung (ZH) process in which the two incoming muons annihilate and the Higgs is emitted by an off-shell Z dominates (the Feynman diagram is shown in Figure 7 left);
- at energies well above 1 TeV, the vector boson fusion (VBF) production process becomes the most relevant (the Feynman diagram is shown in Figure 7 right), where the Higgs is produced

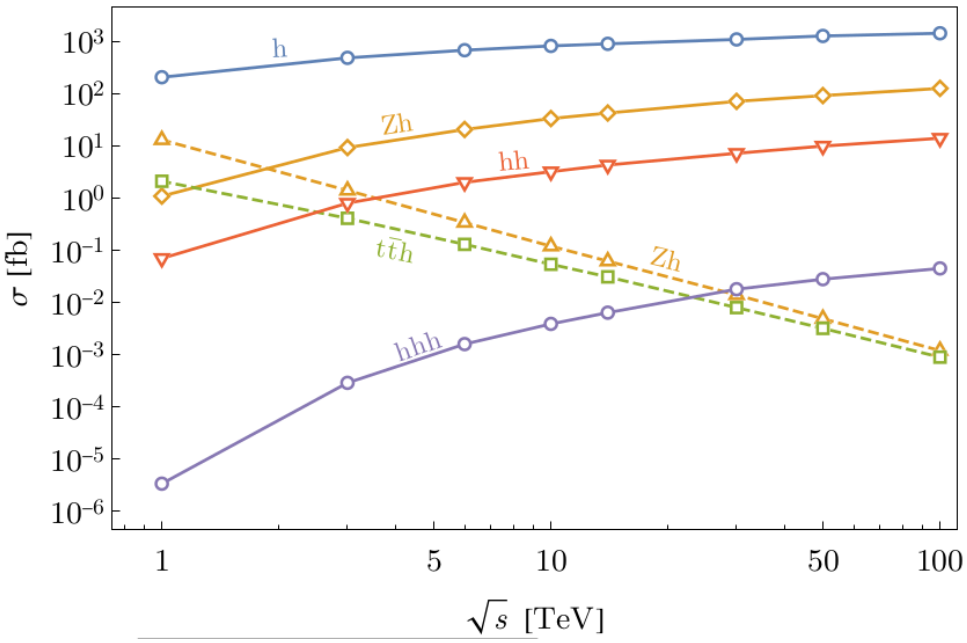

Figure 6: Single Higgs boson production cross section in a $\mu^+\mu^-$ collider as a function of the center of mass energy up to 100 TeV[7]. Dashed lines represents VBF channels, while continuous lines represents ZH channels from Higgs-strahlung and ttH.

by the fusion of two vector bosons. The WW fusion (that is the diagram in the Figure) is the dominant one, while the ZZ fusion mechanism, $l^+l^- \to Z^*Z^*l^+l^- \to Hl^+l^-$, also contributes to the Higgs boson production, with a cross-section suppressed by an order of magnitude with respect to that of WW fusion.

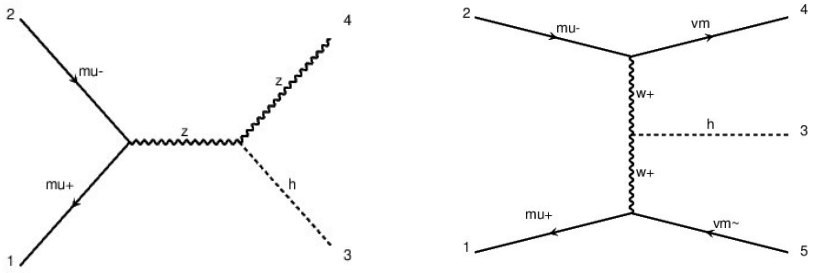

Figure 7: Representative diagrams for single Higgs production at a $\mu^+\mu^-$ colliders.

2.2.4 Double Higgs production at lepton colliders

At lepton colliders, depending on the center of mass energy, the double Higgs can be produced via Higgs-strahlung $l^+l^- \to ZHH$, particularly relevant at low energies (see Feynman dyagrams in Figure 8 up), or via vector boson fusion $l^+l^- \to HH\nu_l\bar{\nu}_l$ (see Feynman dyagrams in Figure 8 low). Only diagrams with the red circle contains the triple higgs coupling. At 3 TeV, that is the center of mass energy muon collider considered in the analysis performed in this thesis, the HH is produced mainly via VBF, while the ZHH is more than one order of magnitude smaller [8]. The HH production cross section via VBF is shown in red in Figure 6 as a function of the center

of mass energy. The double Higgs production mechanism will be studied in this thesis at $\mu^+\mu^-$ colliders.

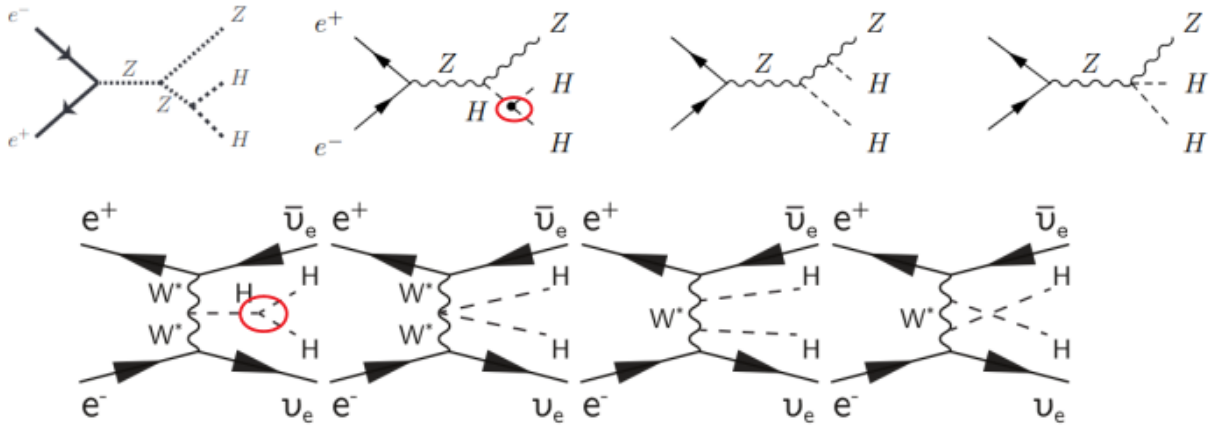

Figure 8: Representative Feynman diagrams of double Higgs production at an electron-positron collider [8].

2.2.5 Higgs boson decays

Figure 9 shows the theoretical variation of the branching ratio as a function of the Higgs mass, in a range between ± 5 GeV around the Higgs boson mass. The width of each curve represents the theoretical error. As seen in 2.1 the HWW and HZZ vertices are proportional to m_W^2 and m_Z^2 while the coupling of the Higgs boson to fermions particles grows linearly with their masses. Therefore, the Higgs boson is favored to decay into the heaviest kinematically accessible particle, which is the b quark and the decay is into $b\bar{b}$ pair. On the other hand, the couplings to electrons, muons and light quarks (u,d,s) are extremely small. The Higgs boson decay modes and the relative branching ratio are summarized in Table 1.

Decay mode	Branching fraction	Rel. uncertainty [%]
$H \to bb$	5.82×10^{-1}	$^{+1.2\%}_{-1.3\%}$
$H \to WW$	2.14×10^{-1}	$\pm 1.5\%$
$H \to \tau\tau$	6.27×10^{-2}	$\pm 1.6\%$
$H \to cc$	2.89×10^{-2}	$^{+5.5}_{-2.0}$
$H \to ZZ$	2.62×10^{-2}	$\pm 1.5\%$
$H \to \gamma\gamma$	2.27×10^{-3}	$\pm 2.1\%$
$H \to Z\gamma$	1.53×10^{-3}	$\pm 5.8\%$
$H \to \mu\mu$	2.18×10^{-4}	$\pm 1.7\%$

Table 1: Branching fraction for different Higgs decay modes [2]

2.3 Higgs properties measurements at LHC

Since 2010 to 2012, LHC [9] produced pp collisions up to 7- 8 TeV centre-of-mass energy for LHCb, ATLAS [10] and CMS [11] experiments, and increased the center of mass energy to 13 TeV

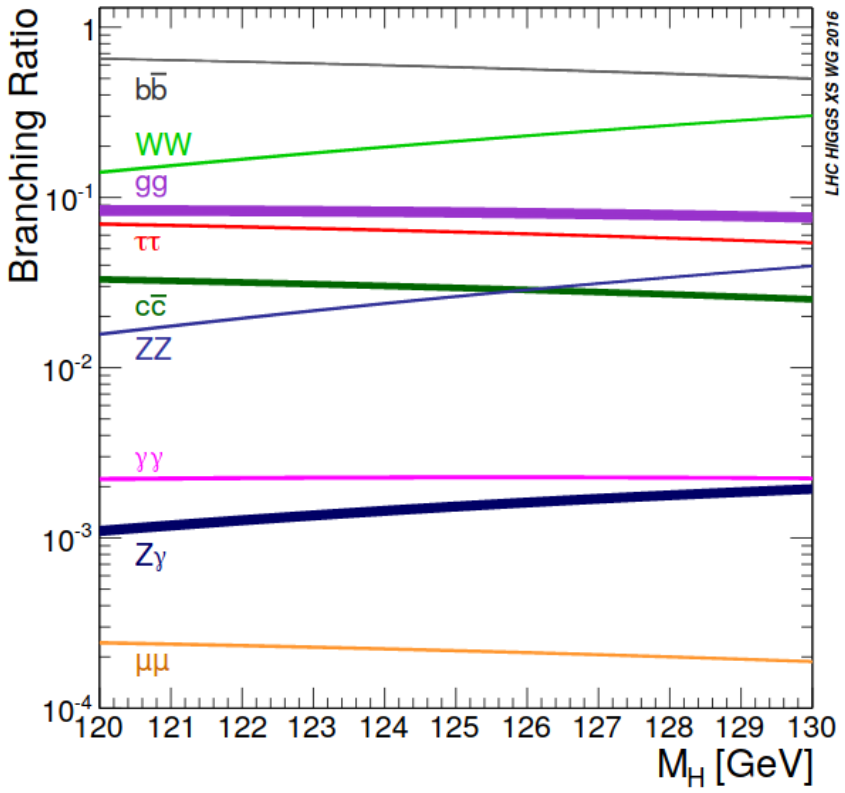

Figure 9: Branching ratios for different Higgs decay modes as a function of the Higgs mass [2]

for the second campaign of data acquisition (2015-2018). Currently it is delivering 13.6 TeV center of mass energy. The upper limits calculations performed in this thesis are based on LHCb data acquired at 13 TeV center of mass energy in 2016. For the upgrade of LHC, the High Luminosity LHC (HL-LHC), pp collisions at 14 TeV are foreseen. The signal strength modifier μ is defined as the ratio between a measured Higgs boson process and its SM expectation value. For each specific process $i \to H \to f$ the μ_i^f is defined as:

$$\mu_i^f = \frac{\sigma_i BR^f}{\sigma_{SM} BR_{SM}}$$

Here σ^i with i=ggH, Vector Boson Fusion (VBF), WH, ZH, ttH and BR^f with f=ZZ, WW,$\gamma\gamma$,$\tau\tau$, bb, $\mu\mu$ are respectively, the production cross section for $i \to H$ and the branching fraction for $H \to f$. Expressions for the production μ_i and the decay μ_f can be separated into:

$$\mu_i = \frac{\sigma^i}{\sigma_{SM}}$$

$$\mu_f = \frac{BR^f}{BR_{SM}}$$

The summary of the production cross sections measurements normalised to the SM expectations, for different Higgs boson final states and production mechanisms are shown in Figure 10. The signal strength measurements μ_i^f at the ATLAS and CMS are shown in yellow and brown respectively. The measurements have been combined separately for ATLAS and CMS, and the combination results for μ_f are shown along the vertical axis, while the combination results for μ_i are shown along the horizontal axis. The Figure is taken from the August 2023 revision of [2] and summarizes the results

obtained by ATLAS and CMS in [12] and [13] with the full Run 2 dataset (integrated luminosity of 139fb^{-1}). In general, the precision in the measurement of a channel depends on the size of the

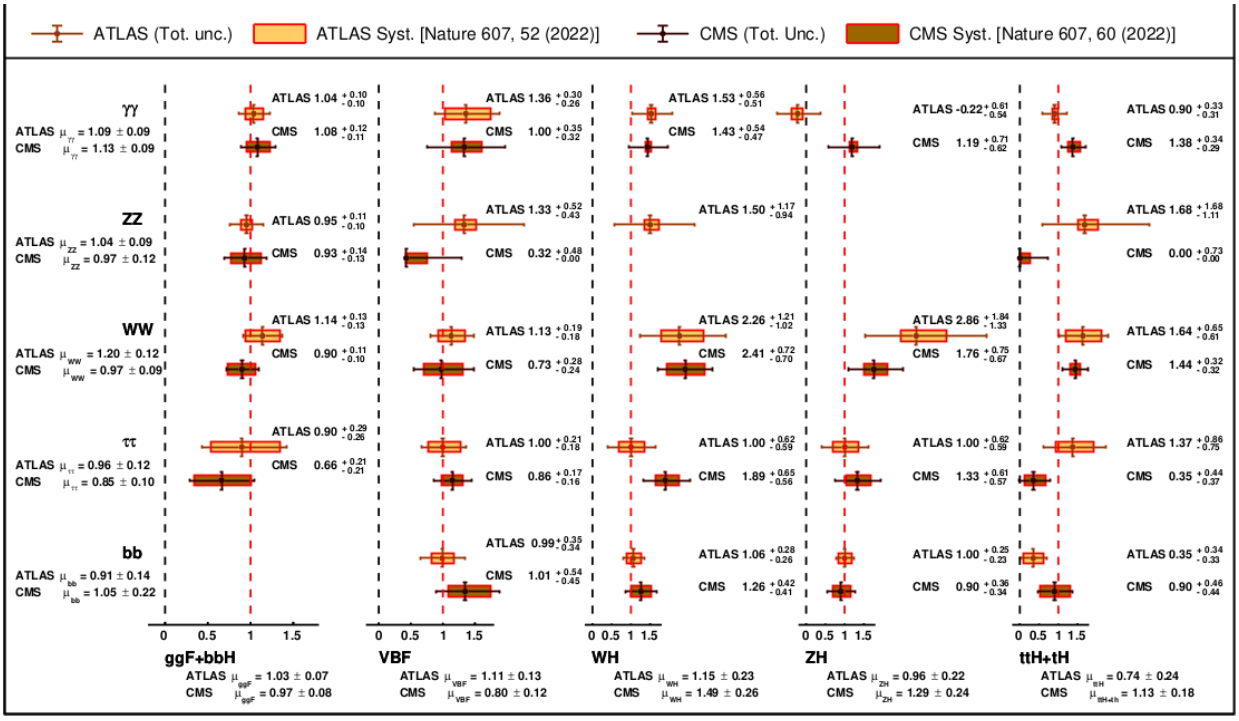

Figure 10: Higgs production signal strength measurements at the ATLAS (yellow) and CMS (brown) experiments, for the five Higgs boson main decay modes and the five main Higgs boson production modes [2].

production cross section of the Higgs boson, its decay branching fraction, the reconstructed mass resolution, the efficiency in the reconstruction and the identification of the final state particles and the level of background in the final state. For example, channels like the $H \to \gamma\gamma$ and $H \to ZZ$, where Z decay into electrons or muons allows for better resolution in the invariant mass of the H ($\sim 1-2\%$) with respect to channels with neutrinos in the final state, like $H \to WW$, where the W decays into leptons and neutrinos, even if this latter channel can benefit from a higher branching ratio. Indeed, since neutrinos escape the detector, a fraction of energy is not measured. $H \to b\bar{b}$ and $H \to \tau\tau$ suffer from large backgrounds due to QCD. The large integrated luminosity available with the Run 2 allowed to improve the measurements on the Higgs couplings: while the $H \to \gamma\gamma$, the $H \to ZZ$, the $H \to WW$ by both ATLAS and CMS already in Run 1, the first clear evidence of direct decays of the $H \to \tau\tau$ was achieved by combining ATLAS and CMS Run 1 results, leading to an excess with a significance of 5.5 σ [14]. ATLAS and CMS analyzed an subsample of the Run 2 dataset to provide independent observations, with significances of 4.4σ [15] and 4.9σ [16] respectively. With the Run 2 dataset CMS [17] reported the first evidence for $H \to \mu\mu$ at 3σ level with a signal strength of

$$\mu = 1.19 \pm 0.40(stat.) \pm 0.15(syst.)$$

, that provide the first direct evidence of the Higgs coupling to second generation fermions. In the next section, a focus on the status of the art on the $H \to b\bar{b}$ and $H \to c\bar{c}$, that are the channel studied in this thesis at LHCb, is done. By now no evidence of deviations from the standard model prediction are found for what concern the Higgs production and branching ratio measurements.

2.3.1 Measurement of the $H \to b\bar{b}$ and $H \to c\bar{c}$

One of the major sources of background for the $H \to b\bar{b}$ and $H \to c\bar{c}$ search is the QCD background that consists of hadronic jets, produced by the interactions between the proton constituents. The most sensitive production mode at ATLAS and CMS to measure the Higgs decay into $b\bar{b}$ pair of b-quarks are the associated WH and ZH processes. The leptonic $W \to l\bar{\nu}_l$, where l=e,μ or τ and Z decays into e^+e^- or $\mu^+\mu^-$ allows to select the events while rejecting QCD background. The direct observation of the Higgs decay into b quarks pair was obtained by ATLAS and CMS independently, observing significant excess corresponding to 4.9 σ [18] and 4.8 σ [19] respectively, by using the Run 2 data only. The observed significance was found to be 5.4 σ for ATLAS and 5.6 σ for CMS when combined to the Run 1 dataset. ATLAS [20] and CMS [21] have also searched for the $H \to b\bar{b}$ in VBF production mode, where the dominant background is the QCD production of multi-jet but still with a limited significance, 2.6 σ and 2.4 σ respectively. However, a 3 σ significance of the Higgs boson decaying to a pair of b quarks in the VBF production mode has been reached by ATLAS [22] by combining the VBF results with the one where the $H \to b\bar{b}$ via VBF is searched in association with a high-energy photon.

The sensitivity in the inclusive search for the Higgs boson in the ggF production mode with $H \to b\bar{b}$ is limited by the large background of $pp \to b\bar{b}+X$. Advanced deep learning techniques have been used to improve the performance in the identification of high transverse momentum $H \to b\bar{b}$, reconstructed as a unique large radius jet.

The measurement of the Higgs coupling to charm quark will be one of the next milestones for Higgs physics, since it would be the first measurement of the second generation quarks. The observation of the $H \to c\bar{c}$ decay is challenging in the LHC environment, given the low $H \to c\bar{c}$ branching ratio, the large amount of multi-jet QCD background and the difficulties in the identification of jets produced by c-quarks. ATLAS and CMS have performed the direct search of the $H \to c\bar{c}$ in association with a vector boson (W or Z), respectively, obtaining observed upper limits of 26 and 14 times the SM cross section at 95% Confidence Level (CL) [23, 24]. A key aspect in this analysis is the ability to develop flavour tagging algorithms to discriminate the c quark from both b quarks and light quarks. Furthermore, techniques for the identification of pairs of charm jets from Higgs boson decay produced at high transverse momentum analogue to the b quark studies have been applied. For example, CMS searched for the inclusive $H \to c\bar{c}$ in the boosted topology where the two jets from the c-quarks fragmentation are merged in a single jet, obtaining an observed upper limit of 47 times the SM predictions at 95% CL [25].

Investigations in the forward region can be made by the LHCb experiment, which is fully instrumented in the range $2 < \eta < 5$, that is a complementary phase space region with respect to ATLAS and CMS. It has been suggested [26] that physics beyond the SM could enhance the $BR(H \to c\bar{c})$ with respect to other decay modes. Moreover, possible enhancement of the Higgs boson forward production is not excluded. LHCb has demonstrated excellent flavour tagging capabilities for jets originating from beauty (b) or charm (c) quarks [27], that can play an important role in the search for a Higgs boson decaying to a $b\bar{b}$ or $c\bar{c}$ pair. These techniques have been used to measure the differential cross sections for $b\bar{b}$- and $c\bar{c}$-dijets production [28], that represent the irreducible backgrounds in the search for $H \to c\bar{c}$. LHCb has also demonstrated its capabilities in measuring a dijet resonance, by extracting the $Z \to b\bar{b}$ signal from the overwhelming multi-jet QCD background. The search for VH production with subsequent decay of the Higgs boson to a pair of b-quarks has been performed even by the LHCb collaboration [29] with 1.98 fb^{-1} of data taken at a center of mass energy of 8 TeV, by requiring two reconstructed b quarks and one lepton in the LHCb acceptance. The upper limit on the cross section is found to be 84 times the SM. In the same study, LHCb has reported its first search for $VH \to c\bar{c}$ production setting a 95% C.L. upper limit of 6.4×10^3 on the VH production when the Higgs boson in a pair of c quarks, while the expected upper limit is 7.9×10^3 times the SM.

This thesis aims to search for inclusive $H \to b\bar{b}$ and $H \to c\bar{c}$ production in the dijet final state, using data collected in the 2016 and corresponding to 1.6 fb^{-1} of integrated luminosity. Although the $H \to b\bar{b}$ process has been already discovered by ATLAS and CMS it can be used as reference channel for the $H \to c\bar{c}$ search. Moreover, in the future upgrades of LHCb, it would be interesting to measure for the first time the $H \to b\bar{b}$ production in the forward region.

This search is different from what previously done by LHCb, since it is performed in the inclusive $H \to b\bar{b}$ and $H \to c\bar{c}$ channels, while the previous one was performed with the Vector Boson associated production. Although the first Run 1 result obtained by LHCb on $H \to c\bar{c}$ was far from the current ATLAS and CMS reach, several improvements are possible. Trivially, using the Run 2 dataset with a larger integrated luminosity and with the higher Higgs production cross section at $\sqrt{s}$ =13 TeV with respect to $\sqrt{s}$ =7,8 TeV will result in a sensitivity improvement. Moreover advanced reconstruction and identification techniques could be employed, as will be discussed in 4.10. The two production channels, inclusive and VH, can be combined in the future, leading to a general improvement in the search. Moreover the inclusive search does not make any assumption on the Higgs production mechanism, therefore it can be compared with any BSM models. This would also be the first limit on the inclusive $H \to c\bar{c}$ in the final state with two resolved jets.

2.3.2 Higgs couplings measurements

The current measurements of the Higgs boson properties are limited to the couplings of the Higgs boson to the SM particles except the neutrinos. The *kappa* framework is used to characterize the Higgs coupling properties in terms of a series of Higgs coupling strength modifier parameters which are defined as the ratio of the couplings of the Higgs boson to a given elementary particle to their corresponding Standard Model values. Within the *kappa* framework the cross section times the branching ratio can be decomposed as:

$$(\sigma \cdot BR)(i \to H \to f) = \frac{\sigma_i \Gamma_f}{\Gamma_H} \tag{2.19}$$

where σ_i is the production cross section through the initial state i, Γ_f is the partial decay width into the final state f and Γ_H is the total width if the Higgs boson.

The κ parameters are introduced to express each component of eq. 2.19 as the Standard Model expectation values multiplied by the square of a coupling strength modifier. In this way the rate relative to the SM expectation μ_i^f becomes:

$$\mu_i^f = \frac{(\sigma \cdot BR)(i \to H \to f)}{\sigma_{SM} BR_{SM}} = \frac{\kappa_i^2 \kappa_f^2}{\kappa_H^2} \tag{2.20}$$

where κ_H^2 is the expression that adjust the SM Higgs width to take into account modifications, κ_i, of the SM Higgs coupling strength:

$$\kappa_H^2 = \sum_j \frac{\kappa_j^2 \Gamma_j^{SM}}{\Gamma_H^{SM}} \tag{2.21}$$

This notation is introduced to parameterize possible deviations from the SM prediction where κ_i are equal to one.

Current results on measurements of the the Higgs boson couplings to gauge bosons, fermions and self-couplings are summarized in Figure 11 where measurements of the Higgs couplings at CMS performed during the Run 2 are compared to SM predictions [13]. On the y axis the couplings to fermions and weak vector bosons, defined as $\kappa_f \cdot \frac{m_f}{v}$, $\kappa_V \cdot \frac{\sqrt{m_V}}{v}$, are shown as a function of the particle mass. The dashed line indicates the predicted dependence of the couplings on the particle mass in the case of the SM. The agreement between the measurements and the SM predictions

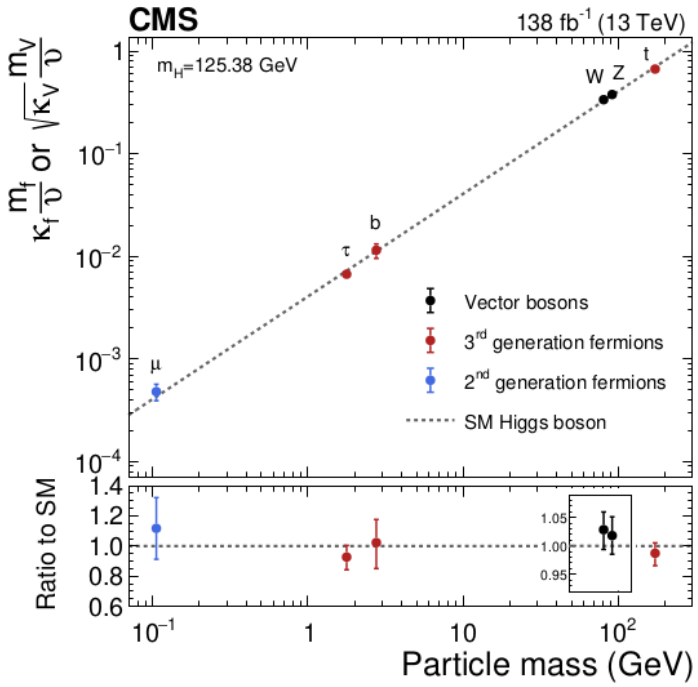

Figure 11: Combined results of the measurements of the Higgs couplings at CMS (Run II), as a function of the particle masses. Dashed line represents the Standard Model prediction [13].

over three orders of magnitude of mass is a powerful test of the Standard Model theory. For what concerns the Higgs self-couplings, as can be seen from Figure 1, the cross section is predicted to be about three orders of magnitude smaller than the single Higgs production at the LHC at 13 TeV (and even at 14 TeV HL-LHC, as will be shown in the next chapter 2.4), which makes these measurements very challenging. For the measurements of these couplings and the improvement of the uncertainty on the Higgs couplings to standard model particles, a Higgs Factory, where high single, double and triple Higgs statistics is delivered, is mandatory.

2.4 Higgs boson physics at Future Colliders

A precise determination of the Higgs boson's properties is one of the priorities at Future Colliders: for what concerns the Higgs couplings, the goal is to push the precision to around 1%. Indeed, a deviation on the Higgs couplings, compared to the SM prediction, may indirectly imply the existence of new physics. The parametrization of possible new physics is done in the SMEFT [30] approach by adding to the Standard Model Lagrangian new terms that depends on the new physics energy scale Λ. In this framework, the couplings deviations Δg, due to a new interaction characterized by the scale of energy Λ can be related to new physics scale with the following relation [31]:

$$g = g_{SM}[1 + \Delta g] : \Delta g = \mathcal{O}(v^2/\Lambda^2) \tag{2.22}$$

where v is the vacuum expectation value, g_{SM} is the Standard Model coupling. Experimental accuracies of $\Delta = 0.2$ down to 0.01 will give us sensitivity to scales of order $\Lambda \sim 550$ GeV up to 2.5 TeV. It is then clear that percent level precision the Higgs couplings will be sensitive to the new physics at multi-TeV regime. The LHC high luminosity phase (HL-LHC) will improve the accuracy of the measurements of the Higgs-boson couplings to the vector bosons and third generation, and access for the first time the couplings to the second generation fermions. Nine coupling modifiers corresponding to the Higgs boson couplings to the standard model particles

$(\kappa_W, \kappa_Z, \kappa_t, \kappa_b, \kappa_\tau, \kappa_\mu, \kappa_\gamma, \kappa_{Z\gamma}, \kappa_g)$ are presented in Figure 12, and are compared for ATLAS and CMS with the LHC data combined for the Run 1 (first column), separately for ATLAS and CMS for the Run 2 (second and third column) and the expected precision at HL-LHC. The results above are obtained under the assumption that the Higgs boson decays only to SM particles.

	LHC Run 1	ATLAS Run 2	CMS Run 2	HL-LHC (expected)
κ_γ	$0.87^{+0.14}_{-0.09}$	1.01 ± 0.06	1.10 ± 0.08	1.8%
κ_W	$0.87^{+0.13}_{-0.09}$	1.05 ± 0.06	1.02 ± 0.08	1.7%
κ_Z	-0.98 ± 0.10	0.99 ± 0.06	1.04 ± 0.07	1.5%
κ_g	$0.78^{+0.13}_{-0.10}$	0.95 ± 0.07	0.92 ± 0.08	2.5%
κ_t	$1.40^{+0.24}_{-0.21}$	0.94 ± 0.11	1.01 ± 0.11	3.4%
κ_b	$0.49^{+0.27}_{-0.15}$	0.89 ± 0.11	0.99 ± 0.16	3.7%
κ_τ	$0.84^{+0.15}_{-0.11}$	0.93 ± 0.07	0.92 ± 0.08	1.9%
κ_μ	–	$1.06^{+0.25}_{-0.30}$	1.12 ± 0.21	4.3%
$\kappa_{Z\gamma}$	–	$1.38^{+0.31}_{-0.36}$	1.65 ± 0.34	9.8%

Figure 12: Relative precision on Higgs coupling modifiers, determined by ATLAS and CMS with the LHC data combined for the Run 1 (first column), separately for ATLAS and CMS for the Run 2 (second and third column) and the precision expected for HL-LHC [2].

For what concern the double Higgs production cross section, ATLAS and CMS experiments have estimated to reach individually a sensitivity of approximately 3.5 σ at HL-LHC [32], and a combined sensitivity of 5σ.

Future colliders projects propose both leptonic and hadronic collisions. For what concerns Higgs physics, precise measurements the couplings and self-couplings will be possible thanks to the large number of Higgs produced. The number of signal events for a certain process at a given collider depends on the luminosity and the cross section:

$$N_{sign} = \sigma_{sign} L_{int}$$

where N_{sign} is the number of signal events, σ_{sign} is the signal cross section, L_{int} is the integrated luminosity ($L_{int} = \mathcal{L} \cdot t$, where $\mathcal{L}$ is the instantaneous luminosity and t the time of data collection).

The increased statistics can be obtained then with: a larger integrated luminosity, that means higher number of interaction per bunch crossing, and/or the exploitation of the fact that a given process has a higher cross section. For example, the cross section of Higgs production processes at hadron colliders are expected to increase with $\sqrt{s}$. Even if the Higgs production cross sections are lower at lepton colliders with respect to hadron colliders, the measurements reach higher precision thanks to the lower background rates. For example, at $\sqrt{s}$=14 TeV, the single Higgs production is around a factor $\sim$ 50 larger at pp machine with respect to $\mu^+\mu^-$. Figure 13 shows in yellow the ratio between the leading single Higgs production cross section at pp colliders ($gg \to H$) and a representative "total" cross section σ_{tot} for the main background (here the $pp \to b\bar{b}$ is taken) as a function of $\sqrt{s}$, while in blue is the ratio between the single Higgs production cross section at $\mu^+\mu^-$ and the sum of inclusive electroweak boson production cross sections, as a function of $\sqrt{s}$. It is evident that lepton colliders may benefit of a much lower background contribution with respect to pp colliders. Furthermore, a lepton collider presents a cleaner environment with respect to hadron colliders due to the fact that leptons are elementary particles, and no secondary interaction arise from the collisions. Beyond the muon collider option, other proposals for Higgs Factories are under evaluation: linear e^+e^- and circular e^+e^- machines. In this chapter an overview of the proposed machines is given. At the moment, the scientific community has not clearly express a preference on

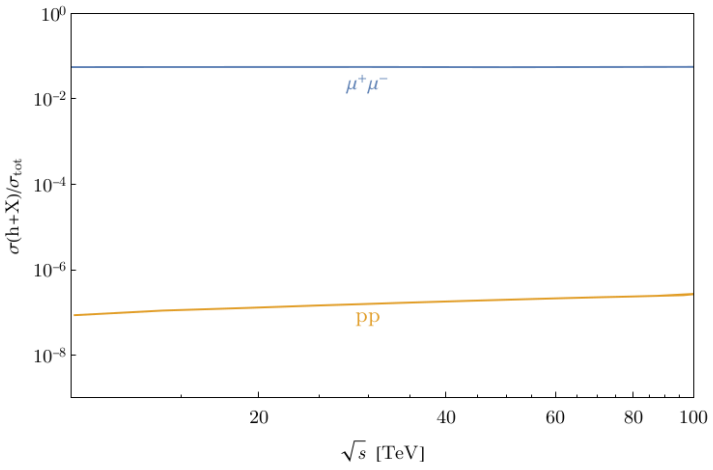

Figure 13: Higgs production cross section $\sigma(h + X)$ as a fraction of a representative "total" cross section σ_{tot} tot for $\mu^+\mu^-$ and pp colliders [7].

one of these options. The Future Circular Collider (FCC) [33] is a proposed particle accelerator, hosted in a 100 km tunnel at CERN which developed and evaluated three accelerator concepts for its conceptual design report:

- FCC-ee: a lepton collider which would operate at multiple centre of mass energies (from 91 GeV to 365 GeV);
- FCC-hh: a proton proton collider with the goal to reach center-of-mass collisions energies of 100 TeV;
- FCC-he: a proton- electron collider with 50 TeV proton beams colliding with 60 GeV electrons.

The Circular Electron Positron Collider (CEPC) [34] is a large international scientific facility to be hosted in China in a circular underground tunnel of approximately 100 km in circumference.

Compact Linear Collider, CLIC [35], is a proposed e^+e^- collider at the TeV scale. CLIC is intended to be built and operate at collision energies of 380 GeV to 3 TeV, for a site length ranging from 11 to 50 km respectively. The International Linear Collider (ILC) [36] is a proposed linear particle accelerator. It is planned to have a collision energy of 250 GeV initially, with the possibility for later upgrades up to 1 TeV.

A brief summary of the proposed Future Colliders and their characteristics in terms of luminosity and time is presented in table 2

The reaches of future collider on the Higgs couplings are evaluated following the same procedure used for LHC and HL-LHC (see section 2.3.2). The results on the sensitivity on coupling modifiers that can be reached by the considered Future Colliders are reported in Table 3. In general at e^+e^- machines most precise measurements of the couplings of the H to the W and Z bosons can be achieved, thanks to the fact that most of the H are produced via diagrams that are sensitive to the HWW (like the VBF) and HZZ (the HZ) couplings and that the collision center of mass energy is well known, since e^+e^- are elementary particle. The proposed designs for e^+e^- machines, except the CLIC at 3 TeV, foreseen stages at lower center of mass energies with respect to the muon collider: the beam power loss in the form of sinchrotron radiation limits the achievable energy and luminosity at circular and even at linear collider. Hadron colliders are complementary to e^+e^- machines, allowing to improve the sensitivity to channel that are not accessible to lepton colliders

Collider	Type	$\sqrt{s}$	$\mathcal{L}_{inst}$ $[10^{-34}cm^{-2}s^{-1}]$	$\mathcal{L}$ $[ab^{-1}]$	Time [years]	Refs	Abbreviation
FCC-hh	pp	100 TeV	30	30.0	25	[33]	FCC-hh
FCC-ee	ee	91 GeV	100/200	150	4	[33]	
		160 GeV	25	10	1-2		
		240 GeV	7	5	3		FCC-ee$_{240}$
		365 GeV	0.8/1.4	1.5	5		FCC-ee$_{365}$
FCC-he	ep	3.5 TeV	1.5	2.0	25	[33]	FCC-eh
CEPC	ee	91 GeV	17/32	16	2	[34]	CEPC
		160 GeV	10	2.6	1		
		240 GeV	3	5.6	7		
CLIC	ee	380 GeV	1.5	1.0	8	[35]	CLIC$_{380}$
		1.5 TeV	3.7	1.5	7		CLIC$_{1500}$
		3.0 TeV	6.0	5.0	8		CLIC$_{1000}$
ILC	ee	250 GeV	1.35/1.7	2.0	11.5	[36]	ILC$_{365}$
		350 GeV	1.6	0.2	1		ILC$_{350}$
		500 GeV	1.8/3.6	4.0	8.5		ILC$_{500}$
		1000 GeV	3.6/7.2	8.0	8.5		ILC$_{1000}$
Muon Collider	$\mu\mu$	3.0	2	1	5	[37]	
		10	20	10.0	5		

Table 2: Summary of the Future Colliders considered in this section in terms of operative center of mass energy $\sqrt{s}$, integrated ($\mathcal{L}$) and instantaneous luminosity $\mathcal{L}_{inst}$ and time years [4]

due to the low statistics, like the coupling to the top quark and the muon. The FCC-ee/eh/hh column is the combination of all the FCC-ee, FCC-eh and FCC-hh stages. Deviations from the

κ	ILC			CLIC			CEPC	FCC-ee/eh/hh
	250	500	1000	380	1500	3000		
κ_W [%]	1.8	0.29	0.24	0.86	0.16	0.11	1.3	0.14
κ_Z [%]	0.29	0.23	0.22	0.5	0.26	0.23	0.14	0.12
κ_g [%]	2.3	0.97	0.66	2.5	1.3	0.9	1.5	0.49
κ_γ [%]	6.7	3.4	1.9	98$\star$	5.0	2.2	3.7	0.29
$\kappa_{Z\gamma}$ [%]	99 $\star$	86$\star$	85$\star$	120$\star$	15	6.9	8.2	0.69
κ_c [%]	2.5	1.3	0.9	4.3	1.8	1.4	2.2	0.95
κ_t [%]	-	6.9	1.6	-	-	2.7	-	1.0
κ_b [%]	1.8	0.58	0.48	1.9	0.46	0.37	1.2	0.43
κ_μ [%]	15	9.4	6.2	320$\star$	13	5.8	8.9	0.41
κ_τ [%]	1.9	0.7	0.57	3.0	1.3	0.88	1.3	0.44

Table 3: Expected relative precision (%) of the κ parameters for Future Colliders after HL-LHC. Colliders are considered independently, not in combination with the HL-LHC. For more details about the entries with $\star$ and $-$, see [4].

trilinear Higgs self-coupling from the Standard Model is usually parametrized as deviation from the Standard Model:

$$\delta\kappa_3 = \kappa_3 - 1 = \frac{\lambda_3}{\lambda_{SM}} - 1 \tag{2.23}$$

where $\lambda_{SM} = \frac{m_H^2}{2v^2}$ is the trilinear or quadrilinear Higgs self-coupling predicted by the Standard Model (see equation 2.10) .

The uncertainty expected at the proposed future colliders on the measurement of $\kappa_3 = \frac{\lambda_3}{\lambda_3^{SM}}$ in combination with the expectation from HL−LHC is summarized in Table 4.

collider	sensitivity on λ_3
HL-LHC	$\sim 50\%$
ILC_{500}	27 %
ILC_{1000}	10 %
$CLIC_{1500}$	36 %
$CLIC_{3000}$	~ 10 %
FCC-ee/eh/hh	5 %

Table 4: Sensitivity at 68% probability on the Higgs cubic self-coupling at HL-LHC and Future Colliders [4].

The Muon Collider is a proposed $\mu^+\mu^-$ collider, that have the potential to reach center of mass energies of tens of TeV. The current stages proposed for the muon collider are at 3 TeV and 10 TeV center of mass energy. The muon collider could be the best solution as Higgs factory: a 3 TeV muon collider with integrated luminosity 1 ab^{-1} is expected to produce a sample of approximately $5.5{\cdot}10^5$ single Higgs bosons and 1000 double Higgs events [37]. At a 10 TeV collider with a dataset of 10 ab^{-1}, approximately 9.3 million of single Higgs bosons and a sample of about 38000 double-Higgs events are expected.

κ	$\mu^+\mu^-$		+ HL-LHC	
	3 TeV	10 TeV	3 TeV	10 TeV
κ_W [%]	0.37	0.10	0.35	0.10
κ_Z [%]	1.2	0.34	0.89	0.33
κ_g [%]	1.6	0.45	1.3	0.44
κ_γ [%]	3.2	0.84	1.3	0.71
$\kappa_{Z\gamma}$ [%]	21	5.5	22	5.5
κ_c [%]	5.8	1.8	5.8	1.8
κ_t [%]	34	53	3.2	3.2
κ_b [%]	0.84	0.23	0.8	0.23
κ_μ [%]	14	2.9	4.7	2.5
κ_τ [%]	2.1	0.59	1.2	0.55

Table 5: Expected relative precision (%) of the κ parameters for muon collider at 3 TeV and 10 TeV and muon collider + HL-LHC [38].

Most of the couplings can be estimated at few percent level at 3 TeV whereas at 10 TeV with 10 ab^{-1} , sensitivities at the permille level are possible, with precisions comparable to future e^+e^- Higgs factories. Some couplings are challenging at Muon Collider even at 10 TeV, like the top Yukawa coupling (y^t), and is significantly below the HL-LHC measurement, as can be understood by the improvement in the uncertainty after the combination, in the last two columns of Table 5. Alternative ways to constrain this channel at muon collider that exploits the channel $W^+W^- \to t\bar{t}$ are under study in [7]. The prospect for the trilinear Higgs self-coupling at muon collider have been calculated in [39] and have been found to be $\sim 25\%$ at 3 TeV center of mass energy. However, at 10 TeV, in 5 years of data taking, the uncertainty on the cubic coupling could be pushed to $\sim 5.6\%$, allowing to reach the highest performance on this coupling at such center of mass energy. This prospect has been calculated with parametric studies, without the full simulation of the detector effects. The four b quarks from the Higgs decay, and the physics backgrounds (four quarks events

from Z and W resonances) are required to be emitted inside a polar angle range between 10° and 70° and a jet energy resolution of ~ 10 % is assumed. The good energy resolution is the key aspect for the uncertainty on the trilinear Higgs self-coupling estimation, since it allows reject almost all the physics background by only imposing an invariant mass requirement of $\pm$ 15 GeV around the Higgs peak for each the two Higgs candidates. The sensitivity on the quadrilinear Higgs self-coupling that can be reached at the end of the entire FCC-hh run [40], under the assumption that $\lambda_3 = \lambda_3^{SM}$:

$$\frac{\lambda_4}{\lambda_4^{SM}} \in [-2, +13]$$

at 2σ with 20 ab^{-1}.

It is obtained by studying HHH production with $6b$ in the final state.

The first analysis of the muon collider potential on the quartic Higgs self-coupling precision has been performed in [41]. This study is performed at phenomenological level, no detailed detector simulation, and the possible physics backgrounds are not considered since they are considered not relevant for the final sensitivity. Assuming $\lambda_3 = \lambda_3^{SM}$ for a 10 TeV muon collider and integrated luminosity of 20 ab^{-1}, uncertainties on the quartic Higgs self-coupling of the order $\sim 50 - 60\%$ are expected to be reached at 1σ C.L.. High energy muon collisions are the only place where such a measurement could me possible.

2.4.1 Muon collider potential

The muon collider potential in the production of New Physics heavy particles has been evaluated in [42] considering $\mu^+\mu^-$ collisions for the single particle final state and has been compared to pp colliders. In single particle final state production, the fact that muons are elementary particles implies that all the beam energy is available for the hard collisions. At a proton collider the relevant interactions occur between the proton constituents, which carry a small fraction of the beam $\sqrt{s_p}$ energy. As mentioned above, at a muon collider, differently from e^+e^- colliders, center of mass energies of the order of 10-14 TeV can be reached. Given a muon collisions cross section $[\hat{\sigma_\mu}]$ for a on threshold process $\mu^+\mu^- \to Y$ ($\sqrt{s_\mu} = M_Y$), the equivalent proton energy collider is defined as the corresponding pp collisions energy $\sqrt{s_p}$ such that the process $pp \to Y$ has $\sigma_p = [\hat{\sigma_\mu}]$. However protons are not elementary particles then the equivalent total cross section σ_p is evaluated in terms of the partons level cross section $\hat{\sigma_p}$, since the colliding particles in $pp \to Y$ process are quarks or gluons.

Figure 14 shows the equivalent pp collider energy for the single particle final state Y with mass $M = \sqrt{s_\mu}$ as a function of the $\mu\mu$ collider energy. Red (blue) lines represent the case in which gluons (quarks) collide. The comparison is done for different assumptions on the couplings involved in the Y production. By labelling colliding partons as (ij):

- for $[\hat{\sigma}_\mu] = [\hat{\sigma}_p]$ (solid line) both $ij \to Y$ and $\mu^+\mu^- \to Y$ processes are governed by the same physics;
- for $[\hat{\sigma}_p] = 10[\hat{\sigma}_\mu]$ (dashed line) the former is governed by QCD and $\mu^+\mu^- \to Y$ by QED;
- for $[\hat{\sigma}_p] = 100[\hat{\sigma}_\mu]$ (pointed line) the former is governed by QCD and the latter by electroweak coupling.

As can be seen, a larger pp collider energy is necessary to achieve the same cross section as a $\mu\mu$ collider.

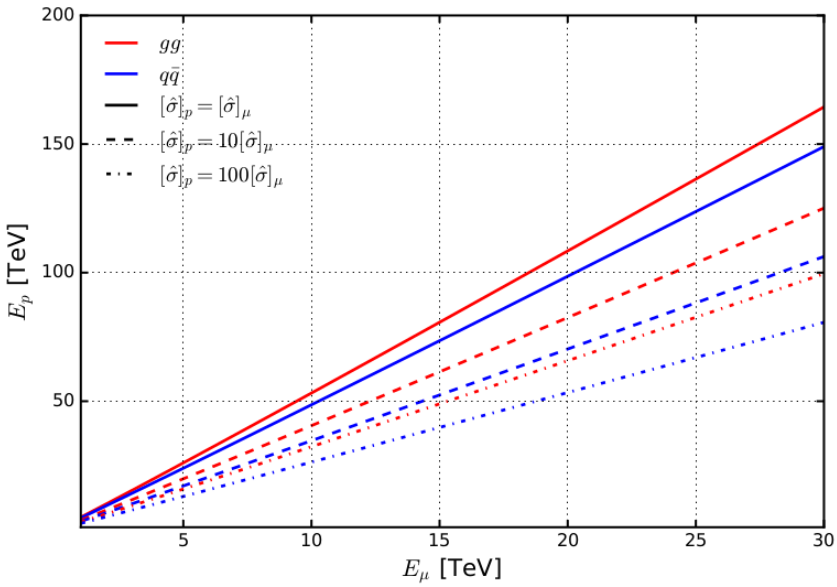

Figure 14: The equivalent proton collider energy E_p [TeV] required to reach the same cross section as a $\mu^+\mu^-$ collider with energy E_μ [TeV] for the annihilation of initial-state particles into a single-particle final state.

2.4.2 Higgs full simulation studies

All these preliminary estimations shown in Table 5 on the Higgs couplings have been obtained via fast simulation, assuming detector performance that do not include the BIB effects on object reconstruction. One peculiar characteristic of the Muon Collider, is the presence of a large amount of backgorund (BIB) generated by muons decay in the beam. A campaign of studies to estimate the precision on the Higgs cross sections via full simulation of the muon collider detector and including the BIB in the reconstruction of the signal and physics background processes, are ongoing. The single Higgs production, with the Higgs decaying in different final states has been studied at 3 TeV center of mass energy $\mu^+\mu^-$ collisions. The goal is to assess the results shown in Table 5. The sensitivity on the $\sigma_H \times BR(H \to XX)$ obtained so far are summarized in Table 6. These results

$\sigma_H \times BR(H \to XX)$	Uncertainty	ref
$\sigma_H \times BR(H \to b\bar{b})$ [%]	0.75	[43]
$\sigma_H \times BR(H \to WW^*)$[%]	2.9	[44]
$\sigma_H \times BR(H \to ZZ^*)$[%]	17	[43]
$\sigma_H \times BR(H \to \gamma\gamma)$ [%]	7.6	[45]
$\sigma_H \times BR(H \to \mu^+\mu^-)$ [%]	38	[46]

Table 6: Sensitivity on the $\sigma_H \times BR(H \to XX)$ obtained so far at a 3 TeV Muon Collider in the presence of BIB.

have been obtained by assuming 1 ab^{-1} dataset, samples have been processed with the detailed detector simulation and reconstructed with the muon collider framework (see section 7.6):

- $H \to b\bar{b}$ events have been reconstructed in [43]. The main physics background is from Z boson decay into $b\bar{b}$ and $c\bar{c}$. A tagging algorithm is used to tag heavy flavour jets. Even in this study the dijet mass resolution (around 18 % for the Higgs) is critical to separate the H and Z peaks and is dominated by BIB effects;

- $H \to W\bar{W}$ has been reconstructed considering the semileptonic final state $qq'\mu\nu_\mu$. This channel provides a good signal-to-background ratio, and the estimated number of Higgs candidates is 2430 over a background of 2600 events;
- $H \to Z\bar{Z}$ is reconstructed in the semileptonic final state $q\bar{q}\mu\mu$, and events are selected with at least two reconstructed jets and two opposite-charge muons. The uncertainty on the cross section has been estimated with 55 signal events and 39 background events;
- the $H \to \gamma\gamma$ search for two high energy final state photons is not significantly affected by the BIB, since a 15 GeV requirement on the photon energy, remove most of the photons from BIB noise. Then, even in presence of BIB, the Higgs invariant mass exhibits a very good resolution of 3.2 GeV. 396 signal events and 484 background events are then used to determine the uncertainty on the cross section;
- the $H \to \mu^+\mu^-$ is reconstructed by selecting events with two muons in the final state. Even in this case the BIB impact on the reconstruction is found to be negligible. 26 signal events and 1114 background events survive the final selection and are used to determine the uncertainty on the cross section.

The study performed in this thesis evaluate the uncertainty on the double Higgs cross section and of the trilinear Higgs self-coupling with a full simulation of the muon collider detector effects on the object reconstruction. Furthermore, the effects of the presence of the BIB on these uncertainties are studied and discussed.

3 LHCb detector description

3.1 LHC overview

The Large Hadron Collider (LHC) is a proton-proton circular collider built from 1998 to 2008 by the CERN (Conseil Européen pour la Recherché Nucléaire) located beneath the France–Switzerland border near Geneva. The LHC lies in a tunnel of 27 km circumference and between 50 and 175 m below the Earth surface. It consists of two rings with counter-rotating beams and is designed to collide proton bunches at nominal center of mass energies up to 14 TeV and peak luminosity up to $\mathcal{L}$= 10^{34} cm^{-2}s^{-1}. The instantaneous luminosity can be written as [47]:

$$\mathcal{L} = \frac{N_b^2 f_{coll} \gamma}{4\pi \epsilon_n \beta^*} \cdot F \tag{3.1}$$

where N_b is the number of particles per bunch, f_{coll} the collision frequency, β^* is the beta function at the collision point, that depends on the magnet lattice and is related to the beam size at the IP and ϵ_n the normalized transverse beam emittance, that is a measure for the average spread of particle coordinates in position and momentum phase space in a beam. In circular machines beam dynamics is described in the Frenet-Serret coordinate systems [48] by using the s coordinate, fixed to the reference particle that moves along the reference path, that is the ideal orbit that pass at the center of the lattice design. The description of the beam dynamics in the transverse phase space is done with coordinates $x(s), x'(s), y(s), y'(s)$, where $x(s), y(s)$ are the displacement perpendicular to the reference orbit $x' = \frac{dx}{ds}$ and $y' = \frac{dy}{ds}$ are the angle with respect to the reference orbit. In an x, x' (or y, y') plane, the beam has an elliptical shape with area $A = \epsilon_x \pi$, where ϵ_x is the emittance in the x, x' plane. In a low-emittance beam, particles are confined to a small phase space area and have nearly the same momentum. In the above equation 3.1 the beam emittances is normalized by the relativistic factor $\epsilon_n = \beta_r \gamma \epsilon$ in which β_r and γ are the relativistic functions. It is assumed that beam profiles are Gaussian in the transverse plane and that the emittances are equal in both x and y [49]. F is the geometric luminosity reduction factor due to the crossing angle at the interaction point (IP). In the case of LHC where the crossing angle is small, it can be expressed as:

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2} \tag{3.2}$$

where θ_c is the full crossing angle at the IP, σ_z the root-mean-square (RMS) bunch length and σ^* the transverse RMS beam size at the IP. Nominal typical values for these parameters at LHC, that allow to achieve nominal luminosity of L= 10^{34} cm^{-2}s^{-1} are listed in Table 7: The CERN

LHC parameters	
N_b	$1.1 \cdot 10^{11}$ particles/bunch
f_{coll}	30 MHz
β^*	0.55 m
σ^*	16.7 μm
σ_z	$\sim$ 7 cm
θ_c	$\sim$ 300 μ rad
γ	$\sim$ 7500
ϵ_n	$\sim$ 3.75 μm

Table 7: List of nominal beam parameters at LHC to achieve the peak luminosity of 10^{34} cm^{-2}s^{-1} [47].

complex acceleration chain [50] consists of several accelerator stages that bring the protons from

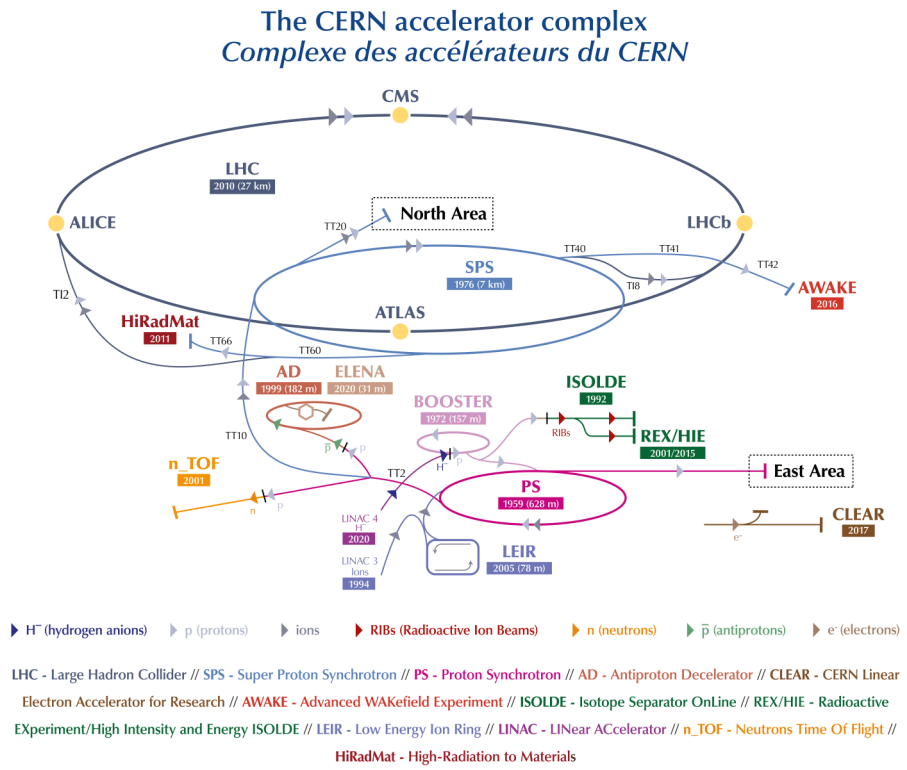

Figure 15: Scheme of LHC and CERN accelerators facility. Yellow points indicate the four interaction points where CMS, ATLAS, LHC and ALICE experiments are host [50].

the production to the nominal colliding energy. Protons are initially produced from an ionizing hydrogenum source and are accelerated up to 50 MeV by LINAC4. Then they enter the Proton Synchrotron Booster (PSB) where they are accelerated up to 1 GeV. The third acceleration stage occurs in the Proton Synchrotron, where proton bunches are formed, with a spacing of 25 ns and are accelerated up to 26 GeV. After that, bunches enter the Super Proton Synchrotron (SPS) where they are accelerated up to 450 GeV and are injected into the LHC. The full scheme of LHC and CERN accelerators facility is shown in Figure 15.

Two of the LHC collision points aims at such high luminosity colliding protons: ATLAS and CMS, that are General Purpose Detector with the aim of investigating the Standard Model and searching for Beyond the Standard Model physics. These detectors can cope with a maximum average number of collisions per bunch crossing (pile-up) of 50. LHCb was initially designed to study heavy quark hadrons (beauty and charm), in particular to search for indirect evidence of new physics in CP violation and rare heavy quarks decays. However, it is used even as General Purpose Detector, like ATLAS and CMS, for example to test perturbative QCD predictions and electroweak physics, new physics models predicting beyond the Standard Model particles or Standard Model physics in a region unexplored by other experiments.

A luminosity leveling technique [51] is applied, and consists in tilting the two proton beams to increase the interaction area. The beam parameters at the interaction point are also tuned in order to lower the peak luminosity: for example, the β^* was around 10 m during the 2015/2016 operations LHCb and 3 m during 2017/2018 [52]).

These techniques reduce the peak luminosity to $\mathrm{L} = 4.5 \cdot 10^{32}\mathrm{cm}^{-2}\ \mathrm{s}^{-1}$ but have the advantage of reducing pile-up around 1.1 in Run 1 and 2. The reduction of multiple pp collisions improves the performance on the identification of beauty and charm hadrons, avoiding the risk of assigning

year	$\mathcal{L}_{int}$ fb^{-1}	μ	$\mathcal{L}$ 10^{32}cm^{-2} s^{-1}
2016	1.67	1.1	2.0
2017	1.71	1.1	3.5
2018	2.01	1.1	4.4

Table 8: Integrated recorded luminosity $\mathcal{L}_{int}$, visible number of collisions per bunch crossing μ and peak luminosity $\mathcal{L}$ at LHCb for each year of Run 2 campaign [53].

the beauty and charm decay vertex to a wrong pp vertex. The fourth experiment is ALICE (A Large Ion Collider Experiment), that is dedicated to heavy-ion collisions and studies the property of quark-gluon plasma, strongly interacting matter at extreme energy densities.

3.2 Detector overview

The overview of the LHCb detector [53] is shown in Figure 16. The coordinate system is right handed with the z axis chosen along the beam line, the y axis parallel to the vertical direction and the x axis (horizontal direction) perpendicular to the y and the z axis. Due to the presence of a magnet that produce a magnetic field in the y direction, that will be described in section 3.2.2, the xz plane is also called bending plane, while the yz plane is called the non-bending plane. The LHCb detector is a single-arm forward spectrometer that covers a forward angular coverage from approximately 10 mrad to 300 (250) mrad in the bending (non-bending) plane. This corresponds to a pseudorapidity [1] (η) range between $2<\eta<5$.

The choice of this geometry is justified by the fact that at high energy pp collisions b and $\bar{b}$ hadrons pairs are mainly produced in the forward or backward directions, with a small angle with respect to the beam direction. In this thesis, a sub-sample of the data acquired during the Run 2 data-taking campaign is used. In the Run 2, LHCb has operated at a center of mass energy of $\sqrt{s}$=13 TeV. The integrated luminosity, the number of visible collision per bunch crossing (μ) and the peak luminosity $\mathcal{L}$ for the Run 2 are shown in Table 8.

The detector is composed by a sequence of sub-detectors:

- The vertex locator (VELO) is the closest sub-detector to the interaction point. It has a fundamental role in detecting signature of b and c hadrons decays, since it is devoted to measure tracks parameters close to the interaction region which are used to identify displaced secondary vertices;
- The tracking system is made by four stations: the closest to the IP, at z=2.5 m, is the Tracker Turicensis (TT) station. TT detector sensors are used to build track segments that can be combined to the tracks segments reconstructed in the VELO, improving the track momentum and position resolutions; The other three stations are located between z= 8 m and 10 m called T1,T2,T3 and comprise a inner tracker (IT) closer to the beam pipe, and an outer tracker (OT), that cover the LHCb angular acceptance.
- The magnet is located between the first (TT) and the second tracking station (T1);
- two ring Imaging Cherenkov counters (RICH1 and RICH2) dedicated to the separation of pions from kaons in the momentum range from 2 to 100 GeV/c;
- the calorimeter system is composed of a Scintillator Pad Detector and Preshower (SPD/PS), and two calorimeters: an electromagnetic calorimeter (ECAL) and a hadronic (HCAL);
- the muon detector system, that is the most forward detector.

[1] The pseudorapidity for a particle is defined as $\eta = -ln tan(\theta/2)$, where θ is the particle polar angle

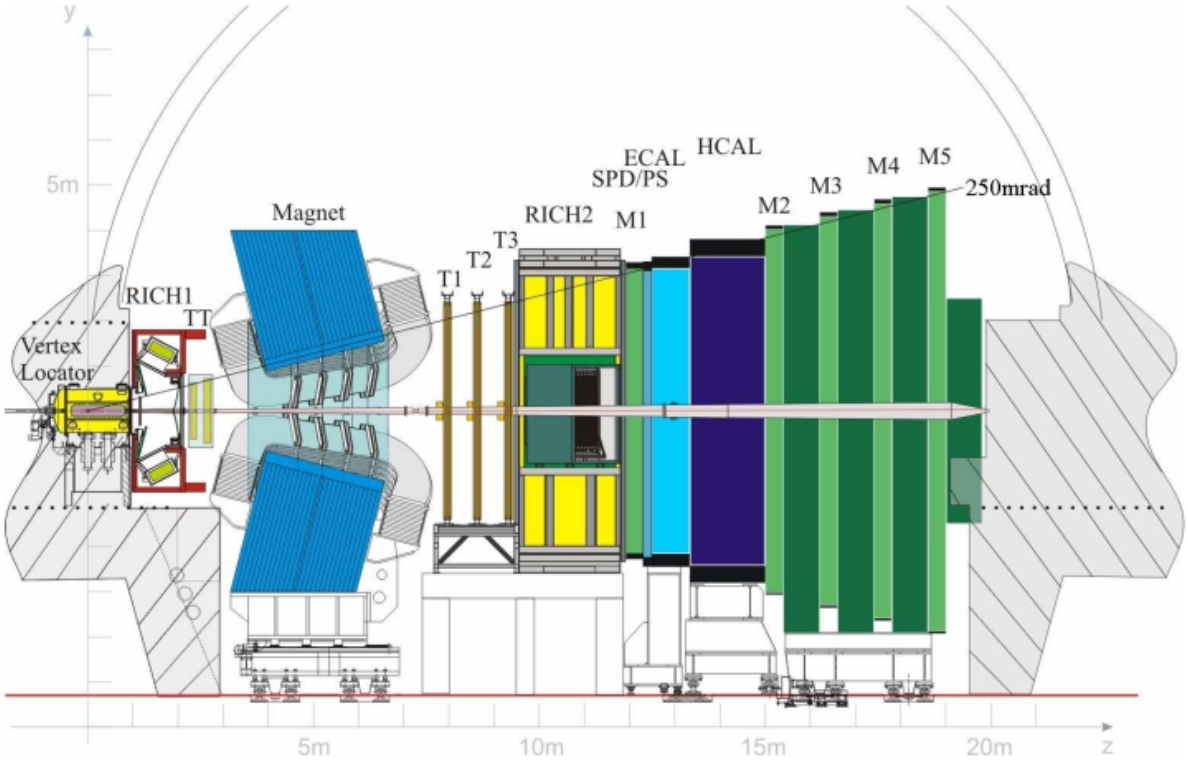

Figure 16: View of the LHCb detector [54].

3.2.1 Vertex Locator

The VELO's layout is shown in Figure 17. It consists of 21 silicon modules arranged along the beam direction over a length of about 1 m. The interaction region is highlighted in the Figure in green: some modules are located backward with respect to the interaction region. Each module appears as a disk of 84 mm diameter and 300 μm thickness, with a hole in the center of 16 mm diameter, centered at the nominal LHC beam position. The detector is separated from the beam pipe only by a thin aluminum foil. During LHC beam injection, VELO modules can be retracted by 29 mm in the horizontal direction for detector safety: this setup is called "VELO fully open" and is schematically shown in the bottom right part of Figure 17. When the beams are in stable condition and VELO modules can be placed in position for data acquisition, the two halves of the VELO disk are "closed" around the nominal LHC beam position, as can be shown in the bottom left of Figure 17.

Two types of semicircular sensors are used in order to measure both the radial r(R sensors) and the azimuthal coordinates ϕ (ϕ sensors). Figure 18 shows the structure of a VELO module, while in in grey are highlighted the R (left half) and ϕ (right half) sensors.

R sensors measure r track coordinate between 8 and 42 mm, their strips are arranged into concentric semicircles, the minimum pitch at the innermost radius is 38 µm, increasing linearly to 102 μm at the outermost radius. ϕ sensors are used to measure azimuthal coordinate between 15 and 390 mrad, and its strips run radially from the inner to the outer radius. It is divided into two regions, an inner disk with radius from 8 to 17.25 mm and an outer disk, with radius from 17.25 to 42 mm. In the inner part the sensor pitch increases with the radial distance from 39.3 μm to 78 μm while in the outer pitch increases from 39 μm to 97 μm.

The performances of the VELO detector have been analyzed using the data collected in 2010 and 2011 [55]. The hit resolution depends on the inter-strip readout pitch and the charge sharing

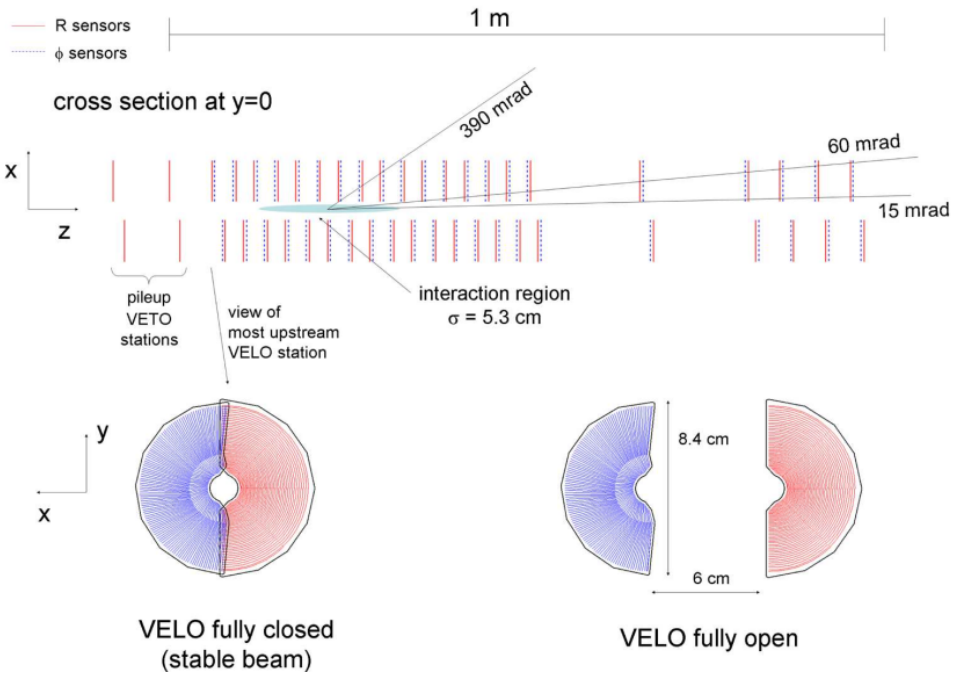

Figure 17: Up: module layout of the LHCb VELO. Bottom: setup of VELO open (left), and VELO closed (right) [53].

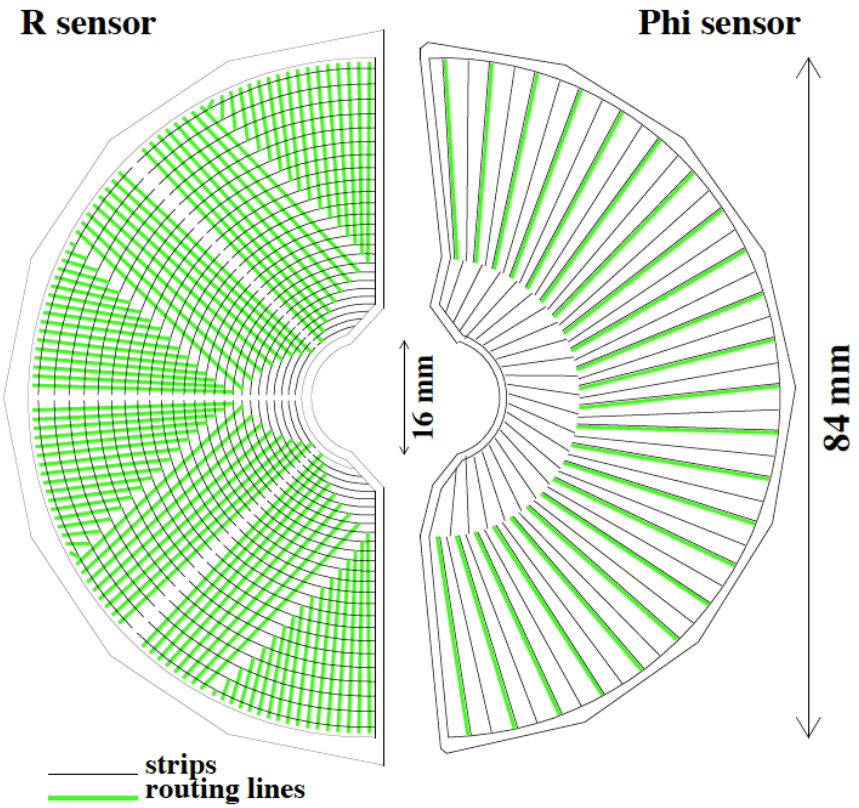

Figure 18: Structure of a VELO module: the two semi-circular modules composing the R and ϕ sensors are shown [55].

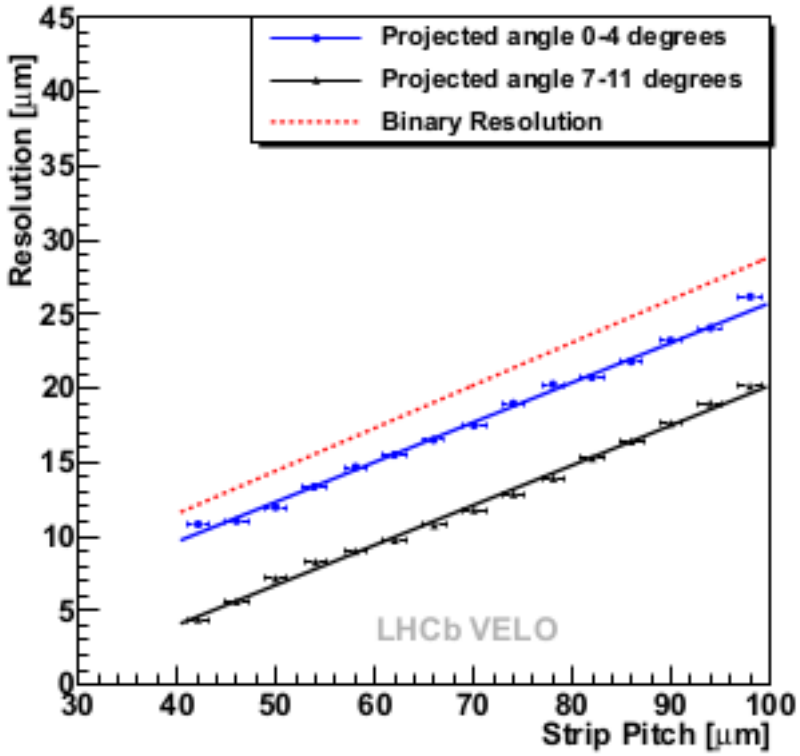

Figure 19: The VELO resolution for two projected angle bins for the R sensors as a function of the readout pitch compared with binary resolution [54].

between strips [56]. The single hit resolution of a digital readout in which all charge is measured by a single strip is called binary resolution and is calculated as pitch/$\sqrt{12}$. At LHCb the single hit resolution is estimated using an ADC count weighted strips. The resolution is improved by exploiting the charge sharing between strips that depends on the projected angle, defined as the angle between the track trajectory and the perpendicular to the strips. As this angle increases, the probability that the charge is distributed over more than one strip, is larger. The cluster position can be then determined from the pulse height weighted average of the strips contributing to the cluster.

Figure 19 shows the resolution for the R sensors as a function of the strip pitch, for projected angle of the tracks between 0° to 4° (blue) and between 7° to 11° (black). The binary resolution is shown in red. The measured hit resolution, once fixed to a certain projected angle range, has a linear dependence on the strip pitch. The hit resolution at small projected angles is close to the binary system resolution, since the charge sharing is minimal. However it is improved for larger projected angles, since the fraction of two strip clusters increases and the advantages in using analogue readout of the pulse height can be exploited. The best resolution is found to be around 4 μm in the smallest strip pitch region (about 40 μm). The ϕ sensor results are found to be compatible with the R sensors results.

3.2.2 Magnet

The warm (non-superconductive) dipole magnet is placed between the TT and the T1-T3 tracking stations, with the aim of measuring the charged particle momentum. It is made by two aluminum conical coils, placed inside the magnet yoke, as shown in Figure 20. The magnetic field is parallel to the y axis, and is generated by the electric current that flows in the coils. By inverting the electric

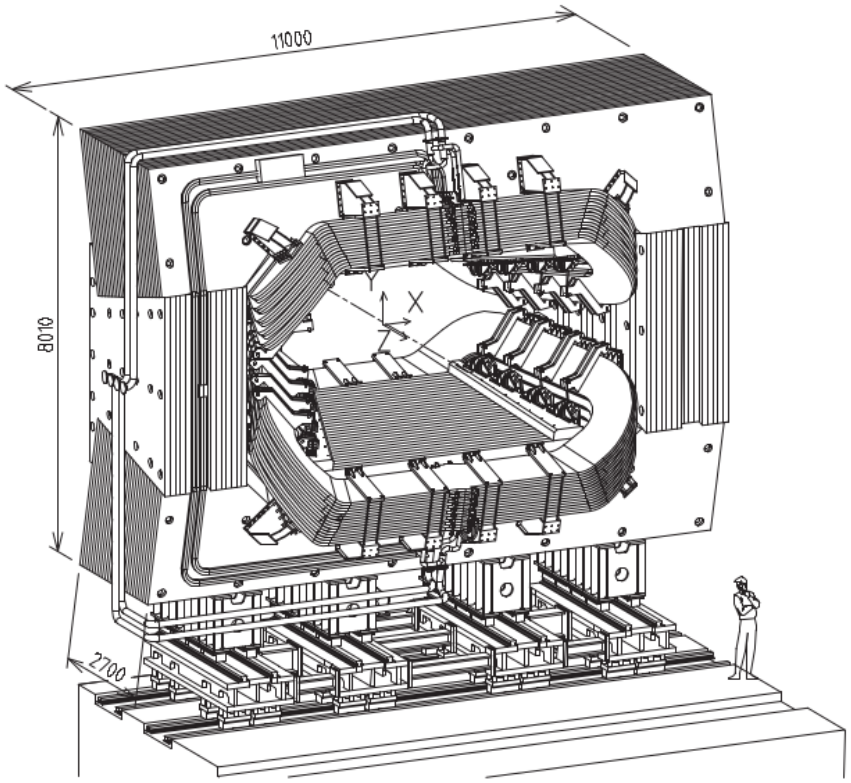

Figure 20: Representation of the magnet inside LHCb (units expressed in mm).

current, the magnetic filed direction can be inverted, allowing measurements with two opposite polarities. The integrated magnetic field for tracks 10 m long is about 4 Tm, while there is still a residual magnetic field of 2 Tm inside the two RICHs.

3.2.3 Tracking system

Figure 21 shows the tracker system layout. It comprises a Silicon Tracker (purple), that covers the region closer to the beam pipe, and an Outer Tracker (cyan). The Silicon Tracker system comprises four planar micro-strip detectors stations arranged perpendicular to the beam axis: the Tracker Turicensis (TT) is placed between the RICH 1 and the magnet, and the Inner Tracker (IT), whose stations are located after the magnet and constitute the inner parts of the stations T1, T2 and T3. They are composed by four layers of 200 μm wide silicon sensors, that differ for the angle at which strips are tilted with respect to the y directions. In the first and fourth stations the strips are parallel to y, while in the second and third stations they are tilted by $+5^\circ$ (u-layer) and -5° (v-layer) respectively. In Figure there is an example of the third TT detection layers. It covers the entire height of the LHCb acceptance, and is composed by half modules, highlighted in red in the figure, that contains seven silicon sensors. Silicon sensors for the TT are 500 μm thick, are 9.64 wide and 9.44 cm long.

The IT station is composed by either one or two silicon modules, shown in Figure, that are 7.6 cm wide and 11 cm long, and carry strips with a strip pitch of 198 μm and thickness 320 μm for the one-sensor module or 410 μm for the two-sensors module.

The OT is composed by four gas-tight straw-tube modules. Each module contains two staggered layers of drift-tubes with inner diameters of 4.9 mm and are filled with a mixture of Argon (70%) and CO_2 (30%). The detector modules are arranged in three stations. Also in this case each module

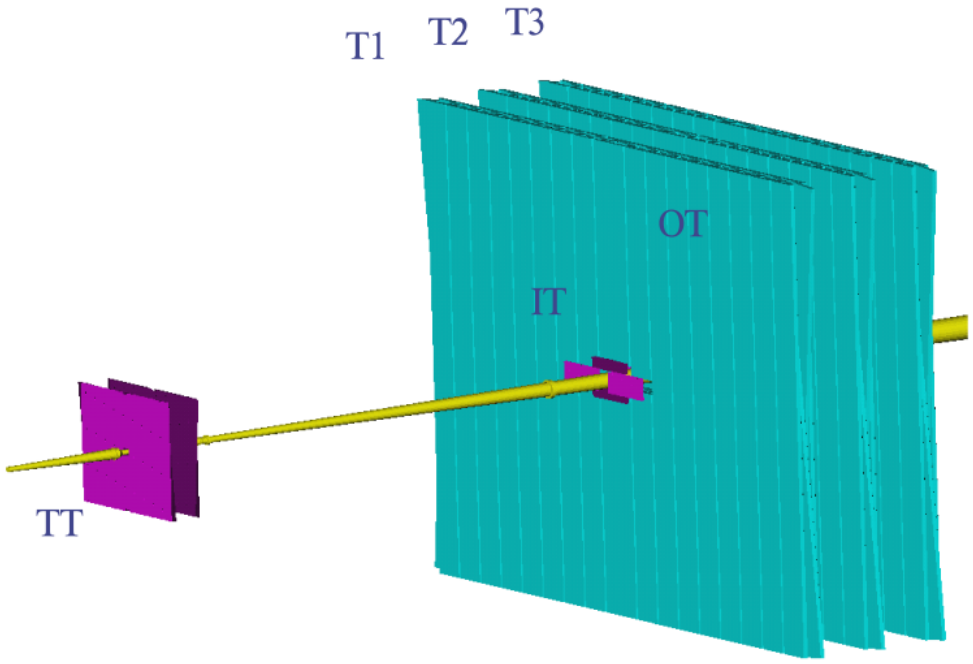

Figure 21: View of the LHCb tracking system. In purple the Tracker Turicensis and the Inner Tracker are shown, the Outer Tracker is coloured in cyan [57].

is composed by four layers, arranged in an x-u-v-x geometry. The angular acceptance covered by the OT is 300 mrad in the magnet bending plane (horizontal) and 250 mrad in the non-bending plane (vertical). The single hit spatial resolution has been determined to be around 50 μm for both the TT and the IT [54].

3.2.4 RICH

Two RICH detectors are used at the LHCb experiment to separate pions from kaons particles in selected B hadron decays. They are based on the Cherenkov effect: a cone of electromagnetic radiation is emitted with axis parallel to the charged particle speed, when it traverses a medium with a velocity higher than the speed of light in that medium. The RICH detectors measures the angle θ_c with respect the direction of the particle direction at which the radiation is emitted. The particle velocity β can be calculated by knowing the Cherenkov angle θ_c and the medium refractive index n with the relation:

$$cos(\theta_c) = \frac{1}{\beta n}$$

Particles can then by identified by combining the measurement of θ_c by the RICH and the momentum measurements by the main tracking system, since the refractive index depends only on the radiator material used to fill the detector. The RICH1, closer to the interaction point, is used to identify low momentum charged particles ($\sim$ 1-60 GeV/c) while RICH2 covers the high momentum range (from 15 GeV/c up to and beyond 100 GeV/c). The former uses a 5 cm thick aerogel layer (refractive index n = 1.03) and a volume filled with C_4F_{10} radiators (refractive index n = 1.03) and cover the full LHCb detector acceptance from 25 mrad to 300 mrad and 250 mrad. The latter uses CF_4 radiator (refractive index n = 1.0014) and covers the region where high momentum particles are produced $\pm$15 mrad to $\pm$120 mrad (horizontal) and $\pm$100 mrad (vertical). The RICH system is schematically drawn in Figure 22. In both the RICHes, Cherenkov photons are reflected

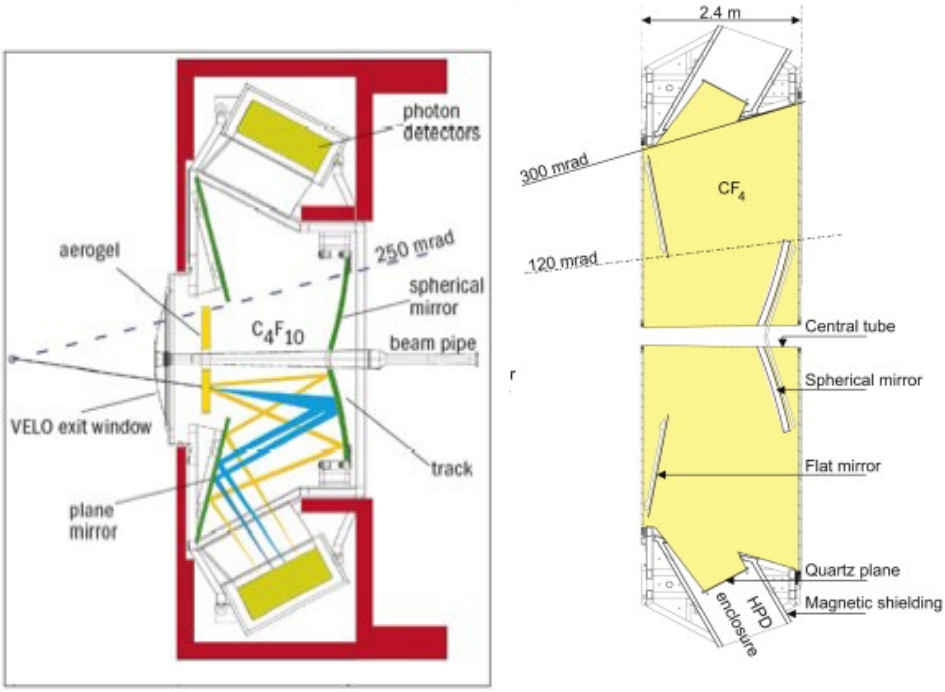

Figure 22: View of the RICH1 and RICH2 systems [53].

by spherical and plane mirrors, coloured in dark green in Figure 22 left and white in Figure 22 right, and are directed to photon detectors (Hybrid Photo Detectors, HPDs, visible in light green in Figure 22 light) where they are converted into electrons. Due to the vicinity to the magnet, in order to avoid electrons to be bent by the residual magnetic field, in RICH1 the HPDs are shielded by iron shields, shown in red in the Figure.

3.2.5 Calorimeters

The calorimeter system [54] [53] [58] is composed by a Scintillating Pad Detector (SPD), a Preshower detector (PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). The system is located between the two muon system stations and has several purposes:

- It measures the transverse energy of hadron, electron and photon candidates for the first trigger level (L0);
- it provides the identification of electrons, photons and hadrons and measures their energies;
- it reconstructs and measure the energy of neutral particles like π^0, photons that is essential for jet reconstruction and flavour tagging.

Each sub-detector is segmented into regions and cells size, as shown in Figure 23. A cell corresponds to a readout electronics group of channels. The SPD and the ECAL are divided into three regions, while the HCAL is divided into two regions. In both the layout cells have increasing sizes going from the innermost region to the outermost region, since the hit density is higher close to the beampipe with respect to the outer region. The dimensions and the number of cells that compose each region are shown in the left boxes in Figure 23 [59]. The SPD and PS detectors are sited in front of the electromagnetic and the hadronic calorimeters and are aimed at tagging the electromagnetic nature of the calorimeter clusters for the first level trigger 3.3.1. They consist of a 12 mm thick lead converter 2.5 radiation length (X_0) thick, that corresponds to the distance over which

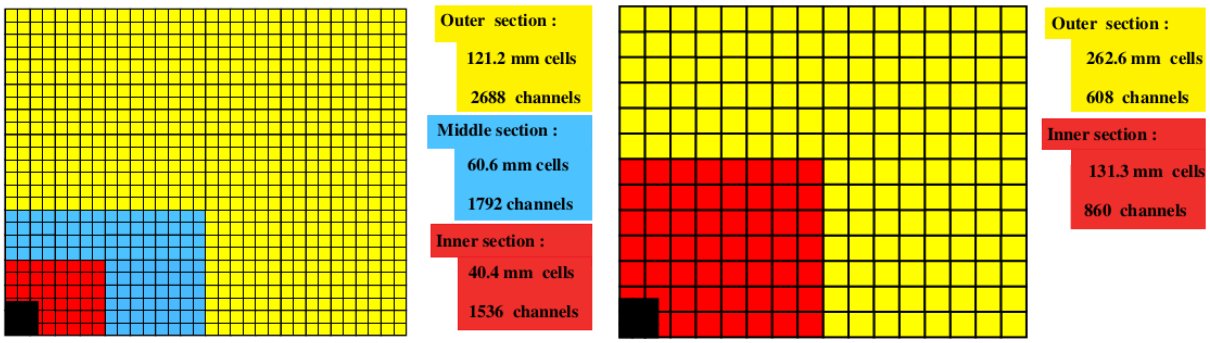

Figure 23: Lateral segmentation of the SPD/PS and ECAL (left) and the HCAL (right). One quarter of the detector front face is shown. In the right the cells dimensions and the number of channels are shown [53].

the energy of an electron is reduced by a factor $1/e$, between two planes of 15 mm thick scintillator pads. Scintillator light is transmitted to a Multi Anode Photo-Multiplier by wavelength-shifting (WLS) fiber. The SPD and PS identify charged particles, and allows electrons to be separated from photons, since they provide information of the electromagnetic shower development. Charge particles are detected in the SPD where neutral ones leave no signal, and this allows to distinguish electrons from photons. The lead thickness initiate the electromagnetic showers for electrons and photons, while the hadronic showers start later. The PS is then used to discriminate electromagnetic showers from hadron showers.

The ECAL calorimeter is placed at 12.5 m from the interaction point and covers the LHCb angular acceptance between 25 mrad up to 300 mrad and 250 mrad in the non-bending and bending plane rispectively. The shashlik electromagnetic calorimeter technology is a sampling scintillator/lead structure readout by plastic WLS fibers, where the 2 mm thick lead layers are alternated to 4 mm thick scintillator tiles of olystyrene. In depth, the 66 Pb and scintillator layers form a 42 cm stack corresponding to 25 X_0. The calorimeter Molier radius is around 3.5 cm.

The energy resolution is

$$\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \bigoplus 1\% \tag{3.3}$$

The HCAL structure is places at a distance from the interaction point of Z=13.33 m, with dimensions of y=8.4 m in height, x=6.8 m in width and z=1.65 m in depth. It is a sampling device made from iron absorbers and scintillating tiles as active material, that are oriented parallel to the beam axis. It is segmented into two section, an inner one close to the beam pipe with square cells of length 131.3 mm and an outer section, with cells of length 262.6 mm, as it is shown in Figure 23. Readout cells are fixed in the rear side of the sampling structure and group together the wavelength shifting fibers, that run along the edges of the scintillator tiles, to one photomultiplier tube. The total depth of the HCAL is 5.6 λ, where λ is the interaction length and its energy resolution is

$$\frac{\sigma_E}{E} = \frac{80\%}{\sqrt{E}} \bigoplus 10\% \tag{3.4}$$

3.2.6 Muon systems

The muon system [60] consists of five stations and covers the LHCb angular acceptance between 20 mrad and 306 mrad in the bending plane and between 16mrad and 258 mrad in the non bending plane. The first station (M_1) is located at 12.1 m from the interaction point, in front of the calorimeter system, and is important for the transverse momentum measurement of the muons track used

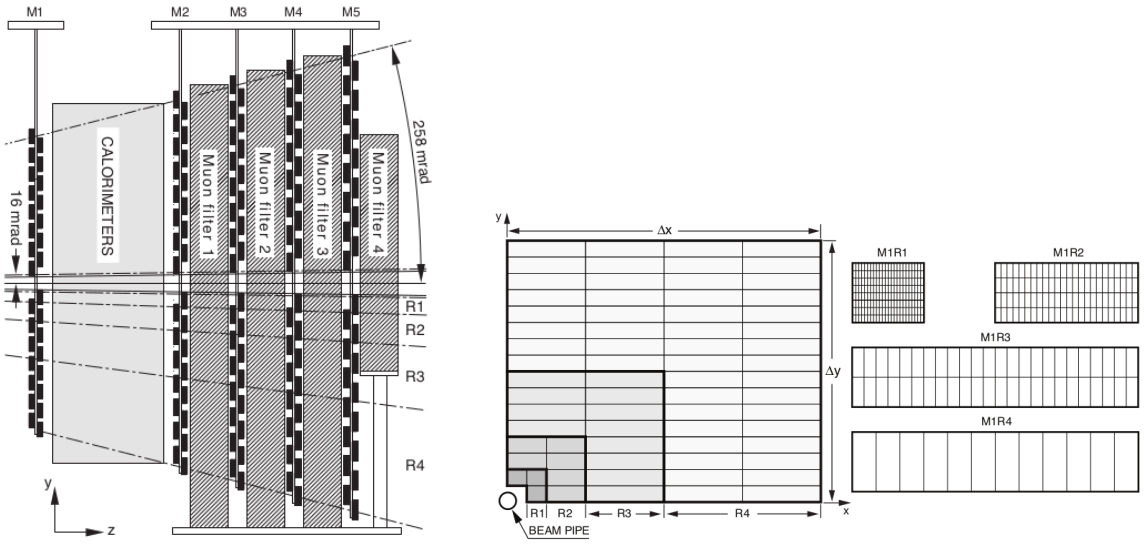

Figure 24: Left: overview of the muon sub-detector in the y- z plane. Right: layout of the muon stations [53].

in the L0 trigger 3.3.1. The other stations (M_2-M_5) are located behind the hadronic calorimeter, between 15.2 m and 18.8 m from the interaction point and are interleaved with three iron 80 cm thickness filters. The minimum momentum of a muon to cross the five stations is approximately 6 GeV/c since the total absorber thickness, including the calorimeters, is approximately 20 interaction lengths [53]. The layout of the muon stations is shown in Figure 24 left. In Figure 24 right is shown the four regions R_1 to R_4 in which all stations are divided into. The dimension of the regions and the segmentation of each region increase as the distance from the beam pipe becomes greater. In particular, the segmentation increases in a ratio 1:2:4:8, in order to keep almost constant the charged-particle flux in each region. Multi-wire proportional chambers are used for all regions except the inner region of station M1 where triple-GEM detectors are used.

3.3 Trigger

The bunch crossing rate is of around 40 MHz (30 MHz of visible interactions). The maximum rate at which all LHCb detectors were read out during the Run 2 operations, determined by the bandwidth and the frequency of the front-end electronic, was 1 MHz. A further reduction from 1 MHz to 12 kHz is required, at which rate the events are written to storage for further offline analysis [61]. All trigger systems consist of a set of algorithms that classify events as either interesting or uninteresting for further analysis, in order to keep interesting physics events only. A single trigger classification algorithm is called "line", so that a trigger consists of a set of trigger lines. The LHCb trigger system is divided into three levels: the first level is the hardware trigger, Level-0 (L0), and the last two levels are the software trigger, that are the High Level Trigger 1 (HLT1) and High Level Trigger 2 (HLT2). A Scheme of the LHCb trigger is shown in Figure 25. The hardware trigger reduces the rate from 40 MHz, to an output of 1 MHz. Afterward, the events are filtered by the High Level Trigger, resulting in an output rate of about 12 kHz.

3.3.1 Level 0 (L0) trigger

The L0 trigger is designed to perform a first filtering on events. It is composed by several trigger lines, that apply requirements on the energy deposited in the SPD, PS, ECAL and HCAL and the

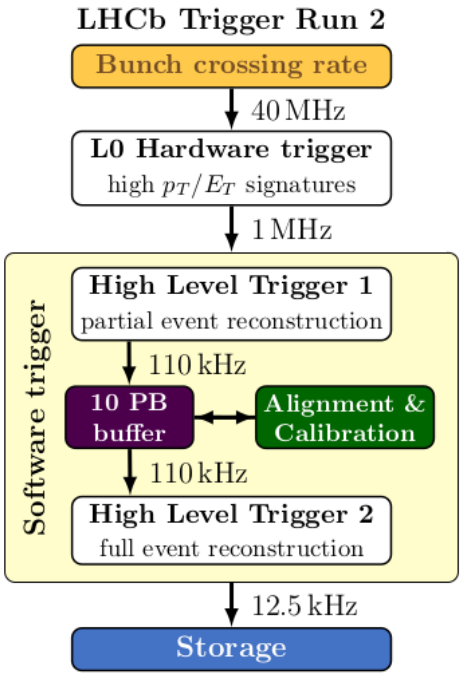

Figure 25: Overview of the LHCb trigger system [61].

muon stations to select events. Indeed energy deposits in the ECAL and HCAL are used to calculate the transverse energy, E_T, deposited in clusters of 2x2 cells defined as:

$$E_T = \sum_{i=1}^{4} E_i sin\theta_i \tag{3.5}$$

where E_i is the energy deposited in the cell i and θ_i is the angle between the z axis and a line from the cell centre and the interaction point. Information from SPD and PS systems is used to distinguish between hadron, photon and electron candidates. Tracks reconstructed in the five muon stations are also used to identify muons candidates. The track direction is used to estimate its p_T, assuming that the particle originated from the interaction point. In order to reject events with multiple interactions, the L0 trigger applies a cut on the number of hits inside the SPD, applying a so called *GlobalEventCut* (GEC). In general, once the GEC cut is applied, the event is accepted if one of the conditions required by the trigger lines is satisfied. The trigger lines are summarized in the following list, specific thresholds are summarized in table 9.

- $L0-Muon$ and $L0-Dimuon$: The trigger decision is based on the two muon candidates with the largest p_T. $L0-Muon$ requires that the largest p_T must be above a certain threshold, the $L0-Dimuon$ lines required that the product of the largest and second largest $p_T^1 \cdot p_T^2$ values must be above a certain threshold;
- $L0-Photon$ a particle release a cluster in the ECAL and has a E_T above a given threshold;
- $L0-Hadron$ a particle release a cluster in the HCAL and has a E_T above a given threshold;
- $L0-Electron$ a particle has a E_T above a given threshold and before leaving a cluster in the ECAL it hits both the SPD and the PS.

In this thesis simulation samples of Standard Model Z and H decaying into b or c dijets processes will be used. As will be shown in section 5.5 the *GlobalEventCut* will have a non-negligible effect in reducing the signal efficiency.

line	p_T [GeV/c]	$p_T^1 \cdot p_T^2$ [GeV^2/c^2]	E_T [GeV]	nSPD
$L0 - Muon$	> 1.76	-	-	-
$L0 - Dimuon$	-	$>(1.6)^2$	-	900
$L0 - Hadron$	-	-	> 3.7	450
$L0 - Photon$	-	-	> 3	450
$L0 - Electron$	-	-	> 3	450

Table 9: L0 trigger thresholds considered in this thesis.

3.3.2 High Level Trigger (HLT)

Once selected by L0, an event is transferred to the Event Filter Farm (EFF) for further selection. The HLT trigger system reduces the rate from 1MHz to a final rate of 12.5 kHz and is divided into two stages:

- HLT1 trigger: partially reconstruct the L0 accepted events. In particular, at this stage trajectories of long tracks, that are the charged particles that traverse the full LHCb tracking system (see section 4.1 for details), are reconstructed. In addition, it used VELO tracks to reconstruct primary vertices and performs muon identification. It reduces the L0 output rate to 110 kHz.

  The trigger lines used in this thesis are used to select events with pairs of jets from heavy quarks. The HLT1 trigger lines used for these purpose are listed in the following with a brief description. Detailed description of the requirements applied by these lines can be found in [62]:

  - *Hlt1TrackMVA*, *Hlt1TwoTrackMVA*,*Hlt1TrackMVATight*,*Hlt1TwoTrackMVATight*: single track lines select displaced tracks with respect the primary vertex with high transverse momentum. The two-track line makes a combination and selects based on the χ^2 of the vertex fit, displacement and summed combination transverse momentum.
  - *Hlt1TrackMuon*, *Hlt1DiMuonHighMass*, *Hlt1DiMuonLowMass*, *Hlt1SingleMuonHighPT*, *Hlt1DiMuonNoL0Decision* that are specialized lines for reconstructing muons;

- The HLT2 trigger [63] exploits the full information from subdetectors to full reconstruct track reconstruction of charged particles, reconstruction of neutral particles and particle identification. It runs inclusive and exclusive algorithms to select b-hadron events. The HLT2 trigger lines used in this thesis look for:

  - events with at least two jets in the final state with $p_T > 17$ GeV, containing a secondary vertex each (*HltQEEJetsDiJetSVSV*);
  - events with at least two jets in the final state with $p_T > 17$ GeV, in which at least one jet in the final state contains a secondary vertex (*HltQEEJetsDiJetSV*).

  More details about jet reconstruction and tagging algorithm will be given in the next chapter.

3.4 Simulations

The main elements that are necessary for a physics analysis at LHCb are the data and the Monte Carlo (MC) simulations. In particular, tracks, jets and particles arising from real or simulated proton-proton collisions have to be reconstructed by using software tools and processed in order to make them available for the analyses. The LHCb framework and data flow consist of several steps:

1. The first step is the acquisition of the energy released by particles from real data or Monte Carlo simulations. For the former, the energy is released by the outgoing particles from the real proton-proton collisions and is acquired by the LHCb detector. For the Monte Carlo simulations, dedicated software packages are used to generate the events and simulate the interaction of particles with the LHCb detector. The simulation is performed by the software GAUSS [64]:

    - the hard proton-proton interaction is simulated by Monte Carlo generators like MADGRAPH [65] [66] or PYTHIA [67], that is used also to simulate the parton shower in the final state quarks fragmentation;
    - particles are propagated to the detector and decays of hadronic particles are simulated using EvtGen;
    - the simulation of the interaction of final state particles through the detector material is simulated by using the GEANT4 [68] software;
    - the simulated hits are converted to signals that mimic the real detector digitization of the front-end electronics, the noise and the inefficiencies by the BOOLE [69] application.

2. In the second steps the software MOORE [69] is used to trigger data or simulated events that pass specific requirements.

3. Data that pass the trigger step and Monte Carlo events are fully reconstructed by the BRUNEL [69] software by applying the reconstruction algorithms that are described in the next chapter 4.

4. the software DAVINCI [69] is used on simulated and real data decay to reconstruct decay chains and jets, combining the objects reconstructed by BRUNEL and applying requirements on them. The software DaVinci allows to apply further selections on data, depending on the object reconstruction and their features, called stripping lines. This will be used in data selection (see section 5.2.4).

3.5 The LHCb experiment upgrade

Figure 26 shows the timeline of the future upgrades foreseen for LHC and LHCb. Two phases of data taking are foreseen with the Upgrade I [70] detector: the Run 3, that is currently undergoing, and the Run 4, that will start in 2027 after the Long Shutdown 3 (LS3).

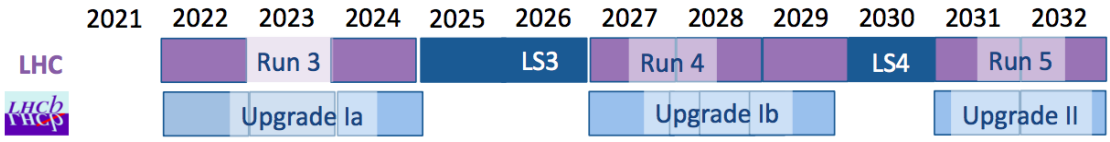

Figure 26: Timeline of the future upgrades of the LHC and the LHCb upgrades [71].

The Upgrade Ia detector includes a new tracking system based on a silicon pixel Vertex Locator detector, a silicon strips upstream tracker and the scintillating fibers. The two latter detectors will substitute the TT and the T1,T2,T3 tracker stations, respectively. The new VELO detector will be closer to the interaction point (at 5.1 mm) and will allow to deal with the pile-up at Run 3 and Run 4 (around 5 times larger than the pile-up in Run 2), while keeping performance comparable to Run 2. Another important improvement that have been done is the removal of the L0 hardware trigger, and the implementation of a fully software trigger. The calorimeter modules and muon chambers are the same as the one used in Run 2 and described in this chapter. During the Upgrade Ib small modifications at the detector are foreseen, like the substitution of calorimeter modules close to the

beam pipe due to radiation damage and the replacement of the RICH electronic. The HL-LHC phase, that will start in 2031 (Run 5), will be preceded by the Upgrade II phase [72]. At the end of the HL-LHC phase, the LHCb experiment is expected to get a total integrated luminosity of $\mathcal{L}_{int}$= 300 fb^{-1}. The center of mass energy collisions will be expected to be increased to 14 TeV during the HL-LHC phase, while instantaneous luminosity will be increased to $\mathcal{L} = 1.5 \cdot 10^{34}$ cm $^{-2}$ s $^{-1}$. The number of visible interactions is expected to be $\sim$40 per bunch crossing. The current goal for the HL-LHC phase is to equipe the VELO detector with timing information: a resolution of few tens of picoseconds would allow to keep the detector performance at the current level while allowing to deal with the improved pile-up for a correct matching between the primary and secondary vertices. As the other subdetector systems, even the ECAL will face high radiation levels, particularly in the region close to the beampipe and larger pile-up. Studies are ongoing the search for new materials and geometries which guarantee radiation tolerance. Finer granularity and the exploitation of the measurement of timing information will be necessary to maintain the same performance as the actual calorimeter. For what concerns the muon system, studies have shown that shielding will be necessary to protect the muon detector from the larger radiation levels. The current proposed solution is to remove the Hadron Calorimeter (HCAL) during LS3 and substituting it with a shielding layer. The primary purpose of the HCAL is to give input to the L0 trigger decision in the current detector. However, since the beginning of Run 3, the L0 trigger has been removed in favour of a fully software trigger. On one side, the absence of the L0 trigger will allow to remove the requirements on the Global Event Cut, that, as will be discussed in section 5.5, reduces the selection efficiencies. On the other side, the substitution of the HCAL with a shielding layer will affect jet reconstruction performance, since, as will be shown in section 4.4 almost 10 % of the particle content of the jet comes from neutral hadrons.

4 Objects reconstruction at LHCb

The experimental signatures of quarks and gluons produced in high energy experiments are jets, collimated streams of particles produced in the fragmentation and hadronization process. In the analysis presented in this thesis events with two jets in the final state containing a secondary vertex (SV) detached from the primary vertex due to the decay of a heavy flavour quarks are selected. In this chapter, the algorithms used to reconstruct tracks and jets at LHCb are described with their performances. Then, the Secondary Vertex Tagging algorithm, used to identify b and c quarks, by reconstructing secondary vertices is described. Last, a Deep Neural Network have been used to identify the jet flavour.

4.1 Tracking performances

Track trajectories are reconstructed at LHCb by combining hits released in the VELO and the Silicon Tracker. Tracks are divided into the following classes [53]:

- Long Tracks: traverse the full tracking system, have hits in all the tracking system detectors: the VELO, the TT and the T1,T2,T3 stations;
- Upstream tracks: have hits only in the VELO and TT stations. They are typically low momentum tracks since they are deflected out of the detector acceptance by the magnetic field;
- Downstream tracks: traversing only the TT and T1,T2,T3 stations. They are tracks produced, for example by the decay of K^0_S and Λ outside the VELO acceptance;
- VELO tracks: have hits in the VELO only, they are useful for the primary vertex reconstruction;
- T tracks: have hits only in the T1,T2,T3 stations and not in other tracking subdetectors.

Figure 27 shows a scheme of the LHCb detector, with the trajectories of the different kinds of tracks.

The track reconstruction steps [54] are here briefly described. Details can be found in references to the algorithms description papers.

- First, VELO tracks are reconstructed as straight lines trajectories having at least three hits in the R sensors and three hits in the ϕ sensors [73].
- Then, there are two algorithms that combine the information on the tracker system with the VELO tracks. The first algorithm, the forward tracking [74], combines the VELO tracks with the information from the T1, T2, T3 stations: a track trajectory candidate is first built by extending the VELO track seed up to a single T station hits, while further hits in the T stations are then searched along this trajectory. The second algorithm is called track matching, that combines tracks segments found in the T stations built by a standalone track finding algorithm, requiring at least two hits in each of the three stations, with the VELO short tracks.
- The tracks candidates found with the two algorithms are then combined, duplicates are removed and the final set of long tracks is defined. The TT hits consistent with the trajectories on the final set of long tracks are then added.
- Two other algorithms are used to build the downstream and upstream tracks. Downstream tracks candidates are built searching for at least three hits in the TT [75] that match to tracks

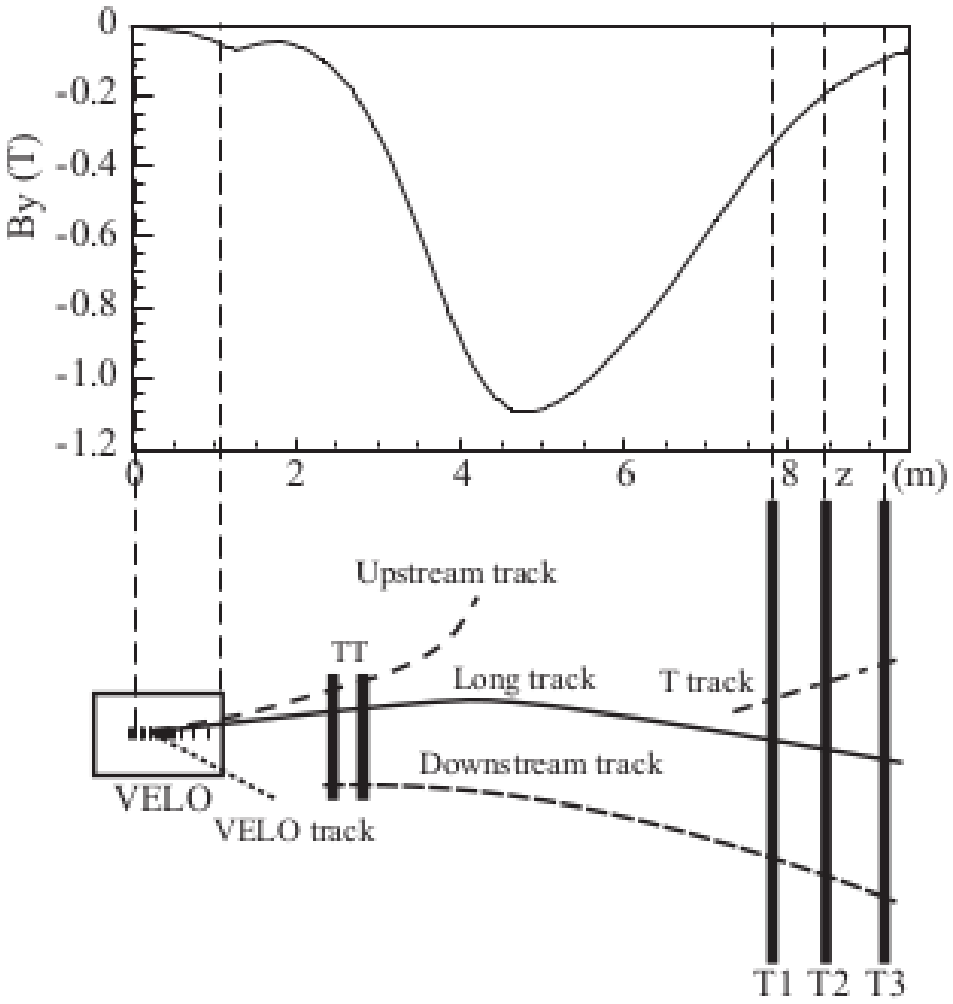

Figure 27: Scheme of the various kinds of tracks reconstructed at LHCb: long, upstream, downstream, VELO and T tracks. In the above plot, the B field component in the y direction is also shown as a function of the z coordinate [53].

segments built in the T stations, and taking into account the presence of the magnetic field in the extrapolation.

- Upstream tracks are built with a similar procedure [76], extrapolating VELO tracks up to the TT and searching for at least three matching hits. The hits used by the forward algorithm are neglected in the hit search performed to build the upstream or downstream tracks.

- As final step, the tracks are fitted using a Kalman filter [77] [78]. The VELO and T seeds that have not been used as part of either a long, upstream, or downstream track, are defined to be of type VELO or T track.

The momentum resolution ($\delta p/p$ for long tracks is determined considering $J/\psi \to \mu^+\mu^-$ decays [79]. The tag-and-probe method has been used: one of the muons, the "tag", is required to be fully reconstructed as a long track, while the other particle is identified as the "probe". Their invariant mass is required to be between 2.6 to 3.6 GeV/c^2, in order to reconstruct the J/ψ mass peak. The reconstruction efficiency for long tracks can be measured by matching the probe track to a fully

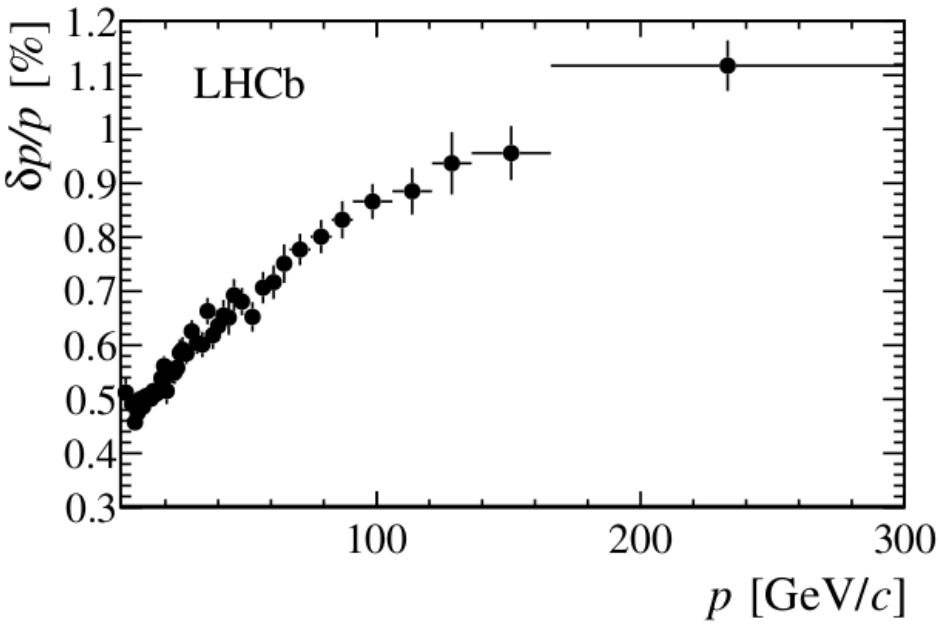

Figure 28: Relative momentum resolution versus momentum for long tracks in data obtained using J/ψ decays [54].

reconstructed long tracks, and is found to be over 95%. The momentum resolution for long tracks has been measured by reconstructing the invariant mass of the two muons and is calculated as

$$\left(\frac{\delta p}{p}\right)^2 = 2\left(\frac{\sigma_m}{m}\right)^2 - 2\left(\frac{p\sigma_\theta}{m\theta}\right)^2 \tag{4.1}$$

where m is the J/ψ mass, σ_m is the Gaussian width obtained from a mass fit and the second term corrects for the opening angle θ between the two muons where σ_θ is the mean per-event error on θ obtained from the track fits. The relative momentum resolution is shown in Figure 28 as a function of the momentum p. It is about 5 per mille for particles below 20 GeV/c and rises about 8 per mille for particles around 100 GeV/c.

4.2 PID variables

Two different typed of PID variables are used in LHCb: the local particle Identification (PID) and the global particle identification (Global PID).

4.2.1 PID

The PID, also called difference log-likelihood (DLL) is provided by the calorimeters, the muon and the RICH sub-detectors [80] [54]. The RICH detectors discriminators is based on the log-likelihood difference $\Delta LL(K-\pi)$ between K and π by building a set of likelihoods relative to the pion mass hypothesis. Furthermore it contribute to the identification of charged leptons (e-μ) complementing the information from calorimeter. Performance in particle identification have been determined on data, using $K^0_S \to \pi^+\pi^-$, $\Lambda \to p\pi^-$ and $D^{*+} \to D^0(K^-\pi^+)\pi^+$. Figure 29 shows the K identification efficiencies and π misidentification rates as a function of the particle p_T, for different $\Delta LL(K-\pi)$ thresholds. The K identification efficiency is defined as the number of K particles correctly identified over the total number of detected K, while the π misidentification rate

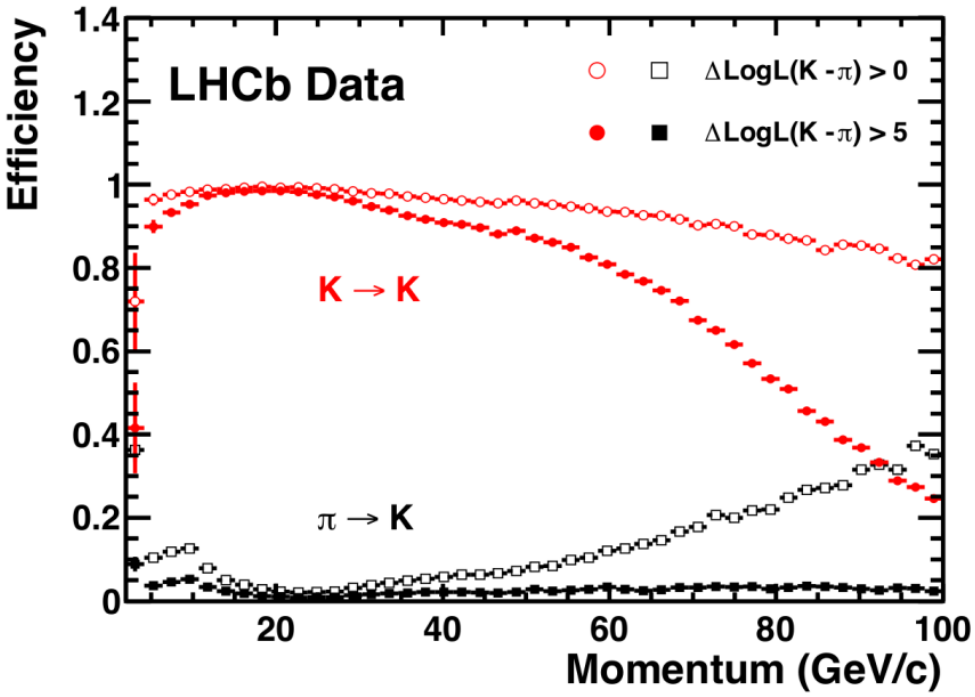

Figure 29: Kaon identification efficiency and pion misidentification rate as measured using data as a function of track momentum [54].

is defined as the number of π wrongly identified as K over the total number of π. As can be seen, the average kaon efficiency for $\Delta LL(K-\pi) > 0$ is of about 95% and the pion misidentification rate is of about 10%, while for a stricter PID requirement $\Delta LL(K-\pi) > 0$, the pion misidentification can be reduced to 3% with a kaon efficiency of 10% on average.

As mentioned, the Calorimeters provide identification of electrons, photons and hadrons in addition to the measurement of their energies and positions. The distinction between charged and neutral particles is performed by studying the presence of absence of tracks in front of the energy deposits in the calorimeter system. Energy deposits relative to neutral particles (photons and π^0) candidates can be distinguished by using the shape of the cluster, and taking into account the hypothesis that the photons may have converted by interacting with the detector material upstream the calorimeter. Two photon hypothesis likelihoods are then built:

- one for non-converted photons, where the difference in log-likelihood is calculated between the photon and the background hypothesis, considering several variables, like the energy deposited in the PS in front of the ECAL cells, the χ^2 between the cluster and any track, the ratio between the energy at the center of the ECAL cluster and the total ECAL energy. A neural network classifier is then used to discriminate between photons and high energy merged π^0 candidates.

- one for converted photons: in this case photons are reconstructed form electron-positron tracks. They are selected on the basis of the electron PID variable, that will be described later, and requiring a minimum p_T and E/pc value within a selected range. Cluster relative to electron-positrons pairs that are close enough and satisfy certain requirements on their transverse momenta and their reconstructed vertex positions, are combined by a proper algorithm. The electron energy is also corrected by including any bremsstrahlung photons on the calorimeter compatible with the electron/positron pair-

For energy deposits relative to charged particles, the electron hypothesis is constructed to distinguish electrons from hadrons, using information derived from ECAL, PS and HCAL. For the ECAL, the log-likelihood difference for electron and hadron hypothesis $\Delta LL^{ECAL}(e-h)$ is computed based on both E/pc and the χ^2 between the cluster and any track estimator. For the PS $\Delta LL^{PS}(e-h)$ and the HCAL $\Delta LL^{HCAL}(e-h)$ are built using the energy deposits in each sub-detector. The overall estimator for the calorimeter system is calculated as the sum of the three $\Delta LL(e-h)$

The identification of a track reconstructed in the tracking system as a muon is based on the association of hits in the muon system around its extrapolated trajectory. Likelihoods for the muon and non-muon hypothesis are computed, based on the average square distance of the hits that are closest to the extrapolation points. The performance in muon identification have been obtained from data $J/\Psi \to \mu^+\mu^-$, while proton misidentification has been evaluated on $\Lambda \to p\pi^-$ events and kaons and pions misidentifications have been evaluated on $D^0 \to K^-\pi^+$. The muon identification efficiencies are found to be over 95% in the momentum range between 20 and 100 GeV, with a misidentification rate below 0.6 and 1% for proton and pions and kaons respectively.

The PID informations in terms of DLL obtained separately from the muon, RICH and calorimeter systems are added linearly to obtain a overall measurement. The electron identification efficiency is 91.9% with DLL calo (e - h) > 2 with a hadron misidentification rate of 4.5%. Including the information from the RICH detectors the electron efficiency raises up to 97% while the misidentification rate becomes lower than 2%.

4.2.2 Global PID

In the Global PID [80] [54] [81] approach is a multiclass machine learning algorithm called ProbNN based on six binary fully-connected neural networks with a single hidden layer implemented in the TMVA library. Each of these networks corresponds to one particle type and is trained to separate a type of particles from the others.

It identifies the charged particle type associated with a given track that are: electron, muon, pion, kaons, protons and ghost tracks, that are charged tracks that arise from errors of the tracking algorithm and are not associated to any real particle passing through the detector. As the PID, it exploits information from the tracking system, the RICHs, calorimeters and muon chambers. Around 50 variables are given as input to train the ProbNN and to classify particles are related to the track kinematic characteristics, and quantities related to their fits (like χ^2), their type (long, downstream, upstream), and binary indicators that refer to wether a track falls into the geometric acceptance of the calorimeter or muon systems. Furthermore, variables related to the RICHs system are given, like the DLL^{rich} quantities described above and a binary indicator that indicate whether the track momentum is above momentum threshold for the different particles type to produce Cherenkov light. For what concerns the calorimeter DLL_{calo} quantities related to the categories mentioned above, the clusters energy in the different calorimeter subdetectors and χ^2 for track/cluster matching values. From the muon systems the DLL_{muon} is provided together with other information related to the hits in the muon chambers. The improvement of the multivariate approach with respect to the simple log likelihood can be seen in Figure 30, where the misidentification rates versus muon and proton identification efficiency curves for the $\Delta LL(X-\pi)$ (black) and ProbNN (red) are shown.

4.3 Primary Vertex and Impact parameter reconstruction

The primary vertex reconstruction algorithm is described in detail in [82]. It consists of two steps:

- Seeding: in this first steps candidates for primary vertices are built, grouping tracks passing close enough to a common points. In practise, a loop over all tracks is performed and for each

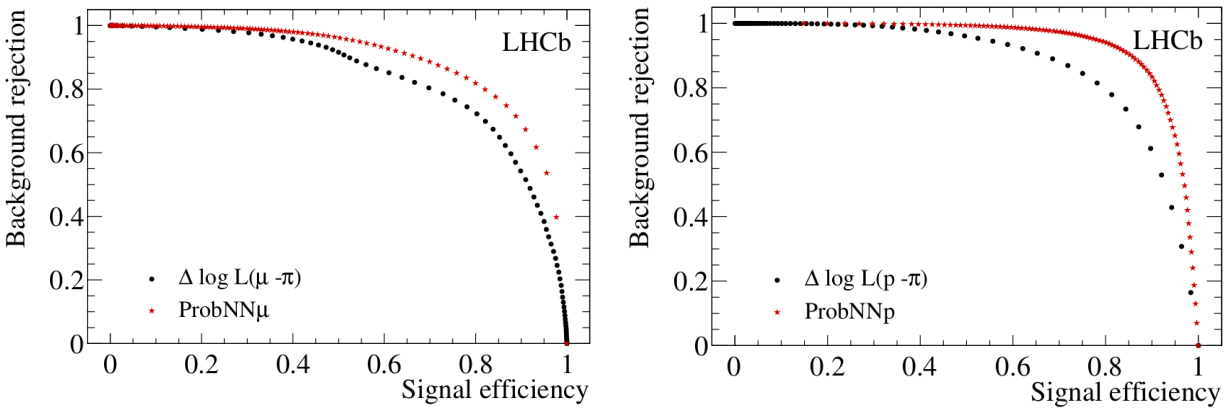

Figure 30: Background misidentification rates versus muon (left) and proton (right) identification efficiency, as measured in the $\Sigma^+ \to p\mu^+\mu^-$ decay study. The variables $\Delta LL(X - \pi)$ (black) and ProbNN (red), the probability value for each particle hypothesis, are compared for 5 - 10 GeV/c muons and 5 - 50 GeV/c protons, using data sidebands for backgrounds and Monte Carlo simulation for the signal [54].

one the number of close tracks is determined. The "close" condition is verified if the distance between the closest points of the two tracks is less than 1 mm. If a track have more than four close tracks, the position of the primary vertex candidate (seed) is determined with a two steps approach: first, the point of closest approach is calculated for each tracks pair in the seed. Then, their weighted average is determined and it is taken as the final position the the PV candidate.

- Fitting: the position of the reconstructed primary vertex is determined by the least square method minimizing the following χ^2_{PV}

$$\chi^2_{PV} = \sum_{i=1}^{n_{tracks}} \chi^2_{PV,i} \cdot W_{T,i} \tag{4.2}$$

where $\chi^2_{PV,i}$ denotes χ^2 of the track impact parameter with respect to the PV, calculated as the value by which the χ^2_{PV} increases when adding the track to the vertex and $W_{T,i}$ is a weight factor that is close to one for tracks with impact parameter with respect to the PV close to zero and decreases as χ^2_{PV} increases [82]. This strategy allows giving lower weights to tracks from secondary vertices wrongly associated to the primary vertex, and badly reconstructed tracks.

Figure 31 shows in the left the primary vertex resolution in x (red) and in y (blue) as a function of the number of tracks used to reconstruct the primary vertex obtained with 2012 data [54]. Typical values for the primary vertex resolution at LHCb where the average number of visible proton proton interactions per event is around 1.3, calculated with 2012 data and the average number of tracks in a minimum bias event containing one PV is 55, are 8-10 μm in the x and y coordinates and 50 μm in z. As the number of reconstructed PVs in the event increases, the resolution degrades with a rate of approximately 5–10 % per additional vertex [55].

The impact parameter (IP) of a track is defined as its distance of closest approach from the primary vertex. b and c hadrons produced from pp collisions have a hard momentum spectrum in the forward region [83], their average momentum is $\sim$ 80 GeV/c. The average decay length of b-hadrons produced at the LHC is about 1 cm, as will be explained in section 4.9.1, so they decay product have a large impact parameter (IP) with respect to their primary vertex (PV).

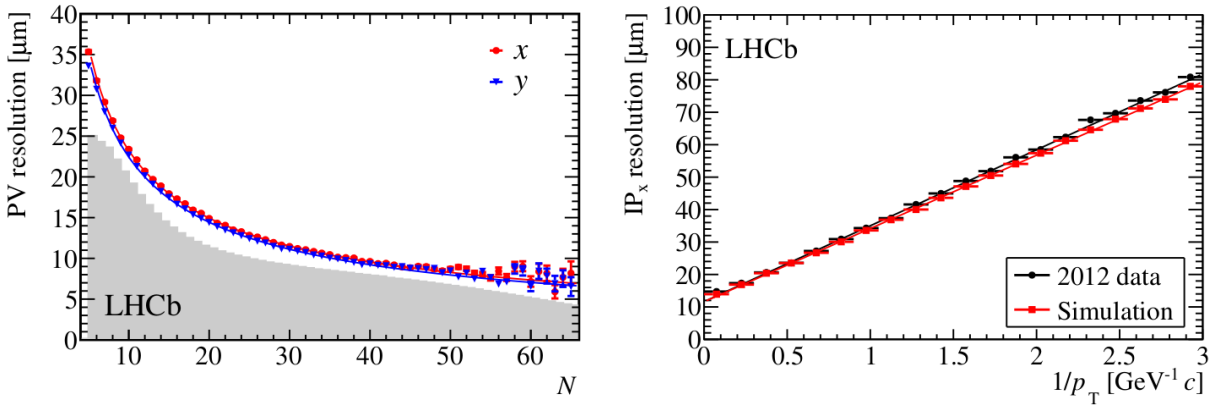

Figure 31: Left: primary vertex reconstruction as a function of track multiplicity in the x and y directions. The resolution in x is shown in red, while the resolution in y is shown in blue. The gray histogram shows the distribution of number of tracks per reconstructed primary vertex for all events that pass the high level trigger [54]. Right: Impact parameter resolution in x as a function of $1/p_T$, where data (black) are compared to simulation (red) [55].

As will be shown in 4.9.1, selection on the IP and the χ^2_{IP}, defined above, are used at LHCb to reduce the contamination from prompt backgrounds in the reconstruction of displaced vertices due to the b and c hadrons decay. The main aspects that affect the IP resolution are: multiple scattering of particles when they pass through the detector material, the hits spatial resolution used to build the track and the distance of the first hit of the track from the PV. As described in 3.2.1 the VELO detector is optimize to reduce these contribution as much as possible, since the sensors are positioned very close to the beams (8 mm for the innermost ones), separated from them by only a thin aluminum foil and the detector provides high-precision hit position measurements. The IP resolution has been measured using good quality long tracks from events with only one reconstructed PV, obtained by fitting at least 25 tracks in order to minimize the contribution of the vertex resolution to the measured IP. Figure 31 left shows the IP projected in the x direction (IP_x) resolution as a function of $1/p_T$. The resolution of IP_y is compatible to the resolution on IP_x identical. The resolution on the 3 dimensional IP can be obtained from the IP_x (see [84] and [85]) and is found $15{+}29/p_T[GeV]$.

4.4 Jets reconstruction algorithms

In a jet, the energy is carried by charged particles, photons, long-lived neutral hadrons and neutrinos.

Figure 32[86] shows the particle composition of a jet according to the Monte Carlo truth as a function of the jet p_T (left) and η (right), obtained from a minimum bias simulated sample at 7 TeV center of mass energy. It can be seen that around 60% of jets constituents are charged particles, detected by the tracking system, around 30% are neutral pions or photons detected by the ECAL and 10% are neutral hadrons detected by HCAL.

The correct reconstruction of particles energy and momentum is fundamental in order to obtain a good jet energy resolution. In a full calorimeter reconstruction approach the jet energy is obtained from the sum of the energies deposited in the electromagnetic and hadronic calorimeters (ECAL and HCAL). In this case the jet energy resolution is limited by the poor energy resolution of the LHCb hadronic calorimeter, is mainly used for triggering on high p_T hadrons. However, as shown in the detector section, the LHCb tracking system allows to reach excellent charged particles momentum resolution (~ 0.5 %) with high reconstruction efficiency (97 %). For this reason, as first step for jet

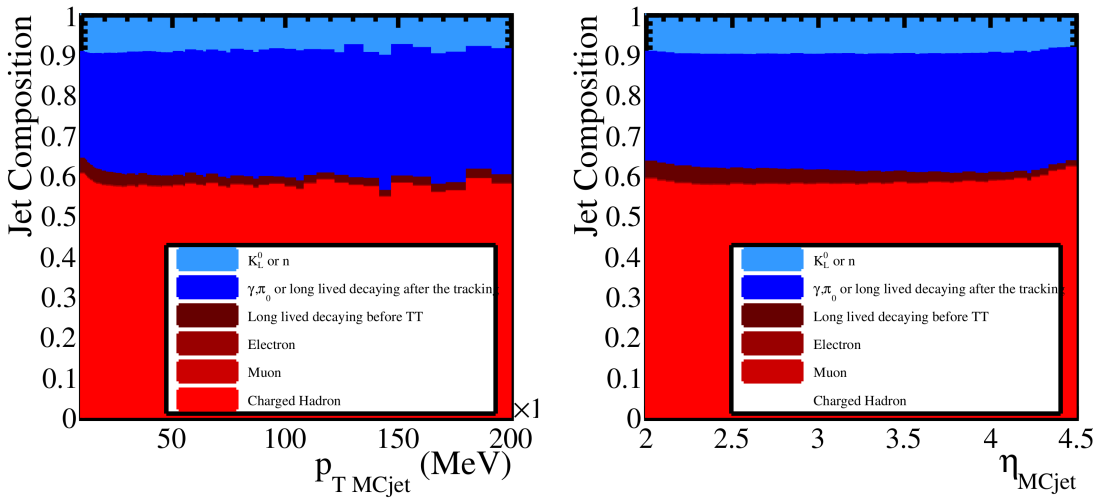

Figure 32: Composition of MC jets according to the MC truth as a function of jet p_T and η [86].

reconstruction at LHCb the tracking system and the calorimeters measurements are combined, by mean of the Particle Flow [87] approach. The LHCb jets reconstruction algorithm [88] has several steps, that will be described in the following:

- Particle Flow algorithm: select input particles by applying specific criteria to tracks and calorimeter clusters;
- Anti-k_T algorithm: particles are selected by the Particle Flow are clustered in jets;
- E-recombination scheme: jet four-momentum is computed as the sum of the jet constituents four-momenta;
- Jet Energy Correction: the jet four-momentum is multiplied by a correction factor that depends on the jet kinematic.

4.5 Particle flow

The Particle Flow algorithm used at LHCb has been described in detail in [86] and [89]. The output of the Particle Flow (PF) algorithm applies selection requirements to jets and tracks before they are given as input to the jet clustering algorithm. Tracks are selected as "charged particles" by the PF algorithm, calorimeter clusters associated to at least one track are selected as "not-isolated neutral particle", while the calorimeter clusters not associated to tracks are selected as "isolated neutral particles".

4.5.1 Charged particles selection

The tracks selection is performed by applying quality requirements to the four categories candidate tracks defined in section 4.1: *long*, *upstream*, *downstream* and *VELO*.

The requirements listed in Table for each kind of track 10:

where:

- p_T is the transverse momentum of the track. Even if the VELO tracks have not a transverse momentum associated to them, they are included among the inputs tracks to the jet clustering algorithm since they are fundamental to associate a primary vertex to jets;

	long	downstream	upstream	VELO
p_T [MeV/c]	-	-	>100	-
χ^2	< 5	< 1.5	< 1.5	< 10
P_{ghost}	<0.5	-	-	-
$\sigma(q/p)/(q/p)$	>10	>10	>2	-

Table 10: List of requirements for all tracks categories in the Particle Flow.

- χ^2 is obtained from the track Kalman fit;
- P_{ghost}, that is the probability for a track to be a fake track, i.e. not associated to a real particle;
- $\sigma(q/p)/(q/p)$ is the momentum resolution: q is the particle charge and p the particle momentum.

By using the information from RICHs, a particle type and mass is also assigned to each track. They can be classified as p/p^-, $\Pi^\pm$, $\mu^\pm$, $e^\pm$, $K^\pm$.

4.5.2 Isolated neutral particles

Isolated neutral particles are identified when calorimeter cluster are detected in the calorimeter system but no tracks are associated to them. If they are detected by the ECAL, they are mainly photons or pairs of photons produced by the π^0. In the latter case, the photon pair can be detected as *merged*, when they are almost collinear and produce a single cluster in the ECAL, or *resolved*, when they are detected as separate clusters. The requirements applied to select ECAL clusters that are then given as input to the jet clustering step, are summarized in the following Table 11:

	merged π^0	resolved π^0	photons
E_T [MeV]	-	-	>200
PhotonID		> -4	> -1 (>-2 with T track)
PhotonID for 1 γ	-	> -2	-
$\chi^2_{track-cluster}$	> 25	>25	>25 (>16 with T track)

Table 11: List of requirements for isolated neutral particles.

where:

- E_T is the cluster transverse energy;
- A likelihood is computed to label the ECAL clusters as the photons-like clusters and π^0-like clusters. *PhotonID* is the likelihood computed to label the ECAL clusters as the photons-like clusters;
- $\chi^2_{track-cluster}$ is a χ^2 associated to each track-cluster combination to evaluate a cluster is likely to be associated to the track;
- requirements to photon clusters associated to T tracks are also reported. This is the case in which a photon split into e^+/e^- pair when interacts with the material in upstream the calorimeter.

If isolated clusters are detected by the HCAL, $\chi^2_{track-cluster}$>25 is required for HCAL cluster with energy below 10 GeV, and $\chi^2_{track-cluster}$>15 for energies above 10 GeV.

4.5.3 Non-isolated neutral particles

Calorimeters clusters associated to a track are selected by applying the following requirements:

- First, a requirement on the $\chi^2_{track-cluster}$ is applied: $\chi^2_{track-cluster} < 25$ is required for ECAL clusters, $\chi^2_{track-cluster} < 25(16)$ for energies below (above) 10 GeV is required for HCAL clusters;
- ECAL and HCAL clusters are grouped such that different groups do not share the same associated tracks;
- E/p, where E is the cluster energy and p the track momentum associated to the cluster is parametrized as a function of p, with a procedure called E/p calibration, that has been performed on data sample of pp collisions at 7 TeV, and is described in detail in [89]. The calibration have been calculated for all the possible tracks-cluster object categories, for example: the track is associated to a cluster in the ECAL but have no clusters in the HCAL and viceversa, the track have clusters in both the calorimeters systems. The expected energy released in the calorimeters by a charged particle with momentum p pointing to a cluster group with energy E is calculated. The cluster group is discarded if the total expected energy of the cluster group (E^{exp}) is larger than 1.8 times the measured energy (E^m), while if $E^{exp}<$ 1.8 E^{exp} then E^{exp} is subtracted from E^m;
- the remaining energy is selected as non-isolated neutral particle if its E^T is greater than 2 GeV and it is used as input in the jet clustering.

4.6 Jet clustering with anti kt algorithm

Tracks selected by the Particle Flow are associated to the primary vertex with the primary vertex reconstruction algorithm described in section 4.3. Then, tracks associated to the same PV are given in input to the jet clustering algorithm. There are two main classes of jet algorithms in use: the cone algorithms and the sequential clustering algorithms. The sequential clustering algorithms, that is the one used in both the analyses of this thesis, assumes that particles within jets have small differences in transverse momenta, then it groups particles based on their momenta. The result is a jet with fluctuating shape in the η-ϕ space. Two important characteristics of all the sequential clustering algorithms are the infrared (IR) and the collinear (UV) safety. From the practical point of view these characteristics ensure that the jet definition is insensitive to the soft gluon emissions and the collinear gluon emissions, respectively. All sequential clustering algorithms follow a similar procedure for jet clustering:

1. a list of all the input particles is created
2. For each pair of particles i, j the distance d_{ij} is calculated:

$$d_{ij} = min(p^a_{ti}, p^a_{tj})\frac{\Delta R^2_{ij}}{R^2} \tag{4.3}$$

   where $\Delta R^2_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, a is an exponent corresponding to a particular clustering algorithm, p_{ti},y_i and ϕ_i are the transverse momentum, rapidity and azimuth of the particle i. R is a jet-radius parameter, that will be discussed later. In the anti-k_t a is equal to -1. For a = 1 the k_t algorithm is obtained while for a=0 the Cambridge/Aachen algorithm is obtained;
3. the distance between each particle i and the beam is calculated with

$$d_{iB} = p^2_{ti}$$

4. The minimum d_{min} of all the d_{ij}, d_{iB} is found. If d_{min} is a d_{ij}, particles i and j are merged into a single particle, summing their four-momenta. If d_{min} it is a d_{iB} then the particle i is declared to be a final jet and is removed from the list.

5. if no more particles are in the list, the algorithm ends, otherwise it restarts from step 2.

   Figure 33 shown the jet clustering in the rapidity-ϕ space, performed on the same data and with the same input radius with the different sequential clustering algorithms. The k_T algorithm orders particles in crescent p_T: the dominance of the low p_T is shown in 2 for a=1, means that the algorithms will cluster soft particles first, and hard particles will tend to be clustered with a soft particle instead of to other hard particles. The anti-k_T algorithm, that is the one used by LHCb, the dominance of the high p_T is shown in 2 for a=-1, means that the algorithms will cluster hard particles first, while soft particles will tend to be clustered with a hard particle instead of to other soft particles. For these reasons jets clustered with a k_T present irregularities due to soft emissions in the boundaries of jets, compared to jets clustered with anti-k_T algorithm. The algorithm is the Cambridge/Aachen (a=0) provides ordering by only using distance in the η-ϕ plane.

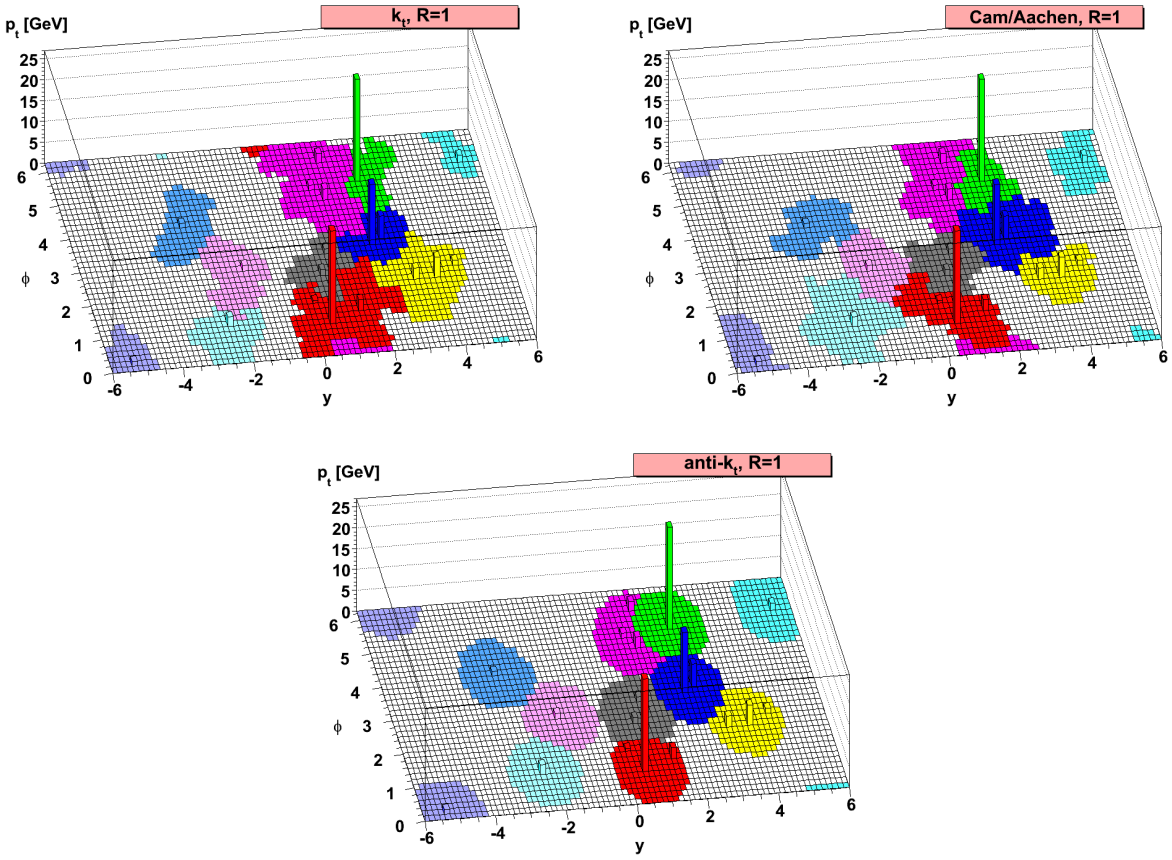

Figure 33: The k_T (top left), anti-k_T (bottom) and Cambridge/Aachen (top right) jet clustering, performed on the same data with the same radius parameter [90].

The last parameter that have to be chosen carefully for the jet clustering is the clustering radius R, since it has to be large enough to contain all the jet energy coming from the hadronization of the final state quarks, while minimizing the contamination from underlying event. At LHCb studies have been performed to determine the best R by minimizing the jet energy resolution and the jet identification efficiency [89]. The optimal radius have been found to be R = 0.5.

4.6.1 E-recombination scheme

Once the particles have been clustered, the jet 4-momenta is computed using the E-recombination scheme: considering the particle four-momenta $p_i = (E_i, \vec{p}_i)$ the E-recombination scheme the jet 4-momentum is computed as:

$$p_{jet} = (E_{jet}, \vec{p}_{jet})$$

where $E_{jet} = \sum_i E_i$ and $p_{jet} = \sum_i p_i$. The performance of the jet reconstruction algorithms have been evaluated by using simulated data. Two classes of jets are reconstructed:

- Monte Carlo jets (jet$_{MC}$) obtained by clustering with the ant-k_T algorithm the stable truth-level particles (with lifetime $\tau > 10^{-8}$s) are characterized by having the true kinematical quantities, as they are not affected by the detector reconstruction efficiencies;
- reconstructed jets (jet$_{reco}$) are reconstructed by using detector information.

The differences between jet energies of jet$_{reco}$ and jet$_{MC}$ are evaluated to assess the jet algorithm performance by using simulated samples at center-of-mass energy of 7 TeV in [88] and selecting the jet$_{MC}$ with a distance from the jet$_{reco}$ in the η-ϕ plane below 0.4 ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ <0.4).

4.7 Jet Energy correction

The reconstructed jet energy is in general different with respect to the energy of the associated Monte Carlo jet. For this reason a multiplicative factor k_{MC}, that takes into account the effects of the pile-up, the noise and the non-uniformity of the detector, is computed to correct the jet$_{reco}$ energy $E_{jet_{reco}}$:

$$E_{jet_{MC}} = k_{MC} E_{jet_{reco}} \tag{4.4}$$

It has been evaluated in simulated events of b, c, light quarks and gluons jets at a centre-of-mass energy of 13 TeV. A parametrization of k_{MC} is obtained as a function of jet p_T, η, ϕ, the fraction of charged particles in the jet (cpf), and the number of primary vertices (nPV), using a cubic function model. For example, Figure 34 shows the k_{MC} obtained using Run 2 simulations as a function of the jet p_T.

4.8 Jet Identification efficiencies

To improve the rejection of fake jets, jets originated from noise and high energy isolated leptons, the following cuts are applied:

- number of tracks in the jet associated to the primary vertex (nPVtrks) ≥ 2
- maximum fraction of transverse momentum carried by a single ParticleFlow particle, mpf $<$ 0.8;
- maximum transverse momentum carried by a track, mpt $\geq$ 1.2 GeV;
- fraction of charged particles on the jet, cpf > 0.1.

The jet identification efficiency is evaluated after these requirements using MC simulations of $Z \to \mu\mu + jet$ events as:

$$\epsilon_{jet} = \frac{n_{reco-jets}}{n_{true-jets}} \tag{4.5}$$

Figure 35 shows the jet identification efficiency as a function of the jet p_T as obtained from Run 1 simulations.

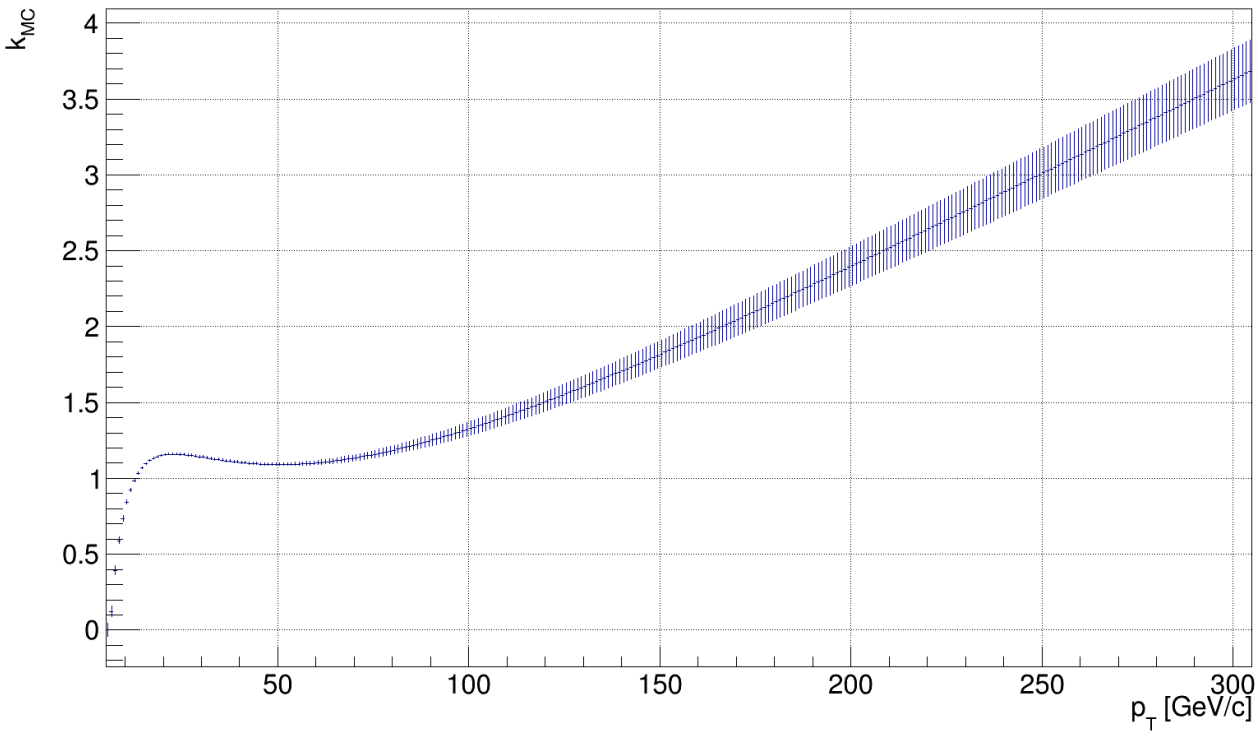

Figure 34: k_{MC} correction factor as a function of the jet p_T [91] calculated with Run 2 simulations.

4.9 Flavour tagging algorithm

The identification of the quark flavour that generates jets is a key aspect of the analysis performed in this thesis. The method used in this thesis to identify the heavy jet flavour is the Secondary Vertex tagging algorithm (SV tagging algorithm). It relies on the the fact that heavy flavour hadrons fly for a distance d before decaying, with $< d >= \beta\gamma c\tau$, where τ is their mean lifetime, c is the speed of light, $\beta = \frac{v}{c}$, where v is the particle velocity and γ is the relativistic factor. For this reason, displaced vertices with respect the primary vertices, that can be found by reconstructing the meson decay products, are searched. Furthermore, as will be described in section 4.10, SV related observables and global observables of the heavy flavour jet can be exploited to separate b jets from c jets.

4.9.1 SV tagging algorithm

The SV tagging algorithm used to reconstruct SV in this thesis is described in detail in [92]. The main steps are listed here:

1. long tracks with $p > 5$ GeV/c, $p_T > 0.5$ GeV/c, $\chi^2/N_{dof} < 3$, where χ^2/N_{dof} is given by the track fit are selected. The χ^2_{IP}, is required to be greater than 16, and the probability that the track is a ghost track $P_{ghost} < 0.3$.

2. All possible 2-tracks SVs combinations are built and the SV position is calculated with a fit. The two tracks associated to a SV are combined to form a particle which flight direction is defined as the vector that points the SV from the PV. Its four-momentum is defined as the sum of tracks four-momenta, assuming the π mass.

3. The following quality requirements are applied to the 2-body particles:

   - a distance of closest approach (DOCA) $d_{DOCA} < 0.2mm$;

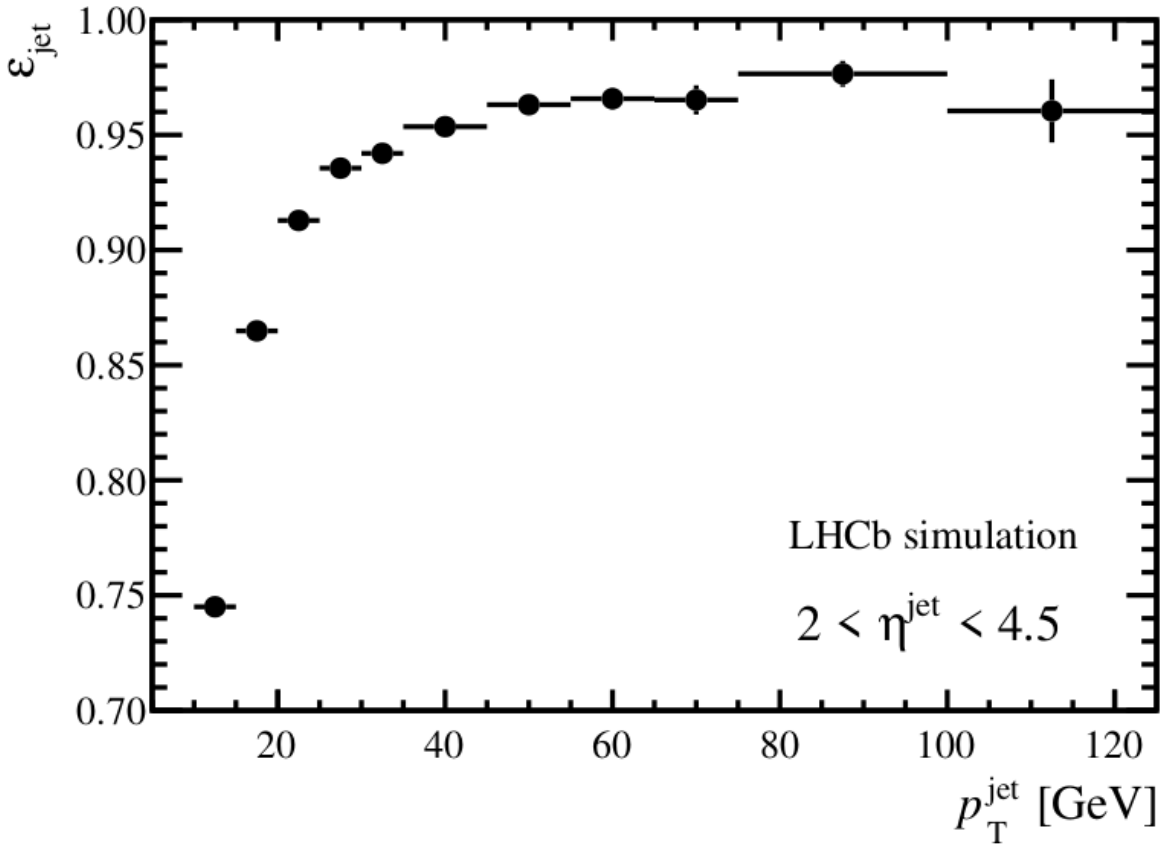

Figure 35: Jet identification efficiency as a function of the jet p_T [88].

- The fit of the secondary vertex chi-square $\chi^2_{SV} < 10$;
- the invariant mass between 400 Mev/c^2, to remove strange-hadrons decays and 5279.4 Mev/c^2, which is the mass of the B^0 meson. This upper requirement to the mass rarely remove the vertices from the B meson decay, since the π mass is assigned to all tracks.

4. The 2-tracks secondary vertices formed as described in the previous point are associated to jets: they are considered inside the jet if the distance between the SV and the jet axis in the $\eta - \phi$ plane is below 0.5. 2-tracks SV that are inside the same jets are merged if they have at least one track in common, forming n-tracks secondary vertices., that are called tag_{SV}. The tag_{SV} position is calculated as the weighted average of the 2-tracks SV position, where the weights are the χ^2 from the 2-tracks vertex fit. The tag_{SV} four-momentum is calculated as the the sum of tracks four-momenta, assuming the π mass.

5. Light jet contamination is reduced by applying requiring that the tag_{SV} has

   - $p_T > 2$ GeV/c,
   - a position along the z axis below 200 mm,
   - $d/p < 1.5$ mm/(GeV/c) where d is the flight distance, that is the distance between the SV and the PV,
   - the flight distance χ^2_d, where χ^2_d is the flight distance χ^2 obtained from the PV fit when tag_{SV} tracks are added to the fit result;
   - the tag_{SV} is rejected if it has a mass compatible with the K_S and is formed by only two tracks;

- it must have at most one track with $\Delta R > 0.5$
- a corrected mass M_{corr} greater than 600 MeV/c^2 in order to remove kaon or hyperon. The corrected mass is defined as the invariant mass M, corrected for the missing momentum:

$$M_{corr} = \sqrt{M^2 + p^2 sin^2\theta} + psin\theta \tag{4.6}$$

  where M and p are respectively the invariant mass and the momentum of the tag SV, θ is the angle between the tag_{SV} momentum and the vector that points its position from the PV.
- If more than one $tag_{SV}$is found inside a jet, the one with greater p_T is chosen, and if at least one tag_{SV} inside a jet, the jet is defined as generated by a heavy flavour quark ("SV tagged").

The standard tool used to separate between b, c and light jets flavour uses the output of two Boosted Decision Tree (BDTs) that uses SV-related observables as inputs: one for the heavy/light jet separation ($\text{BDT}_{bc|udsg}$) and the other for the b/c jet separation ($\text{BDT}_{b|c}$). The observables in input to the BDTs are those related to secondary vertex: the SV mass, the SV corrected mass, the transverse flight distance of the 2-body particle closest to the PV and built with tracks that belong to the SV, the fraction of the jet p_T carried by the SV, the number of tracks that form the SV, the number of tracks that form the SV with $\Delta R < 0.5$ from the jet axis, the total charge of tracks in SV, the SV flight distance χ^2 and the sum of χ^2_{IP} of all the tracks in the SV. In this thesis the Machine Learning set of algorithms presented in chapter 4.10 will be used for jet identification, but the BDTs templates have been used in the next section 4.9.2 to study the SV tagging algorithm efficiency differences between data and Monte Carlo.

4.9.2 Tagging Performance

The SV tagging performances have been evaluated on simulation of $b\bar{b}$, $c\bar{c}$ and light partons samples generated with PYTHIA 8 for pp collisions at 8 TeV.

The $b(c)$ tagging efficiency has been calculated as $\epsilon_{b(c)}$:

$$\epsilon_{b(c)} = \frac{N^{tag}_{b(c)}}{N^{tot}_{b(c)}} \tag{4.7}$$

where $N^{tag}_{b(c)}$ is the number of tagged $b(c)$ jets, while $N^{tot}_{b(c)}$ is the total number of reconstructed $b(c)$ jets, while the light jets misidentification is calculated as ϵ_q (where q stands for u, d, s quarks and g):

$$\epsilon_q = \frac{N^{tag}_q}{N^{tot}_q} \tag{4.8}$$

where N^{tag}_q is the number of tagged light jets, while N^{tot}_q is the total number of reconstructed light jets. The SV tagging algorithm performance have been evaluated by simulating b,c and light dijet samples coming from pp collisions at 8 TeV [92]. The tagging efficiencies and misidentification probability have been determined as a function of the jet p_T and η and are shown in Figure 36. As can be seen in Run 1 the SV-tagging efficiency was of about 65%(25%) for $b(c)$-jets with a transverse momentum greater than 20 GeV, with a light jet misidentification below 0.3%, almost constant over the pseudo-rapidity range. For jets with p_T less than 20 GeV the b(c) tagging efficiencies are significantly lower.

It has been verified in simulation that in Run 2 the SV-tagging efficiency for $b(c)$-jets is of the same order ($< 10\%$ relative loss with respect to Run 1), while the light jet misidentification is higher (approximately 1% with respect to 0.3%)

The performance of jet tagging algorithm have been measured also on data, to check differences with the simulation. Data events with a high p_T fully reconstructed b or c hadron are selected and is labeled as "event-tag". Different event-tag processes have been used to compute $b(c)$ tagging efficiencies and q jets mistag like B+jet, that is a b-jets enriched sample, D+jet, that is a b and c jets enriched sample, since the B hadrons decay into D, μ+jet, where the b and c jets presence is enriched by the fact that a high p_T displaced muon is selected. W+ jet events are also used to measure the light jets misidentification, since this sample is composed at 95% by light parton jets. The highest p_T jet associated to the same PV of the event-tag, with a $\Delta\phi$ with respect to the event-tag greater than 2.5 and p_T in the [10,100] GeV/c range is selected and used as test jets. The b and c SV tagging efficiencies have been then calculated by using the $\mathrm{BDT}_{b|c}$ and $\mathrm{BDT}_{bc|udsg}$ distributions for b,c and light jets templates obtained using MC events: by fitting the BDTs distributions of event-tag where the test jet is tagged, the yield of tagged b,c and light jets ($n^{tag}_{b(c)}$ and n^{tag}_{q} respectively) events is measured. To obtain the total number of b,c and light jets ($n^{tot}_{b(c)}$ and n^{tot}_{q} respectively), the χ^2_{IP} of the track with the highest p_T in the jet ($\chi^2_{IP}(max-pt)$) distributions are obtained for b,c and light jets using MC events. They are used to fit the $\chi^2_{IP}(max-pt)$ distributions of event-tag where the SV tagging is not applied to the test jet are.

The $b(c)$ tagging efficiency has been then calculated as:

$$\epsilon_{b(c)} = \frac{n^{tag}_{b(c)}}{n^{tot}_{b(c)}} \tag{4.9}$$

Ratio between the SV tagging efficiencies measured in data and those obtained from simulation samples, for b and c jets is shown in Figure 37 , while Table 38 shows the measured efficiencies and the data/MC scale factors for the SV tagger in different p_T intervals.

4.10 Machine learning description

At the LHCb experiment, the standard jet SV-tagging algorithm is used to properly distinguish jets originating from b, c, and light quarks. Further discrimination between the quark flavour can be obtained by using a set of Machine Learning algorithms, in particular of Neural Networks that exploits correlation between features to cluster or classify data. The tagging technique used in this thesis uses the whole information coming from the jet substructure to provide further discrimination to distinguish heavy from light, and b from c jets. The set of Neural Networks used in this study resembles the *DeepJet* algorithm [93] developed at the CMS experiment. Particularly, the substructure features of the jet are separated into four categories:

- *charged particles*: for each jet, 10 charged particles with the highest IP are selected and ordered in descending order. For each charged particle, 22 features are considered, namely the result of the PID described in section 4.2 `ID`, the impact parameter `IP`, the charge `Q`, the energy `E`, the transverse momentum `pT`, the momentum components along the x, y and z axes `pX`, `pY`, `pZ`, the pseudorapidity `Eta` and azimuthal angle `Phi`, `Chi2`, the ratio between the charge associated to a track and its momentum `QoverP`, the ProbNN output for electrons, kaons, pions and muons `NNe`, `NNk`, `NNpi`, `NNmu`, `trackX`, `trackY`, `trackZ`, `trackVX`, `trackVY`, `trackVZ`;

- *neutral particles*: for each jet, 15 neutral particles with the highest transverse energy are selected and ordered in descending order. For each neutral particle, 12 features are considered, namely `ID`, the energy `E`, the transverse momentum `pT`, the momentum components along the x, y and z axes `pX`, `pY`, `pZ`, the pseudorapidity `Eta` and azimuthal angle `Phi`, ECAL cluster energy associated to the particle `CaloNeutralEcal`, Hcal/Ecal energy ratio associated to the particle `CaloNeutralHcal2Ecal`, the ratio of the energy between 2×2 cells and 3×3 cells `CaloNeutralE49`, the PS digits associated to the particle`CaloNeutralPrs`;

- *SV features*: for the highest p_{T} SV identified inside a jet 14 features are considered, namely the ΔR between the SV flight direction and the jet `fdrMin`, the ratio of p_T of the SV over the p_T of the jet `ptSvrJet`, ΔR between the jet axis and the SV flight direction `drSvrJet`, the absolute value of the sum of the charges of tracks coming from the SV `absQSum`, the number of tracks coming from the vertex `nTrk`, `nTrkJet`, the mass M of the SV `m`, the SV corrected mass `mCor`, the flight distance χ^2 `fdChi2`, the sum of χ^2_{IP} for all the tracks in the SV `ipChi2Sum`, the time of flight of the SV `tau`, the position of the SV along the z axis `z`, the SV transverse momentum `pt`;

- *global features:* for each jet, 7 global features are selected, namely the jet momentum components along the x, y and z axes `PX`, `PY`, `PZ`, the jet energy `PE`, the jet mass `M`, the number of charged particles in the jet `NPartChg`, the number of neutral particles in the jet `NPartNtr`.

The number of features is fixed, for a total number of 421 inputs. If the number of particles is not enough to fill the features, then the remaining ones are put to 0. An important advantage of this set of Neural Networks is that it can be applied to a jet even if a SV is not found. In that case the SV features are put to 0.

The structure of the set of Neural Networks is shown in Fig. 39. The structure has the following steps:

1. *batch normalization*, where all the features are properly normalized to be used by the network;

2. *convolutional layer*, a 1D Convolutional Network applied only to the charged and neutral features. The depth of each convolutional layer is set to 64.

3. *recursive layer*, a LSTM layer applied only to the charged and neutral features. This is a Recurrent Neural Network (RNN) [94], used for example in language modeling, that allows to exploit correlation between the particle features.

4. *dense layer*, a final dense layer, that consists of a Deep Neural Network (DNN) with six layers. The last layer, the output layer, is composed by three neurons.

The output of the DNN algorithm is whether a jet has originated from a b, c or light quark. Particularly, three probabilities are defined, namely P_b, P_c and P_q, which represents the probability for a jet to come from a b, c or light quark.

4.11 Neural Network performance

To evaluate the Neural Network performance, the algorithm is applied on MC simulation of $b\bar{b}$, $c\bar{c}$ and light partons samples generated with PYTHIA 8 for pp collisions at 13 TeV. SV-tagging requirements are not applied in the selection of these samples. The original dataset is properly separated into training, evaluation, and testing datasets. To show qualitatively the performance of the DNN, Fig. 40 shows the outputs (namely P_b, P_c and P_q) applied to the testing dataset for different samples of bb, cc and qq di-jets. Figure 41 shows a similar comparison for jets that are SV-tagged. It is evident that the probabilities distribution are well separated between b, c and light jets.

The standard figure of merit to properly assess the DNN performance is the Receiving Operating Characteristic (ROC) curve. The ROC usually evaluates the ability of a binary classifier to classify events depending on the variation of a discrimination threshold. Therefore, Fig. 42 shows the ROC curves for each pair of classes (b vs c, b vs q, c vs q and bc vs q) for the DNN. The best discrimination is obtained in the separation of b from light jets and for bc from light.

Finally, the comparison between the DNN and the SV-tagging algorithm is performed in terms of tagging efficiency. The tagging efficiency $\varepsilon_{\mathrm{tag}}$ is defined as

$$\varepsilon_{\text{tag}} = \frac{N_{\text{tag}}}{N_{\text{rec}}} \tag{4.10}$$

where N_{rec} (N_{tag}) is the number of reconstructed (tagged as b or c) jets. To have a fair comparison between DNN and SV-tagging, the requirements on the DNN probabilities are optimized in order to have the same light jet mis-identification rate. Figure 43 shows the comparison between DNN and SV-tagging for identifying a b or a c jet, as a function of the jet p_{T}. For each bin of jet p_{T}, the efficiency improvement coming from the DNN with respect to SV-tagging is also shown. Particularly for the c-jets, this improvement is always above 40%.

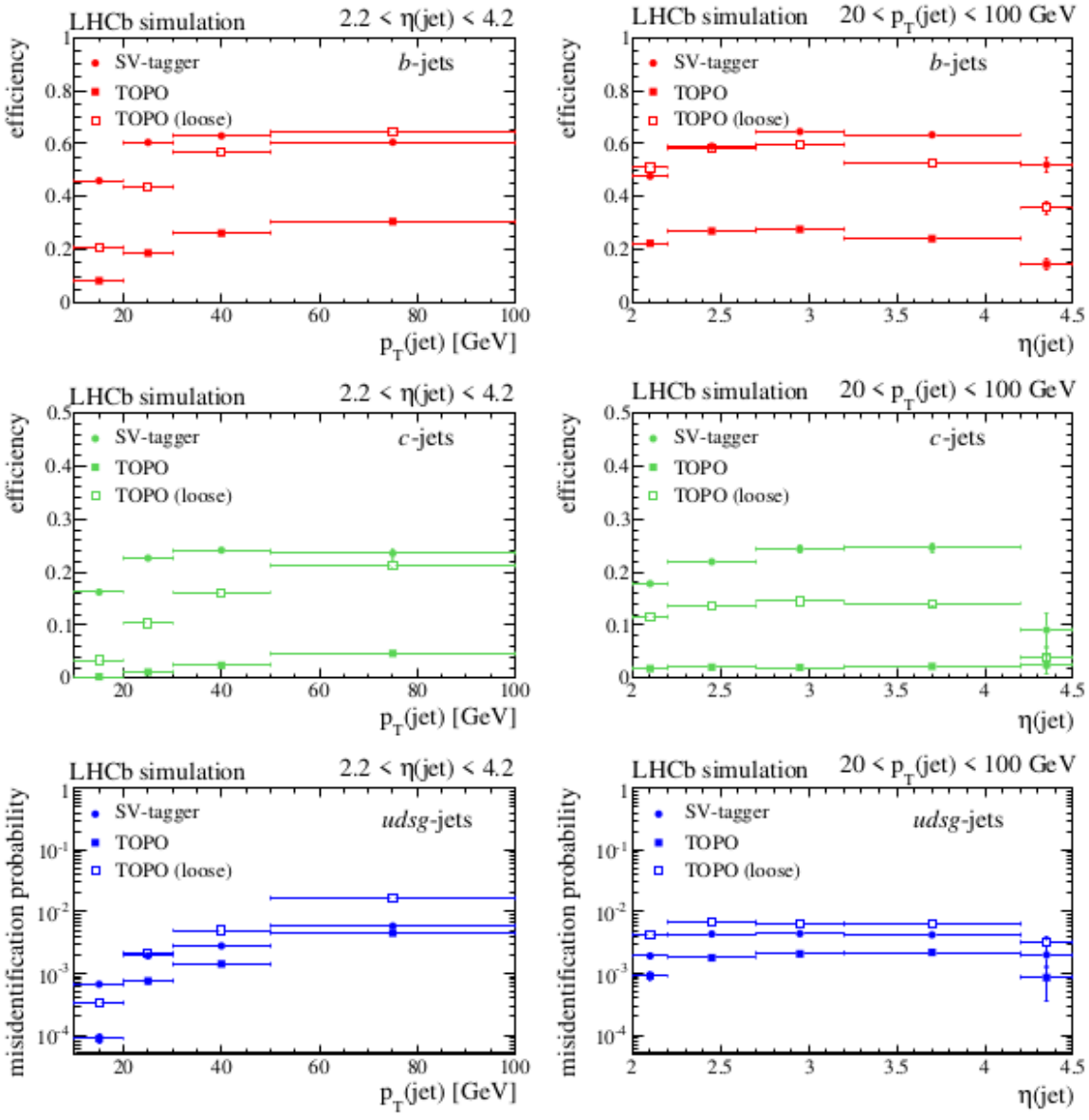

Figure 36: SV tagging efficiencies (filled dots) on b jets (first row), c-jets (middle row) and misidentification probability of light parton jets (last row) calculated as a function of the jet p_T and η. Other points refer to the topological trigger and its loose version [92].

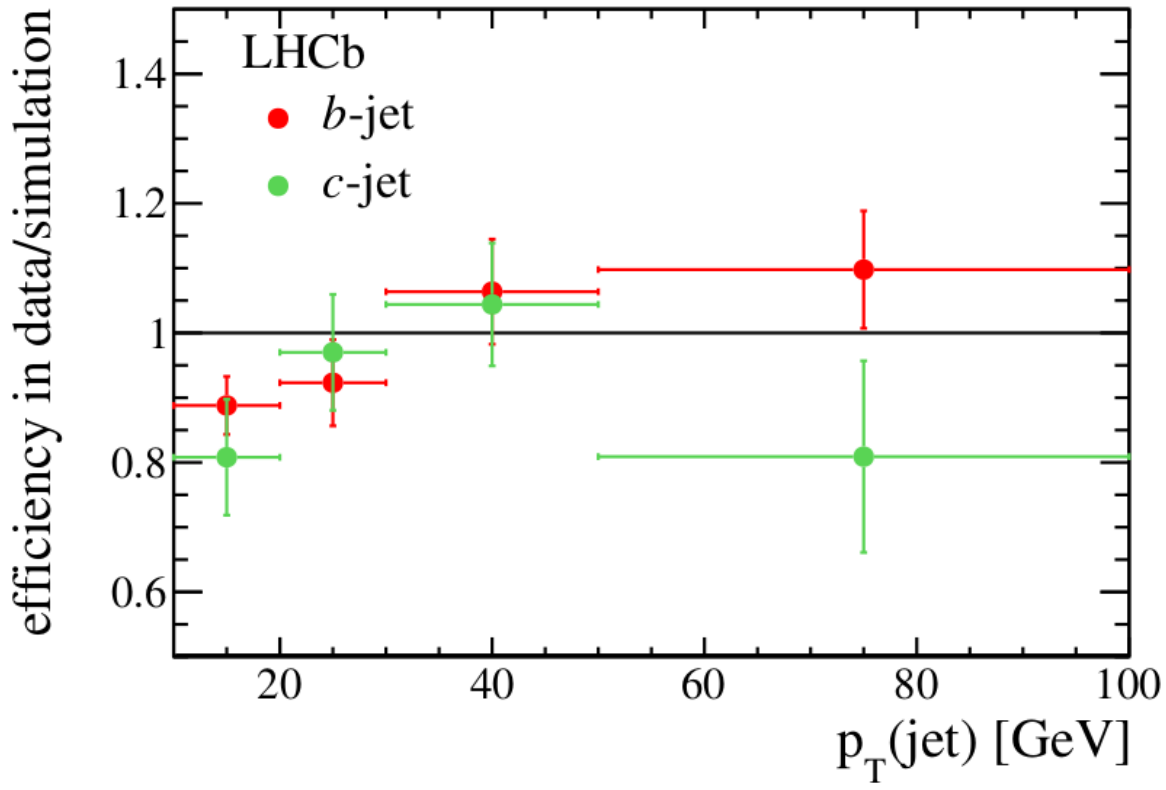

Figure 37: Ratio between the SV tagging efficiencies measured in data and those obtained from simulation, for b and c jets [92].

		ϵ(data)/ϵ(simulation)			ϵ(data) (%)	
jet p_{T} (GeV)	jet η	b jets	c jets	(b,c) jets	b jets	c jets
10–20	2.2–4.2	0.89 ± 0.04	0.81 ± 0.09	0.91 ± 0.04	38 ± 2	14 ± 1
20–30	2.2–4.2	0.92 ± 0.07	0.97 ± 0.09	0.97 ± 0.04	61 ± 3	23 ± 1
30–50	2.2–4.2	1.06 ± 0.08	1.04 ± 0.09	0.97 ± 0.04	65 ± 3	25 ± 1
50–100	2.2–4.2	1.10 ± 0.09	0.81 ± 0.15	1.05 ± 0.06	70 ± 4	28 ± 4
20–100	2–2.2	1.00 ± 0.07	1.12 ± 0.10	1.05 ± 0.03	56 ± 2	20 ± 1

Figure 38: SV tagging algorithm efficiencies measured on data over the SV tagging algorithm efficiencies measured on Monte Carlo as a function of the jet p_T [92].

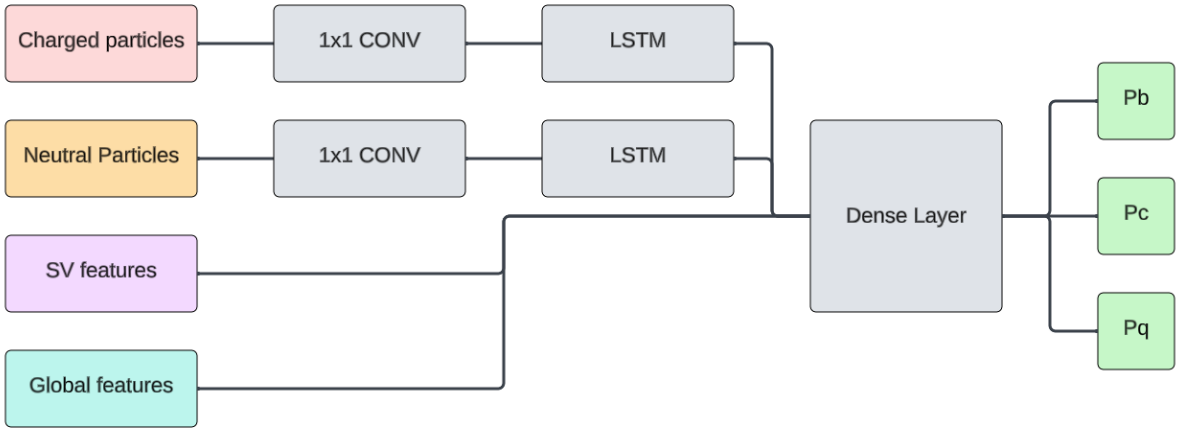

Figure 39: Schematic representation of the DNN used in this study.

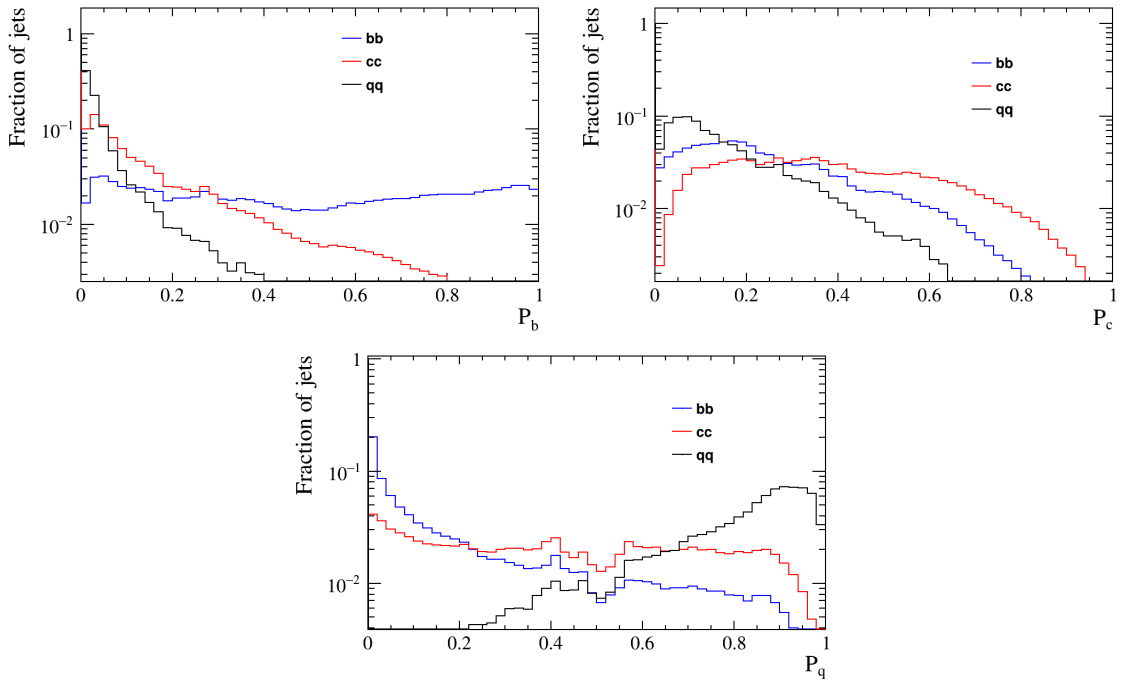

Figure 40: Single-jet probabilities P_b, P_c and P_q shown for bb, cc and qq di-jets samples where no SV tagging is applied.

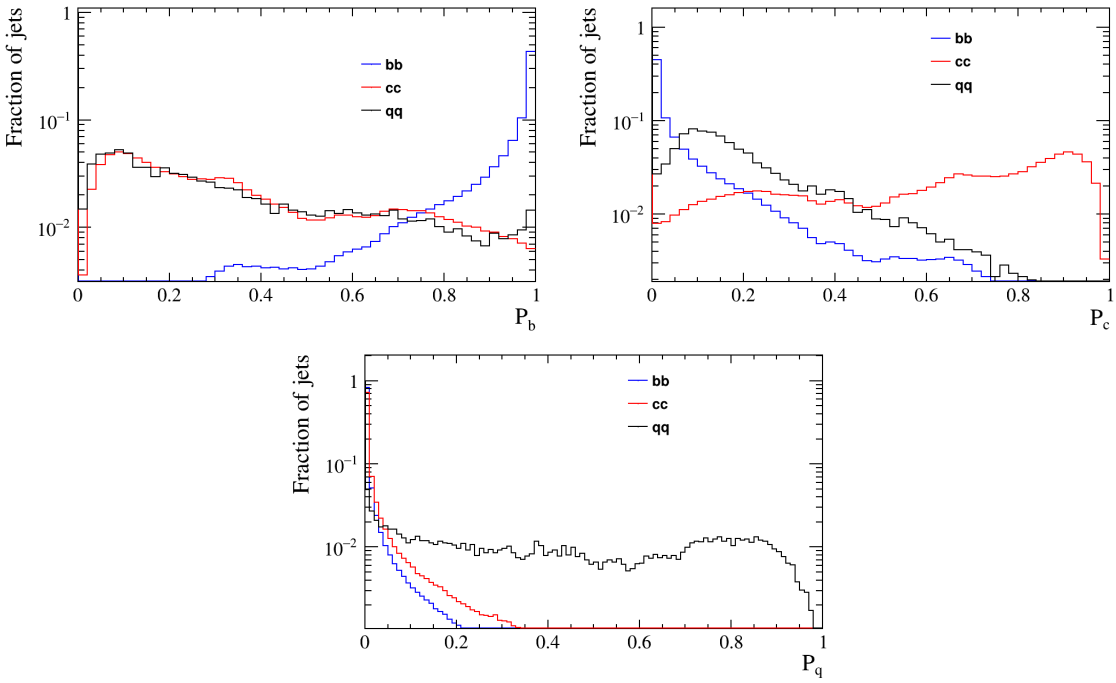

Figure 41: Single-jet probabilities P_b, P_c and P_q shown for bb, cc and qq di-jets samples where SV tagging is applied.

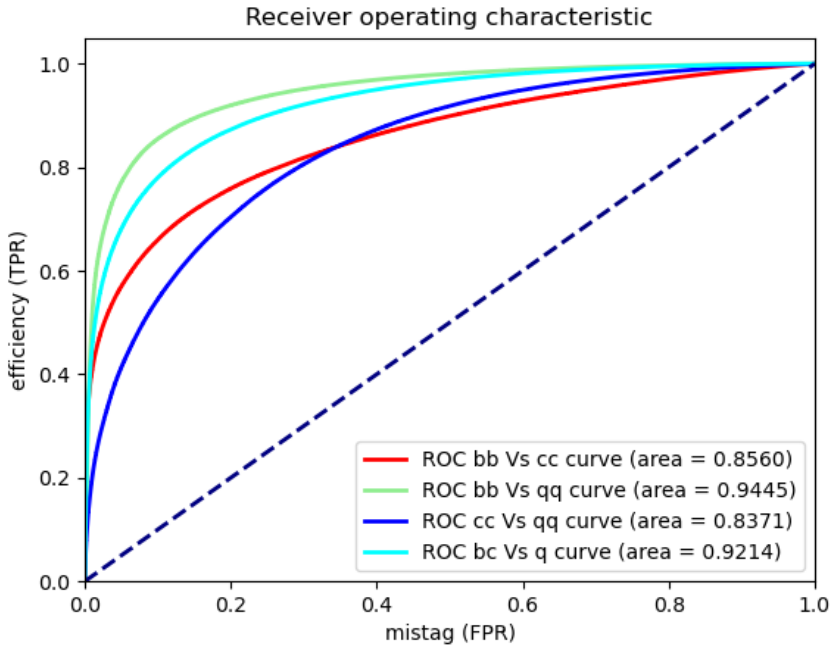

Figure 42: ROC curves for DNN for different classification tasks.

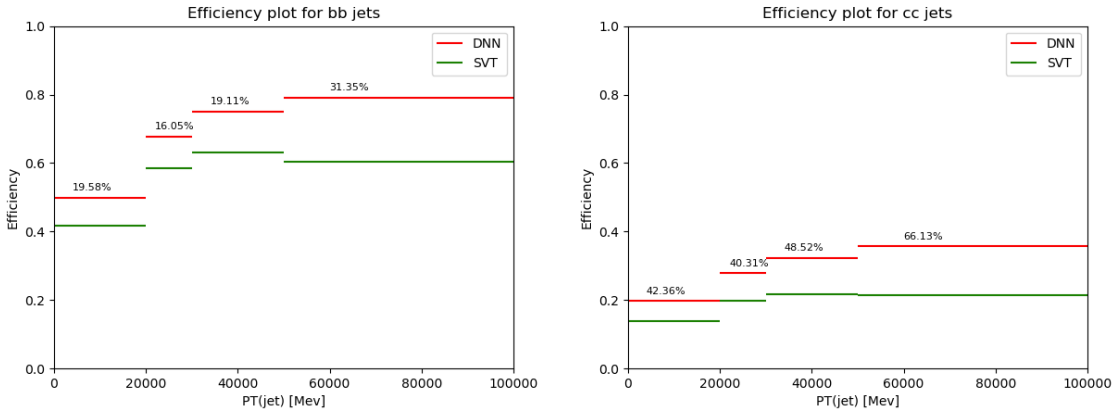

Figure 43: Single-jet heavy flavour tagging efficiency ε_{tag} for DNN and SV-tagging, in the *bb* sample (left) and *cc* sample (right) jets, as a function of jet p_{T}.

5 Search for $H \to b\bar{b}$ and $H \to c\bar{c}$ at LHCb

5.1 Introduction

In this chapter the determination of upper limits on the inclusive $H \to b\bar{b}$ and $H \to c\bar{c}$ production processes at LHCb is presented. The analysis is performed with the 2016 dataset, a sub-sample of the Run 2 dataset, that corresponds to 1.6 fb^{-1} integrated luminosity. It aims to test the jet identification strategy that uses the Neural Network algorithms described in section 4.10. The main challenge of this analysis is the description of the QCD multi-jet background, since MC simulation fails to reproduce such background processes. A data-driven approach is then followed, as it will be described in section 5.3. Other sources of background are the $Z \to b\bar{b}$ and $Z \to c\bar{c}$ processes.

This chapter is structured in the following way:

- Section 5.2 presents the Monte Carlo and data samples used for the analysis and the selection requirements applied to the reconstructed events;
- Section 5.3 presents an overview of the analysis strategy;
- From section 5.4 to section 5.7 the analysis is presented in detail;
- Section 5.8 presents the results of upper limit calculation with the CLs tecnique without the systematic uncertainties;
- The assessment of the systematic uncertainties in the Higgs searches is discussed in section 6.

5.2 Monte Carlo and data samples

All Monte Carlo and data samples used in this thesis have been produced at $\sqrt{s}$=13 TeV center of mass energy with the standard LHCb simulation framework described in section 3.4.

5.2.1 $\mathbf{H \to b\bar{b}}$ and $\mathbf{H \to c\bar{c}}$ samples

236000 (2100000) events of $H \to b\bar{b}$ ($H \to c\bar{c}$) have been generated at Leading Order with the PYTHIA8 Monte Carlo by requiring the two $b(c)$ quarks with pseudorapidity $2 < \eta < 5$. They have been generated considering the gluon-gluon fusion (ggF) production mechanism, that is the dominant one, and have been simulated to determine efficiencies, the signal dijet invariant mass models and to evaluate the working point for the tagging DNN requirements, as will be explained in section 5.4. The impact of neglecting Vector Boson Fusion (VBF) and the Higgs associated production with Vector boson (VH), that have a cross section of more than one order of magnitude smaller than ggF, has been calculated in section 5.5.

5.2.2 $b\bar{b}$, $c\bar{c}$ and light dijet samples

In this analysis simulation samples of $b\bar{b}$ dijets, $c\bar{c}$ dijets and dijets generated from light partons (u, d, s and g, indicated in the following as q) are used in the $H \to b\bar{b}$ and $H \to c\bar{c}$ searches for the evaluation of the transfer function from the control region to the signal region and for the determination of the multi-jet QCD background (section 5.6).

In order to have a significant amount of events for the different ranges of jet transverse momentum, samples with different values of transverse momentum exchanged in the hard interaction ($\hat{p}_{\mathrm{T}}$) have been generated: [10,15], [15,20], [20,50] and [50,+∞] GeV.

When they are used together, one out of four possible weights is applied to each event, depending on the $\hat{p}_{\mathrm{T}}$ range. These weights have been taken proportional to the cross sections and have been evaluated with PYTHIA 8. They are reported in table 12. About 2M events have been generated

for each $\hat{p}_{\rm T}$ interval and dijet flavour. $b\bar{b}$, $c\bar{c}$ have been generated by asking the two $b(c)$ quarks with pseudorapidity $2 < \eta < 5$ while for light dijets samples, two q-partons are required to have pseudorapidity $2 < \eta < 5$.

MC sample	Weight
$q\bar{q}$, $\hat{p}_T$ in [10,15] GeV	1.0
$q\bar{q}$, $\hat{p}_T$ in [15,20] GeV	0.22
$q\bar{q}$, $\hat{p}_T$ in [20,50] GeV	0.12
$q\bar{q}$, $\hat{p}_T > 50$ GeV	0.0035
$c\bar{c}$, $\hat{p}_T$ in [10,15] GeV	1.0
$c\bar{c}$, $\hat{p}_T$ in [15,20] GeV	0.21
$c\bar{c}$, $\hat{p}_T$ in [20,50] GeV	0.12
$c\bar{c}$, $\hat{p}_T > 50$ GeV	0.035
$b\bar{b}$, $\hat{p}_T$ in [10,15] GeV	1.0
$b\bar{b}$, $\hat{p}_T$ in [15,20] GeV	0.25
$b\bar{b}$, $\hat{p}_T$ in [20,50] GeV	0.15
$b\bar{b}$, $\hat{p}_T > 50$ GeV	0.0045

Table 12: List of $b\bar{b}$, $c\bar{c}$ and light dijets samples used in this study. The label q indicates a u, d, s quark or a gluon. All the samples have been generated with PYTHIA 8, and about 2M events have been produced for each sample. Weights applied to each simulation sample, proportional to the PYTHIA 8 cross sections as explained in the text, are also reported.

5.2.3 $Z \to b\bar{b}$, $Z \to c\bar{c}$ and $Z \to \mu\mu$+jet samples

Apart from the multi-jet QCD background, other important background contributions in the Higgs searches are represented by the $Z \to b\bar{b}$ and $Z \to c\bar{c}$ processes. To study their properties MC samples have been produced. 236000 (214000) events of $Z \to b\bar{b}$ ($Z \to c\bar{c}$) have been generated with the PYTHIA Monte Carlo by asking the two $b(c)$ quarks with pseudorapidity $2 < \eta < 5$ and have been fully simulated within the LHCb framework.

A sample of $Z \to \mu\mu$+jet have been also generated to evaluate the systematic uncertainty relative to the jet energy scale and jet identification. A sample of 2.29M event has been generated with the PYTHIA Monte Carlo by asking the two μ with pseudorapidity $2 < \eta < 5$ and $p_T(\mu) > 4$ GeV.

5.2.4 Data samples

Four different datasets have been used in this analysis, based on four different selection requirements.

1. events used to extract the $H \to b\bar{b}$ and $H \to c\bar{c}$ limits are those that have passed the *HltQEEJetsDiJetSVSV* HLT2 line requirements (see section 3.3.2), requiring two jets, both reconstructed with transverse momentum ($p_{\rm T}$) greater than 17 GeV and with a secondary vertex (SV) in the jet cone. The total integrated luminosity analyzed is 1.6 fb^{-1};

2. a data sample of events that passed the HLT2 trigger line *HltQEEJetsDiJetSV* (see section 3.3.2) is used for background studies in the $H \to b\bar{b}$ and $H \to c\bar{c}$ searches. This trigger line selects events with two reconstructed jets with $p_{\rm T} > 17$ GeV and requires that at least one

of the two jets must contain a SV. The *HltQEEJetsDiJetSV* line is prescaled with a prescale factor of 1%[2];

3. as will be explained in section 4.10, the DNN identification technique can be applied to all jets, including those that do not have a reconstructed SV. In Run 2 data the inclusive Higgs searches are limited by the fact that only the dijet line *HltQEEJetsDiJetSVSV* is available without prescaling. In order to evaluate the improvement that can be achieved by applying the tagging DNN at trigger level, another dataset is used: the sample of events that passed the HLT2 trigger line *HltQEEJetsDiJet*. This HLT2 trigger line selects events with two reconstructed jets with $p_{\mathrm{T}} > 17$ GeV, without any requirement on the SV. The *HltQEEJetsDiJet* line is even prescaled because of the high rate, with a prescale factor of 0.1%. A further selection is applied at the stripping level (see section 3.4), that applies a further cut on the jets transverse momentum, $p_{\mathrm{T}} > 20$ GeV, and have a pre-scale factor of 0.13%;

4. some systematic studies have been performed with a data sample of Z+jet events. Events containing two tracks identified as muons at L0, HLT1 and HLT2 levels and with a transverse momentum greater than 20 GeV are required, with a dimuon invariant mass compatible with the Z boson.

These four samples are respectively labeled as dijetSVSV, dijetSV, dijet and Zjet samples in the rest of the thesis.

5.2.5 Event selection and yield prediction

The selection is applied to data events that passed the $HltQEEJetsDiJetSVSV$ trigger line and that belong to the dijetSVSV sample defined in section 5.2.4. Two reconstructed jets coming from the same Primary Vertex (PV) are selected as dijet candidates. The two jets must be SV-tagged. The jets pseudorapidity must be in the range $2.2 < \eta < 4.2$, the jet transverse momentum must be greater than 20 GeV and an angular difference between the two jets in the azimuthal plane with respect to the beam axis ($\Delta\phi$) greater than 1.5 radians. The latter requirement is applied to remove jet duplicated and associated to the same true jet. The reduced η range with respect to the full LHCb acceptance is chosen to ensure a flat jet reconstruction and identification efficiency.

L0 and HLT1 trigger decisions at LHCb are classified into two categories: Trigger on signal (TOS) or trigger independent of signal (TIS), and are recorded for each jet of the event that fires the lines. A jet is triggered on signal if it is part of the signal that has fired the trigger line. For example, a jet that is TOS for the $L0Muon$ line, will contain the muon with the thresholds listed in Table 9, that has fired the trigger line. On the contrary, if a jet is TIS for the $L0Muon$, means that a muon is found in the event that pass the trigger, but it is not contained in the jet. A jet is defined as *L0Chain* TOS if it is TOS for one of the following L0 lines: $L0Muon$, $L0MuonEW$, $L0Electron$, $L0Photon$, $L0Hadron$, $L0DiMuon$. A jet is defined as *HLT1Chain* TOS if it is TOS for one of the HLT1 lines described in section 3.3.2. At least one of the two jets is required to be $L0Chain$ TOS and $HLT1Chain$ TOS. After applying all the requirements, if multiple dijet candidates exist in the event, the jet pair with maximum $p_{\mathrm{T}}(j_0) + p_{\mathrm{T}}(j_1)$ is selected. A total number of 24.4M data events are selected. The summary of the selection requirements is shown in table 13.

A similar selection has been applied to the dijetSV and dijet samples described in section 5.2.4, with the only difference that at least one SV is required for the dijetSV sample, and no requirements on the SVs are applied to the dijet sample.

Some reconstructed observables are compared between data and Higgs simulation after the selection requirements. These features are: the leading jet pseudorapidity, the leading jet transverse

[2]Prescale is applied to high rate data. A prescale factor of 1% means that 1 over 100 events that pass the line requirement is stored.

Table 13: List of requirements used to select data events for the $H \to b\bar{b}$ and $H \to c\bar{c}$ searches.

$HltQEEJetsDiJetSVSV$ HLT2 line
two reconstructed jets
both jets SV-tagged
one of the two jets $L0Chain$ TOS and $HLT1Chain$ TOS
$p_{\mathrm{T}}(j_0) > 20$ GeV
$p_{\mathrm{T}}(j_1) > 20$ GeV
$2.2 < \eta(j_0) < 4.2$
$2.2 < \eta(j_1) < 4.2$
$\lvert\Delta\phi\rvert > 1.5$

momentum, the dijet invariant mass and the separation in the azimuthal angle between the two jets ($\Delta\phi$). The dijet invariant mass distributions is of paramount importance since it is used to extract the $H \to b\bar{b}$ and $H \to c\bar{c}$ upper limits. The distributions of the four observables in the selected data events are shown in figures 44, 45, 46 and 47. Just a small fraction of data events have been used to produce these plots ($< 1\%$).

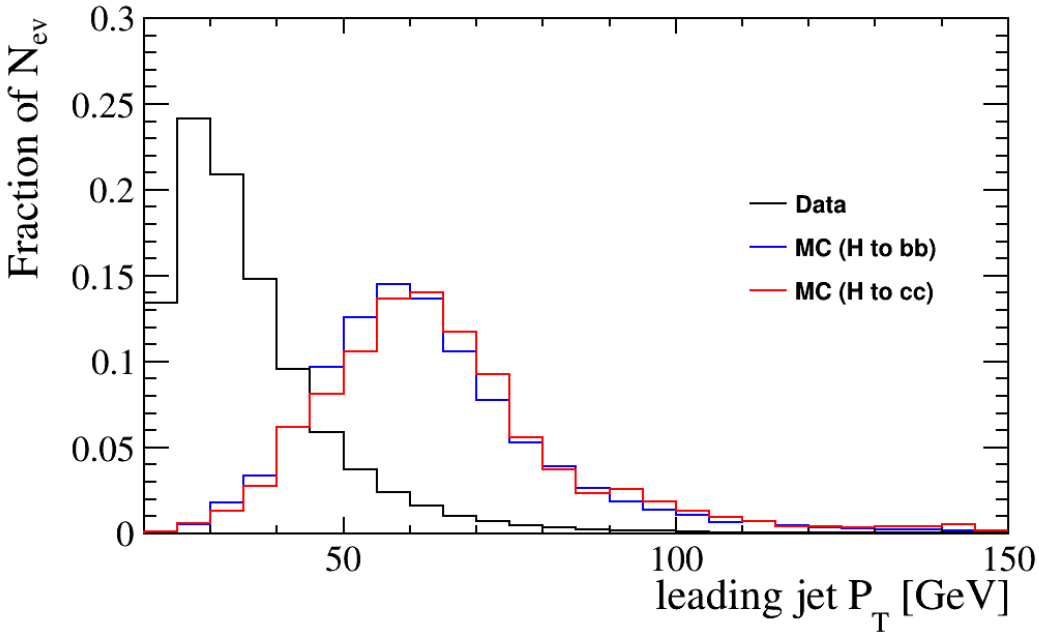

Figure 44: Leading jet transverse momentum distribution of heavy flavour dijet candidates in a small subset (0.8 %) of 2016 data and Higgs MC samples.

The number of Monte Carlo expected $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ events with these requirements has been calculated using the following formula:

$$N_{exp}^{SM} = \mathcal{L} \cdot \sigma_{SM} \cdot A \cdot \epsilon, \tag{5.1}$$

where:

- σ_{SM} is the theoretical SM cross section. For $H \to b\bar{b}$ and $H \to c\bar{c}$ it has been taken from [95] at N3LO, while for $Z \to b\bar{b}$, $Z \to c\bar{c}$ the NLO estimation of the cross section, as obtained with MADGRAPH, is used;

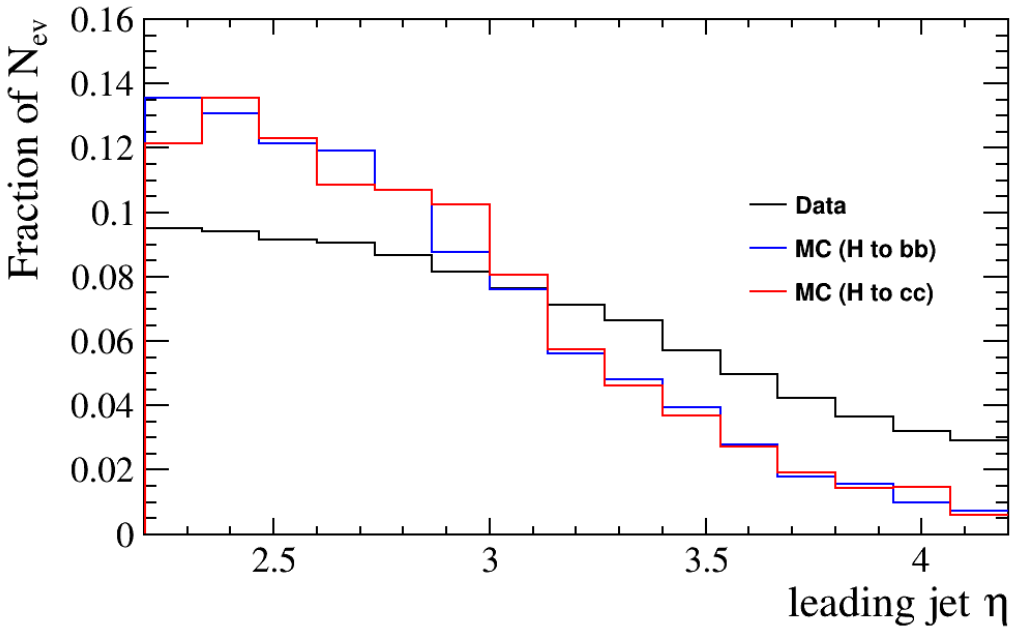

Figure 45: Leading jet pseudorapidity of heavy flavour dijet candidates in a small subset (0.8 %) of 2016 data and Higgs MC samples.

- A is the LHCb acceptance factor computed with PYTHIA for the H and with MADGRAPH for the Z;
- ϵ is the selection efficiency and is defined as

$$\epsilon = \frac{n_{MC}^{sel}}{n_{MC}^{gen}} \tag{5.2}$$

  where $n_{MC,sel}$ is the number of MC events selected after applying selection requirements, while $n_{MC,gen}$ is the number of generated MC events.

The $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ cross sections, acceptances, efficiencies and expected events after the selection described in this section, are reported in Tab. 14. Possible kinematic differences between the Higgs produced via ggF at LO and NLO have also been checked, since they would affect the acceptance factor. Figure 48 shows the transverse momentum (left) and the pseudo-rapidity distributions (right) of the Higgs produced via ggF at LO (red) and NLO (blue) using MADGRAPH Monte Carlo. As can be seen, the two distribution are similar, then acceptance variations due to higher order effects are considered negligible.

Process	σ_{SM} [pb]	A	ϵ (%)	N_{exp}^{SM}
$Z \to b\bar{b}$	6526	0.128	12	160947
$Z \to c\bar{c}$	5130	0.120	0.6	5484
$H \to b\bar{b}$	31.8	0.074	10	368
$H \to c\bar{c}$	1.6	0.073	1.3	2.4

Table 14: $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ SM cross sections, acceptances, efficiencies and expected events with $\mathcal{L} = 1.6$ fb^{-1} after the selection described in this section.

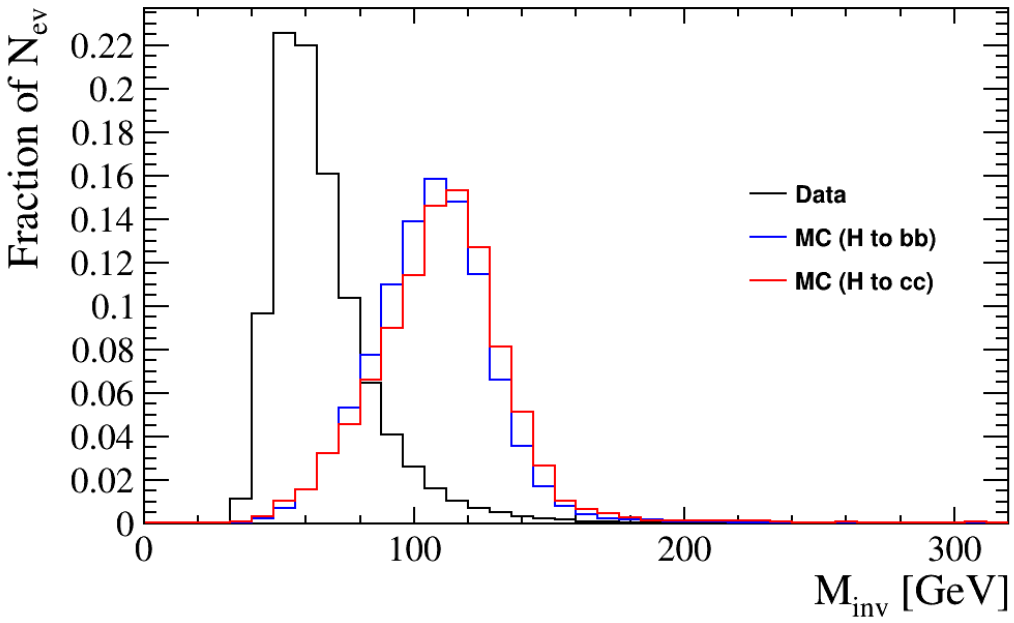

Figure 46: Invariant mass distribution of heavy flavour dijet candidates in a small subset (0.8 %) of 2016 data and Higgs MC samples.

5.2.6 Z$\to \mu\mu$ + jet selection

The $Z \to \mu\mu$ + jet process has been studied by LHCb in [96]. It has been demonstrated that this sample has high purity, above 97%. A stripping pre-selection is required ($StrippingZ02MuMuLine$) where two tracks identified as muons, from the same PV, with p_T greater than 3 GeV and invariant mass greater than 40 GeV are selected. Then further cuts are applied to the muons p_T, requiring them to be greater than 20 GeV. The $2.0 < \eta < 4.5$ and requirement is also applied to both muons. Events with a single jets have been selected and it is required to have a p_T greater than 20 GeV and $2.2 < \eta < 4.2$ and have the SV.

The selection is applied to a data sample corresponding to 1.6 fb^{-1}, collected in the year 2016. A $Z \to \mu\mu$ + jet MC sample is also used: it has been generated with PYTHIA 8 within the GAUSS framework with 2016 conditions.

5.2.7 Jet energy resolution and jet energy scale

The energy of simulated jet in Monte Carlo samples have been calibrated to take into account for possible data/MC differences: the jet energy have been scaled for a certain factor, called jet energy scale (JES), and have been smeared for a certain resolution factor, the jet energy resolution (JER). These factors have been determined by using the Zjet sample and the low mass sideband of the dijetSVSV sample, (<45 GeV) respectively. The JES and JER corrections may change the event selection efficiency if selection requirement are applied on jet kinematic quantities like the jet p_T, then these factors have been applied to Monte Carlo events before the event selection. The jet resolution and scale factors variations will be evaluated as systematic uncertainty in section 6.4.

The Zjet sample has been used for determining the jet energy scale correction and its systematic uncertainty, by comparing the $p_{\rm T}({\rm jet})/p_{\rm T}(Z)$ distributions in data and MC, as done in [28]. Z+jets events have been selected by requiring the muons transverse momentum $p_{T,\mu} > 20$ GeV, the jet transverse momentum $p_{T,jet} > 10$ GeV, the di-muons invariant mass between 60 and 120 GeV and the angular aperture in the transverse plane $\Delta\phi > 2.3$. Furthermore, the jet is required to be

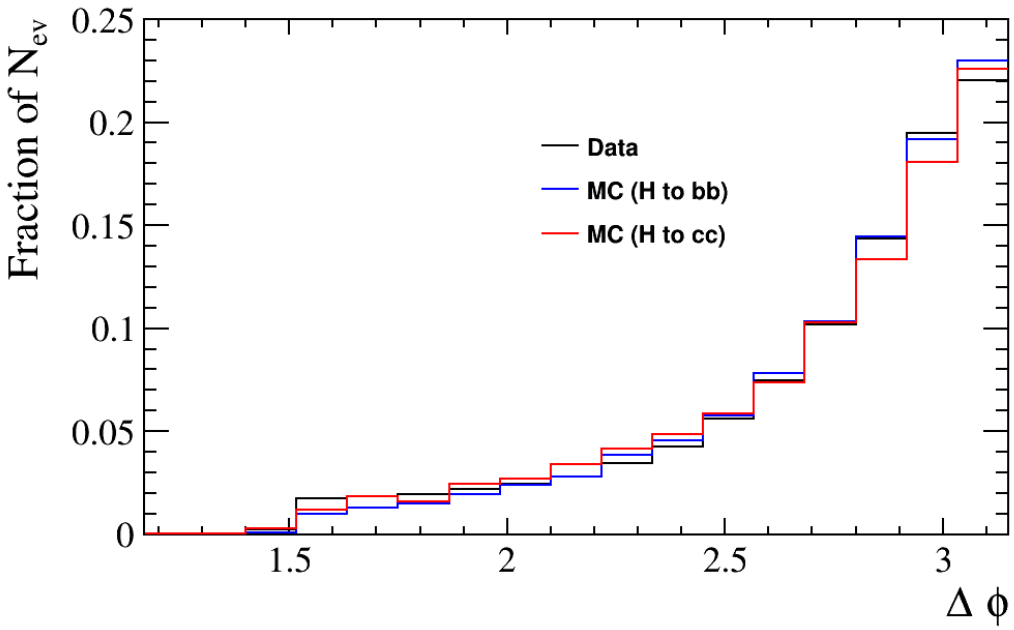

Figure 47: Separation in the azimuthal angle between the two jets ($\Delta\phi$) in a small subset (0.8 %) of 2016 data and Higgs MC samples.

tagged. To calculate the JES factor, $p_{\mathrm{T,jet}}$ has been scaled by a certain factor and for each factor the mean of the $p_{\mathrm{T}}(\mathrm{jet})/p_{\mathrm{T}}(Z)$ distributions in data and MC are compared. The uncertainty interval associated with the jet energy scale is computed as the factor one needs to apply to $p_{T,jet}$ to have an agreement with data less than 1 σ and is found to be $[0.098, 1.056]$. The centroid of the interval is 1.015. Figure 49 shows the mean values of the $p_{\mathrm{T}}(\mathrm{jet})/p_{\mathrm{T}}(Z)$ distributions as a function of the JES scale factor applied to Monte Carlo events. Figure 50 shows the $p_{\mathrm{T}}(\mathrm{jet})/p_{\mathrm{T}}(Z)$ distributions before (left) and after (right) the central JES factor (1.015) is applied to Monte Carlo events.

The jet energy resolution has been evaluated on the low mass sidebands of the dijetSVSV sample, not used in the fit to the invariant mass that will be described in section 5.7 (< 45 GeV). In Figure 51 left, the jet p_{T} asymmetry $\frac{p_{\mathrm{T}}(\mathrm{jet1})-p_{\mathrm{T}}(\mathrm{jet2})}{p_{\mathrm{T}}(\mathrm{jet1})+p_{\mathrm{T}}(\mathrm{jet2})}$ for the low mass sideband is compared between data (black dots) and dijet simulation samples (red). Data and Monte Carlo have been selected with the same kinematic requirements listed in Table 13 and one SV for each jet. The Monte Carlo jets $p_{\mathrm{T}}(\mathrm{jet1})$ and $p_{\mathrm{T}}(\mathrm{jet2})$ have been smeared as:

$$P_T^{smeared} = N(p_T, \sigma_{p_T}) \tag{5.3}$$

where $N(\mu, \sigma)$ draws from a normal distribution with mean μ and standard deviation σ. σ is the factor used for the smearing resolution. Figure 52 shows the plot of the tested σ_{p_T} values (x-axis) versus the jet asymmetry standard deviation for data (black) and Monte Carlo (red). The agreement between the standard deviation of data and simulation is found for a resolution factor of 0.069.

Distributions of $\frac{p_{\mathrm{T}}(\mathrm{jet1})-p_{\mathrm{T}}(\mathrm{jet2})}{p_{\mathrm{T}}(\mathrm{jet1})+p_{\mathrm{T}}(\mathrm{jet2})}$ in the low mass sideband < 45 GeV, compared between data (black) and simulation (red), before (left) and after (right) the JER scale is applied to Monte Carlo are shown in Figure 51. The Monte Carlo error-bars after the JER smearing are the bin-per-bin difference of the Monte Carlo distribution before and after the JER factor is applied.

5.3 Analysis strategy

The tagging DNN described in section 4.10 is applied to the search for inclusive $H \to b\bar{b}$ and $H \to c\bar{c}$ production. In this thesis the 2016 LHCb dataset is used to extract the upper limits on

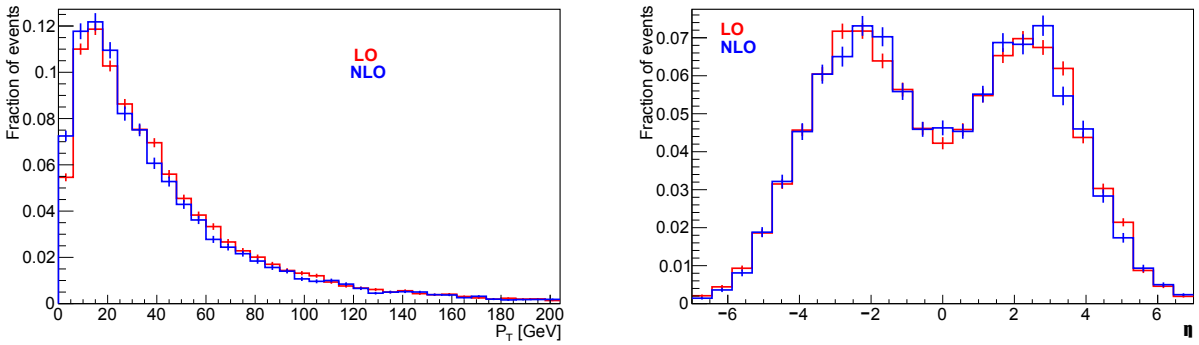

Figure 48: Left: transverse momentum distributions of the Higgs produced via ggF at LO (red) and NLO (blue) using MADGRAPH Monte Carlo. Right: pseudo-rapidity distributions of the Higgs produced via ggF at LO (red) and NLO (blue) using MADGRAPH Monte Carlo.

the cross sections.

The analysis strategy used to search for $H \to b\bar{b}$ and $H \to c\bar{c}$ can be summarized in the following steps:

- the selection described in section 5.2.5 is applied to the dijetSVSV dataset defined in section 5.2.4. In this way $H \to b\bar{b}$ and $H \to c\bar{c}$ dijet candidates are selected;
- the tagging DNN is evaluated for the two jets. Requirements on the DNN outputs are then applied to maximise the signal significance. These requirements are different for the $H \to b\bar{b}$ search and for the $H \to c\bar{c}$ search: in the former they have to keep $b\bar{b}$-dijets while removing the other jet flavours, in the latter they have to keep $c\bar{c}$-dijets. These requirements define the Signal Region (SR) for the $H \to b\bar{b}$ or $H \to c\bar{c}$ searches;
- a Control Region (CR), orthogonal to the SR, is defined by applying requirements to the DNN output, selecting dijet events with mixed flavours: bq, bc and cq. It has been assumed that, at first order, SM resonances are not present in the CR and we use these events to model the invariant mass distribution of the multi-jet QCD background;
- the invariant mass distribution of the QCD multi-jet background is obtained from the CR, and a Trasfer Function (TF) is employed to describe differences between SR and CR. The TF is obtained from the $b\bar{b}$ and $c\bar{c}$ simulation samples;
- the full dijet invariant mass model of the background is obtained by summing the templates of the multi-jet QCD background obtained from the CR, and the templates of the $Z \to b\bar{b}$ and $Z \to c\bar{c}$ backgrounds obtained from the simulation;
- the observed signal yield ($H \to b\bar{b}$ or $H \to c\bar{c}$) is determined with a fit to the dijet invariant mass distribution in the SR. In this fit the multi-jet QCD background and the signal yields are left free, while $Z \to b\bar{b}$ and $Z \to c\bar{c}$ yields are fixed to the SM expectation. A correction function for the TF that takes into account data/MC differences in its determination is also introduced in the fit, with floating parameters;
- the upper limits to the signal yields of $H \to b\bar{b}$ and $H \to c\bar{c}$ are calculated by using the CLs technique. The selection efficiencies are then applied to determine the upper limits to the $H \to b\bar{b}$ and $H \to c\bar{c}$ cross sections.

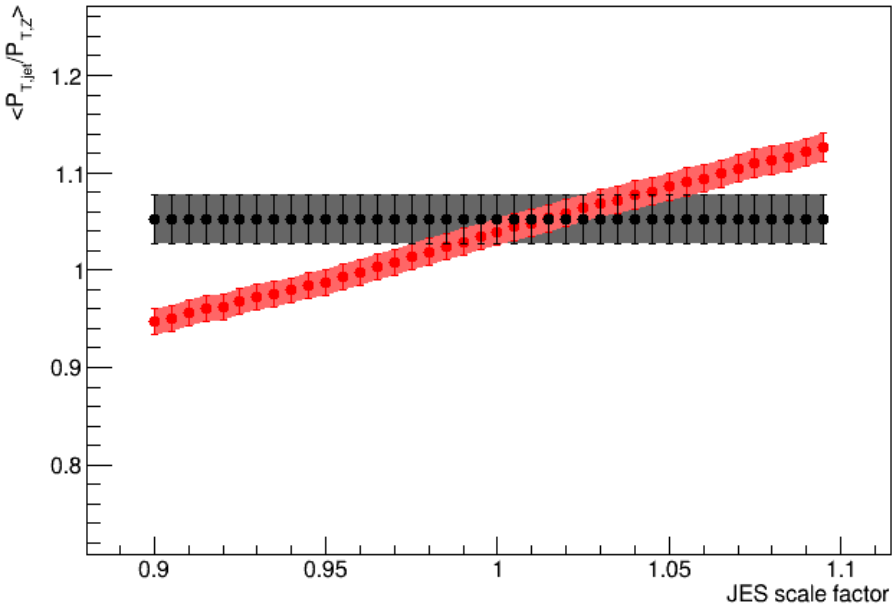

Figure 49: Jet energy scale in data (black dots) and Z→ $\mu^+\mu^-$ jets simulations (red dots) as a function of the scaling factor.

5.4 Definition of signal and control regions

The main challenge in the search for $H \to b\bar{b}$ and $H \to c\bar{c}$ is the precise description of the multi-jet QCD background. Indeed, Monte Carlo simulation fails to reproduce this background [28], therefore a data-driven approach has been used to derive the multi-jet QCD background. The tagging DNN outputs evaluated on the two selected jets have been used to define the signal region (SR) and the control region (CR). In Figures 53, 54 and 55 the DNN outputs distribution for jets in the dijetSVSV data, H $\to b\bar{b}$ and H $\to c\bar{c}$ samples can bee seen. From these figures it is evident that most of the data sample is formed by b-jets, with a smaller contributions from c-jets. The contamination from light jets is almost negligible, consistently with the mis-identification rate of the SV-tagging algorithm.

In order to define the SR for the $H \to b\bar{b}$ search, requirements to the DNN probabilities are applied to remove the contamination from c and light jets. In the same way the SR for the $H \to c\bar{c}$ search is defined by removing as much as possible the contamination from b and light jets. Requirements on the three DNN outputs (P_b,P_c,P_q) are determined in order to maximize the following quantity (S):

$$S = \frac{N_{sig}^{SR}}{\sqrt{N_{data}^{SR}}}, \tag{5.4}$$

where N_{data}^{SR} is the number of selected data events in the SR, and N_{sig}^{SR} is the number of selected signal MC events ($H \to b\bar{b}$ or $H \to c\bar{c}$) in the SR.

Since two out of three probability distributions P_b, P_c, P_q are independent ($P_b + P_c + P_q = 1$), the significance on the $H \to b\bar{b}$ ($H \to c\bar{c}$) signal is maximized by optimizing the requirement on P_b (P_c) and P_q. In practice, the number of expected signal and data events having both jets with $P_{b(c)} > p_{b(c),SR}$ and $P_q < p_{q,SR}$, where $p_{b(c),SR}$ and $p_{q,SR}$ runs over the range [0,1] by steps of 0.01,

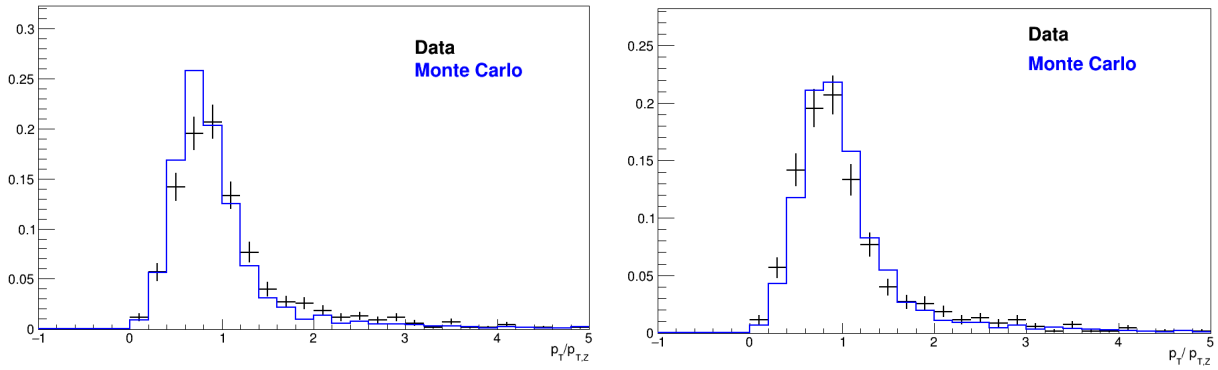

Figure 50: $p_{\mathrm{T}}(\mathrm{jet})/p_{\mathrm{T}}(Z)$ distributions before (left) and after (right) the JES factor (1.015) is applied to Monte Carlo events. Monte Carlo distribution is shown in blue, Data distribution is shown in black.

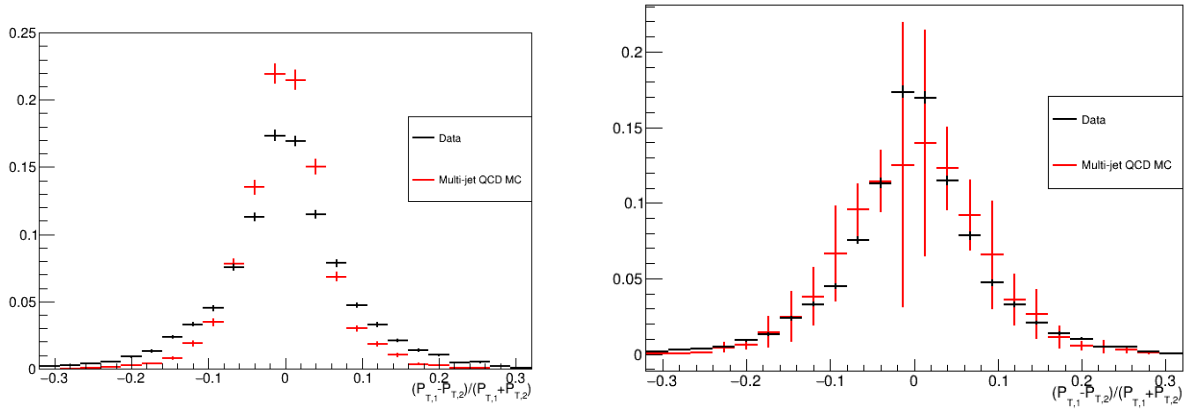

Figure 51: Distributions of $\frac{p_{\mathrm{T}}(\mathrm{jet1})-p_{\mathrm{T}}(\mathrm{jet2})}{p_{\mathrm{T}}(\mathrm{jet1})+p_{\mathrm{T}}(\mathrm{jet2})}$ in the low mass sideband < 45 GeV, compared between data (black) and simulation (red), before (left) and after (right) the JER factor is applied to Monte Carlo.

have been calculated for all possible combinations of $p_{b(c),SR}$ and $p_{q,SR}$. Then, the values of $p_{b(c),SR}$ and $p_{q,SR}$ that maximize S are kept.

The Control Region (CR) is defined as the region where just one of the two jets passes the SR DNN requirements, while a different selection is applied to the other jet. In this way mainly dijet events with mixed flavours are selected: since SM resonances are not expected, the CR is used to model the invariant mass distribution of the multi-jet QCD background.

The main goal for the determination of the CR, is to have a data sample with enough data statistics while minimizing the number of events from signal or resonances. The following figure of merit is considered for the determination of the CR working point:

$$F = \frac{N_{sig}^{CR}}{(N_{Data}^{CR})^{\frac{3}{2}}}. \tag{5.5}$$

where N_{data}^{CR} is the number of selected data events in the CR, and N_{sig}^{CR} is the number of selected signal MC events ($H \to b\bar{b}$ or $H \to c\bar{c}$) in the CR. The idea behind this last equation is to increase the data statistics ($1/\sqrt{N_{Data}^{CR}}$) while keeping the signal contamination ($N_{sig}^{CR}/N_{data}^{CR}$) close to the minimum.

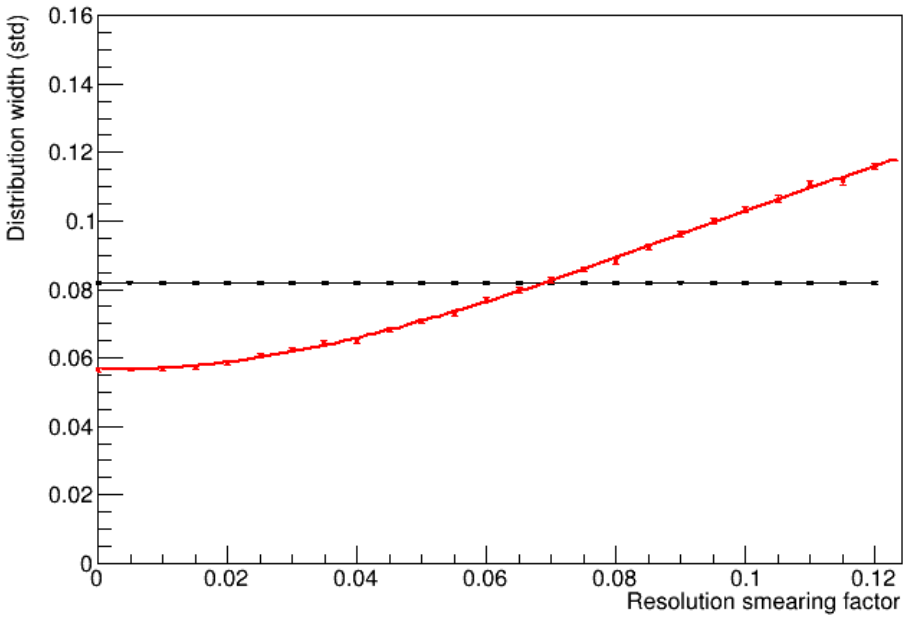

Figure 52: Tested σ values (x-axis) versus the jet asymmetry standard deviation for data (black) and Monte Carlo (red).

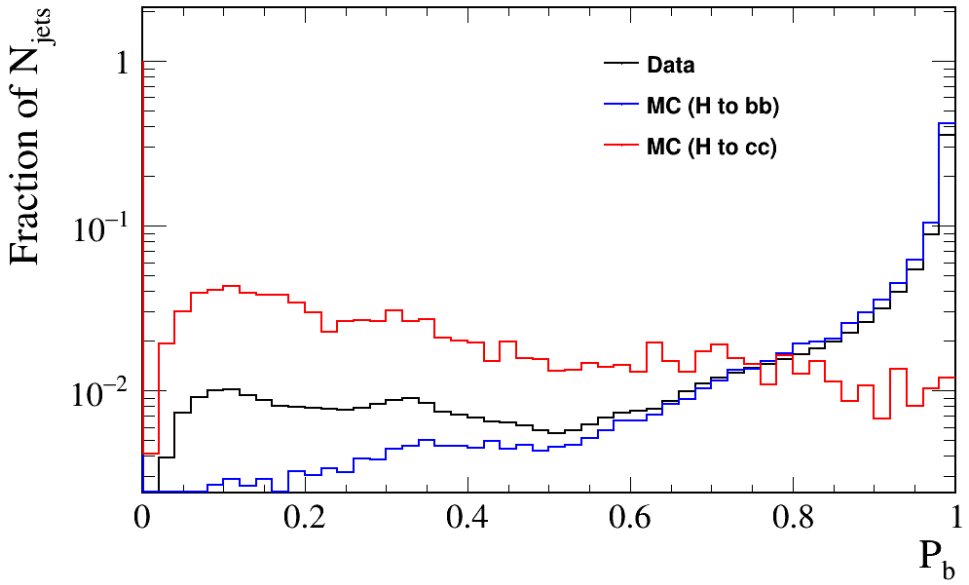

Figure 53: P_b DNN output distribution for data from the dijetSVSV data (black), $H \to b\bar{b}$ (blue) and $H \to c\bar{c}$ (red) selected with requirements listed in Table 13. Distributions are normalized to unit area.

For the determination of the CR we proceed in the following way:

1. events that satisfy the SR requirements are first discarded from the signal and dijetSVSV samples;

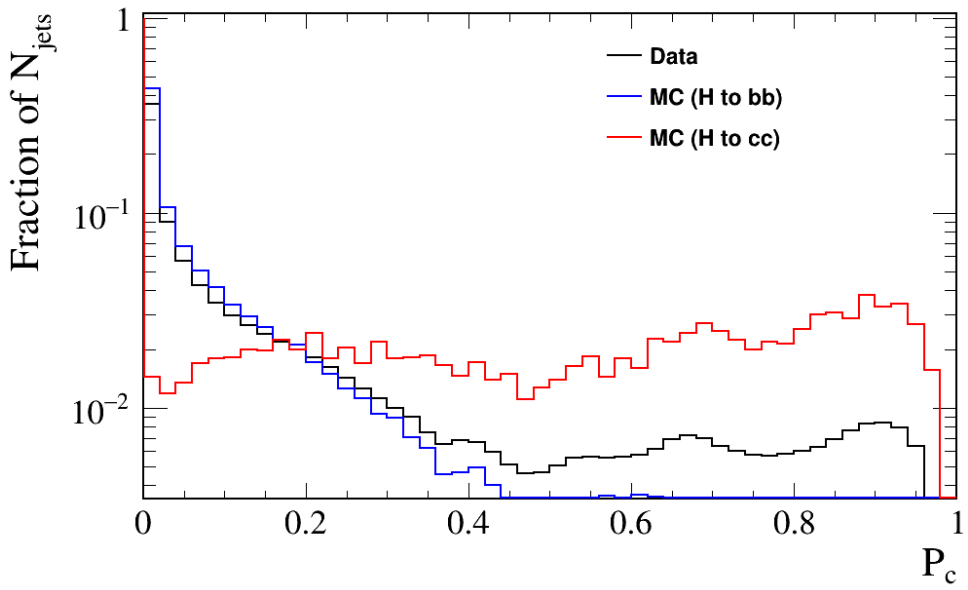

Figure 54: P_c DNN output distribution for data from the dijetSVSV data (black), $H \to b\bar{b}$ (blue) and $H \to c\bar{c}$ (red) selected with requirements listed in Table 13. Distributions are normalized to unit area.

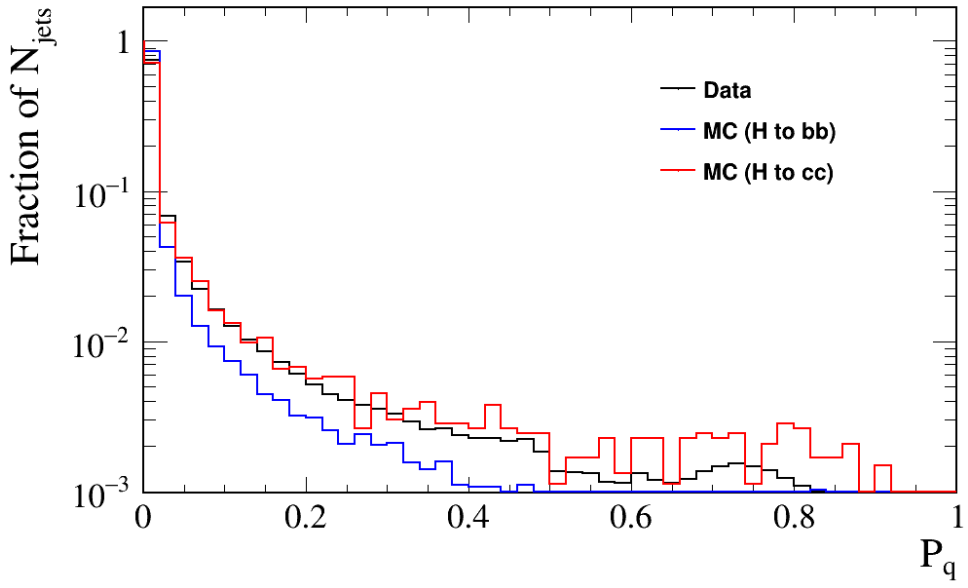

Figure 55: P_q DNN output distribution for data from the dijetSVSV data (black), $H \to b\bar{b}$ (blue) and $H \to c\bar{c}$ (red) selected with requirements listed in Table 13. Distributions are normalized to unit area.

2. signal and data events having one jet with $P_{b(c)} > p_{b(c),SR}$ and $P_q < p_{q,SR}$, where $p_{b(c),SR}$ and $p_{q,SR}$ are the requirements determined by the SR definition, are selected. In this way we ensure that at least one jet in the dijet is a real $b(c)$-jet.

3. a requirement of the other jet, $P_{b(c)} < p_{b(c),CR}$ is applied. This requirement is applied for having a second jet with a different flavour with respect to the first one;

4. $p_{b(c),CR}$ is varied in the range [0,1] by steps of 0.01, for each point F is calculated;

5. the values of $p_{b(c),CR}$ that minimizes F is kept.

As mentioned in section 4.10, the DNN flavour identification is not limited to jets with a SV. In the dijetSV sample, where just one SV is required, the contamination from light jets is higher and this can help in the definition of the CR. For this reason the CR procedure has been applied both to the dijetSVSV and the dijetSV samples: in the latter case the jet that satisfies the SR requirements must have a SV, while no SV requirements on the other jet are applied. The CR sample, dijetSVSV or dijetSV, with the highest purity ratio with respect to the SR is then used in the analysis:

$$\text{Purity ratio} = \frac{N_{sig}^{SR}/N_{Data}^{SR}}{N_{sig}^{CR}/N_{Data}^{CR}}. \tag{5.6}$$

In Figures 56, 57 and 58 the DNN outputs distribution for jets in data from the dijetSV data and $H \to b\bar{b}$ and $H \to c\bar{c}$ MC samples with the selection requirements listed in Table 13 except that only one jet in the pair is required to have a SV, can bee seen. It is evident that the number of light jets is higher with respect to the dijetSVSV sample.

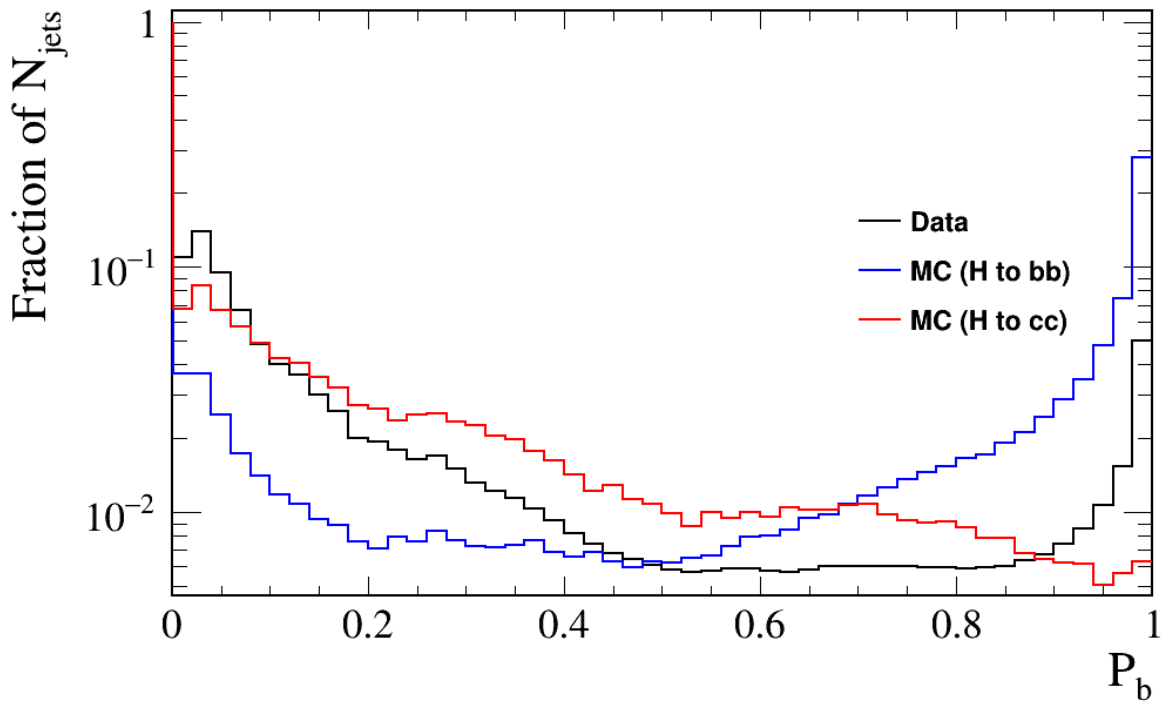

Figure 56: P_b DNN output distribution for dijetSV data (black),H $\to b\bar{b}$ (blue) and H $\to c\bar{c}$ (red) samples. Distributions are normalized to unit area.

5.4.1 Signal and control region definitions for $H \to b\bar{b}$

The maximum S for the $H \to b\bar{b}$ search is obtained when both jets have P_b greater than 0.25 and no requirement on P_q is found to further increase S. This requirement ($P_b > 0.25$) applied to both jets defines the SR of the $H \to b\bar{b}$ search. With these cuts the number of selected data events, expected signal events, the significance and purity (assuming SM cross sections) are summarized in Table 15. The improvement on the Higgs significance after applying the DNN cuts is 3%.

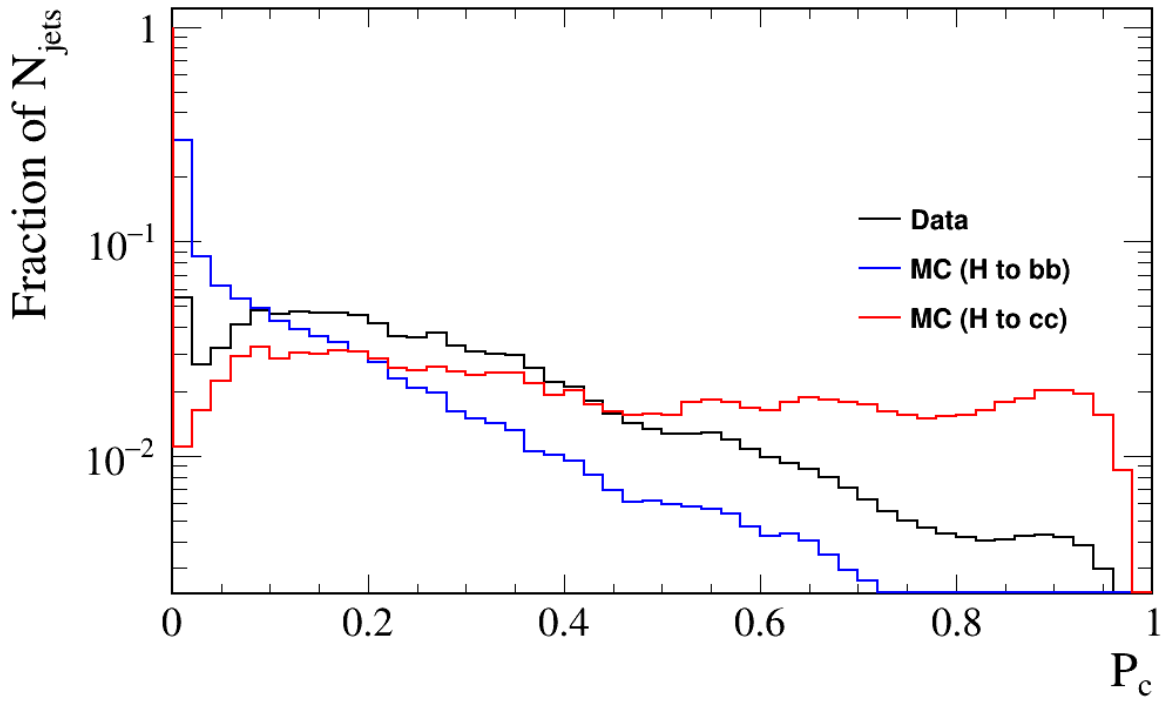

Figure 57: P_c DNN output distribution for dijetSV data (black), H $\to b\bar{b}$ (blue) and H $\to c\bar{c}$ (red) samples. Distributions are normalized to unit area.

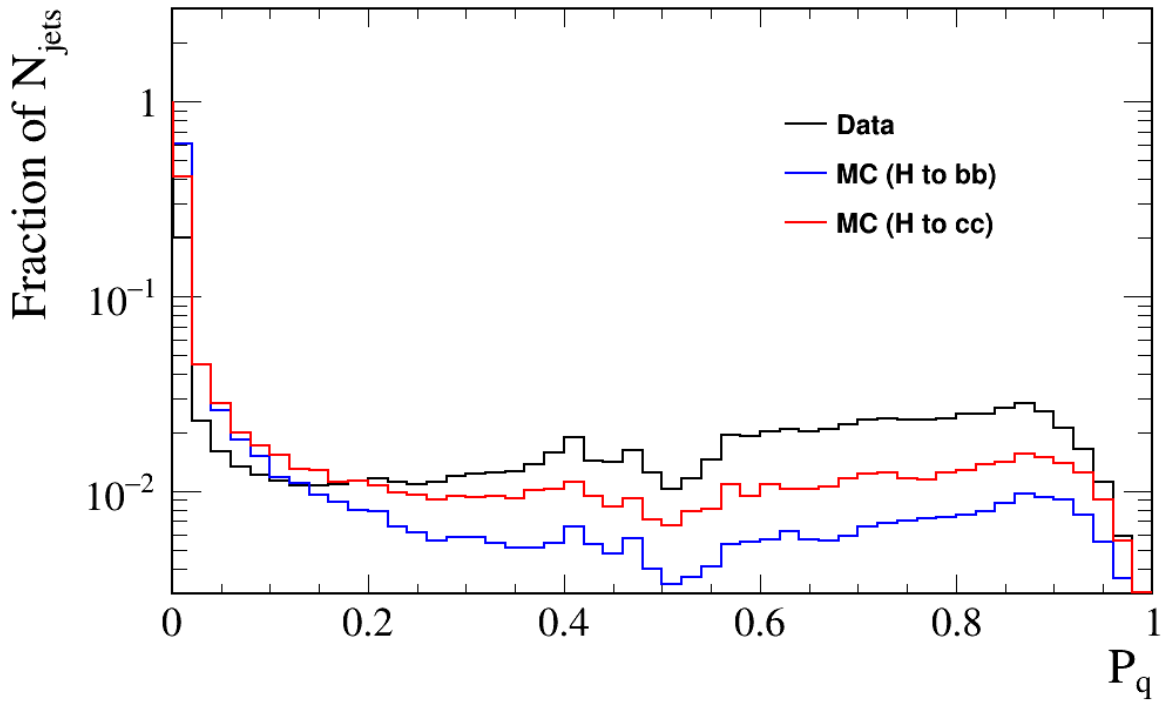

Figure 58: P_q DNN output distribution for dijetSV data (black),H $\to b\bar{b}$ (blue) and H $\to c\bar{c}$ (red) samples. Distributions are normalized to unit area.

SR Requirements	$N^{SM}_{H\to b\bar{b}}$	N^{SR}_{data}	Significance	Purity
$P_b > 0.25$	346	2.1e+07	0.08	$1.7 \cdot 10^{-5}$

Table 15: Requirement for the SR definition, number of selected data events, expected signal events, significance and purity (assuming SM cross sections) in the SR of the $H \to b\bar{b}$ search.

Once the SR has been defined, the procedure for the definition of the CR has been performed. Events where one of the two jets satisfies the $P_b > 0.25$ requirement are selected, and P_b requirements on the other jet have been tuned to minimize F (Eq. 5.5). The configurations that minimize F are reported in the Tab. 16 for the dijetSVSV and dijetSV samples. The highest purity ratio is

	dijetSVSV	dijetSV
CR requirements on DNN outputs	$P_b > 0.25$ on one jet $P_b < 0.25$ the other jet	$P_b > 0.25$ on one jet with a SV $P_b < 0.22$ the other jet
N^{CR}_{data}	3.2e+06	1.1e+06
$N^{CR}_{H\to b\bar{b}}$	22	3.6
Purity ratio	2.5	5.5

Table 16: CR requirements, data events in the CR and purity ratio with respect to SR for the $H \to b\bar{b}$ search.

obtained with the dijetSV sample, therefore it is used for the CR of the $H \to b\bar{b}$ search. The invariant mass distribution of data from the dijetSV sample in the CR is shown in Figure 59. Table

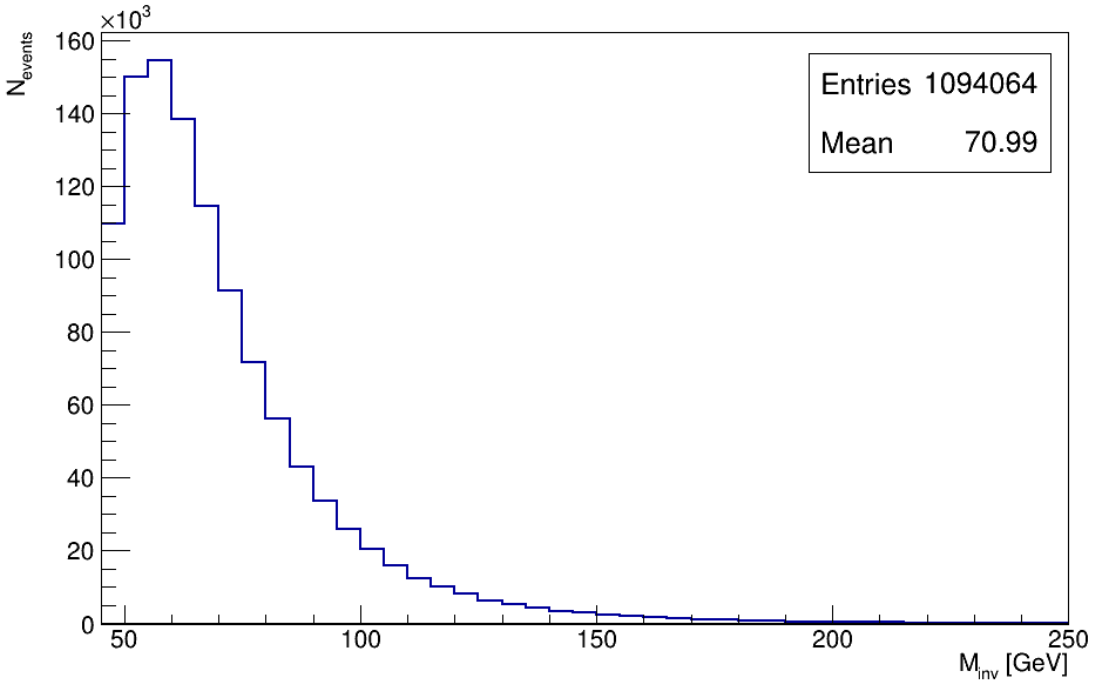

Figure 59: Invariant mass distribution in the CR of the $H \to b\bar{b}$ search, obtained with the dijetSV sample.

17 shows the expected number of $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ events in the SR and CR, and with the requirements of having the invariant mass in the range $[45, 250]$ GeV, that is the

range chosen for the final fit. Possible contamination from the $W \to qq'$ in this CR has also been

Process	N_{exp}^{SR}	N_{exp}^{SR} with $m_{jj} \in [45, 250]$ GeV	N_{exp}^{CR}	N_{exp}^{CR} with $m_{jj} \in [45, 250]$ GeV
$Z \to b\bar{b}$	153529	153171	1053	1040
$Z \to c\bar{c}$	1907	1787	193	192
$H \to b\bar{b}$	346	346	4	3
$H \to c\bar{c}$	0.9	0.9	0.1	0.1

Table 17: Table with the number of expected $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ in SR and CR of the $H \to b\bar{b}$ search, and with the requirements of having the invariant mass in the range $[45, 250]$ GeV.

evaluated: 800000 events has been generated with Pythia 8, and the $W \to qq'$ cross section in the LHCb acceptance has been calculated with MADGRAPH Monte Carlo. Table 18 shows the NLO cross section, the acceptance, the total efficiency and the number of events expected in the CR in the invariant mass range [45,250] GeV, taking into account the pre-scale factor. The efficiency includes the requirements on the trigger and the jet kinematic applied to the dijetSV sample (Table 13), the presence of at least one jet tagged in the pair and the requirements on P_b and P_q defined for the $H \to b\bar{b}$ CR. The effect of the contamination of the CR with H, Z and W renonances on

Process	σ_{SM} [pb]	A	ϵ (%)	$N_{CR,exp}^{SM}$
$W \to qq'$	99832	0.127	1.5	3050

Table 18: $W \to qq'$ SM cross section, acceptance, efficiency and expected events with $\mathcal{L} = 1.6$ fb^{-1} in the CR after the selection described in this section.

the upper limit calculation is evaluated in section 6.6.2.

5.4.2 Signal and control region definitions for $H \to c\bar{c}$

The maximum S for the $H \to c\bar{c}$ search is obtained when both jets have P_c greater than 0.15 and no requirement on P_q is found to further increase the significance. The selected data events in the SR, the expected number of signal events, the significance and purity (assuming SM cross sections) calculated on the entire dijet invariant mass spectrum are summarized in Table 19. The

Requirements	$N_{H\to c\bar{c}}^{SM}$	N_{data}^{SR}	Significance	Purity
$P_c > 0.15$	1.8	4.1e+06	0.0009	$4.6 \cdot 10^{-07}$

Table 19: Requirement for the SR definition, number of selected data events, expected signal events (assuming SM cross sections), the significance and purity in the SR of the $H \to c\bar{c}$ search.

improvement on the $H \to c\bar{c}$ significance after applying the DNN cuts is 87%. It is clear that the large improvement is due to the removal of a fraction of the b-jet background, that forms most of the data sample.

The working point for $H \to c\bar{c}$ CR has been calculated using the dijetSVSV and dijetSV data samples. The same procedure of the $H \to b\bar{b}$ have been followed: events where one of the two jets satisfies the $P_c > 0.15$ requirement are selected, and P_c requirements have been tuned on the other jet in order to minimize F (Eq. 5.5). The results are shown in table 20. The dijetSVSV sample is chosen for the analysis since the purity ratio is higher (35). It is worth notice that this technique is more effective for the $H \to c\bar{c}$ search with respect to $H \to b\bar{b}$ search, since in the latter the purity ratio was 5.5. The dijet invariant mass distribution of data in the CR is shown in Figure 60. Table 21 shows the expected number of $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ events in the SR and CR,

	dijetSVSV	dijetSV
CR requirements on DNN outputs	$P_c > 0.15$ on one jet $P_c < 0.02$ the other jet	$P_c > 0.15$ on one jet with a SV $P_c < 0.02$ the other jet
N^{CR}_{data}	4.5e+06	1.3e+05
$N^{CR}_{H\to c\bar{c}}$	0.06	0.002
Purity ratio	35	34

Table 20: CR requirements, data events in the CR and purity ratio with respect to SR for the $H \to c\bar{c}$ search.

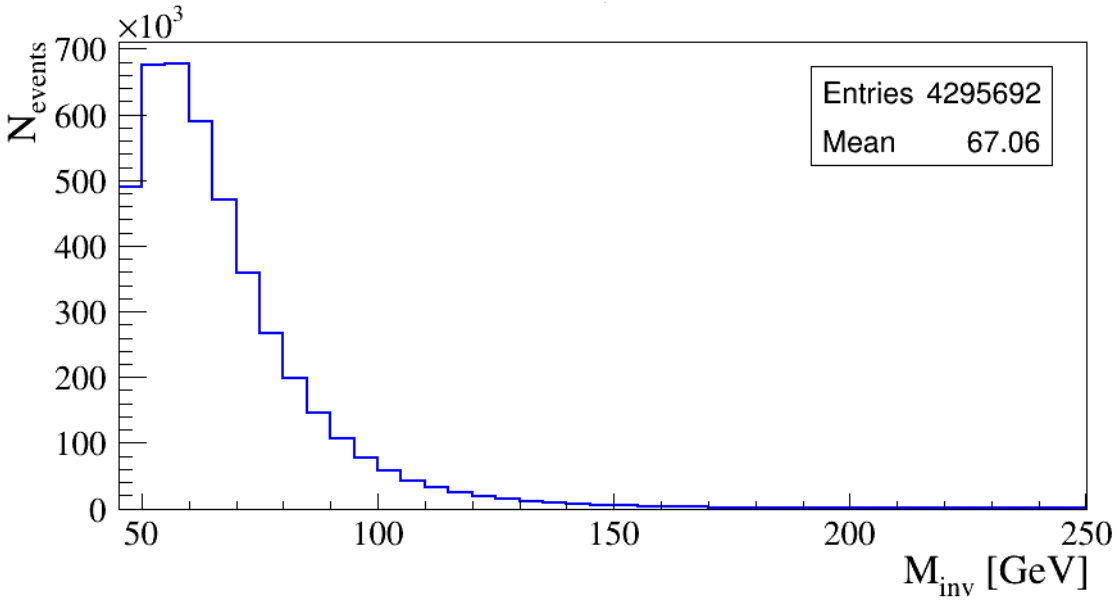

Figure 60: Invariant mass distribution in the CR of the $H \to c\bar{c}$ search, obtained with the dijetSVSV sample.

and with the requirements of having the invariant mass in the range $[45, 250]$ GeV, that is the range chosen for the final fit. Even in this case the contamination from the $W \to qq'$ in the $H \to c\bar{c}$ CR

Process	N^{SR}_{exp}	N^{SR}_{exp} with $m_{jj} \in [45, 250]$ GeV	N^{CR}_{exp}	N^{CR}_{exp} with $m_{jj} \in [45, 250]$ GeV
$Z \to b\bar{b}$	8930	8896	31938	31858
$Z \to c\bar{c}$	4571	4372	83	78
$H \to b\bar{b}$	19	18	72	72
$H \to c\bar{c}$	1.9	1.8	0.06	0.06

Table 21: Table with the number of expected $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ in SR and CR of the $H \to c\bar{c}$ search, and with the requirements of having the invariant mass in the range $[45, 250]$ GeV.

has been evaluated: with a selection efficiency of 0.03 %, 6100 $W \to qq'$ events are found.

5.5 Simulated events correction

Other differences between real data and simulated events affects the events selection efficiency used to calculate the Monte Carlo yields in Table 14. This is due for example to difference in the jet hadronization, detector and trigger simulation modelling. These corrections may affect both

the shape and the normalization of the Monte Carlo sample and are here discussed. Details about how these scale factors have been calculated will be given in the next sections. The uncertainties on these corrections are also calculated and are considered as systematic uncertainties, as will be shown in section 6. The corrections have been applied for:

1. the SV tagging efficiency, that as shown in section 4.9.2, it is different between data and Monte Carlo. The ratio values between the efficiencies measured in data and those obtained in simulation depends on the jet p_T, as has been shown in Figure 37, and have been used to weight jets in the events, as will be described in detail in section 5.5.1;

2. selection efficiency on Deep Neural Network outputs is found to be different between data and Monte Carlo events. The ratio values between the efficiencies measured in data and those obtained in simulation is calculated in section 5.5.2 as a function of the jet p_T and have been used as weight for jets in the event counting, on top of the event weighting due to the SV tag;

3. An overall factor has been calculated to take into account for the fact that the Global Event Cut (GEC) (see section 3.3.1) efficiency is higher for Monte Carlo simulation with respect to data. This correction takes into account for differences in the multiplicity in the SPD detector (nSPD), not well simulated in the LHCb Monte Carlo. As can be seen from Figure 61 the average nSPD of simulated Monte Carlo events is lower than data. The number of Monte Carlo events that are selected by the L0 requirements on nSPD is then over estimated. The procedure followed to calculate the corrections and their uncertainties is described in section 5.5.3.

4. A further correction to the number of H expected events have to be applied to take into account for the fact that the Higgs ggF production have been used in the simulation, neglecting VBF and VH. The impact of neglecting VBF and VH have been calculated in the following way: MADGRAPH has been used to generate the VBF, VH and ggF Higgs production processes at NLO, with the Higgs decaying into $b\bar{b}$. The cross sections and the acceptance calculated by requiring the two b quarks between $2 < \eta < 5$ are summarized in the following Table 22:

Process	σ (pb) at NLO	A
$V + H(H \to b\bar{b})$	2.1	0.06
$VBF(H \to b\bar{b})$	3.6	0.04
$ggF(H \to b\bar{b})$	32.3	0.04

Table 22: Higgs process cross sections and acceptance calculated with MADGRAPH at NLO.

The weighted acceptance, calculated as: $A_w = \frac{\sigma_{VH} \cdot A_{VH} + \sigma_{VBF} \cdot A_{VBF} + \sigma_{ggF} \cdot A_{ggF}}{\sigma_{VH} + \sigma_{VBF} + \sigma_{ggF}}$ is found to be a factor 1.04 larger than the ggF acceptance. This factor is used to scale the number of Higgs events.

5.5.1 SV tagging correction

The differences between the efficiencies measured in data and those obtained in simulation and corresponding uncertainties have been measured in [92] in events with fully reconstructed B and D in addition to a jet and have been shown in section 4.9.2. Despite these weights having been measured with Run 1 data, in [97] it has been demonstrated using a $B \to J/\psi K$ data sample that Run 1 tagging efficiencies are compatible with Run 2 at a 3% level. In this thesis, as conservative approach, a 3% uncertainty have been added in quadrature to the nominal Run 1 uncertainty, to take into account the Run 1 - Run 2 compatibility, as it has been done

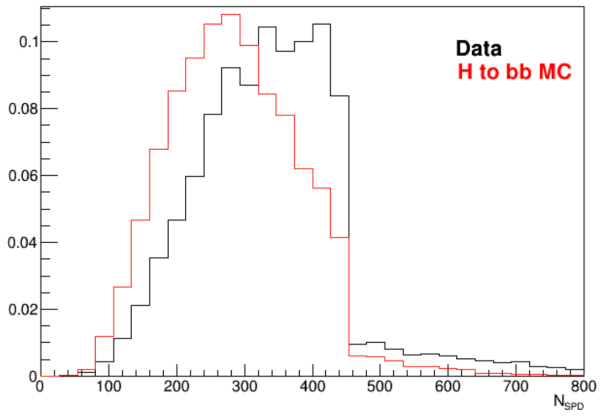

Figure 61: Comparison between nSPD distributions for Monte Carlo (red) and data (black).

in [28]. Correction factors have been calculated and used to estimate the Higgs and Z yields for the fit setup: the mean value of the ratio calculated in Table 38 is used to weight jets as a function of their p_T. Two weights are applied, one for each jet in the pair and their product has been used as weight associated to the event. For the ($\pm 1\sigma$) variation with respect to the mean correction, the mean $\pm$ its error is used as weight. Table 23 and 24 show the $\pm 1\sigma$ variation in the number of events respect to the number of events found with the mean. The effect of this corrections is also to vary the H and the Z Monte Carlo sample invariant mass shapes. For example, Figures 62 show invariant mass distributions of the $H \to b\bar{b}$ (left) and $H \to c\bar{c}$ (right) Monte Carlo, in the $H \to b\bar{b}$ SR and $H \to c\bar{c}$ SR respectively. Different colors correspond to the invariant mass: before the tagging correction (blue), after the tagging correction (blue), with $+1\sigma$ variation with respect to the correction (black) and with -1σ variation with respect to the correction (red). Both the shape and normalization variations have been considered as systematic uncertainties on the final upper limit calculation.

Process	Correction	$+1\sigma$ w.r.t. corrected N_{ev}	-1σ w.r.t. corrected N_{ev}
$H \to b\bar{b}$	1.15	1.17	0.84
$Z \to b\bar{b}$	1.07	1.17	0.84
$H \to c\bar{c}$	0.81	1.30	0.74
$Z \to c\bar{c}$	0.96	1.22	0.80

Table 23: Tagging efficiency correction factors and $\pm 1\sigma$ variation in the number of events, for $H \to b\bar{b}$ search

Process	Correction	$+1\sigma$ w.r.t. corrected N_{ev}	-1σ w.r.t. corrected N_{ev}
$H \to b\bar{b}$	1.13	1.17	0.84
$Z \to b\bar{b}$	1.05	1.17	0.84
$H \to c\bar{c}$	0.82	1.30	0.74
$Z \to c\bar{c}$	0.98	1.22	0.80

Table 24: Tagging efficiency correction factors and $\pm 1\sigma$ variation in the number of events for the $H \to c\bar{c}$ search

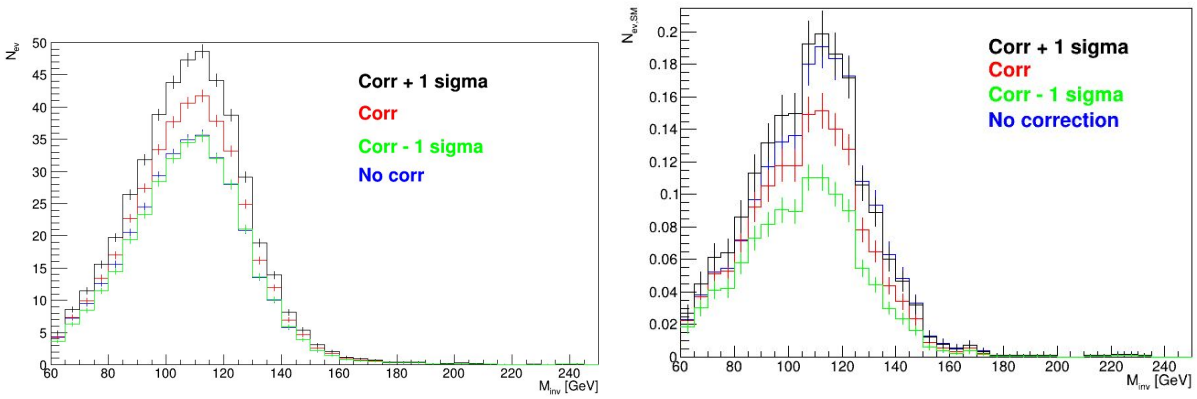

Figure 62: Invariant mass distributions of the $H \to b\bar{b}$ (left) and $H \to c\bar{c}$ (right) Monte Carlo, in the $H \to b\bar{b}$ SR and $H \to c\bar{c}$ SR respectively. Different colors correspond to the invariant mass: before the tagging correction (blue), after the tagging correction (blue), with $+1\sigma$ variation with respect to the correction (black) and with -1σ variation with respect to the correction (red).

5.5.2 Deep Neural Network correction

The signal efficiency of the DNN requirements applied for the SR definition can differ between data and MC. A related correction and systematic uncertainty should be evaluated. A tag-and-probe technique has been employed using the dijetSVSV sample and the Multi-jet QCD Monte Carlo. The tag is selected by applying a tight cut on the DNN output P_b (or P_c) for one jet in the sample, while the probe is selected by requiring a jet with $\Delta\phi > 2.7$ with respect to the tag. In this way we will obtain a sample of probe jets with enriched b (or c) component. The efficiency of applying the SR analysis cut is then calculated with the probe sample. The systematic uncertainty has been determined by varying the P_b (or P_c) requirement on the tag jet, obtaining in this way different levels of c (or b) and light jets contamination. Figure 63 shows the ratio between the DNN efficiency obtained from data and Monte Carlo, with probe jets enriched in b component. Tag jets are selected by applying $P_b > 0.7$ (blue), $P_b > 0.8$ (black), $P_b > 0.9$ (red) while the efficiency is evaluated by requiring SR $H \to b\bar{b}$ DNN cut ($\mathrm{P}_b > 0.25$) to the probe jet. Figure 64 shows the ratio between the DNN efficiency obtained from data and Monte Carlo, with probe jets enriched in c component. Tag jets are selected by applying $P_c > 0.7$ (blue), $P_c > 0.8$ (black), $P_c > 0.9$ (red) while the efficiency is evaluated by requiring SR $H \to b\bar{b}$ DNN cut ($\mathrm{P}_b > 0.25$) to the probe jet. Figure 65 shows the ratio between the DNN efficiency obtained from data and Monte Carlo, with probe jets enriched in c component. Tag jets are selected by applying $P_c > 0.7$ (blue), $P_c > 0.8$ (black), $P_c > 0.9$ (red), while the efficiency is evaluated by requiring SR $H \to c\bar{c}$ DNN cut ($\mathrm{P}_c > 0.15$) to the probe jet. Figure 66 shows the ratio between the DNN efficiency obtained from data and Monte Carlo, with probe jets enriched in b component. Tag jets are selected by applying $P_b > 0.7$ (blue), $P_b > 0.8$ (black), $P_b > 0.9$ (red) , while the efficiency is evaluated by requiring SR $H \to c\bar{c}$ DNN cut ($\mathrm{P}_c > 0.15$) to the probe jet.

The mean efficiency for each p_T interval has been applied to weight jets of H and Z Monte Carlo samples and to correct the number of events with the same weighing procedure explained for the SV tagging. The weights for the ($\pm 1\sigma$) variation with respect to the mean correction have been calculated by using the difference between the variation in the mean efficiency for the different levels of c (or b) contamination. Table 23 and 24 show the correction factor and the $\pm 1\sigma$ variation on the number of events.

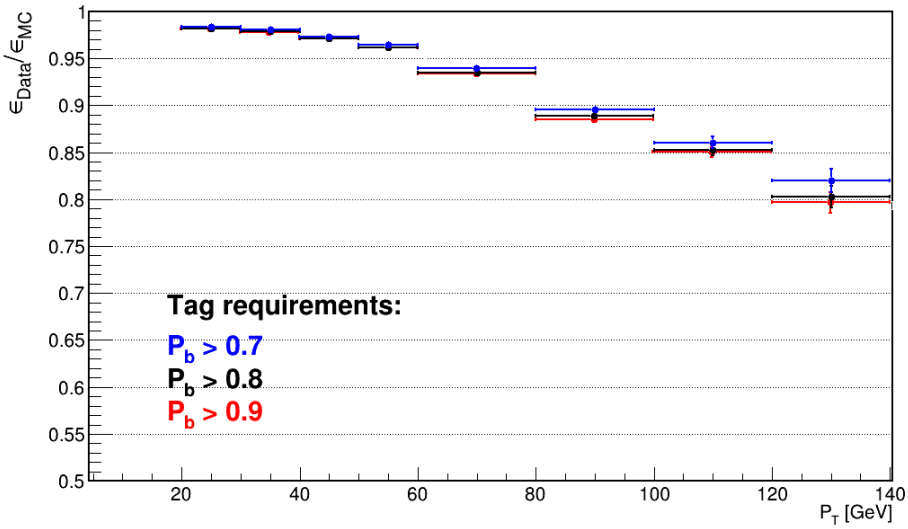

Figure 63: Ratio between the DNN efficiency obtained from data and Monte Carlo, with probe jets enriched in b component. Tag jets are selected by applying $P_b > 0.7$ (blue), $P_b > 0.8$ (black), $P_b > 0.9$ (red) while the efficiency is evaluated by requiring SR $H \to b\bar{b}$ DNN cut ($P_b > 0.25$).

Process	Correction factor	$+1\sigma$ w.r.t. corrected N_{ev}	-1σ w.r.t. corrected N_{ev}
$H \to b\bar{b}$	0.911	1.006	0.996
$Z \to b\bar{b}$	0.948	1.004	0.997
$H \to c\bar{c}$	1.11	1.08	0.96
$Z \to c\bar{c}$	1.16	1.14	0.96

Table 25: DNN correction factor on the number of events and the $\pm$ 1 σ variation for the $H \to b\bar{b}$ search

Process	Correction factor	$+1\sigma$ w.r.t. corrected N_{ev}	-1σ w.r.t. corrected N_{ev}
$H \to b\bar{b}$	1.68	0.98	1.01
$Z \to b\bar{b}$	1.46	0.99	1.01
$H \to c\bar{c}$	0.75	0.93	1.04
$Z \to c\bar{c}$	0.80	0.94	1.04

Table 26: DNN correction factor on the number of events and the $\pm$ 1 σ variation for the $H \to c\bar{c}$ search.

5.5.3 Global Event Cut correction

The correction for the Global Event Cut efficiency (ϵ^{GEC}) and associated systematic uncertainty have been determined as in the $b\bar{b}$ and $c\bar{c}$ cross sections measurement [28]. Data sample with events having fired different trigger selections with loose or no nSPD cuts are used. There are *L0MuonEW* (no nSPD cut), *L0JetPh* and *L0JetEl* (nSPD<10000). No Monte Carlo simulated events have been found with nSPD>10000. The cumulative distribution of nSPD in data for samples triggered by *L0MuonEW* (blue), *L0JetPh* (red) and *L0JetEl* (green) is shown in Figure 67.

The data GEC efficiencies have been determined in events having triggered *L0MuonEW*,

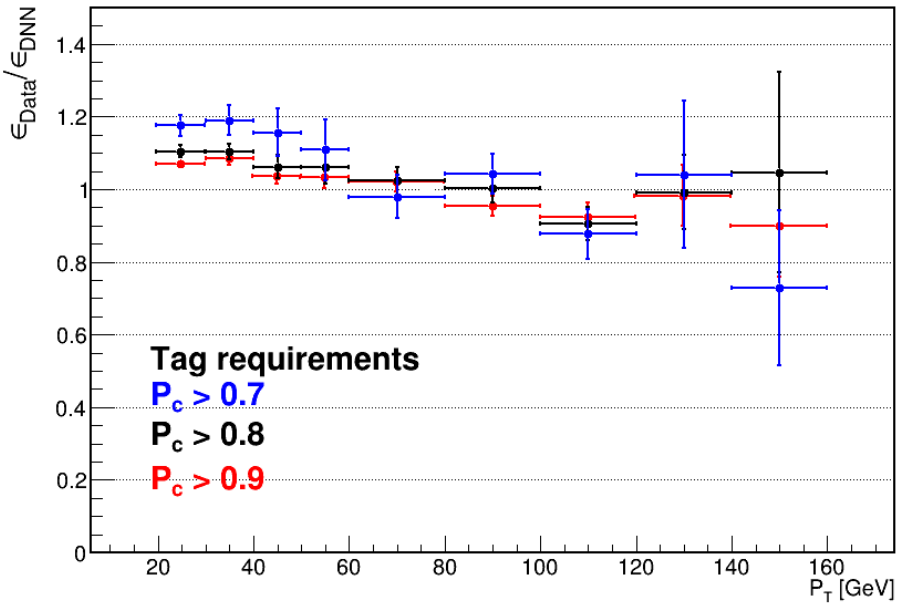

Figure 64: Ratio between the DNN efficiency obtained from data and Monte Carlo, with probe jets enriched in c component. Tag jets are selected by applying $P_c > 0.7$ (blue), $P_c > 0.8$ (black), $P_c > 0.9$ (red) while the efficiency is evaluated by requiring SR $H \to b\bar{b}$ DNN cut ($P_b > 0.25$).

L0JetPh or *L0JetEl*. Data efficiency of the nSPD cuts (<450 and <900) in samples selected by the different trigger lines are shown in Table 27.

In order to apply this correction in simulation, the nSPD distribution in MC are corrected to match that of data, since they are underestimed in simulation. The distributions of nSPD for data events that pass the *L0MuonEWDecision*, *L0JetEl* and *L0JetPh* lines, in different intervals of dijet invariant mass ([45,83]GeV, [83,130] GeV, [130,250] GeV, corresponding to the Higgs peak region and to the sidebands), are considered. A scale factor has been applied to nSPD in Higgs Monte Carlo to match the distribution in data. The scale factors are obtained by minimizing the χ^2 between the two histograms. The distribution of the Monte Carlo nSPD before and after the correction for the scale factor and the data distribution are shown for all the three trigger lines and for each invariant mass range from Figure 68 to 72. The nSPD distribution of Monte Carlo before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization are shown in red. The nSPD distribution of data is shown in black.

Table 27 shows: the efficiencies on data and Monte Carlo after the nSPD correction, obtained by applying the requirement nSPD<450 and nSPD<900 to events that pass each the three trigger lines, for the three intervals of invariant mass, and the last column shows the scale factors applied to the nSPD for each trigger lines and invariant mass interval.

Three corrected efficiencies are then calculated for the Higgs and Z Monte Carlo samples, one for each of the three trigger lines *L0MuonEW* ($\epsilon_{MC,L0MuonEW,tot}$), *L0JetPh* ($\epsilon_{MC,L0JetPh,tot}$), *L0JetEl* ($\epsilon_{MC,L0JetEl,tot}$): the scale factors relative to each line are applied to the nSPD of each Monte Carlo event depending on the invariant mass. The requirement of nSPD<450 or nSPD<900 is applied to the corrected nSPD depending on the trigger lines fired by the Monte Carlo event. Events that pass the nSPD requirement are weighted for the data/MC

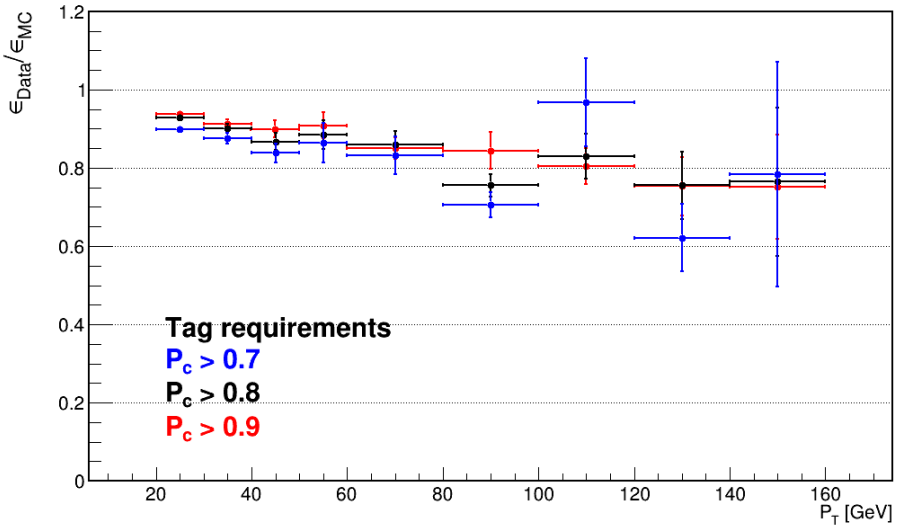

Figure 65: Ratio between the DNN efficiency obtained from data and Monte Carlo, with probe jets enriched in c component. Tag jets are selected by applying $P_c > 0.7$ (blue), $P_c > 0.8$ (black), $P_c > 0.9$ (red), while the efficiency is evaluated by requiring SR $H \to c\bar{c}$ DNN cut ($P_c > 0.15$).

M_{inv} [0,83] GeV	nSPD<450	nSPD<900	Scale factor
$\epsilon_{Hbb,L0MuonEW}$	0.49 ± 0.02	0.962 ± 0.007	1.49
$\epsilon_{Hbb,L0JetPh}$	0.53 ± 0.02	0.987 ± 0.005	1.47
$\epsilon_{Hbb,L0JetEl}$	0.49 ± 0.01	0.982 ± 0.004	1.47
$\epsilon_{data,L0MuonEW}$	0.502 ± 0.002	0.967 ± 0.0009	-
$\epsilon_{data,L0JetPh}$	0.533 ± 0.006	0.983 ± 0.002	-
$\epsilon_{data,L0JetEl}$	0.488 ± 0.003	0.979 ± 0.0008	-
M_{inv} [83,130] GeV	nSPD<450	nSPD<900	Scale factor
$\epsilon_{Hbb,L0MuonEW}$	0.507 ± 0.008	0.971 ± 0.003	1.49
$\epsilon_{Hbb,L0JetPh}$	0.557 ± 0.008	0.978 ± 0.002	1.45
$\epsilon_{Hbb,L0JetEl}$	0.517 ± 0.005	0.977 ± 0.001	1.48
$\epsilon_{data,L0MuonEW}$	0.498 ± 0.006	0.971 ± 0.002	-
$\epsilon_{data,L0JetPh}$	0.549 ± 0.008	0.981 ± 0.002	-
$\epsilon_{data,L0JetEl}$	0.503 ± 0.004	0.980 ± 0.001	-
M_{inv} [130,250] GeV	nSPD<450	nSPD<900	Scale factor
$\epsilon_{Hbb,L0MuonEW}$	0.54 ± 0.02	0.985 ± 0.006	1.27
$\epsilon_{Hbb,L0JetPh}$	0.53 ± 0.02	0.984 ± 0.004	1.37
$\epsilon_{Hbb,L0JetEl}$	0.55 ± 0.01	0.983 ± 0.003	1.32
$\epsilon_{data,L0MuonEW}$	0.54 ± 0.02	0.978 ± 0.005	-
$\epsilon_{data,L0JetPh}$	0.58 ± 0.02	0.987 ± 0.005	-
$\epsilon_{data,L0JetEl}$	0.535 ± 0.009	0.983 ± 0.002	-

Table 27: Efficiencies on data and Monte Carlo after the nSPD correction, obtained by applying the requirement nSPD<450 and nSPD<900 to events that pass each the three trigger lines, for the three intervals of invariant mass, and the last column shows the scale factors applied to the nSPD for each trigger lines and invariant mass interval.

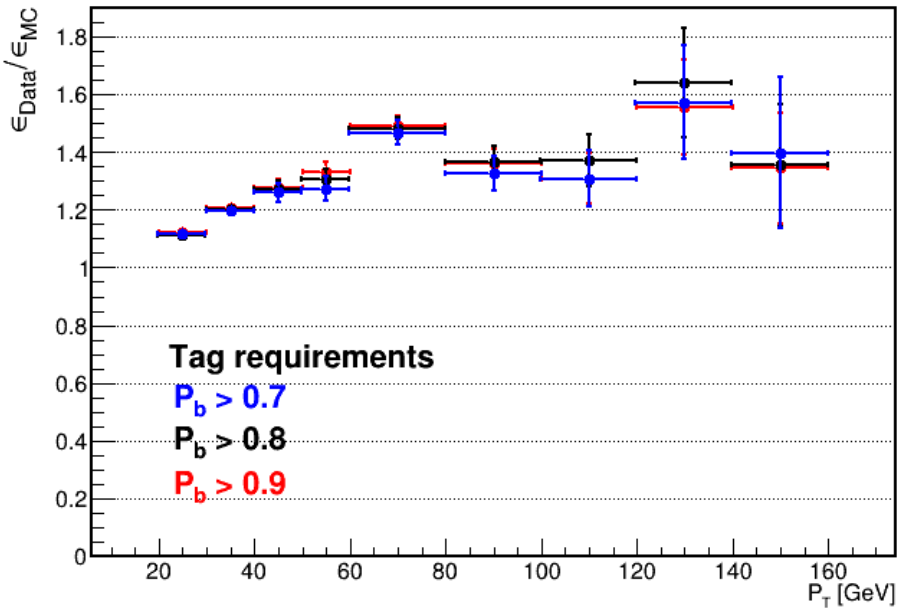

Figure 66: Ratio between the DNN efficiency obtained from data and Monte Carlo, with probe jets enriched in b component. Tag jets are selected by applying $P_b > 0.7$ (blue), $P_b > 0.8$ (black), $P_b > 0.9$ (red) , while the efficiency is evaluated by requiring SR $H \to c\bar{c}$ DNN cut ($P_c > 0.15$).

efficiency ratio relative to the trigger line and invariant mass range, to take into account for the data/MC differences.

The resulting efficiencies and the $\pm$ 1 σ variations are shown in Table 28. The mean values are calculated as the mean of the two outermost values, while the distance of the mean from the outermost values is used as $\pm$ 1 σ variation and will be evaluated as systematic uncertainty.

Process	$\epsilon_{MC,L0MuonEW,tot}$	$\epsilon_{MC,L0JetPh,tot}$	$\epsilon_{MC,L0JetEl,tot}$	Mean + upper/lower variation
$H \to c\bar{c}$	0.64	0.66	0.63	0.65 $\pm$ 0.02
$H \to b\bar{b}$	0.63	0.65	0.62	0.64$\pm$ 0.02
$Z \to b\bar{b}$	0.69	0.70	0.68	0.694$\pm$ 0.009
$Z \to c\bar{c}$	0.69	0.71	0.68	0.70$\pm$ 0.01

Table 28: Efficiency correction calculated for all Higgs and Z samples. In the last column the mean between the highest and the lowest efficiencies has been calculated together with the $\pm$ 1 σ variation. This variation has been used as systematic for the limit calculation.

The number of Monte Carlo events in the invariant mass range [45,250] GeV have been calculated with the requirements on the DNN that define the SR and the correction factors calculated in this section. First three rows of Table 29 and 30 show: the correction factors to be applied to the number of events due to the SV tagging, the SV tagging and the DNN weights considered together, and the GEC, for the $H \to b\bar{b}$ and $H \to c\bar{c}$ search, respectively. The total correction factor is written in the fourth row. The acceptance factor used to scale the Higgs acceptance is shown in row five of Tables 29 and 30. The last two rows show the number of events before and after the corrections. The number of events after corrections are the number of Higgs and Z events used in the following analysis.

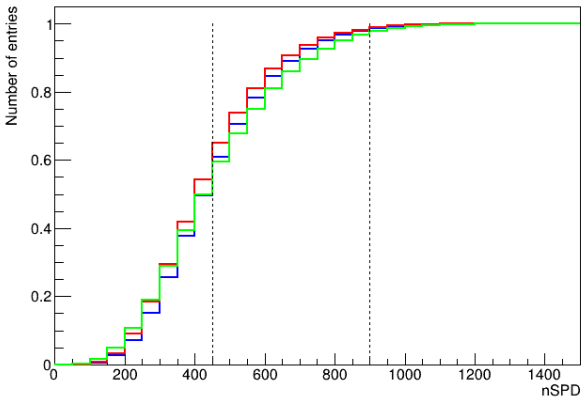

Figure 67: Cumulative distribution of nSPD in data for samples triggered by *L0MuonEW* (blue), *L0JetPh* (red) and *L0JetEl* (green). The dashed lines indicates the nSPD cut at 450 and 900.

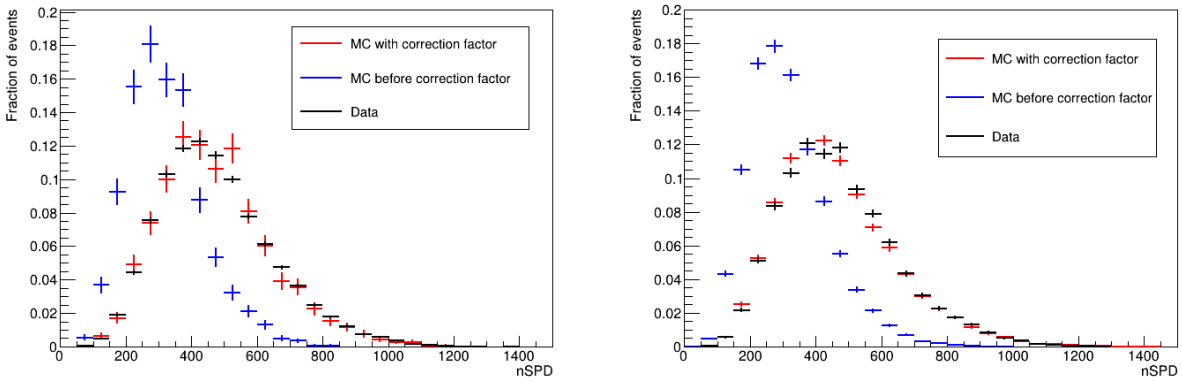

Figure 68: nSPD distributions for invariant mass range [45,83]GeV (left) and [83,130](right) of events that have passed the *L0JetEl* line. The nSPD distribution of Monte Carlo before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization is shown in red. The nSPD distribution of data is shown in black.

The shape of H and Z boson invariant mass distributions in the SR after the jet weighting due to the SV tag and the DNN have also been used.

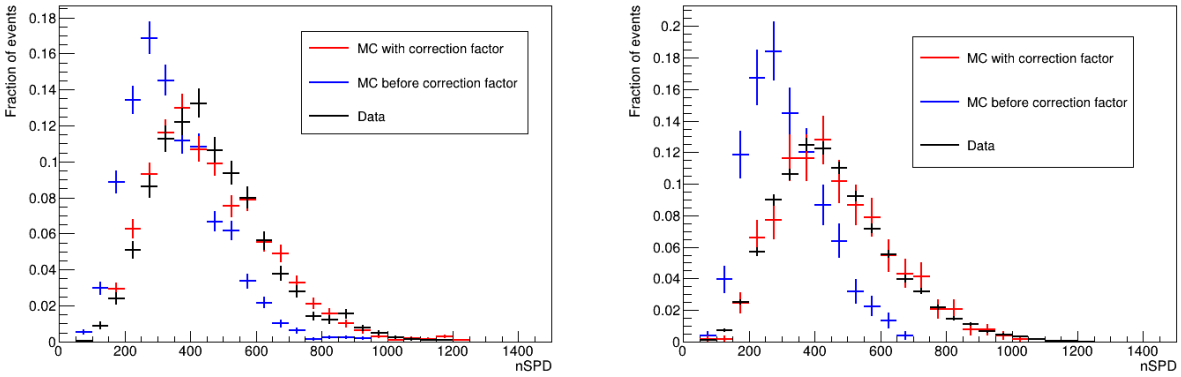

Figure 69: nSPD distributions for invariant mass range [130,250]GeV (left) and [45,83](right) of events that have passed the *L0JetEl* line and *L0JetPh* line respectively. The nSPD distribution of Monte Carlo before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization is shown in red. The nSPD distribution of data is shown in black.

Correction factors	$H \to b\bar{b}$	$H \to c\bar{c}$	$Z \to b\bar{b}$	$Z \to c\bar{c}$
SV tagging	1.15	0.81	1.07	0.96
SV tagging + Deep Neural Network	1.04	0.91	1.02	1.12
Global Event Cut	0.64	0.65	0.69	0.70
Total correction (SV+DNN+GEC)	0.67	0.59	0.70	0.78
Acceptance factor	1.04	1.04	-	-
Number of events before correction	346	0.87	153171	1787
Number of events after corrections	239	0.53	107920	1391

Table 29: Correction factors for selection efficiencies of $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ in the $H \to b\bar{b}$ SR: the SV tagging, the SV tagging and the DNN weights considered together, and the GEC. Total correction (SV+DNN+GEC) shows the total correction factor to the efficiency. The acceptance factor used to scale the number of Higgs events is also shown. The last two rows show the number of events before and after the corrections. The number of events after corrections are the number of H and Z events used in the following analysis.

5.6 Transfer Function

The Transfer Function (TF) is used to model the multi-jet QCD background in the SR starting from the dijet invariant mass distribution of the CR. It is evaluated by using the dijet Monte Carlo samples, where resonances are not included, as the ratio between the number of dijet events in the SR (n_{SR}) over the number of events in the CR (n_{CR}), per bin of dijet invariant mass:

$$TF = \frac{n_{SR}}{n_{CR}}(m_{jj}) \tag{5.7}$$

The TF is then fitted with a polynomial curve.

5.6.1 Transfer function for the $H \to b\bar{b}$ search

The $b\bar{b}$ and $c\bar{c}$ dijet simulation samples are used to calculate the TF as a function of the dijet invariant mass for the H$\to$ $b\bar{b}$ search. The truth-matching to real b-jets is not applied for this

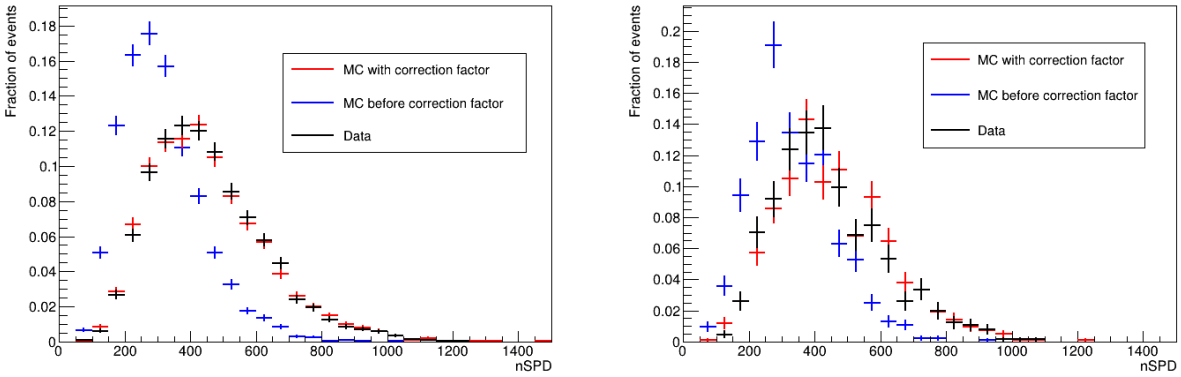

Figure 70: nSPD distributions for invariant mass range [83,130]GeV (left) and [130,250](right) of events that have passed the *L0JetPh* line. The nSPD distribution of Monte Carlo before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization is shown in red. The nSPD distribution of data is shown in black.

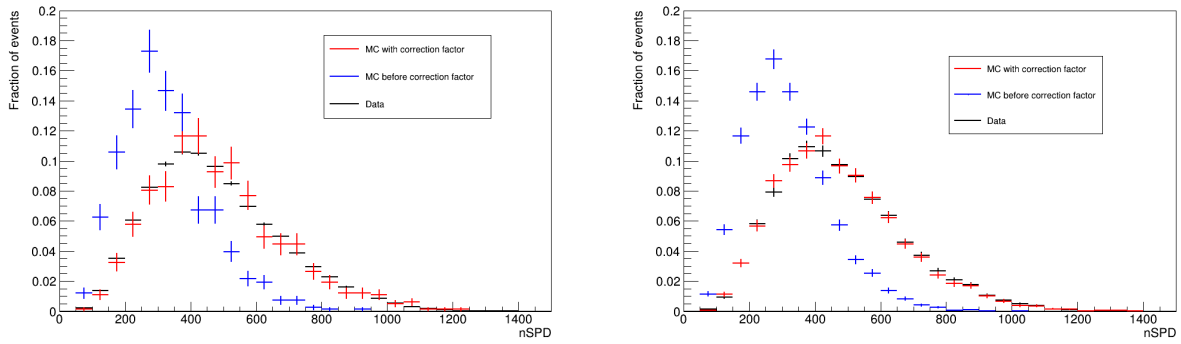

Figure 71: nSPD distributions for invariant mass range [45,83]GeV (left) and [83,130](right) of events that have passed the *L0MuonEW* line. The nSPD distribution of Monte Carlo before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization is shown in red. The nSPD distribution of data is shown in black.

purpose. n_{SR} is the number of events in the $b\bar{b}$+$c\bar{c}$ dijet MC sample that fulfill the SR requirements, while n_{CR} is the number of events in the $b\bar{b}$+$c\bar{c}$ dijet MC sample that are selected by applying the CR requirements. The transfer function $TF_{MC,b}(x)$ is parametrized by a polynomial of degree 4:

$$TF_{MC,b}(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3 + a_4 \cdot x^4 \tag{5.8}$$

where x is the invariant mass of the dijets candidate and a_0, a_1, a_2, a_3, a_4 are the polynomial coefficients. In Figure 74 the $TF_{MC,b}(x)$ profile is shown in the invariant mass range between 45 and 250 GeV, with the fitted function. The TF will be used in section 5.7 multiplied to the CR invariant mass distribution of Figure 59. As will be shown later, a correction function will be also introduced in order to take into account for further differences between data and MC in the determination of the TF.

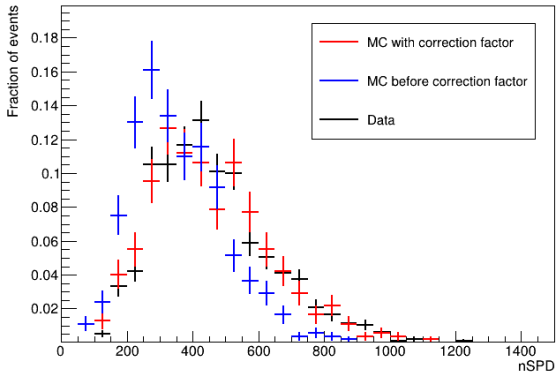

Figure 72: nSPD distributions for invariant mass range [130,250] of events that have passed the *L0MuonEW* line. The nSPD distribution of Monte Carlo before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization is shown in red. The nSPD distribution of data is shown in black.

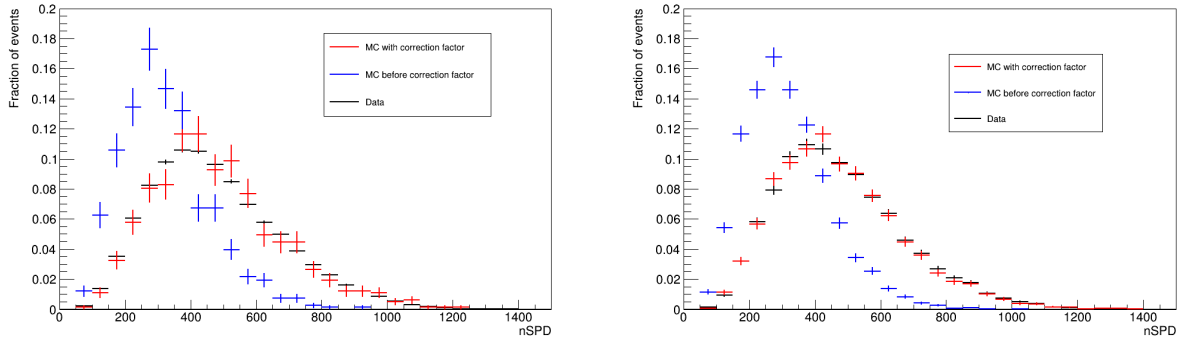

Figure 73: nSPD distributions for invariant mass range [45,83]GeV (left) and [83,130](right) of events that have passed the *L0MuonEW* line. The nSPD distribution of Monte Carlo before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization. The nSPD distribution of data is shown in black.

5.6.2 Transfer function for the $H \to c\bar{c}$ search

For modeling the multi-jet QCD background in the $H \to c\bar{c}$ search, $b\bar{b}$ and $c\bar{c}$ dijet simulation samples are used to calculate the Transfer Function as a function of the dijet invariant mass. Even in this case the truth-matching to real b and c-jets is not applied. n_{SR} is the number of events in the $b\bar{b}$+$c\bar{c}$ dijet MC sample that are selected by applying the SR requirements shown in Table 19, while n_{CR} is the number of events in the $b\bar{b}$+$c\bar{c}$ dijet MC sample that are selected by applying the CR requirements shown in Table 20. The transfer function $TF_{MC,c}(x)$ is parametrized with a polynomial of degree three:

$$TF_{MC,c}(x) = b_0 + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 \tag{5.9}$$

where x is the invariant mass of the dijets candidate and b_0, b_1, b_2, b_3 are the polynomial coefficients. The fitted TF is shown in figure 75.

The transfer function will be used in section 5.7 multiplied to the CR invariant mass distribution of Figure 60. Also in this case, a correction function will be also introduced in order to take into

Correction Factors	$H \to b\bar{b}$	$H \to c\bar{c}$	$Z \to b\bar{b}$	$Z \to c\bar{c}$
SV tagging	1.13	0.82	1.05	0.98
SV + Deep Neural Network	1.91	0.62	1.55	0.79
Global Event Cut	0.64	0.65	0.69	0.70
Total correction (SV+DNN+GEC)	1.21	0.40	1.08	0.54
Acceptance factor	1.04	1.04	-	-
Number of events before corrections	19	1.9	8896	4372
Number of events after corrections	23	0.8	9568	2379

Table 30: Correction factors for selection efficiencies of $Z \to b\bar{b}$, $Z \to c\bar{c}$, $H \to b\bar{b}$ and $H \to c\bar{c}$ in the $H \to c\bar{c}$ SR: the SV tagging, the SV tagging and the DNN weights considered together, and the GEC. Total correction (SV+DNN+GEC) shows the total correction factor to the efficiency. The acceptance factor used to scale the number of Higgs events is also shown. The last two rows show the number of events before and after the corrections. The number of events after corrections are the number of H and Z events used in the following analysis.

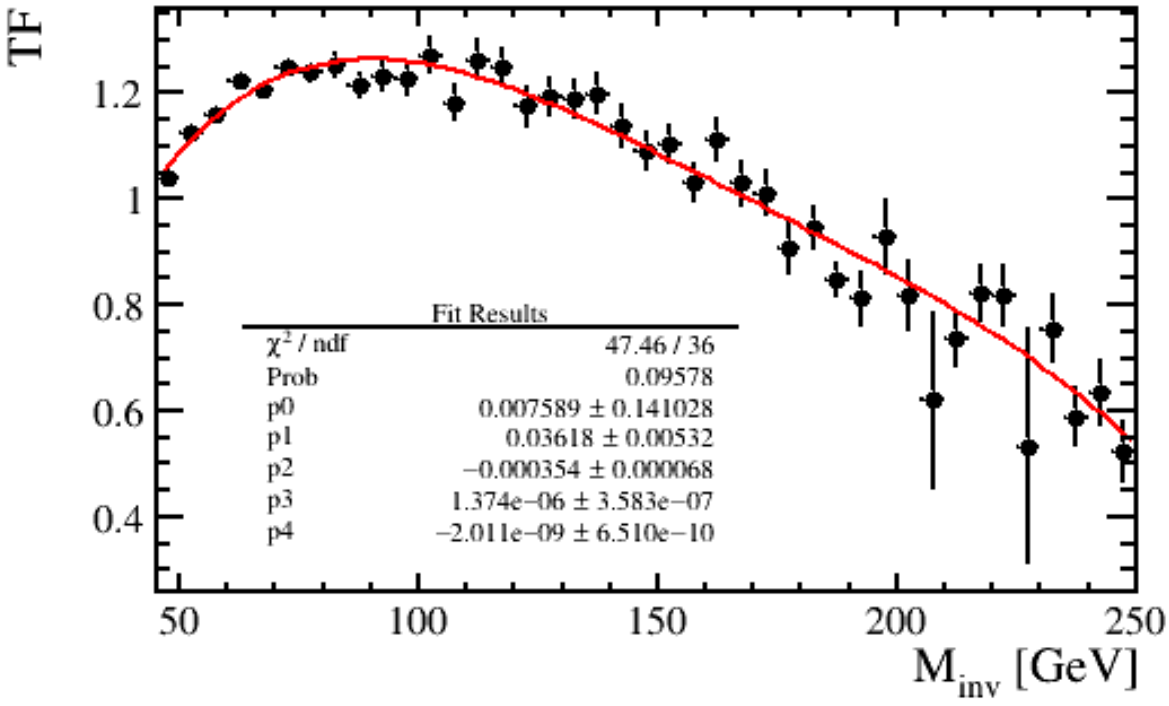

Figure 74: TF obtained for the $H \to b\bar{b}$. The polynomial fit with the fitted parameters are shown.

account for further differences between data and MC in the determination of the TF.

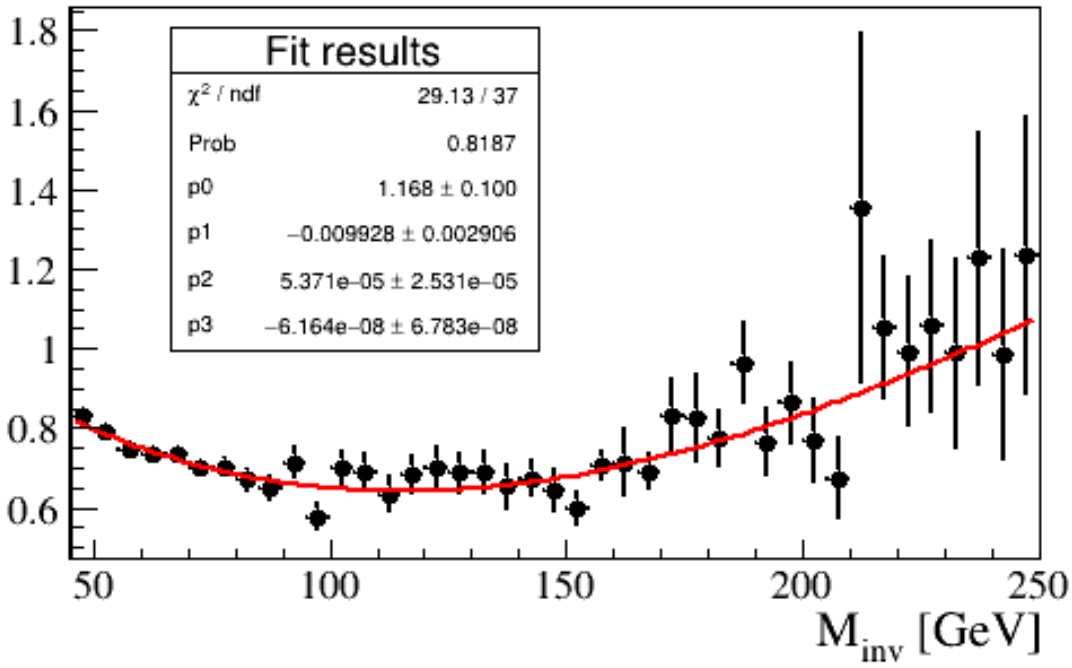

Figure 75: TF obtained for the $H \to c\bar{c}$ search. The polynomial fit with the fitted parameters are shown.

5.7 Fit to the signal region

A binned maximum likelihood approach is followed to fit data in the SR. The distributions in the fit model are therefore obtained as histograms: 41 bins of equal width in the mass range [45,250] GeV are considered. Events outside this mass range are discarded from the analysis. The model f used in the fit is the following:

$$f(x) = N_{QCD,SR} \cdot B_{CR}(x) \cdot TF_{MC}(x) \cdot BP(x) + N_{Hbb,SR} \cdot B_{Hbb}(x) + N_{Hcc,SR} \cdot B_{Hcc}(x) + N^{SM}_{Zbb,SR} \cdot B_{Zbb}(x) + N^{SM}_{Zcc,SR} \cdot B_{Zcc}(x),$$

where:

- x is the dijet invariant mass;
- B_{Hbb}, B_{Hcc}, B_{Zbb}, B_{Zcc} are the dijet invariant mass binned pdf of respectively the H$\to b\bar{b}$, H$\to c\bar{c}$, Z$\to b\bar{b}$, Z$\to c\bar{c}$ and have been obtained from the MC by applying the SR selection;
- $N^{SM}_{Zbb,SR}$, $N^{SM}_{Zcc,SR}$ are the number of expected $Z \to b\bar{b}$ and $Z \to c\bar{c}$ events in the SR within the invariant mass range [45,250] GeV. These normalization factors are fixed in the fit and systematic uncertainties will be assigned in section 6.8;
- $N_{Hbb,SR}$ and $N_{Hcc,SR}$ are the number of $H \to b\bar{b}$ and $H \to c\bar{c}$ events in the SR within the invariant mass range [45,250] GeV. In the $H \to b\bar{b}$ search $N_{Hcc,SR}$ is fixed to the SM expectation, while in the $H \to c\bar{c}$ search $N_{Hbb,SR}$ is fixed to the SM expectation. This choice leads to a systematic uncertainty that will be discussed in section 6.8;
- $N_{QCD,SR}$ is the number of multi-jet QCD events in the SR, it is a free parameter of the fit;
- $B_{CR}(x) \cdot TF_{MC}(x) \cdot BP(x)$ is the multi-jet QCD background pdf, composed by:

1. $B_{CR}(x)$ is the pdf histogram obtained in the CR data. Contaminations from signal and SM resonances will be discussed in section 6.6;
2. $TF_{MC}(x)$ is the transfer function obtained using the dijet simulation;
3. $BP(x)$ is the correction function mentioned in 5.6.2 and 5.6.1 that takes into account further differences between the TF in data and MC. The $BP(x)$ function has been modeled as a Bernstein polynomial and is

$$BP(x) = N \cdot \sum_{i=0}^{n} c_i B_{i,n}(x) \tag{5.10}$$

where, N is a normalization factor, c_i are the function coefficients, $B_{i,n}$ is the Bernstein basis polynomials of degree n:

$$B_{i,n}(x) = \binom{n}{i} x^i \cdot (1-x)^{n-i}. \tag{5.11}$$

A Fisher's Test is performed to determine the proper number of coefficients of the Bernstein polynomial that are needed to fit data (Appendix A).

5.7.1 Fit to the signal region for the $H \to b\bar{b}$ search

The Fisher's Test determined that a nine coefficients Bernstein polynomial is sufficient to fit data in the $H \to b\bar{b}$ search (Appendix A). The fit free parameters are $N_{QCD,SR}$, $N_{Hbb,SR}$ and the coefficients of the Bernstein polynomial c_1, c_2, c_5, c_6, c_7, c_8. The constant coefficient c_0 is fixed to 1 and the coefficients c_3 and c_4 have been fixed to their fitted value since they have high correlation with other coefficients and lead to fit instabilities in the Monte Carlo validation. c_4 coefficient is set to 0 since the fitted value is at limit (of the order of 10^{-10}). The results for $N_{Hbb,SR}$ and the fitted coefficients c_i are shown in the Table 31. The $\chi^2/ndof$ obtained from the fit is 9.5. The high $\chi^2/ndof$ may be due to the fact that for the moment the systematics related to signal/background modeling are not considered yet. It can be seen that the $N_{Hbb,SR}$ uncertainty (4227) is of about 18 times the expected SM signal events reported in Table 29. Figure 76 shows the fit residuals, defined as the difference between the number of data ($N_{data,i}$) in each bin i of the invariant mass and the number of fitted data ($N_{fit,i}$) in that bin:

$$\text{Residual}_i = N_{data,i} - N_{fit,i}. \tag{5.12}$$

Figure 77 shows the fit pulls, defined as the residuals over the Poissonian error on $N_{data,i}$, for each bin of invariant mass:

$$\text{Pull}_i = \frac{N_{data,i} - N_{fit,i}}{\sqrt{N_{data,i}}}. \tag{5.13}$$

From the pull plot it is evident that fluctuations larger than 2 sigmas are present, however systematic uncertainties should be considered for a proper test.

The same model used to fit data has been employed to generate 1000 Monte Carlo pseudo-experiments of the dijet invariant mass distribution, in order to validate the fit. The Bernstein polynomial initial parameters have been set to the fitted ones, while the expected SM signal yields are used to generate the pseudo-experiments. Figure 78 shows the result of the fit with one single pseudo-experiment, Figures 79 and 80 show the residual and pull of this fit.

The stability of the fit have also been checked by computing the pull P_{NHbb} on the number of fitted $H \to b\bar{b}$ with respect to the SM expectation for several pseudo-experiments, calculated as:

$$P_{NHbb} = \frac{N_{Hbb}^{fit} - N_{Hbb}^{SM}}{\sigma_{Hbb}^{fit}}. \tag{5.14}$$

Parameter	Fit Result
$\chi^2/ndof$	9.5
$N_{Hbb,SR}$	-2223 ± 4227
c_0	1.
c_1	0.848 ± 0.003
c_2	0.185 ± 0.003
c_3	0.90
c_4	0.
c_5	0.59 ± 0.02
c_6	0.19 ± 0.05
c_7	0.44 ± 0.04
c_8	0.36 ± 0.02

Table 31: Fit to data results for the $H \to b\bar{b}$ search.

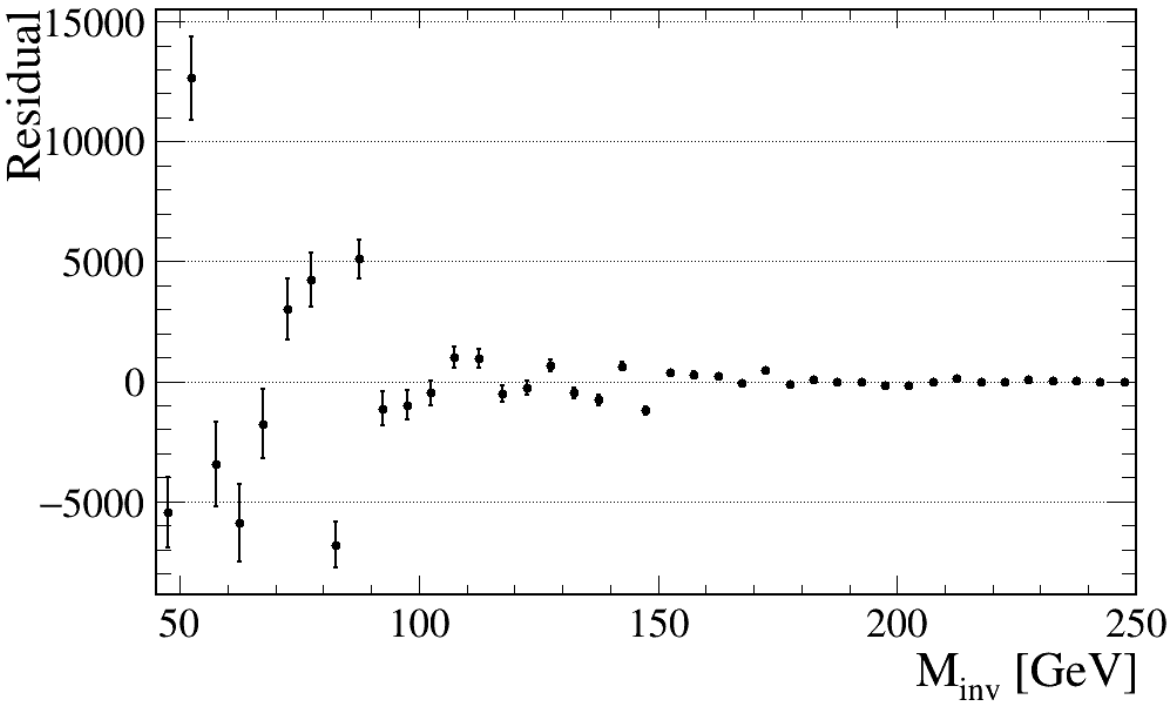

Figure 76: Residuals from the fit to data for the $H \to b\bar{b}$ search.

The distribution of P_{NHbb} is in Figure 81, where N^{fit}_{Hbb} and σ^{fit}_{Hbb} are the number of $H \to b\bar{b}$ fitted and its error, respectively. The standard deviation of the distribution is around one, and this means that the error is not over-estimated or under-estimated. Figure 82 shows the σ^{fit}_{Hbb} distribution. The error obtained from the fit to data is found compatible with the one obtained from pseudo-experiments. The pull of the other free parameters of the fit are shown in appendix B.

5.7.2 Fit to the signal region for the $H \to c\bar{c}$ search

The Fisher's Test determined that a ten coefficients Bernstein polynomial is sufficient to fit data, as reported in Appendix A. The fit free parameters are $N_{QCD,SR}$, $N_{Hcc,SR}$ and the coefficients of the Bernstein polynomial c_1, c_2, c_4, c_6, c_7, c_8, c_9. The constant coefficient c_0 is fixed to 1, while c_3, c_5 coefficients are set to zero, since they reach the lower limit in the fit and this leads to instabilities

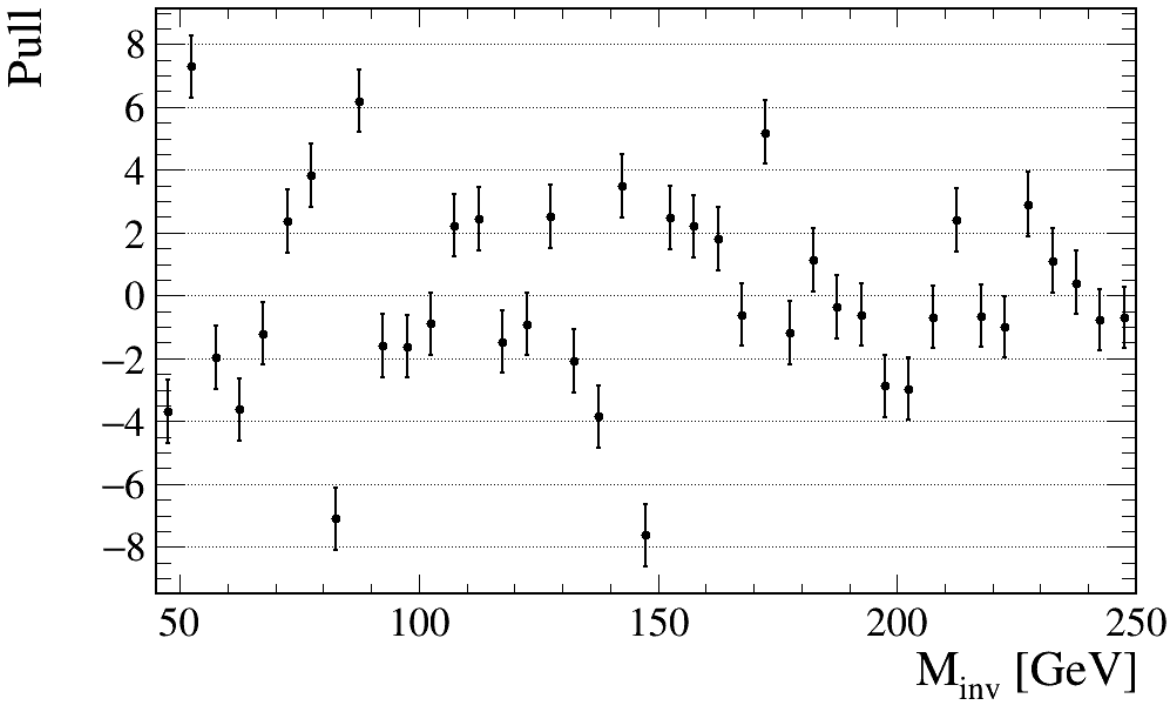

Figure 77: Pulls from the fit to data for the $H \to b\bar{b}$ search.

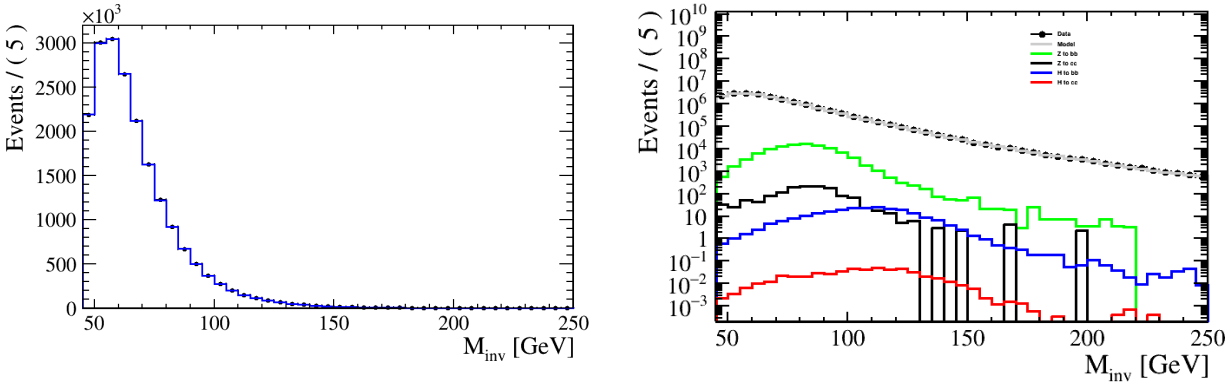

Figure 78: Left: Pseudo-experiment with the fitted model superimposed, for the $H \to b\bar{b}$ search. Right: The same fit result but in logarithmic scale, with Z boson and Higgs distributions superimposed.

in the Monte Carlo validation. The fit results are reported in Table 32. The uncertainty on $N_{Hcc,SR}$ is of about 3500 times the SM expectation reported in Table 30. The $\chi^2/ndof$ obtained from the fit is 1.6, therefore a good fit quality is achieved.

Fit residuals and pulls are shown in Figure 83 and 84.

The same model used to fit data has been used to generate 1000 pseudo-experiments, in order to validate the fit stability. In the pseudo-experiments generation the Bernstein polynomial parameters have been set to the ones from the data fit, and the number of signal $H \to c\bar{c}$ events is set to the SM expectation. The results of one pseudo-experiment are shown in Figure 85, where the fitted model with the pseudo-experiment superimposed is plotted. Figures 86 and 87 show the residual and pull of this fit. The pull of the other free parameters of the fit are shown in appendix C.

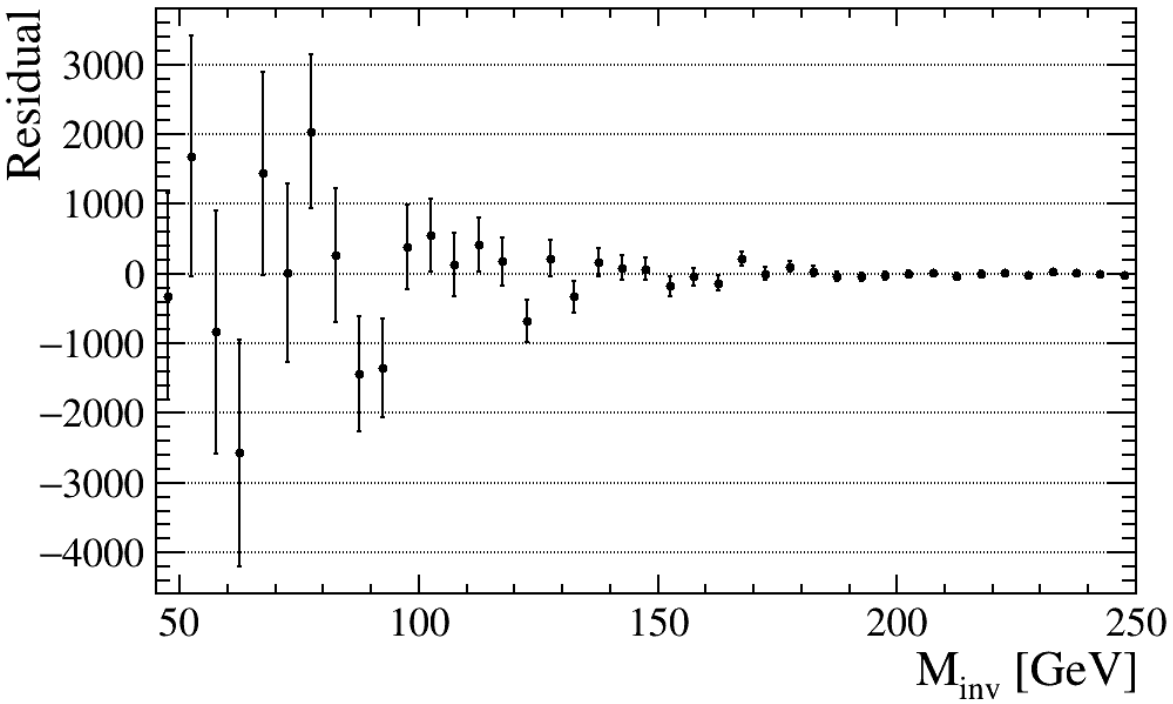

Figure 79: Residual of the model from the fit to the pseudo-experiment for the $H \to b\bar{b}$ search.

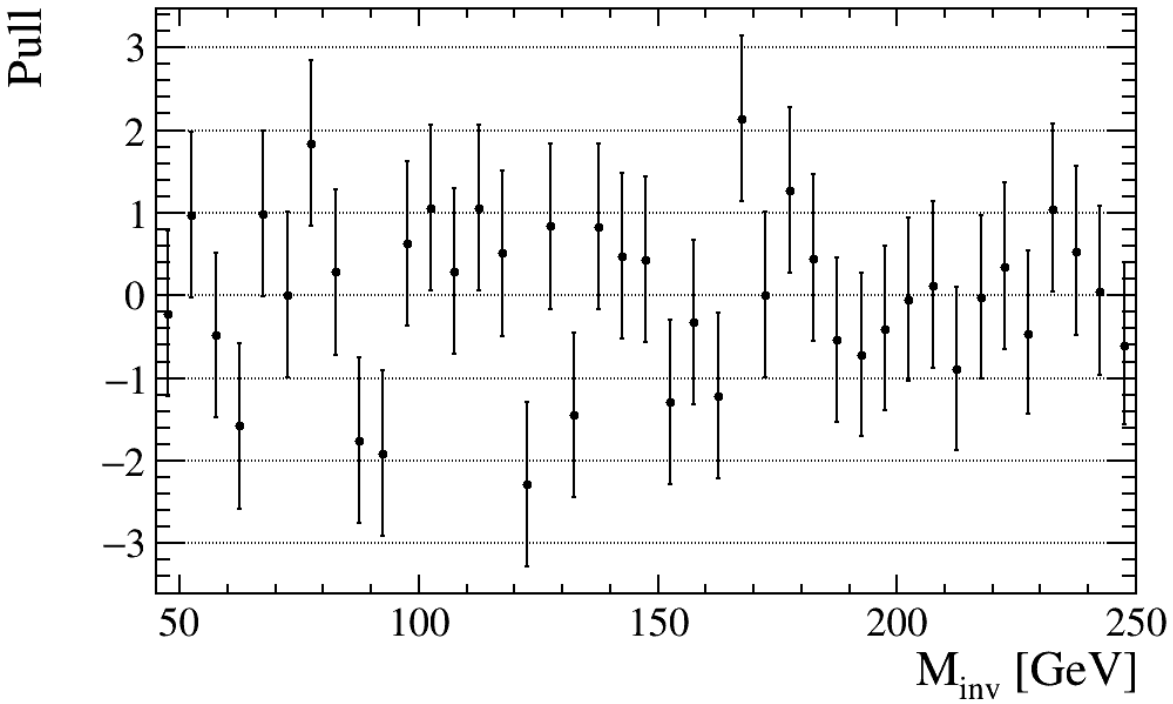

Figure 80: Pull of the model from the fit to the pseudo-experiment for the $H \to b\bar{b}$ search.

The stability of the $H \to c\bar{c}$ fit has also been checked in Figure 88, that shows the pull P_{NHcc}

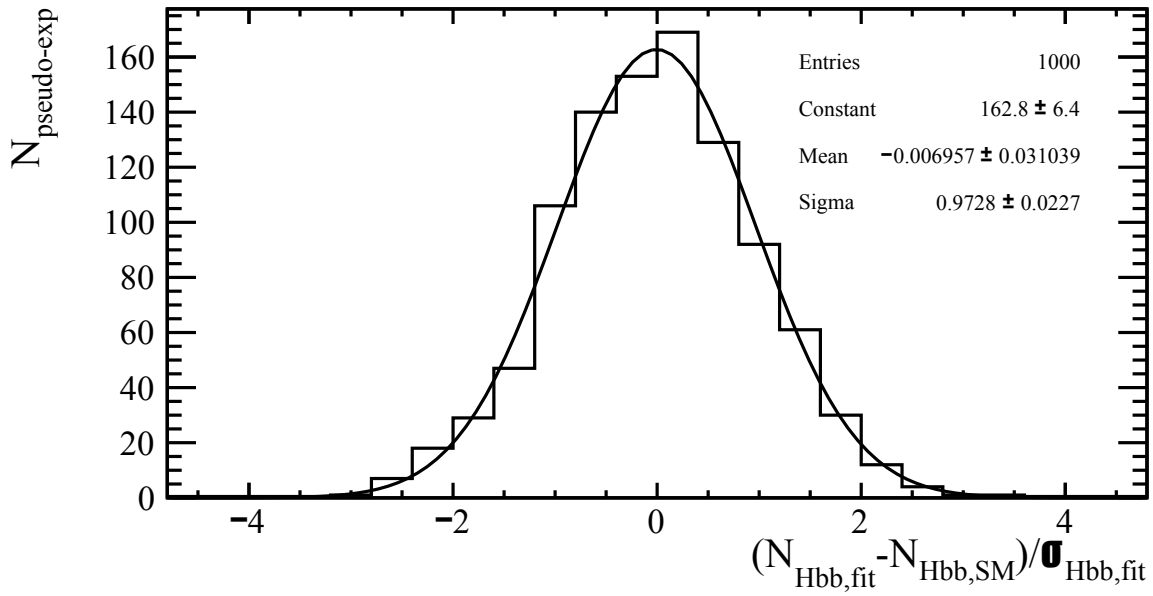

Figure 81: P_{NHbb} obtained from several pseudo-experiments for the $H \to b\bar{b}$ search.

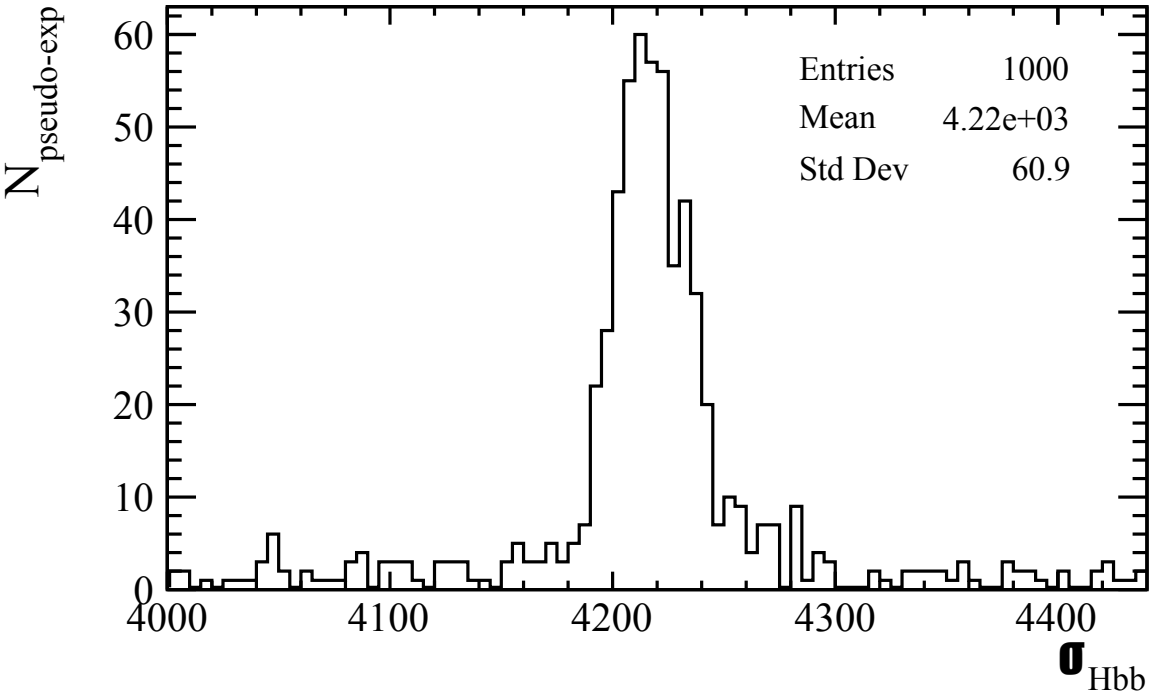

Figure 82: Error on the number of Higgs from the fit to the pseudo-experiment.

on the number of fitted $H \to c\bar{c}$ with respect to the SM expectation, calculated as:

$$P_{NHcc} = \frac{N_{Hcc}^{fit} - N_{Hcc}^{SM}}{\sigma_{Hcc}^{fit}} \tag{5.15}$$

where N_{Hcc}^{fit} and σ_{Hcc}^{fit} are the number of $H \to c\bar{c}$ fitted and its error, respectively. The standard deviation of the distribution is around one, and this means that the error is not over-estimated or

Parameter	Fit Result
$\chi^2/ndof$	1.6
$N_{Hcc,SR}$	-971 ± 2816
c_0	1.
c_1	0.65 ± 0.01
c_2	1.49 ± 0.02
c_3	0.
c_4	2.97 ± 0.07
c_5	0.
c_6	1.77 ± 0.18
c_7	1.08 ± 0.34
c_8	1.75 ± 0.28
c_9	1.09 ± 0.11

Table 32: Fit to data results for the $H \to c\bar{c}$ search.

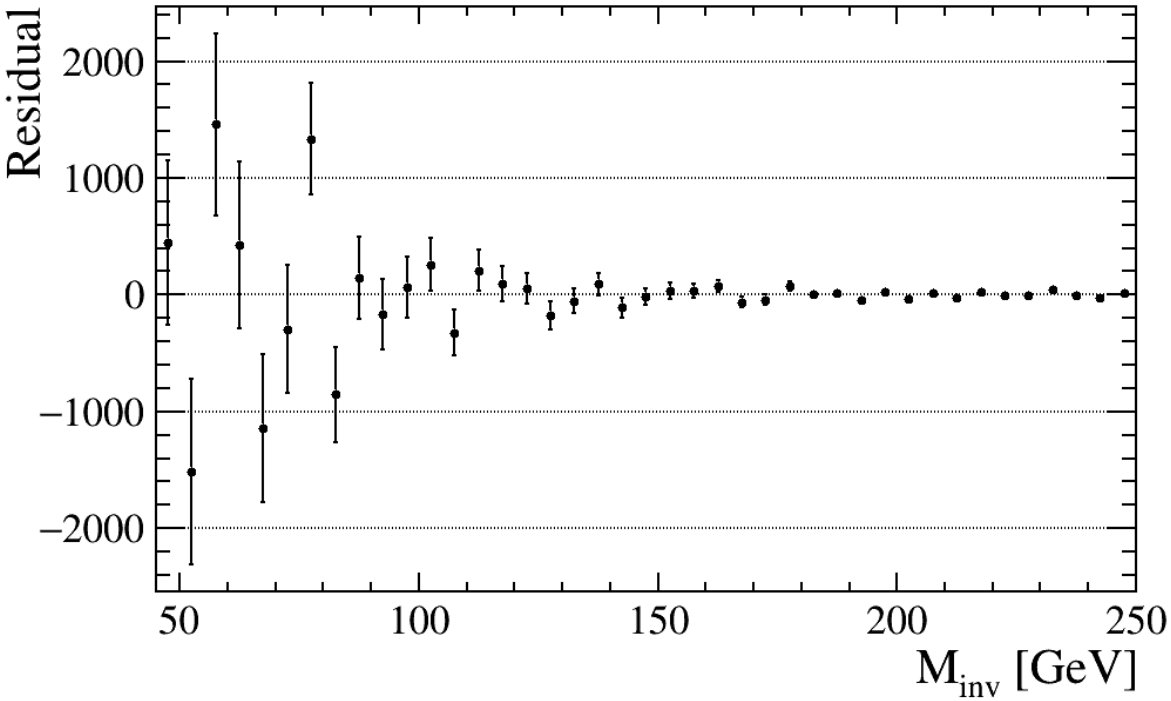

Figure 83: Residuals from the fit to data for the $H \to c\bar{c}$ search.

under-estimated.

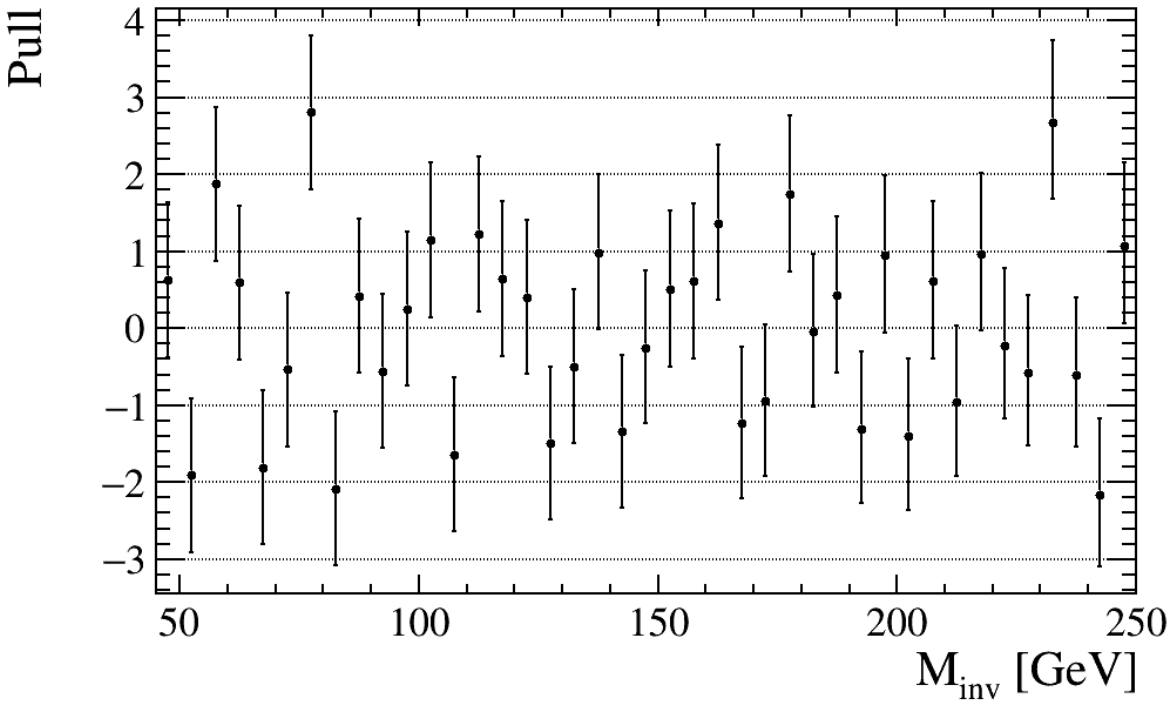

Figure 84: Pulls from the fit to data for the $H \to c\bar{c}$ search.

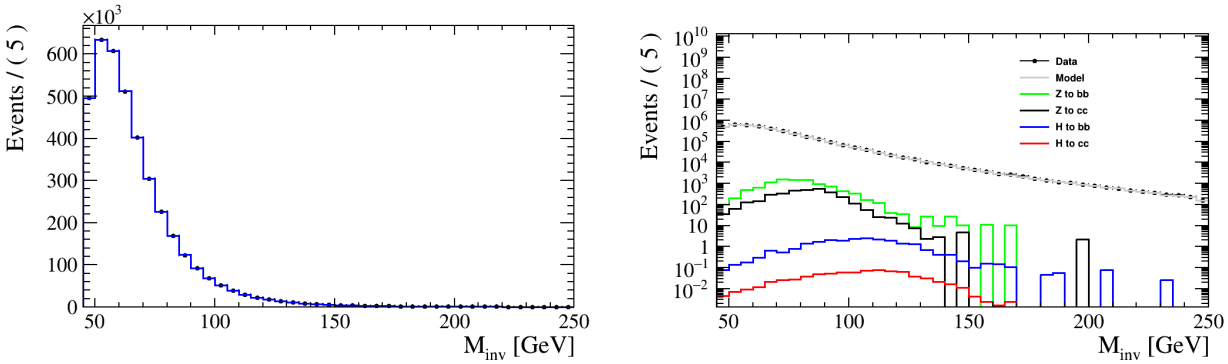

Figure 85: Left: Pseudo-experiment with the fitted model superimposed, for the $H \to c\bar{c}$ search. Right: The same fit result but in logarithmic scale, with Z boson and Higgs distributions superimposed.

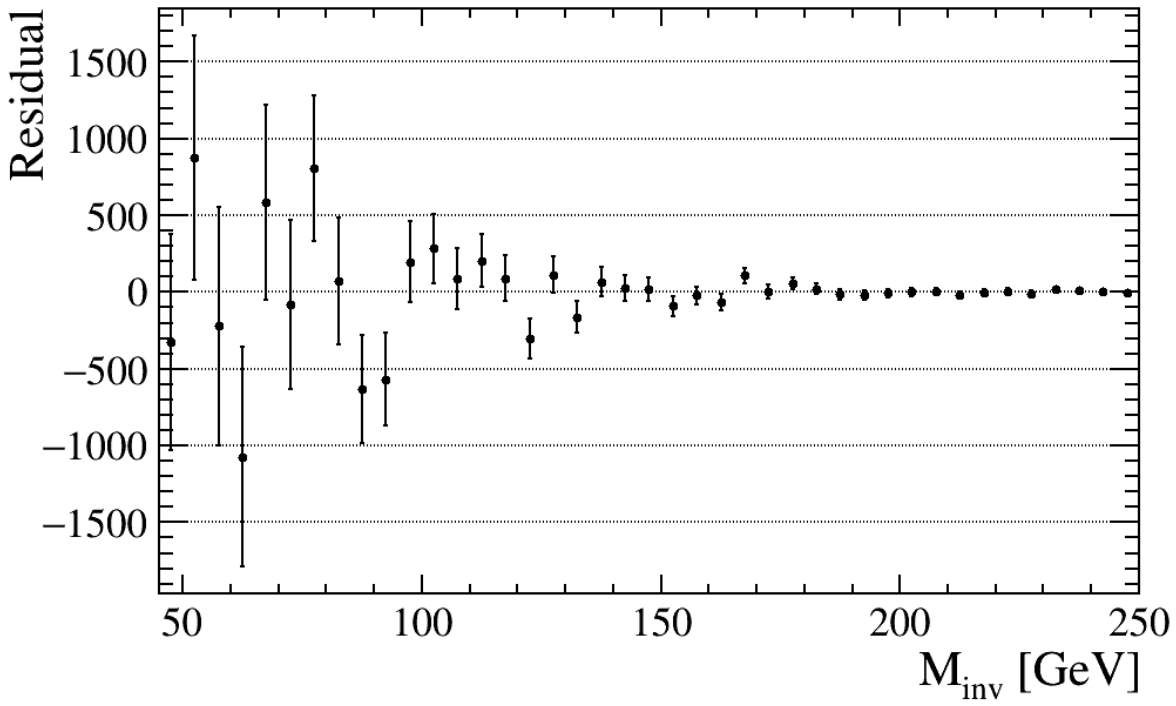

Figure 86: Residual of the model from the fit to the pseudo-experiment for the $H \to c\bar{c}$ search.

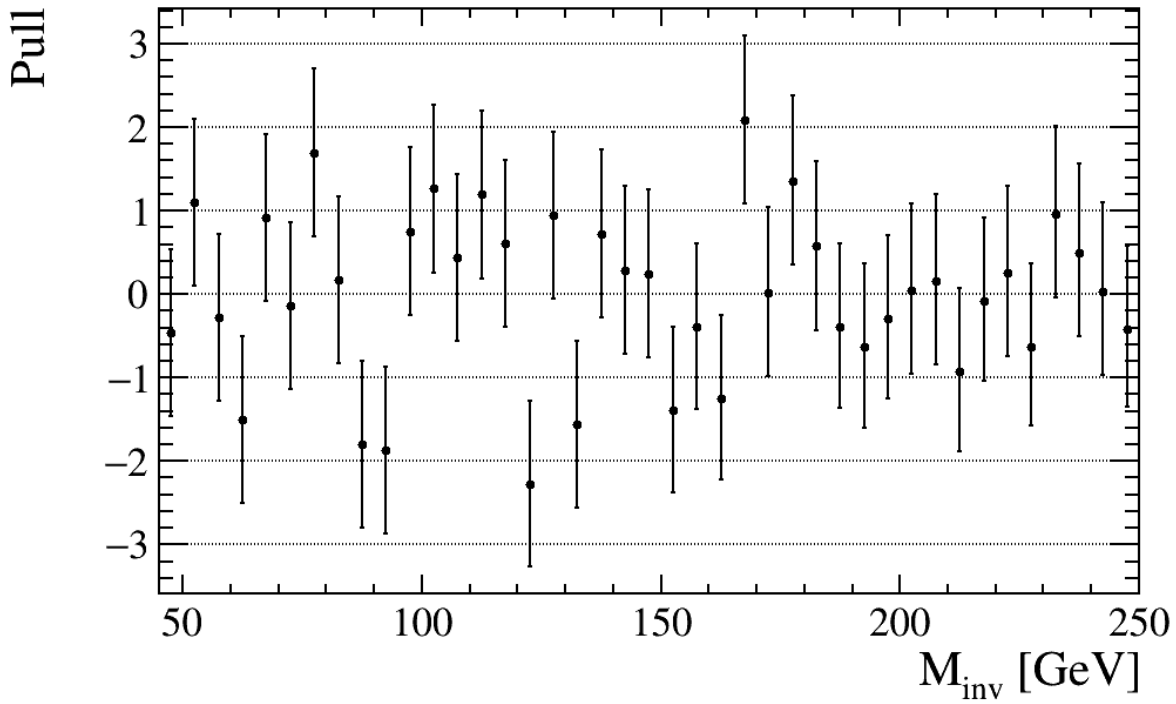

Figure 87: Pull of the model from the fit to the pseudo-experiment for the $H \to c\bar{c}$ search.

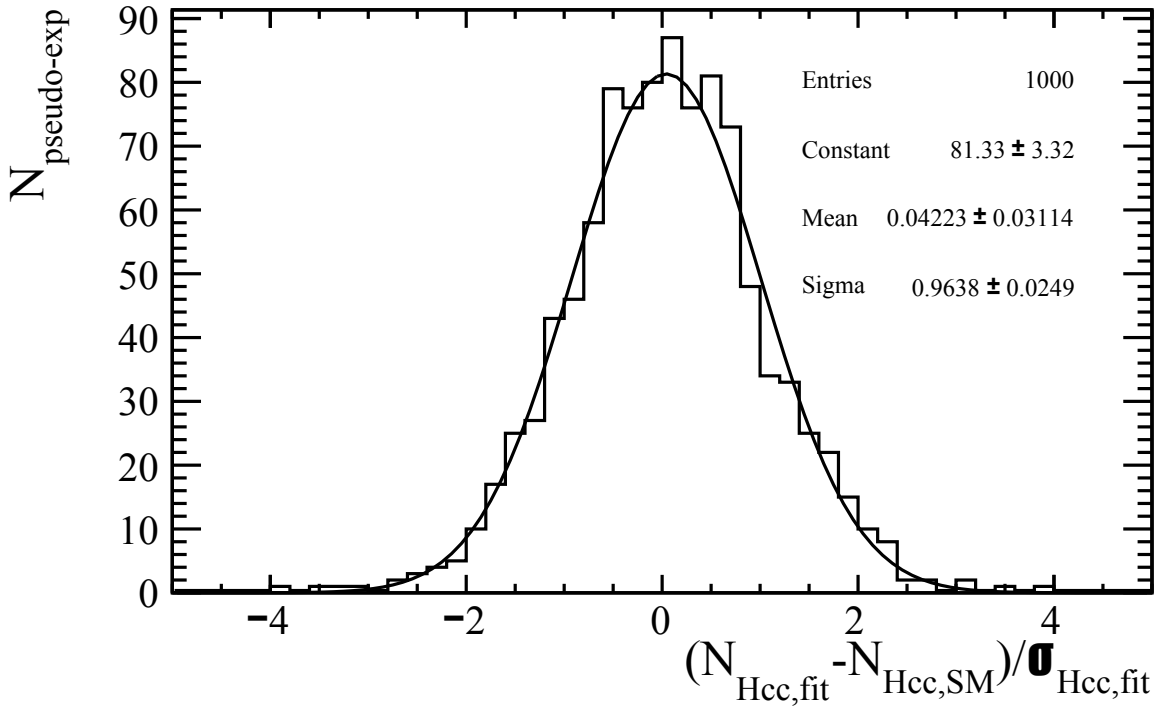

Figure 88: P_{NHcc} obtained from several pseudo-experiments for the $H \to c\bar{c}$ search.

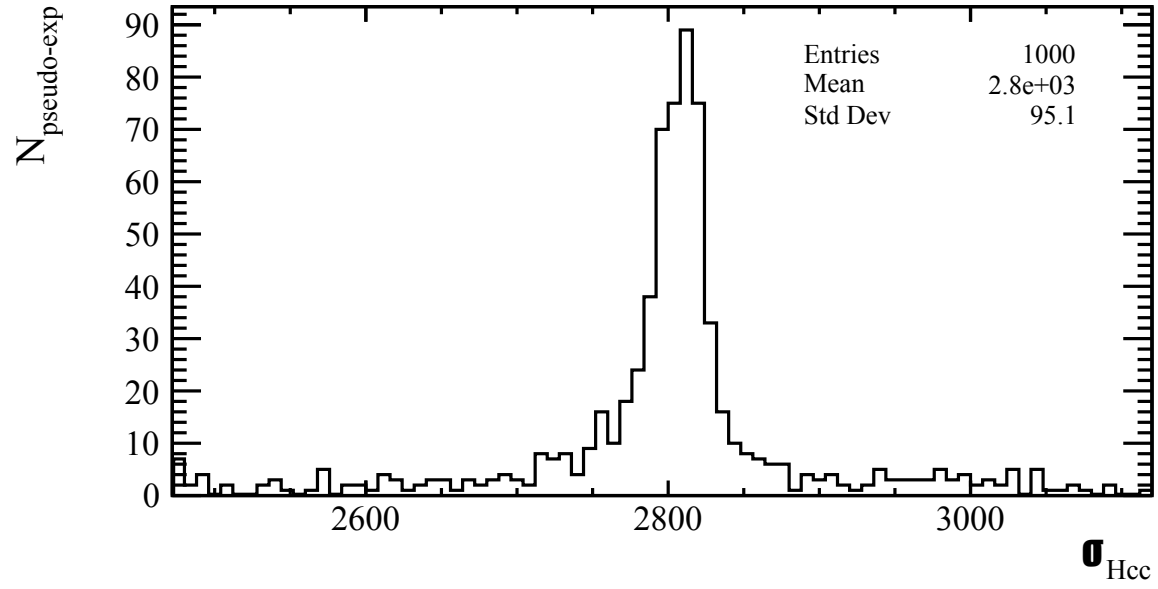

Figure 89: Error on the number of Higgs from the fit to the pseudo-experiment.

5.8 Upper limits calculation

Limits on the production cross section have been set since there is no evidence of signal in data. For this purpose, the CLs method is employed, described in section 5.8.1. In section 5.8.2 the expected upper limit using the background model calculated in the previous section, without systematic uncertainties, has been calculated.

5.8.1 CLs method

Hypothesis tests [98] provide a rule for accepting or rejecting hypotheses depending on the outcome of the measurement.

In the CLs method two hypothesis are considered:

- Null hypothesis: signal is absent;
- Alternate hypothesis: signal is present in the data.

A test-statistics is a random variable that depends on the parameters of the signal model (such as the cross section of the Higgs boson) and is built, as:

$$Q = \frac{\mathcal{L}(H_{s+b})}{\mathcal{L}(H_b)} \tag{5.16}$$

where is the $\mathcal{L}(H_{s+b})$ and $\mathcal{L}(H_b)$ is likelihood for the alternative and null hypothesis respectively.

The p-value of the signal+background hypothesis is given by the probability that the test statistic is less than or equal to the value observed in the experiment Q_{obs}:

$$CL_{s+b} = P_{s+b}(Q <= Q_{obs})$$

where

$$P_{s+b} = \int_{Q_{obs}}^{-\infty} \frac{dP_{s+b}}{dQ} dQ$$

and where $\frac{dP_{s+b}}{dQ}$ is the probability distribution function (p.d.f.) of the test-statistic for signal+background experiments. Small values of CL_{s+b} indicate poor compatibility with the signal+background hypothesis and favor the background hypothesis. Similarly, the p-value for the background hypothesis is given by the probability that the test-statistic is less or equal to the value observed in the experiment, Q_{obs}.

$$CL_b = P_b(Q <= Q_{obs})$$

where

$$P_b = \int_{Q_{obs}}^{-\infty} \frac{dP_b}{dQ} dQ$$

Small values of CL_{s+b} indicate poor compatibility with the signal+background hypothesis and favor the background hypothesis. In what is called the "CL_{s+b}" method, the signal model is regarded as excluded at a confidence level of 1 - α = 95 % if one finds $P_{s+b}(Q \leq Q_{obs}) < \alpha$, where, $\alpha = 0.05$. In this way, for example, a confidence interval at confidence level CL = 1 - α for the signal process cross section can be constructed from those values of the cross section that are not excluded, and the upper limit σ_{up} is the largest value of σ not excluded. In cases when the expected number of signal events is much less than that of background (s much smaller than b), as in this thesis, the CLs method is used. The CL_s [99] is defined as:

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{5.17}$$

When using CL_s a signal model is regarded as excluded at 1-α confidence level, if one finds

$$CL_s < \alpha \tag{5.18}$$

The upper limit on the Higgs cross section is found when the exclusion limit is greater than the specified confidence level for all values of the parameter above the confidence limit. CL_{s+b} and CL_b can be computed with different methods: in this thesis the *asymptotic calculator* [100] is used, which uses the asymptotic formula for the test-statistics distribution. Other methods forseen the usage of Monte Carlo toys (like the *frequentist calculator*).

5.8.2 Upper limits

Upper limits on the number of observed Higgs events have been calculated with the CLs method. The same model described for the fit in section 5.7 is used. The coefficients of the Bernstein Polynomial are fixed to the ones obtained from the fit to data. The upper limits results are normalized to the SM expectation as:

$$\frac{N_{up}}{N_{SM}} = \frac{\sigma_{up}}{\sigma_{SM}}, \tag{5.19}$$

where N_{up} is the upper limit on the Higgs yield, N_{SM} is the expected number of Higgs events, σ_{up} is the upper limit on the cross section and σ_{SM} is the SM cross-section. For this reason the limits can be expressed in terms of $\frac{\sigma_{up}}{\sigma_{SM}}$. Expected CLs in the background-only hypothesis and 95% expected upper limits are shown in Figures 90 and 91 for $H \to b\bar{b}$ and $H \to c\bar{c}$ respectively. The expected upper limit with 13 TeV collisions and 1.6 fb^{-1} integrated luminosity is 13 times the SM cross section for the $H \to b\bar{b}$, and 1913 times the SM cross section for the $H \to c\bar{c}$. Expected upper limits are summarized in Table 33.

Process	σ_{UP}/σ_{SM}	N_{up}	σ_{UP} [pb]
$H \to b\bar{b}$	13	3110	31
$H \to c\bar{c}$	1913	1453	222

Table 33: Expected upper limits on the $H \to b\bar{b}$ and $H \to c\bar{c}$ cross sections.

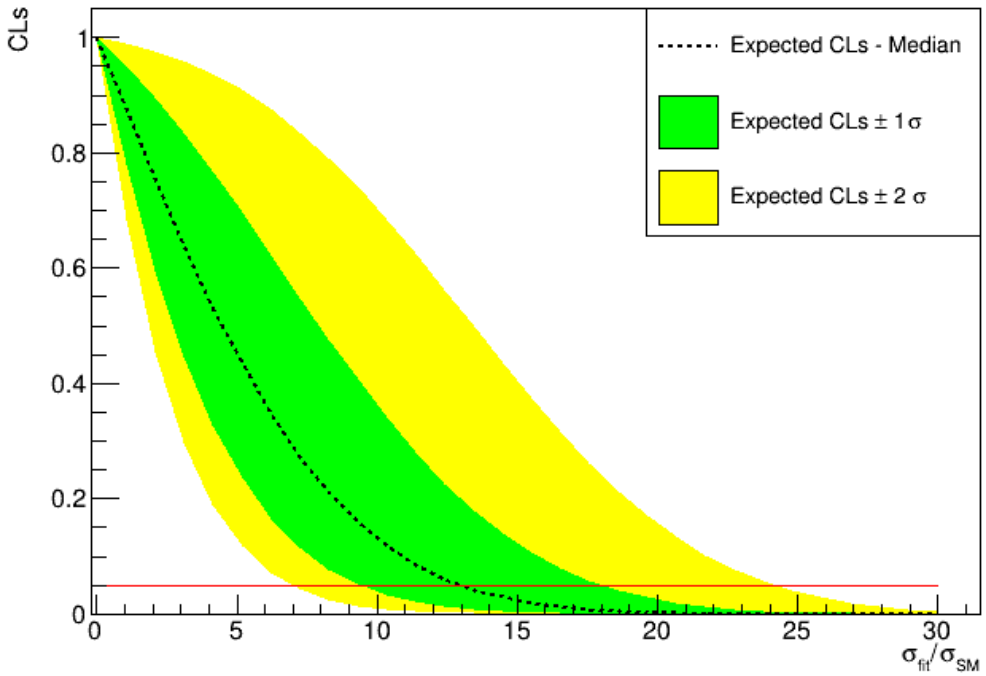

Figure 90: CLs values for several $H \to b\bar{b}$ cross-section hypotheses. The dotted black line represents the expected-limit in the background-only hypothesis. The green (yellow) bands define the $\pm$ 1 σ ($\pm$ 2 σ) uncertainties. The red line defines the 95 % CL limit.

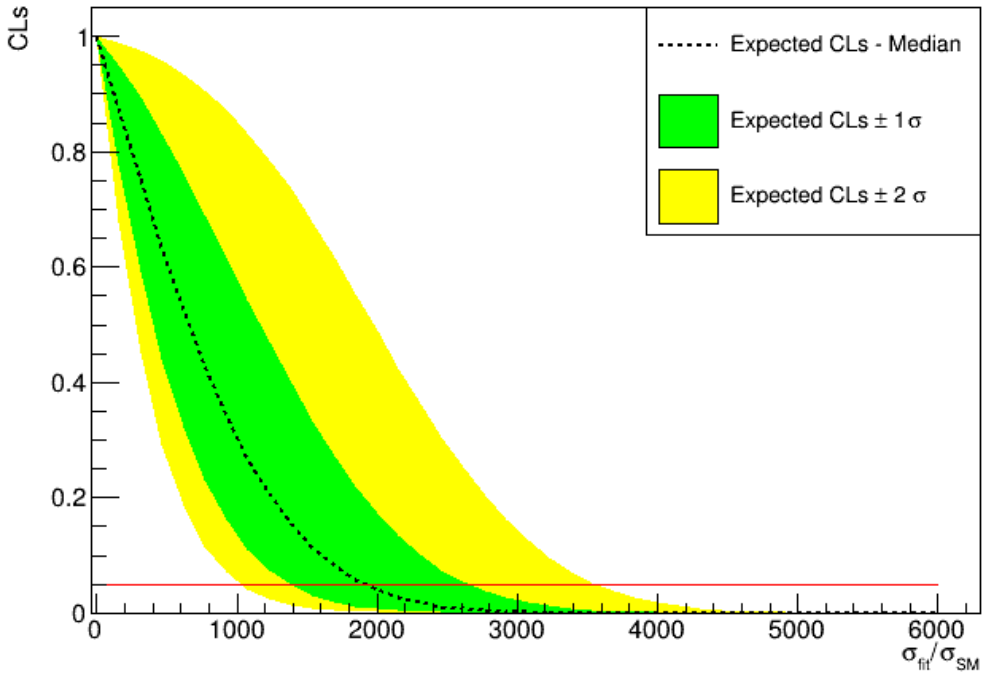

Figure 91: CLs values for several $H \to c\bar{c}$ cross-section hypotheses. The dotted black line represents the expected-limit in the background-only hypothesis. The green (yellow) bands define the $\pm$ 1 σ ($\pm$ 2 σ) uncertainties. The red line defines the 95 % CL limit.

6 Upper limits on the H $\to b\bar{b}$ and H $\to c\bar{c}$ search and future prospects

Systematic uncertainties may affect the analysis at different levels: they are related to data/MC differences and may affect the templates shapes (T_{shape}) used in the fits and/or the selection efficiency, and consequently the templates normalization (T_N). Other sources are related to the physical quantities used for the calculation of the Z and H Monte Carlo yields, like the cross section evaluation, that affect the templates normalization. In this section the procedure to estimate the systematic uncertainties on the upper limit calculation are studied. The list of the systematic sources discussed in this section is shown in Table 34.

Systematic source	T_N	T_{shape}
L0 nSPD (GEC)	X	-
L0 Trigger	X	-
HLT1	X	-
HLT2	X	-
Jet SV-tagging	X	X
Jet Identification	X	X
Jet Energy Resolution	X	X
Jet Energy Scale	X	X
Background model	-	X
SM cross sections	X	-
DNN	X	X
Acceptance	X	-
Luminosity	X	-

Table 34: List of systematics uncertainties studied.

6.1 L0

Differences in the efficiencies due to data/MC differences on the L0 requirements may have two sources:

- global event cut;
- other L0 requirements.

6.1.1 Global Event Cut efficiency

The uncertainty related to the efficiency correction factor shown in Table 28, calculated in section 5.5 are considered as systematic uncertainties on the GEC efficiency. Nuisance parameters relative to the Monte Carlo distributions normalization are added to the likelihood for the upper limit calculation, in order to account for the variation on the expected number of events due to the GEC uncertainty. The expected upper limit worsen by 0.4 % for the $H \to b\bar{b}$ and less 0.1 % for the $H \to c\bar{c}$.

6.1.2 L0 and HLT1 efficiencies

As for the GEC, the efficiencies of the L0 and HLT1 requirements can be different in data and MC, therefore data-driven weights should be applied. The Zjet sample is used for this purpose,

as done in the $b\bar{b}$ and $c\bar{c}$ cross sections measurement [28]. A tag-and-probe technique has been employed: the tag is the reconstructed Z boson in the dimuon final state, at least one muon is required to pass the $L0MuonEW$ line and the $Hlt1SingleMuonHighPT$ line, while the whole event is required to pass the $StrippingZ02MuMuLineDecision$.

The probe is the associated jet with $\Delta\phi$>2.3 with respect to the Z candidate. The same requirement on p_T and η as Table 13 are applied to the jet, including the SV tagging, except the requirement on the $L0$ and $HLT1$ chain. The efficiency of the $L0ChainTOS$ and $Hlt1ChainTOS$ requirements has been computed on the probe jet, and compared between data and simulation. When the $L0ChainTOS$ requirement is applied to the probe jet in simulation, the $nSPD$ is previously scaled to match the data distribution and the GEC cut corresponding to the passed L0 lines is applied (<450 or <900 or no cut). For this purpose, events passing the $L0MuonEW$ line are selected for both Monte Carlo and data. The Monte Carlo and data $nSPD$ distributions are compared, and the Monte Carlo $nSPD$ are scaled in order to match data, with the same χ^2 minimization procedure described in section 5.5.3. The scale factor is found to be 1.43. Figure 92 shows the nSPD distributions of events that have passed the $L0MuonEW$ line. The nSPD distribution of Monte Carlo Zjet before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization is shown in red. The nSPD distribution of data Zjet is shown in black. The efficiencies of L0 has been calculated as the ratio between events where

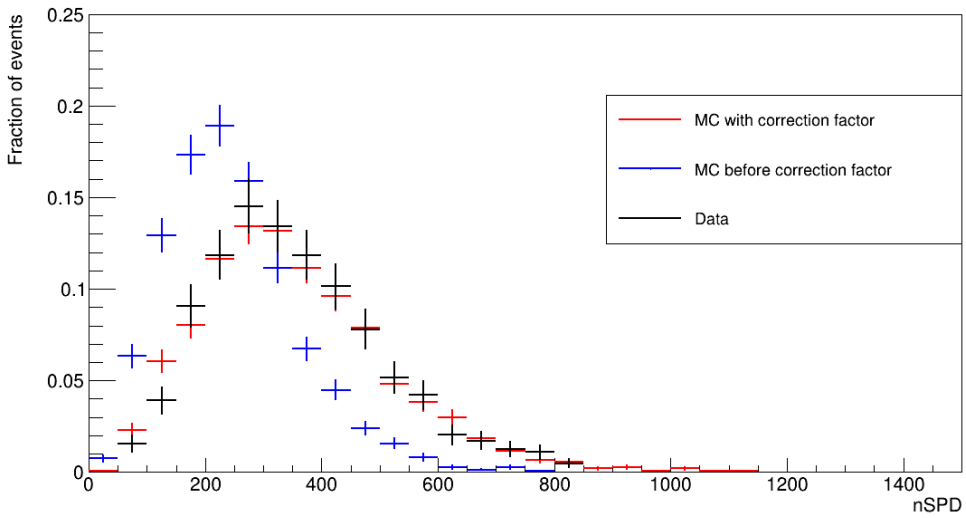

Figure 92: nSPD distributions of events that have passed the $L0MuonEW$ line. The nSPD distribution of Monte Carlo Zjet before the correction is shown in blue, the distribution after having applied the scale factor determined by the χ^2 minimization. The nSPD distribution of data Zjet is shown in black.

the tag and the probe jet pass the cuts described above and the $L0ChainTOS$, N_{L0}, over all events selected by the tag and probe requirements (N_{tag}):

$$\epsilon_{L0} = \frac{N_{L0}}{N_{tag}} \tag{6.1}$$

The efficiency of the $L0ChainTOS$ requirements on the Monte Carlo events has then been calculated after the nSPD correction and the GEC cut have been applied. Figure 93 shows the per-jet $L0ChainTOS$ efficiency, calculated in Zjet data (black) and simulation (red) events, as a function of the jet p_T.

The efficiencies are compatible, then no correction is needed to account for data/MC differences. The Monte Carlo statistical uncertainty is used as systematic uncertainty related to the $L0ChainTOS$.

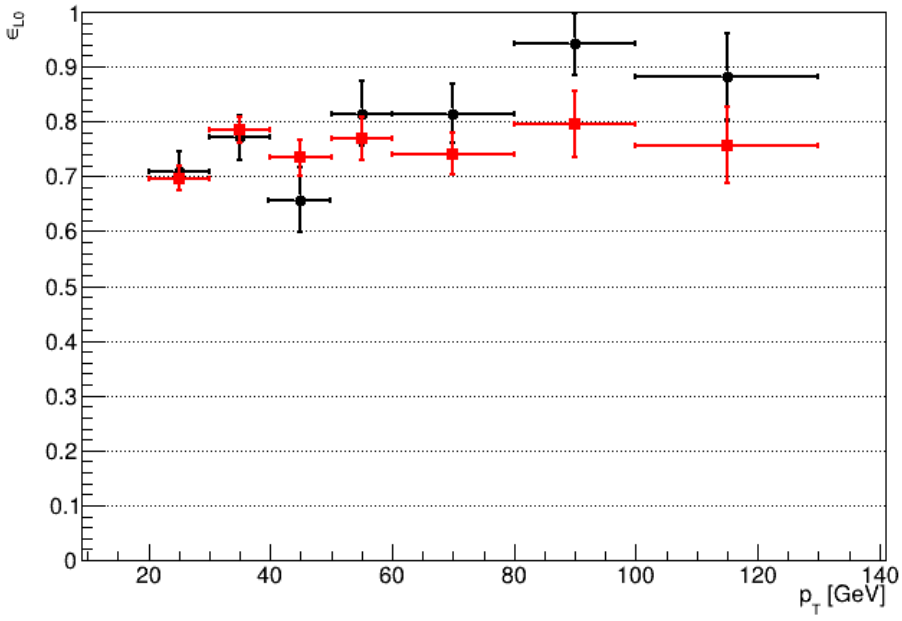

Figure 93: Per-jet $L0ChainTOS$ efficiency, calculated in Zjet data (black) and simulation (red) events using the tag-and-probe techique described in the text.

The efficiencies of HLT1 trigger has been calculated as the ratio between events tha pass both the $L0ChainTOS$ and the $Hlt1ChainTOS$ ($N_{HLT1,L0}$) over all the events that pass the $L0ChainTOS$ (N_{L0}):

$$\epsilon_{HLT1} = \frac{N_{HLT1,L0}}{N_{L0}} \tag{6.2}$$

Figure 94 shows the per-jet $Hlt1ChainTOS$ efficiency, calculated in Zjet data (black) and simulation (red) events.

The overall uncertainty on the HLT1+L0 efficiency, as a function of the jet p_T, has been determined by summing in quadrature the statistics uncertainty on the Monte Carlo for the L0 efficiency and the HLT1 efficiency. The upper (lower) variation on the expected number of events related to the HLT1+L0 efficiency have been determined then by weighting each jet of the event for the efficiency $+1\sigma$ (-1σ) variation. The weight associated to the event is, as for the SV tagging, the product of the two weights associated to the jets. The variation for the $H \to b\bar{b}$ and the $H \to c\bar{c}$ search are shown in Table 35.

A nuisance parameter relative to the Monte Carlo distributions normalization is added to the likelihood in the expected upper limit calculation, in order to account for the variation on the number of events shown in Table 35. The uncertainty is expected to increase the $H \to b\bar{b}$ and $H \to c\bar{c}$ expected upper limits by 3.2% and 3.6% respectively.

6.2 Jet SV-tagging efficiency

A nuisance parameter is added to the likelihood in the upper limit calculation, associated to the variation on the H and Z number of events and on the variation of their invariant mass shapes due

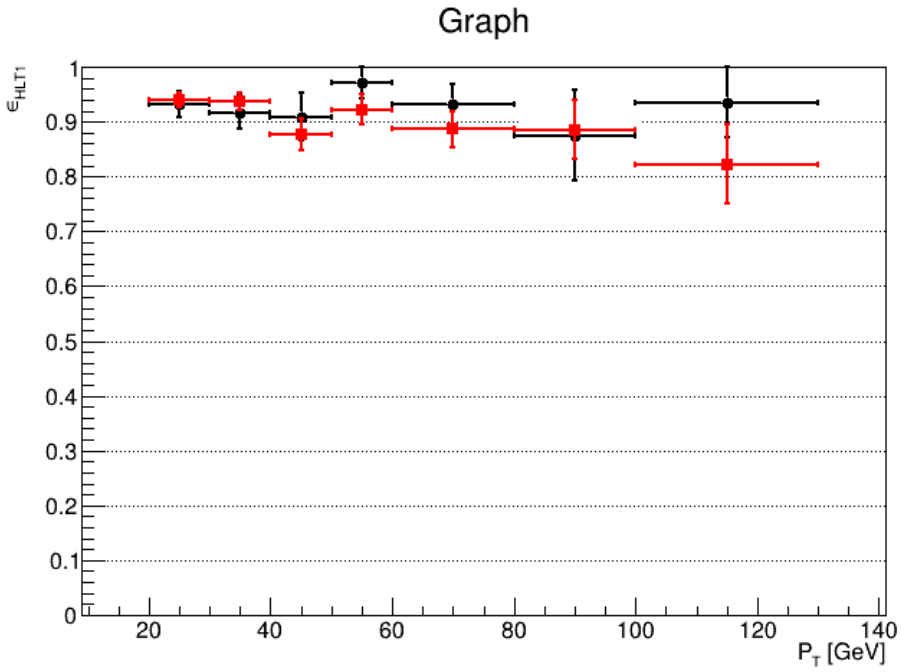

Figure 94: Per jet $Hlt1ChainTOS$ efficiency, calculated in Zjet data (black) and simulation (red) events using the tag-and-probe techique described in the text.

Process	$+1\sigma$ w.r.t. corrected N_{ev}	-1σ w.r.t. corrected N_{ev}
$H \to b\bar{b}$	1.096	0.91
$Z \to b\bar{b}$	1.07	0.93
$H \to c\bar{c}$	1.096	0.91
$Z \to c\bar{c}$	1.07	0.93

Table 35: L0+HLT1 upper and lower variation factors on the number of events for the $H \to b\bar{b}$ search and the $H \to c\bar{c}$ search.

to the SV tagging. This systematic uncertainty is expected to increase the $H \to b\bar{b}$ and $H \to c\bar{c}$ upper limits by 5.4% and 23% respectively, and it is the dominant systematic uncertainty for both searches.

6.3 HLT2 efficiency

The HLT2 line considered requires two reconstructed jets with two associated SVs, therefore the systematic uncertainty for this requirement is already considered in 6.2. A residual systematic uncertainty could arise from HLT2 from differences between the online and offline reconstruction algorithms. This difference has been already determined in [28], and it is below 0.3% for the reconstruction efficiency, therefore it is considered negligible.

6.4 Jet energy resolution and jet energy scale systematic uncertainties

The systematic uncertainty on the JES has been evaluated by calculating the variation on the number of H and Z events that one would have if a different JES scale factor was used to scale the jet $p_T(jet)$ before the event selection.

The number of events obtained by scaling $p_T(jet)$ for the two extremes of the uncertainty interval $[0.098, 1.056]$, found with the procedure described in section 5.2.7, is first calculated. The selections requirements in Table 13 have been applied to the events after the scaling procedure, and the correction for the SV tagging, the DNN, the GEC and the H cross section described in section 5.5 have been applied on the new samples. The variation with respect the corrected number of events calculated with the JES central factor is considered as systematic uncertainty. The variation on the number of events with respect the one calculated with the JES central factor for the $H \to b\bar{b}$ search are shown in Table 36, while for the $H \to c\bar{c}$ search are shown in Table 37. These numbers will be used to evaluate the systematic uncertainty on the Jet Energy Scale.

Process	$+1\sigma$ w.r.t. central N_{ev}	-1σ w.r.t. central N_{ev}
$H \to b\bar{b}$	1.005	0.993
$Z \to b\bar{b}$	1.03	0.97
$H \to c\bar{c}$	0.98	1.02
$Z \to c\bar{c}$	0.99	1.01

Table 36: Jet energy scale $\pm 1\sigma$ variation on the number of events with respect to the N_{ev} calculated by applying the centroid of the JES factor interval to the $p_T(jet)$, for the $H \to b\bar{b}$ search.

Process	$+1\sigma$ w.r.t. central N_{ev}	-1σ w.r.t. central N_{ev}
$H \to b\bar{b}$	1.03	0.96
$Z \to b\bar{b}$	1.05	0.95
$H \to c\bar{c}$	0.98	1.03
$Z \to c\bar{c}$	0.997	1.003

Table 37: Jet energy scale $\pm 1\sigma$ variation on the number of events with respect to the N_{ev} calculated by applying the centroid of the JES factor interval to the $p_T(jet)$, for the $H \to c\bar{c}$ search.

A nuisance parameter is added to the likelihood for the expected upper limit calculation, associated to the variation on the H and Z number of events and to the variation of their invariant mass shapes to account for the uncertainty related to the JES. The variation of the Monte Carlo shape for the $\pm 1\sigma$ configuration has been included in the fit but it is found to be negligible. The expected upper limit is found to be worsen by 1.2 % for the $H \to b\bar{b}$ and less of 0.3 % for the $H \to c\bar{c}$. The systematic uncertainty on the JER has been evaluated by calculating the variation on the number of H and Z events that one would have if a different JER smearing factor was used to smear the $p_T(jet1)$ and $p_T(jet2)$ before the event selection. The number of events obtained by smearing the $p_T(jet1)$ and $p_T(jet2)$ of Monte Carlo events for the two extremes of the uncertainty interval $[0., 0.138]$, found with the procedure described in section 5.2.7, has been first calculated. The selections requirements in Table 13 have been applied to the events after the smearing procedure, and the correction for the SV tagging, the DNN, the GEC and the H cross section described in section 5.5 have been applied on the new samples. The variation with respect the corrected number of events calculated with the JER central factor is considered as systematic uncertainty. The upper and lower variations on the number of events with respect the one calculated with the JER central factor are shown in Table 38, for the $H \to b\bar{b}$ search, and 39, for the $H \to c\bar{c}$ search. These numbers will be used as systematic uncertainties on the JER.

A nuisance parameter is added to the likelihood for the expected upper limit calculation, associated to the variation on the H and Z number of events and to the variation of their invariant mass shapes to account for the uncertainty related to the JER. Even in this case the impact of the variation of the Monte Carlo shape for the $\pm 1\sigma$ configuration is found to be negligible. The expected upper limit is found to be worsen by 0.03 % for the $H \to b\bar{b}$ and less 0.2 % for the $H \to c\bar{c}$.

Process	$+1\sigma$ w.r.t. central N_{ev}	-1σ w.r.t. central N_{ev}
$H \to b\bar{b}$	0.988	1.004
$Z \to b\bar{b}$	0.974	1.007
$H \to c\bar{c}$	1.009	1.007
$Z \to c\bar{c}$	0.994	1.009

Table 38: Jet energy resolution $\pm 1\sigma$ variation on the number of events with respect to the N_{ev} calculated by applying the centroid of the JER factor interval to smear the $p_T(jet1)$ and $p_T(jet2)$, for the $H \to b\bar{b}$ search.

Process	$+1\sigma$ w.r.t. central N_{ev}	-1σ w.r.t. central N_{ev}
$H \to b\bar{b}$	0.983	1.003
$Z \to b\bar{b}$	0.976	1.006
$H \to c\bar{c}$	1.005	1.008
$Z \to c\bar{c}$	0.983	1.004

Table 39: Jet energy resolution $\pm 1\sigma$ variation on the number of events with respect to the N_{ev} calculated by applying the centroid of the JER factor interval to smear the $p_T(jet1)$ and $p_T(jet2)$, for the $H \to c\bar{c}$ search.

6.5 Jet Identification

The dominant contribution to the jet reconstruction and identification efficiency comes from the jet identification requirements. As explained in section 4.8, jet identification is based on cuts applied to the number of tracks inside the jet pointing to the PV (N_{trk}), the maximum fraction of transverse momentum carried by a single particle inside the jet (mtf) and the maximum transverse momentum carried by a track inside the jet (cpf). Figure 95 shows the distributions of data and simulated Z+ jet events for N_{trk}, mtf and cpf. Zjets events have been selected by requiring the same requirements listed in section 5.2.6 requiring also the jet to be tagged. A systematic uncertainty may arise due to the mis-modeling of these observables in MC simulations. The systematic uncertainty associated with the jet identification requirements is determined as explained in [88] by tightening these requirements and comparing the fraction of events rejected in data and simulation:

- the cut on N_{trk} is tightened to $N_{trk} \geq 3$ (default: $N_{trk} \geq 2$);
- the cut on mtf is tightened to mtf$<$ 0.7 (default: mtf$<$ 0.8);
- the cut on cpf is tightend to cpf$>$ 0.2 (default: cpf$>$ 0.1).

The level of agreement l between data and Monte Carlo is then evaluated as:

$$l = \frac{|f_{data} - f_{sim}|}{f_{sim}} \tag{6.3}$$

where f_{data} and f_{sim} is the fraction of events rejected by tightening the jet identification requirements in data and simulation respectively. It is found to be:

- 0.013 $\pm$ 0.006 for N_{trk}
- 0.18 $\pm$ 0.03 for mtf
- 0.05 $\pm$ 0.02 for cpf

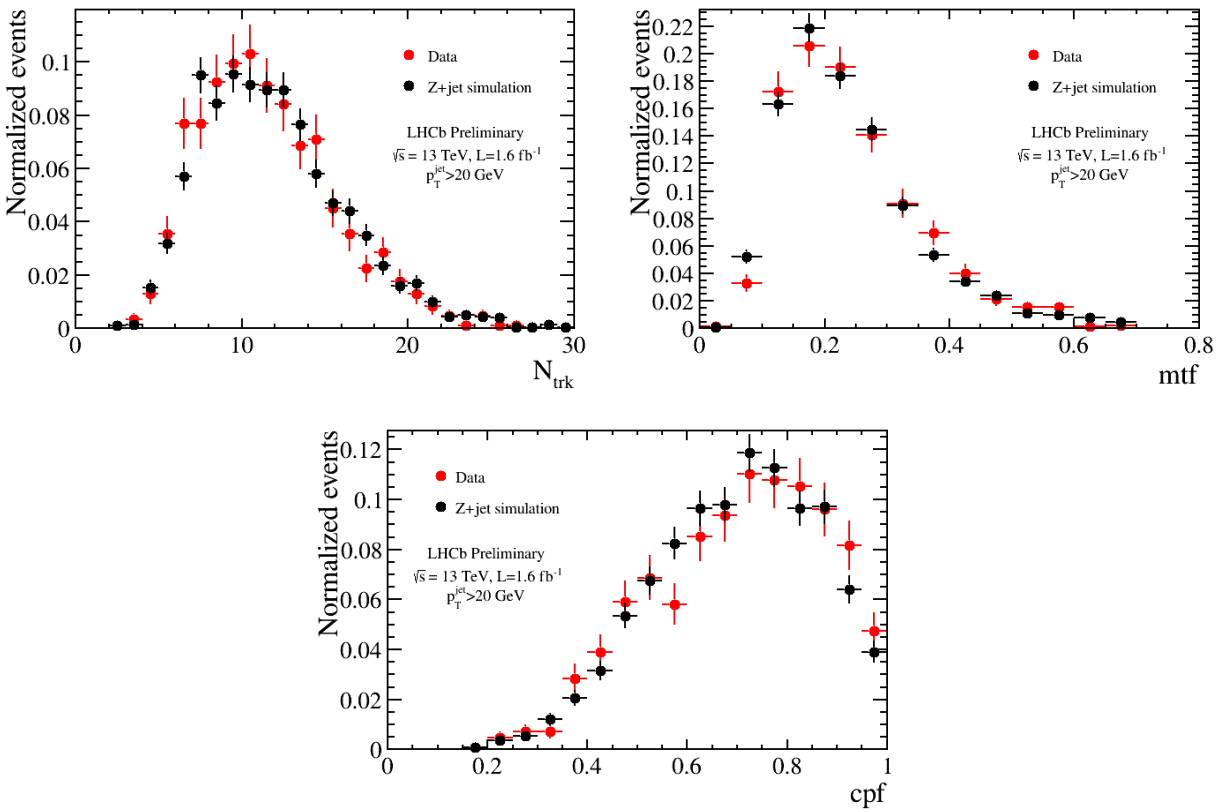

Figure 95: Distribution of N_{trk} (upper left plot), mtf (upper right plot) and cpf(lower plot) for Z +jet in data and simulations

As a conservative approach the level of agreement is set by the variable that has the worse agreement, mtf. The systematic uncertainty on the jet reconstruction efficiency is then computed as:

$$\delta_{ID} = (1 - \epsilon_{jet}) \cdot l \tag{6.4}$$

where ϵ_{jet} is the total jet reconstruction and identification efficiency, that is found to be around 92 % [101]. The δ_{ID} is found to be around 1.4 %. A nuisance parameter relative to the Monte Carlo distributions normalization is added to the likelihood in the expected upper limit calculation, in order to account for the variation on the number of events. The expected upper limit is found to be worsen by 0.7 % for the $H \to b\bar{b}$ and less 0.1 % for the $H \to c\bar{c}$ by this systematic uncertainty.

6.6 Multi-jet QCD background modelling

The following systematic sources related to the multi-jet QCD background modeling, that is the core of the $H \to b\bar{b}$ and $H \to c\bar{c}$ analyses, are considered:

- Transfer Function correction: the uncertainty has been evaluated by varying the degree of the Bernstein Polynomial by 1 grade;
- CR contamination: the CR can be contaminated with the signal and SM resonances, since it is basically impossible to completely remove them. In order to determine the associated systematic uncertainty, Z and H events are subtracted from the CR. The limit computation is repeated with this configuration, and the difference with respect to the nominal one is taken as uncertainty;

Process	$H \to b\bar{b}$ search	$H \to c\bar{c}$ search
$H \to b\bar{b}$	2.4	29
$H \to c\bar{c}$	0.06	0.09
$Z \to b\bar{b}$	728	35284
$Z \to c\bar{c}$	149	41
$W \to qq'$	3050	6100

Table 40: The number of Z, H and W expected in the $H \to b\bar{b}$ and $H \to c\bar{c}$ CR.

- CR statistics: this systematic uncertainty has been evaluated by re-sampling the CR events with a bootstrap technique and repeating the limit computation.

6.6.1 Transfer Function correction

The systematic uncertainty related to the transfer function has been valuated by re-calculating the upper limit on the $H \to b\bar{b}$ and $H \to c\bar{c}$ cross sections by varying the degree of the Bernstein Polynomial by 1 grade. For the $H \to b\bar{b}$ search, a ten degree Bernstein Polynomial has been considered. The variation of the upper limits calculated by using the nine and ten degree polynomials is found to be $< 1.0\%$.

For the $H \to c\bar{c}$ search, a 11 degree Bernstein Polynomial has been considered. The variation of the upper limits calculated by using the ten and eleven degree polynomials is found to be $< 0.05\%$.

6.6.2 CR contamination

This uncertainty is determined by subtracting the expected $Z \to b\bar{b}$, $Z \to c\bar{c}$ and $W \to qq'$ SM contributions from the CR. New templates for the multi-jet QCD background are obtained, and the fit and limit computation is repeated. The difference with respect to the nominal limit is considered as uncertainty. In a similar way the Higgs contamination is subtracted from the CR, by conservatively assuming a cross section that is 100 times the upper limit. The number of Z, H expected in the CR, corrected by the factors shown in Table 29 and 30 are shown in Table 40: The number of W events in the two CR has also been calculated. For the number of events in the $H \to b\bar{b}$ search CR, the prescale is taken into account. Variation on the expected upper limit between 0.3 and 1 % is found for the $H \to b\bar{b}$ search, 0.4 and 1.2 % for the $H \to c\bar{c}$ search.

6.6.3 CR statistics

The systematic uncertainty due to the limited statistics in the Control Region have been evaluated by randomly re-sampling the Control Region events using a bootstrap technique. For each re-sampling a different multi-jet QCD template is obtained and the upper limit computation is repeated. The upper limits on the $H \to b\bar{b}$ cross section distribution obtained by re-sampling the Control Region can be seen in Figure 96. Most of the fits (94%) converge and have an accurately computed error matrix. The mean of the distribution coincides with the upper limit found in section 5.8, while the width of the distribution is of about 0.4% of the mean value. For the $H \to c\bar{c}$ the width of the distribution is of about 0.4% of the mean value.

6.7 DNN

The systematic uncertainty related to the DNN selection efficiency, have been calculated in 5.5.2, and have been added as nuisance parameters in the likelihood for the upper limit calculation taking into account both the variation on the normalization and the variation on the invariant mass shape.

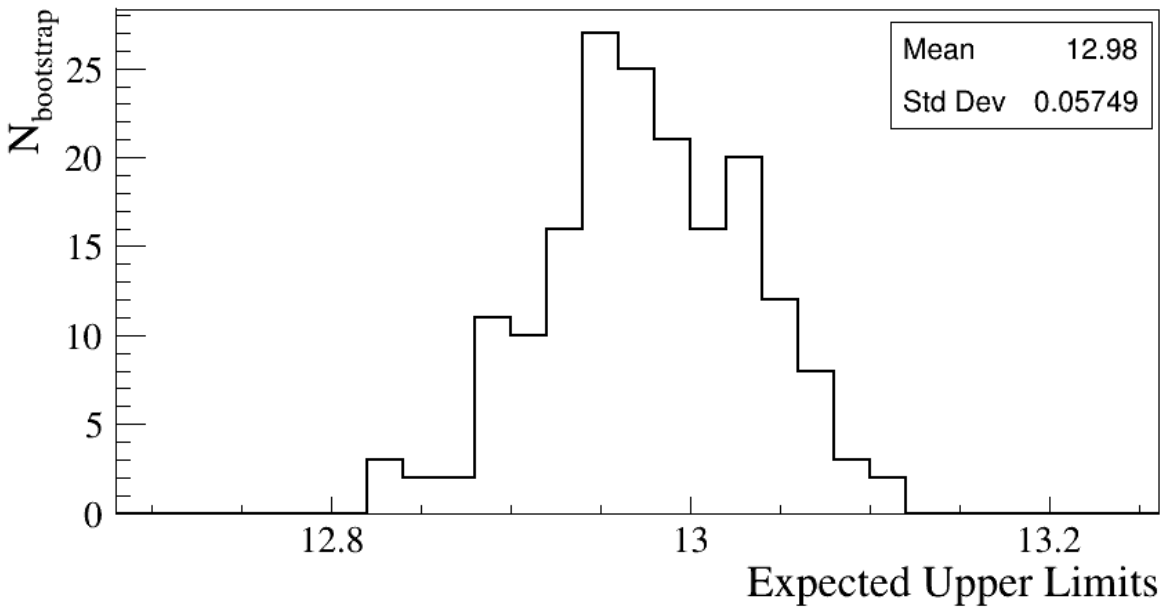

Figure 96: Distribution of $H \to b\bar{b}$ expected upper limits obtained by re-sampling the Control Region dataset.

The expected upper limit is found to be worsen by 0.14 % for the $H \to b\bar{b}$ and 1.8 % for the $H \to c\bar{c}$ by this systematic uncertainty.

6.8 Theoretical uncertainty on the Z cross section

$Z \to b\bar{b}$ and $Z \to c\bar{c}$ contributions are fixed to the SM expectation at leading order (NLO) calculated with MADGRAPH. Their yields should be varied as nuisance parameters to take into account the uncertainty on the prediction. The systematic uncertainty related to the SM cross section calculation and have been found with MADGRAPH to be around $+5\%$ and -10% and has been added nuisance parameters in the likelihood for the upper limit calculation, relative to the Z Monte Carlo normalization. The expected upper limit is found to be worsen by 2.1 % for the $H \to b\bar{b}$ and 0.8 % for the $H \to c\bar{c}$ by this systematic uncertainty.

6.9 Theoretical uncertainty on the Higgs cross section and acceptance

Even the $H \to c\bar{c}$ ($H \to b\bar{b}$) contribution in the SR of the $H \to b\bar{b}$ ($H \to c\bar{c}$) search is fixed to the SM expectation. In this case a systematic uncertainty is calculated by considering the prediction uncertainty [95].

$$\sigma_{H,ggF} = 48.52^{+4.6}_{-6.7} \pm 3.9 \pm 3.2 \pm 1.9 \pm 2.6 pb$$

The total $+1\sigma$ variation is found to be 7.5 %, the total -1σ variation is found to be 9 %. The systematic uncertainty related to the SM cross section calculation has been added as nuisance parameters in the likelihood for the upper limit calculation, relative to the H Monte Carlo normalization. The expected upper limit is found to be worsen by 1.5 % for the $H \to b\bar{b}$ and 1.2 % for the $H \to c\bar{c}$ due to this systematic uncertainty. A further 4% variation has been summed in quadrature to the systematic uncertainty due to the cross section, to take into account for the acceptance correction factor described in section 5.5. The expected upper limit is found to be worsen by 1.6 % for the $H \to b\bar{b}$ and 1.7 % for the $H \to c\bar{c}$ due to this systematic uncertainty.

6.10 Luminosity

The luminosity determination has a precision of 2.0% [102].It has been implemented in the upper limit computation as a common parameter to all Monte Carlo samples, and is found to increase the expected upper limit on the $H \to b\bar{b}$ by 1.2 % and on the the $H \to c\bar{c}$ by 0.2 %

6.11 Systematic overview

The systematic uncertainties and their individual variation on the upper limit calculation are summarized in Table 41. The observed and expected upper limits have been calculated by including the dominant systematic uncertainties as nuisance parameters and the results are shown in the next section 6.12.

Systematic source	Variation upper limit $H \to b\bar{b}$	Variation upper limit $H \to c\bar{c}$
GEC	0.4 %	0.1 %
L0 + HLT1	3.2 %	3.6 %
HLT2	-	-
Jet SV-tagging	5.4 %	23 %
Jet Identification	0.7 %	0.1 %
Jet Energy Resolution	0.03 %	0.2 %
Jet Energy Scale	1.2 %	0.3 %
Background model	$\sim$ 1 %	$\sim$ 1.2 %
Z Cross section	2.1 %	0.8 %
DNN	0.13 %	1.8 %
Higgs Cross section + Acceptance	1.6%	1.7%
Luminosity	1.2 %	0.2 %

Table 41: List of systematics uncertainties studied.

6.12 Upper limits results

The dominant systematic uncertainty is related to the SV tagging. The L0 and HLT1 are the second dominant systematic uncertainty. The inclusive observed $H \to b\bar{b}$ and $H \to c\bar{c}$ upper limits have been computed using the CLs technique and are shown in Figure 97 left and right respectively as a function of the cross section normalized to the standard model expectation. The SV tagging and L0+HLT1 nuisance parameters are considered in the computation of the upper limit on the $H \to b\bar{b}$ cross section, while the SV tagging nuisance parameter has been considered for the computation of the upper limit on the $H \to c\bar{c}$ cross section. The observed and expected upper limits are compatible within the 2σ uncertainty bar.

In table 42 and 43 the numerical values of the expected and observed upper limits are summarized.

Process	σ_{UP}/σ_{SM}	N_{up}	σ_{UP} [pb]
$H \to b\bar{b}$	14	3277	32
$H \to c\bar{c}$	2360	1793	274

Table 42: Expected upper limits on the $H \to b\bar{b}$ and $H \to c\bar{c}$ cross sections.

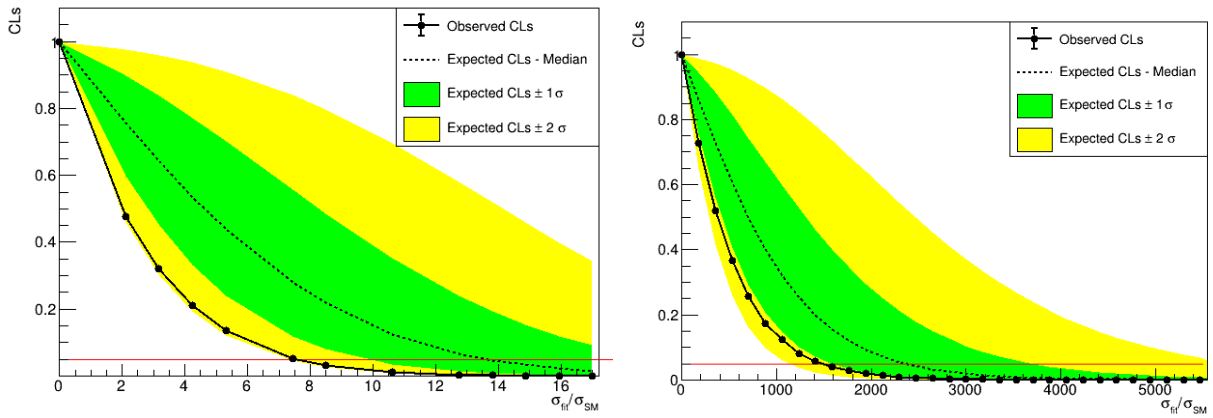

Figure 97: Left: observed and expected upper limits on $H \to b\bar{b}$ cross section. Right: observed and expected upper limits on $H \to c\bar{c}$ cross section.

Process	σ_{UP}/σ_{SM}	N_{up}	σ_{UP} [pb]
$H \to b\bar{b}$	7.5	1830	18
$H \to c\bar{c}$	1475	1121	171

Table 43: Observed upper limits on the $H \to b\bar{b}$ and $H \to c\bar{c}$ cross sections.

6.13 Higgs limits extrapolation to full Run 2 and Upgrade 2

In this section prospects on the LHCb sensitivity are discussed, with the full Run 2 dataset, where $\mathcal{L} = 5.4$ fb^{-1} integrated luminosity have been collected at 13 TeV, and with the Upgrade 2 conditions, where $\mathcal{L} = 300$ fb^{-1} at 14 TeV are expected to be collected. The upper limit roughly scales with the inverse of the Higgs significance $\mathcal{S} = \frac{S}{\sqrt{S+B}}$, where S and B are the number of signal and background events respectively. In general, the scale factor f between the upper limit calculated in Sec. 5.3 and the upper limit calculated at $\sqrt{s'}$ center of mass energy and $\mathcal{L}'$ integrated luminosity can be calculated as:

$$f = \frac{\mathcal{S}(13\text{ TeV}, 1.6\text{ fb}^{-1})}{\mathcal{S}(\sqrt{s'}, \mathcal{L}')} = \frac{\sigma_H(13\text{ TeV})}{\sigma_H(\sqrt{s'})} \cdot \frac{\epsilon_H(13\text{ TeV})}{\epsilon_H(\sqrt{s'})} \cdot \sqrt{\frac{\sigma_{bkg}(\sqrt{s'})}{\sigma_{bkg}(13\text{ TeV})}} \cdot \sqrt{\frac{\epsilon_{bkg}(\sqrt{s'})}{\epsilon_{bkg}(13\text{ TeV})}} \cdot \sqrt{\frac{1.6\text{ fb}^{-1}}{\mathcal{L}'}}$$

where ϵ_H and ϵ_{bkg} are the Higgs and the background efficiencies, while σ_H and σ_{bkg} are the respective cross sections in the fiducial region ($2 < \eta < 5$).

The scale factor for the full Run 2 is calculated assuming that the signal and background cross sections and selection efficiencies are the same for the 2016, 2017 and 2018 datasets:

$$f_{\text{Run2}} = \frac{\mathcal{S}(13\text{ TeV}, 1.6\text{ fb}^{-1})}{\mathcal{S}(13\text{ TeV}, 5.4\text{ fb}^{-1})} = \sqrt{\frac{1.6\text{ fb}^{-1}}{5.4\text{ fb}^{-1}}} = 0.54 \tag{6.5}$$

For the Upgrade 2 scale factor, we assume the same detector performance of Run 2, therefore we consider the same selection efficiencies of the 2016 analysis. The signal and background cross sections in the fiducial region at 13 TeV and 14 TeV have been obtained with PYTHIA. It has been found that the cross section ratio is similar for signal and background, therefore

$$f_{\text{U2}} = \frac{\mathcal{S}(13\text{ TeV}, 1.6\text{ fb}^{-1})}{\mathcal{S}(14\text{ TeV}, 300\text{ fb}^{-1})} \simeq \sqrt{\frac{1.6\text{ fb}^{-1}}{300\text{ fb}^{-1}}} = 0.073 \tag{6.6}$$

The upper limits that can be obtained with the full statistics of Run 2 and prospects for Upgrade 2 are shown in Table 44.

Process	σ_{UP}/σ_{SM} 2016	σ_{UP}/σ_{SM} Run2	σ_{UP}/σ_{SM} U2	σ_{UP}/σ_{SM} U2 w/o HCAL
$H \to b\bar{b}$	13	7	0.9	1.1
$H \to c\bar{c}$	2359	1274	172	203

Table 44: Extrapolation of Higgs expected upper limits with the full Run 2 dataset and with the Upgrade 2.

Process	σ_{UP}/σ_{SM} 2016	σ_{UP}/σ_{SM} Run2	σ_{UP}/σ_{SM} U2	σ_{UP}/σ_{SM} U2 w/o HCAL
$H \to b\bar{b}$	7.5	4	0.5	0.6
$H \to c\bar{c}$	1475	796	108	127

Table 45: Extrapolation of Higgs observed upper limits with the full Run 2 dataset and with the Upgrade 2.

6.13.1 Impact of HCAL removal in Upgrade 2

The removal of the HCAL and its substitution with a shielding for the Muon System has been proposed for Run 5/Upgrade 2. The impact of HCAL removal on the jet reconstruction has been studied in [103]. It has two effects that are relevant for this analysis: a degradation of the jet p_{T} resolution at high p_{T}, and a loss of jet reconstruction efficiency. The impact of the HCAL removal on the Higgs invariant mass is shown in figure 98. The combined impact of these two effects in the inclusive Higgs search, is a reduction of about a factor 0.85 on the Higgs significance [103]. This factor is applied to the $H \to b\bar{b}$ and $H \to c\bar{c}$ limits extrapolated to the Upgrade 2, and reported in Table 44 as a comparison.

6.13.2 Improvements with dedicated trigger lines

The 2016 analysis is limited by the fact the just the HLT2 dijet line with two reconstructed SVs is available without pre-scaling. The DNN tagging can be applied also to jets without a reconstructed SV. The possible improvement of the signal significance on a sample of data without any requirement on the secondary vertex and by only applying the identification power of the DNN to all jets, a sample of dijets has been evaluated. The highly pre-scaled sample (prescale:$1.3 \cdot 10^{-5}$) described in section 5.2.4 has been used and $H \to b\bar{b}$ and $H \to c\bar{c}$ samples have been produced with requirements mentioned in Table 13 without the requirement of secondary vertices. The DNN P_c and P_q have been optimized to maximize the significance of the $H \to c\bar{c}$ signal. The maximum significance with the dijet sample without any requirement on the secondary vertices, S_{dijet}, is found to be a factor 1.24 and 1.39 times the respective maximum significance found by optimizing the DNN P_b and P_c on dijet data with two secondary vertices, for the $H \to b\bar{b}$ and $H \to c\bar{c}$ search respectively.

6.13.3 Improvements with L0 GEC removal

In the HL-LHC stage, the *GlobalEventCut* will be removed from the *L*0 analysis. In order to evaluate the improvement on the upper limit without the GEC, the nSPD distribution before the *GEC* correction has been fitted with the following function:

$$f = (nSPD) = Ae^{-\gamma(nSPD-\alpha)}(nSPD-\alpha)^{\beta} \tag{6.7}$$

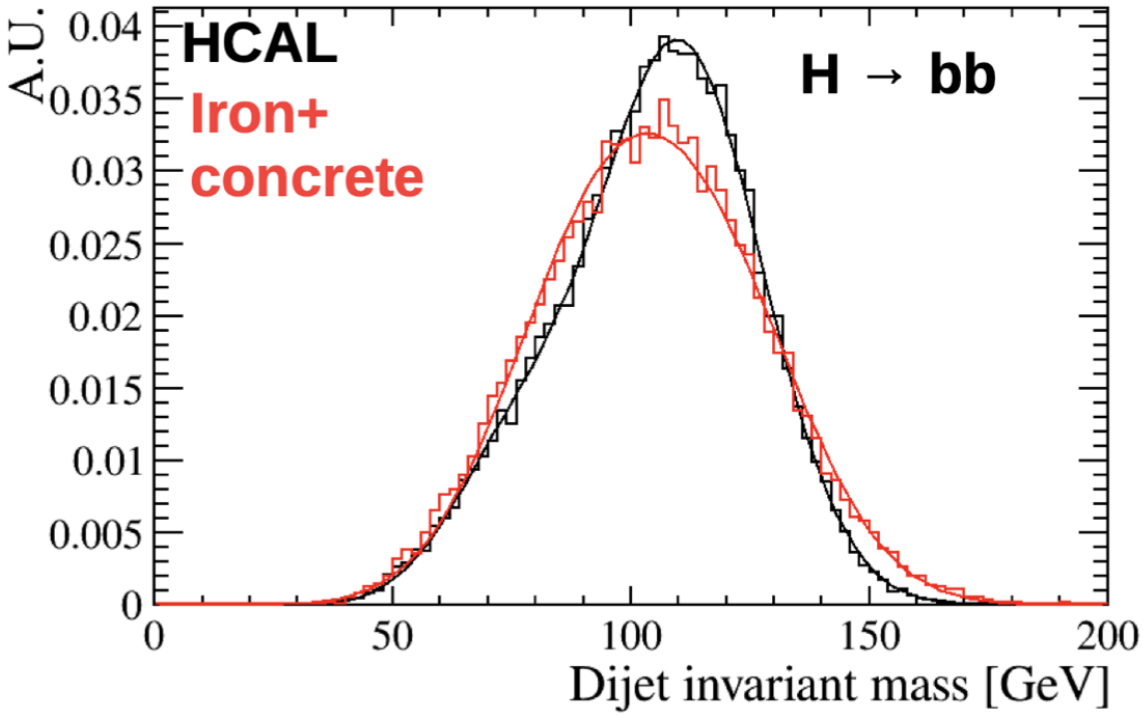

Figure 98: Reconstructed $H \to b\bar{b}$ dijet invariant mass obtained with the Run 4 simulation when HCAL is included (black) and when HCAL is removed and substituted with an iron+concrete shielding [103].

where A,α, β, γ are free parameters. In this way, the approximated shape of the nSPD distribution for events with nSPD>450 and nSPD>900 can be obtained. Figure 99 shown the $nSPD$ distribution of the $H \to b\bar{b}$ Monte Carlo fitted with equation 6.7. The tail of the $nSPD$ distribution for $nSPD$>450 is due to events that pass one of the $L0$ trigger lines listed in Table 9 that require $nSPD$<900 or that do not require any requirement on the $nSPD$. The nSPD distribution before the GEC correction has been taken for this exercise, since it has a larger statistics and allows for a more stable fit.

By comparing the integral of the function and the $nSPD$ distribution, it is found that without any requirement on the GEC the integral of the $nSPD$ distribution without the $nSPD$ correction would increase by a factor 1.13. The total improvement factor, taking into account the $\sim$ 63% further reduction of the $nSPD$ integral due to the $nSPD$ correction applied on the analysis, is then $\sim$ 1.8. By considering the same improvement even in the number of data, the significance would improve by a factor 1.34.

6.13.4 Regression technique for energy correction

As described in section 4.7 the discrepancy between truth-jets and reconstructed jets is corrected by multiplying the energy of the reconstructed jet by a correction factor, calculated as a function of the p_T,η,ϕ the charge particle fraction and the number of primary vertices found in the event. An alternative way to correct the reconstructed jet energy is under investigation, and is based on the application of machine learning technique, the Gradient Boosting Regressor (GBR), to perform jet energy correction with the aim of improving the invariant mass resolution of the $H \to b\bar{b}$ and $H \to c\bar{c}$. The inputs are the jet kinematic properties, like the jet p_T, η and mass, the jet composition in terms of number of daughters and tracks inside the jet and their kinematic properties and the jet energy distribution inside the region defined by ΔR. This algorithm aims to find the best correction

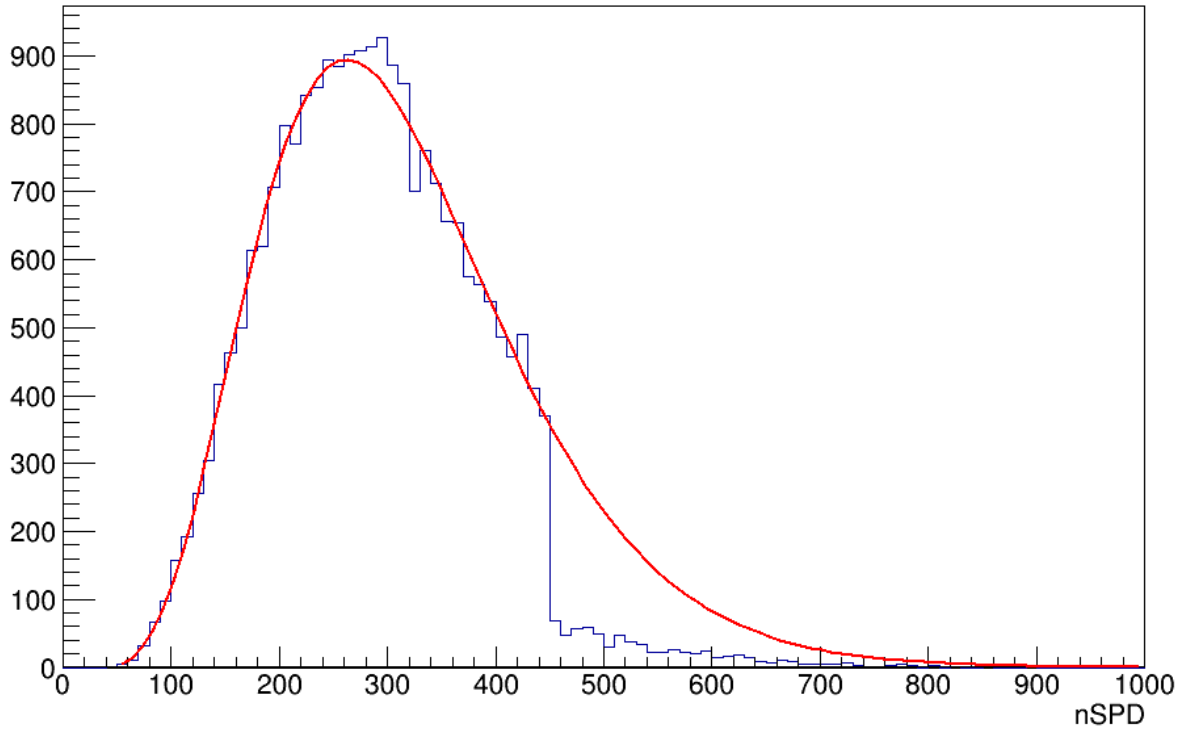

Figure 99: $nSPD$ distribution of the $H \to b\bar{b}$ Monte Carlo fitted with equation 6.7.

factors separately for each the two jets in the pair. Figure 100 shows the invariant mass calculated at truth level, with the default jet energy correction and with the GBR method. The GBR has been applied even to a small sample of data, in order to calculate the percentage of sensitivity improvements. From preliminary studies an improvement of a factor 1.22 is found by calculating the significance $S/\sqrt{S+B}$ after the GBR correction with respect to the default jet correction.

6.13.5 Final prospects

The final prospects on the observed upper limits foreseen for the Upgrade 2 are shown in Table 46. The upper limits have been scaled to include the improvements on the signal efficiency by not requiring the Global Event Cut at L0 trigger level, the employment of the Deep Neural Network at trigger level to select jets without requirements on the secondary vertex (+DNN), and the employment of the regression technique to correct the jet energy (+regression).

Process	σ_{UP}/σ_{SM} U2 w/o HCAL	σ_{UP}/σ_{SM} No GEC	σ_{UP}/σ_{SM} (+DNN)	σ_{UP}/σ_{SM} (+regression)
$H \to b\bar{b}$	0.64	0.48	0.38	0.31
$H \to c\bar{c}$	127	95	68	56

Table 46: Extrapolation of Higgs observed upper limits with improvements related to the GEC removal, the employment of the DNN at trigger level and the regression technique.

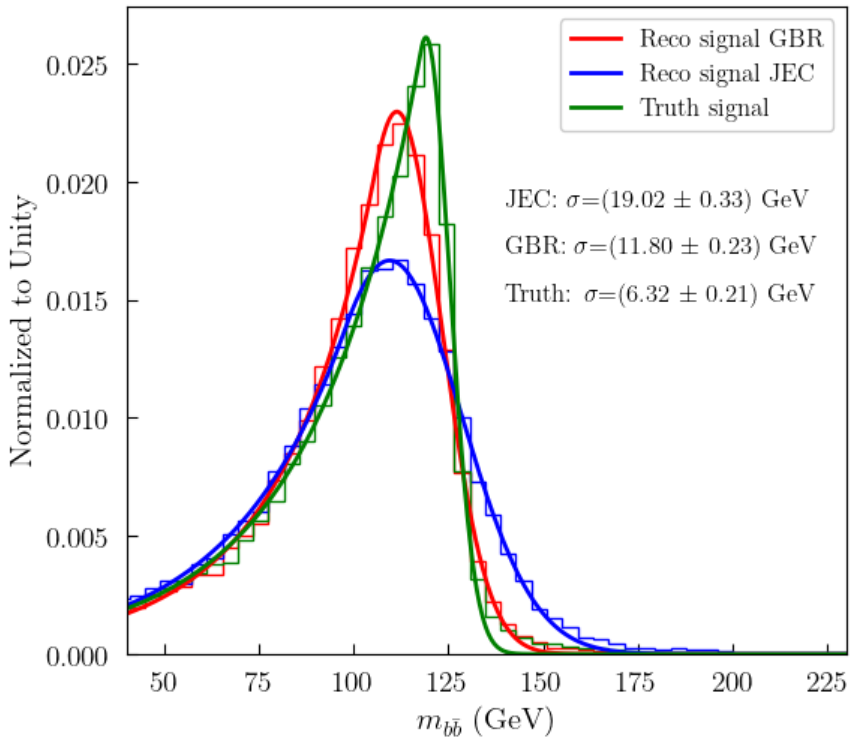

Figure 100: Fitting of the three invariant mass peaks (truth level and reconstruction level with default correction and GBR correction) to the crystal-ball function

7 The Muon Collider machine and detector

In the previous chapters the single Higgs production at LHCb, where pp are made collided at 13 TeV center of mass energy, has been studied by searching for heavy quarks final states $H \to b\bar{b}$ and $H \to c\bar{c}$. In the next chapters, the Higgs reconstruction at the Muon Collider, a future collider machine where $\mu^+\mu^-$ will collide at multi-TeV center of mass energies, is studied. In particular, the uncertainty on the HH production cross section and the trilinear Higgs self-couplings will be evaluated. The $H \to b\bar{b}$ decay channel is selected, in order to exploit the high decay branching ratio. As will be shown in section 7.5 the Muon Collider detector foreseen a cylindrical shape surrounding the collision point, differently from LHCb which is forward. However, the techniques used in this thesis to reconstruct HH events at Muon Collider are similar to the one employed by experiments at hadron collider that have been presented in section 4, like the Particle Flow approach and jet clustering for jets, tracks and secondary vertices reconstruction algorithms. As will be shown in section 8, these techniques have been optimized to deal with the Muon Collider environment, where high levels of beam-induced background due to muons decay are present.

7.1 The Muon Collider machine

The idea of a Muon Collider has been discussed since 1960 [104] and its design was developed over the years. Between 2011 and 2018 the Muon Accelerator Program (MAP) [105] in the US focused its efforts in the development of a concept for the Muon Collider, setting the baseline and the tentative target parameters, that is now the starting point of the International Muon Collider Collaboration (IMCC). The Collaboration was born in 2022 and aims to define the feasibility of the Muon Collider.

Indeed, the Muon Collider accelerator presents several technical challenges [37]. As shown in equation 3.1, in order to reach high luminosity, it is necessary to produce high collimated and compact beams. This is particularly difficult at a Muon Collider, since with the muon production method proposed by the MAP collaboration (see next section 7.2), that is the one adopted by the IMCC and assumed in these physics studies, muons are produced with high spread in position and energy. Due to the short lifetime, muons have to be produced and accelerated before a significant loss has occurred. At relativistic velocities, a muon with energy E has a relativistic factor $\gamma = \frac{E}{m_\mu}$, where m_μ is the mass of muons, and its lifetime in the laboratory frame τ_{lab} will increase as:

$$\tau_{lab} = \tau \cdot \gamma$$

where τ is the lifetime in the rest frame. However even at high center of mass energies, the rate of muon decays along the beam pipe is not negligible, and their decay products give origin to high beam-induced background level [106]. This chapter gives an overview of the present baseline design of the Muon Collider machine and the Muon Collider detector. The beam-induced background and its main effects on the detector design and the objects reconstruction are also described.

7.2 Overview of the facility

The conceptual scheme of a Muon Collider is driven by the need of high luminosity. The tentative target parameters for the proposed Muon Colliders at 3 and 10 TeV center of mass energies are summarized in Table 47, parameter definitions are the same of section 3.1. The muon production method proposed by the MAP collaboration is the proton driver scheme, that consists of a proton beam hitting a thick target. The $\pi^\pm$ produced decay into muons $\pi^\pm \to \mu^\pm + \overset{(-)}{\nu}$. Other methods have been proposed, like the Low EMittance Muon Accelerator scheme [107], where muons are produced by the process $e^+e^- \to \mu^+\mu^-$ by sending a positron beam at the threshold energy of 45 GeV to a target. This scheme would allow to produce highly collimated muon beams with a very

Center of mass energy	3 TeV	10 TeV
Luminosity	$2 \cdot 10^{34}$ cm^{-2} s^{-1}	$20 \cdot 10^{34}$ cm^{-2} s^{-1}
N_b	$2.2 \cdot 10^{12}$	$1.8 \cdot 10^{12}$
Fill rate	5 Hz	5 Hz
β^*	5 mm	1.5 mm
σ^*	3 μm	0.9 μm
σ_z	5 mm	1.5 mm
θ_c	-	-
Lorentz factor	14196	47322
ϵ_n	25 μm	25 μm

Table 47: Tentative target parameters for 3 TeV and 10 TeV Muon Colliders [37].

small emittance, but it is still under study since with the actual configuration a very high intense proton is needed to compensate for the low cross section of the process ($\sigma(e^+e^- \to \mu^+\mu^-) \sim 1\mu b$ for $\sqrt{s} \sim 230$ GeV) and produce enough muons. The conceptual scheme of the MAP configuration of the Muon Collider facility, that is the one adopted by the International Muon Collider Collaboration and for the physics studies done in this thesis, is shown in Figure 101.

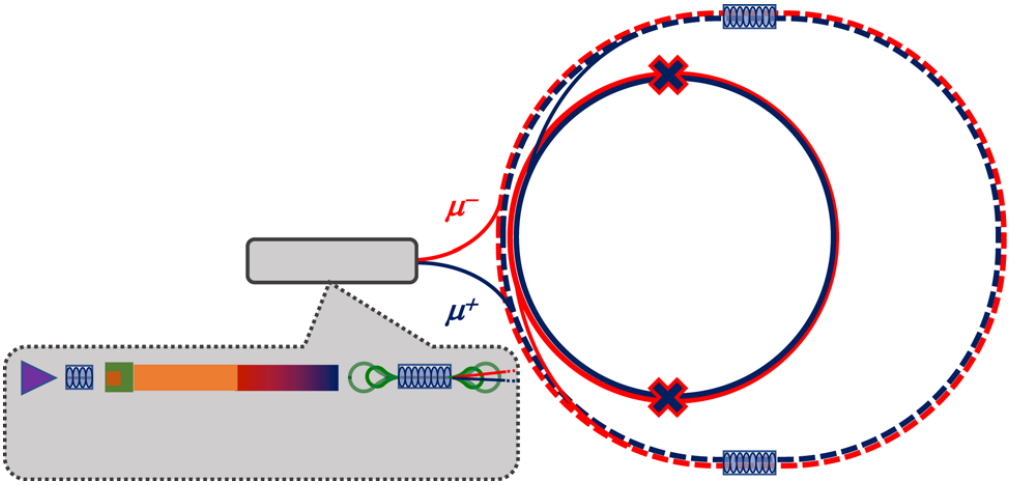

Figure 101: Conceptual scheme of a Muon Collider [37].

The main sections of the facility are:

1. Target and decay channel: in the Proton Driver, high intensity proton beams, with power between 2 and 4 MW are produced, with energy indicatively between 5 and 30 GeV are produced. Negative hydrogen ions are firstly created in a negative ion source, are converted into proton with charge-exchange stripping. The acceleration takes place by mean of a series of drift-tube linacs, and, in some designs, with Rapid Cycling Synchrotrons or Fixed Field Alternating gradient accelerators. The current design assumes $\sim$ 1-3 ns bunch length bunches at a 15 Hz repetition rate. The protons are then impinged into a high-Z target ($\sigma \sim$ mb) immersed in a high solenoid field (15-20 T) producing high transverse momenta pions, kaons, that decay into muons. The high magnetic field is necessary to transversely confine the charged pions. The target is subjected to significant instantaneous shock that can lead to damage. For this reason the target material have to be properly chosen. Graphite or heavy liquid metal target solutions are currently under investigation.

2. Muon front end: buncher and phase rotator. After the decay channel, muon beams are

captured and enter the buncher section, where a series of RF cavities with frequencies of the order of 300-350 Hz form a train of bunches of both signs. Then the train of bunches enters a phase rotator section, that reduced the energy spread between the bunches by decelerating high-energy bunches and accelerating low-energy bunches. A charge selection system then split the beam into a positive and a negative muon line.

3. Muon ionization cooling:

   - Transverse cooling: the muon beams produced by the muon front end have a large emittance, that is reduced via ionization cooling, in order to achieve high luminosity. The transverse phase space volume of the muon beam is reduced in this step by passing beams through a material, where the muons lose both transverse and longitudinal momentum via ionization. The lost energy in the longitudinal direction is then restored by mean of RF cavities. After several passages through the absorber and RF cavities, the transverse emittance will be reduced. This process is summarized in Figure 102. In order to limit the multiple Coulomb scattering of the muons off nuclei in the absorber, which would increase the beam transverse momentum, a low atomic number and the presence of an external magnetic field are required. The Muon Ionization Cooling Experiment demonstrated transverse ionization cooling of muons [108], using both liquid hydrogen and lithium hydride absorbers.

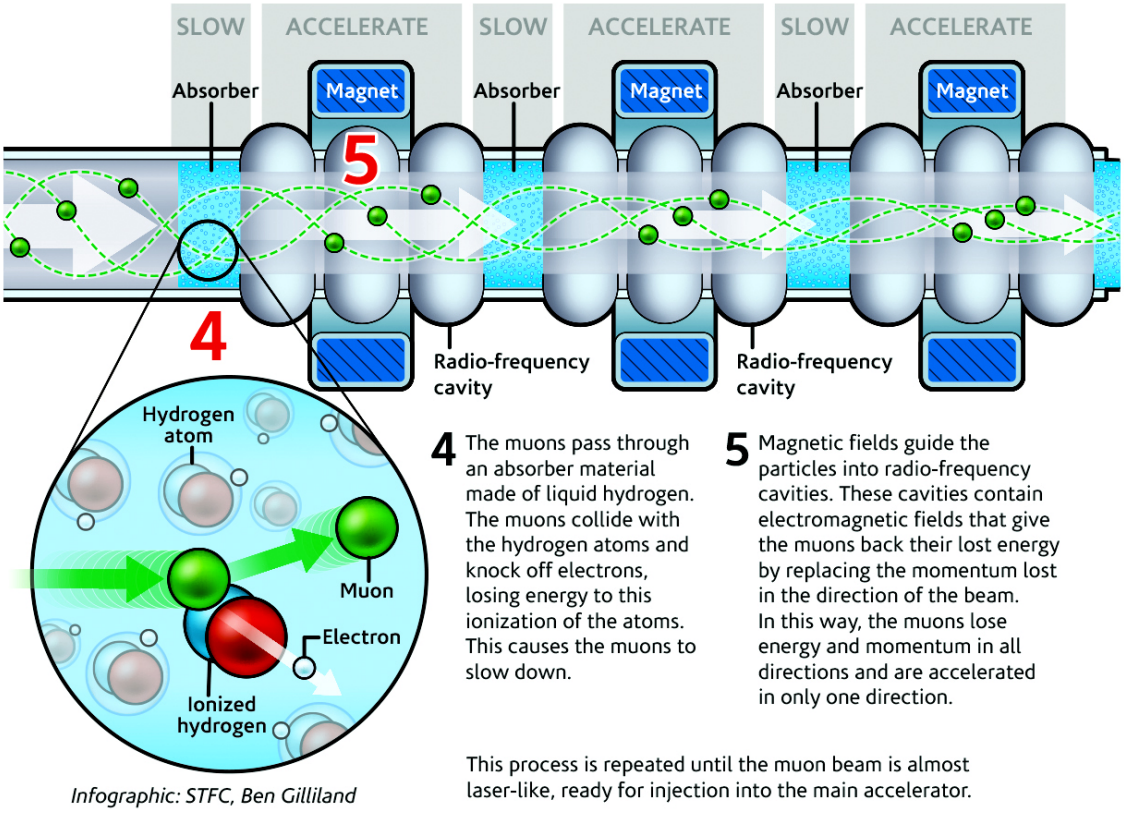

Figure 102: Conceptual picture of ionization cooling. Particles loose momentum in all direction via ionization by passing through an absorber. A RF cavity restores the longitudinal momentum only [109].

   - Longitudinal cooling: the reduction of longitudinal emittance can be achieved by mean of emittance exchange from the longitudinal to the transverse phase space. Its concept is represented in Figure 103: a bending magnet is used to separate particles with different

momentum, while the particle beam pass through a wedge absorber with variable thickness, properly designed to have grater energy loss at higher momenta. The result will be that the energy spread (longitudinal emittance) between particles is reduced, while the position spread (transverse emittance) is increased. The 6D cooling, which is assumed for the Muon Collider concept, will be a combination of the transverse ionization cooling and emittance exchange sections.

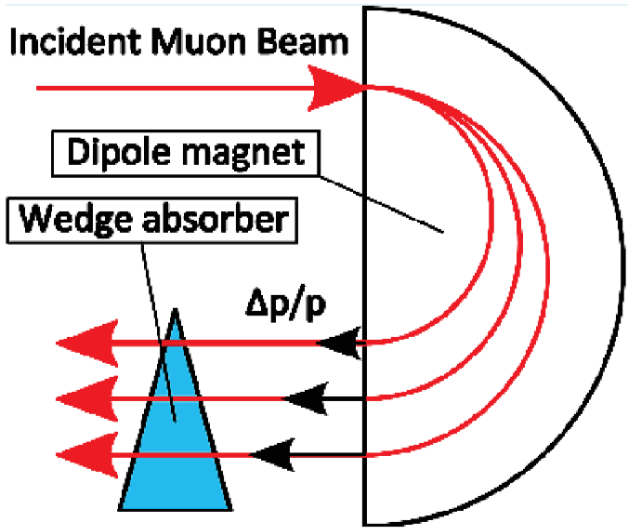

Figure 103: Schematic view of emittance exchange. Figure taken from [110].

- Acceleration: after the ionization cooling section the baseline design foresees a series of linacs followed by a series of pulsed synchrotrons to accelerate the beam before the collision. A design for this region in not available yet and some proposed approaches are still under study, like Recirculating Linear Accelerators (RLA) up to 100 GeV beam energies [111] and Rapid Cycling Synchrotron(RCS) [112] or Fixed Field Alternating Gradient accelerators (FFAs) [113] to improve the beam energy up to its final colliding energy (TeV). Preliminary estimates indicate that acceleration up to 3 TeV centre-of-mass energy would require a ring of circumference around 10 km, using 10 TeV bending magnets, while for a 10 TeV Muon Collider would require a circumference around 35 km, using 16 TeV bending magnets.
- Collider Ring: after the acceleration, the positive and negative muon beams are sent to the collision ring. High field bending magnets are required in order to keep a collider ring circumference as low as possible and keep the highest possible luminosity. 10 T dipole magnets for a 4.5 km circumference ring are assumed for a 3 TeV while 16 T dipole for a 10 km circumference ring are assumed for the 10 TeV Muon Collider. At present the strongest accelerator-style dipoles have field of 14.5 TeV, then further technological improvements are required to fulfill the requirements. Collider rings have been studied by MAP for different center of mass energies: 125 GeV, 1.5 TeV, 3 TeV and 6 TeV

7.3 The Machine-Detector Interface

The MAP collaboration developed the detailed design of the interaction region (IR) and of the machine detector interface at $\sqrt{s}$ of 1.5 TeV and 125 GeV. Figure 104 shows Muon Collider IR and the material of the components, from a distance of $\sim$ 14 m at left of the interaction point (IP), set at z =0 m to 21 m at the right of the IP.

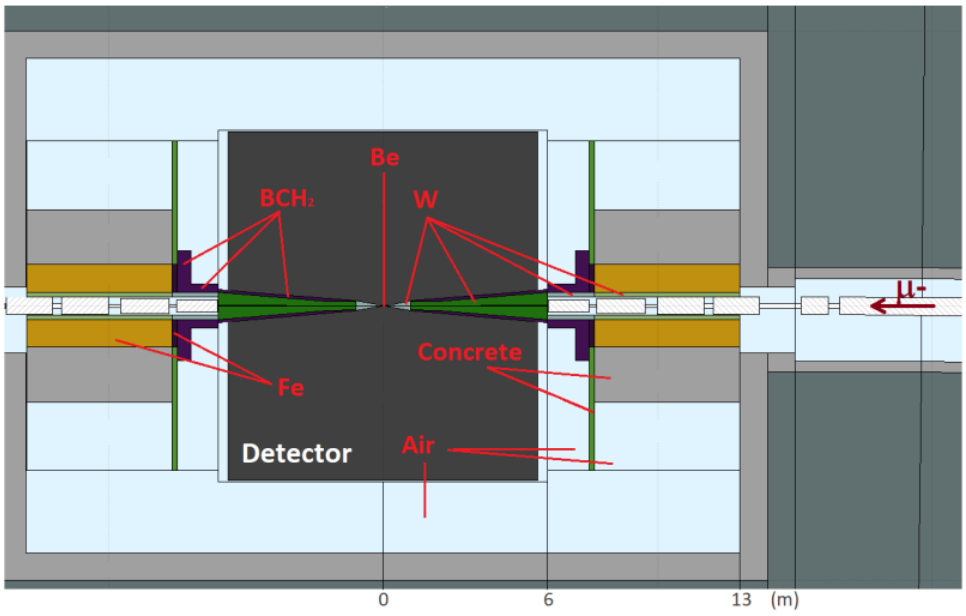

Figure 104: Interaction region. The passive elements, the nozzles and the pipe around the interaction point are constituted by iron (Fe), borated polyethylene (BCH_2), berillium (Be), tungsten (W) and concrete [114].

The incoming muon beam direction, that coincides with the z axis, is shown by the red arrow at the right of the picture. Going from right to the left of the picture, we find:

- the last $\sim$ 7 m of the arc of the collider ring before entering the interaction region. The white boxes represent the collider ring dipole magnets, located inside the vacuum vessel;
- at $\sim$ 14 m from the interaction point, the beam enters the interaction region, where the final superconducting quadrupole magnets, with nominal field of around 11 T, provide the final focus of the beam before the collision (white boxes). Different colors represent different materials that surround the beam pipe: iron (yellow and red) and concrete (gray);
- At 6 m from the IP the beam enters the detector region, that is represented by the black region in the picture, a cylinder of 11.28 m long and with radius 6.3 m. In green are represented two cone-shaped tungsten nozzles, whose function is fundamental for the background reduction produced by the muon decay along the beam pipe, as it will be explained in section 7.4. Figure 105 shows a zoom on the nozzle geometry and material description: at left the entire structure of the nozzle (up to 6 m from the IP) is shown, at right a zoom on the last 35 cm from the IP is shown. The tip of the nozzle is located at 6 cm from the interaction point. They are characterized by two sections, one closest to the IP (between 6 and 100 cm) with an angular aperture of $\sim 10°$ while between 100 cm to 6 m the angular aperture is reduced to $\sim 5°$, in order to minimize the impact on the detector angular acceptance. In the outer surface, the nozzles are coated by a layer of borated polyethylene (BCH_2) (in red in Figure 105 and in purple in Figure 104). The reason for the choice of these materials will be discussed in the next section 7.4.

7.4 The Beam-Induced Background

The beam-induced background at a Muon Collider is generated by the muon decay along the collider beam pipe. At $\sqrt{s}$=1.5 TeV, a single 0.75 TeV muon bunch with $2 \cdot 10^{12}$ particles produce

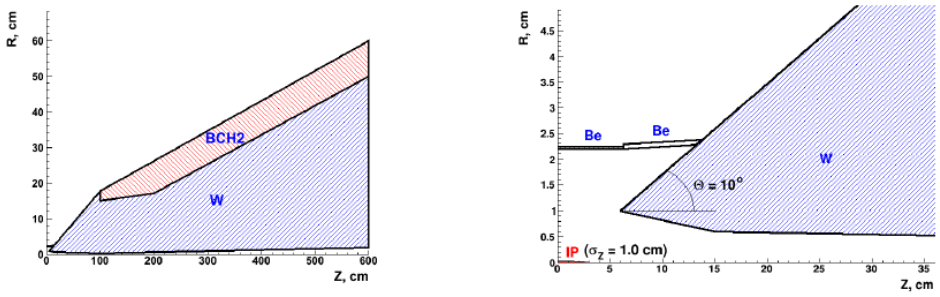

Figure 105: Detailed geometry and material description of the nozzle [114].

$4.28 \cdot 10^5$ decays per meter of lattice for a single muon beam. This number decreases to $2.1 \cdot 10^5$ decays per meter of lattice for a single muon beam at 3 TeV center of mass energy. The BIB that have been simulated in the physics studies performed in this thesis is the one obtained at 1.5 TeV by the MAP collaboration, that used the MARS15 code to simulate the geometry, materials distribution of the magnetic field lattice elements, the detector and the MDI in the $\pm$ 200 m region from the IP, to characterize the beam-induced background, in order to find solutions to reduce it. Even if the $\mu^+\mu^-$ collisions for the HH analysis performed in this thesis are at 3 TeV center of mass energy, the BIB at 1.5 TeV center of mass energy generated with the MAP software has been used, since it was the only available at the moment of the thesis. However, the MAP results at 1.5 TeV have been reproduced in [114] with the FLUKA Monte Carlo multi-particle transport code. The FLUKA software [115] [116] have been used to generated BIB even at 3 TeV and 10 TeV, and, as will be shown in section 7.4.2, found that the BIB level and characteristics at 3 and 1.5 TeV are very similar.

In the next subsection 7.4.1 the results of MAP studies on the BIB are summarized and the characteristics of the BIB used in this analysis are shown. In subsection 7.4.2 the studies performed at 3 and 10 TeV with FLUKA are shown.

7.4.1 1.5 TeV Beam Induced Background

Energetic electrons produced by the muon decay along the beam pipe emit synchrotron photons due to the presence of the high magnetic field in the ring magnets and neutrinos. The interaction of electrons and synchrotron photons with the beam pipe and the surrounding material produces electromagnetic showers, hadrons via photo-nuclear interactions and muons via Bethe-Heitler process. All these secondary and tertiary particles interact in turn with the lattice and the detector components, producing other electromagnetic showers and hadronic cascades via nuclear interaction processes. The high flux of particles produced from these interactions lead to a deposit of energy on the surrounding machine components. Particles from muons that decay in proximity of the machine-detector interface may reach the interaction point and lead to energy deposits on the detector. Without any protection, the large flux of particles, around 0.5-1 kW/m [117], can cause hardware problems, like magnet quench and high heat load to the cryogenic system in the collider ring, while may affect the detector performance and consequently the physics measurements.

The simulations conducted by MAP demonstrated the effectiveness of the two nozzles in the reduction of the high flux of particles which reach the IR and the detector. Furthermore, they optimized the design (like the aperture angle and the distance from the interaction point) and the material of the nozzles to maximize the beam-induced background absorption. The usage of tungsten, which is a high-Z material, combined with the detector solenoid magnetic field, was found particularly effective for the absorption of the electromagnetic component of the beam-induced

background, while the borated polyethylene shell is added to reduce the flux of low-energy neutrons. At 1.5 TeV the presence of the nozzles reduces by the $\sim$ 85 % and $\sim$ 97 % the number of photon and electrons reaching the detector, absorbing the most energetic particles, and by 60 % the number of charge hadrons [114]. Most of the BIB particles reaching the detector are produced by muons decaying in a range of $\pm$ 25 m around the IP, while the Bethe Heitler muons can be produced by muons that decay up to $\pm$100 m from the IP [117]. The particles that survive the nozzles and reach the detector are mainly photons (94%) and neutrons (4%), followed by electrons (1%) and positrons, charged hadrons and muons (<1%) [114]. Figure 106 shows the number of background particles entering the detector through the external surface of the tungsten nozzles as a function of the distance from IP to muon decay point along the beam can be seen. The different particles type are shown in different colours: photons in red, neutrons in blue, electrons in black, charged hadrons in pink and muons in green. Figure 107 shows the particle momentum spectra. The average

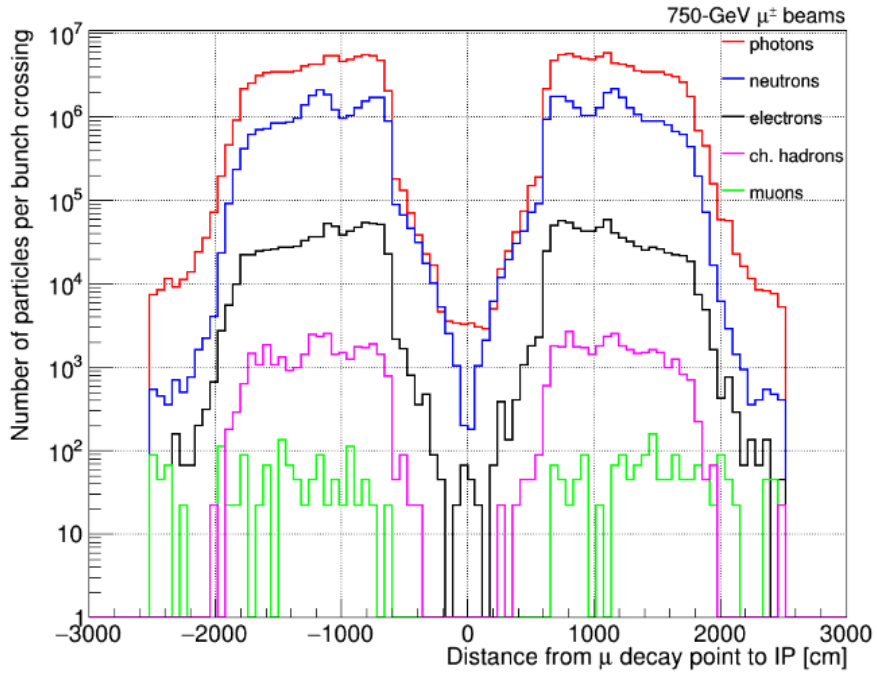

Figure 106: Number of background particles as a function of distance from IP to muon decay point along the beam [118].

momentum of the electromagnetic component is 1.7 MeV for photons, and 6.4 MeV for electrons, while it is 477 MeV for neutrons and 481 for charged hadrons. Muons are the component with the highest values of the momenta. The low momentum BIB particles can be rejected, as will be shown in the next chapter, by applying a cut on the energy released in the detector components or by using double-layer silicon detectors, since have higher probability to release energy in only one of the two layers. Another characteristic that is exploited for BIB reduction is that BIB particles are characterized by a broad arrival time at the detector. Figure 108 shows the time of flight of BIB particles entering the MDI surface with respect to the bunch crossing time. The time distribution shows a peak around $t = 0$ that is mainly due to muon decays close to the interaction point, that arrive at the same time with the bunch crossing. The long tails up to tens of ns are due to particles that are produced far from the IP or that are produced after multiple interaction of other particles with the MDI components. A proper choice of the acquisition time window will allow to reduce the amount of beam-induced background during the data taking.

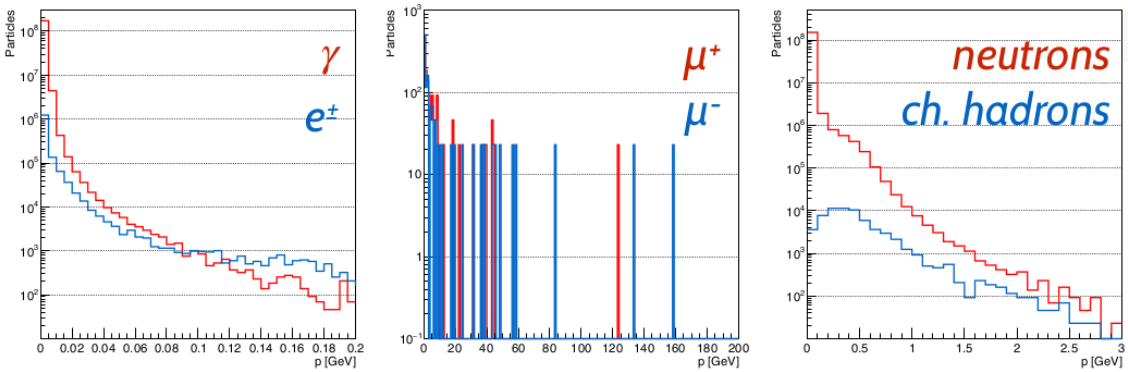

Figure 107: Momentum distribution of beam-induced background particles [119].

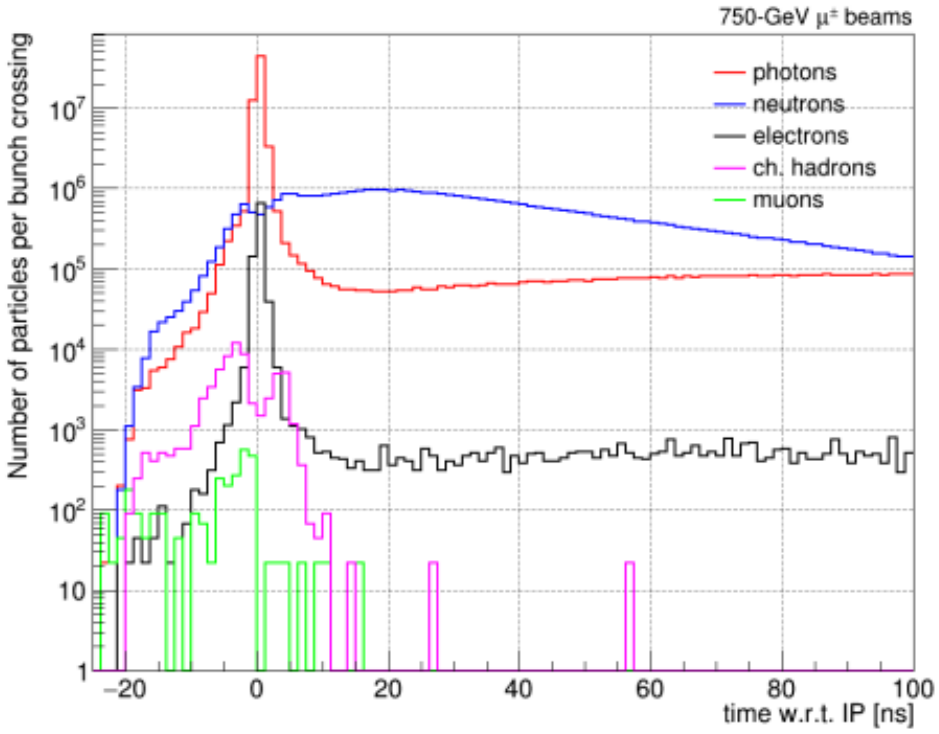

Figure 108: Time of arrival distribution of beam-induced background particles at the detector surface [118].

7.4.2 The Beam-Induced Background at Multi-TeV Muon Collider

Dedicated studies are necessary for each center of mass energy: on one side the BIB characteristics and the MDI design are interdependent. Furthermore, products of muon decays may reach the interaction region from distances that depend on the center of mass energy. In [114] the MAP MDI design at 1.5 TeV have been implemented in the FLUKA software and the BIB properties in this configuration have been reproduced and are found to be in agreement to those obtained with MARS15 code. This result is important since MARS15 and FLUKA are two different software that are based on different tools to simulate particle-matter interactions. Studies for 3 TeV and 10 TeV are in its early stage, but here main results obtained up to now are summarized. The $\sqrt{s}$=3 TeV machine configuration including the interaction (IR) has been designed by the MAP program, but

does not include the nozzle design optimized for that energy. The configuration of the IR at $\sqrt{s}$ = 3 TeV by using the same structure of $\sqrt{s}$ = 1.5 TeV has been implemented with FLUKA and the BIB properties are compared to those at 1.5 TeV center of mass energy. As explained above, this is not fully realistic but allows for a first assessment of the BIB levels dependence on the collider energy. The distance from IP of primary muons decay considered at 3 TeV to include all possible decay on detector surface is around twice the distance considered for the 1.5 TeV case, around 55 m. Comparison between the 1.5 TeV and the 3 TeV BIB are shown in Figure 109 [120]. In the left picture, the time of arrival of different types of particles are compared between 1.5 and 3 TeV in the right picture the energy distributions of the particles that reach the detector are compared, in the time window [-1ns,15ns] with respect to the bunch crossing. The arrival time and the energy distribution shapes look very similar for all particle types in the two energies cases.

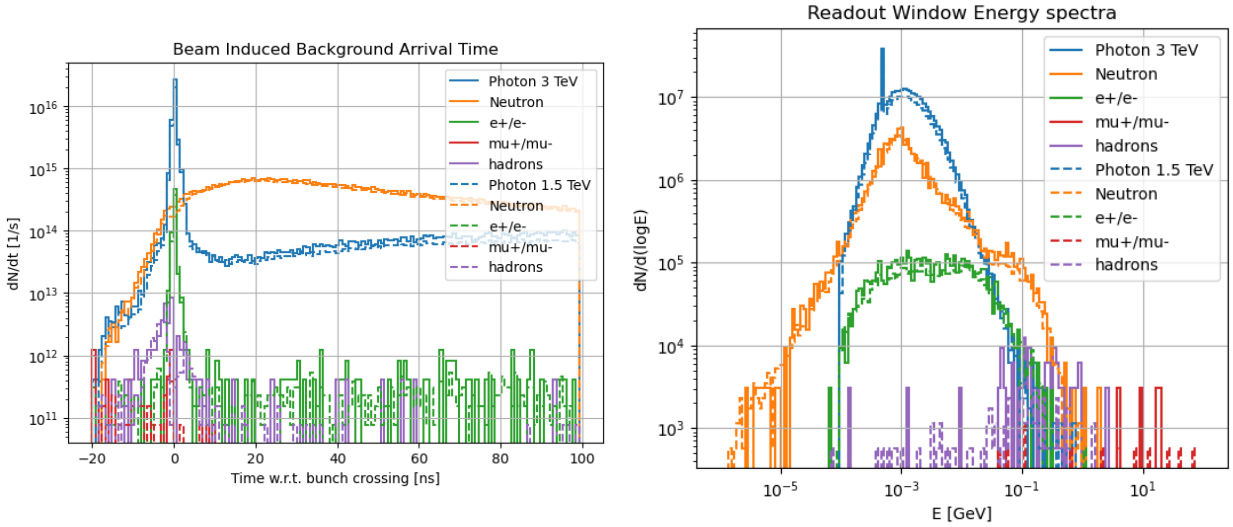

Figure 109: Left: Distribution of the arrival time of the variuos particle types for $\sqrt{s}$ = 1.5 TeV (dashed line) and $\sqrt{s}$ = 3 TeV (solid line). Different colors represent different types of particles. Right: Energy distribution of particles arriving at the detector for the two center of mass energies [120].

The preliminary design at $\sqrt{s}$=10 TeV has been done by the IMCC [121] The configuration of the IR at $\sqrt{s}$ = 10 TeV by using the same structure of nozzles $\sqrt{s}$ = 1.5 TeV has been implemented with FLUKA and the BIB properties are compared to those at 3 TeV center of mass energy. The distance from IP of primary muons decay considered at 10 TeV to include all possible decay on detector surface is around 100 m. Also in this case, the distributions in time and energy are similar for the two center of mass colliders. This demonstrates that thanks to the presence of the nozzles the characteristics of the BIB particles that arrive at the detector are similar for the three center of mass energies. A larger flux of particles is expected at 10 TeV with respect to 3 TeV. However studies for the optimization of the MDI at 10 TeV center of mass energy are ongoing, and a reduction of the amount of BIB is expected. For example, as shown in [122], a dependency of the secondary particle fluence on the distance between IP and final focus magnets is found, then a possible reduction could be obtained by increasing this distance. Another possibility for optimizing the physics performance at 10 TeV is the nozzle design: small modifications of the nozzle tip affect the particle multiplicity around the IP. In addition to the adjustment of the nozzle tip and the nozzle aperture, several other geometrical variables of the nozzle design can be modified to tune the BIB suppression, like the modification of the thickness of the tungsten shield or extending the borated polyethylene layer [123]. Furthermore, also the detector magnetic field produced by the detector plays an important role. For example, using a 5 T magnetic field instead of the 3.57 T expected for the 1.5 TeV can

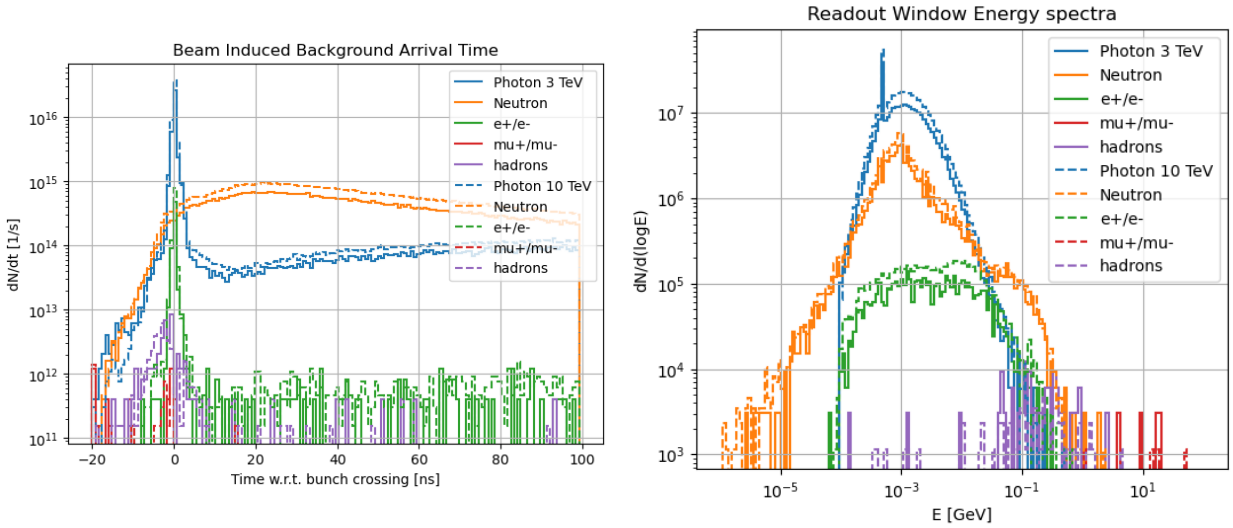

Figure 110: Left: Distribution of the arrival time of the various particle types for for $\sqrt{s} = 3$ TeV (dashed line) and $\sqrt{s} = 10$ TeV (solid line). Different colors represent different types of particles. Right: Energy distribution of particles arriving at the detector for the two center of mass energies [120].

reduce the fluxes of e^+/e^- BIB particles in the 10 TeV case [120].

7.5 Muon Collider detector

The Muon Collider detector design used for the full simulation of the physics events used in the analysis described in this thesis, is explained in this section. It has been implemented in the Muon Collider software framework, initially branched from ILCSoft [124]. This detector was originally designed and implemented in the framework by the CLIC collaboration, in the version that can be found in [125]. For Muon Collider studies, several detector components were modified and implemented in the code according to the studies performed by MAP [126]: the tungsten nozzles that absorb part of the BIB have been added to the geometry, the design of the vertex and tracker detectors was changed, and the two small electromagnetic calorimeters LumiCal and BeamCal designed in the forward region of the CLIC detector were removed from the design, to host the nozzles.

The detector has a cylindrical shape, 11.4 m long and with a diameter of 12.8 m. The interaction point is located at the center of the detector, that covers an angular region of pseudorapidity $-2.44 < \eta < 2.44$. It is composed, by going from the interaction point to the outside, from:

- A vertex and a tracker detector
- An electromagnetic calorimeter
- A hadronic calorimeter
- A superconducting solenoid which provide a magnetic field of 3.57 T
- An iron yoke instrumented with a resistive plate chamber for the muons detection

The coordinate system is right-handed, with the z-axis direction parallel to the beam pipe, the y-axis parallel to the gravity direction and the x-axis is defined as perpendicular to the y and z axes.

7.5.1 Vertex detector and tracking system

The vertex detector and the tracker system, aimed to the measurement of the charged particles transverse momenta that are produced in the muon collision, are composed by barrels and endcaps. Barrels are cylindrical surface with variable length and radius, whose axis coincides with the beam pipe and cover the central part of the detector. The endcaps are annulus centered on the z axis, with variable distance from the interaction point and radius which cover the forward part of the detector.

The vertex detector (VTX) is the closer to the beam pipe and is used to identify primary and displaced secondary vertices.

The vertex detector is composed by:

- four barrel layers with silicon pixel of size 25×25 μm^2, and thickness 50 μm with radius between 3.1 to 10.2 cm.
- eight endcaps layers, four for each side of the interaction point are composed by silicon pixel of size 25×25 μm^2 and thickness 50 μm and 16 modules. They are located at a distance from 8 cm to 28 cm from the interaction point along the z axis.

The timing resolution of the vertex detector sensors is assumed to be 30 ps.

Figure 111 left shows the four layers of the vertex barrel (black cylinders). Figure 111 right shows the half section of the longitudinal view of the vertex detector, with $z \in [-17$ cm, 28 cm] with respect to the interaction point. The half section of two endcap layers out of the four are shown at the left of the IP, while all the four endcap layers are shown at right. The beam pipe is coloured in purple.

The building blocks of the barrel layers are rectangular staves of sensors, arranged to form a cylinder (see Figure 111 left), while the endcaps are constituted by trapezoidal modules of sensors, arranged as "petal" to form the disk. They are all composed by double-sensor layers, two sensitive layers fixed on one support structure, in both barrel and forward region, with 2 mm gap.

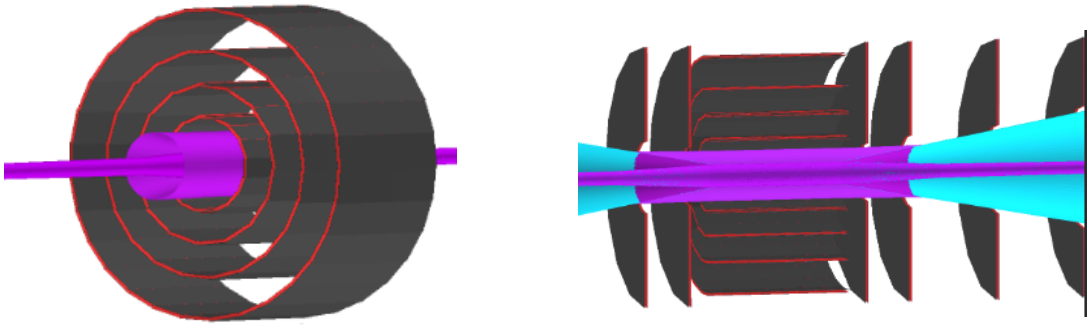

Figure 111: Left: longitudinal view of the vertex detector barrel layers. Right: longitudinal section of the vertex detector with $z \in [-17$ cm, 28 cm].

Surrounding the vertex detector, there are the inner tracker barrel and the outer tracker system. The scheme of the full tracker system is shown in Figure 112. They are composed by a single layer of silicon sensors of 100μ m thickness.

The inner tracker is composed by:

- three barrel layers with radius between 12.7 to 55.4 cm. Strips on the barrels have size 50 μm $\times$ 1 mm and thickness 100 μm are oriented with the long side parallel to the beam axis.

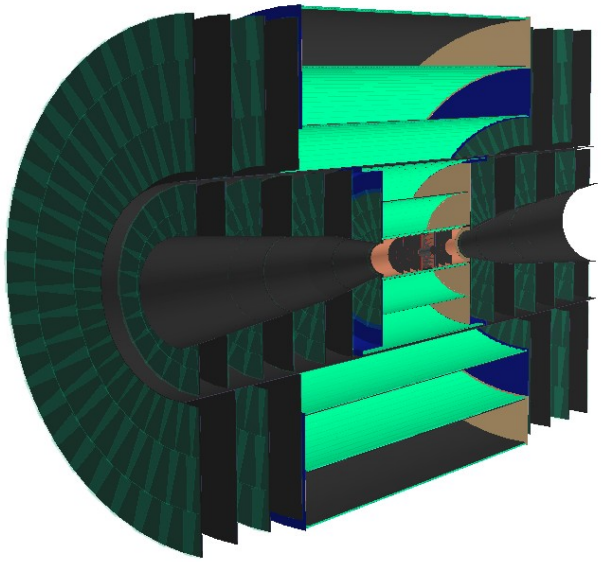

Figure 112: Longitudinal section of the full tracker system. The inner tracker is located in the central part, surrounded by the outer tracker layers.

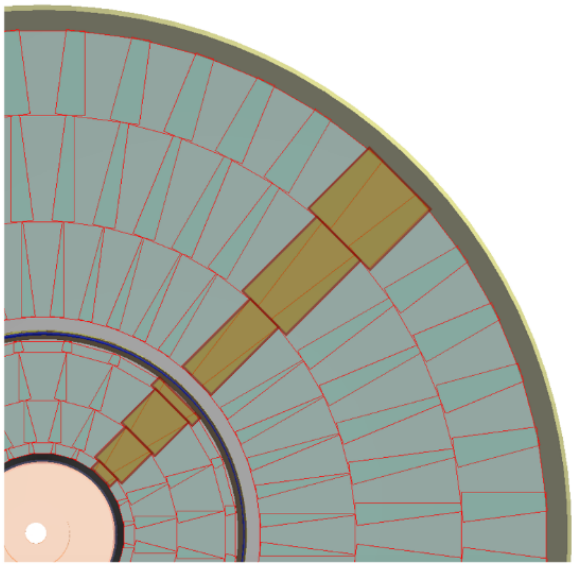

Figure 113: Transverse view of the sixth inner disk and the third outer tracker disk.

- fourteen endcaps, seven for each side of the interaction point. Each of them is composed by 26 modules. They are set along z axis, between 52.4 cm and 229 cm from the interaction point. The first endcap is composed by pixel sensors of size 25×25 μm^2, while all other endcaps by strips of size 50 μm $\times$1 mm. Strips are oriented along the radius of the disk. All sensors have thickness 100 μm. The inner tracker and outer tracker endcaps are composed by radial modules composed by rectangular pads. In Figure 113 is presented a transverse view of the sixth inner disk and the third outer tracker disk [127]. Pads are highlighted in brown. The overlay between petals is visible as darker, smaller wedges. Their outermost radius is around 43 cm for the endcap pair closer to the interaction point, and 55 cm for the other endcaps.

In the outermost part is located the Outer Tracker. All the subdetectors are composed by silicon

strips of size $50\mu m \times 10mm$. In particular in the Outer tracker there are:

- three barrel layers whose radius goes from 82 cm to 150 cm. They are composed by strips of size 50 μm $\times$10 mm and thickness 100 μm. As in the inner tracker, the strips are oriented with the long side parallel to the beam axis.
- eight endcaps, four for each side of the interaction point. Their positions along the z direction goes from 131 cm to 219 cm from the interaction point and the outermost radius of the annulus is around 143 cm. They are composed by 48 modules with sensor strips of size 50 μm $\times$10 mm and thickness 50 μm. As in the inner tracker, strips are oriented along the radius of the disk.

The timing resolution assumed for the inner and the outer tracker sensors is assumed to be 60 ps.

7.5.2 Calorimetry

The calorimeter detector is divided in the electromagnetic and the hadronic calorimeter part. Figure 114 shows both systems: the electromagnetic calorimeter in yellow and the hadronic calorimeter in magenta.

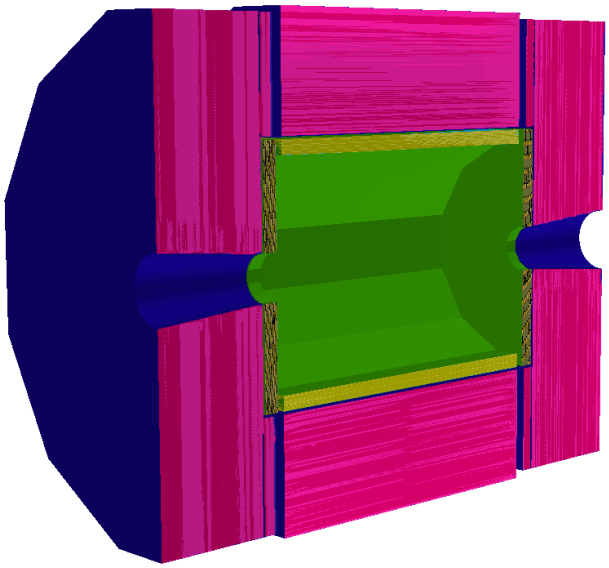

Figure 114: Longitudinal section of the calorimeter system: the electromagnetic calorimeter is in yellow, the hadronic calorimeter is in purple.

7.5.3 Electromagnetic and hadronic calorimeter

The electromagnetic calorimeter (ECAL) consists of a dodecagonal barrel and two endcaps systems, located inside the hadronic calorimeter. It is composed by 40 layer of interlaced layers of Tungsten as absorber material 1.9 mm thick and Si sensor as active material with 5×5 mm^2 silicon detector cells. Tungsten is a dense material with a large ratio of interaction length (192 g cm^{-2}) to radiation length (6.76 g cm^{-2}) [2]. A small radiation length will promote the start of the electromagnetic shower earlier in the calorimeter, while a large interaction length will reduce the fraction of hadronic showers starting in the ECAL. The choice of thin silicon layers offers an optimal spatial resolution. The total thickness of the ECAL corresponds to about 22 X_0. The ECAL total barrel and endcap depths is 20.2 cm. The barrel starts at a radius of 150 cm from the IP and has

an outermost radius of 170 cm. The ECAL encaps cover the region along the z direction between 230 up to 250 cm from the interaction point and covers the forward region up to a radius of 170 cm. The role of the hadronic calorimeter (HCAL) consists of a dodecagonal barrel and two endcaps systems. It is composed by 60 layers of interlaced layers of steel absorber 19 mm thick and plastic scintillating tiles with cell size 30×30 mm^2. Both, the endcap and the barrel HCAL, are around 7.5 λ_I deep, which brings the combined thickness of ECAL and HCAL to 8.5 λ_I. Both the barrel and the endcaps are 159 cm depth. The barrel innermost radius is 174 cm, and the outermost radius is 233 cm. The HCAL endcaps start at z position of 254 cm up to 413 cm, and has an innermost radius of 25 cm, and an outermost radius of 325 cm.

7.5.4 Solenoid and Muon detector

A large superconducting solenoid located outside the calorimeters, contains a coil of radius 382 cm, producing a nominal 3.57 T magnetic field. The solenoid is represented in cyan in Figure 115. The iron yoke return the magnetic flux, and has a magnetic field of 1.5 T pointing in the opposite direction with respect to the inner field. The iron yoke is instrumented with Resisitive Plate Chamber (RCP) sensor layers to act as a muon detector. There are seven RCP layers in the barrel and six layers in the endcaps, with 30×30 mm^2 cell size. The free space between yoke steel layers is 40 mm.

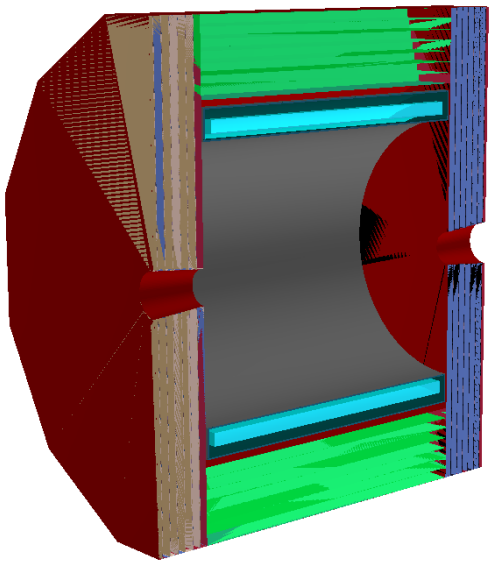

Figure 115: Longitudinal section of the solenoid and muon detectors. In cyan is represented the solenoid. In green and blue are represented the Resistive Plate Chamber layers for the muon detection.

7.6 Detector simulation software

The full simulation steps of μ^+ μ^- collisions is similar to the one described in section for the Monte Carlo simulation. At Muon Collider different software tools are used and are here listed:

1. first particles entering the detector are generated: at the Muon Collider these are the particles produced from the μ^+ μ^- collision, that are generated via Whizard or Madgraph Monte Carlo, and the BIB particles, that as mentioned in section 7.4 are generated with MARS15 (or FLUKA). The MARS15 code provides the list of BIB particles at the external surface of the detector and the nozzle surface, together with their timing, energy and momentum.

2. DD4HEP [128] is the tool that provides the geometry description of the detector and the simulation. It provides and interface with the GEANT4 toolkit and other reconstruction tools.

3. GEANT4 as mentioned in section 3.4 is used to simulate the passage of particles trough the passive and sensitive material of the detector. The output of GEANT4 are the hits (energy deposits) of particles on the tracker layers and on the calorimeter cells. The propagation and the passage through matter of BIB particles and of particles from μ^+ μ^- collisions are simulated separately and are put together (overlaid) before the digitization and full reconstruction steps. More details about the BIB simulation can be found in [129].

4. the toolkit Marlin performs the full reconstruction framework. It is based on processors dedicated to specific tasks from the digitization of the simulated hits to the event reconstruction. The digitization of the simulated hits in the pixels/strips detectors, like the tracking system, is the following: the hits position is smeared in the x and y coordinates with a gaussian with a σ set to the single point resolution of the subdetector. The timing response is smeared with the same procedure assuming a timing resolution of $\sigma = 30$ ps in the vertex detector and $\sigma = 60$ ps in the inner and the outer trackers. In the calorimeter system the digitization procedure is the following: the energies released by particles in a single cell are summed, while only the time of the first particle hitting the cell is assigned to the whole digitized hit. The same digitization approach is used for the Muon Detector.

The tools for the event simulation and reconstruction (point 2. and 3.) are provided by the is Muon Collider software framework.

7.7 Effects of BIB on Muon Collider detector and mitigation strategies

The main effects of BIB particles in the tracker layers is to produce a large number of hits in particular in the innermost layer, where BIB particles release ~ 500000 hits in each bunch crossing, corresponding up to more than 1000 hits/cm^2. Figure 116 shows in blue the average hits density on the layers of the tracking detector per each bunch crossing, that arrive at the detector layers within a time window of [-360,480] ps with respect to the bunch crossing.

In order to reduce the average number of hits, the difference in the arrival time to the detector layers between BIB and prompt particles can be exploited. In Figure 117 the arrival time to the tracker layers is compared between BIB particles and muons generated at the IP. The hits of both samples are smeared for the two timing resolutions assumed for the vertex detector (filled red distribution for the muons and empty red for BIB) and inner or outer trackers (filled gray distribution for the muons and empty black for BIB). The arrival time in the x axis is defined as:

$$T_{arrival} = T_{hit} - \frac{|\vec{r}_{hit}|}{c} \tag{7.1}$$

where T_{hit} is the arrival time of the particles that hits the detector in the position defined by the $|\vec{r}_{hit}|$, $\frac{|\vec{r}_{hit}|}{c}$ is the time of flight of a photon. By applying a timing window of $-3\sigma_t$ and $+5\sigma_t$ a reduction by a factor three on the number of BIB particle hits can be achieved. The average hit density on the layers after this time requirement is shown in yellow in Figure 116. The granularity and timing requirements of both the vertex detector and the tracker system sensors described in section 7.5.1 have been optimized via simulation studies to keep the BIB hit occupancy under the 1% level. Despite the reduction of the occupancy on the tracking layers, a large number of hits is still present in the silicon layers. Tracks reconstruction within the detector acceptance is then highly computational demanding, as will be explained in the next section 8. Another strategy that has been used in the event reconstruction of this thesis exploits the fact that BIB charged particles,

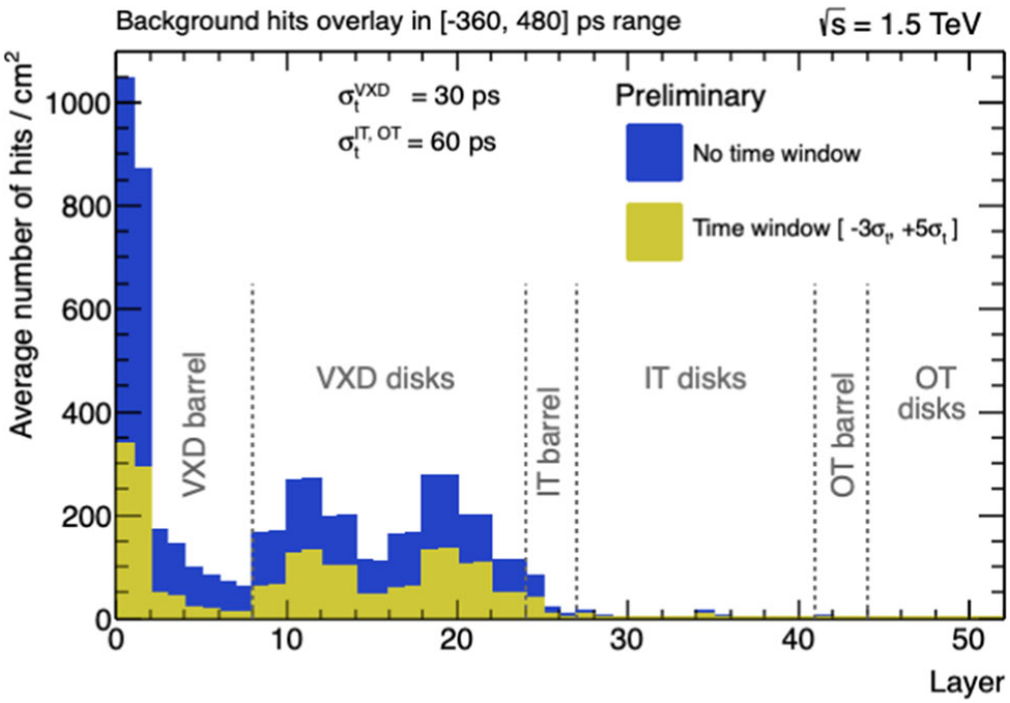

Figure 116: Average hit density per bunch crossing in the tracker as a function of the detector layer [37].

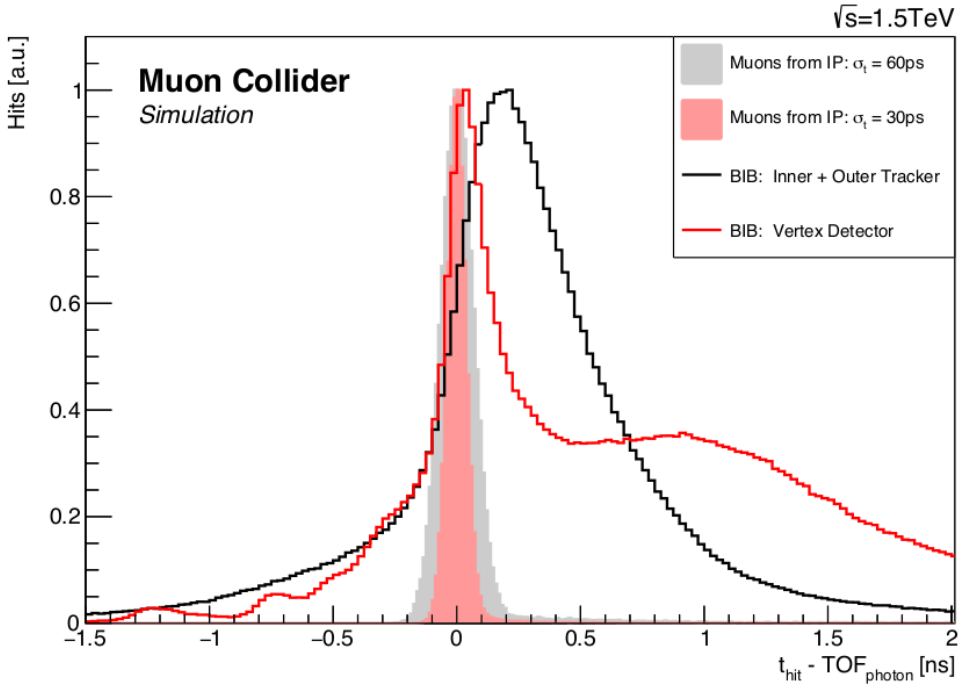

Figure 117: Average hit density per bunch crossing in the tracker as a function of the detector layer [130].

like electrons, have a very low momentum and are not prompt particles, since they mostly enter the detector from the nozzles surfaces. As mentioned in section 7.5.1, the vertex detector is composed by double-layers (DL). BIB electrons are very likely to either stop in the first layer of the doublet,

or to cross the two layers at shallow angles, creating doublets not aligned to the IP. A filter on the hits is then applied to DL of the vertex detector, as shown in Figure 118: the blue black lines represent the double layers of pixel sensors that are crossed by signal particles (green) and BIB (red) particles. Due to the shallow crossing angle and more displaced origin, the BIB particles have a larger angular difference that signal particles.

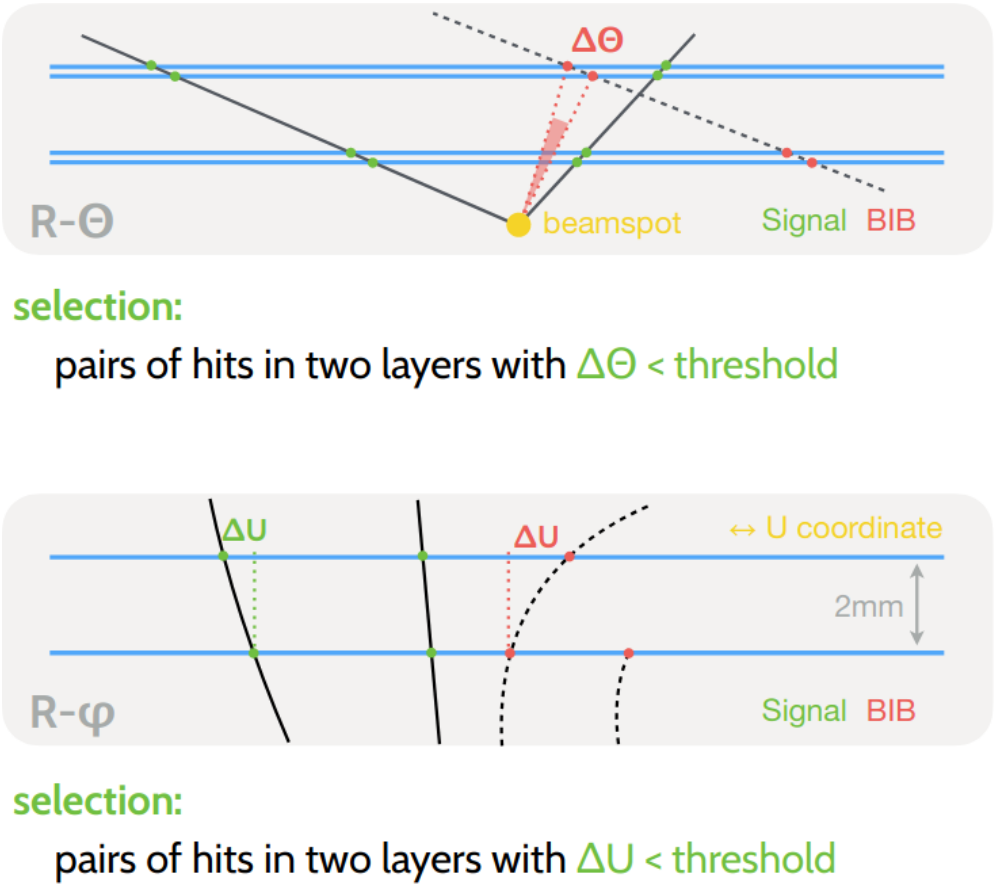

Figure 118: Illustration of the double-layer filter used to reject hits released by BIB particles in the Vertex Detector. ΔU cuts are calculated the transverse plane (upper figure), while $\Delta\theta$ is calculated in the longitudinal-radial one (bottom figure) [131] .

The DL filter implemented in the Muon Collider software, calculate the distance ΔU in the transverse plane in the second layer with respect the normal direction to the two layers passing through the position of the hit in the first layer and the polar angle $\Delta\theta$) between two hits on the subsequent layers from the interaction point. For reference, the requirements on ΔU and $\Delta\theta$ are reported in the following table 48, for the barrel layers of the Vertex Detector. Similar requirements are applied to the endcap layers, and can be found in [131]. The combination of hits rejection via the timing cuts and the DL allows a reduction of the number of hits of one order of magnitude.

Furthermore, a large number of "fake" tracks are produced by the reconstruction algorithms, built from the BIB hits combinatorial. This will require additional requirements on the tracks parameters, in order to reject part of them while keeping tracks associated to particles of the physics events coming from the $\mu^+\mu^-$ collision.

In the calorimeters, the beam-induced background generates a diffused distribution of energy. The BIB in the calorimeter region is mainly formed by photons (96%) and neutrons (4%). A flux of about 300 particles per cm^2 is present at the ECAL barrel surface, with an average photon energy of about 1.7 MeV. Figures 119 left and right show the occupancy, defined as the number of hits per

N layers	Δ U$_{max}$	Δ θ_{max}
0-1	0.55	0.3
2-3	0.55	0.2
4-5	0.5	0.15
6-7	0.4	0.12

Table 48: Double layer filter cuts on the Vertex detector Barrel layers [131].

mm^2 in the calorimeter layers, as a function of the calorimeter depth in the barrel system and of the distance from the IP in the z coordinates in the endcap system respectively. The ECAL system absorbs most of the BIB energy, resulting in a significantly lower occupancy for the HCAL system. Furthermore, BIB particles deposit most of their energy in the innermost layer of the calorimeter system. As will be explained in section 8.2, requirements on the energy deposited by particles in the calorimeter cells will be crucial for jet reconstruction.

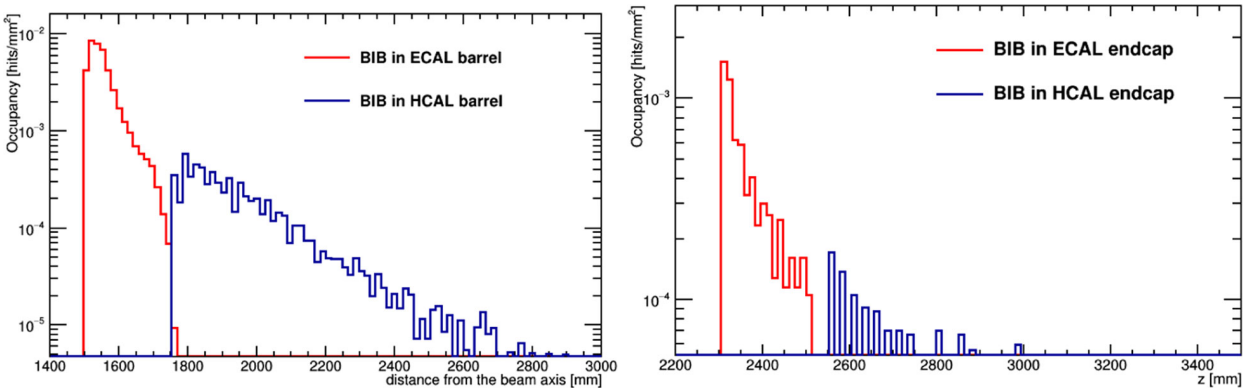

Figure 119: BIB hit occupancy in the calorimeter barrel region (left) and in the calorimeter endcap region (right) in a single bunch-crossing [37].

A similar procedure to the tracking system is applied to the calorimeter system: the arrival time to the ECAL barrel is compared between b-jets (blue) and BIB (red) in Figure 120. Also in this case, a time window read-out of $\pm$250 ps is applied to suppress the soft BIB particles. The presence of energy from BIB particles in the calorimeter system affects jet reconstruction, as will be explained in section 8.2, and requires quality selections to reject fake jets arising from the spurious combination of BIB energy deposits.

7.7.1 Preliminary studies for a 10 TeV Muon Collider

Studies for the design of a 10 TeV Muon Collider detector are ongoing. The study of the characteristics of events from $\sqrt{s}$=10 TeV collisions are crucial to determine the detector characteristics. For example, Figure 121 left shows the b-quark p_T distribution from the $H \to b\bar{b}$ decays at $\sqrt{s}$=3 TeV and $\sqrt{s}$=10 TeV, as obtained at generator level with MadGraph5. Despite the higher center-of-mass energy, it is evident that the distributions are quite similar. On the other hand, as shown in Figure 121 right Higgs bosons at $\sqrt{s}$ = 10 TeV are emitted at smaller polar angles compared to 3 TeV, therefore the forward region of the detector should be designed carefully.

As mentioned in the previous chapter, the 10 TeV Muon Collider may give access to heavy new particles, then a proper detector may be designed in order to reconstruct them. For example, Figure 122 shows the distributions of leptons and jets p_T from the decays of a Z' with a mass of 9.5 TeV produced at $\sqrt{s}$=10 TeV muons collisions. If on one side the SM background are expected to have a very lower impact at 10 TeV with respect to lower center of mass energies, this distributions exhibit

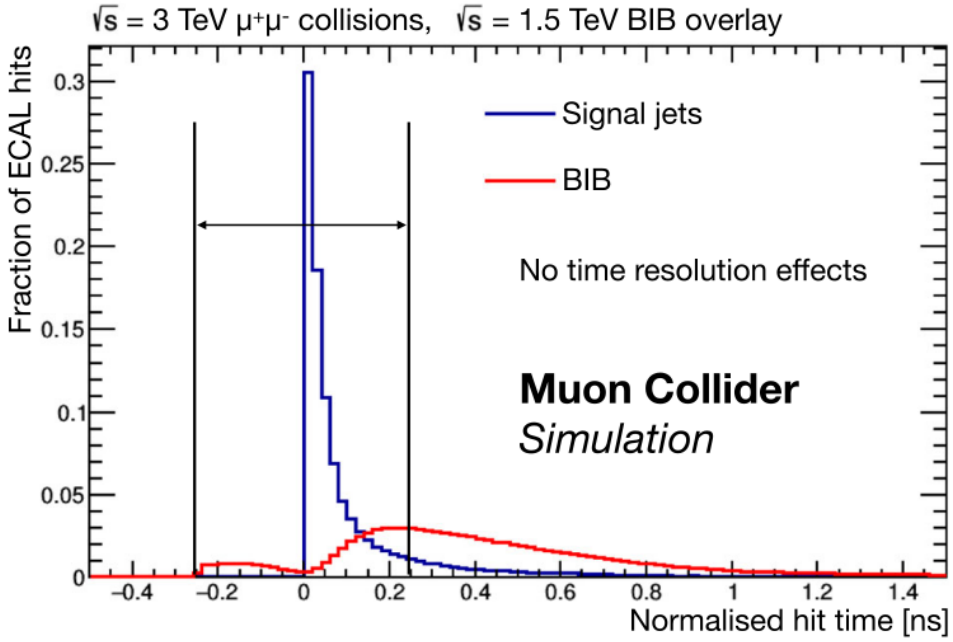

Figure 120: Comparison of signal and BIB hit properties in the ECAL Barrel: TOF-corrected time [37].

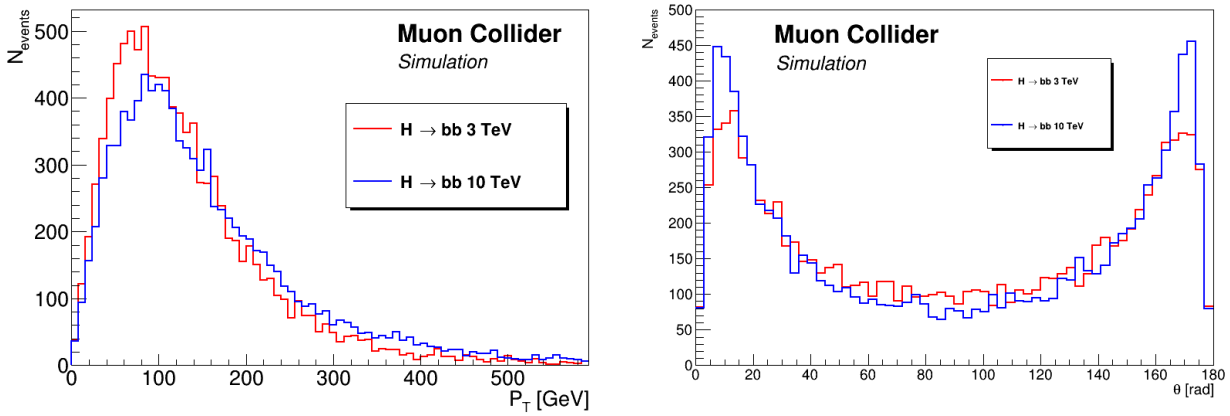

Figure 121: Left:b-quark p_T distributions from the $H \to b\bar{b}$ decays at $\sqrt{s}$= 3 and 10 TeV, as obtained at generator level. Right: angle between the Higgs and the beam axis at Muon Collider distribution at $\sqrt{s}$= 3 and 10 TeV, as obtained at generator level.

a peak at $p_T \sim 5$ TeV, and a long tail at lower p_T, then it is important to design the detector to have a good p_T resolution also in these p_T regimes.

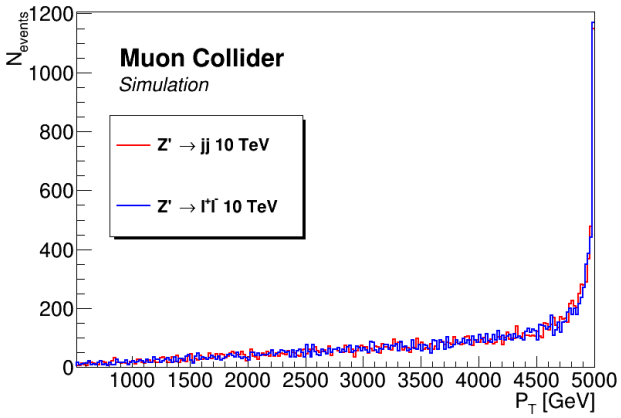

Figure 122: Distribution of leptons and jets p_T from the decay of a Z' produced at $\sqrt{s}$= 10 TeV with a mass of 9.5 TeV, as obtained at generator level.

8 Event reconstruction at Muon Collider

In the next chapter, the analysis will be focues in the search of double Higgs events, where the Higgs decays into $b\bar{b}$ pairs at 3 TeV center of mass energy. In this chapters, a description of the algorithms used to reconstruct the four final state jets produced by the Higgs decay and the physics background events are presented. Furthermore, in section 8.4.2, studies on secondary vertices reconstruction at muon collider that have been done in the context of this thesis work, are shown. The strategy to reduce the BIB effects on tracks, jets and secondary vertices reconstruction level will be also discussed. The jets reconstruction starts from the tracks reconstruction using the energy hits deposited in the tracking system, then a particle flow algorithm, whose concept has already been introduced in section 4.4, is used to match the information obtained from the tracking system to the energy deposited in the calorimeter system to create reconstructed particles, and then a clustering algorithm is used to reconstruct jets. The jet flavour identification at a muon collider is performed by reconstructing vertex displaced from the primary vertex. A dedicated algorithm, optimized to maximize the b-jet tagging efficiency while minimize the tag of c-jet, that is a source of background for this analysis, and the mis-tag arising from light and fake jets due to BIB is presented.

8.1 Track reconstruction

Hit surviving the time window requirement mentioned in section 7.6 are given as input to the track reconstruction algorithm. Two approaches have been studied for track reconstruction:

- a tracking algorithm that makes use of Combinatorial Kalman Filter, referred here as CKF algorithm, developed for track reconstruction in busy environment, like that of pp collisions. This algorithm has been used to reconstruct jets, as will be explained in section 8.2 .
- the second one is the Conformal Tracking (CT) [132]. This algorithm was developed to reconstruct tracks at the clean environment of e^+e^- colliders. In the presence of the BIB it becomes computationally high demanding, and may need long times to reconstruct single events of BIB (weeks). As will be discussed in section 8.1.2 it has been employed by selecting hits in regions of interests (for example it reconstructs secondary vertices inside jets cone) and with further tracks filtering to reduce the computing time;

8.1.1 Combinatorial Kalman Filter

The description of the Combinatorial Kalman Filter algorithm can be found in detail [133], here the main steps are summarized.

- The first step is the Track Seed Finder, that takes in input three- dimensional hits positions to form all possible track candidates. The seeds are formed from hit triplets in the four layers of the Vertex Detector, by considering only hits in the outer half of the doublet layer. Selection criteria are applied in order to minimize the number of seeds that do not correspond to a particle;
- A Combinatorial Kalman filter algorithm is used to estimate the track parameters by using an iterative process incorporating individual hits assigned to the track and fitting it, while, at the same time, perform the search of new hits belonging to the track. Tracks are propagated to the outer layers with respect to the Vertex Detector and the parameters are re-calculated by adding the hit found in the subsequent layer within a certain search window width.

The output of the tracks fit are the tracks parameters, namely:

1. D_0 is the distance between the track helix and the interaction point in the $x - y$ plane
2. Φ is the azimuthal angle of the reference point with respect to the center of the helix
3. Ω the signed curvature of the track, defined as $\Omega = p_T / \ cBQ$, where B is the magnetic field and Q is the charge of the particle
4. Z_0 the distance between the helix and the reference point in the z direction
5. $tan(\lambda)$ that is the dip angle, i.e., the angle of the helix to the x-y plane

Around 100000 fake tracks are built per bunch crossing, but most of them have low p_T, < 3 GeV and a small number of hits associated to the track N_{hits}, < 6. The performance of this algorithm is assessed in the presence of BIB using a set of prompt muons with fixed momentum of 10 GeV and uniform distribution in *theta*. The reconstruction efficiency is shown in Figures 123 as a function of the muon p_T (left) and as a function of the muon θ (right). The efficiency is found to be close to 100% with and without BIB for $p_T > 3$ GeV and in the central angular region of the detector. The drop in the efficiency is due to the larger presence of BIB tracks at low p_T and the closeness to the two nozzles at very low and very large θ.

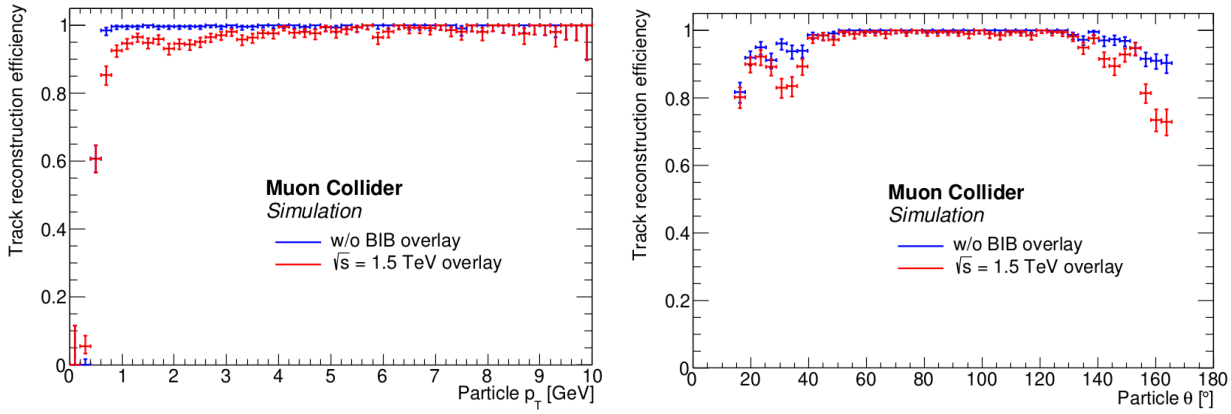

Figure 123: Track reconstruction efficiency for events containing a single muon with (red) and without (blue) the BIB overlay as a function of the true muon p_T (left) and θ (right) [37].

8.1.2 Conformal Tracking

The description of the Conformal Tracking algorithm can be found in detail [132], here the main steps are summarized. It is divided into the following steps: first a conformal algorithm performs pattern recognition to identify hits that may belong to a track, then track fitting is done by using a Kalman method, at last good tracks selection is performed on built tracks.

- The conformal algorithm maps the hits coordinates in the Euclidean space into the conformal space (CS) coordinate. The advantage is that prompt tracks bent by the solenoid magnetic field of the detector are straight lines in the CS. In order to increase the efficiency in reconstructing displaced tracks, that are not straight lines in the CS, it is combined with cellular automaton. The cellular automaton proceeds as follows:

  1. It builds cellular tracks candidates, taking as input the hits belonging to the innermost layer of the vertex detector (*seed hits* and connect the *seed hits* with the all the possible hits in the subsequent layer that lie within a certain polar angle in the CS and a length smaller than a certain length. The track segments are then prolonged in the radial

direction of the CS to search for new hits than can be connected to it. A minimum number of hits is required to each track candidates and a preliminary fit is performed to reject tracks with χ^2 larger than a certain threshold.

2. The tracks surviving the fit are then extended according to the particle transverse momentum.

These two steps are applied iteratively: first tracks segments are built with the hits in the subsequent layers of the vertex barrel, then tracks are extended using the hits in the vertex endcap, then other tracks segments are built by using the remaining hits in the vertex barrel+endcap detectors. As final steps, all cellular track candidates are extended to the tracker detectors and remaining hits of the entire tracking system are used to reconstruct remaining tracks, most of which are generated by displaced particles.

- The outputs of the pattern recognition steps are all tracks candidates. A Kalman filter algorithm is used to perform track fit

- A selection procedure is then applied to ensure good quality tracks based on the length of the tracks and on the χ^2. Furthermore, at this level, minimum number of three hits is required.

  The distribution of these parameters will be studied in section 8.4.2 in order to reject part of BIB tracks.

8.2 Jet reconstruction

In this chapter techniques for jet reconstruction [37] used in the data analysis of Section 9 are presented. The algorithm for jet reconstruction has been optimized to reconstruct jets in the presence of BIB and is here described. It consists of the following steps:

- tracks are reconstructed using the CKF algorithm described in sect. 8.1.1. In order to reduce fake tracks they are required to have at least three hits in the Vertex Detector and at least two hits in the Inner Tracker;

- Calorimeter hits are filtered with the timing window requirement described in section 7.6. A calorimeter hits energy threshold of 2 MeV is then applied to calorimeter hits of both ECAL and HCAL, this reduced the average number of ECAL barrel hits from 1.5 M to less than 10^4. The effect of this high energy threshold and its optimization will be discussed in the jet reconstruction section 8.3.1;

- tracks and calorimeter hits are used as inputs in the PandoraPFA algorithm [134] to obtain reconstructed particles;

- the reconstructed particles are clustered into jets with the k_T algorithm [135].

- E-recombination scheme gives the scheme to compute the final jet four momentum;

- the energy of the jets passing the BIB-removal selection are calibrated.

8.2.1 PandoraPFA algorithm

The concept of Particle Flow approach have been described in section 4.4. Digitized calorimeter hits and charged particles tracks are given as input to the PandoraPFA, the Particle Flow algorithm used to reconstruct events by the muon collider software, that produces as output reconstructed particles known as particle flow objects. The PandoraPFA algorithm is described in detail in [136], while a summary can be found in [137]. It performs calorimeter clustering and the Particle Flow

reconstruction in eight main stages (default values used in the algorithms can be found in the repository [138]):

- Track Selection/Topology: tracks are projected into the front face of the electromagnetic calorimeter. Tracks topologies such as kinks (charged particle decaying into a single charged particle plus neutral particles) or decays of neutral particles into a pair of charged particles are identified.
- Calorimeter Hits Selection and Ordering: a energy cut is applied to the calorimeter the list of digitized calorimeter hits and their energy is corrected in order to take into account the invisible energy not detected in the hadronic calorimeter.
- Clustering: hits are grouped into clusters by using a cone-based clustering algorithm, exploiting the fact that the direction of the released calorimeter hits is almost the same of the original particle. If seed calorimeter hits are not associated to projection of any track, the initial cone direction is taken from the interaction point to the calorimeter hit. A cone with a opening angle is formed around the direction of the seed and hits that lie within cone defined by existing cluster, and are sufficiently close, are added to the cluster. Unmatched hits, are used to seed new clusters.
- Topological Cluster Merging: neutral clusters that are not already associated to tracks are merged with clusters that have an associated tracks (charged clusters).
- Statistical Re-clustering: if there are discrepancy between the cluster energy and any associated track momenta clustering algorithms are run again, with different configuration parameters.
- Photon Recovery and identification: clusters of calorimeter hits consistent with photons are identified.
- Fragment Removal: remaining neutral clusters that are not identified as photons are merged to charged particle hadronic showers
- Formation of particle Flow objects: reconstructed particle are created. The energy for charged particles is calculated from the track momenta, while neutral particle energies are taken from calorimeter measurements.

8.3 Jet clustering with the k_T algorithm

Particles selected by the Particle Flow are used as inputs in the jet clustering algorithm. As mentioned in 4.6 several jet clustering algorithms exist and have been there described. The one used in the jet reconstruction at muon collider is the "k_T" [139] algorithm with radius ΔR=0.5, and is provided by the FastJet package [135] within the Muon Collider framework.

8.3.1 Jet reconstruction efficiency and resolution

The jet performance [37] have been evaluated on simulated sample of $b\bar{b}$, $c\bar{c}$ and $q\bar{q}$ dijets (where q stand for a light quark u,d or s) These samples have been generated with an almost uniform dijet $p_T$distribution from 20 to 200 GeV. The jet performance have been calculated by matching the reconstructed jets with true level b (c or light)-jets by requiring ΔR<0.5. The BIB effects on jet reconstruction is to produce a large number of "fake jets", on average 13 fake jets per event. However, a requirement on the number of tracks associated to the jet (N_{trk}>0) is applied and is

found to reduce the rate of fake jets well below 1%, while keeping most (90-95%) of the reconstructed jets. The jet reconstruction efficiency has been calculated for the three classes of jets as:

$$\epsilon_{jet} = \frac{N_{jet,reco}}{N_{jet,tot}} \tag{8.1}$$

where $N_{jet,reco}$ is the number of reconstructed jets matched with a true level b (c or light)-jets, while $N_{jet,tot}$ is the total number of true level b (c or light)-jets. The jet reconstruction efficiency in the central region is shown in Figure 124 left for b-jets(c and light)-jets in black(red and blue). The reconstruction efficiency is found to be between 85% at low p_T and 100% at high p_T. The dependency of the selection efficiency on the jet polar angle θ is shown in Figure 124 right. The efficiency is around 90% in the central region while a significant drop is observed for $\theta < 0.5$. This effect is mainly due to the track requirement, since many jets without tracks are found in the forward region. This is due to the fact that many tracks reconstructed by the CKM algorithm in the forward region are rejected by applying the selection requirements on the number of hits.

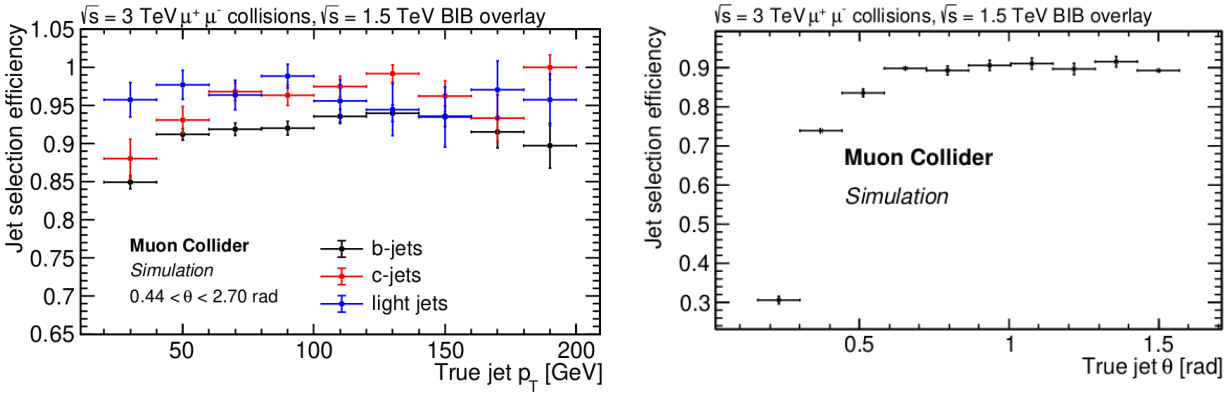

Figure 124: Left: jet selection efficiency as a function of the true jet p_T for b-jets,c-jets and light jets in the detector central region $|\eta|$<1.5, colored in black, red and blue respectively. The differences between the jet flavors are mainly due to different jet $|\eta|$ distributions [37]. Right: Efficiency of jet selection as a function of truth-level jet θ.

The Jet Energy Correction have been evaluated to recover energy lost by reconstruction inefficiencies, non sensitive material and BIB contamination. This correction has been determined by comparing the reconstructed jet p_T with the corresponding truth-level jet p_T. The correction have been evaluated in five intervals in η between 0 and 2.5 and 19 intervals of p_T between 10 and 200 GeV. For each interval the average and the standard deviation of the true jet p_T distribution is calculated. The jet energy correction function are then found by fitting the average truth-level jet p_T as a function of the reconstructed jet p_T. Figure 125 show the jet correction function for θ in the forward (left) and central regions (right). The jet reconstruction performance in terms of jet resolution have then been calculated after the JEC correction. The jet p_T resolution as a function of the true jet p_T is shown in Figure 126 left, and goes from 35% for jet p_T around 20 GeV, to 20% for high jets p_T. The dijet invariant mass reconstruction obtained with this procedure have been evaluated on $\mu^+\mu^- \to H(\to b\bar{b}) + X$ and $\mu^+\mu^- \to Z(\to b\bar{b}) + X$ at $\sqrt{s}$ =3 TeV , by requiring jets to have p_T>40 GeV and 0.44<θ<2.7. Figure 126 right shows the fitted shapes of the invariant masses, having relative width of 27% and 29% for the Higgs and Z respectively.

The energy threshold set at 2 MeV value is high, and it is the compromise between needed computing time and jet reconstruction performance. Its effect in the jet recontruction performance can be seen in Figure 127 left. In the Figure, the invariant mass of the two b-jet is shown when reconstructed without the BIB overlayed after having applied a threshold of 2 MeV (red) or 200

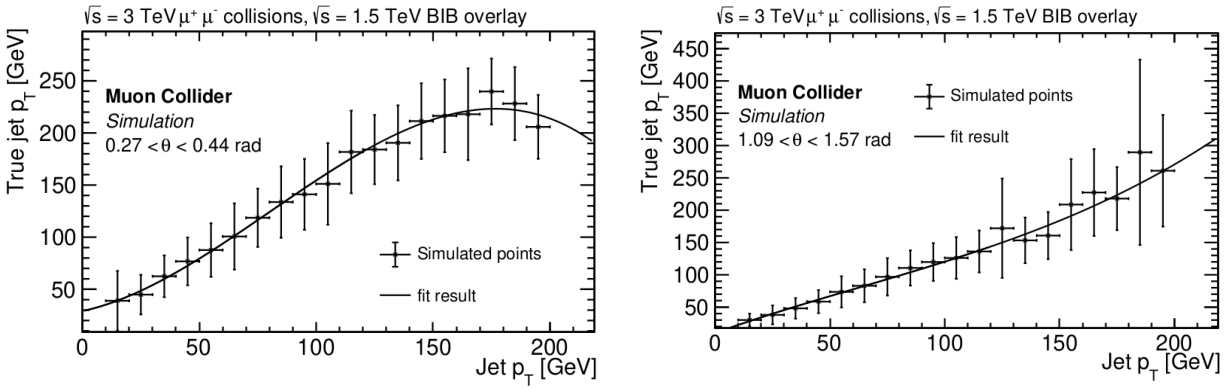

Figure 125: Jet energy correction function used for jet momentum corrections, for $0.27< \theta<0.44$ (left) and $1.09< \theta<1.57$ (right) [37].

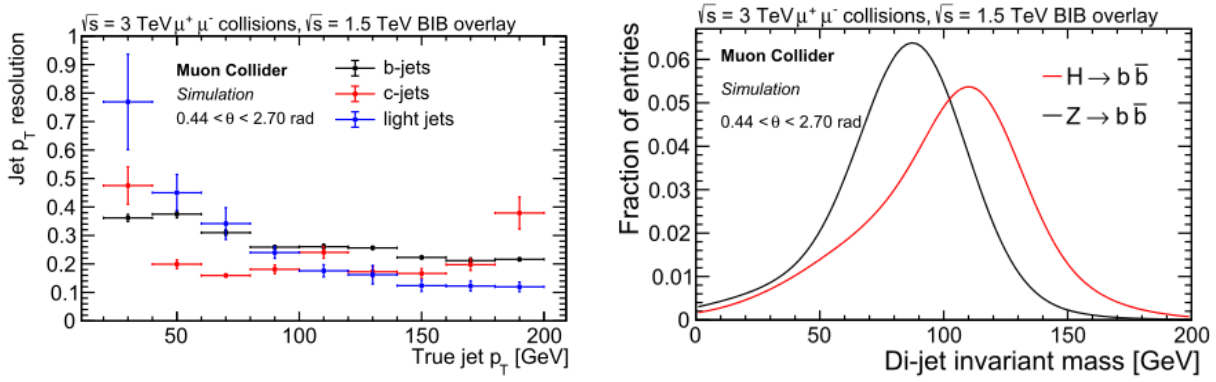

Figure 126: Left: Jet p_T resolution as a function of the true jet p_T for b-jets, c-jets and light jets in the detector central region $0.44<\theta<2.7$, colored in black, red and blue respectively. Right: Fitted di-jet invariant mass distributions for $H \to b\bar{b}$ (red) and $Z \to b\bar{b}$ (black) [37].

keV (black) to the calorimeter cells. The high energy threshold worsen the jet energy determination and degrades the invariant mass resolution. In the presence of BIB, the effect is even worse, by reducing the energy threshold the jet energy resolution is degraded and the tails in the invariant mass distribution increase due to the additional BIB hits clustered in the original jets or forming new jets. This can be seen in Figure 127 right, where the invariant mass of the two b-jet is reconstructed with the BIB after having applied a threshold of 2 MeV (black) or 1 MeV (red) to the calorimeter cells. The solution to this problem is the development of an optimized algorithm for the jet reconstruction starting from hits selection where an energy threshold as a function of the calorimeter depth should be used.

8.4 Primary and secondary vertex tagging algorithm

The CT algorithm have been used to reconstruct tracks inside jets, as described in [131]. The identification of jets arising from hadronization of b and c hadrons is evaluated using the secondary vertex tagging as discriminator.

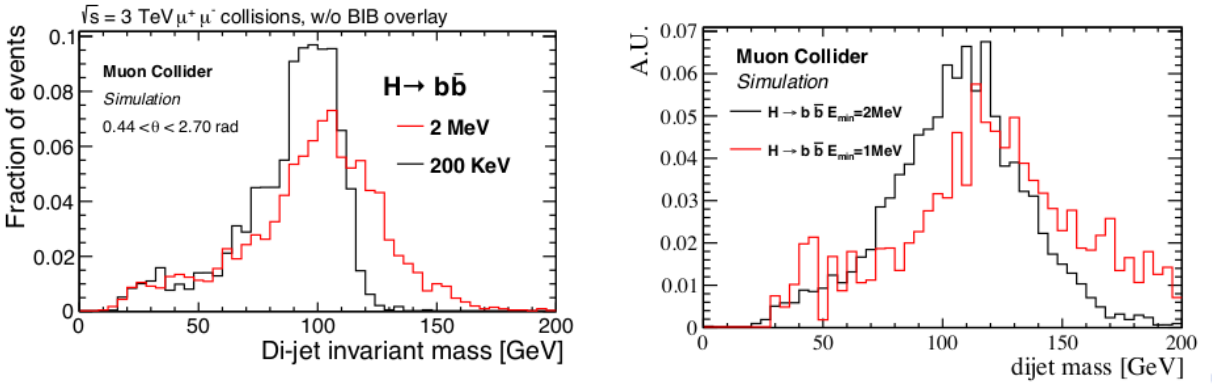

Figure 127: Left: $H \to b\bar{b}$ dijet invariant mass, reconstructed without the presence of the BIB and with 2 MeV (red) and 200 keV (black) calorimeter hit energy thresholds [37]. Right: $H \to b\bar{b}$ dijet invariant mass reconstructed with the BIB after having applied a threshold of 2 MeV (black) or 1 MeV (red) to the calorimeter cells [140].

8.4.1 Vertex algorithm description

In order to reduce the amount of tracks due to BIB, requirements are applied to tracks given in input to the algorithm for primary and secondary vertices reconstruction. A vertex fitter algorithm is used both for the PV and SV position calculation, and proceed by minimizing the χ^2_{PV} value, which is constructed by adding to the χ^2 the contribution of every track to the fitted vertex point. In the following, the χ^2_{PV} is the minimized χ^2 value of the primary vertex, while χ^2_{track} is the individual contribution to the χ^2_{PV} of individual tracks. The algorithm used is the one included in the Muon Collider framwork, the main steps of the algorithm are here summarized:

- The vertex fitter algorithm is used to calculate the primary vertex. Tracks with $D_0 < 0.1$ mm and $Z_0 < 0.1$ mm with respect the interaction point, set at the origin of the coordinate system defined in section 7.5, and that have at least four hits in the vertex detector are used to calculate the vertex χ^2_{PV}. A $\chi^2_{threshold}$= 10 is set, and the track with the highest χ^2_{track} is then removed from the set of tracks if it exceeds the $\chi^2_{threshold}$. The vertex fitter algorithm is then repeated until no tracks with χ^2_{track}>$\chi^2_{threshold}$ are found in the track sample;

- Tracks not used to reconstruct the primary vertex are used as input to the the secondary vertex fitter algorithm. A set of requirements has been applied to the input tracks in order to reduce the number of tracks due to BIB. $b\bar{b}$ and $c\bar{c}$ di-jet samples with the BIB overlay have been used to study the characteristics of tracks coming from b and c hadrons decay. Tracks associated to truth particles coming from the decay of a b and c hadron are matched to particles inside the jet by minimizing the χ^2 built between the Monte Carlo and the reconstructed tracks parameters, as described in [131]. Their characteristics have been compared to all other tracks. Most of BIB tracks are found to have minor number of hits in the vertex detector with respect to tracks coming from b and c hadrons decay as can be seen in Figure 128 left, where the distribution of the total number of hits in the vertex detector of BIB tracks are shown in red and the same distribution for tracks matched at Monte Carlo level with particles generated by b or c hadrons decays is shown in blue and black. A minimum number of 4 hits in the vertex detector is then required to reject most of BIB tracks. Further selections on the track p_T, the maximum track z_0 and d_0 and the z_0 and d_0 errors are applied in order to further reduce the amount of BIB tracks. Figure 128 right shows for example the relative transverse momentum distribution of BIB tracks, compared to the one of particles coming from b or c hadrons decays. As can be seen, BIB tracks have a very low momentum compared

to tracks coming from b or c hadrons. Then, a requirement on $p_T > 0.8$ GeV reject $\sim 80\%$ of BIB tracks.

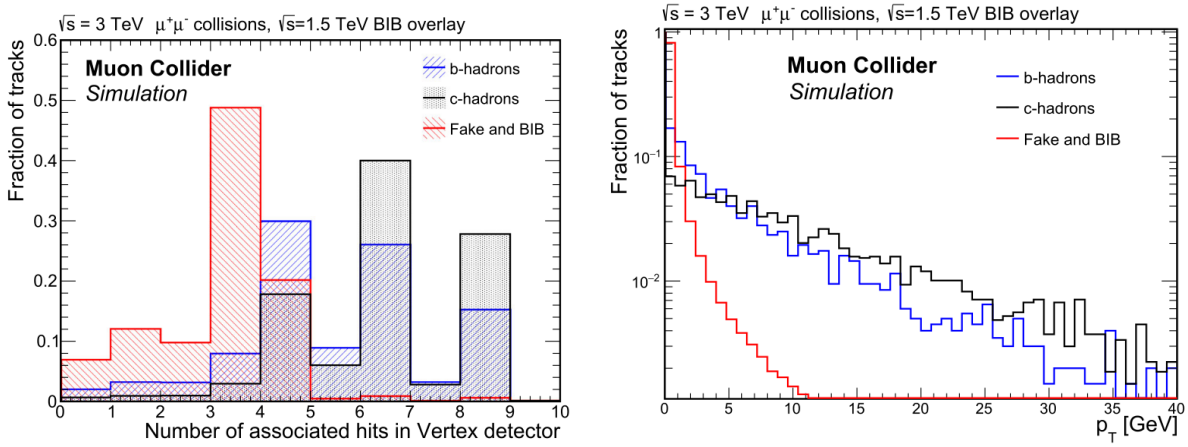

Figure 128: Left: number of hits in the vertex detector of BIB tracks (BIB) and of tracks matched with Monte Carlo truth particles coming from b or c hadrons (blue and black respectively). Right: transverse momentum distribution of tracks coming from b (blue), c (black) hadrons and BIB (red) tracks. The distributions are normalized to unit area [37].

- The tracks passing the requirements are used to build two-tracks vertex candidates, that must have: an invariant mass below 10 GeV and must be smaller than the energy of each track, the position of the primary vertex must lie in the same side of the sum of the tracks momenta and the χ^2_{track} of all tracks with respect to the secondary vertex position must be below 5. The track pairs are also required to not be compatible with coming from the decay of neutral long lived particles. Additional tracks are added to the two-tracks vertices if they satisfy the above requirements. Tracks associated to more than one SV are associated to the vertex with lowest χ^2 and removed from other vertices.

8.4.2 SV tagging performance

The studies presented in this section concern the characterization of primary and secondary vertices reconstruction algorithm that was implemented in the Muon Collider framework, the optimization of the selection requirements to distinguish BIB tracks from displaced tracks from $b-$ and $c-$ hadrons, and the secondary vertices reconstruction efficiency determination. The SV tagging performance have been evaluated on simulated $b\bar{b}$, $c\bar{c}$ and light $q\bar{q}$ partons samples generated with PYTHIA 8 for $\mu^+\mu^-$ collisions at 3 TeV including the BIB.

The characteristics of secondary vertices inside reconstructed jets have been studied in order to maximize the tag efficiency of b quarks while minimize the c quarks and the mis-tag of light quarks contribution. Figure 129 shows the distribution of the secondary vertices proper lifetime (τ). As it can be seen, secondary vertices of b jets have a larger proper lifetime, then a selection requirement of $\tau >$ 0.2 ns rejects $\sim$ 30 % of both c and q jets, while retaining the 90% of b-jets. The b tagging efficiency has been calculated as ϵ_b:

$$\epsilon_b = \frac{N_b^{tag}}{N_b^{tot}} \tag{8.2}$$

where N_b^{tag} is the number of tagged b jets surviving the requirement on the SV τ, while N_b^{tot} is the total number of reconstructed b jets. As mentioned in section 8.3.1 fake jets may be reconstructed due to the presence of the BIB, and a fake SV is likely to be reconstructed into such jets contributing

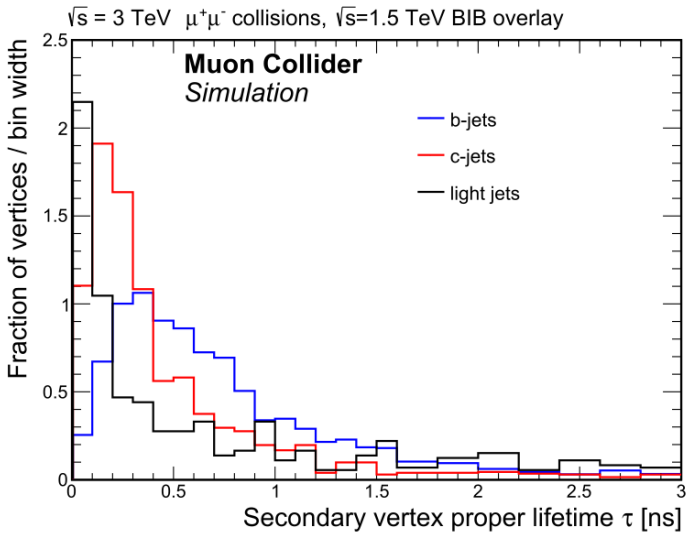

Figure 129: Distribution of the secondary vertex proper lifetime for b (blue), c (black) and light+ fake tagged jets. Distributions are normalised to the unit area.

to a possible significant rate of mis-tagged jets. This mis-tag rate is included in the light jets mis-tag. The $c-+$(light+fake) jet mis-tag is defined as $\epsilon_{c(q)}$ (where q stands for u, d, s quarks and fake jets):

$$\epsilon_{c(q)} = \frac{N^{tag}_{c(q)}}{N^{tot}_{c(q)}} \tag{8.3}$$

where $N^{tag}_{c(q)}$ is the number of tagged $c-+$(light+fake) jets surviving the requirement on the SV τ, while $N^{tot}_{c(q)}$ is the total number of reconstructed $c-+$(light+fake) jets. The effect of the Double Layer filter on the secondary vertex efficiency has been calculated in [131]. Events of $b\bar{b}$, $c\bar{c}$ and light $q\bar{q}$ partons samples generated with PYTHIA 8 for $\mu^+\mu^-$ collisions at 3 TeV without the BIB, with and without the DL filter. The ratio of the number of tagged jets without any double layer filter and the number of tagged jets passing the double layer filter is determined as a function of the jet p_T and θ and is used to correct the tagging efficiencies. Figure 130 shows the b-tagging efficiency, evaluated on b-jets, as a function of the jet p_T (left) and θ (right). It is around 50% at low p_T and increases up to 70% at high p_T and is higher in the central region with respect to the forward region.

Figure 131 shows the c-jets mis-identification efficiency, evaluated on c-jets, as a function of the jet p_T (left) and θ (right). It is around 20 % in the whole p_T spectra and is higher in the central region with respect to the forward region.

Figure 132 shows the q-jets mis-identification efficiency, evaluated on q-jets and fake, as a function of the jet p_T (left) and θ (right). It is below 1 % for $p_T <$ 50 GeV, and increases to 5% at higher p_T spectra. The light and fake jets mis-identification is higher in the forward region of the detector as expected.

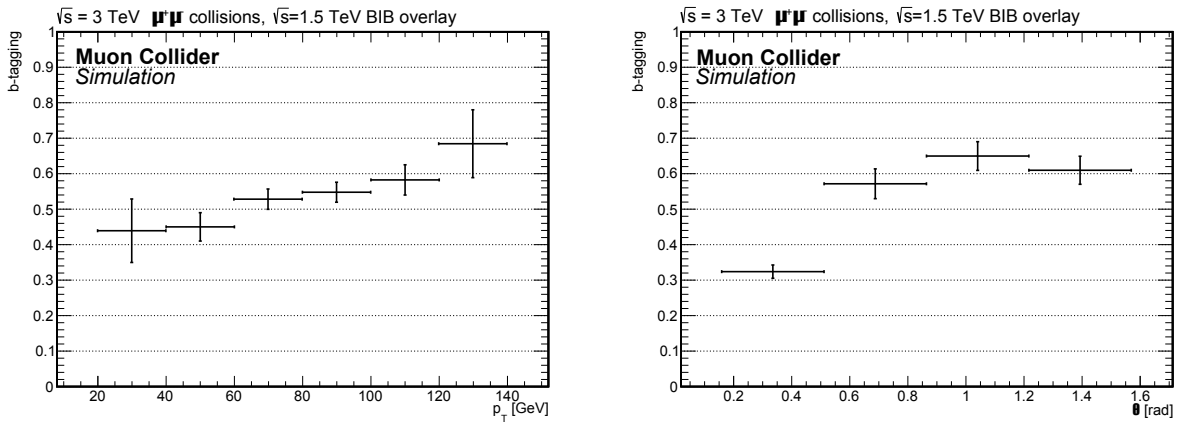

Figure 130: Efficiency of the b-tagging algorithm as a function of the jet p_T (left) and θ (right), evaluated on $b\bar{b}$ dijet events.

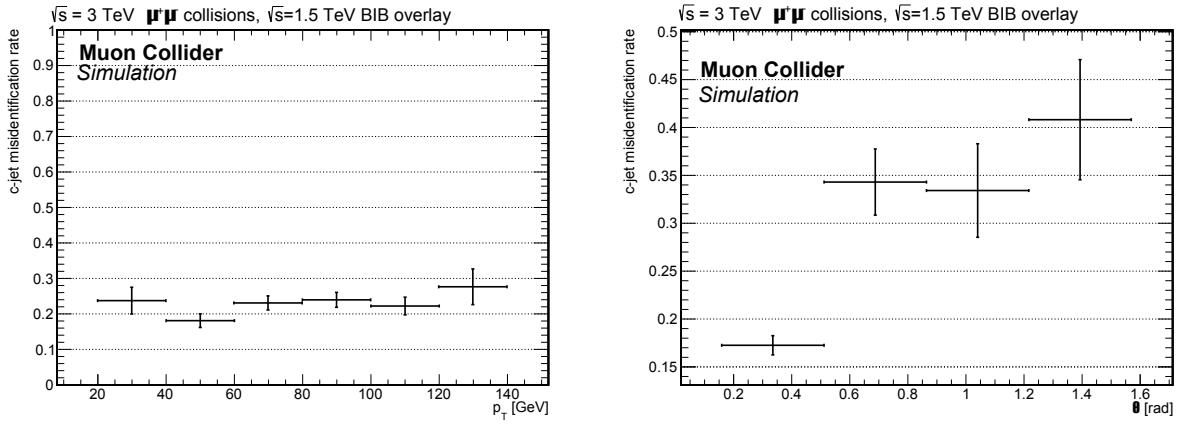

Figure 131: Efficiency of the b-tagging algorithm as a function of the jet p_T (left) and θ (right), evaluated on $c\bar{c}$ di-jet events.

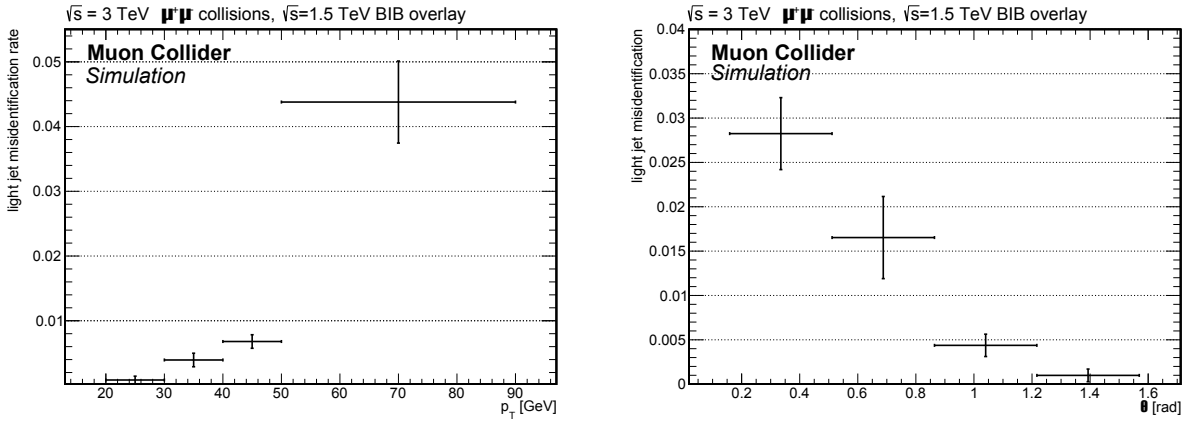

Figure 132: Efficiency of the b-tagging algorithm as a function of the jet p_T (left) and θ (right) evaluated on q-jets including fake jets.

9 Double Higgs production and Higgs self-coupling at muon collider

The purpose of this chapter is to describe the determination of the statistical uncertainty on the measurement of the double Higgs production cross section and on the trilinear Higgs self-coupling by using the Higgs decaying into $b\bar{b}$ pair at 3 TeV center of mass collisions. The chapter is organized as follows:

- in section 9.1 a review of the measurement on the $H \to b\bar{b}$ coupling is presented;
- the analysis strategy followed by the determination of the statistical uncertainty on the HH cross section is discussed in section 9.2. The signal and physics background events generation together with their reconstruction and the signal selection strategy are then presented.
- in section 9.3 the statistical uncertainty on HH cross section is calculated by using a Multivariate Analysis technique to classify signal and backgrounds events.
- in section 9.4 the statistical uncertainty on the trilinear Higgs self-coupling is calculated. First the strategy followed to extract the uncertainty is presented. The Multivariate Analysis techniques are used to classify signal and background events and to separate HH events sensitive to the trilinear coupling from all other HH events. Results assuming improvements on jet energy resolution and on jet tagging performances are also presented.

9.1 Reconstruction of Higgs to bb at 3 TeV muon collider

The statistical sensitivity on the measurement of the $H \to b\bar{b}$ cross section has been determined in [43]. The signal $\mu^+\mu^- \to XH(\to b\bar{b})$ and the physics background $\mu^+\mu^- \to Xq\bar{q})$, where $q = b, c$ (the contributions from light jets are considered negligible) have been considered, since the b-jet identification is applied to both jets in the analysis. A minimum p_T of 40 GeV and pseudo-rapidity $|\eta| < 2.5$ is required to both jets. The number of signal and background event expected in 1 ab$^-$1 of integrated luminosity are respectively, 59500 and 65400. The di-jet invariant mass distributions for signal and background are fitted with double-gaussian probability distribution function, in order to obtain the signal and background models which are used to generate pseudo-data. From the fit to pseudo-data distribution a statistical uncertainty on the $\sigma(H \to b\bar{b})$ of 0.75 % is found.

9.2 Measurement of the uncertainty on the double Higgs cross section at Muon Collider

The analysis strategy for the calculation of the uncertainty on the HH cross section is the following:

- signal and physics background events are generated and reconstructed with the detailed detector simulation including the BIB;
- selection requirements are applied to reconstructed events;
- observable quantities are given in input to a Multivariate Analysis method to separate signal from background;
- the MultiVariate Analysis distributions are used to generate pseudo-data and by fitting it, to calculate the relative uncertainty on the HH cross section.

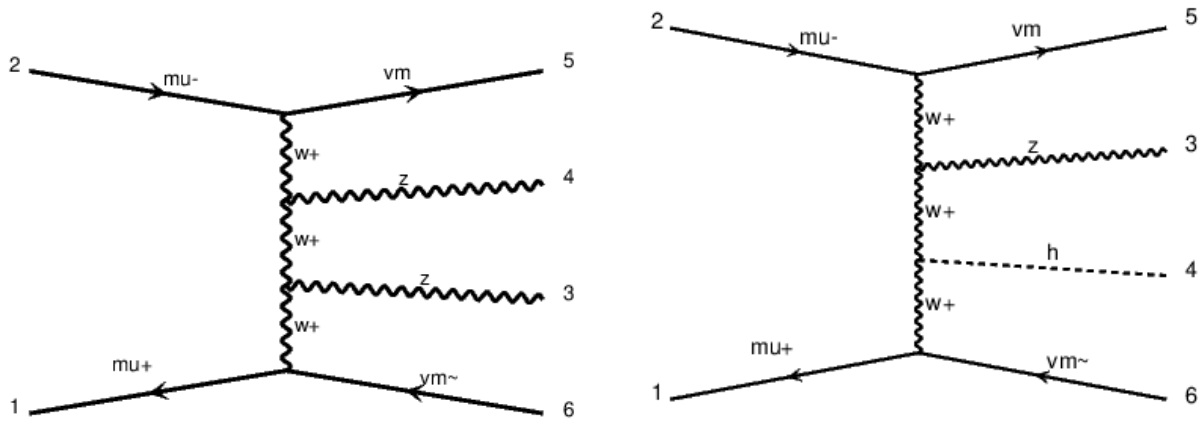

Figure 133: Example of the Feynman diagrams of the two main sources of physics background: $\mu^+\mu^- \to (Z \to q_h\bar{q_h})(Z \to q_h\bar{q_h})\nu\bar{\nu}$ (left) and $\mu^+\mu^- \to H(\to b\bar{b})Z(\to q_h\bar{q_h})\nu\bar{\nu}$ (right).

9.2.1 Signal and background events generation

This chapter presents the study of the double Higgs production, where the Higgs decays into $b\bar{b}$ pairs: $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$. Since the $H \to b\bar{b}$ decay has the highest branching ratio, the four b channel is the one with the highest statistic for the HH cross section measurement. The b-jets identification efficiencies and c-jets mis-tag calculated in section 8.4 have been applied to the reconstructed jets, depending on the flavour composition of the final state. The backgrounds sources arising from light quarks are considered negligible. On the other hand, mis-tag of fake jets is found to be low. Furthermore, several improvements in the reconstruction, as mentioned in section 8.3.1 can be applied in the future to further reduce the number of fake jets due to BIB. For these reasons, the fake jets tagging probability is assumed to be negligible. The signal and the physics background are generated by using WHIZARD Monte Carlo at LO. PYTHIA 8 was used to simulate the parton showering, hadronization and fragmentation of hadrons. Background processes with four $b-$ jets final state are dominated by VBF, with two neutrinos. The main physics background contribution comes from processes with four heavy jets (q_h=b or c) in the final state and two neutrinos $\mu^+\mu^- \to q_h\bar{q_h}q_h\bar{q_h}\nu\bar{\nu}$, that comprises multiple intermediate electroweak gauge bosons. WHIZARD options have been tuned in order to not include the HH signal in this process. Most of this background is produced from the decay of two on-shell Z bosons produced via Vector Boson Fusion as shown in the diagram in Figure 133 left. The background produced by the heavy flavour decay of the W boson $\mu^+\mu^- \to (W^+ \to c\bar{b})(W^- \to b\bar{c})\nu\bar{\nu}$ is found around 10^3 times smaller with respect to the Z, since the W decay into b and c quarks is suppressed due to the CKM Matrix.

The other important source of background is the process $\mu^+\mu^- \to Hq_h\bar{q_h}\nu\bar{\nu} \to b\bar{b}q_h\bar{q_h}\nu\bar{\nu}$ right includes also the signal, that has been subtracted in the analysis that is shown in Figure 133 right.

Other sources of background, like $\mu^+\mu^- \to W^{\pm}(\to qq')Z(\to q_h\bar{q_h})\mu^{\pm}\nu_\mu$ background, that arise from W/Z boson fusion. This background has been fully simulated and the number of events with four heavy jets in the final state is found to be negligible with respect to other background sources. In total 10000 events for each process have been generated.

The cross sections of these process, as obtained with WHIZARD at $\sqrt{s} = 3$ TeV center of mass energy are reported in Table 49.

9.2.2 Events reconstruction

The event reconstruction steps are here listed:

Process	σ $[fb]$
$\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$	0.28
$\mu^+\mu^- \to H(\to b\bar{b})q_h\bar{q_h}\nu\bar{\nu}$	2.8
$\mu^+\mu^- \to q_h\bar{q_h}q_h\bar{q_h}\nu\bar{\nu}$	5.4

Table 49: Cross section of signal and background processes used for the determination of the double Higgs cross section uncertainty, with $\mu^+\mu^-$ collisions at 3 TeV. q_h stands for heavy quarks final states, since the contribution to the background from light quarks is assumed to be negligible.

- The interaction of final state particles from the physics processes listed in Table 49 with the detector has been simulated with the Muon Collider framework;
- Hits have been selected by applying the timing windows defined in section 7.6 to the tracker system and the calorimeter systems;
- The track reconstruction have been performed with the CKF algorithm;
- Jets have been reconstruction with the algorithm described in 8.3.1, optimized for the 1.5 TeV BIB, and the jet energy has been corrected by using the jet energy correction function shown in Figure 125;
- In order to select real jets coming from the b or c quarks hadronization, truth-level jets associated to b or c quark are matched to reconstructed jets (jet$_{reco}$). A reconstructed jet is matched to a true-jet if is the jet$_{reco}$ in the event that minimize the distance in the $\eta - \phi$ plane ΔR, and if ΔR<0.5. By using the jet matching the fake jet contribution is assumed to be negligible;
- The b and c tagging efficiencies in section 8.4 have been applied to the matched jets with p_T>20 GeV on a statistical basis: a random number between 0 and 1 is extracted from a uniform distribution. If the number is higher than the b-tagging or the c mis-tag probability calculated for the jet p_T interval, the jet is labeled as tagged, if the number is lower, the jet is labeled as non tagged;
- All possible two-jet combinations are formed, by requiring at least one jet for each pair to be tagged.

As mentioned in section 7.4, the BIB for a 3 TeV muon collider is not ready to be used for physics studies. In this analysis the physics signal and background processes listed in the next section at 3 TeV have been fully simulated with the 1.5 TeV BIB produced by the MAP collaboration. As shown in section 7.4, the BIB at 1.5 and 3 TeV are expected to have a very similar behavior in terms of time of arrival to the Muon Collider detector. The timing windows set to the tracker system and the calorimeter systems, to reduce the BIB at 1.5 TeV are then supposed to allow a comparable background reduction also for the 3 TeV physics analysis. Furthermore, similar performance in terms of jets reconstruction, fake rate and secondary vertex reconstruction are expected, since the number of particles and their average energy is expected to be similar between 1.5 and 3 TeV BIB.

9.2.3 Events selection

Two pairs of jets are then chosen among all the jet pairs reconstructed in the last section 9.2.2 in the following way:

- The invariant mass is calculated for each jet pair:

$$m_{ij} = \sqrt{(E_i + E_j)^2 - |p_{ij}||p_{ij}|} \tag{9.1}$$

where i,j are the indices relative to the two jets,

$$|p_{ij}| = \sqrt{(p_{i,x} + p_{j,x})^2 + (p_{i,y} + p_{j,y})^2 + (p_{i,z} + p_{j,z})^2} \tag{9.2}$$

is the module of the vectors sum of the jets momenta and $E_{i(j)}$ is the energy of the i-th(j-th) jet.

- Among all the calculated invariant masses, the two pairs (i,j) and (k,l), where i,j,k,l are associated to different jets, that minimize the following relation:

$$\sqrt{(m_{ij} - m_H)^2 + (m_{kl} - m_H)^2}$$

with $m_H = 125$ GeV are selected.

No further selection requirements have been applied to increase the signal to background ratio, since all variables relative to signal and background have been given in input to a Multi-variate analysis, that exploits full information about the events and correlation between different variables, to separate the signal from the background. The number of signal and background events is calculated by using the formula:

$$S = \sigma \cdot \mathcal{L}_{int} \cdot t \cdot \epsilon_{sel} \tag{9.3}$$

where σ is the cross section of the process, $\mathcal{L}_{int}$ the integrated luminosity, that is assumed to be the one foreseen for the 3 TeV muon collider (1 ab^{-1}) and ϵ_{sel} is the selection efficiency. In Table 50 all the processes are listed, together with the reconstruction efficiency and the number of expected events.

Process	$\epsilon[\%]$	σ $[fb]$	N_{exp}
$\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$	27.50 ± 0.45	0.28	77
$\mu^+\mu^- \to H(\to b\bar{b})q_h\bar{q_h}\nu\bar{\nu}$	24.72 ± 0.43	2.8	698
$\mu^+\mu^- \to q_h\bar{q_h}q_h\bar{q_h}\nu\bar{\nu}$	18.1 ± 0.39	5.4	976

Table 50: Selection efficiency, cross section and expected events for signal and background processes in the $HH \to b\bar{b}b\bar{b}$ analysis with $\mu^+\mu^-$ collisions at 3 TeV and $\mathcal{L} = 1$ ab^{-1}.

9.3 Determination of the HH cross section precision

A multivariate data analysis method is used to discriminate the HH from the background. The choice of the input variables has been done by looking at the more discriminating ones that are described here. The multivariate data analysis method is used within the Toolkit for Multivariate Data Analysis (TMVA) is a package that provides a machine learning environment in ROOT for signal/background classification. In general a MVA classification with a supervised learning method, consists of these steps:

- training of the model on a data sample. Each event is characterized by a set of variables called features variables, that have different distributions for the signal and the background. On the basis of their distributions the model tunes its parameters in order to learn how to efficiently distinguish events into the two categories. For this purpose a sample of 10000 events of signal ($\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$) and 10000 events of four heavy quark background ($\mu^+\mu^- \to q_hq_hq_hq_h\nu\bar{\nu}$) have been generated, reconstructed and selected with the procedure described in section 9.2.2 for the training;

- test of the model by using a sample of data different from the one used in the training step to evaluate its performance.

Several classification algorithms provided by TMVA are tested, and the best with the highest performance have been used for the analysis.

9.3.1 Study of kinematic properties

The choice of the input variables has been done by looking at the more discriminating ones, that are described here.

- The mass of the two pairs of jet M_{H1} and M_{H2}, namely the invariant mass relative to the pair of jets with the highest transverse momentum (leading candidate) and the one relative to the pair of jets with the lowest transverse momentum (sub-leading candidate). In Figure 134 left and right the invariant mass distributions of the leading and sub-leading candidates are shown for signal (red) and the background (black) are presented. In the signal a peak at higher invariant mass is evident, since the jet pair is associated to the Higgs candidate. For the background, most of the jet pairs are produced by the decay of an electroweak boson, whose invariant mass peaks at 80-90 GeV;

- Figure 135 left shows the distributions of the module of the vectorial sum of the four jets momentum $|P|$:
$$|P| = |\vec{p}_1 + \vec{p}_2 + \vec{p}_3 + \vec{p}_4| \tag{9.4}$$
where p_1 and p_2 are respectively the highest and the lowest momentum of the jets relative to the leading candidate, p_3 and p_4 are respectively the highest and the lowest momentum of the jets relative to the sub-leading candidate. Figure 135 right shows the sum of the four jets energy for signal (red) and background (black) events.

- Figure 136 left shows the distributions of the angle between the two jets relative to the leading candidates. The angle between each jet pair is calculated for any four jets combination and the maximum angle is chosen. The distribution for the signal (red) and background (black) is shown in Figure 136 right. From this last distribution, it can be seen that the HH events have a larger aperture angle with respect to the background.

- Figure 137 shows the angular variable relative to the angle between the highest p_T jet in the pair relative to the leading (left) and sub-leading candidate (right) with respect to the z axis. From these distributions, we can see that background events are more central with respect to the signal events.

- Figure 138 shows the p_T distribution of the highest p_T jet (left) and the lowest p_T jet (right) relative to the leading candidate.

- Figure 139 shows the p_T distribution of the highest p_T jet (left) and the lowest p_T jet (right) relative to the sub-leading candidate.

9.3.2 Multivariate analysis

The signal and background events kinematic distributions presented in the previous section 9.3 are given as input to TMVA. The full sample of signal and background events is split randomly in two. Half of the samples are used as training sample, while the others are used as test sample. TMVA provides several machine-learning algorithms with which perform a multivariate analysis, for example Boosted Decision Tree (BDT), BDT with gradient boost (BDTG) Multilayer Perceptron

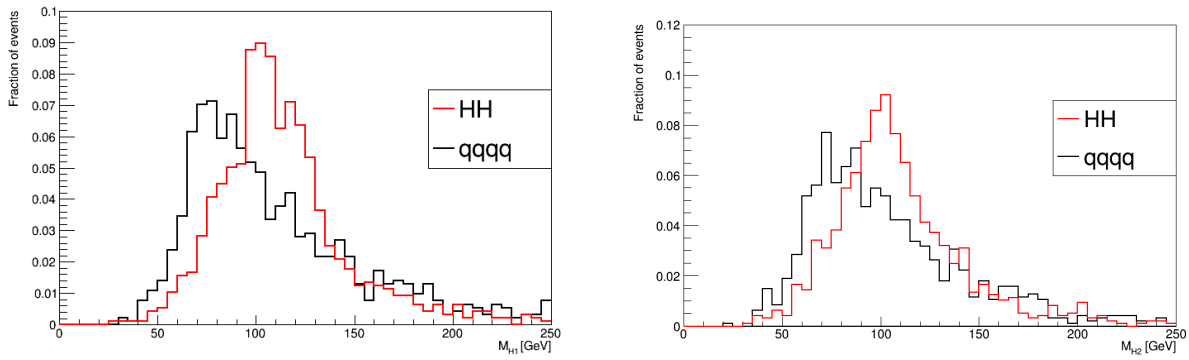

Figure 134: Left: distributions of the invariant mass of the leading candidate for signal (red) and background (black) events. Right: distributions of the invariant mass of the sub-leading candidate for signal (red) and background (black) events. The distributions are normalized to the unity area.

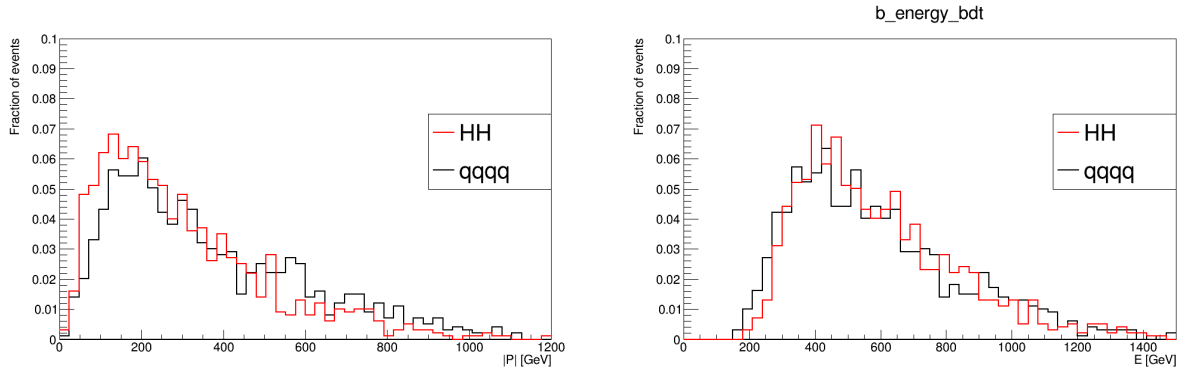

Figure 135: Left: distributions of the module of the vectorial sum of the four jets momentum for signal (red) and background (black) events. Right: distributions of the module of the sum of the four jets energy for signal (red) and background (black) events. The distributions are normalized to the unity area.

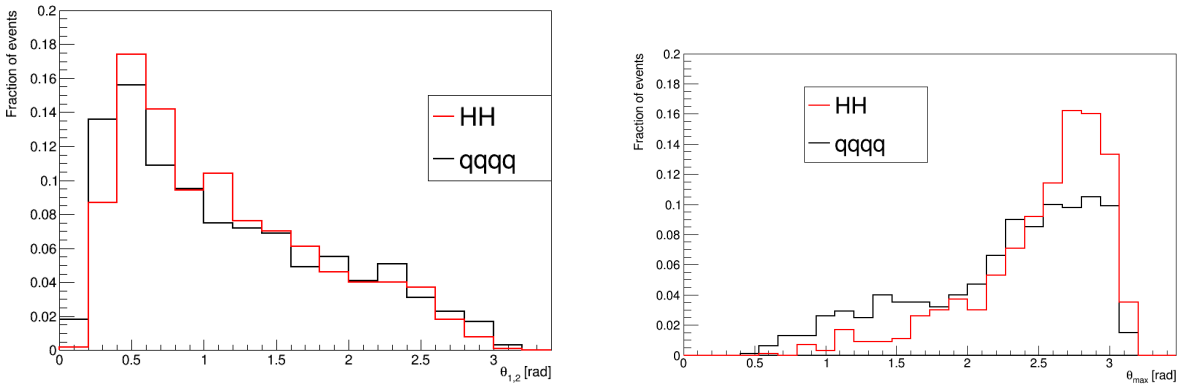

Figure 136: Left: Distributions of angle between the two jets from the leading candidate for signal (red) and background (black) events. Right:Distributions of the maximum angle for signal (red) and background (black) events. All angles are calculated in the laboratory frame. The distributions are normalized to the unity area.

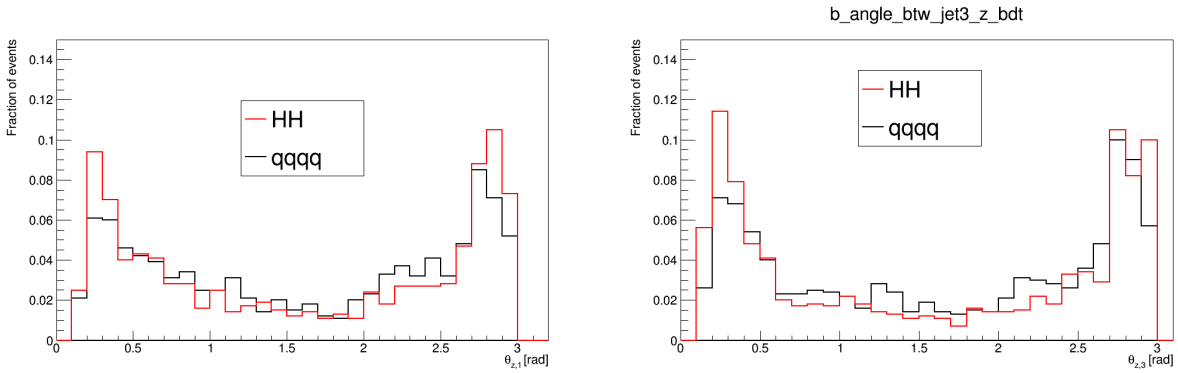

Figure 137: Left: distributions of the angle between the highest p_T jet in the pair relative to the leading candidate with respect to the z axis for signal (red) and background (black) events. Right: distributions of the angle between the highest p_T jet in the pair relative to the sub-leading candidate with respect to the z axis for signal (red) and background (black) events. All angles are calculated in the laboratory frame. The distributions are normalized to the unity area.

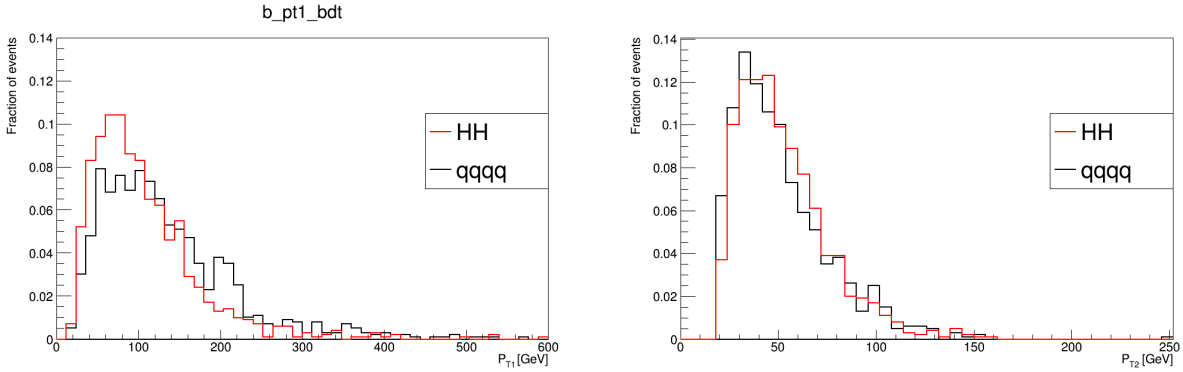

Figure 138: Left: distributions of the highest p_T jet in the pair relative to the leading candidate for signal (red) and background (black) events. Right: distributions of the angle between the lowest p_T jet in the pair relative to the leading candidate for signal (red) and background (black) events. The distributions are normalized to the unity area.

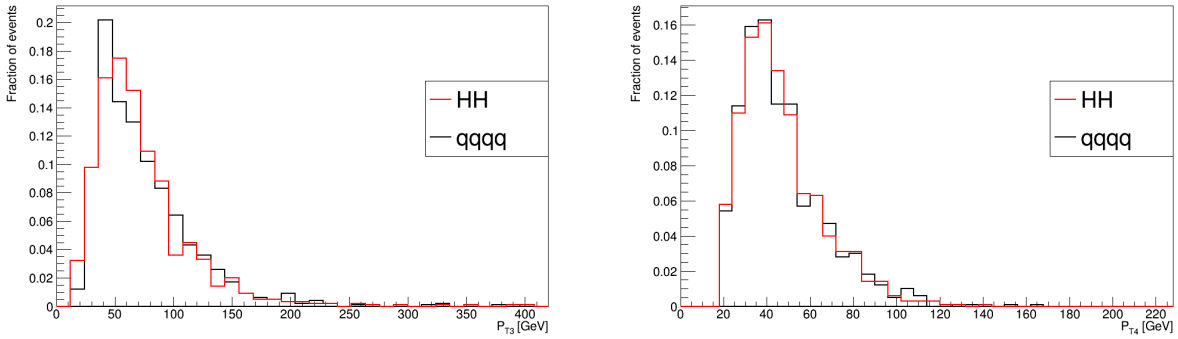

Figure 139: Left: distributions of the highest p_T jet in the pair relative to the leading candidate for signal (red) and background (black) events. Right: distributions of the angle between the lowest p_T jet in the pair relative to the sub-leading candidate for signal (red) and background (black) events. The distributions are normalized to the unity area.

(MLP), Cut optimisation, Fisher discriminant and Likelihood. The different classifiers can be compared by calculating area under the Receiver Operating Characteristic ROC curve (Area Under Curve, AUC), that shows the signal efficiency as a function of the background rejection. The larger the area under the curve (called AUC), the better is the separation of signal from background that can be achieved. In Figure 140 the ROC curve is shown for different classification methods used with TMVA. In this case it can be seen that the BDT, the BDTG and the MLP gives the best separation with an AUC value of 0.817 for the BDT, 0.801 for the BDTG and 0.796 for the MLP.

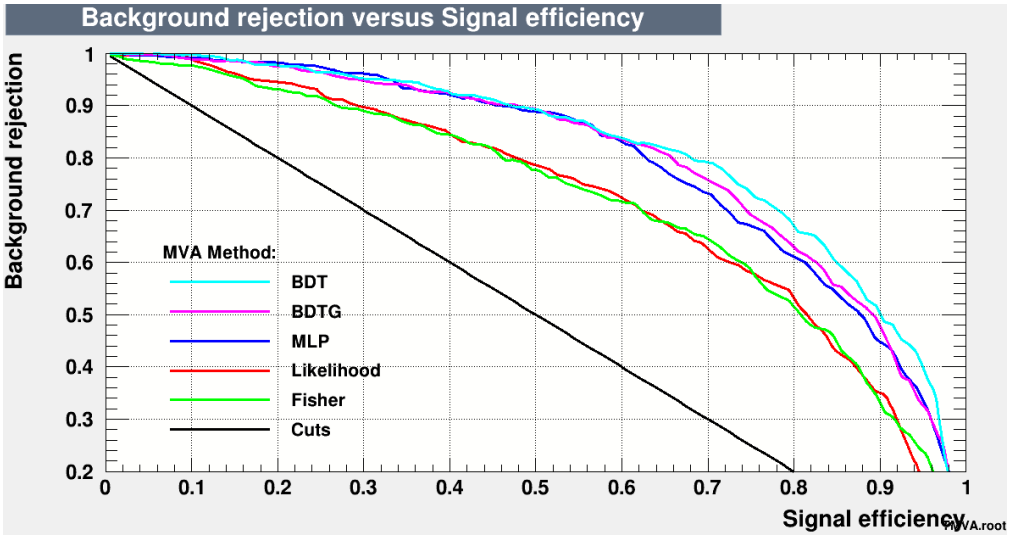

Figure 140: Receiver operating characteristic (ROC) for different MVA methods: BDT with gradient boost (BDTG) Multilayer Perceptron (MLP), Cut optimisation, Fisher discriminant and Likelihood.

Another aspect that must be checked to evaluate the quality of a discriminator is the overtraining. It occurs when too many parameters of the algorithms are adjusted to classify a data sample with few events. This leads to a low mis-classification error in the training procedure, but it is not able to perform classification in the testing sample. Overtraining can be checked by comparing the distributions of the classifier output on the training and test sample, each for signal and background, separately for the signal and the background samples. In Figure 141 the distributions of the MLP (left) output and the BDT (right) for the training (filled histograms) and test samples (dots) are compared. Signal distributions are colored in blue, background distributions are colored in red. It can be seen that for the BDT the training distributions are not fully compatible with the test distributions. This is confirmed by the Kolmogorov-Smirnov (KS) test, that is used to check the compatibility between two shapes, gives a zero p-value both for the signal and the background. On the other hand, for the MLP the two distributions are found to compatible.

Figure 142 and 143 show the linear correlation coefficients between variables in percentage. Coefficients close to 100(-100) % means that two variables are highly correlated (uncorrelated). Correlation close to 100% or -100 % means that two variables do not add information to the training of the discriminator algorithm. In this case all correlation coefficients are below 70%.

9.3.3 Event classification

The MLP has been applied to classify events of the simulated signal and background processes mentioned in Table 49. Figure 144 shows the distribution of the MLP output for the signal $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$ (blue) and the two backgrounds $\mu^+\mu^- \to H(\to b\bar{b})q_h\bar{q_h}\nu\bar{\nu}$ (green) and $\mu^+\mu^- \to q_h\bar{q_h}q_h\bar{q_h}\nu\bar{\nu}$ (red). As can be seen, the signal and four heavy quarks background follows the distributions found in Figure 141 right, as expected. The qqH background, on the other hand

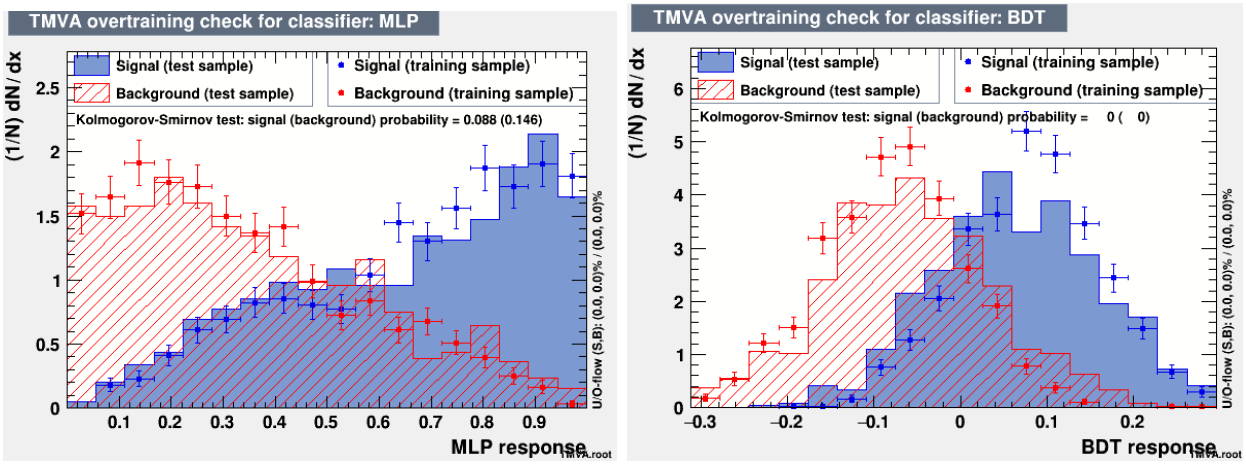

Figure 141: Overtraining check for the MLP (left) and the BDT (right) classifier. Points are used for output on the training samples, histograms are used for the test samples. The shapes of the training and test samples distributions are compared for the signal (blue) and the background (red) events.

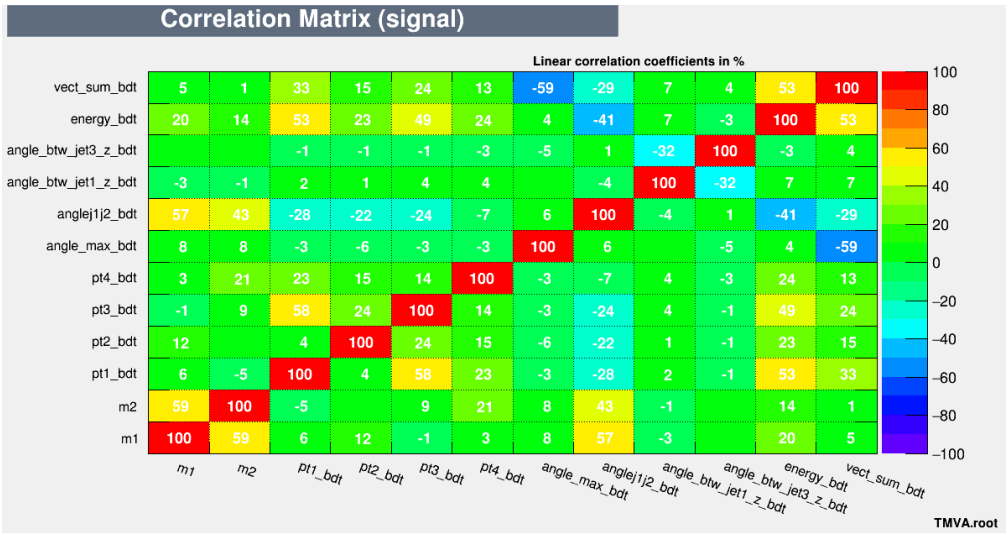

Figure 142: Signal correlation matrix among MLP input variables.

shows an enhancement for MLP close to 1, since the process includes the signal.

9.3.4 Determination of the HH cross section precision

Histograms shown in Figure 144 have been used to build the pdf of the model of the signal and background for the determination of the uncertainty on the HH cross section. The three background distributions have been weighted for the number of events expected for 1 ab^{-1} shown in Table 49, and the signal distribution have been subtracted to the qqH distribution. The signal and background invariant mass models and the expected yields of events are used to generate 2000 pseudo-data. The pseudo-data are then fitted with the invariant mass models, by using a binned maximum likelihood fit, and by letting the signal yields floating. The physics background, since it is generated by electroweak process, it is assumed to be determined with a high theoretical precision

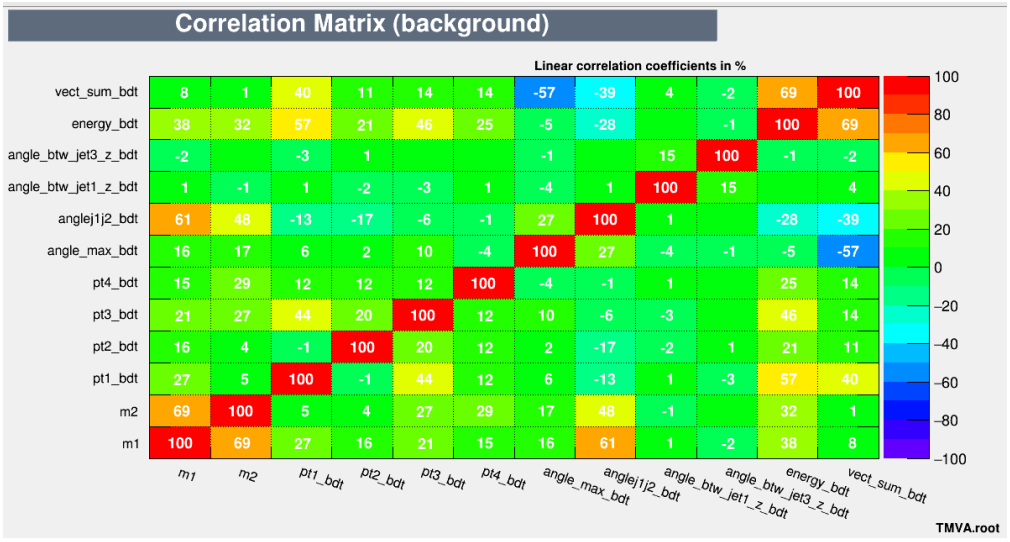

Figure 143: Background correlation matrix among MLP input variables.

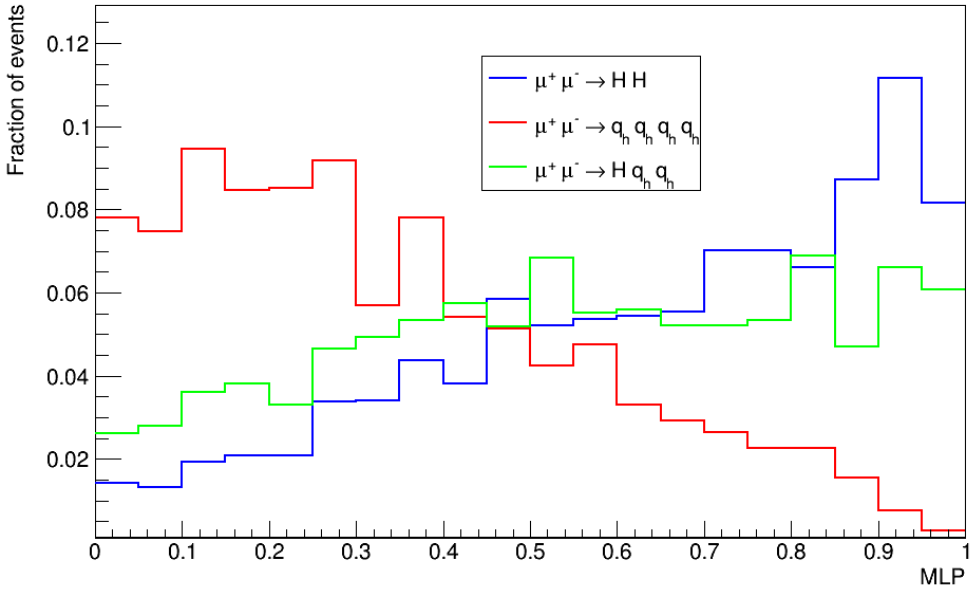

Figure 144: MLP output distributions of the signal $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$ (blue) and the two backgrounds $\mu^+\mu^- \to H(\to b\bar{b})q_h\bar{q_h}\nu\bar{\nu}$ (green) and $\mu^+\mu^- \to q_h\bar{q_h}q_h\bar{q_h}\nu\bar{\nu}$ (red).

and is fixed in the fit. The uncertainty on the cross section has then be estimated as:

$$\sigma(HH \to b\bar{b}b\bar{b}) = \frac{N_S}{\epsilon \cdot \mathcal{L}} \tag{9.5}$$

where N_S is the number of signal events extracted from the fit and ϵ is the signal efficiency. Assuming the uncertainty on the efficiency and the luminosity negligible, the statistical uncertainty on the cross section is the uncertainty on the signal yield obtained from the fit. Figure 145 shows an example of the MLP fit used to extract the number of signal yield and uncertainty. The physics background template is colored in blue, while the signal one is colored in red. Dots are the pseudo-experiments. The relative uncertainties of each pseudo-experiment have been plotted in Figure 146. The Landau distributions is used to fit the distribution, and the values are shown in Table 51. The most probable value is found to be 0.34.

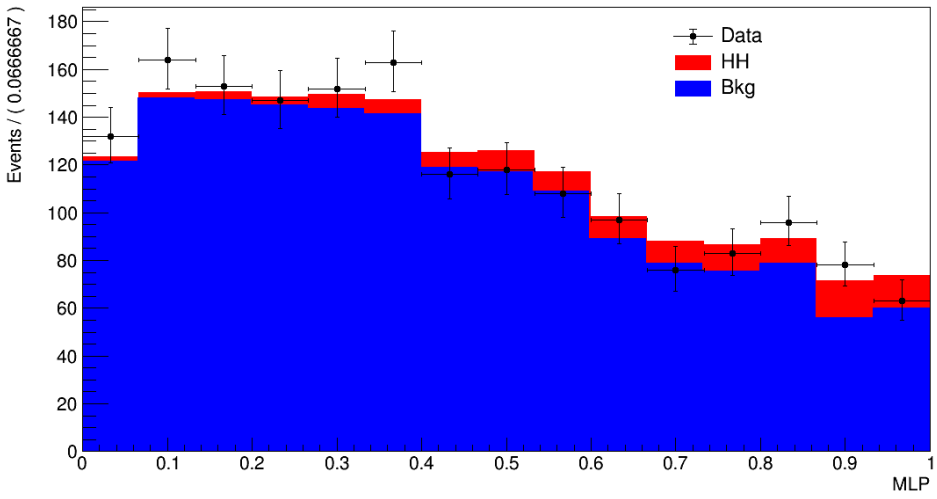

Figure 145: Example of the MLP fit used to extract the number of signal yield and uncertainty.

Constant	723± 25
Most Probable Value (MPV)	0.339 ± 0.003
Sigma	0.056± 0.002

Table 51: Parameters of the Landau distribution used to fit the $\Delta\sigma/\sigma$ distribution.

The fit results are listed in Table 51.

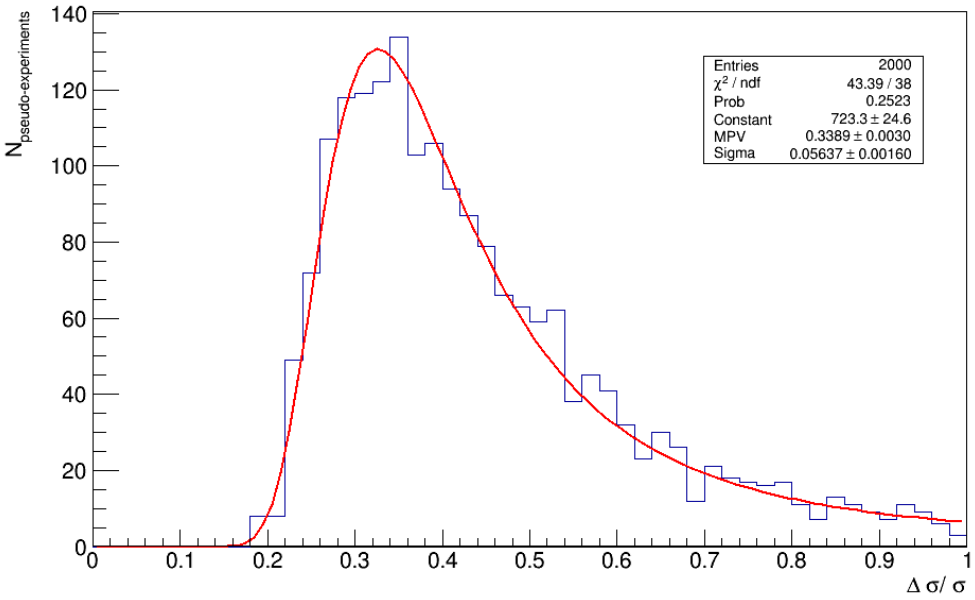

Figure 146: Relative uncertainties on the Higgs cross section ($\Delta\sigma/\sigma$) of all pseudo-experiments.

This result can be compared with CLIC.

In [8] CLIC determined the uncertainty on the $e^+e^- \to HH\nu_e\bar{\nu}_e$ cross section, assuming 5 ab^{-1} and the polarized beams. Considering the $HH \to b\bar{b}b\bar{b}$ channel, an uncertainty of 7.4 % is found. The main difference between the result that have been obtained at the muon collider and CLIC is the integrated luminosity considered for the estimation, that is five times more for CLIC, and the polarized beams, that imply a total increment of the $S/\sqrt{S+B}$ ratio of a factor 2.72. By scaling

the 7.4 % for the different luminosity and the polarization factor, the uncertainty on the double Higgs cross section found by CLIC is $\sim$ 20 %, that is lower, but comparable, to the 34% found in this thesis. In order to understand how much the result at the muon collider can be improved assuming the BIB effects on the reconstruction negligible and with better tagging performances, the same analysis will be repeated in section 9.5.2 on samples reconstructed without the BIB.

9.4 Determination of the sensitivity on the Higgs self-coupling

The analysis for the determination of the uncertainty on the trilinear Higgs self-coupling have been carried out by studying $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$ events. The total cross section of VBF double Higgs boson production is sensitive to the value of the trilinear Higgs self-coupling: Figure 147 shows the three Feynman diagrams of double Higgs production. The first one on the left is the only process directly affected by the value of the trilinear Higgs self-coupling. However, the total cross section of the these processes have to be used for the coupling measurement, since interference between them affects their cross sections.

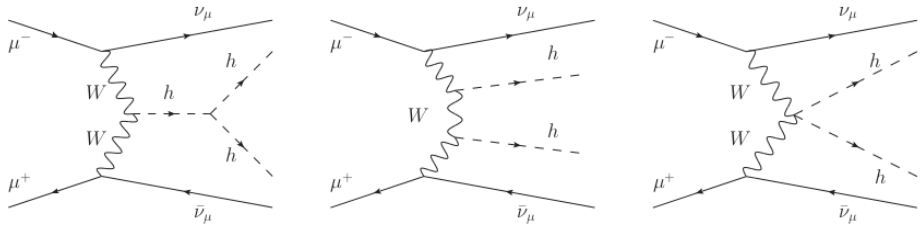

Figure 147: Feynman diagrams of the three double-Higgs production modes [141].

The analysis strategy followed to extract the uncertainty on the trilinear Higgs self-coupling is the following:

- Samples of double Higgs events are generated for different values of

$$\kappa_{\lambda_3} = \frac{\lambda_3}{\lambda_{3,SM}} = (0.2, 0.4, 0.6, 0.8, 0.9, 1.0, 1.1, 1.2, 1.4, 1.6, 1.8)$$

  with WHIZARD, and subsequently simulated and reconstructed with the procedure followed in section 9.2.2. The $k = 1$ case is the Standard Model signal. Table 52 shows the expected number of events for all κ_{λ_3} hypothesis:

κ_{λ_3}	$\sigma(H \to b\bar{b})$ [fb]	ϵ	N_{events} (1 ab^{-1})
0.2	0.52	30	155
0.4	0.44	29	129
0.6	0.38	28	106
0.8	0.32	28	89
0.9	0.30	28	84
1.0	0.28	27	77
1.1	0.27	26	68
1.2	0.25	26	65
1.4	0.23	27	62
1.6	0.22	25	56
1.8	0.23	25	56

Table 52: Cross sections values, selection efficiencies and number of events for all κ_{λ_3} hypothesis.

- Two MLPs are trained independently:

  1. The first is the same MLP used for the cross section extraction (section 9.3.3), that have been trained to separate the SM signal ($k = 1$) from the physics background;

2. the second MLP is trained on four variables, to separate the SM signal from a sample of $\mu^+\mu^- \to HH\nu\bar{\nu}$ with $H \to b\bar{b}$, where the Higgs pair is produced only via trilinear Higgs self-coupling (first diagram in Figure 147). Note that since this diagram interferes with the others double Higgs production diagrams, it is only used to train the MLP, while all the samples with different κ_{λ_3} hypothesis mentioned above, that have been used for the uncertainty calculation, contains all the processes. The idea behind this approach is to build a MLP in that exploits the kinematic characteristics of the Higgs pair produced only via trilinear Higgs self-coupling to separate it from the other double Higgs production. The Multi-variate analysis performed for this purpose is shown in section 9.4.2.

The two MLP scores on these samples have been arranged in 2-dimensional histograms.

- A 2-dimensional template with the MLPs scores is built for all the samples: the SM signal, the two physics backgrounds, and all the samples with different κ_{λ_3} hypothesis.
- Total signal+background 2-dimensional templates are then built for each κ_{λ_3} hypothesis: the HH templates for different κ_{λ_3} hypothesis, weighted for the expected yields are summed to the physics background templates weighted for the SM expected background yields. The SM HH template is also subtracted from the qqH template to avoid double counting.
- The total 2D template for the $\kappa_{\lambda_3} = 1$ hypothesis SM is then used to generate pseudo-datasets, and the logarithm of the likelihood between the κ_{λ_3} hypothesis+ background templates and each pseudo-experiment is calculated.
- A Log Likelihood scan is then obtained. The log likelihood profile have been fitted with a polynomial function of fourth degree, and the uncertainty on κ_λ at 68% C.L. is obtained by taking the points along the curve that are as far as ΔLL=0.5 from the minimum.

9.4.1 HH vs HH from trilinear coupling classification

The choice of the input variables to the MLP used to separate the SM signal from a sample of $\mu^+\mu^- \to HH\nu\bar{\nu}$ with $H \to b\bar{b}$, where the Higgs pair is produced only via trilinear Higgs self-coupling has been done by looking at the more discriminating ones, that are described here.

- Figure 148 left shows the angle between the two Higgs momenta in the full HH signal (red) and the sample with the HH from trilinear (black). They have been calculated by finding the total momentum of the two jets associated to the leading and sub-leading Higgs candidate and by calculating the angle in the laboratory frame between them. As can be seen, when the two Higgs are produced from trilinear coupling, they are closer with respect to other processes. Figure 148 right shows the cosine of the helicity angle of the two Higgs, which is defined as the angle between the Higgs momentum in its mother frame and the mother momentum in the laboratory frame. In the case of the two Higgs produced from trilinear coupling, their mother is the off-shell Higgs, and since it has spin 0, the distribution of the cosine of the angle is flat. In the full HH, the Higgs is produced also by W boson propagators, then the distribution of the cosine of the helicity angle is different.
- Figure 149 shows the angle between the highest p_T jet (left), the sub-leading candidate (right), in the pair with respect to the z axis. From these distributions, we can see that jets from the Higgs from trilinear (black) are more central with respect to the one from the full HH signal (red).

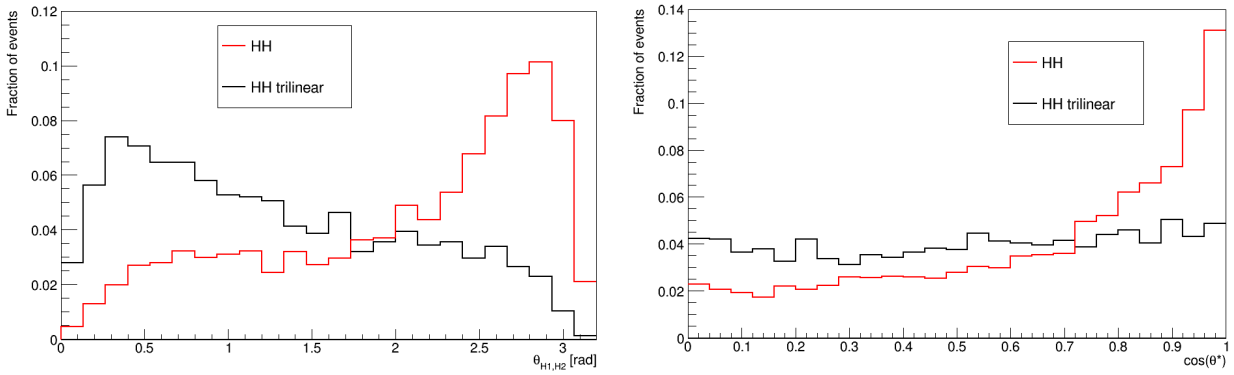

Figure 148: Angular aperture between the two Higgs (left) and the cosine of the Higgs helicity angle (right), calculated in the laboratory frame. Distribution from the HH from trilinear are shown in black, while SM full HH events are shown in red.

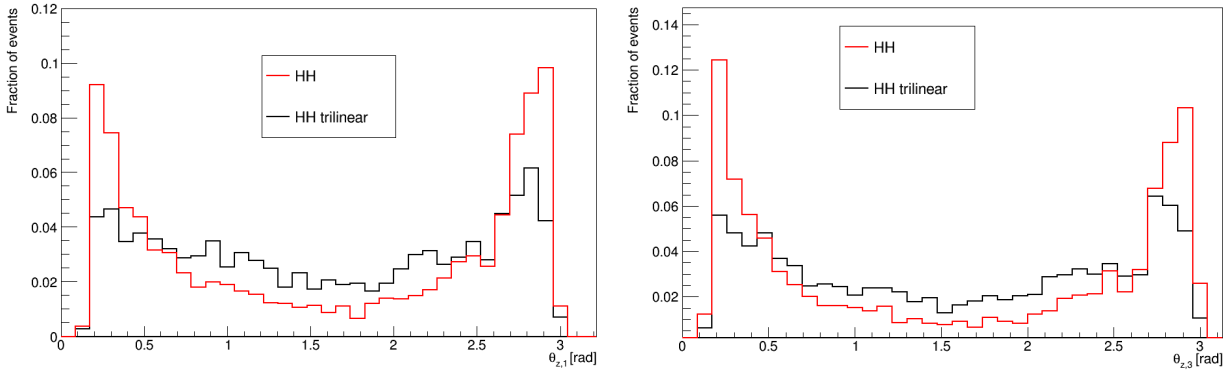

Figure 149: Left: angle between the highest p_T jet in the pair with respect to the z axis, calculated in the laboratory frame. Right: same variable for the sub-leading jet. Distributions from the HH from trilinear are shown in black, while SM full HH events are shown in red.

9.4.2 Event Classification

The ROC curve obtained by training the TMVA algorithms with 10000 events of each sample are shown in Figure 150. The best AUC is obtained with the MLP (0.827). Figure 151 shows the distributions of the MLP output for the training (filled histograms) and test samples (dots) are compared. Distributions relative to the full HH are colored in red, while distributions relative to the HH from trilinear are colored in blue. As can be seen there is no overtraining in the model. The signal and background linear correlation coefficients between variables are shown Figure 169 and Figure 170 in Appendix D.

Figure 152 shows the Multi-Layer Perceptron (HH vs HH trilinear) output, for all the κ_{λ_3} samples. Distributions are normalized to the number of events listed in Table 52.

9.5 Studies on the uncertainty on the trilinear Higgs self-coupling at muon collider

The analysis strategy described at the beginning of section 9.4 has been applied to the signal, background and different κ_3 hypothesis samples reconstructed in the presence of BIB and requiring two tagged jets per event, assuming the BIB performance found in section 8.4.2, one for each pair, as described in section 9.2.2. Pseudo-datasets have been generated using the expected SM signal and background yields and the 2-dimensional template for $\kappa_{\lambda_3} = 1$. A binned likelihood scan is

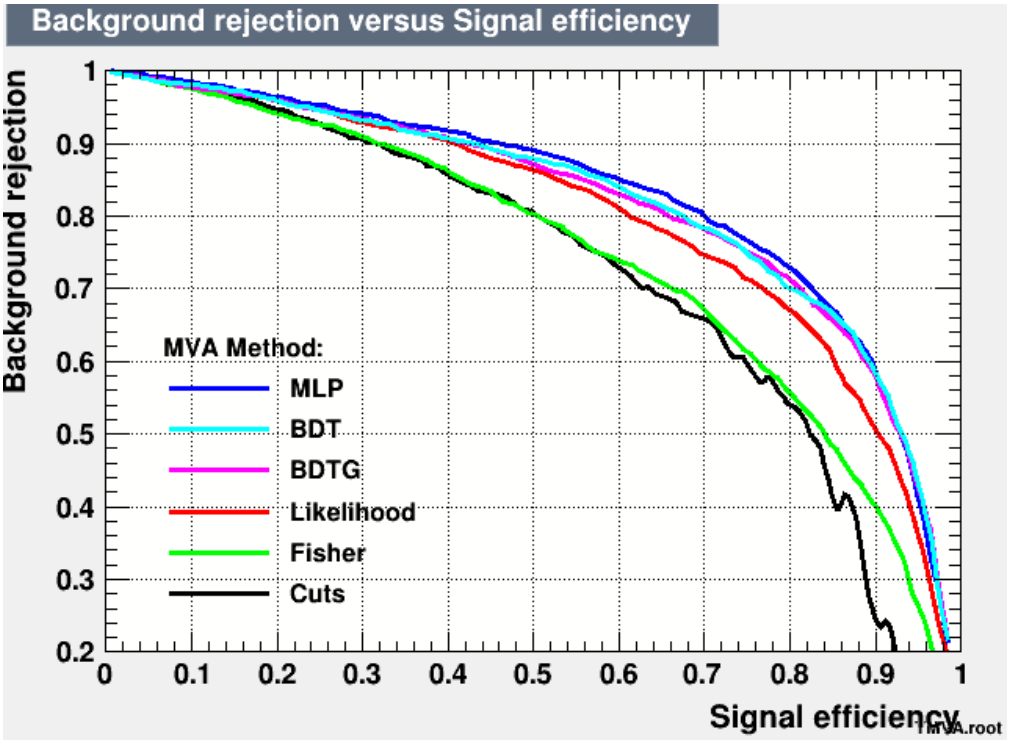

Figure 150: Receiver operating characteristic (ROC) for different MVA methods.

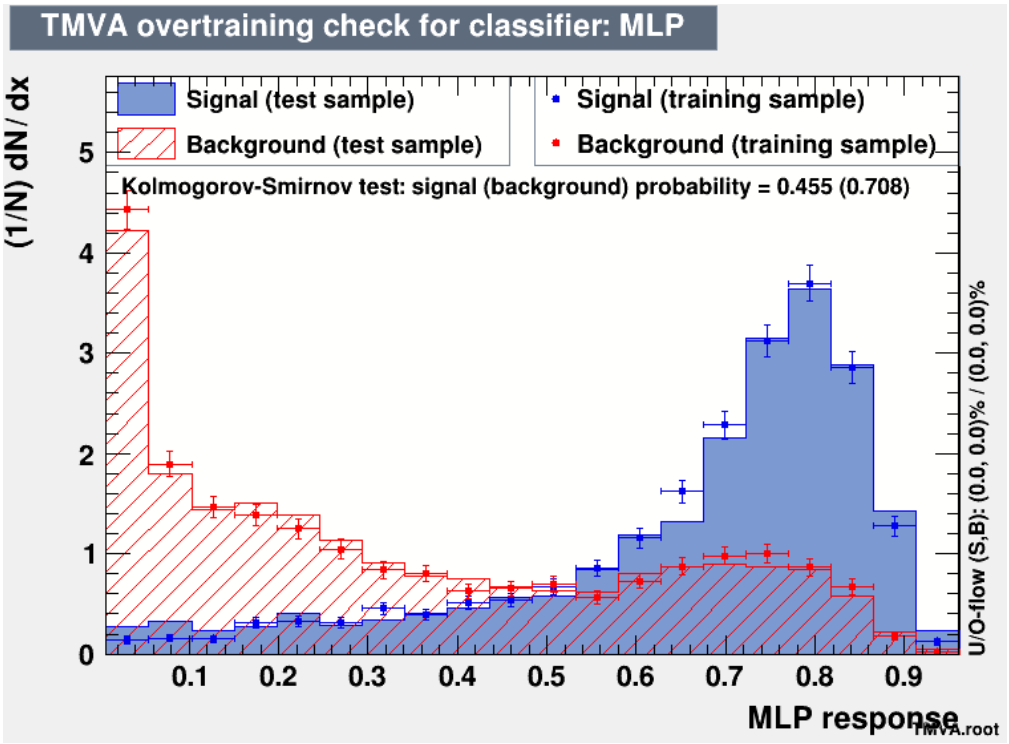

Figure 151: Overtraining check for the MLP. Points are used for output on the training samples, histograms are used for the test samples. Distributions relative to the full HH are colored in red, while distributions relative to the HH from trilinear are colored in blue.

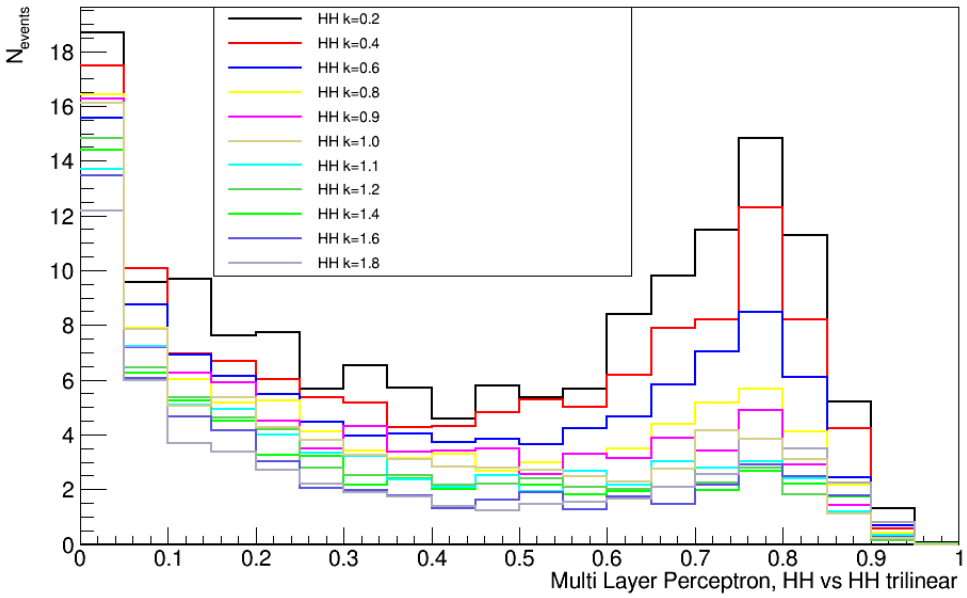

Figure 152: MLP trained to saparate the HH vs the HH sensitive to the trilinear coupling output distributions for all the κ_{λ_3} samples, normalized to the expected number of events calculated in Table 52.

calculated between pseudo-experiments and each κ_{λ_3} hypothesis template. The large number of background events introduces too much noise to be sensitive to the differences between the κ_{λ_3} hypothesis. The 68 % C.L. is found to be around [-40%, 90%]. This result is much higher with respect to the one obtained at the same center of mass energy of 3 TeV by the e^+e^- collider CLIC. The expected sensitivity for CLIC on the HH cross section and on the trilinear Higgs self-coupling, has been calculated in [8]. Using the same process that has been used in this analysis, $\mu^+\mu^- \rightarrow HH\nu\bar{\nu} \rightarrow b\bar{b}b\bar{b}$, CLIC will be able to measure the trilinear Higgs self-coupling with a relative uncertainty of -9% and +12% at 68% C.L. with an integrated luminosity of 5 ab^{-1}. The first difference is that the integrated luminosity assumed by CLIC is five times larger than the Muon Collider. The beam-polarization is also assumed, that increases the number of selected signal and background events by a factor 1.48. This imply a total improvement on the HH significance of a factor 2.72. However, even scaling the uncertainty obtained in this section to the same CLIC luminosity, the difference is still high.

In the next sections, possible improvements that can be reached by assuming negligible the BIB effects on the reconstruction are investigated. In the next sections, the effects of the BIB on the uncertainty on the cross section uncertainty and on the trilinear Higgs self-coupling have been evaluated in two ways:

- the tagging performance calculated with the BIB have been improved, by assuming the one calculated for LHCb in section 8.4.2;
- the same samples have been simulated without the BIB, in order to improve the invariat mass resolution and allow for a better separation of the H and Z invariant mass resonances.

9.5.1 Improvements on tagging performance

The b and c-jets tagging efficiencies and mist-tag rates presented in section 8.4.2 are not fully optimized as already mentioned, but these are the only ones developed so far and the optimization process is in progress. Their usage in the HH measurements limits the possible Muon Collider

performance giving underestimated results, which are not representing the actual situation. These results are expected to be improved in many ways. For example, the hits in the tracker system could be reduced by applying a state-of-the-art clustering algorithm which can exploit the fact that BIB particles are low momentum therefore more ionizing. This will reduce the number of fake tracks. A dedicated jet reconstruction algorithm has to be developed to minimize the effect of the BIB. Then, proper Machine Learning algorithm that combines the full information of jets and the tracks belonging to it, and not only the presence of the secondary vertex, like it was shown for LHCb in section 4.11 could be developed. Therefore, it is reasonable to assume that when all the improvements will be in place, the Muon Collider jet tagging efficiencies and mistag rates will be comparable to the LHCb ones. In the following the b tagging efficiency and c mistag of the LHCb experiment shown in Figure 42will be used. The red curve is the b tagging efficiency as a function the c mis-tag calculated for different values of P_b. A working point for the muon collider is chosen by maximizing the HH significance ($S/\sqrt{S+B}$) for different b-tagging c mis-tag configurations. For each tagging configuration, the significance is calculated by requiring three jets tagged in the event and at least one jet per pair tagged. The working point with the highest significance is found to be: 3 tagged jet, a total b-tagging efficiency of 76% and a c mis-tag of 20%. The b-tagging efficiency and c mis-tag curves, as a function of the jet p_T are shown in Figure 153 left and right, respectively. These values are comparable to those assumed by other Futrure Colliders. CLIC for

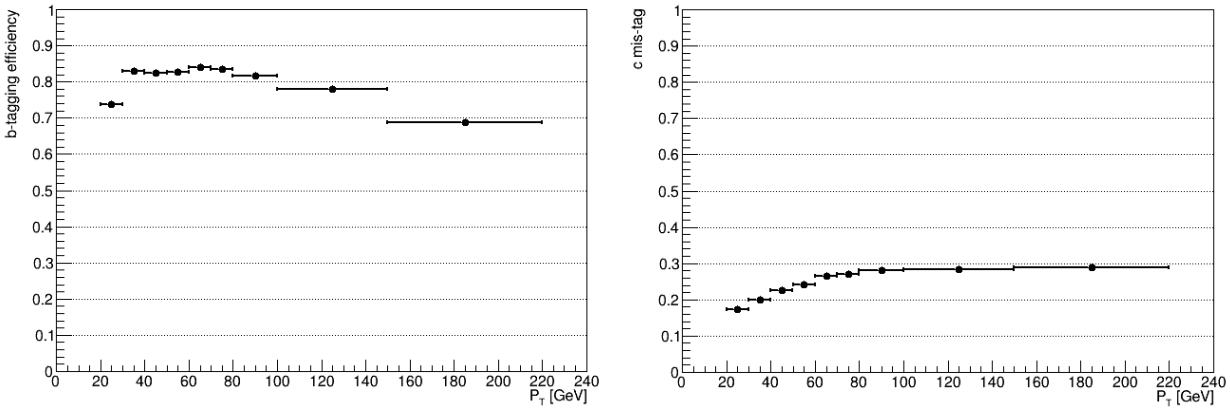

Figure 153: Left: b-tagging efficiencies as a function of the p_T at the highest significance working point. Right: c mis-tag efficiencies as a function of the p_T at the highest significance working point.

example, declare the same b tagging efficiency with a mis-tag due to c quarks of 10%, better than the Muon Collider assumptions.

The number of signal and background events at different κ hypothesis determined by using the new tagging efficiencies and mistag rates are reported in Table 54. The MLP for the HH vs background and HH vs trilinear optimization have been repeated using the same configuration. Figure 154 left shows the average ΔLL over the 1000 pseudo-experiments calculated for each κ hypothesis. The uncertainty on the trilinear Higgs self-coupling improves by around 20% for the upper bound and 10%in the lower bound, and the total uncertainty is [-32%, +70%]. By further reducing the c mis-tag to the CLIC level, keeping constant the b tagging efficiency, the uncertainty on the trilinear Higgs self-coupling is further improved up to [-31%, +62%]. The scan is shown in Figure 154 right.

9.5.2 Beam-induced background effects

As shown in section 8.3.1, the selections applied to calorimeter cluster reconstruction to reduce the BIB affect the jet energy resolution which, then, worsen the di-jets invariant mass resolution.

Process	N_{events} (1 ab^{-1})
NHH	97
4q	921
Hqq	794
κ_{λ_3}	N_{events} (1 ab^{-1})
0.2	197
0.4	162
0.6	134
0.8	113
0.9	105
1.0	97
1.1	90
1.2	82
1.4	76
1.6	70
1.8	70

Table 53: Number of signal and background events found at the working point of best significance assuming the LHCb performance on the b tagging and c mis-tag.

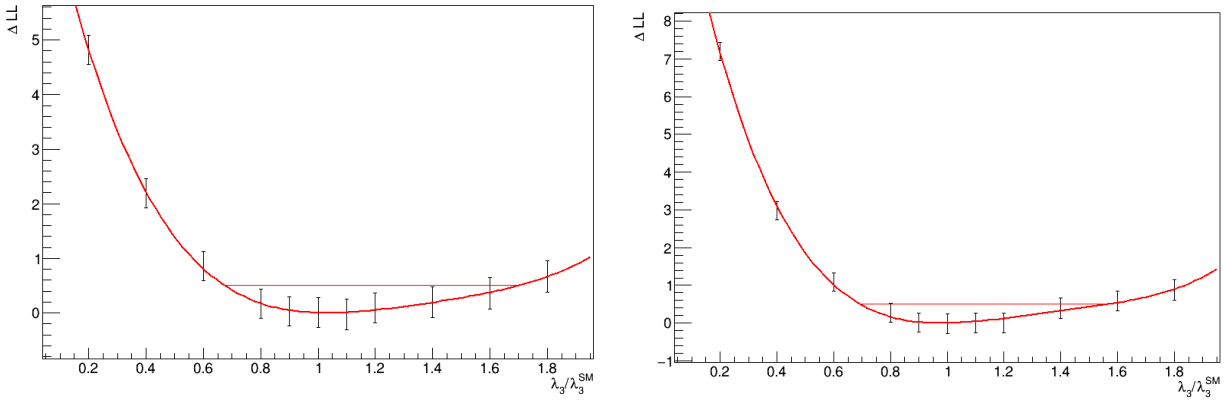

Figure 154: Left: ΔLL as a function of κ hypothesis for samples selected assuming LHCb tagging performance (b-tagging efficiency of 76% and a c mis-tag of 20%), by requiring three tagged jets in the event. Right: ΔLL as a function of κ hypothesis for samples selected assuming LHCb tagging performance and CLIC c mis-tag (b-tagging efficiency of 76% and a c mis-tag of 10%), requiring three tagged jets in the event.

The comment made previously on the tagging can be repeated here: the impact of the BIB on the HH measurements is more detrimental than what it will be when the reconstruction algorithms will be optimized. The evaluation of the final effect of the BIB on the reconstruction is very difficult to determine, therefore a limit case is considered. The assumption is that the effects of BIB on reconstruction performance are negligible. The data samples described above is reconstructed without the BIB. The conformal tracking algorithm, the Pandora Particle Flow toolkit and the k_T-algorithms with R parameter R=0.5 is used to reconstruct jets. Events are then been selected with the following requirements, based on the optimizations found in section 9.5.1, not new optimization is performed:

- In order to select jets coming from the b or c quarks hadronization, truth-level jets associated to b or c quark are matched to reconstructed jets (jet$_{reco}$). A reconstructed jet is matched to

a true-jet if it is the jet$_{reco}$ in the event that minimize the distance in the $\eta - \phi$ plane ΔR, and if ΔR<0.5.

- The b and c tagging efficiencies in section 9.5.1 have been applied to the matched jets with p_T>20 GeV with the same procedure followed in section 9.2.2. The working point of 20% and 10% of c mis-tag are tested;
- At least three jets per event are required to be tagged;
- All possible two-jet combinations are formed;
- The jets are then paired by calculating the invariant mass among all combinations and minimizing equation 9.1.

The improvements on the jet energy resolution can be shown in 155 left, where the invariant mass resolution of the leading Higgs candidate is compared between HH events and $q_h\bar{q}_h q_h\bar{q}_h$ events without the BIB. The Multi-variate Analysis to separate HH from the four heavy quarks physics

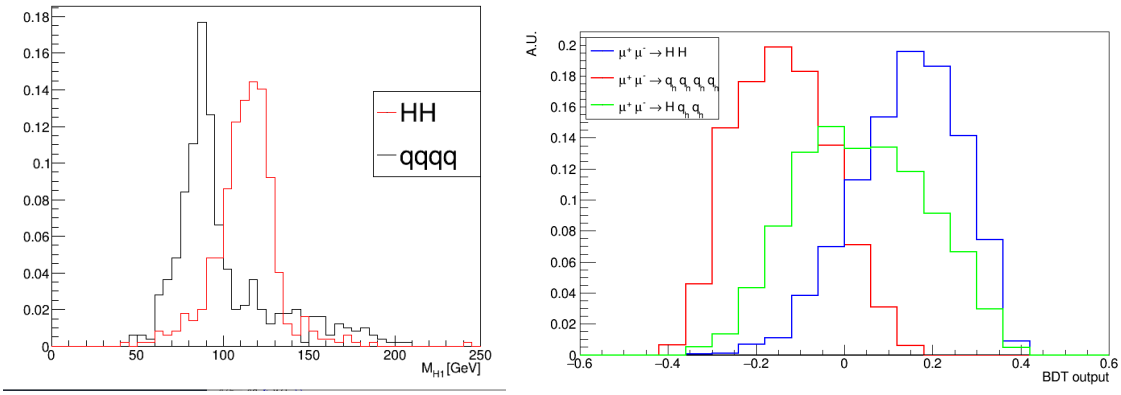

Figure 155: Left:Invariant mass of the leading Higgs candidate compared between HH events and $q_h\bar{q}_h q_h\bar{q}_h$ events, by using samples reconstructed without the BIB. Right:BDT output distributions of the signal $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$ (blue) and the two backgrounds $\mu^+\mu^- \to H(\to b\bar{b})q_h\bar{q}_h\nu\bar{\nu}$ (green) and $\mu^+\mu^- \to q_h\bar{q}_h q_h\bar{q}_h\nu\bar{\nu}$ (red).

background and the total HH from the HH sensitive to the trilinear coupling is repeated. Similar variables to the one selected in section 9.3 are used to separated the events: the invariant mass of the leading and sub-leading Higgs, the highest p_T jets associated to the two Higgs candidates, the maximum angle between jets in the event, the angle between the highest p_T jets associated to the two Higgs candidates with the z-axis, the sum of the four jet energies in the event and the module of the sum of the four jets momenta. In this case the MLP and the Boosted Decision Tree (BDT) method show similar performances. The BDT is found to have a higher AUC (0.934) with respect to the MLP, then it is used. Figure 155 right shows the BDT output distributions of the signal $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$ (blue) and the two backgrounds $\mu^+\mu^- \to H(\to b\bar{b})q_h\bar{q}_h\nu\bar{\nu}$ (green) and $\mu^+\mu^- \to q_h\bar{q}_h q_h\bar{q}_h\nu\bar{\nu}$ (red). As can be seen, the separation between the samples is much better with respect the samples with the BIB. 67 events of signal and 1000 of background have been reconstructed.

The same procedure followed to calculate the uncertainty on the cross section described in 9.3.4 is followed and the final distribution of the $\Delta\sigma/\sigma$ assuming LHCb performance on the tagging, is shown in Figure 156. The Most Probable Value is about 28 %. By further reducing the background contribution, assuming a c mistag at the 10 % level, the Most Probable Value is around 26 %, that is 8% better than the uncertainty obtained including the BIB and is close to the CLIC uncertainty on the cross section.

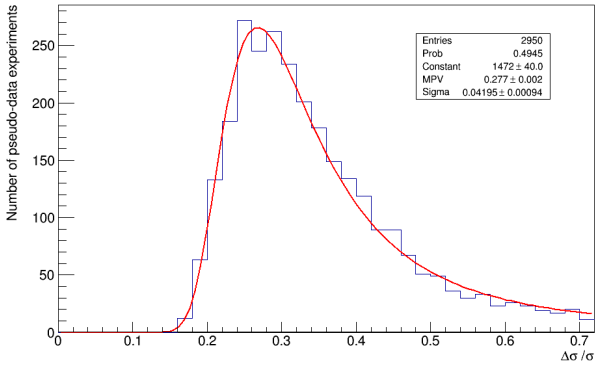

Figure 156: Relative uncertainties on the $\Delta\sigma/\sigma$ of all pseudo-experiments.

Process	N_{events} (1 ab^{-1})
NHH	67
4q	526
Hqq	543

κ_{λ_3}	N_{events} (1 ab^{-1})
0.2	140
0.4	120
0.6	101
0.8	80
0.9	74
1.0	67
1.1	63
1.2	59
1.4	55
1.6	48
1.8	48

Table 54: Number of events for samples reconstructed without the BIB, selected assuming LHCb tagging performance (b-tagging efficiency of 76% and a c mis-tag of 20%), requiring three tagged jets in the event.

The BDT used for the uncertainty on the cross section calculation, has been used to build the 2D templates used for the likelihood scan. The number of events used for this likelihood scan, by assuming $\sim$ 76% of b tagging efficiency, $\sim$ 20% of c mis-tag and by requiring three tagged jets is shown in Table 54. The likelihood scan obtained with this configuration is shown in Figure 157. The uncertainty on the trilinear Higgs self-coupling is found to be [-20 %,45 %] at 68 % C.L.. Assuming to be able to further reduce the physics background contribution by reducing the c mis-tag to a 10% level, the uncertainty can be reduced up to [-20 %,40%] at 68 % C.L.. In conclusion, the reduction of the BIB effects in the reconstruction and the tagging can improve the uncertainty on the trilinear Higgs self-coupling from [-31 %, +62 %] to [-20 %,40%] at 68 % C.L.. These are the best results that can be obtained with this analysis, and are comparable to the one obtained by CLIC (-9% and +12% at 68% C.L. assuming 5 ab^{-1} and polarized beams [8]).

Table 55 summarizes the results obtained in this section for both the statistical uncertainty on $\sigma_{HH} \cdot BR(H \to b\bar{b})^2$ and the trilinear Higgs self-coupling. CLIC results are scaled to take into account for differences in the integrated luminosity and beam polarization. Results shown in Table 55 for the Muon Collider are obtained in the most optimistic scenario in which BIB effects on the

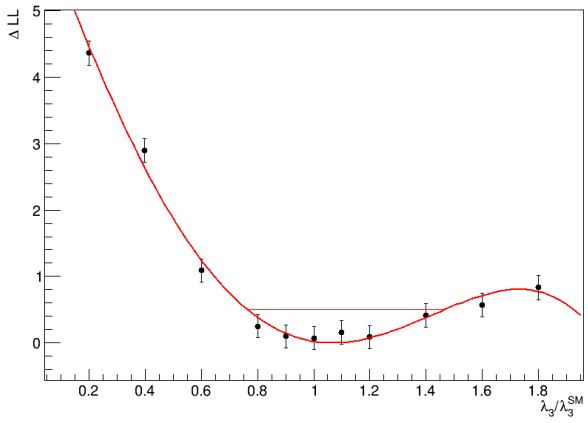

Figure 157: ΔLL as a function of κ hypothesis for samples reconstructed without the BIB, selected by assuming LHCb tagging performance (b-tagging efficiency of 76% and a c mis-tag of 20%), by requiring three tagged jets in the event.

Statistical uncertainty on $\sigma_{HH} \cdot BR(H \to b\bar{b})^2$	
Muon Collider	26 %
CLIC	20 %
Statistical uncertainty on κ_3	
Muon Collider	[-20% ,+40%]
CLIC	[-24% ,+33%]

Table 55: Summary of the results obtained at the Muon Collider and CLIC experiments with 1 ab^{-1}. Muon Collider results are obtained assuming the BIB effects negligible, CLIC results are taken from [8] and scaled by a factor 2.72 to take into account differences in integrated luminosity and beam polarization.

events reconstruction are assumed to be negligible. The current reconstruction tools are far from being fully optimized and do not allow to reach these performance, but many proposals to improve the detector design and the reconstruction techniques are under development. Furthermore, a more effective analysis strategy to determine the statistical uncertainty on κ_3 could be adopted after the BIB effects will be reduced. Here the same analysis strategy is followed to stay on the conservative side. For example, the procedure followed by CLIC to select events and calculate the uncertainty on the trilinear coupling is different from the one adopted in this thesis due to the not-optimal resolution: the di-Higgs invariant masses for the different couplings hypothesis are used to perform the χ^2 minimization with respect to the Standard Model hypothesis and find the uncertainty interval on the trilinear Higgs self-coupling.

For these reasons, performance close to the one obtained neglecting the BIB effects can be expected after the inclusion of the all these optimizations. It is also interesting to compare the [-20% ,+40%] uncertainty interval on κ_3 obtained with the full simulation of the detector, with the $\sim$ 25% uncertainty found in [39], assuming a parametric simulation of the detector response, details on the reconstruction assumptions have been described in section 2.4. The two results are comparable, demonstrating that the prospects calculated in [39] are not too far from what can be experimentally obtained. This gives confidence on the prospects on the trilinear coupling at 10 TeV center of mass energy evaluated in the same paper. By running a 10 TeV Muon Collider for 5 years a $\sim$ 6% uncertainty is expected to be reached with 10 ab^{-1} which is the best value any future collider can achieve.

10 Conclusions

In this thesis several properties of the Higgs boson have been studied. The decays $H \to b\bar{b}$ and $H \to c\bar{c}$ have been reconstructed at the LHCb experiment to measure their cross sections. Then the double Higgs production cross section sensitivity at Muon Collider is evaluated.

LHCb is a forward spectrometer detector, that would allow to measure the H properties in a phase space region complementary to ATLAS and CMS. The Muon Collider is a proposed future machine that can be seen as Higgs Factory since a large number of H and HH events are produced by $\mu^+\mu^-$ collisions thanks to the multi-TeV center of mass energies, and the high luminosity conditions possible for this machine. Techniques used to identify the flavour of the jets have been applied at both experiments. At the Muon Collider, a vertex reconstruction algorithm has been set up to identify the decay point of a b-jets by using only the tracks not compatible with the primary vertex of the interaction. At LHCb a more sophisticated Machine Learning techniques exploit information about the charged and neutral particles inside the jet and the presence of displaced vertices and allow to identify b and c jets.

Data taken by LHCb experiment during Run II campaign in 2016, at 13 TeV center of mass energy, have been analyzed to search for inclusive production of $H \to b\bar{b}$ and $H \to c\bar{c}$. The contribution of the QCD background is evaluated with a data-driven technique based on the identification of signal and control regions. For each H decay process, the observed upper limit on the production cross section times the branching ratio in $b\bar{b}$ or $c\bar{c}$ is set at 95% confidence level. The observed upper limit on $H \to b\bar{b}$ is found to be 8 times the SM value while the upper limit for the $H \to c\bar{c}$ process is 1475 times the SM value. Assuming the expected improvements discussed in section 6.13 regarding the data selection requirements, the employment of the Machine Learning algorithm to identify b and c flavour jets even without the presence of secondary vertex at trigger level, the increment of integrated luminosity expected for the HL-LHC stage, and the usage of a technique to better correct jet energy, the upper limits will be reduced to 0.3 times the SM value for the $H \to b\bar{b}$ process and 56 times the SM value or the $H \to c\bar{c}$ process.

The former result indicates that with the techniques used in this thesis, by only studying the $H \to b\bar{b}$ inclusive production, in the HL-LHC phase LHCb will be close to the $H \to b\bar{b}$ observation, and that will be the first measurement in the forward region of the pp collision. The upper limit on the $H \to c\bar{c}$ process is the first ever set in the inclusive production with resolved jets. The prospects indicate that the $H \to c\bar{c}$ via inclusive production is still far from the observation. However, additional improvements can be obtained by apply the tagging technique used in this thesis also to the VH production and combining the results.

In LHCb the upper limits on the $V + H(\to b\bar{b})$ and $V + H(\to c\bar{c})$ production [142], studied in the analysis performed with the Run 1 dataset (2 fb^{-1}) at 8 TeV center of mass [29], has been extrapolated to 300 fb^{-1} at 14 TeV, without any improvement on the analysis or on the detector. Here the expected upper limit on the $H(\to c\bar{c})$ cross section arrives to 50 times the Standard Model. Assuming further improvements in the detector performance, the expected limit reaches 5-10 times the Standard Model, which correspond to a limit of 2-3 times the Standard Model prediction on the charm Yukhawa coupling.

At ATLAS and CMS an integrated luminosity of 3000 fb^{-1} is expected for pp collision data at $\sqrt{s} = 14$ TeV. Starting from the results obtained by ATLAS with Run 2 data on the upper limit on the $H \to c\bar{c}$ cross section considering the ZH associated production, an expected upper limit at the 95% confidence level of 6.3 times the Standard Model is estimated [142]. Recently, CMS [143] [144] evaluated the $H \to c\bar{c}$ considering the VH, with $V = W, Z$ associated production and found that a constrain on the Higgs-charm coupling to 1.7 times the SM value is expected with HL-LHC data that corresponds to an upper limit on the signal strength of ~ 3.

From this comparison it is evident that LHCb can give an important contribution to the $H(\to b\bar{b})$ and $H(\to c\bar{c})$ cross section measurements and this thesis constitutes a proof of that.

The uncertainty on the double Higgs cross section and the trilinear Higgs self-coupling measurements at a Muon Collider at 3 TeV center of mass energy have been evaluated for the first time by using the detailed simulation of the detector and including the BIB. The study has been done by simulating the $\mu^+\mu^- \to HH\nu\bar{\nu} \to b\bar{b}b\bar{b}\nu\bar{\nu}$ process and all the relevant physics backgrounds that include heavy quarks in the final state. The performance of the identification of heavy quarks via secondary vertex reconstruction have been evaluated in the presence of the beam-induced background. It has been found that the presence of the BIB highly affects the reconstruction algorithms performance and these create an important bias on the measurements with four jets in the final state. The two main limiting factors are the tagging performance and the jet energy resolution. The tagging performance are limited by the fake tracks presence and the requirements to reduce them and the by the very simple b-jet identification algorithm based only on secondary vertices. The jet energy resolution and consequently the Higgs invariant mass resolution are limited due to energy threshold required in the calorimeter system cells to remove part of the energy deposited by the BIB particles. While the $H(\to b\bar{b})$ is reconstructed with very good performance, final states with four jets in presence of physics background is highly impacted by the above limitations.

The analysis has been repeated by using the the same analysis strategy under different assumptions. The b-jet tagging efficiency and mistag rate are assumed comparable to LHCb ones and then a mis-tag from c reduced to 10% levels. Then it is assumed that BIB effects are negligible.

Even though a huge effort has been made to develop and study the tracks and jets reconstruction algorithms, these, are indeed, not optimized for the Muon Collider environment. This thesis demonstrates that the development of dedicated algorithms are a priority for the future detector and physics studies. When the BIB effects are included and the Muon Collider $b-$jet tagging are used in the reconstruction of HH, the uncertainty on the double Higgs cross section is found be 34%, while under the two assumption just discussed, it is reduced to 26 %. For what concerns the trilinear Higgs self-coupling uncertainty the situation is much more complicated. Indeed in this case it is necessary to separate HH diagrams sensitive to the trilinear Higgs self-coupling from the other HH productions. With no assumptions, the trilinear self-coupling is found to be [-40%, 90%], while it is improved up to [-20 %, 40%] with the two assumptions.

These results have been compared to CLIC, the only other future collider that studied the HH production at 3 TeV center of mass energy. Considering the same production and decay channel of the analysis performed in this thesis, CLIC found a 7.4 % uncertainty on the HH cross section, and -9% and +12% at 68% C.L. on the trilinear Higgs self-coupling, assuming 5 ab^{-1} and beam polarization [8]. Scaling the CLIC results to no polarization beams and to the same luminosity of the Muon Collider, the uncertainties obtained by the two experiments are similar, and are reported in Table 56:

Statistical uncertainty on $\sigma_{HH} \cdot BR(H \to b\bar{b})^2$	
Muon Collider	26 %
CLIC	20 %
Statistical uncertainty on κ_3	
Muon Collider	[-20% ,+40%]
CLIC	[-24% ,+33%]

Table 56: Summary of the results obtained at the Muon Collider and CLIC experiments with 1 ab^{-1}. Muon Collider results are obtained assuming the BIB effects negligible, CLIC results are taken from [8] and scaled by a factor 2.72 to take into account differences in integrated luminosity and beam polarization.

The results shown in 56 are obtained under the two assumptions that can be considered too optimistic in particular the scenario that assumes the BIB effects are negligible. But the results

obtained with the current Muon Collider $b-$jet tagging and the jet reconstruction including the BIB are highly biased by the poor performance of these algorithms with final states with four jets in presence of physics background due to W/Z decays. The two assumptions are, indeed, more close to what can be achieved at Muon Collider with optimized algorithms and can be considered as target detector performance to be reached. In fact, several improvements on the detector design, reconstruction tools and data acquisition are ongoing that will allow to reduce the BIB impact on the measurements.

The results obtained in this thesis and the studies performed are the first one ever on HH with the detailed detector simulation including the BIB effects and the fact that results compatible to those obtained via parametric simulation of the detector response in [39] gives confidence on the expectations at a Muon Collider at 10 TeV center of mass energy: $\sim 6\%$ uncertainty in 10 ab^{-1} of data collected in only 5 years of data taking. This is the best value any future collider can achieve.

A Fisher Test for Transfer Function correction modeling

In the $H \to b\bar{b}$ and $H \to c\bar{c}$ fit to the dijet invariant mass, a correction function is introduced to take into account possible data/MC differences in the Transfer Function (Sec. 5.7). A Fisher's Test [145] is performed in this analysis to determine the proper number of coefficients of the Bernstein function that are needed to fit data. In this section the Fisher's Test procedure is described. This test is used to determine if a more complex model, with more parameters, is better than a simpler one, with less parameters, to fit a set of data. With the Fisher's test we are testing the null hypothesis, that the additional parameter is useless. Null hypothesis always pertains to the reduced model, the alternative hypothesis always pertains to the full model. The steps followed to perform the Fisher's test are:

- define of a larger full model, with more parameters;
- define a smaller reduced model, with fewer parameters;
- decide a size of the test, for example in this analysis α =0.01 have been chosen;
- use an F-statistic to decide whether to reject the smaller reduced model in favor of the larger full model. F is calculated as:

$$F = \frac{\frac{\sum_i (y_i - f_1(x_i))^2 - \sum_i (y_i - f_2(x_i))^2}{p_2 - p_1}}{\frac{\sum_i (y_i - f_2(x_i))^2}{n - p_2}} \tag{A.1}$$

  where f_1 is the reduced model, f_2 is full model, p_1 is the number of free parameters of f_1, p_2 is the number of free parameters of f_2, n is the number of bins. F is distributed as a Fisher-Snedecor distribution ($\mathcal{FS}$), that depends on two parameters: d_1 and d_2, that in the test are $p_2 - p_1$ and $N - p_2$ respectively. In Figure 158 some examples of Fisher-Snedecor distributions for different combinations of d_1 and d_2 are shown;

- calculate the F statistic from data and the p value as:

$$p = 1 - \int_0^F \mathcal{FS}(F', p_2 - p1, N - p_2) dF'; \tag{A.2}$$

- reject the null hypothesis if $p < \alpha$.

When this procedure is applied the fit results are kept blind, just the F and p values are checked.

A.1 Fisher Test for the H $\to b\bar{b}$ mass fit

Bernstein polynomials with 8,9,10 coefficients are tested in the fit. The result of the Fisher's test p value for each hypothesis is reported in Table 57: The Fisher-Snedecor distribution with 1 and 31

Reduced Model	Full Model	F Statistic	p value
Bernstein 8 coeff	Bernstein 9 coeff	8.8	0.006
Bernstein 9 coeff	Bernstein 10 coeff	5.5	0.03

Table 57: Fisher's test results for H $\to b\bar{b}$ mass fit.

degrees of freedom, corresponding to the Fisher's test by fitting data with Bernstein polynomial of 9 and 10 coefficients, is shown in Figure 159. The p-value is highlighted in blue. As can be seen, the null hypothesis (the additional parameter is useless) is rejected for all the tests except for the last, where a Bernstein with 9 coefficients is found sufficient to fit data.

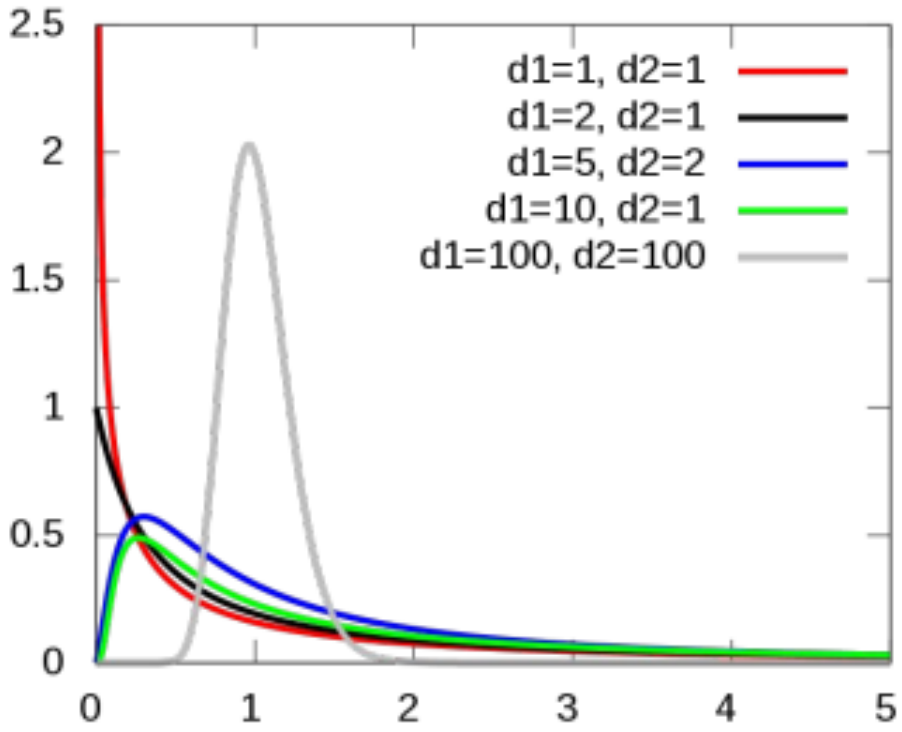

Figure 158: Examples of Fisher-Snedecor distributions for different combinations of d_1 and d_2 [145].

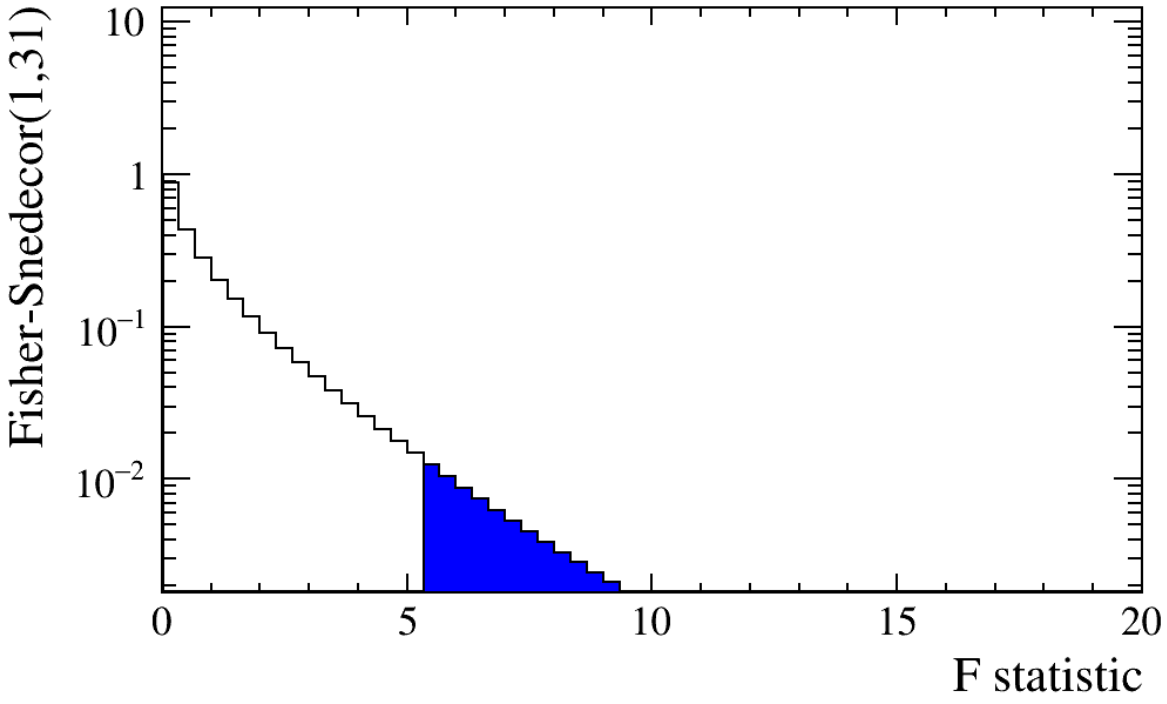

Figure 159: Fisher-Snedecor distribution with 1 and 31 degrees of freedom. In blue is highlighted the p-value calculated with the Fisher's test by fitting data with Bernstein polynomial of 9 and 10 coefficients.

A.2 Fisher Test for the H $\to c\bar{c}$ mass fit

The result of the Fisher's test p value for each hypothesis tested is reported in Table 58: The

Reduced Model	Full Model	F Statistic	p value
Bernstein 9 coeff	Bernstein 10 coeff	7.9	0.009
Bernstein 10 coeff	Bernstein 11 coeff	3.5	0.07

Table 58: Fisher's test results for H $\to c\bar{c}$ mass fit.

Fisher-Snedecor distribution with 1 and 30 degrees of freedom, corresponding to the Fisher's test by fitting data with Bernstein polynomial of 10 and 11 coefficients, is shown in Figure 160. The p-value is highlighted in blue. As can be seen, the null hypothesis (the additional parameter is useless) is rejected for all the tests except for the last, where a Bernstein with 10 coefficients is found sufficient to fit data.

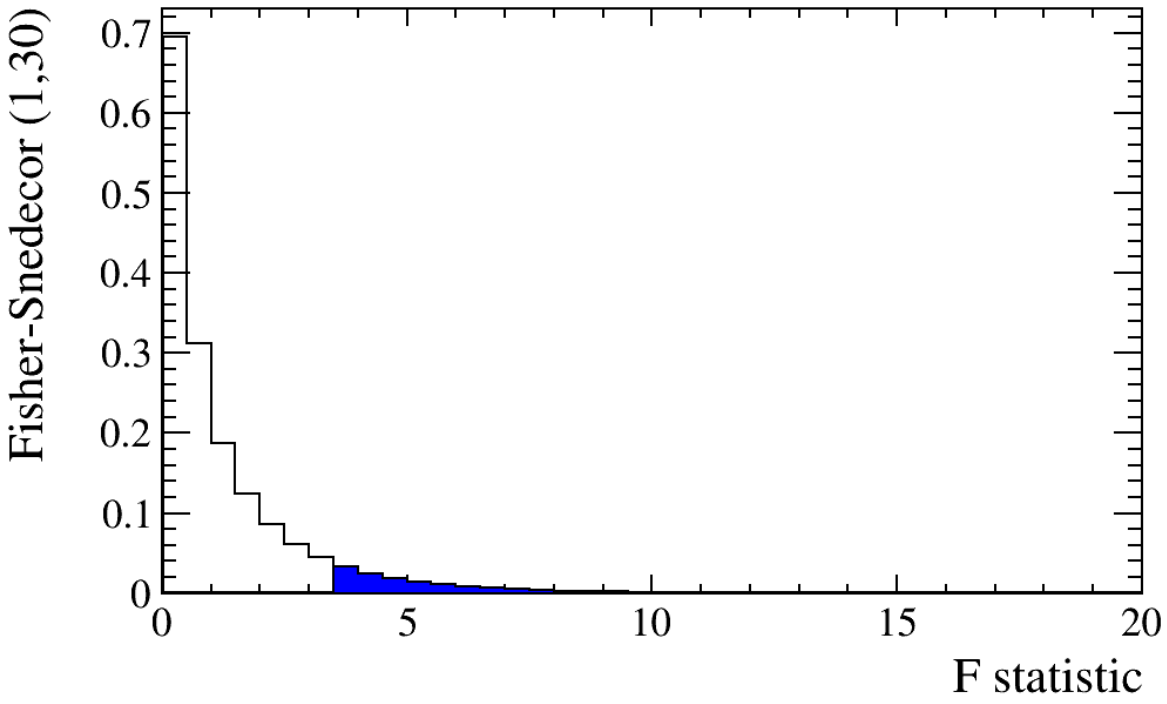

Figure 160: Fisher-Snedecor distribution with 1 and 30 degrees of freedom. In blue is highlighted the p-value calculated with the Fisher's test by fitting data with Bernstein polynomial of 10 and 11 coefficients.

B Fit H→ $b\bar{b}$

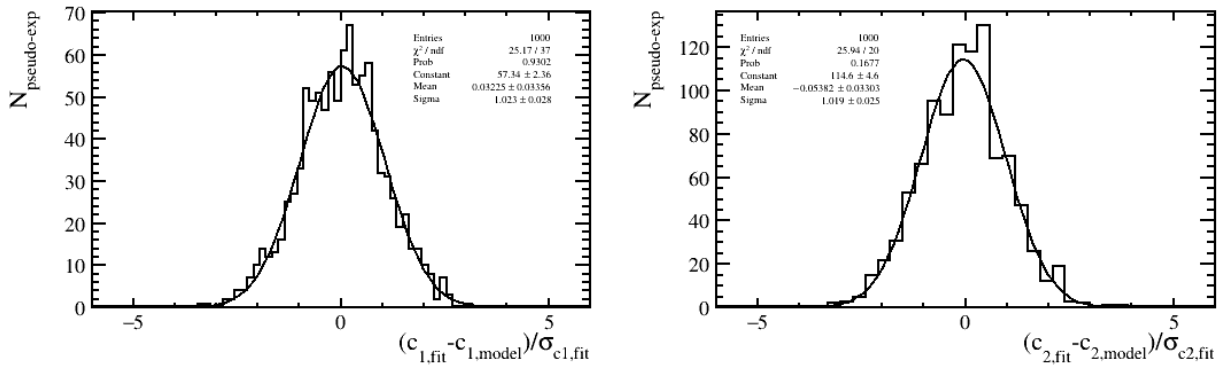

Figure 161: Pull of the coefficients c_1 and c_2 of the model used to fit pseudo-experiments in the H $\to b\bar{b}search$.

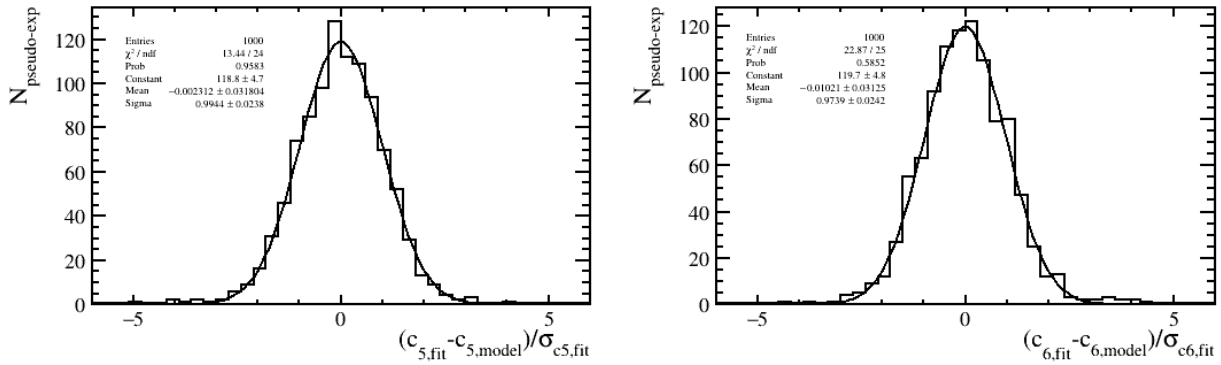

Figure 162: Pull of the coefficients c_5 and c_6 of the model used to fit pseudo-experiments in the H $\to b\bar{b}search$.

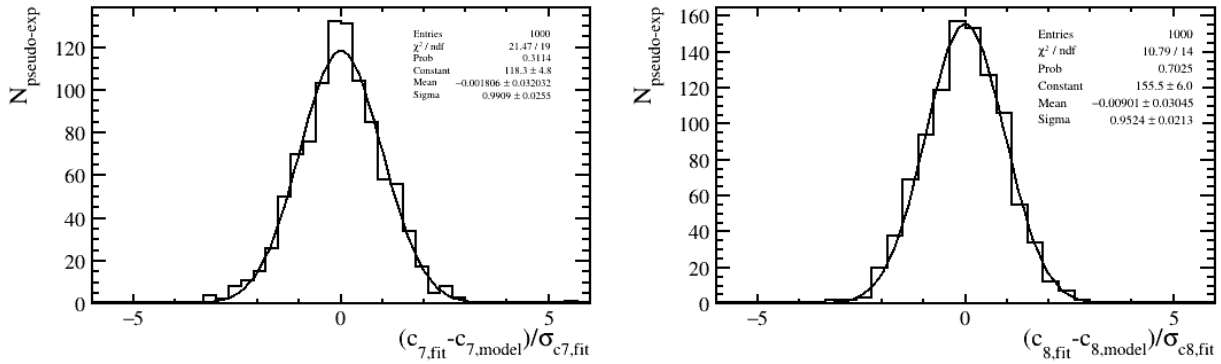

Figure 163: Pull of the coefficients c_7 and c_8 of the model used to fit pseudo-experiments in the H $\to b\bar{b}search$.

C Fit H→ $c\bar{c}$

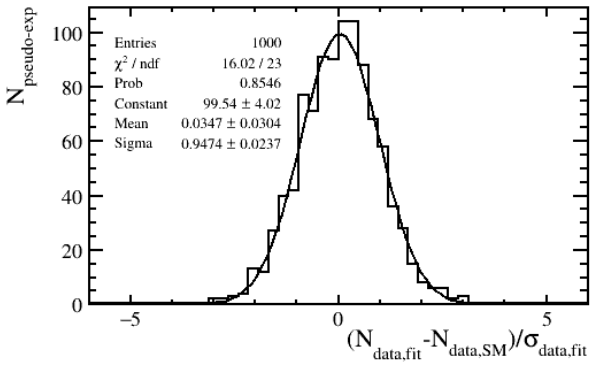

Figure 164: Pull of the number of data from the fit to pseudo-experiments in the H $\to b\bar{b}search$.

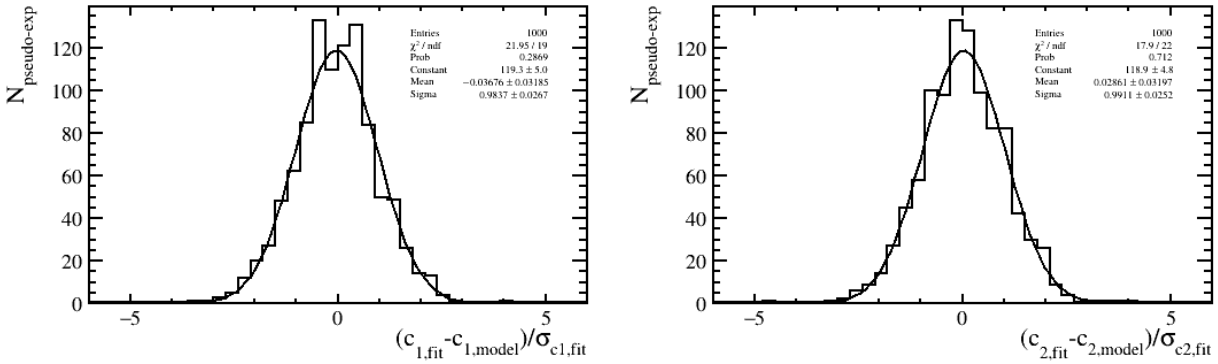

Figure 165: Pull of the coefficients c_1 and c_2 of the model used to fit pseudo-experiments in the H $\to c\bar{c}$ search.

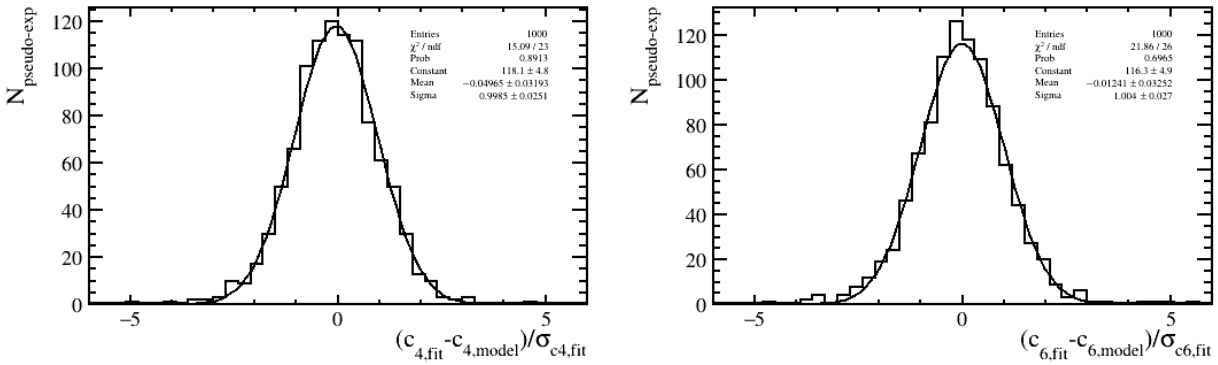

Figure 166: Pull of the coefficients c_4 and c_6 of the model used to fit pseudo-experiments in the H $\to c\bar{c}$ search.

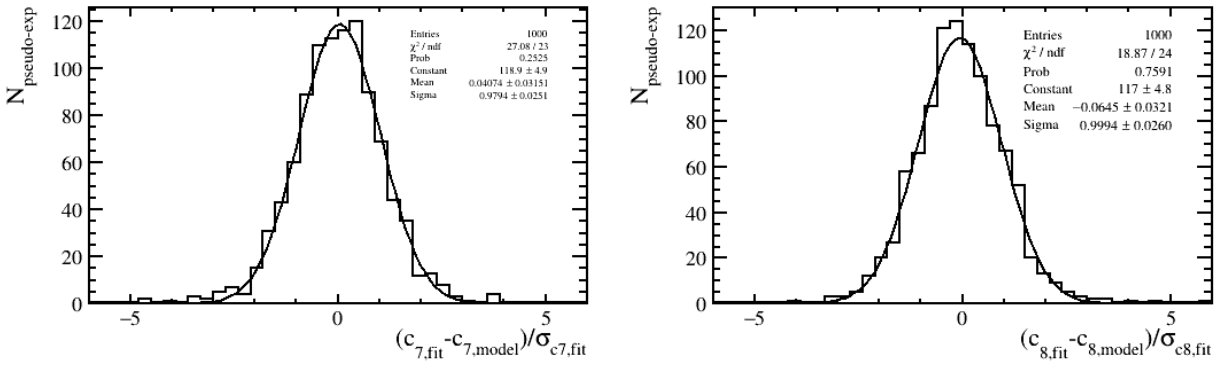

Figure 167: Pull of the coefficients c_7 and c_8 of the model used to fit pseudo-experiments in the H $\to c\bar{c}$ search.

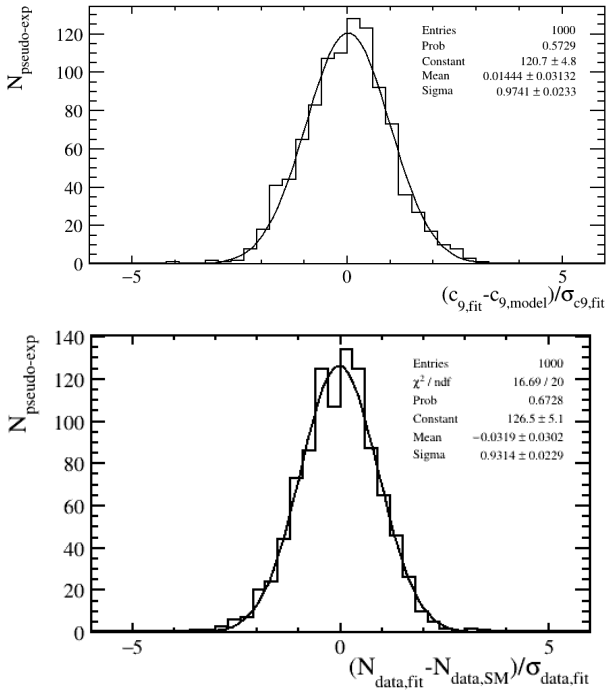

Figure 168: Pull of the coefficients c_9 of the model used to fit pseudo-experiments and the number of data from the fit to pseudo-experiments in the H $\to c\bar{c}$ search.

D Appendix

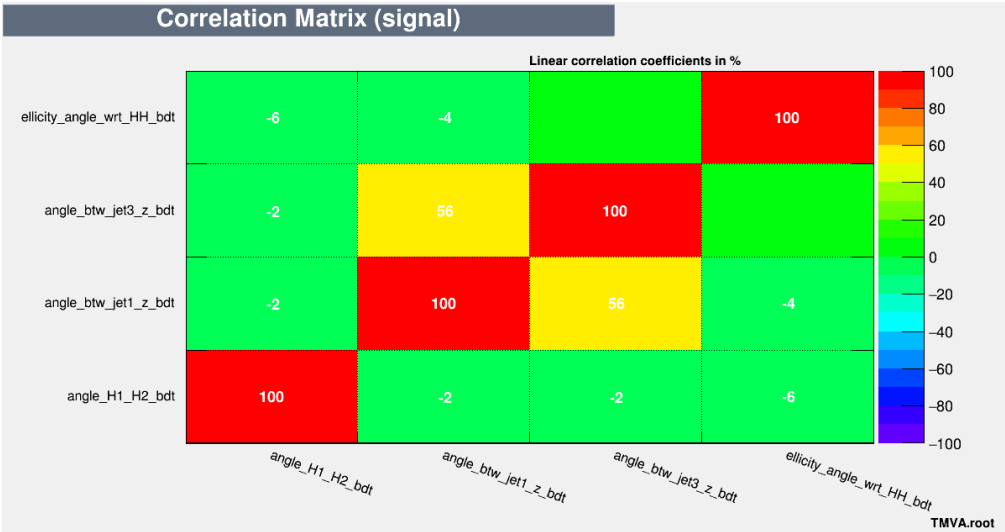

Figure 169: Signal correlation matrix among MLP input variables for HH vs HH trilinear separation.

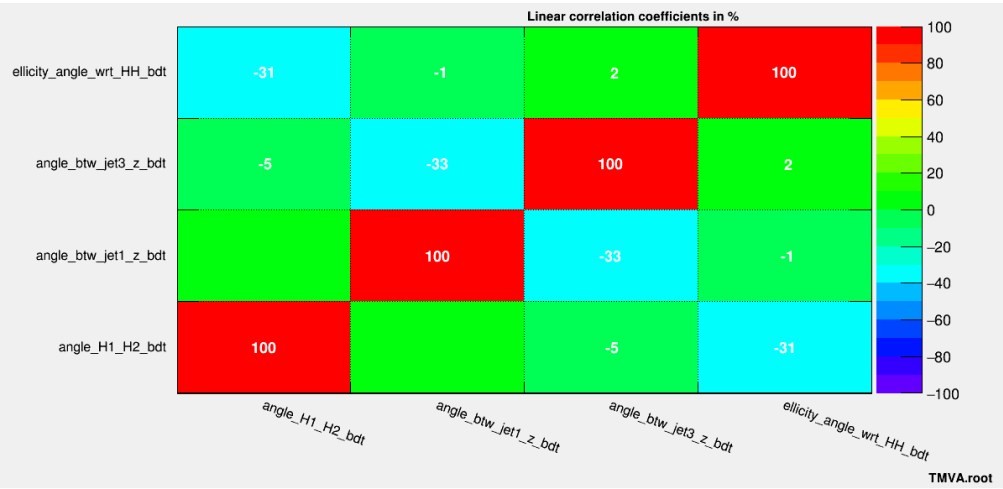

Figure 170: Background correlation matrix among MLP input variables for HH vs HH trilinear separation.

References

[1] Graham Shaw Franz Mandl. *Quantum field theory*. 2nd edition, Cap. 17-18.

[2] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[3] Georges Aad et al. "Combined measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels with the ATLAS detector using $\sqrt{s} = 7$, 8 and 13 TeV pp collision data". In: (Aug. 2023). arXiv: 2308.04775 [hep-ex].

[4] J. de Blas et al. "Higgs Boson Studies at Future Particle Colliders". In: *JHEP* 01 (2020), p. 139. DOI: 10.1007/JHEP01(2020)139. arXiv: 1905.03764 [hep-ph].

[5] LHC Higgs Cross Section Working Group. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG..

[6] J. Alison et al. "Higgs boson potential at colliders: Status and perspectives". In: *Rev. Phys.* 5 (2020). Ed. by Biagio Di Micco et al., p. 100045. DOI: 10.1016/j.revip.2020.100045. arXiv: 1910.00012 [hep-ph].

[7] Hind Al Ali et. al. "The muon Smasher's guide". In: *Reports on Progress in Physics* 85.8 (July 2022), p. 084201. DOI: 10.1088/1361-6633/ac6678. URL: https://dx.doi.org/10.1088/1361-6633/ac6678.

[8] Philipp Gerhard Roloff et al. "Double Higgs boson production and Higgs self-coupling extraction at CLIC". In: *Eur. Phys. J. C* 80.11 (2020). 16 pages, 12 figures, v3 for submission to EPJC, p. 1010. DOI: 10.1140/epjc/s10052-020-08567-7. arXiv: 1901.05897. URL: https://cds.cern.ch/record/2649438.

[9] "LHC Machine". In: *JINST* 3 (2008). Ed. by Lyndon Evans and Philip Bryant, S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[10] Georges 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: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[11] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[12] Georges Aad et al. "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". In: *Nature* 607.7917 (2022). [Erratum: Nature 612, E24 (2022)], pp. 52–59. DOI: 10.1038/s41586-022-04893-w. arXiv: 2207.00092 [hep-ex].

[13] Armen Tumasyan et al. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery." In: *Nature* 607.7917 (2022), pp. 60–68. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043 [hep-ex].

[14] Georges Aad et al. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex].

[15] Morad Aaboud et al. "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 99 (2019), p. 072001. DOI: 10.1103/PhysRevD.99.072001. arXiv: 1811.08856 [hep-ex].

[16] Albert M Sirunyan et al. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector". In: *Phys. Lett. B* 779 (2018), pp. 283–316. DOI: 10.1016/j.physletb.2018.02.004. arXiv: 1708.00373 [hep-ex].

[17] Albert M Sirunyan et al. "Evidence for Higgs boson decay to a pair of muons". In: *JHEP* 01 (2021), p. 148. DOI: 10.1007/JHEP01(2021)148. arXiv: 2009.04363 [hep-ex].

[18] Morad Aaboud et al. "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector". In: *Phys. Lett. B* 786 (2018), pp. 59–86. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex].

[19] A. M. Sirunyan et al. "Observation of Higgs boson decay to bottom quarks". In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[20] Georges Aad et al. "Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 81.6 (2021), p. 537. DOI: 10.1140/epjc/s10052-021-09192-8. arXiv: 2011.08280 [hep-ex].

[21] Aram Hayrapetyan et al. "Measurement of the Higgs boson production via vector boson fusion and its decay into bottom quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: (Aug. 2023). arXiv: 2308.01253 [hep-ex].

[22] Georges Aad et al. "Search for Higgs boson production in association with a high-energy photon via vector-boson fusion with decay into bottom quark pairs at $\sqrt{s}$=13 TeV with the ATLAS detector". In: *JHEP* 03 (2021), p. 268. DOI: 10.1007/JHEP03(2021)268. arXiv: 2010.13651 [hep-ex].

[23] Georges Aad et al. "Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector". In: *Eur. Phys. J. C* 82 (2022), p. 717. DOI: 10.1140/epjc/s10052-022-10588-3. arXiv: 2201.11428 [hep-ex].

[24] Armen Tumasyan et al. "Search for Higgs Boson Decay to a Charm Quark-Antiquark Pair in Proton-Proton Collisions at s=13 TeV". In: *Phys. Rev. Lett.* 131.6 (2023), p. 061801. DOI: 10.1103/PhysRevLett.131.061801. arXiv: 2205.05550 [hep-ex].

[25] Armen Tumasyan et al. "Search for Higgs Boson and Observation of Z Boson through their Decay into a Charm Quark-Antiquark Pair in Boosted Topologies in Proton-Proton Collisions at s=13 TeV". In: *Phys. Rev. Lett.* 131.4 (2023), p. 041801. DOI: 10.1103/PhysRevLett.131.041801. arXiv: 2211.14181 [hep-ex].

[26] Gilad Perez et al. "Constraining the charm Yukawa and Higgs-quark coupling universality". In: *Phys. Rev.* D92 (2015), p. 033016. DOI: 10.1103/PhysRevD.92.033016. arXiv: 1503.00290 [hep-ph].

[27] Roel Aaij et al. "Measurement of differential $b\bar{b}$- and $c\bar{c}$-dijet cross-sections in the forward region of pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 02 (2021), p. 023. DOI: 10.1007/JHEP02(2021)023. arXiv: 2010.09437 [hep-ex].

[28] Roel Aaij et al. "Measurement of differential $b\bar{b}$- and $c\bar{c}$-dijet cross-sections in the forward region of pp collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 02 (2021), p. 023. DOI: 10.1007/JHEP02(2021)023. arXiv: 2010.09437 [hep-ex].

[29] "Search for $H^0 \to b\bar{b}$ or $c\bar{c}$ in association with a W or Z boson in the forward region of pp collisions". In: (2016). URL: https://cds.cern.ch/record/2209531.

[30] Chung Ngoc Leung, S. T. Love, and S. Rao. "Low-Energy Manifestations of a New Interaction Scale: Operator Analysis". In: *Z. Phys. C* 31 (1986), p. 433. DOI: 10.1007/BF01588041.

[31] C. Englert et al. "Precision Measurements of Higgs Couplings: Implications for New Physics Scales". In: *J. Phys. G* 41 (2014), p. 113001. DOI: 10.1088/0954-3899/41/11/113001. arXiv: 1403.7191 [hep-ph].

[32] *Projected sensitivity of Higgs boson pair production combining the $b\bar{b}\gamma\gamma$ and $b\bar{b}\tau^+\tau^-$ final states with the ATLAS detector at the HL-LHC*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2022-005. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2802127.

[33] A. Abada et al. "FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1". In: *Eur. Phys. J. C* 79.6 (2019), p. 474. DOI: 10.1140/epjc/s10052-019-6904-3.

[34] CEPC Study Group Collaboration. "CEPC Conceptual Design Report: Volume 2- Physics & Detector". In: *IHEP-CEPC-DR-2018-02, IHEP-EP-2018-01, IHEP-TH-2018-01* (2019). arXiv: 1811.10545[hep-ex].

[35] CLIC Collaboration CLICdp. "The Compact Linear Collider (CLIC) - 2018 Summary Report". In: CERN Yellow Rep. Monogr. *1802 (2018), 1-19* (2018). DOI: https://doi.org/10.23731/CYRM-2018-002.

[36] P. Bambade et al. *The International Linear Collider: A Global Project*. 2019. arXiv: 1901.09829[hep-ex].

[37] Carlotta Accettura et al. *Towards a Muon Collider*. 2023. arXiv: 2303.08533 [physics.acc-ph].

[38] Matthew Forslund and Patrick Meade. "High precision higgs from high energy muon colliders". In: *JHEP* 08 (2022), p. 185. DOI: 10.1007/JHEP08(2022)185. arXiv: 2203.09425 [hep-ph].

[39] Tao Han et al. "Electroweak couplings of the Higgs boson at a multi-TeV muon collider". In: *Phys. Rev. D* 103.1 (2021), p. 013002. DOI: 10.1103/PhysRevD.103.013002. arXiv: 2008.12204 [hep-ph].

[40] Andreas Papaefstathiou, Gilberto Tetlalmatzi-Xolocotzi, and Marco Zaro. "Triple Higgs boson production to six b-jets at a 100 TeV proton collider". In: *Eur. Phys. J. C* 79.11 (2019), p. 947. DOI: 10.1140/epjc/s10052-019-7457-1. arXiv: 1909.09166 [hep-ph].

[41] Mauro Chiesa et al. "Measuring the quartic Higgs self-coupling at a multi-TeV muon collider". In: (Mar. 2020). arXiv: 2003.13628 [hep-ph].

[42] Antonio Costantini et al. "Vector boson fusion at multi-TeV muon colliders". In: May 2020. arXiv: 2005.10289 [hep-ph].

[43] Lorenzo Sestini et al. "Higgs Physics at a Muon Collider with detailed detector simulation". In: *PoS* ICHEP2022 (2022), p. 515. DOI: 10.22323/1.414.0515.

[44] L. Castelli. *Study of $H \to WW$ reconstruction and coupling precision determination at Muon Collider*. URL: https://hdl.handle.net/20.500.12608/28559.

[45] M. Casarsa. *Higgs physics prospects at Muon Collider with a detailed detector simulation* . URL: https://indico.desy.de/event/34916/contributions/147292/attachments/84406/111835/casarsa_EPS-HEP2023.pdf.

[46] A. Montella. *Study of the Physics Potential of the $H \to \mu\mu$ Direct Decay Channel at a 3 TeV Muon Collider*. URL: Master%E2%80%99s%20Thesis,%20Univ.%20of%20Trieste,%20Trieste,%20Italy%20(2021).

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

[48] H. Wiedemann. *Particle Accelerator Physics*. Berlin : Springer, 4th ed., 2015.

[49] Harms USPAS notes Barletta Spentzouris. URL: https://uspas.fnal.gov/materials/10MIT/Emittance.pdf.

[50] "The CERN accelerator complex in 2019". In: *General Photo* (). Ed. by E. Mobs.

[51] F. Follin and D. Jacquet. "Implementation and experience with luminosity levelling with offset beam". In: (2014). Comments: 5 pages, contribution to the ICFA Mini-Workshop on Beam-Beam Effects in Hadron Colliders, CERN, Geneva, Switzerland, 18-22 Mar 2013, pp. 183–187. DOI: 10.5170/CERN-2014-004.183. arXiv: 1410.3667. URL: http://cds.cern.ch/record/1955354.

[52] Stephane Fartoukh and et al. Kostoglou. *LHC Configuration and Operational Scenario for Run 3*. Tech. rep. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2790409.

[53] The LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005. URL: https://doi.org/10.1088/1748-0221/3/08/s08005.

[54] Roel Aaij et al. "LHCb Detector Performance". In: *Int. J. Mod. Phys. A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex].

[55] R Aaij et al. "Performance of the LHCb Vertex Locator". In: *Journal of Instrumentation* 9.09 (Sept. 2014), P09007. DOI: 10.1088/1748-0221/9/09/P09007. URL: https://dx.doi.org/10.1088/1748-0221/9/09/P09007.

[56] T. Szumlak, C. Parkes, and T. Ruf. "Reconstruction of cluster positions in the LHCb VELO". In: (Dec. 2007).

[57] L B A Hommels. *The LHCb Outer Tracker Detector Design and Production*. Tech. rep. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/828990.

[58] S et al. Amato. *LHCb calorimeters: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/494264.

[59] Eduardo Picatoste Olloqui. "LHCb Preshower(PS) and Scintillating Pad Detector (SPD): commissioning, calibration, and monitoring". In: *J. Phys.: Conf. Ser.* 160 (2009), p. 012046. DOI: 10.1088/1742-6596/160/1/012046. URL: https://cds.cern.ch/record/1293075.

[60] et al Barbosa-Marinho. *LHCb muon system: Technical Design Report*. Technical design report. LHCb. Geneva: CERN, 2001. URL: https://cds.cern.ch/record/504326.

[61] R. Aaij et al. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC". In: *Journal of Instrumentation* 14.04 (Apr. 2019), P04013. DOI: 10.1088/1748-0221/14/04/P04013. URL: https://dx.doi.org/10.1088/1748-0221/14/04/P04013.

[62] *HLT1 lines description*. URL: https://gitlab.cern.ch/lhcb/Hlt/-/tree/hlt2016-patches/Hlt/Hlt1Lines/python/Hlt1Lines?ref_type=heads.

[63] Roel et al. Aaij. "Design and performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC. Performance of the LHCb trigger and full real-time reconstruction in Run 2 of the LHC". In: *JINST* 14.04 (2019). 46 pages, 35 figures, 1 table. All figures and tables are available at https://cern.ch/lhcbproject/Publications/LHCbProjectPublic/LHCb-DP-2019-001.html, P04013. DOI: 10.1088/1748-0221/14/04/P04013. arXiv: 1812.10790. URL: https://cds.cern.ch/record/2652801.

[64] G. Barrand et al. "GAUDI — A software architecture and framework for building HEP data processing applications". In: *Computer Physics Communications* 140.1 (2001). CHEP2000, pp. 45–55. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(01)00254-5. URL: https://www.sciencedirect.com/science/article/pii/S0010465501002545.

[65] 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: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[66] R. Frederix et al. "The automation of next-to-leading order electroweak calculations". In: *JHEP* 07 (2018). [Erratum: JHEP 11, 085 (2021)], p. 185. DOI: 10.1007/JHEP11(2021)085. arXiv: 1804.10017 [hep-ph].

[67] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[68] S. Agostinelli et al. "GEANT4–a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[69] Gloria Corti et al. "Software for the LHCb experiment". In: *IEEE Trans. Nucl. Sci.* 53 (2006), pp. 1323–1328. URL: https://cds.cern.ch/record/913834.

[70] Roel Aaij et al. "The LHCb upgrade I". In: (May 2023). arXiv: 2305.10515 [hep-ex].

[71] I. Belyaev et al. "The history of LHCb". In: *Eur. Phys. J. H* 46.1 (2021), p. 3. DOI: 10.1140/epjh/s13129-021-00002-z. arXiv: 2101.05331 [physics.hist-ph].

[72] Nicola Neri. "The LHCb Upgrade II". In: *PoS* BEAUTY2018 (2018). Ed. by Robert Fleischer et al., p. 057. DOI: 10.22323/1.326.0057.

[73] D Hutchcroft. *VELO Pattern Recognition*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1023540.

[74] O Callot and S Hansmann-Menzemer. *The Forward Tracking: Algorithm and Performance Studies*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1033584.

[75] O Callot. *Downstream Pattern Recognition*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1025827.

[76] O Callot, M Kucharczyk, and M Witek. *VELO-TT track reconstruction*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1027834.

[77] J Van Tilburg. "Track simulation and reconstruction in LHCb". Presented on 01 Sep 2005. 2005. URL: https://cds.cern.ch/record/885750.

[78] R. Frühwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: https://doi.org/10.1016/0168-9002(87)90887-4. URL: https://www.sciencedirect.com/science/article/pii/0168900287908874.

[79] The LHCb collaboration. "Measurement of the track reconstruction efficiency at LHCb". In: *Journal of Instrumentation* 10.02 (Feb. 2015), P02007. DOI: 10.1088/1748-0221/10/02/P02007. URL: https://dx.doi.org/10.1088/1748-0221/10/02/P02007.

[80] LHCb Collaboration. *LHCb PID Upgrade Technical Design Report*. Tech. rep. 2013. URL: https://cds.cern.ch/record/1624074.

[81] NIKITA KAZEEV. *Machine learning for particle identification in the LHCb detector*. URL: https://hdl.handle.net/11573/1448881.

[82] Marcin Kucharczyk, Piotr Morawski, and Mariusz Witek. *Primary Vertex Reconstruction at LHCb*. Tech. rep. Geneva: CERN, 2014. URL: https://cds.cern.ch/record/1756296.

[83] Monica Pepe Altarelli. "LHCb status and early physics prospects". In: *Nuovo Cimento C* 32.5-6 (2009). Comments: 9 pages, 5 figures, proceedings of Les Rencontres de Physique de la Vallee d'Aoste 2009, pp. 193–201. DOI: 10.1393/ncc/i2010-10536-x. arXiv: 0907.0926. URL: https://cds.cern.ch/record/1188876.

[84] P.C. Tsopelas. "A silicon pixel detector for LhCb". English. Exacte Wetenschappen Naam instelling promotie: Vrije Universiteit Amsterdam Naam instelling onderzoek: Vrije Universiteit Amsterdam. PhD thesis. Vrije Universiteit Amsterdam, 2016. ISBN: 9789462334151.

[85] Aras Papadelis. "Characterisation and commissioning of the LHCb VELO detector". Presented on 17 Jun 2009. 2009. URL: http://cds.cern.ch/record/1186697.

[86] William Barter et al. *Jets reconstruction and performances at LHCb*. Tech. rep. Geneva: CERN, 2012. URL: https://cds.cern.ch/record/1446552.

[87] Florian Beaudette. "The CMS Particle Flow Algorithm". In: *International Conference on Calorimetry for the High Energy Frontier*. 2013, pp. 295–304. arXiv: 1401.8155 [hep-ex].

[88] Roel Aaij et al. "Study of forward Z + jet production in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 01 (2014), p. 033. DOI: 10.1007/JHEP01(2014)033. arXiv: 1310.8197 [hep-ex].

[89] Lorenzo Sestini. "Search for a Higgs boson decaying to a pair of b quarks in the forward region of pp collisions with the LHCb detector". Presented 09 Mar 2017. Padua U., 2017. URL: https://cds.cern.ch/record/2302043.

[90] Ryan Atkin. "Review of jet reconstruction algorithms". In: *Journal of Physics: Conference Series* 645.1 (Sept. 2015), p. 012008. DOI: 10.1088/1742-6596/645/1/012008. URL: https://dx.doi.org/10.1088/1742-6596/645/1/012008.

[91] LHCb Jets and EW bosons subgroup. *Jets and EW bosons subgroup*. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/EWWG2.

[92] The LHCb collaboration. "Identification of beauty and charm quark jets at LHCb". In: *Journal of Instrumentation* 10.06 (June 2015), P06013. DOI: 10.1088/1748-0221/10/06/P06013. URL: https://dx.doi.org/10.1088/1748-0221/10/06/P06013.

[93] Emil Bols et al. "Jet Flavour Classification Using DeepJet". In: *JINST* 15.12 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].

[94] *Understanding LSTM Networks*. URL: http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[95] LHC Higgs Cross Sections WG. *Higgs Cross Sections*. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageAt13TeV.

[96] R. Aaij et al. "Measurement of forward W and Z boson production in association with jets in proton-proton collisions at $\sqrt{s}$=8 Tev". In: *JHEP* 05 (2016), p. 131. DOI: 10.1007/JHEP05(2016)131. arXiv: 1605.00951 [hep-ex].

[97] R. Aaij et al. "Measurement of forward top pair production in the dilepton channel in pp collisions at $\sqrt{s}$ =13 Tev". In: *JHEP* 08 (2018), p. 174. DOI: 10.1007/JHEP08(2018)174. arXiv: 1803.05188 [hep-ex].

[98] Particle Data Group et al. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf. URL: https://doi.org/10.1093/ptep/ptaa104.

[99] Thomas Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth. A* 434 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006.

[100] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[101] D. Zuliani. "Measurement of W boson production in association with jets and W W vector boson pair production at the LHCb experiment." In: (2021/2022), Phd thesis.

[102] Roel Aaij et al. "Precision luminosity measurements at LHCb". In: *JINST* 9.12 (2014), P12005. DOI: 10.1088/1748-0221/9/12/P12005. arXiv: 1410.0149 [hep-ex].

[103] *Impact of HCAL removal on Physics measurements* . URL: https://indico.cern.ch/event/1245669/contributions/5276008/attachments/2598891/4487097/HCAL_physics_impact.pdf.

[104] G. I. Budker. *Proceedings of the VII International Conference on High Energy Accelerators*. Yerevan, 1969 (Academy of Sciences of Armenia, Yerevan, 1970), Vol. I,pp. 33–39.

[105] M. A. Palmer. "An Overview of the US Muon Accelerator Program". In: *International Workshop on Beam Cooling and Related Topics*. June 2013.

[106] N.V. Mokhov and S.I. Striganov. "Detector Background at Muon Colliders". In: *Physics Procedia* 37 (Apr. 2012). DOI: 10.1016/j.phpro.2012.03.761.

[107] N. Amapane et al. "LEMMA approach for the production of low-emittance muon beams". In: *Nuovo Cim. C* 42.6 (2020), p. 259. DOI: 10.1393/ncc/i2019-19259-6.

[108] M. Bogomilov et al. "Demonstration of cooling by the Muon Ionization Cooling Experiment". In: *Nature* 578.7793 (2020), pp. 53–59. DOI: 10.1038/s41586-020-1958-9. arXiv: 1907.08562 [physics.acc-ph].

[109] *MICE infographic,* URL: news.fnal.gov/wp-content/uploads/2020/02/mice-infographic-stfc.jpg.

[110] R. P. Johnson et al. "A six-dimensional muon beam cooling experiment". In: *Conf. Proc. C* 060626 (2006). Ed. by Christopher Prior, pp. 2409–2411.

[111] S.A. Bogacz et al. "RECIRCULATING LINEAR ACCELERATORS FOR FUTURE MUON FACILITIES". In: *Proceedings of IPAC'10, Kyoto, Japan* (2012), THOAMH01.

[112] J. Eldred, V. Lebedev, and A. Valishev. "Rapid-cycling synchrotron for multi-megawatt proton facility at Fermilab". In: *Journal of Instrumentation* 14.07 (July 2019), P07021–P07021. DOI: 10.1088/1748-0221/14/07/p07021.

[113] D. J. SUMMERS. *MUON ACCELERATION USING FIXED FIELD, ALTERNATING-GRADIENT (FFAG) RINGS*. International publisher of Modern Physics A.

[114] Francesco Collamati et al. "Advanced assessment of beam-induced background at a muon collider". In: *JINST* 16.11 (2021), P11009. DOI: 10.1088/1748-0221/16/11/P11009. arXiv: 2105.09116 [physics.acc-ph].

[115] *FLUKA*. URL: https://fluka.cern.

[116] *Fluka code* . http://www.fluka.org/fluka.php .

[117] N.V. Mokhov and S.I. Striganov. "Detector Backgrounds at Muon Colliders". In: *Physics Procedia* 37 (2012). Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011), pp. 2015–2022. ISSN: 1875-3892. DOI: https://doi.org/10.1016/j.phpro.2012.03.761. URL: https://www.sciencedirect.com/science/article/pii/S187538921201927X.

[118] Nazar Bartosik et al. "Preliminary Report on the Study of Beam-Induced Background Effects at a Muon Collider". In: (May 2019). arXiv: 1905.03725 [hep-ex].

[119] N. Bartosik. *Beam background simulation with the CLIC framework*. URL: https://indico.cern.ch/event/876797/contributions/3706171/attachments/1974853/3286503/2020_01_23_bartosik_v0.pdf.

[120] *Machine-Detector interface for multi-TeV Muon Collider*. URL: https://agenda.infn.it/event/38110/contributions/215945/attachments/113078/161774/mdi_frascati.pdf.

[121] Massimo Casarsa, Donatella Lucchesi, and Lorenzo Sestini. "Experimentation at a muon collider". In: (Nov. 2023). DOI: 10.1146/annurev-nucl-102622-011319. arXiv: 2311.03280 [hep-ex].

[122] D Calzolari et al. "Lattice and detector studies for the MDI of a 10 TeV muon collider". In: *JACoW IPAC* 2023 (2023), MOPA090. DOI: 10.18429/JACoW-IPAC2023-MOPA090. URL: http://cds.cern.ch/record/2884558.

[123] Daniele Calzolari and Kyriacos Skoufaris. "Machine-detector interface studies for a multi-TeV muon collider". In: *PoS* ICHEP2022 (2022), p. 063. DOI: 10.22323/1.414.0063.

[124] *ILCSoft*. URL: https://github.com/iLCSoft.

[125] Dominik Arominski et al. "A detector for CLIC: main parameters and performance". In: (Dec. 2018). arXiv: 1812.07337 [physics.ins-det].

[126] C. Gatto et al. *The IlcRoot FrameWork*. URL: http://www.dmf.unisalento.it/~danieleb/IlcRoot/.

[127] Niloufar Alipour Tehrani et al. "CLICdet: The post-CDR CLIC detector model". In: (2017), CLICdp-Note-2017–001.

[128] *DD4Hep*. URL: https://dd4hep.web.cern.ch/dd4hep/.

[129] Nazar Bartosik et al. "Full Detector Simulation with Unprecedented Background Occupancy at a Muon Collider". In: *Comput. Softw. Big Sci.* 5.1 (2021), p. 21. DOI: 10.1007/s41781-021-00067-x.

[130] *Detector Design using BIB simulation data*. URL: https://indico.cern.ch/event/1175126/contributions/5024030/.

[131] G. Da Molin. *Study of b- and c- jets identification for Higgs coupling measurement at muon collider*. URL: https://hdl.handle.net/20.500.12608/3238.

[132] E. Brondolin et al. "Conformal tracking for all-silicon trackers at future electron–positron colliders". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 956 (Dec. 2019), p. 163304. DOI: 10.1016/j.nima.2019.163304.

[133] Xiaocong Ai et al. "A Common Tracking Software Project". In: *Comput. Softw. Big Sci.* 6.1 (2022), p. 8. DOI: 10.1007/s41781-021-00078-8. arXiv: 2106.13593 [physics.ins-det].

[134] M.A. Thomson J.S. Marshall. *The Pandora Software Development Kit for Pattern Recognition*. 2015. DOI: https://doi.org/10.1140/epjc/s10052-015-3659-3.

[135] Matteo Cacciari et al. *FastJet user manual*. CERN-PH-TH/2011-297, 2011.

[136] M. Thomson. "Particle Flow Calorimetry and the PandoraPFA Algorithm". In: *Nuclear Instruments and Methods in Physics Research Section A Accelerators Spectrometers Detectors and Associated Equipment* 611 (July 2009). DOI: 10.1016/j.nima.2009.09.009.

[137] L. Buonincontri. *Study of mitigation strategies of beam induced background and Higgs boson couplings measurements at a muon collider*. URL: https://hdl.handle.net/20.500.12608/22861.

[138] PandoraPFA repository. URL: https://github.com/PandoraPFA.

[139] S. Catani et al. “Longitudinally invariant K_t clustering algorithms for hadron hadron collisions”. In: *Nucl. Phys. B* 406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M.

[140] N. Bartosik et al. “Simulated Detector Performance at the Muon Collider”. In: (Mar. 2022). arXiv: 2203.07964 [hep-ex].

[141] Alexander Conway et al. “Measuring the Higgs Self-Coupling Constant at a Multi-TeV Muon Collider”. In: (May 2014). arXiv: 1405.5910 [hep-ex].

[142] M. Cepeda et al. “Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC”. In: *CERN Yellow Rep. Monogr.* 7 (2019). Ed. by Andrea Dainese et al., pp. 221–584. DOI: 10.23731/CYRM-2019-007.221. arXiv: 1902.00134 [hep-ph].

[143] *CMS Collaboration*. URL: https://cms-results.web.cern.ch/cms-results/public-results/publications/HIG-21-008/index.html.

[144] Andrey Pozdnyakov. “Constraints on the Higgs-charm coupling by CMS”. In: *PoS* ICHEP2022 (2022), p. 505. DOI: 10.22323/1.414.0505.

[145] *F-distribution*
. https://en.wikipedia.org/wiki/F-distribution.