Measurement of Higgs boson hadronic decays at FCC-ee

Andrea Del Vecchio[1,2], Jan Eysermans[5], Loukas Gouskos[2], George Iakovidis[3], Alexis Maloizel[4], Giovanni Marchiori[4] and Michele Selvaggi[2]

[1]La Sapienza, Roma, Italy.
[2]European Organization for Nuclear Research (CERN), Geneva, Switzerland.
[3]Brookhaven National Laboratory, Upton, NY, USA.
[4]Laboratoire AstroParticule et Cosmologie, (CNRS/IN2P3), Paris, France.
[5]Massachusetts Institute of Technology, Boston, USA.

Contributing authors: andrea.del.vecchio@cern.ch; jan.eysermans@cern.ch; loukas.gouskos@cern.ch; george.iakovidis@cern.ch; alexis.maloizel@ens-paris-saclay.fr; giovanni.marchiori@cern.ch; michele.selvaggi@cern.ch;

Abstract

The sensitivity of the measurement of the Higgs boson branching ratios to $b\bar{b}$, $c\bar{c}$, $s\bar{s}$ and gg final states with a modified version of the IDEA detector concept at the Future Circular Collider is studied using simulated event samples of the main signal and background processes. The study assumes four identical detectors leading to an integrated luminosity of $\mathbf{10.8\,ab^{-1}}$ of e^+e^- collisions with a centre-of-mass energy $\sqrt{s} = \mathbf{240\,GeV}$, and to an integrated luminosity of $\mathbf{3.0\,ab^{-1}}$ of e^+e^- collisions with a centre-of-mass energy $\sqrt{s} = \mathbf{365\,GeV}$, The signal consists mainly of $e^+e^- \to ZH$ Higgs-strahlung events followed by one of the four Higgs boson decays under study. The study considers Z boson decays to pairs of leptons, either charged (e^+e^-, $\mu^+\mu^-$) or neutral ($\nu\bar{\nu}$), or to a quark-antiquark pair ($b\bar{b}$, $c\bar{c}$, $s\bar{s}$ and $q\bar{q}$, $q = u, d$). Vector boson fusion ($ee \to eeZZ \to eeH$, $ee \to \nu_e\nu_eWW \to \nu_e\nu_eH$) also contributes to the final states under study. Its impact is estimated separately from that of ZH events.

1 Introduction

In 2012, using the data collected during the Run 1 of the CERN Large Hadron Collider (LHC), the ATLAS and CMS experiments discovered a new particle H [1, 2], with a mass m_H around 125 GeV The properties of that particle were found to be consistent with those predicted by the Standard Model (SM) of particle physics for a Higgs boson [3, 4] of the same mass [5].

While the Higgs boson was observed mainly through its decays to di-boson final states ($\gamma\gamma$, ZZ, WW), decays to $q\bar{q}$ (mainly $q = b$) or gg pairs, leading to copious hadron production, represent the most probable final states for a Higgs boson with $m_H = 125$ GeV in the SM. The branching ratios (B) predicted by the SM for $m_H = 125$ GeV are $B(H \to b\bar{b}) = 58.24\%$, $B(H \to c\bar{c}) = 2.89\%$, $B(H \to gg) = 8.19\%$, summing up to a total of 69.32%, *i.e.* more than two thirds of the total. The branching ratio $B(H \to s\bar{s})$ is expected to be around 2.4×10^{-4}, with a large uncertainty related to the limited knowledge of the mass of the strange quark.

Measuring with high precision these branching ratios is therefore mandatory in order to gain better insights about the Higgs particle and to verify whether its properties agree with the the predictions of the Standard Model. In particular, the measurements of these branching ratios are necessary for the determination of some fundamental SM parameters such as the couplings of the Higgs boson to bottom, charm and strange quarks (induced by tree-level diagrams), as well as its effective coupling to gluons (induced by loop-mediated amplitudes). Physics beyond the SM could enhance such couplings significantly. For instance [6], simplified models featuring two vector-like quark states could at the same time be in agreement with current constraints from electroweak precision and higgs data, and lead to first-generation quark Yukawa couplings enhanced by several hundred times their Standard Model value, while the Higgs couplings to charm (strange) quarks could be increased by factors of a few (few tens).

With the larger dataset provided by the LHC Run 2, both the ATLAS and CMS collaborations observed the $H \to b\bar{b}$ decays [7, 8], while no evidence of the rarer $H \to c\bar{c}$ decays has been found in dedicated searches [9, 10], and no attempt to measure the $H \to gg$ or $H \to s\bar{s}$ branching ratios has yet been performed. These results underline the difficulty of the LHC experiments to measure with precision the hadronic branching ratios of the Higgs boson for various reasons.

Firstly, despite the large Higgs boson production cross section in the LHC proton–proton (pp) collisions (around 55 pb at a centre-of-mass energy $\sqrt{s} = 13$ TeV), the signal-to-background ratio is extremely small. This is due to the fact that Higgs bosons are mainly produced at the LHC (about 94% of the total) by processes (gluon–gluon or vector-boson fusion) that lead to purely hadronic final states, that can also be produced by strong interactions of the colliding protons with cross sections that are larger by several orders of magnitude. For this reason, the most sensitive processes at the LHC for the identification of hadronic Higgs boson decays are Higgs-strahlung events, $pp \to WH, ZH$ followed by a leptonic decay of the W or Z boson. This corresponds to less than 1% of the total amount of Higgs bosons produced and reduces significantly the signal yield, affecting in particular the searches for rarer decays such as $H \to gg$ and $H \to c\bar{c}$.

Secondly, the LHC detectors, while providing excellent performance for the identification ("tagging") of jets of hadrons from b-quarks ("b-jets") in samples including also jets from gluons or light and charm quarks, were not designed for efficient c-, s- and g-jet tagging. While the algorithms used for b-tagging can, to some extent, be

re-tuned in order to identify c-jets, the purity of such algorithms is rather small. In addition, g-tagging at the LHC seems extremely challenging, though recent efforts seem to indicate some potential for quark/gluon discrimination [11]. Identification of jets induced by the hadronisation of strange quarks is also unfeasible at the LHC.

Even with an approximately 20 times larger dataset than the Run 2 one, corresponding to 3000 fb^{-1} of pp collisions to be collected at the high-luminosity phase of the LHC by around 2040, the perspectives for the measurement of the Higgs boson hadronic branching ratios by ATLAS and CMS are quite limited [12]: combining the results of the two experiments, the projected uncertainty on $B(H \to b\bar{b})$ is 4.4%, while sensitivity to $B(H \to c\bar{c})$ might just reach the SM value, and there is yet no demonstration of any potential significant constraint on $B(H \to gg)$ nor on $B(H \to s\bar{s})$[1] It should be noted, in addition, that these measurements rely on the assumption that the production cross sections assume the SM predicted values, thus introducing a certain degree of model dependence in the results.

An extremely high-luminosity e^+e^- collider operating at centre of mass energies around 240 GeV or above, such as the FCC-ee [14], the proposed first stage of the Future Circular Collider, could produce a large number of Higgs bosons in a clean environment, without pile-up collisions, underlying event, and much larger signal-to-background ratio than at the LHC. The Higgs bosons would be mainly produced through the Higgsstrahlung process ($e^+e^- \to ZH$), whose cross-section σ_{ZH} as a function of $\sqrt{s}$ rises sharply for $\sqrt{s} > m_H + m_Z \approx 216$ GeV, and peaks at a value greater than 200 fb for $\sqrt{s} \approx 255$ GeV. Since the instantaneous luminosity $\mathcal{L}$ of a circular collider such as the FCC is expected instead to decrease steeply with $\sqrt{s}$, the maximum value of the production rate $\mathcal{L}\sigma_{ZH}$ of ZH events would be obtained for $\sqrt{s} = 240$ GeV, where $\sigma_{ZH} \approx 200$ fb. Around two million ZH events would thus be produced with an integrated luminosity $\mathcal{L} = 10.8$ ab^{-1}, expected to be collected in three years with four detectors located at four interaction points (IPs) along the FCC-ee ring. An additional five years of data taking at an e^+e^- center-of-mass energy around 365 GeV, above the $t\bar{t}$ production threshold, as planned in the nominal FCC-ee running scenario, would lead to about 330 thousand Higgs boson decays produced through Higgsstrahlung for an integrated luminosity of about 3.0 ab^{-1}, given $\sigma_{ZH} \approx 125$ fb.

A further advantage of such a "Higgs boson factory", compared to the LHC, would be the possibility to measure the total Higgs boson production cross-section, through the use of the recoil mass technique as described later, without looking at a specific Higgs boson decay. Reconstructing each boson decay separately, the corresponding branching ratio could then be measured, and a joint analysis of these results would yield, in a model-independent way, the Higgs boson width and its couplings to the various elementary particles.

Finally, the progress on the particle detection techniques of the past decades after the design of the LHC detectors was finalised and construction was started would allow building next-generation devices that can perform efficient tagging of jets produced by the hadronisation of gluons, charm quarks and maybe even strange quarks.

[1] A recent ATLAS update [13] indicates that these uncertainties could be about 15% better than in the previous extrapolation, but still far from the FCC-ee potential highlighted in the following.

1.1 Outline

This note describes a study of the sensitivity of the measurement of Higgs hadronic branching ratios using a modified version of the IDEA detector concept [14, 15] at FCC-ee, based on an integrated luminosity of 10.8 (3.0) ab^{-1} of e^+e^- collisions at $\sqrt{s} = 240$ (365) GeV. These numbers correspond to the luminosities to be collected by four detectors in the FCC-ee run plan; it is assumed here that the detectors will have similar performance to that of the IDEA detector being simulated.

The target production mode of the Higgs boson is in association with a Z boson, decaying either to charged lepton pairs (either electrons or muons; decays to tau leptons are ignored), to $\nu\bar{\nu}$, or to quark-antiquark pairs of any flavour that is kinematically allowed.

Vector boson fusion also contributes to the final states under study in the case of Z bosons reconstructed to ee (Z boson fusion: $ee \to eeH$) or to $\nu\nu$ (W boson fusion: $WW \to \nu_e\nu_e H$). The analyses is performed in two ways. First is done inclusively were it does not separate the contribution to the signal from ZH and vector-boson-fusion; the selection largely suppresses the latter, and only inclusive signal strengths, determined as the ratio between observed and expected signal yields, inclusive over Higgs boson production mode, are estimated. The second way is performed in ZH and the vector-boson-fusion enriched processes and determines separate signal strengths for the ZH and the vector-boson-fusion processes. This is done at least for the $\nu\nu H$ analysis, as it is sensitive to different Higgs boson couplings (ZH and Z-boson fusion amplitudes are proportional to the HZZ coupling strength modifier κ_Z while the W-boson fusion amplitude is proportional to the HWW coupling strength modifier κ_W).

The analyses proceed in the following way:

- an event selection is applied to identify events that are consistent with the signature of the final state (Z and H boson decays) under study, without distinguishing between the different hadronic Higgs boson decays;
- events are classified into mutually orthogonal categories that are enriched in one of the different Higgs decays;
- a simultaneous fit to one or more observables discriminating between signal and background in each of the event categories is performed in order to determine the signal strengths in each Higgs decay channel.

The note is organised as follows. Section 2 summarises briefly the main features of the detector concept. The simulated signal and background event samples used for this study are described in Section 3. Section 4 illustrates how particle candidates and events are reconstructed. The following four sections describe the analyses at $\sqrt{s} = 240$ GeV and $\sqrt{s} = 365$ GeV: the analysis of the $Z(\ell\ell)H(jj)$ final state is presented in Sections 5; in the Sections 6 and 7 two independent analyses of the $Z(\nu\bar{\nu})H(jj)$ final states are described; section 8 details the analysis of the $Z(jj)H(jj)$ final state. The combination of the analyses of the three final states is given in section 9.

2 The detector concept

The results presented in this study are based on the latest projections of the performance of a modified version of the IDEA detector concept [14]. The detector, with

overall height and length of about 11 m and 13 m respectively, would consist of several nested sub-detectors arranged around one of the FCC-ee interaction points.

The charged-particle tracking system would include a silicon-pixel vertex detector, surrounded by a low-mass drift chamber (total amount of material traversed by particles emitted in the transverse plane about 1.6% X_0) and an external layer of silicon micro-strip detectors. The tracking detectors are immersed in a uniform, axial 2 T magnetic field parallel to the beam direction, generated by a superconducting solenoid. Compared to the initial detector concept, the updated design used in the event samples that have been simulated for this study assumes a smaller beam-pipe with a radius of 1 cm and a thickness of 0.67% X_0, an innermost layer of the vertex detector that is closer to the beam ($R = 1.2$ cm instead of 1.7 cm), and three single layers of silicon sensors rather than two double-layers for the inner disks.

Outside the tracker, a preshower based on lead and Micro Pattern Gas Detector (MPGD), surrounded by a calorimeter (composed of electromagnetic and hadronic compartments) would allow the measurement of the position and energy of electromagnetic and hadronic showers and thus the reconstruction of photons, electrons, and charged and neutral hadrons. The simulation includes a proposed modification of the baseline IDEA detector design [15], in which a dual-readout crystal electromagnetic calorimeter is added upstream of a dual-readout hadronic calorimeter based on lead-fibres, and the (thin) solenoid is located between the electromagnetic and the hadronic calorimeters, to avoid spoiling the excellent energy resolution for electromagnetic showers.

Finally, a system consisting of layers of μ-Rwell chambers embedded in the magnet return yoke would surround the other systems and provide an efficient and cost-effective solution for the detection of muons that are not absorbed by the inner sub-detectors.

3 Simulated event samples

Monte Carlo event samples of the main signal and background processes have been generated with either the WHIZARD [16] or Pythia 8 [17] event generators and passed through a parametric simulation of the detector response implemented in DELPHES [18–20]. The Higgs boson mass is set to 125 GeV. MC samples were produced centrally as part of the `winter2023` campaign.

Events have been generated at both e^+e^- center-of-mass energies of 240 GeV and 365 GeV. The samples are normalised, using their theoretical cross-sections, to an integrated luminosity of 10.8 (3.0) ab^{-1} at 240 (365) GeV. Halving the number of IPs and corresponding detectors, the integrated luminosity in the same running time and thus the Higgs boson yield would decrease by a factor 1.7.

Electrons and positrons collide head-on with the same nominal energy $\sqrt{s}/2$ GeV. A random gaussian smearing of their energy with a given width (222 MeV at 240 GeV) is performed to simulate the momentum spread of the particles in the beam bunches. The nominal IP corresponds with the nominal centre of the detector, which is also the origin of the detector coordinate system[2]. The simulated beamspot has a finite size (for events generated at 240 GeV it has an elongation along the z axis of

[2] A right-handed coordinate system is used, with its origin at the nominal interaction point in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the ring, and the y-axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the azimuthal angle around the z-axis. The pseudorapidity is defined in terms of the polar angle θ as $\eta = -\ln\tan(\theta/2)$.

0.64 mm, while its RMS sizes along the x and y axes are much smaller, 9.8 μm and 25.4 nm).

The processes that have been simulated, their cross-sections, the number of events generated and the corresponding equivalent luminosity are given in Table 1 for $\sqrt{s} =$ 240 GeV and Table 2 for $\sqrt{s} = 365$ GeV. The numbers for the signal are calculated using values of the cross-sections from Whizard and of branching ratios from the LHC Higgs working group (with the exception of $B(H \to s\bar{s})$ for which there is no recent public estimate). The Higgs boson branching ratios are assumed to be:

- $B(H \to b\bar{b}) = 58.24\%$
- $B(H \to gg) = 8.187\%$
- $B(H \to c\bar{c}) = 2.891\%$
- $B(H \to s\bar{s}) = 0.024\%$
- $B(H \to WW) = 21.37\%$
- $B(H \to ZZ) = 2.619\%$
- $B(H \to \tau\tau) = 6.272\%$

The Higgs boson cross-sections at 240 GeV are assumed to be:

- $\sigma(eeH) = 7.17\,\text{fb}$
- $\sigma(\mu\mu H) = 6.76\,\text{fb}$
- $\sigma(\nu\bar{\nu}H) = 46.2\,\text{fb}$
- $\sigma(q\bar{q}H) = 136.35\,\text{fb}$

The Higgs boson cross-sections at 365 GeV are assumed to be:

- $\sigma(eeH) = 7.39\,\text{fb}$
- $\sigma(\mu\mu H) = 4.19\,\text{fb}$
- $\sigma(\nu\bar{\nu}H) = 53.94\,\text{fb}$
- $\sigma(q\bar{q}H) = 84.36\,\text{fb}$

Process	σ [fb]	N_{gen}	L_{gen} [fb^{-1}]	L_{gen}/L
$\nu\bar{\nu}H(b\bar{b})$	26.900000000	1200000	44610	4.131
$\nu\bar{\nu}H(c\bar{c})$	1.335000000	1100000	823970	76.294
$\nu\bar{\nu}H(s\bar{s})$	0.011090000	1008052	90897385	8416.425
$\nu\bar{\nu}H(gg)$	3.782000000	1055845	279176	25.850
$\nu\bar{\nu}H(\tau\tau)$	2.897000000	1200000	414222	38.354
$\nu\bar{\nu}H(WW)$	9.940000000	1200000	120724	11.178
$\nu\bar{\nu}H(ZZ)$	1.220000000	1200000	983607	91.075
$\nu\bar{\nu}H(u\bar{u})$	0.000004158	4900000	1178451178451	109115849.857
$\nu\bar{\nu}H(d\bar{d})$	0.000009702	4979640	513259121831	47523992.762
$\nu\bar{\nu}H(b\bar{d})$	1000.000000000	1000000	1000	0.093
$\nu\bar{\nu}H(b\bar{s})$	1000.000000000	1000000	1000	0.093
$\nu\bar{\nu}H(s\bar{d})$	1000.000000000	1000000	1000	0.093
$\nu\bar{\nu}H(c\bar{u})$	1000.000000000	1000000	1000	0.093
$eeH(b\bar{b})$	4.171000000	400000	95900	8.880
$eeH(c\bar{c})$	0.207000000	400000	1932367	178.923

$eeH(s\bar{s})$	0.001718000	352836	205376019	19016.298
$eeH(gg)$	0.586300000	400000	682245	63.171
$eeH(\tau\tau)$	0.449100000	400000	890670	82.469
$eeH(WW)$	1.541000000	400000	259572	24.034
$eeH(ZZ)$	0.189100000	400000	2115283	195.860
$eeH(u\bar{u})$	0.000000645	400000	620422580367	57446535.219
$eeH(d\bar{d})$	0.000001504	400000	265895392764	24619943.774
$eeH(b\bar{d})$	1000.000000000	450000	450	0.042
$eeH(b\bar{s})$	1000.000000000	450000	450	0.042
$eeH(s\bar{d})$	1000.000000000	450000	450	0.042
$eeH(c\bar{u})$	1000.000000000	450000	450	0.042
$\mu\mu H(b\bar{b})$	3.940000000	300000	76142	7.050
$\mu\mu H(c\bar{c})$	0.195600000	400000	2044990	189.351
$\mu\mu H(s\bar{s})$	0.001624000	400000	246305419	22806.057
$\mu\mu H(gg)$	0.553800000	400000	722282	66.878
$\mu\mu H(\tau\tau)$	0.424300000	400000	942729	87.290
$\mu\mu H(WW)$	1.456000000	400000	274725	25.438
$\mu\mu H(ZZ)$	0.178600000	400000	2239642	207.374
$\mu\mu H(u\bar{u})$	0.000000609	400000	656797614761	60814593.959
$\mu\mu H(d\bar{d})$	0.000001421	400000	281484696662	26063397.839
$\mu\mu H(b\bar{d})$	1000.000000000	450000	450	0.042
$\mu\mu H(b\bar{s})$	1000.000000000	450000	450	0.042
$\mu\mu H(s\bar{d})$	1000.000000000	450000	450	0.042
$\mu\mu H(c\bar{u})$	1000.000000000	450000	450	0.042
$q\bar{q}H(b\bar{b})$	31.070000000	500000	16093	1.490
$q\bar{q}H(c\bar{c})$	1.542000000	200000	129702	12.009
$q\bar{q}H(s\bar{s})$	0.012800000	400000	31250000	2893.519
$q\bar{q}H(gg)$	4.367000000	400000	91596	8.481
$q\bar{q}H(\tau\tau)$	3.346000000	200000	59773	5.535
$q\bar{q}H(WW)$	11.480000000	1100000	95819	8.872
$q\bar{q}H(ZZ)$	1.409000000	1200000	851668	78.858
$s\bar{s}H(b\bar{b})$	17.450000000	200000	11461	1.061
$s\bar{s}H(c\bar{c})$	0.866100000	300000	346380	32.072
$s\bar{s}H(s\bar{s})$	0.007190000	300000	41724618	3863.391
$s\bar{s}H(gg)$	2.453000000	400000	163066	15.099
$s\bar{s}H(\tau\tau)$	1.879000000	400000	212879	19.711
$s\bar{s}H(WW)$	6.447000000	1200000	186133	17.235
$s\bar{s}H(ZZ)$	0.791200000	600000	758342	70.217
$c\bar{c}H(b\bar{b})$	13.590000000	200000	14717	1.363
$c\bar{c}H(c\bar{c})$	0.674700000	400000	592856	54.894
$c\bar{c}H(s\bar{s})$	0.005607000	300000	53504548	4954.125
$c\bar{c}H(gg)$	1.911000000	400000	209314	19.381
$c\bar{c}H(\tau\tau)$	1.464000000	400000	273224	25.299
$c\bar{c}H(WW)$	5.023000000	1200000	238901	22.120
$c\bar{c}H(ZZ)$	0.616400000	1200000	1946788	180.258
$b\bar{b}H(b\bar{b})$	17.450000000	100000	5731	0.531
$b\bar{b}H(c\bar{c})$	0.866400000	400000	461681	42.748
$b\bar{b}H(s\bar{s})$	0.007193000	400000	55609620	5149.039
$b\bar{b}H(gg)$	2.454000000	200000	81500	7.546
$b\bar{b}H(\tau\tau)$	1.880000000	400000	212766	19.701

$b\bar{b}H(WW)$	6.450000000	1000000	155039	14.355
$b\bar{b}H(ZZ)$	0.791500000	1000000	1263424	116.984
$\nu_e\bar{\nu}_e Z$	33.274000000	2000000	60107	5.565
WW	16438.500000000	373375386	22713	2.103
ZZ	1358.990000000	56162093	41326	3.827
$Z/\gamma^*(q\bar{q})$	52653.900000000	100559248	1910	0.177
$Z/\gamma^*(ee)$	8305.000000000	85400000	10283	0.952
$Z/\gamma^*(\mu\mu)$	5288.000000000	53400000	10098	0.935

Table 1: For each simulated signal and background process, the theoretical cross-section σ, the number of generated events N_{gen}, the equivalent luminosity $L_{\text{gen}} = N_{\text{gen}}/\sigma$ and its ratio to the nominal luminosity $L = 10.8$ ab^{-1} for FCC-ee at $\sqrt{s} = 240$ GeV.

Process	σ [fb]	N_{gen}	L_{gen} [fb^{-1}]	L_{gen}/L
$\nu\bar{\nu}H(b\bar{b})$	31.430000000	1200000	38180	12.727
$\nu\bar{\nu}H(c\bar{c})$	1.560000000	1200000	769231	256.410
$\nu\bar{\nu}H(s\bar{s})$	0.010790000	1200000	111214087	37071.362
$\nu\bar{\nu}H(gg)$	4.418000000	1200000	271616	90.539
$\nu\bar{\nu}H(\tau\tau)$	3.385000000	1200000	354505	118.168
$\nu\bar{\nu}H(WW)$	11.610000000	900000	77519	25.840
$\nu\bar{\nu}H(ZZ)$	1.425000000	1200000	842105	280.702
$eeH(b\bar{b})$	4.303000000	1200000	278875	92.958
$eeH(c\bar{c})$	0.213600000	900000	4213483	1404.494
$eeH(s\bar{s})$	0.001478000	1122800	759675237	253225.079
$eeH(gg)$	0.604900000	1200000	1983799	661.266
$eeH(\tau\tau)$	0.463400000	1200000	2589555	863.185
$eeH(WW)$	1.590000000	1100000	691824	230.608
$eeH(ZZ)$	0.195100000	1200000	6150692	2050.231
$\mu\mu H(b\bar{b})$	2.438000000	1000000	410172	136.724
$\mu\mu H(c\bar{c})$	0.121000000	1100000	9090909	3030.303
$\mu\mu H(s\bar{s})$	0.000837100	1000000	1194600406	398200.135
$\mu\mu H(gg)$	0.342600000	900000	2626970	875.657
$\mu\mu H(\tau\tau)$	0.262500000	900000	3428571	1142.857
$\mu\mu H(WW)$	0.900700000	1100000	1221272	407.091
$\mu\mu H(ZZ)$	0.110500000	800000	7239819	2413.273
$q\bar{q}H(b\bar{b})$	19.220000000	1200000	62435	20.812
$q\bar{q}H(c\bar{c})$	0.954000000	1100000	1153040	384.347
$q\bar{q}H(s\bar{s})$	0.006599000	1100000	166691923	55563.974
$q\bar{q}H(gg)$	2.701000000	1100000	407257	135.752
$q\bar{q}H(\tau\tau)$	2.070000000	1200000	579710	193.237
$q\bar{q}H(WW)$	7.101000000	1100000	154908	51.636
$q\bar{q}H(ZZ)$	0.871500000	1200000	1376936	458.979
$s\bar{s}H(b\bar{b})$	10.800000000	1200000	111111	37.037
$s\bar{s}H(c\bar{c})$	0.535900000	900000	1679418	559.806
$s\bar{s}H(s\bar{s})$	0.003708000	1200000	323624595	107874.865
$s\bar{s}H(gg)$	1.518000000	1200000	790514	263.505

$s\bar{s}H(\tau\tau)$	1.163000000	1200000	1031814	343.938
$s\bar{s}H(WW)$	3.989000000	1000000	250689	83.563
$s\bar{s}H(ZZ)$	0.489600000	1100000	2246732	748.911
$c\bar{c}H(b\bar{b})$	8.407000000	900000	107054	35.685
$c\bar{c}H(c\bar{c})$	0.417300000	1100000	2635993	878.664
$c\bar{c}H(s\bar{s})$	0.002887000	1100000	381018358	127006.119
$c\bar{c}H(gg)$	1.182000000	1200000	1015228	338.409
$c\bar{c}H(\tau\tau)$	0.905400000	1200000	1325381	441.794
$c\bar{c}H(WW)$	3.107000000	1200000	386225	128.742
$c\bar{c}H(ZZ)$	0.381300000	1000000	2622607	874.202
$b\bar{b}H(b\bar{b})$	10.710000000	1200000	112045	37.348
$b\bar{b}H(c\bar{c})$	0.531600000	1200000	2257336	752.445
$b\bar{b}H(s\bar{s})$	0.003678000	1200000	326264274	108754.758
$b\bar{b}H(gg)$	1.506000000	1200000	796813	265.604
$b\bar{b}H(\tau\tau)$	1.153000000	1000000	867303	289.101
$b\bar{b}H(WW)$	3.957000000	1200000	303260	101.087
$b\bar{b}H(ZZ)$	0.485700000	1000000	2058884	686.295
$\nu_e\bar{\nu}_e Z$	126.240000000	1400000	11090	3.697
WW	10716.500000000	11754213	1097	0.366
ZZ	642.800000000	11470944	17845	5.948
$Z/\gamma^*(q\bar{q})$	21414.900000000	6000000	280	0.093
$Z/\gamma^*(ee)$	1527.000000000	3000000	1965	0.655
$Z/\gamma^*(\mu\mu)$	2285.800000000	6600000	2887	0.962
$t\bar{t}$	800.000000000	2700000	3375	1.125

Table 2: For each simulated signal and background process, the theoretical cross-section σ, the number of generated events N_{gen}, the equivalent luminosity $L_{\text{gen}} = N_{\text{gen}}/\sigma$ and its ratio to the nominal luminosity $L = 3.0$ ab^{-1} for FCC-ee at $\sqrt{s} = 365$ GeV.

WHIZARD has been used for the production of Higgs boson event samples, with Higgs boson decaying to $b\bar{b}$, $c\bar{c}$, gg, $s\bar{s}$ or the other main final states ($\tau\tau$, WW, ZZ). WHIZARD calculates the kinematic distributions of the Higgs boson and of the accompanying leptons, while the Higgs boson decay and the parton showering are handled by Pythia 6 [21].

The $\mu^+\mu^- H$ final state is produced only by $e^+e^- \to ZH$ events followed by $Z \to \mu^+\mu^-$ decays, while the $e^+e^- H$ final state can be either produced by Higgsstrahlung ($e^+e^- \to ZH$, $Z \to e^+e^-$) or by Z-boson fusion, in which two Z bosons radiated by each of the colliding particles (in $e \to eZ$) fuse into a Higgs boson ($ZZ \to H$). Both processes are generated in the simulated $e^+e^- H$ event samples. At 240 GeV, the cross-section of the latter (0.40 fb) is almost 20 times smaller than that of the former (6.76 fb); at 365 GeV, on the other hand, the two contributions are comparable (3.20 fb for Z boson fusion compared to 4.19 fb for $Z(ee)H$).

Similarly, the $\nu\bar{\nu}H$ final state can be produced both by $e^+e^- \to ZH$ events followed by $Z \to \nu\bar{\nu}$ decays (for any neutrino flavour) and by W-boson fusion, $WW \to H$, in events in which the colliding particles radiate a virtual W and an electron neutrino ($e^+e^- \to \nu_e\bar{\nu}_e H$). Both processes are generated in the simulated $\nu\bar{\nu}H$ event samples. At 240 (365) GeV, the cross-section of the W-boson fusion process, ≈ 6.1 fb (≈ 29.2 fb) is about 15% (117%) of that of $ee \to ZH \to \nu\bar{\nu}H$, 40.1 fb (29.2 fb)

WHIZARD has been used for the production of some non-Higgs background samples such as $ee \to ee$, $ee \to \mu\mu$, and $ee \to \nu_e\bar{\nu}_e Z$ (via W boson fusion only, to avoid double-counting with $Z(\nu_e\bar{\nu}_e)Z$ events in the inclusive ZZ sample), while PYTHIA 8 has been used for the production of the other non-Higgs backgrounds (inclusive WW, ZZ, $Z/\gamma^*(q\bar{q})$ production). In the PYTHIA 8 samples, both the matrix element calculation and the decays of unstable particles (including W and Z bosons) are handled by the same programme.

It should be noted that since only a few processes contribute to the main backgrounds for the studies presented in this note, for the future FCC-*ee* experiments it should not be a problem to generate significantly larger samples than those used in this study, making the impact of MC statistical uncertainties negligible.

4 Object reconstruction

Charged particle tracks are assumed to be reconstructed in the IDEA tracking system with perfect efficiency (100%) for transverse momentum above 100 MeV and $|\eta| < 2.56$. The momenta and impact parameters of the reconstructed tracks are smeared to account for the intrinsic resolution of the detector and for multiple scattering in its material.

Electromagnetic and hadronic energy deposits are expected to be reconstructed in the dual-readout calorimeters (for $|\eta| < 3$) with the following resolutions:

$$\frac{\sigma_E}{E} = \frac{3\%}{\sqrt{E}} \oplus \frac{0.2\%}{E} \oplus 0.5\% \tag{1}$$

for the electromagnetic component, and

$$\frac{\sigma_E}{E} = \frac{30\%}{\sqrt{E}} \oplus \frac{5\%}{E} \oplus 1\% \tag{2}$$

for the hadronic one.

A particle-flow algorithm uses the charged track candidates and the electromagnetic and hadronic energy deposits in the calorimeter to define three lists of reconstructed particle candidates: charged particles, photons and neutral hadrons.

Charged particle candidates reconstructed from electrons or muons with $E > 2$ GeV and $|\eta| < 3$ are assumed to be properly identified by dedicated algorithms with an efficiency of 99%. Similar efficiency performance are assumed for the identification of photon candidates. Particle mis-identification is neglected.

The missing momentum vector $\mathbf{p}^{\text{miss}}$ is computed as the opposite of the vectorial sum of the momenta of all particle-flow (p-flow) candidates.

The partons produced by the Higgs boson hadronic decays hadronise producing streams ("jets") of hadrons that interact with the detector. A clustering algorithm is used to combine the reconstructed particles into jet candidates. The jets' four-momenta are then calculated as the sum of the four-momenta of the jets' constituents. The exclusive $N = 2$ "Durham" algorithm [22], *i.e.* a sequential spherical k_t clustering algorithm that combines at each step the two closest particles (or the closest particle and jet) together and that stops when all input particles have been grouped into two jets, as implemented in the `FastJet` package, is used for this study. The distance between two particles (or a particle and a jet) is defined as:

$$d_{ij} = 2\min(E_i^2, E_j^2)(1 - \cos\theta_{ij}). \tag{3}$$

The E-recombination scheme is adopted, and jets returned by the algorithm are sorted in decreasing order of their energies. The leptons identified as the products of a $Z \to \ell\ell$ decay as described later are excluded from the inputs for the clustering.

To distinguish the flavour of the parton initiating a hadronic shower reconstructed as a jet, a tagging algorithm is used. This is based on a graph-neural-network algorithm exploiting various properties of the p-flow inputs clustered into the jets. The main features of the algorithm are detailed in Ref. [23], though a more recent version of its training is used. The network outpus seven weights (or scores), whose sum equals one, representing the probabilities assigned by the algorithm to each jet to originate from the hadronisation of either a b, c, s u, d quark, or from a gluon, or from the (hadronic) decay of a τ lepton. The network was trained on simulated $Z(\nu\bar{\nu})H$ events, clustering the particles in two jets with the same exclusive algorithm used in this analysis. The true flavour assigned to the jets was based on the generated decay of the Higgs boson.

5 Analysis of the $Z(\ell\ell)H(jj)$ final state at $\sqrt{s} = 240$ GeV and $\sqrt{s} = 365$ GeV

5.1 Event selection at $\sqrt{s} = 240$ GeV

Event selection begins by searching for a $Z \to \ell\ell$ decay. At $\sqrt{s} = 240$ GeV the momentum of the Z boson prodxuced in ZH events is about 50 GeV, leading to leptons that are produced in the laboratory frame with momenta between about 25 and 80 GeV, while ZZ events lead to more energetic Z bosons (momenta close to 80 GeV) producing in turn leptons with a broader momentum spectrum, mainly between 20 and 100 GeV. The events are thus required to contain two leptons (electrons or muons)[3] with momenta between 25 and 80 GeV, with same flavour (SF) and opposite-sign (OS) charges. In case of multiple SFOS lepton pairs satisfying the previous requirements, the one with invariant mass closest to the nominal Z boson mass of 91.2 GeV is selected.

The selected di-lepton pair is required to have invariant mass within a 20 GeV-wide range (81–101 GeV) around the nominal Z boson mass.

In signal events, the Z boson is expected to recoil against a Higgs boson. For this reason, the recoil mass m_{recoil}, calculated as

$$m^2_{\text{recoil}} = |P_i - P_{\ell\ell}|^2 \tag{4}$$

from the reconstructed four-momentum $P_{\ell\ell} = (E_{\ell\ell}, \vec{p}_{\ell\ell})$ of the dilepton pair and the initial four-momentum $P_i = (\sqrt{s}, \vec{0})$, is expected to be peaked near the Higgs boson mass. Events are therefore required to pass the requirement $120 < m_{\text{recoil}} < 140$ GeV.

After the lepton candidates from the $Z \to \ell\ell$ decays have been identified, jets are reconstructed by clustering all the reconstructed p-flow candidates after excluding the two lepton candidates.

A veto for events containing additional leptons with momenta above 25 GeV and a requirement that the variables $d_{2,3}$ and $d_{3,4}$ are greater than zero suppresses background events from Higgs decays to non-hadronic final states, such as $H \to \tau\tau$ with one of the τ leptons decaying to leptons, from ZZ events and from inclusive $Z(\ell\ell)$ production.

The event cutflow, with expected yields and efficiencies of each selection step for signal and background processes, is summarised in Table 3. The final selection efficiencies for $\ell\ell H$ events are 60% / 62% / 63% / 63% for Higgs bosons decaying to $b\bar{b}$, $c\bar{c}$, $g\bar{g}$, or $s\bar{s}$, respectively. The efficiency is a bit lower for $H \to b\bar{b}$ because of

[3] charge conjugation is implicitly assumed

the additional lepton veto; it might be considered to veto additional isolated leptons instead. The efficiency for other Higgs boson decays is 46% for WW, 54% for ZZ and 41% for $\tau\tau$; when considering only fully hadronic decays of the Higgs boson decay products, the efficiency is 63% for WW, 59% for ZZ and 56% for $\tau\tau$. Due to the different electron and muon resolutions, the expected signal yield is about 3.5% larger in the $\ell = \mu$ channel than in the $\ell = e$ channel, despite the slightly larger cross-sections for the $\ell = e$ channel (that are due to Z boson fusion, which is however suppressed by the lepton kinematic requirements). A sample of $ee \to Z(ee)H(b\bar{b}$ events was produced with WHIZARD to study the contribution to the eeH signal from the Higgsstrahlung process alone. The efficiency is estimated to be 59% for ZH events, and (by subtraction) 34% for the Z-fusion events, which contribute - after the selection – about 3% to the total signal.

The distributions of the variables used in the selection, for simulated signal and background events, after applying the selection criteria on the variables in the previous steps of the cutflow, are shown in Figure 1. The distributions are normalised to unit area.

The expected recoil mass distributions from the sum of signal and background processes after reconstruction and at the end of the selection are shown in Figure 2.

5.2 Event classification

After the selection described in the previous section, events are classified in ten mutually orthogonal categories targeting the different Higgs boson decays: $b\bar{b}$, $c\bar{c}$, $s\bar{s}$, gg, WW, ZZ and $\tau\tau$, and the the main background components: ZZ, WW, $Z/\gamma^*(\ell\ell)$. The classification is performed by a fully connected neural network (NN) regressor with a certain number of input nodes equal to the number of input variables and seven output nodes corresponding to the seven targets. The training of the neural network is performed with Keras using the Tensorflow backend. For the training, 600000 events each are selected for each Higgs boson decay category from the signal samples. The samples are then split in two, half of the events are used to train the network and the other half of the sample are used to validate it. Once trained, the network is evaluated on the full samples of Section 3, which are then used to extract the final results as described later in this Section.

The input variables used by the neural networks are the seven output scores of the flavour tagging algorithm for each of the two jets in the selected events (14 inputs), the dijet invariant mass, the missing momentum, and the $d_{2,3}$ and $d_{3,4}$ merge distance variables. The latter four quantities are transformed prior to the training in the following way:

- $d_{2,3} \to \log d_{2,3}/8$
- $d_{3,4} \to \log d_{3,4}/8$
- $m_{jj} \to (m_{jj} - 100)/40$.
- $p_{\text{miss}} \to \log(1 + p_{\text{miss}}/30)$

Additionally, considering the different spins of the Z ($J = 1$) and Higgs ($J = 0$) boson and the different production mechanisms, we expect different angular distributions of the Z boson in ZH and ZZ events [24], as shown in Figure 3 for the polar angle $\theta_{\ell\ell}$ of the dilepton pair. For this reason, we also include this polar angle variable as an input of the Neural Network.

The distributions of the NN input variables are shown in Figures 4–6.

Table 3: Event cutflow for the $\ell\ell H$ analysis, assuming an integrated luminosity $L = 10.8\ \mathrm{ab}^{-1}$ of ee collisions at $\sqrt{s} = 240$ GeV.

Selection	$\ell\ell H(b\bar{b})$		$\ell\ell H(c\bar{c})$		$\ell\ell H(gg)$		$\ell\ell H(s\bar{s})$		$\ell\ell H(WW)$		$\ell\ell H(ZZ)$		$\ell\ell H(\tau\tau)$		ZZ		WW		$Z/\gamma^*(\ell\ell)$		$Z/\gamma^*(q\bar{q})$	
	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)
No cuts	87599	-	4348	-	12313	-	36	-	32368	-	3971	-	9433	-	14677092	-	177535800	-	146804400	-	568662120	-
one $Z(\ell\ell)$ candidate	78728	90	3906	90	11058	90	32	90	29314	91	3594	91	8546	91	1223530	8	1667859	1	28920864	20	304657	0
$m_{\ell\ell}$ in 81–101 GeV	70832	90	3514	90	9948	90	29	90	26247	90	3244	90	7662	90	784799	64	410769	25	12264100	42	16264	5
m_{recoil} in 120–140 GeV	67630	95	3356	96	9502	96	28	96	24948	95	3059	94	7281	95	96931	12	179156	44	2014189	16	3161	19
no leptons with $p > 25$ GeV	64531	95	3325	99	9495	100	28	100	18326	73	2757	90	5398	74	87009	90	178913	100	2014187	100	2839	90
$d_{2,3} > 0$, $d_{3,4} > 0$	64531	100	3325	100	9495	100	28	100	18136	99	2612	95	4781	89	69857	80	8072	5	71760	4	2839	100
$\ell = e$	31730	-	1633	-	4664	-	14	-	8904	-	1284	-	2369	-	35366	-	5048	-	67285	-	1425	-
$\ell = \mu$	32801	-	1692	-	4831	-	14	-	9232	-	1328	-	2412	-	34491	-	3024	-	4475	-	1414	-

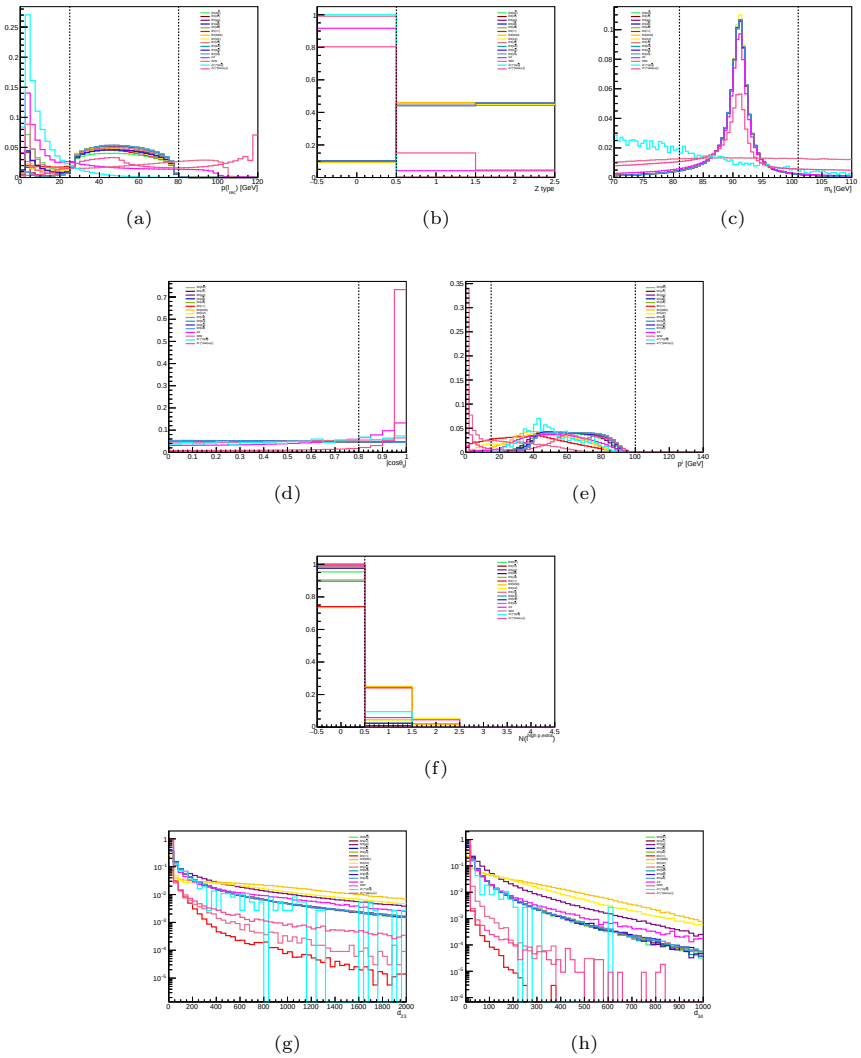


Fig. 1: Distributions (normalised to unit area) for signal and backgrond simulated events of the variables used in the $\ell\ell H$ event selection: (a) lepton momentum, (b) type of reconstructed $Z(\ell\ell)$ candidate (1: ee, 2: $\mu\mu$, 0: none), (c) invariant mass and (d) recoil mass, (e) jet momentum, (f) dijet invariant mass, (g) missing momentum, (h) number of additional leptons with $p > 25$ GeV. The distributions are plotted after applying selections on the preceding variables.

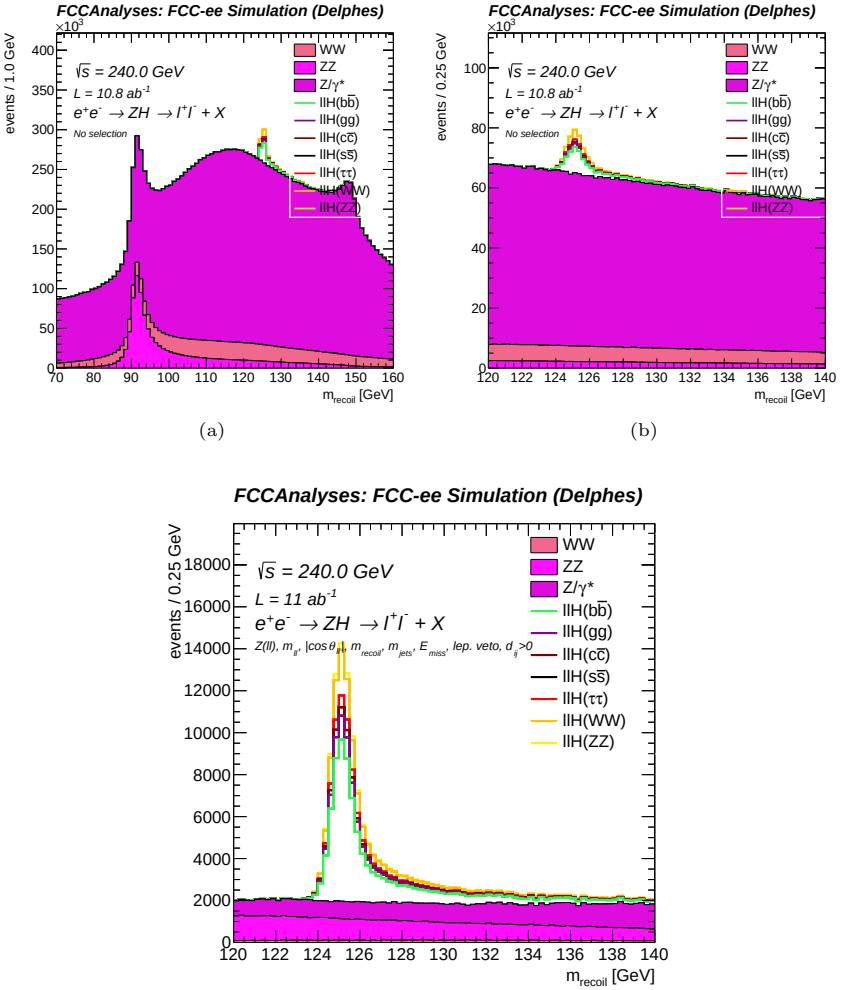


Fig. 2: Recoil mass distribution of signal and background superimposed before selection (after the reconstruction of one $Z(\ell\ell)$ candidate) (a) in a wider range than that (b) used in the selection; (c) m_{recoil} distribution after the final selection.

The training is performed over 50 epochs. The overall accuracy of the trained network is 80%. The accuracy and loss of the training vs the epoch and the final confusion matrix evaluated on the validation sample of the training are shown in Figure 7.

To increase the final sensitivity of the analysis, we further split the seven signal categories into up to three sub-categories based on the their weights, leading to a

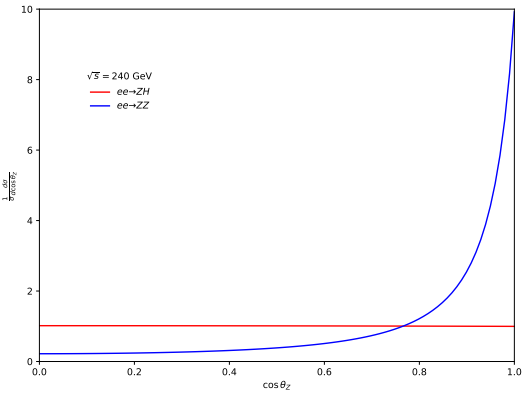


Fig. 3: Expected polar angle distributions of Z bosons produced in ZH (red) or ZZ (blue) events, computed with the formulae in Ref. [25].

possible total of $3 * 7 + 3(bkg) = 24$ categories in the fit. A demonstration of a sub-categorization if shown on figure 8.

The number of events expected in each category is summarised in Table 4.

5.3 Signal strength measurement

Once the events are classified in the seven categories, $b\bar{b}$, $c\bar{c}$, $s\bar{s}$, gg, WW, ZZ, $\tau\tau$ (the three background categories are mostly use to eliminate background as much as possible from the signal for the fit), the signal strengths for each of Higgs boson decays under study are determined by means of a simultaneous maximum likelihood signal+background fit to the recoil mass distribution of events in each category. A binned fit is performed in the interval 120–140 GeV, with a bin width of 0.25 GeV.

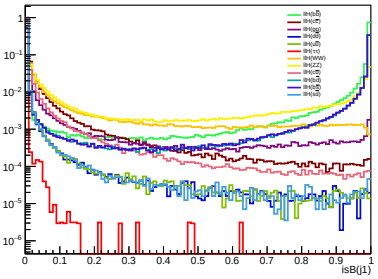
The global likelihood function is a product of likelihood functions for each category. In each category the likelihood is built from the product of Poisson probabilities for each bin, calculated from the event count in that bin and the expected yield from the sum of signal and background.

All the signal yields, for each Higgs boson decay mode, and the background yields, in each bin are calculated as the product of the expected signal yield times a signal strength μ_{XX} $(XX = b\bar{b}, c\bar{c}, ..)$ times the fraction of signal events predicted by the m_{recoil} signal model. There are no fixed or manually constrained parameters in the fit, as well as no systematics uncertainties. There is however an implementation of the Monte-Carlo statistical uncertainty.

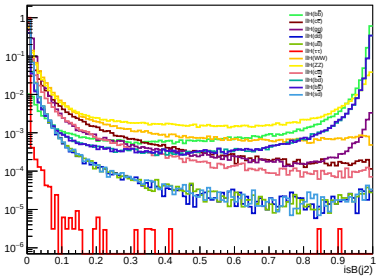
The fit is performed simultaneously on all sub-categories using the CombineTF tool

The results from the fits are summarized in Table 5

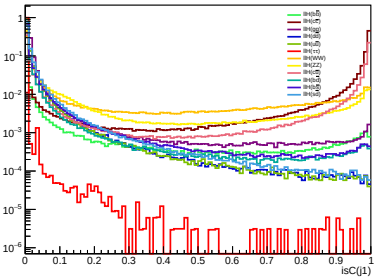
The $H \to b\bar{b}$ signal strength is thus expected to be measured with a statistical uncertainty close to 0.6%; for $H \to c\bar{c}$ the uncertainty is around 3.5%; for $H \to gg$ the uncertainty is slighlty better than 3.%. We also obtain high precision on other Higgs decay (ZZ, WW and $\tau\tau$) which are by-product of our analysis. We can also point that those results are extremely close to independent in terms of number of signal strenghts we simultaneously measure (*i.e.* if we manually fix one or more signal



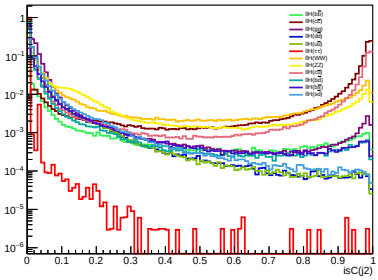
(a)



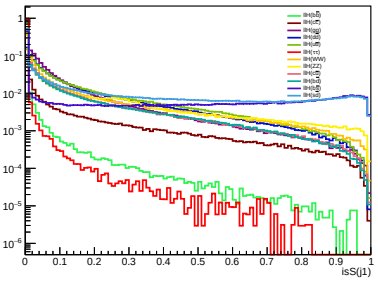
(b)



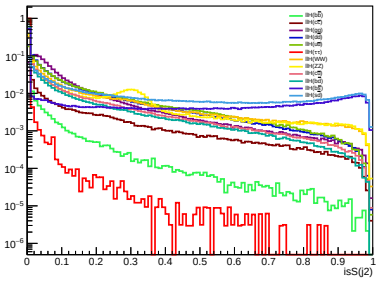
(c)



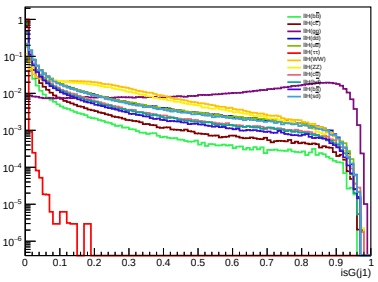
(d)



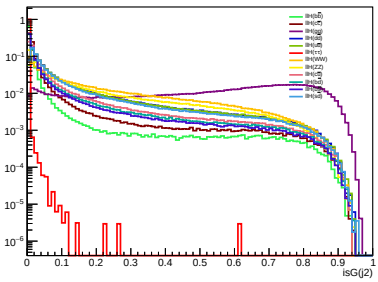
(e)



(f)



(g)

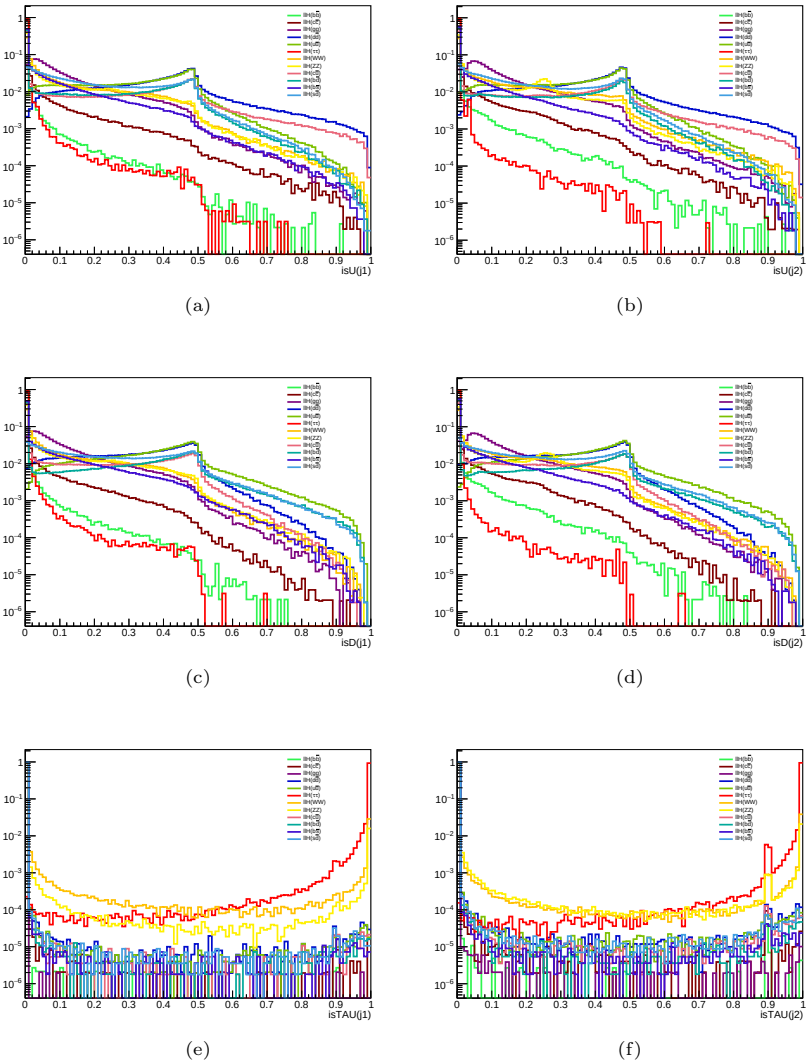


(h)

Fig. 4: Distributions (normalised to unit area) in $\ell\ell H$ simulated events for the input variables used in the neural network classification algorithm.

Table 4: Expected event yield from the different Higgs boson decays and from the background in each category, assuming an integrated luminosity $L = 10.8$ ab^{-1} of ee collisions at $\sqrt{s} = 240$ GeV. The numbers in parentheses are counting significances (with the approximate formula $S/\sqrt{S+B}$) for each Higgs process type. The high, mid, and low labels refer to the sub-categories.

Category	$\ell\ell H(b\bar{b})$	$\ell\ell H(c\bar{c})$	$\ell\ell H(gg)$	$\ell\ell H(s\bar{s})$	$\ell\ell H(WW)$	$\ell\ell H(ZZ)$	$\ell\ell H(\tau\tau)$	ZZ	WW	$Z/\gamma^*(\ell\ell)$	Total
$b\bar{b}_{low}$	3702.9 (50)	0.7 (0)	41.3 (1)	0.0 (0)	6.1 (0)	63.4 (1)	0.0 (0)	1571.7 (21)	5.2 (0)	1.1 (0)	5392.3
$b\bar{b}_{mid}$	6885.3 (74)	0.5 (0)	51.0 (1)	0.0 (0)	4.8 (0)	51.1 (1)	0.0 (0)	1591.5 (17)	1.9 (0)	0.0 (0)	8586.0
$b\bar{b}_{high}$	48179.9 (214)	0.7 (0)	28.0 (0)	0.0 (0)	1.9 (0)	26.8 (0)	0.0 (0)	2670.3 (12)	3.8 (0)	0.0 (0)	50911.5
$c\bar{c}_{low}$	63.7 (1)	557.8 (9)	103.7 (2)	0.0 (0)	304.3 (5)	73.0 (1)	0.0 (0)	2438.8 (41)	55.6 (1)	0.0 (0)	3596.9
$c\bar{c}_{mid}$	17.4 (0)	635.2 (15)	21.8 (0)	0.0 (0)	21.5 (0)	8.6 (0)	0.0 (0)	1197.7 (27)	8.1 (0)	0.0 (0)	1910.1
$c\bar{c}_{high}$	5.5 (0)	1768.8 (34)	7.4 (0)	0.0 (0)	2.4 (0)	1.9 (0)	0.0 (0)	950.5 (18)	0.5 (0)	0.0 (0)	2736.9
gg_{low}	239.3 (5)	15.3 (0)	1071.0 (21)	0.7 (0)	477.9 (9)	63.6 (1)	0.0 (0)	671.9 (13)	1.9 (0)	0.0 (0)	2541.5
gg_{mid}	145.8 (3)	9.0 (0)	1779.3 (34)	0.5 (0)	316.3 (6)	42.6 (1)	0.0 (0)	486.6 (9)	0.0 (0)	0.0 (0)	2780.0
gg_{high}	41.2 (1)	4.5 (0)	4428.7 (63)	0.3 (0)	156.7 (2)	20.9 (0)	0.0 (0)	253.5 (4)	0.0 (0)	0.0 (0)	4905.9
$s\bar{s}_{low}$	0.6 (0)	12.7 (0)	325.6 (5)	6.3 (0)	125.7 (2)	67.7 (1)	0.0 (0)	3196.4 (52)	16.6 (0)	0.0 (0)	3751.8
$s\bar{s}_{high}$	0.3 (0)	4.4 (0)	55.0 (1)	17.1 (0)	5.5 (0)	5.9 (0)	0.0 (0)	1590.0 (39)	0.0 (0)	0.0 (0)	1678.2
$(H)WW_{low}$	39.8 (0)	63.5 (1)	153.6 (1)	0.1 (0)	8649.6 (83)	171.0 (2)	1.6 (0)	1384.6 (13)	370.4 (4)	30.5 (0)	10864.6
$(H)WW_{mid}$	1.2 (0)	0.7 (0)	1.3 (0)	0.0 (0)	1017.1 (31)	3.4 (0)	0.0 (0)	32.7 (1)	23.8 (1)	0.0 (0)	1080.1
$(H)WW_{high}$	0.8 (0)	0.2 (0)	0.3 (0)	0.0 (0)	1138.0 (33)	2.0 (0)	0.0 (0)	17.5 (1)	23.8 (1)	0.0 (0)	1182.5
$(H)ZZ_{low}$	1818.9 (20)	24.1 (0)	117.8 (1)	0.1 (0)	581.7 (7)	723.4 (8)	2.2 (0)	4045.7 (45)	645.7 (7)	1.1 (0)	7960.6
$(H)ZZ_{mid}$	26.0 (1)	0.1 (0)	2.7 (0)	0.0 (0)	28.6 (2)	135.4 (8)	0.0 (0)	81.5 (5)	32.8 (2)	1.1 (0)	308.3
$(H)ZZ_{high}$	2.9 (0)	0.0 (0)	0.3 (0)	0.0 (0)	17.5 (1)	235.1 (13)	0.0 (0)	50.4 (3)	24.2 (1)	0.0 (0)	330.4
$\tau\tau$	0.1 (0)	0.0 (0)	0.0 (0)	0.0 (0)	407.3 (4)	40.8 (0)	4739.1 (48)	2938.4 (30)	549.7 (6)	963.4 (10)	9638.9
ZZ	672.6 (4)	18.2 (0)	71.2 (0)	0.5 (0)	261.7 (1)	153.8 (1)	14.0 (0)	32522.8 (177)	102.2 (1)	119.8 (1)	33936.8
WW	0.1 (0)	0.3 (0)	0.0 (0)	0.0 (0)	28.4 (1)	7.2 (0)	0.1 (0)	144.8 (3)	2712.7 (48)	263.0 (5)	3156.7
$Z/\gamma^*(\ell\ell)$	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	4.1 (0)	0.4 (0)	6.1 (0)	322.5 (1)	1661.4 (6)	68786.2 (259)	70780.6
TOTAL	61844.3	3116.6	8259.9	25.7	13556.9	1897.9	4763.1	58159.7	6240.3	70166.0	

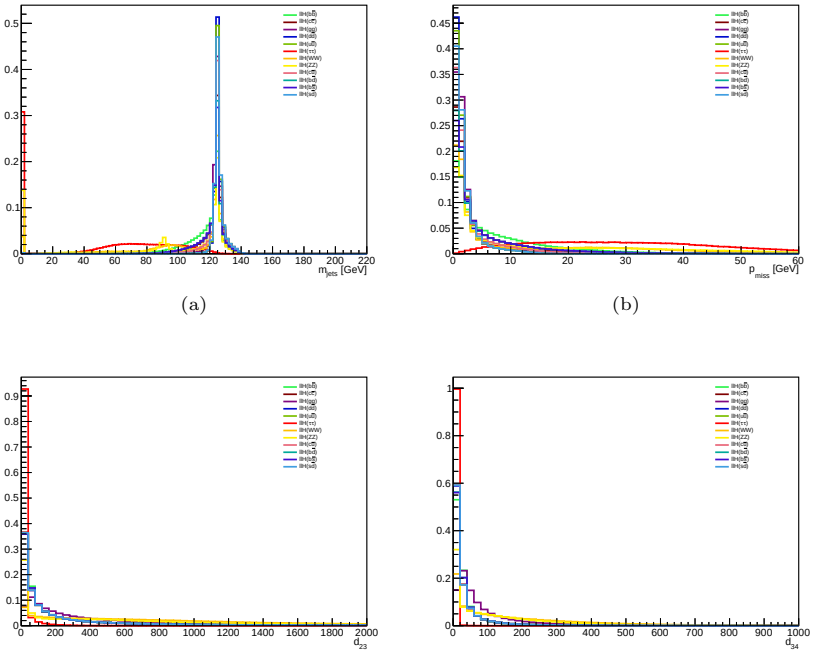


(a) (b) (c) (d) (e) (f)

Fig. 5: Distributions (normalised to unit area) in $\ell\ell H$ simulated events for the input variables used in the neural network classification algorithm.

Table 5: Expected statistical uncertainties on the signal strength in the $\ell\ell H$ production mode assuming an integrated luminosity $L = 10.8$ ab^{-1} of ee collisions at $\sqrt{s} = 240$ GeV, for three different configurations of the POIs.

Signal strength	Categories						
	$b\bar{b}$	$c\bar{c}$	gg	$s\bar{s}$	ZZ	WW	$\tau\tau$
Uncertainty (%)	0.60	3.47	1.93	223	7.65	1.49	2.54



(a) (b) (c) (d)

Fig. 6: Distributions (normalised to unit area) in $\ell\ell H$ simulated events for the input variables used in the neural network classification algorithm.

strenght to the Standard Model expectation). Assuming any known sensitivity on the background also does not provide any improvement or loss in the tabulated results.

5.4 Analysis at $\sqrt{s} = 365$ GeV

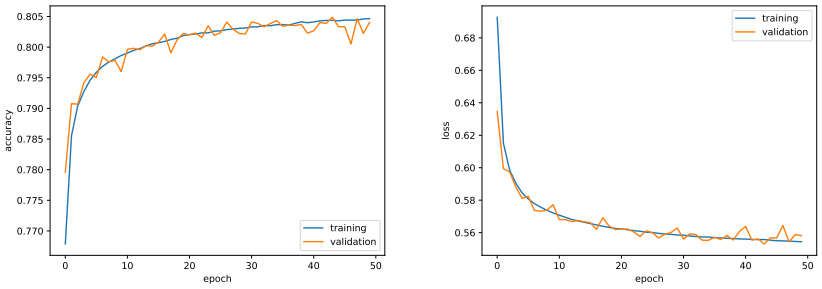
The analysis at 365 GeV is strictly the same as the one previously described. The only modifications corresponds to a tuning of the selection criteria to the new energy in the center of mass, and the consideration of an additional $t\bar{t}$ background component. The event cutflow for all selection steps for signal and background processes, is summarised in Table 6. We obtain final efficiencies around 50% / 54% / 54 % / 54% for $\ell\ell H$ events with the Higgs boson decay respectively to $b\bar{b}$, $c\bar{c}$, $g\bar{g}$, or $s\bar{s}$, and 32% / 38% / 20% for the WW, ZZ and $\tau\tau$ decays. If we consider only the fully hadronic decays for these last three cases, we obtain final efficiencies of 55% / 51% / 47%. The resolution discrepancy at this center of mass energy amounts to only a single percent difference in favor of the muonic decay channel compared to the electronic one.

The categorization is performed using the same neural network structure, and the same inputs, except that we include an additional background category, resulting in a total of $7 + 4$ categories.

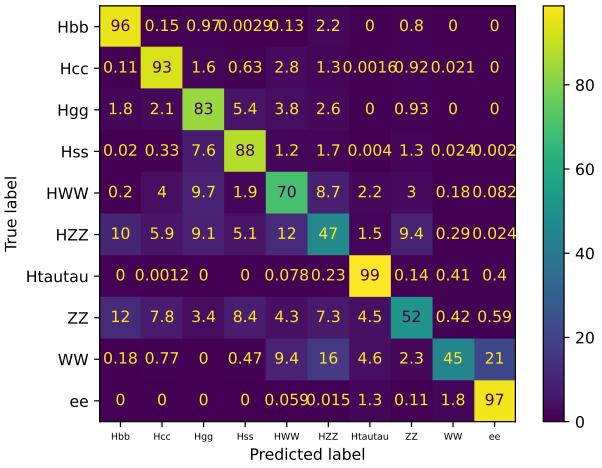
The confusion matrix evaluated on the validation sample of the training are shown in Figure 9.

Table 6: Event cutflow for the $\ell\ell H$ analysis, assuming an integrated luminosity $L = 3.0$ ab^{-1} of ee collisions at $\sqrt{s} = 365$ GeV.

Selection	$\ell\ell H(b\bar{b})$		$\ell\ell H(c\bar{c})$		$\ell\ell H(gg)$		$\ell\ell H(s\bar{s})$		$\ell\ell H(WW)$		$\ell\ell H(ZZ)$		$\ell\ell H(\tau\tau)$		ZZ		WW		$t\bar{t}$		$Z/\gamma^*(\ell\ell)$		$Z/\gamma^*(q\bar{q})$	
	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)		
No cuts	20223	-	1004	-	2842	-	8	-	7472	-	917	-	2178	-	1928400	-	32149500	-	2400000	-	11451000	-	49254270	-
> 0 iso-leptons with $p > 40$ GeV	19913	98	988	98	2799	98	8	98	7377	99	904	99	2150	99	357188	19	11839778	37	609966	25	9385275	82	54598	0
one $Z(\ell\ell)$ candidate	18514	93	916	93	2591	93	8	93	6880	93	844	93	2008	93	261035	73	734012	6	90463	15	5995494	64	39609	73
$m_{\ell\ell}$ in 81–101 GeV	12288	66	605	66	1707	66	5	66	4574	66	572	68	1339	67	161006	62	63740	9	11799	13	2828161	47	25	0
m_{recoil} 120–200 GeV	11027	90	547	90	1547	91	5	91	4076	89	498	87	1188	89	35765	22	23096	36	520	4	275305	10	0	0
$E(j_2) > 15$ GeV	10995	100	546	100	1546	100	5	100	3873	95	453	91	906	76	33567	94	21597	94	516	99	195619	71	0	0
≤ 2 iso. leptons	10117	92	539	99	1536	99	5	100	2438	63	368	81	464	51	29797	89	21465	99	476	92	195619	100	0	0
$d_{23} > 0$, $d_{34} > 0$	10117	100	539	100	1536	100	5	100	2435	100	350	95	447	96	24433	82	11400	53	476	100	4788	2	0	0
$\ell = e$	5022	-	267	-	759	-	2	-	1206	-	173	-	223	-	12672	-	6028	-	243	-	3635	-	0	-
$\ell = \mu$	5095	-	272	-	777	-	2	-	1229	-	177	-	225	-	11761	-	5372	-	234	-	1153	-	0	-



(a) (b)



(c)

Fig. 7: (a) Accuracy and (b) loss of the NN vs the learning epoch for both the training and validation subsamples. (c) Confusion matrix of the NN after the final training, evaluated on the validation subsample.

Using the same fit strategy, we obtaine the expected sensitiviy summarized in the Table 7

6 Analysis of the $Z(\nu\bar{\nu})H(jj)$ final state at $\sqrt{s} = 240$ GeV (I)

6.1 Event selection

Event selection begins by vetoing events containing at least one lepton with momentum between 20 and 100 GeV, ensuring orthogonality with the $\ell\ell H$ analysis.

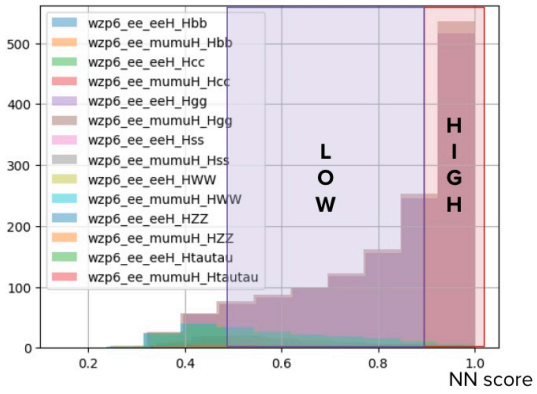


Fig. 8: Example of a sub-categorization based on the Neural Network output weight. Up to three categories between $[0.5, 1.0]$ were considered.

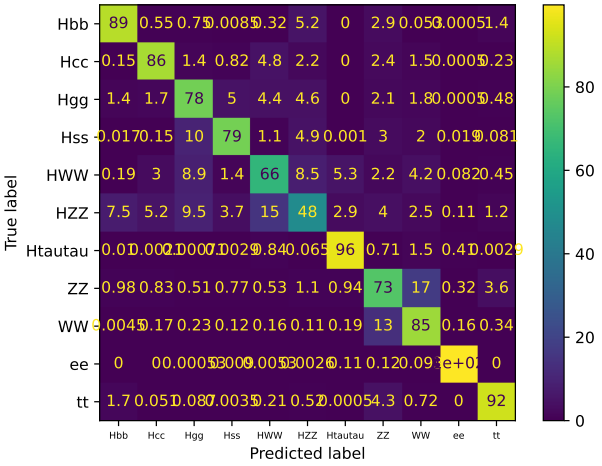


Fig. 9: Confusion matrix of the NN after the final training at $\sqrt{s} = 365$ GeV, evaluated on the validation subsample.

Table 7: Expected statistical uncertainties on the signal strength in the $\ell\ell H$ production mode assuming an integrated luminosity $L = 3.0$ ab^{-1} of ee collisions at $\sqrt{s} = 365$ GeV, for three different configurations of the POIs.

Signal strength	Categories						
	$b\bar{b}$	$c\bar{c}$	gg	$s\bar{s}$	ZZ	WW	$\tau\tau$
Uncertainty (%)	1.35	9.69	4.61	912	23.84	3.44	6.26

All reconstructed particles are clustered in two jets. The selection requires the energy of the leading and subleading jets (sorted by energy in descending order) are in the ranges 45–105 GeV and 20–70 GeV, respectively.

The number of constituents of the two jets are required to be greater than 10 and 6 to suppress background from hadronic tau-lepton decays or leptons misreconstructed as jets.

To suppress events from inclusive $Z(q\bar{q})$ production in which the two jets are back-to-back in azimuth, the cosine of the azimuthal separation between the two jets is required to be lower than 0.999.

Finally, for the two-dimensional fit to the missing mass m_{miss} and the *visible mass* m_{vis} to discriminate signal from background, m_{miss} is required to be in the range 50–140 GeV, while m_{vis} is required to be in the range 70–150 GeV. The visible mass is calculated after rescaling the dijet 4-momentum by a scale factor α such that the missing mass is equal to the Z-boson mass:

$$\begin{aligned}
(\sqrt{s}, \vec{0}) &= \alpha(E_{jj}, \vec{p}_{jj}) + (E_{\text{miss}}, \vec{p}_{\text{miss}}) \\
\Rightarrow\ (\sqrt{s} - \alpha E_{jj}, -\alpha\vec{p}_{jj}) &= (E_{\text{miss}}, \vec{p}_{\text{miss}}) \\
\Rightarrow\ \alpha^2 E_{jj}^2 - 2\alpha E_{jj}\sqrt{s} + s - \alpha^2 p_{jj}^2 &= m_Z^2 \\
\Rightarrow\ \alpha^2 m_{jj}^2 - 2\alpha E_{jj}\sqrt{s} + (s - m_Z^2) &= 0 \\
\Rightarrow\ \alpha = \frac{E_{jj}\sqrt{s} - \sqrt{\Delta}}{m_{jj}^2}&,
\end{aligned} \tag{5}$$

where

$$\begin{aligned}
\Delta &= E_{jj}^2 s - m_{jj}^2 (s - m_Z^2) \\
&= E_{jj}^2 s \left(1 - \left(\frac{m_{jj}}{E_{jj}}\right)^2 \left(1 - \frac{m_Z^2}{s}\right)\right)
\end{aligned} \tag{6}$$

The quantity that is then used for analysis is thus:

$$\begin{aligned}
m_{\text{vis}} = \alpha m_{jj} &= \frac{E_{jj}}{m_{jj}}\sqrt{s} - \sqrt{\frac{\Delta}{m_{jj}^2}} \\
&= \frac{E_{jj}}{m_{jj}}\sqrt{s}\left[1 - \sqrt{1 - \left(\frac{m_{jj}}{E_{jj}}\right)^2 \left(1 - \frac{m_Z^2}{s}\right)}\right]
\end{aligned} \tag{7}$$

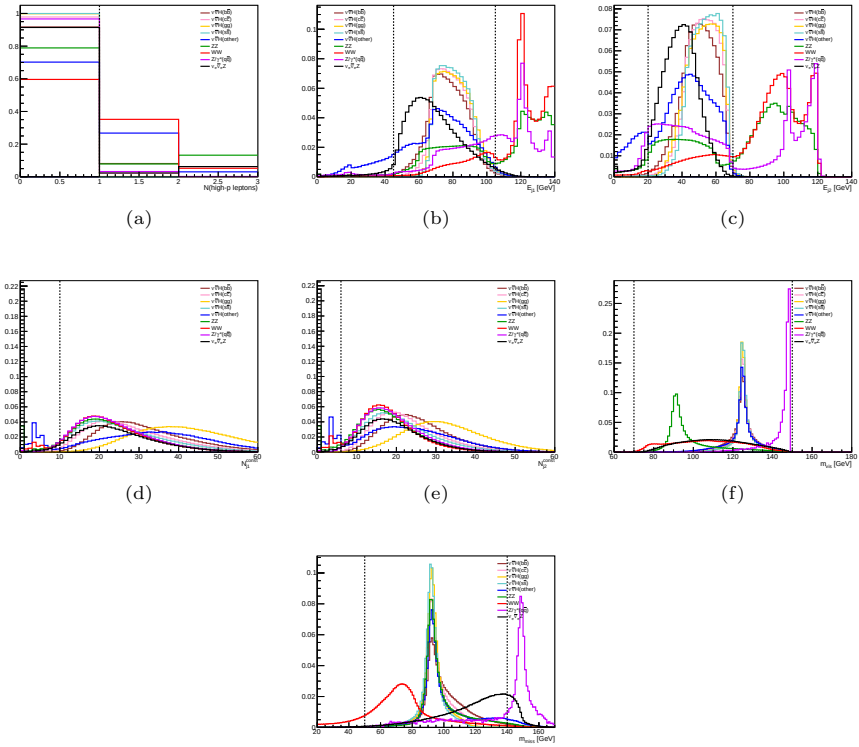
The event cutflow, with expected yields and efficiencies of each selection step for signal and background processes, is summarised in Table 8. The final selection efficiencies for $\nu\bar{\nu}H$ events are 86% / 91% / 92% / 92% / 62% / 70% / 44% for Higgs boson decaying to $b\bar{b}$, $c\bar{c}$, $g\bar{g}$, $s\bar{s}$, WW, ZZ, or $\tau\tau$, respectively.

The distributions of the variables used in the selection, for simulated signal and background events, after applying the selection criteria on the variables in the previous steps of the cutflow, are shown in Figure 10. The distributions are normalised to unit area.

The expected missing mass and visible mass distributions from the sum of signal and background processes after reconstruction and at the end of the selection are shown in Figure 11.

Table 8: Event cutflow for the $\nu\bar{\nu}H$ analysis, assuming an integrated luminosity $L = 10.8$ ab^{-1} of ee collisions at $\sqrt{s} = 240$ GeV.

Selection	$\nu\bar{\nu}H(b\bar{b})$		$\nu\bar{\nu}H(c\bar{c})$		$\nu\bar{\nu}H(gg)$		$\nu\bar{\nu}H(s\bar{s})$		$\nu\bar{\nu}H(WW)$		$\nu\bar{\nu}H(ZZ)$		$\nu\bar{\nu}H(\tau\tau)$		qqH		$\nu_e\bar{\nu}_e Z$		$Z/\gamma^*(q\bar{q})$		ZZ		WW	
	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)	Yield	ε (%)
No cuts	290520	-	14418	-	40846	-	120	-	107352	-	13176	-	31288	-	1472084	-	359359	-	568662120	-	177535800	-	14677092	-
No lepton with $p > 20$ GeV	266467	92	14108	98	40775	100	120	100	73746	69	11525	87	21221	68	1253254	85	328555	91	549585416	97	105842575	60	11446338	78
$15 < E_{j1} < 105$ GeV, $10 < E_{j2} < 70$ GeV	263683	99	13909	99	39962	98	118	98	71007	96	10632	92	16851	79	36239	3	318835	97	236447033	43	14074079	13	3913290	34
$\lvert\cos(\phi_{j1} - \phi_{j2})\rvert < 0.999$	253194	96	13352	96	38482	96	113	96	68687	97	10276	97	16058	95	34321	95	307331	96	59252457	25	13483356	96	3763882	96
$70 < m_{\text{vis}} < 150$ GeV, $60 < m_{\text{miss}} < 220$ GeV	250915	99	13177	99	37910	99	111	98	66423	97	9281	90	14681	91	31523	92	227747	74	32022838	54	10449862	78	3291598	87
$d_{2,3} > 0$, $d_{3,4} > 0$	250915	100	13177	100	37910	100	111	100	66372	100	9267	100	14043	96	31523	100	227512	100	32021097	100	10402375	100	3288059	100



(a) (b) (c) (d) (e) (f) (g)

Fig. 10: Distributions (normalised to unit area) for signal and backgrond simulated events of the variables used in the $\nu\bar{\nu}H$ event selection: (a) number of leptons with $p > 20$ GeV, (b,c) leading and sub-leading jet energies, (d,e) number of constituents of leading and sub-leading jets, (f) visible mass, (g) missing mass. The distributions are plotted after applying selections on the preceding variables. Histograms with less than 300 entries are not shown because of their large statistical fluctuations.

6.2 Event classification

Similarly to the analysis of the $\ell\ell H$ process in the previous Section, the selected events are classified by means of a neural network in $7 + 5$ mutually orthogonal categories targeting seven Higgs boson decays: $b\bar{b}$, $c\bar{c}$, $s\bar{s}$, gg, WW, ZZ, $\tau\tau$ and the five main background components. The network inputs are also the same than for the previous section, and network structure and the training procedure are identical.

For the training, 600000 events each are selected for each Higgs boson decay category from the signal samples, and 1000000 for the background categories. The samples are then split in two, half of the events are used to train the network and the other half of the sample are used to validate it. Once trained, the network is

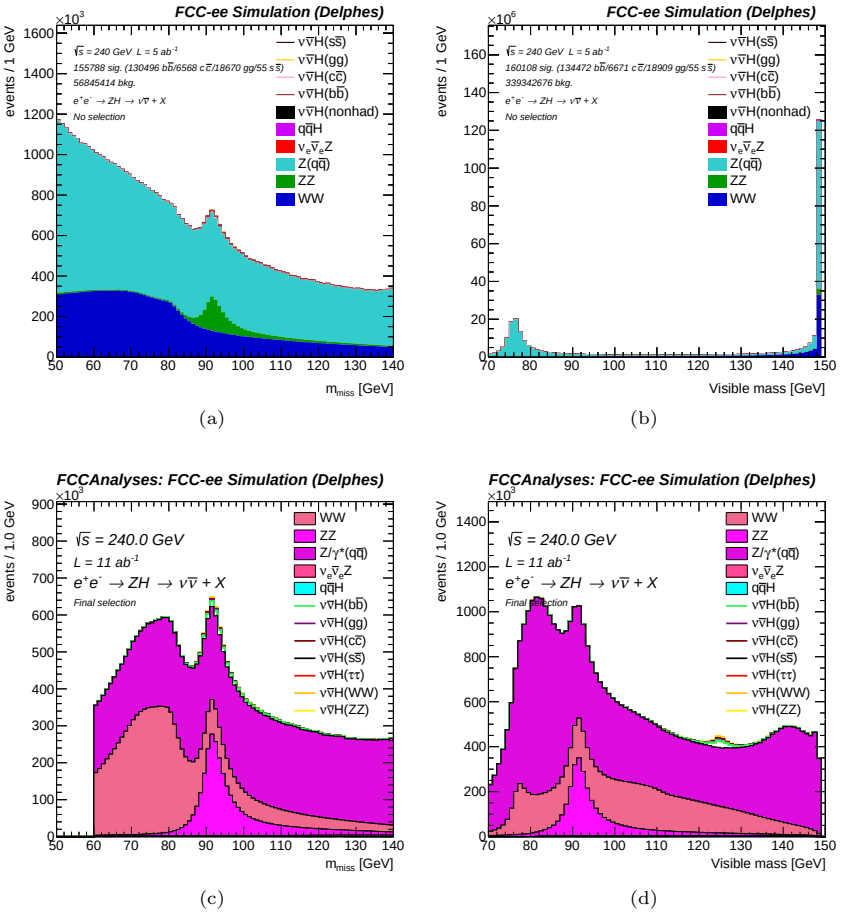


Fig. 11: Distributions of signal and background superimposed for (a,c) the missing mass and (b,d) the visible mass, (a,b) before and (c,d) after selection.

evaluated on the full samples of Section 3, which are then used to extract the final results as described later in this Section.

The input variables used by the neural networks are the five output scores of the flavour tagging algorithm for each of the two jets in the selected events (10 inputs), and the $d_{2,3}$ and $d_{3,4}$ merge distance variables. The latter two quantities are transformed prior to the training in the following way:

- $d_{2,3} \to \log d_{2,3}/6$
- $d_{3,4} \to \log d_{3,4}/6$

The different kinematic distributions of the polar angle of Z and H bosons in ZH events vs ZZ events are once again exploited as the last input of the neural network.

The distributions of the NN input variables are shown in Figures 12 and 13.

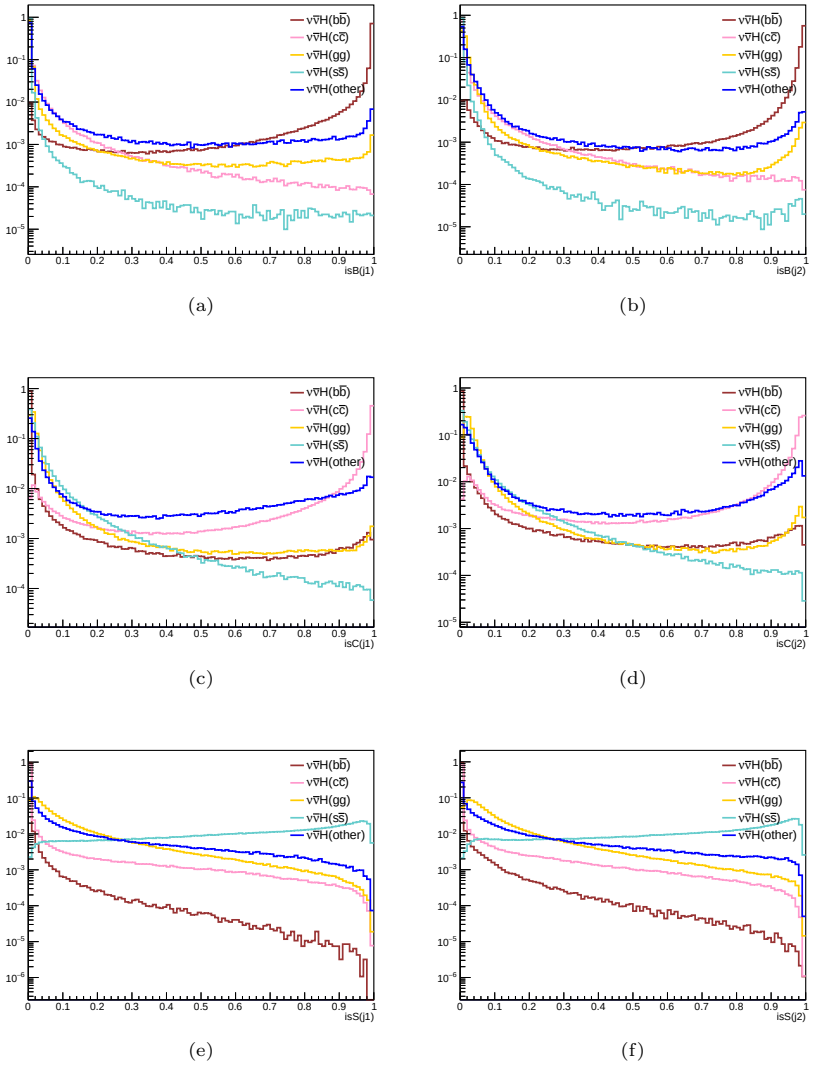


Fig. 12: Distributions (normalised to unit area) in $\nu\bar{\nu}H$ simulated events for the input variables used in the neural network classification algorithm.

The accuracy and loss of the training vs the epoch and the final confusion matrix evaluated on the validation sample of the training are shown in Figure 14. The overall accuracy of the network is about 88%.

As for the leptonic channels, the sensitivity of the analysis is further increased by splitting each category into up to threesub-categories based on the NN score of the events, in the $[0.5, 1]$ interval.

The number of events expected in each category is summarised in Table 9.

Table 9: Expected event yield from the different Higgs boson decays and from the background in each category, assuming an integrated luminosity $L = 10.8$ ab^{-1} of ee collisions at $\sqrt{s} = 240$ GeV. The numbers in parentheses are counting significances (with the approximate formula $S/\sqrt{S+B}$) for each Higgs process type. The high, mid, and low labels refer to the sub-categories.

Category	$\ell\ell H(b\bar{b})$	$\ell\ell H(c\bar{c})$	$\ell\ell H(gg)$	$\ell\ell H(s\bar{s})$	$\ell\ell H(WW)$	$\ell\ell H(ZZ)$	$\ell\ell H(\tau\tau)$	$Z/\gamma^*(qq)$	ZZ	WW	$q\bar{q}H$	nu_enu_eZ	Total
$b\bar{b}_{\text{low}}$	61092.0 (137)	5.4 (0)	306.2 (1)	0.0 (0)	18.9 (0)	460.5 (1)	0.0 (0)	104917.1 (235)	28678.3 (64)	2748.8 (6)	673.8 (2)	729.1 (2)	199630.1
$b\bar{b}_{\text{mid}}$	55486.2 (187)	1.5 (0)	26.2 (0)	0.0 (0)	0.6 (0)	31.0 (0)	0.0 (0)	24836.7 (84)	7484.4 (25)	219.2 (1)	34.7 (0)	112.3 (0)	88232.8
$b\bar{b}_{\text{high}}$	101039.0 (302)	0.4 (0)	6.2 (0)	0.0 (0)	0.0 (0)	5.7 (0)	0.0 (0)	8018.8 (24)	2795.5 (8)	16.2 (0)	3.0 (0)	7.0 (0)	111891.7
$c\bar{c}_{\text{low}}$	143.3 (0)	2314.7 (7)	298.7 (1)	0.2 (0)	542.8 (2)	217.1 (1)	0.0 (0)	47609.4 (137)	17316.0 (50)	52492.5 (151)	306.5 (1)	383.1 (1)	121624.4
$c\bar{c}_{\text{mid}}$	48.4 (0)	3276.3 (17)	67.7 (0)	0.0 (0)	40.9 (0)	23.5 (0)	0.0 (0)	15166.7 (79)	8229.2 (43)	9472.7 (50)	28.7 (0)	81.2 (0)	36435.3
$c\bar{c}_{\text{high}}$	10.4 (0)	5640.8 (50)	10.3 (0)	0.0 (0)	1.8 (0)	2.1 (0)	0.0 (0)	3189.4 (28)	2603.9 (23)	1194.4 (11)	1.1 (0)	4.9 (0)	12659.1
gg_{low}	1109.5 (6)	58.8 (0)	4677.8 (26)	4.8 (0)	1501.9 (8)	227.8 (1)	0.0 (0)	6265.7 (35)	4214.0 (24)	13795.8 (77)	204.3 (1)	27.0 (0)	32087.5
gg_{mid}	520.0 (3)	34.3 (0)	8634.7 (57)	3.4 (0)	973.4 (6)	146.6 (1)	0.0 (0)	3585.3 (24)	2790.3 (18)	6123.8 (40)	109.2 (1)	10.1 (0)	22931.0
gg_{high}	63.2 (0)	9.0 (0)	15309.4 (113)	1.2 (0)	307.9 (2)	45.8 (0)	0.0 (0)	893.5 (7)	870.0 (6)	991.4 (7)	26.3 (0)	1.1 (0)	18518.8
$s\bar{s}_{\text{low}}$	0.0 (0)	0.8 (0)	20.1 (1)	0.4 (0)	5.2 (0)	2.5 (0)	0.0 (0)	220.5 (8)	242.0 (9)	214.9 (8)	2.0 (0)	2.5 (0)	711.0
$s\bar{s}_{\text{mid}}$	0.2 (0)	31.2 (0)	449.8 (3)	25.2 (0)	65.8 (0)	53.9 (0)	0.0 (0)	5864.2 (44)	6840.4 (51)	4654.6 (35)	35.1 (0)	69.0 (1)	18089.6
$s\bar{s}_{\text{high}}$	0.0 (0)	7.3 (0)	81.9 (1)	54.5 (1)	3.2 (0)	3.9 (0)	0.0 (0)	1238.4 (18)	3042.2 (43)	526.4 (7)	1.4 (0)	10.4 (0)	4969.7
$(H)WW_{\text{low}}$	12.1 (0)	232.7 (1)	973.8 (2)	0.8 (0)	24637.2 (58)	724.2 (2)	0.0 (0)	2895.4 (7)	2868.9 (7)	147891.7 (348)	388.9 (1)	63.2 (0)	180689.0
$(H)WW_{\text{mid}}$	0.0 (0)	3.6 (0)	10.9 (0)	0.0 (0)	3108.9 (41)	12.8 (0)	0.0 (0)	17.0 (0)	18.8 (0)	2563.8 (34)	4.7 (0)	0.0 (0)	5740.5
$(H)WW_{\text{high}}$	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	20.3 (4)	0.0 (0)	0.0 (0)	0.0 (0)	1.0 (0)	8.6 (2)	0.0 (0)	0.0 (0)	29.9
$(H)ZZ_{\text{low}}$	6844.7 (28)	123.7 (1)	566.7 (2)	1.4 (0)	982.7 (4)	1745.3 (7)	0.0 (0)	6180.9 (26)	14588.2 (61)	22763.5 (95)	4001.6 (17)	65.0 (0)	57863.8
$(H)ZZ_{\text{mid}}$	43.6 (2)	0.3 (0)	5.4 (0)	0.0 (0)	4.9 (0)	75.2 (3)	0.0 (0)	28.3 (1)	167.8 (7)	197.3 (8)	65.0 (3)	0.4 (0)	588.2
$(H)ZZ_{\text{high}}$	0.2 (0)	0.0 (0)	0.2 (0)	0.0 (0)	0.0 (0)	0.3 (0)	0.0 (0)	0.0 (0)	1.6 (1)	1.4 (1)	0.1 (0)	0.0 (0)	3.8
$\tau\tau_{\text{high}}$	0.0 (0)	0.0 (0)	0.0 (0)	0.0 (0)	1.7 (0)	2.0 (0)	5775.6 (11)	62.2 (0)	13751.7 (26)	268433.4 (500)	0.0 (0)	523.6 (1)	288550.2
$Zq\bar{q}_{\text{high}}$	587.8 (0)	14.3 (0)	12.8 (0)	0.1 (0)	72.2 (0)	17.5 (0)	0.2 (0)	17127438.3 (4121)	75842.1 (18)	66730.6 (16)	122.1 (0)	4349.0 (1)	17275187.0
ZZ_{high}	3919.8 (2)	67.0 (0)	78.2 (0)	0.5 (0)	660.3 (0)	455.8 (0)	6.8 (0)	154833.8 (94)	2130044.4 (1287)	351728.5 (212)	1555.0 (1)	96605.3 (58)	2739955.5
WW_{high}	172.1 (0)	98.9 (0)	155.9 (0)	0.3 (0)	6474.4 (2)	151.9 (0)	40.0 (0)	267356.9 (94)	137506.6 (49)	7619188.7 (2688)	2852.0 (1)	2366.0 (1)	8036363.8
qqH_{high}	81.1 (0)	2.2 (0)	9.3 (0)	0.0 (0)	19.3 (0)	17.4 (0)	0.0 (0)	5570.2 (24)	14655.6 (64)	23294.7 (102)	8412.8 (37)	4.0 (0)	52066.5
$\nu_e\nu_eZ_{\text{high}}$	490.7 (1)	15.7 (0)	3.1 (0)	0.0 (0)	91.2 (0)	243.1 (0)	0.0 (0)	26369.2 (51)	180656.4 (351)	7342.5 (14)	222.1 (0)	49183.3 (96)	264617.4
TOTAL	231664.5	11938.8	31705.4	93.0	39535.7	4666.0	5822.8	17812558.0	2655209.4	8602595.5	19050.2	154597.4	

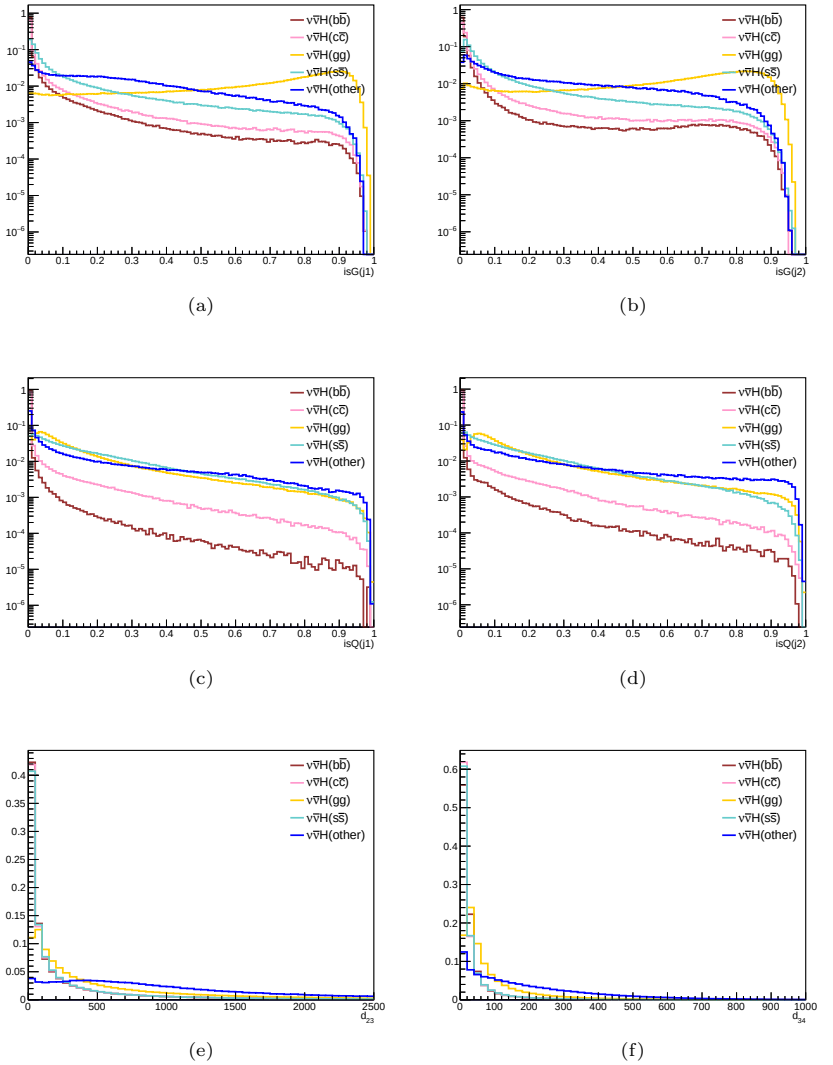
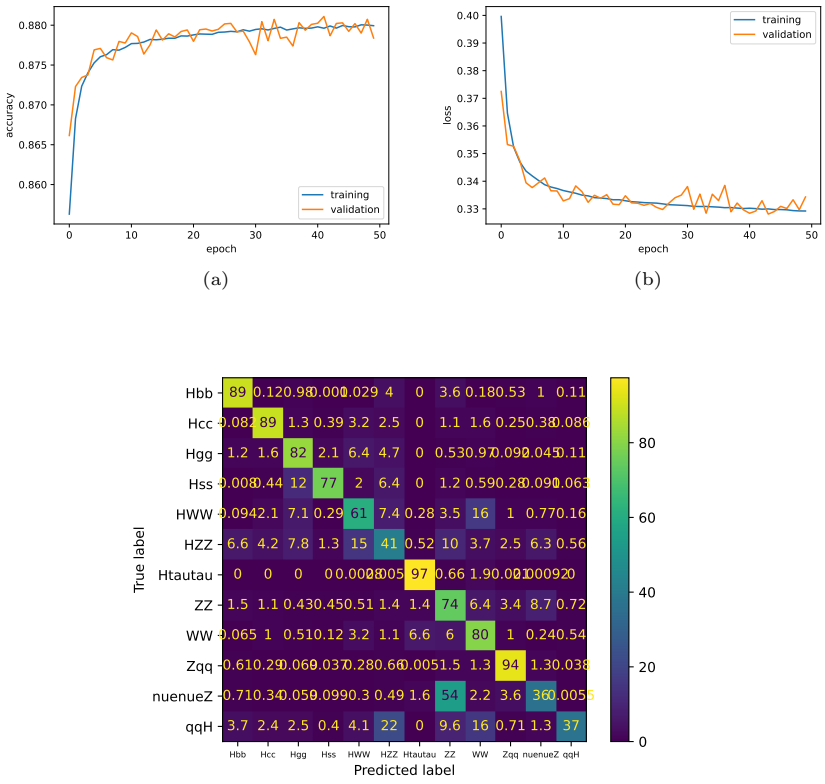


Fig. 13: Distributions (normalised to unit area) in $\nu\bar{\nu}H$ simulated events for the input variables used in the neural network classification algorithm.

6.3 Signal strength measurement

The signal strengths are measured by means of a simultaneous maximum likelihood signal+background fit to the distributions of two discriminant variables in each category. The observables used for the fit are the visible mass, which allows discriminating between the ZH signals (peaking around the Higgs boson mass) and all backgrounds



(a) (b)

(c)

Fig. 14: (a) Accuracy and (b) loss of the NN vs the learning epoch for both the training and validation subsamples. (c) Confusion matrix of the NN after the final training, evaluated on the validation subsample.

(which are non resonant), and the missing mass, which allows discriminating the signal (peaking around the Z boson mass) and non-ZZ backgrounds, as well as against $\nu\bar{\nu}H$ production from W-boson fusion.

A binned fit is performed in the intervals 60–140 GeV for m_{miss} and 70–150 GeV for m_{vis}, with fixed bin widths of 3 GeV and 2 GeV, respectively. The signal and background distributions in each category are modelled with binned templates determined directly from the simulation. Similarly to the $Z(\ell\ell)$ analysis, the fit is performed using Combine on all sub-categories simultaneously, with all signal strenght parameters for all Higgs decays and background let free. Only Monte Carlo statistical uncertainties were included in the fit.

The results from the fits are summarized in Table 10.

The $H \to b\bar{b}$ signal strength is thus expected to be measured with a statistical uncertainty better than 0.4%; for $H \to c\bar{c}$ the uncertainty is 2% and for $H \to gg$ it is about 1%. The same additional tests with alternative number of fitted parameters

Table 10: Expected statistical uncertainties on the signal strength in the $\nu\bar{\nu}H$ production mode assuming an integrated luminosity $L = 10.8$ ab^{-1} of ee collisions at $\sqrt{s} = 240$ GeV, for three different configurations of the POIs.

Signal strength	Categories						
	$b\bar{b}$	$c\bar{c}$	gg	$s\bar{s}$	ZZ	WW	$\tau\tau$
Uncertainty (%)	0.35	2.06	1.01	99.31	11.35	1.28	10.58

performed for $Z(\ell\ell)$ led to the same conclusions for this channel (*i.e.* no consequence on the sensitivities).

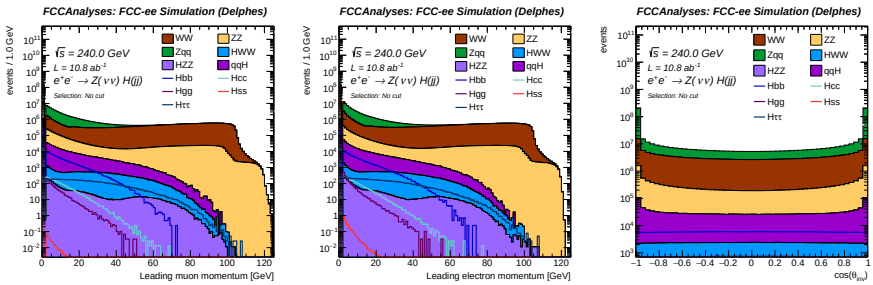
7 Analysis of the $Z(\nu\bar{\nu})H(jj)$ final state at $\sqrt{s} = 240$ GeV and $\sqrt{s} = 365$ GeV(II)

7.1 ZH + VBF inclusive analysis at $\sqrt{s} = 240$ GeV

In this section the couplings are studies without splitting the ZH and Vector boson fusion ($ee \to eeZZ \to eeH$, $ee \to \nu_e\nu_e WW \to \nu_e\nu_e H$) production modes.

7.1.1 Event selection

The event selection is similar to the one presented in Section 6. Events containing more than two leptons (e or μ) with $p > 20$ GeV are vetoed and the cosine of the angle between the visible momentum and the beam axis is required to be $|\cos\theta_{jj}| < 0.85$ to reject the single and di-boson backgrounds. The three corresponding distributions before cuts are shown in Figure 15 Moreover, cuts on $d_{1,2}$ and $d_{2,3}$ merge distance variables are applied to make sure that a two jet topology is selected. The selection cut-flow is shown in Table 11.



(a) (b) (c)

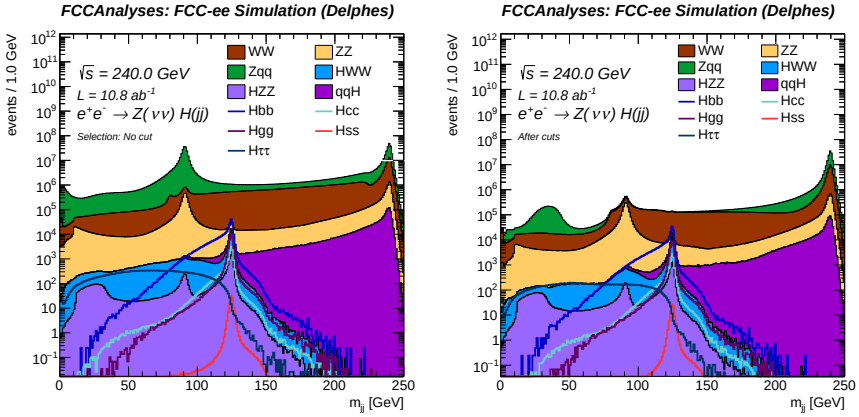
Fig. 15: Distributions for signal and background simulated events of the variables used in the $\nu\bar{\nu}H$ event pre-selection: (a) leading muon momentum, (b) leading electron momentum, (c) cosine of the angle between the visible momentum and the beam axis

Table 11: Selection cut flow.

		Before selection	Lepton cuts	$\|\cos\theta\| < 0.85$	kinematics & d_{ij}	efficiency(%)
Hbb	Yield(10^5)	2.91	2.67	2.28	2.28	78.4
	Sig.	10.59	10.34	16.01	54.0	
Hcc	Yield(10^4)	1.44	1.41	1.21	1.21	84.0
	Sig.	0.52	0.54	0.84	2.87	
Hgg	Yield(10^4)	3.59	3.59	3.08	3.07	85.5
	Sig.	1.30	1.39	2.16	7.27	
Hss	Yield	110	110	93.9	93.9	85.5
	Sig.	0.004	0.004	0.006	0.02	
$H\tau\tau$	Yield(10^4)	2.73	1.97	1.67	1.38	50.5
	Sig.	0.99	0.76	1.17	3.27	
HWW	Yield(10^4)	10.4	7.34	6.28	6.10	58.7
HZZ	Yield(10^4)	1.25	1.10	0.94	0.80	64.0
qqH	Yield(10^5)	14.7	12.6	8.86	0.56	3.8
WW	Yield(10^7)	17.3	10.6	6.35	1.26	7.3
ZZ	Yield(10^6)	14.0	11.0	6.93	2.60	18.6
Zqq	Yield(10^7)	56.6	54.7	13.1	0.22	0.4

As explained in previous sections, at lepton colliders, the recoil mass method can be used to reconstruct the mass of a particle without measuring its decay products. In the $\nu\bar{\nu}H$ final state, the Z boson decays in a pair of neutrinos and therefore its invariant mass can be reconstructed from the recoil method against the H invariant mass which is reconstructed from the two jets.

In Figures 16 and 17 the impact on the di-jet invariant mass M_{jj} and recoil mass M_{rec} of the event pre-selection discussed above is shown. The M_{jj} is visible to peak at $\sim 125\,\text{GeV}$, while the M_{rec} at $\sim 91\,\text{GeV}$



(a) (b)

Fig. 16: Impact of the pre-selection on the invariant mass M_{jj} distribution: (a) before applying any selection (b) after pre-selection.

7.1.2 Event Categorization

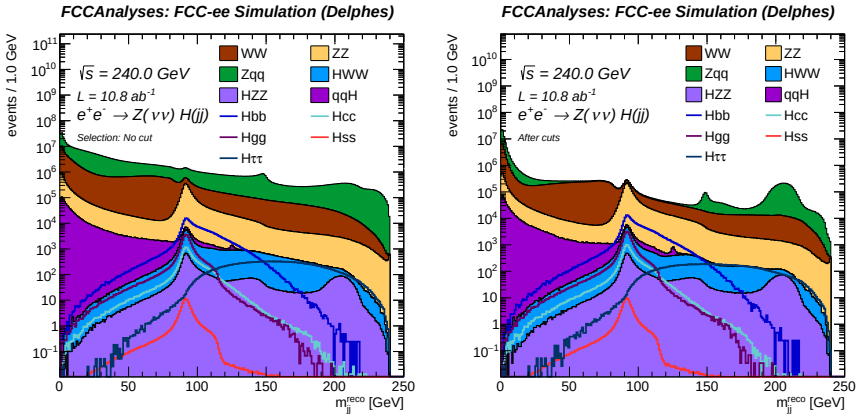
The jet flavour probabilities $K_j = (b_j, c_j, s_j, g_j, \tau_j, u_j, d_j)$, $j = 1, 2$ returned by the jet tagger in each event are summed $K = K_1 + K_2$. This way we form for event-wise observables $K = (B, C, S, G, T, U, D)$. The corresponding distributions of K values are plotted in Figure 18). The K-scores are then used to orthogonally categorize the events. An event is categorized in a given enriched category if the K value for that category is larger than for the other categories. For example, an event is labelled as C-like if $\max\{B, C, S, G, T, U, D\} = C$. This classification of events leads to the confusion matrix shown in Figure 19, which displays relatively pure categories and low contamination across different signals.

The yields resulting from isolating each category are shown in 12.

The events are further subdivided by defining three orthogonal categories with varying signal purity: Low (LP), Medium (MP) and High (HP) purity for each of the 4 previous enriched categories, thus leading to a total of (B-like, C-like, S-like, G-like) $\times$ (LP, MP, HP) 12 categories. The boundaries that define LP, MP and HP have been optimized by maximizing the individual significance for each signal in their respective enriched category. The chosen boundaries are shown in Figure 20 and numerically summarized in Table 13.

Table 12: Flavour-like category selection

	Hbb	Hcc	Hss	Hgg	$H\tau\tau$	qqH	HZZ	HWW	Zqq	WW	ZZ
B-like	219733	217	0	1126	1174	4468	2344	6720	1020071	491651	596278
C-like	3211	11289	1	1348	8859	12507	1766	26857	566307	1681120	597243
G-like	5034	374	13	29985	1112	3881	2588	20854	278283	588011	597210
S-like	138	185	87	2483	4042	1221	1717	8564	287095	732166	930994



(a) (b)

Fig. 17: Impact of the pre-selection on the recoil mass M_{rec} distribution: (a) before applying any selection (b) after pre-selection.

Table 13: Purity based subdivision of events.

	B	C	S	G
L	< 1.1	< 1.0	< 1.1	< 1.2
M	$[1.1, 1.9]$	$[1.0, 1.8]$	$[1.1, 1.7]$	$[1.2, 1.5]$
H	> 1.9	> 1.8	> 1.7	> 1.5

7.1.3 Fit Results

The background estimation model and the signal extraction strategy relies on a combined fit on both the Higgs candidate mass (denoted as m_{jj}) and the recoil mass (denoted as m_{rec}). The fit automatically uses background enriched regions to predict accurately the shape and normalization of the background processes. In a nutshell, the advantages of this approach are:

- The effective sideband on the analysis is two dimensional. The different background processes and the signal has different shapes in the m_{jj} - m_{rec} plane. Simultaneously using both m_{rec} and m_{jj} sidebands from the fit, results in better constraint of the backgrounds, and therefore better sensitivity.
- Adding nuisance parameters that affect simultaneously allows to account for mismodeling of their correlations.
- Instead of a selection in m_{rec} and m_{jj}, we exploit the full mass line-shape to extract the signal. This is very important because W and Z backgrounds have very close mass values and a cutoff-based selection is not optimal in separating their contributions.
- Lastly, in the case of signal we expect a resonance in both dimensions. This fit model allows to exploit this property and further improve the discrimination between signals and backgrounds, yielding improved sensitivity. Further

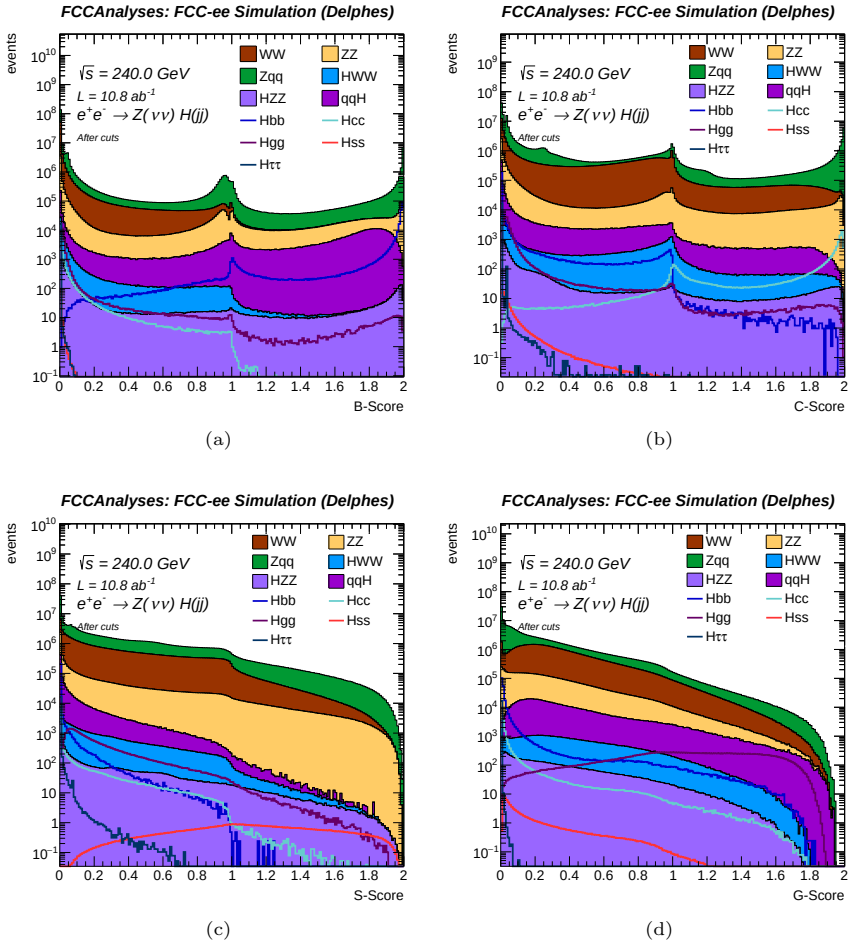


Fig. 18: Distributions of the sum of the two same flavor jet score (the K-scores) after the event pre-selection.

discrimination between the signals that defer by their decay mode, is achieved by selections on the jet tagging output and working points.

The signal is extracted by a binned likelihood fit in the m_{rec} and m_{jj} plane. The templates for each process are obtained from simulation. The fitted range for m_{rec} (m_{jj}) is from 40 (60) to 145 (150) GeV, with a bin width of 5 (1) GeV. The values are chosen to maximize the analysis sensitivity after accounting for the systematic uncertainties discussed next. The fit is performed simultaneously over all analysis categories.

Systematic uncertainties related to the normalization of the signal and background processes are included. For each signal (background) process we consider 0.1% (5%) log-normal uncertainty, uncorrelated between the processes, but correlated across the various analysis categories. It should be noted that due to the design

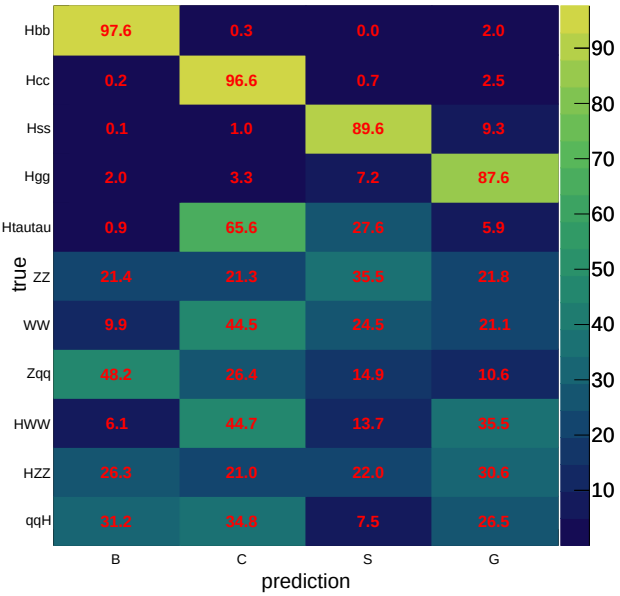


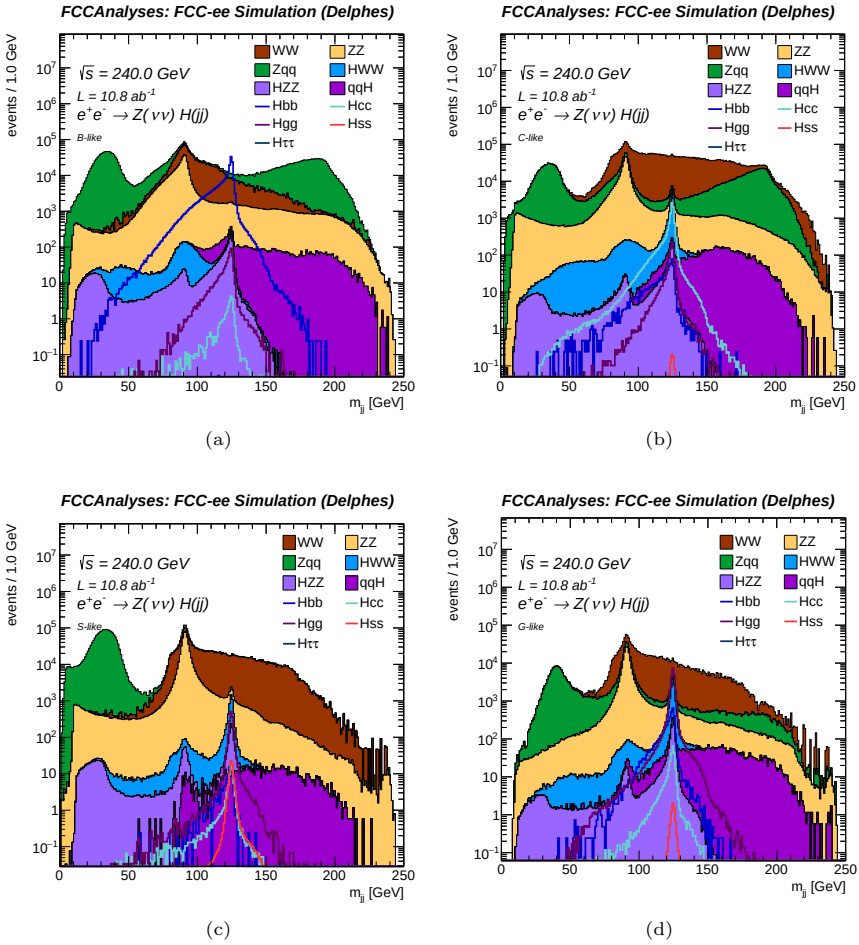
Fig. 19: Confusion matrix

of the fit, the uncertainties assigned to the background processes are constrained by the fit data in the sidebands to better than a percent level. Systematic uncertainties related to the limited simulation sample size are also considered.

The relative uncertainty on the signal strength of the Higgs boson decay bosons considered in this analysis are summarized in Table 14, and correspond for one interaction point and $10.8\,\mathrm{ab}^{-1}$ of integrated luminosity. In the case of H$\to$ s$\bar{\mathrm{s}}$ we calculated the upper limit at 95% confidence level on $B(H \to \mathrm{s\bar{s}})$ at 2 times the SM prediction. Figure 21 shows the negative log likelihood scans and the Hessian matrix computed signal strengths for the main processes.

Table 14: Expected Precision for the $\sigma(ZH) \times \mathrm{BR}(H \to jj)$ at 68 % CL

ZH final state		240 (%)	240 & MCStat (%)
$Z \to \nu\nu$	$H \to b\bar{b}$	0.26	0.33
	$H \to c\bar{c}$	1.74	2.27
	$H \to gg$	0.90	0.94
	$H \to s\bar{s}$	120.87	136.79
	$H \to \tau\tau$	17.92	21.76
	$H \to WW$	1.62	1.89
	$H \to ZZ$	16.74	19.83



(a) (b) (c) (d)

Fig. 20: Distribution of invariant mass M_{jj} over each category of events, respectively for (a) B-like (b) C-like (c) S-like (d) G-like events.

7.2 Splitting production modes (ZH, VBF) at $\sqrt{s} = 240$ GeV

In this section the ZH and Vector boson fusion (VBF) ($ee \to eeZZ \to eeH$, $ee \to \nu_e\nu_e WW \to \nu_e\nu_e H$) production modes at $\sqrt{s} = 240$ GeV are split. To be able to split the production modes, additional samples were produced. For ZH a muon neutrino sample was produced which is scaled by a factor of 3 accounting for the three neutrino variations. For VBF, an electron neutrino sample was produced from which the muon neutrino sample is subtracted. This results in a pure ZH region and a region which includes the VBF component and interference. Table 15 shows the samples used along with the events generated and the cross section for each process. For this study, the WW, ZZ, Zqq samples were used as in the inclusive analysis.

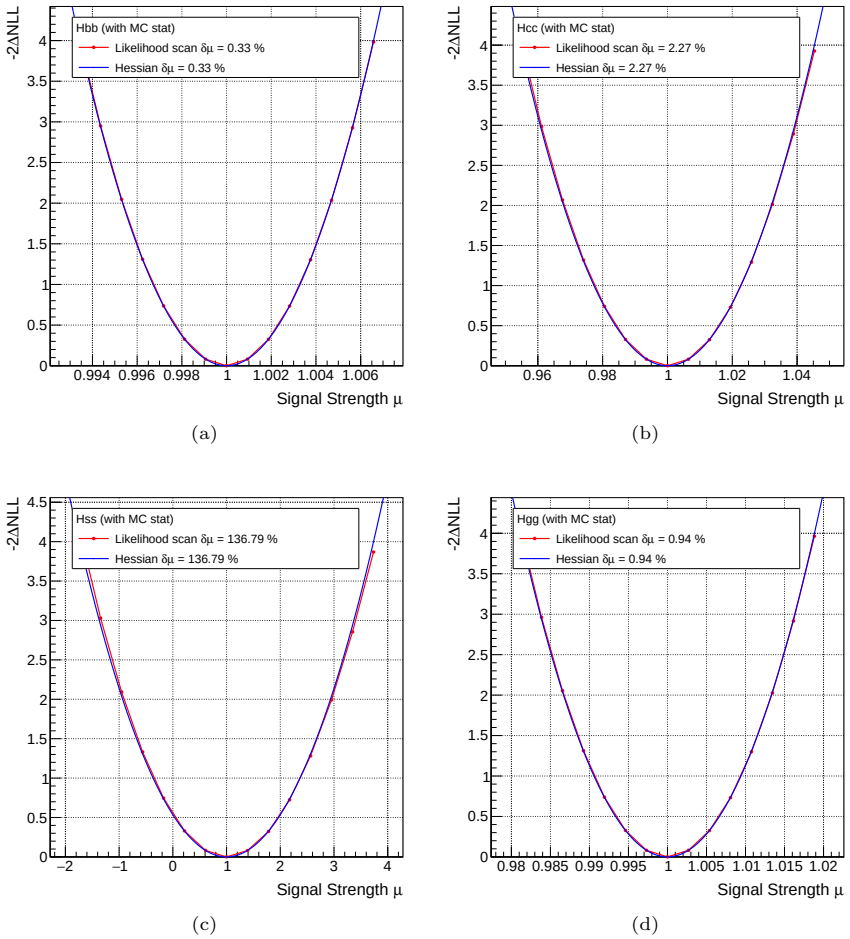


Fig. 21: Signal strength for Negative log-likelihood scans(red) and from Hessian computed matrix (blue) for the (a)$H \to b\bar{b}$, (b)$H \to c\bar{c}$, (c) $H \to s\bar{s}$ and (d) $H \to gg$ including bin by bin Monte-Carlo statistical uncertainties.

Figure 22 shows the m_{miss} at the generator (left) and the reconstruction level (right) for the ZH, VBF+interference and the inclusive samples which are in agreement.

The analysis and cuts described in section 7 remain unchanged. To extract the sensitivities, a 2D fit is performed on the m_{miss} versus the $m_{H_{jj}}$ and the VBF component (incl. interference) and ZH are extracted simultaneously. Table 16 shows the relative uncertainties as obtained from the simultaneous fit.

Table 15: Number of events and cross sections for the muon neutrino and electron neutrino samples at $\sqrt{s} = 240$ GeV.

Process	Number of Events	Cross Section (pb)
$\nu_\mu\bar{\nu}_\mu H(b\bar{b})$	1200000	0.00778
$\nu_\mu\bar{\nu}_\mu H(c\bar{c})$	1200000	0.0003862
$\nu_\mu\bar{\nu}_\mu H(s\bar{s})$	1200000	3.209e-06
$\nu_\mu\bar{\nu}_\mu H(gg)$	1200000	0.001094
$\nu_\mu\bar{\nu}_\mu H(\tau\tau)$	1200000	0.0008378
$\nu_\mu\bar{\nu}_\mu H(ZZ)$	1200000	0.0003528
$\nu_\mu\bar{\nu}_\mu H(WW)$	1200000	0.002875
$\nu_e\bar{\nu}_e H(b\bar{b})$	1200000	0.01132
$\nu_e\bar{\nu}_e H(c\bar{c})$	1200000	0.0005621
$\nu_e\bar{\nu}_e H(s\bar{s})$	1200000	4.67e-06
$\nu_e\bar{\nu}_e H(gg)$	1200000	0.001592
$\nu_e\bar{\nu}_e H(\tau\tau)$	1200000	0.001219
$\nu_e\bar{\nu}_e H(ZZ)$	1200000	0.0005135
$\nu_e\bar{\nu}_e H(WW)$	1200000	0.004184

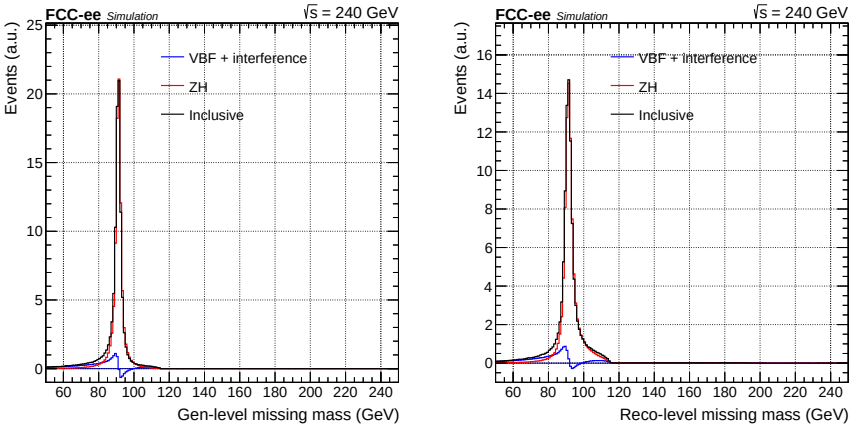


Fig. 22: the m_{miss} at the generator (left) and the reconstruction level (right) for the ZH, VBF+interference and the inclusive samples at $\sqrt{s} = 240$ GeV.

7.3 $\nu\bar{\nu}$ H(jj) at $\sqrt{s} = 365$ GeV

In this section the analysis of the $\nu\nu jj$ channel at $\sqrt{s} = 365$ GeV is described. At this center of mass energy the analysis was directly performed with the split in ZH and VBF production modes since the contribution of VBF is becoming significant. The analysis remains the same as in $\sqrt{s} = 240$ GeV and the equivalent samples listed in Table 17 were used.

Figure 23 shows the m_{miss} at the generator (left) and the reconstruction level (right) for the ZH, VBF+interference and the inclusive samples which are in agreement.

Table 18 shows the relative uncertainties as obtained from the simultaneous fit.

Table 16: Relative Uncertainty (%) for the $\sigma_{ZH} \times \mathcal{B}(H \to XX)$ and $\sigma_{\nu_e\bar{\nu}_e} \times \mathcal{B}(H \to XX)$ at 68 % CLin $\nu\bar{\nu}jj$ channel.

$\sqrt{s}$	240 GeV	
Integrated luminosity	10.8 ab^{-1}	
Channel	ZH	$\nu_e\bar{\nu}_e H$
$H \to b\bar{b}$	± 0.38	± 1.99
$H \to c\bar{c}$	± 2.66	± 20.52
$H \to s\bar{s}$	± 159.3	∞
$H \to gg$	± 1.06	± 5.68
$H \to W^+W^-$	± 1.87	± 165
$H \to ZZ$	± 20.44	± 143
$H \to \tau^+\tau^-$	∞	∞

Table 17: Number of events and cross sections for the muon neutrino and electron neutrino samples at $\sqrt{s} = 365$ GeV.

Process	**Number of Events**	**Cross Section (pb)**
$\nu_\mu\bar{\nu}_\mu H(b\bar{b})$	1200000	0.004814
$\nu_\mu\bar{\nu}_\mu H(c\bar{c})$	1200000	0.0002389
$\nu_\mu\bar{\nu}_\mu H(s\bar{s})$	1200000	1.984e-06
$\nu_\mu\bar{\nu}_\mu H(gg)$	1200000	0.0006767
$\nu_\mu\bar{\nu}_\mu H(\tau\tau)$	1200000	0.0005184
$\nu_\mu\bar{\nu}_\mu H(ZZ)$	1200000	0.0002183
$\nu_\mu\bar{\nu}_\mu H(WW)$	1200000	0.001779
$\nu_e\bar{\nu}_e H(b\bar{b})$	1200000	0.02181
$\nu_e\bar{\nu}_e H(c\bar{c})$	1200000	0.001083
$\nu_e\bar{\nu}_e H(s\bar{s})$	1200000	8.988e-06
$\nu_e\bar{\nu}_e H(gg)$	1200000	0.003066
$\nu_e\bar{\nu}_e H(\tau\tau)$	1200000	0.002349
$\nu_e\bar{\nu}_e H(ZZ)$	1200000	0.000989
$\nu_e\bar{\nu}_e H(WW)$	1200000	0.008059

8 Analysis of the $Z(qq)H(jj)$ final state at $\sqrt{s} = 240$ GeV and at $\sqrt{s} = 365$ GeV

8.1 Event selection at $\sqrt{s} = 240$ GeV

The event selection is similar to the one presented in Section 7, orthogonal to the $lljj$ analysis. The event topology requires four jets using Durham k_t exclusive algorithm. Events containing more that two leptons (e or μ) with $p > 20$ GeV are vetoed and the cosine of the angle between the visible momentum and the beam axis is required to be $|\cos\theta_{\text{vis}}| < 0.85$ to reject the single and di-boson backgrounds. To reject the $\nu\bar{\nu}jj$ events an additional cut on the visible mass, $M_{\text{vis}} > 150$ GeV is applied. Additional cuts on $d_{1,2}$, $d_{2,3}$ and $d_{3,4}$ merge distance variables are applied for such that a four jet topology is selected. The corresponding distributions are shown in Figures 24 and 25.

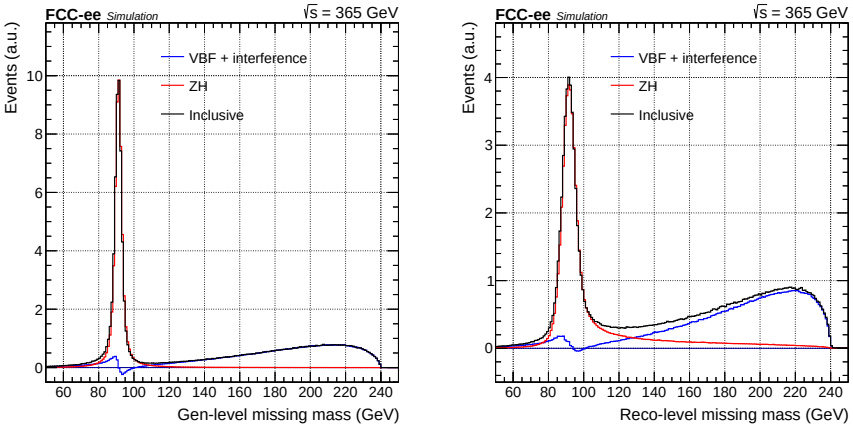


Fig. 23: the m_{miss} at the generator (left) and the reconstruction level (right) for the ZH, VBF+interference and the inclusive samples at $\sqrt{s} = 365$ GeV.

Table 18: Relative Uncertainty (%) for the $\sigma_{ZH} \times \mathcal{B}(H \to XX)$ and $\sigma_{\nu_e\bar{\nu}_e} \times \mathcal{B}(H \to XX)$ at 68 % CL in $\nu\bar{\nu}jj$ channel.

$\sqrt{s}$	365 GeV	
Integrated luminosity	10.8 ab^{-1}	
Channel	ZH	$\nu_e\bar{\nu}_e H$
$H \to b\bar{b}$	±0.73	±0.68
$H \to c\bar{c}$	±4.91	±3.56
$H \to s\bar{s}$	±482	±295
$H \to gg$	±3.00	±2.68
$H \to W^+W^-$	±5.48	±165
$H \to ZZ$	±40.9	±38.6
$H \to \tau^+\tau^-$	±20.01	±28.5

In e^+e^- colliders the final state energy and momentum is known: $(\sqrt{s}, 0, 0, 0)$. Hence in a four jet topology energy and momentum is conserved:

$$\begin{aligned} E_1 + E_2 + E_3 + E_4 &= \sqrt{s} \\ P_1^{x,y,z} + P_2^{x,y,z} + P_3^{x,y,z} + P_4^{x,y,z} &= 0 \end{aligned} \tag{8}$$

The jet directions ($\beta_i = p_i/E_i$ are very well measured:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ \beta_1^x & \beta_2^x & \beta_3^x & \beta_4^x \\ \beta_1^y & \beta_2^y & \beta_3^y & \beta_4^y \\ \beta_1^z & \beta_2^z & \beta_3^z & \beta_4^z \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{bmatrix} = \begin{bmatrix} \sqrt{s} \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

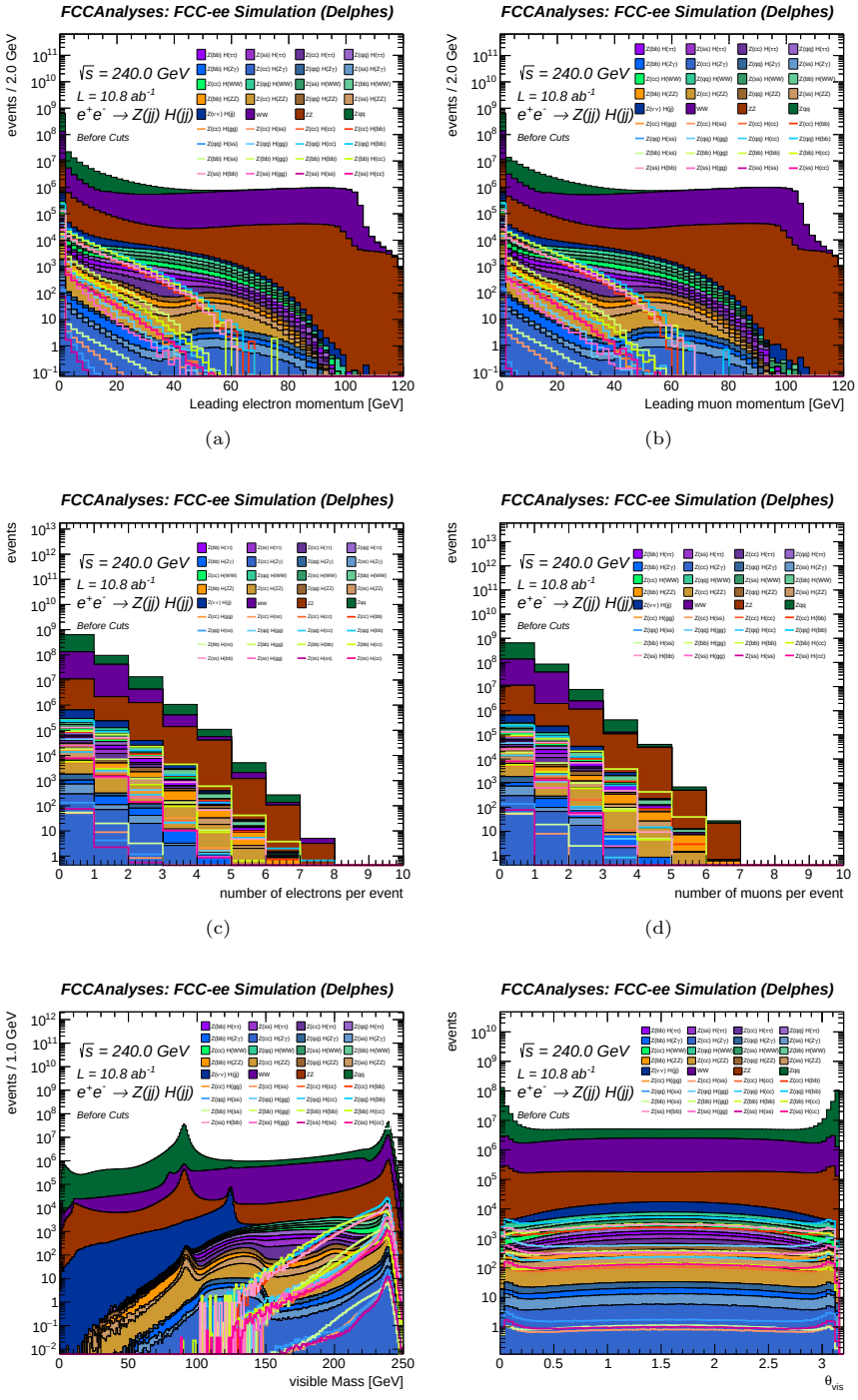


Fig. 24: Distributions for signal and background simulated events of the variables used in the jjH event pre-selection: (a) leading e momentum, (b) leading μ momentum, (c) number of electrons per event (d) number of muons per event (e) visible mass, (f) polar angle of the visible momentum.

The jet energies then can be corrected by inverting the above matrix. For each event the χ^2 is computed and a cut is applied for values less than 30. Moreover if the inversion of the matrix is not successful or return non-physical values (<1% of the events), the nominal values are kept such that no event is cut through this calculation. The selection cut-flow in terms of efficiency in percentage with respect to the sample size is shown in Table 11.

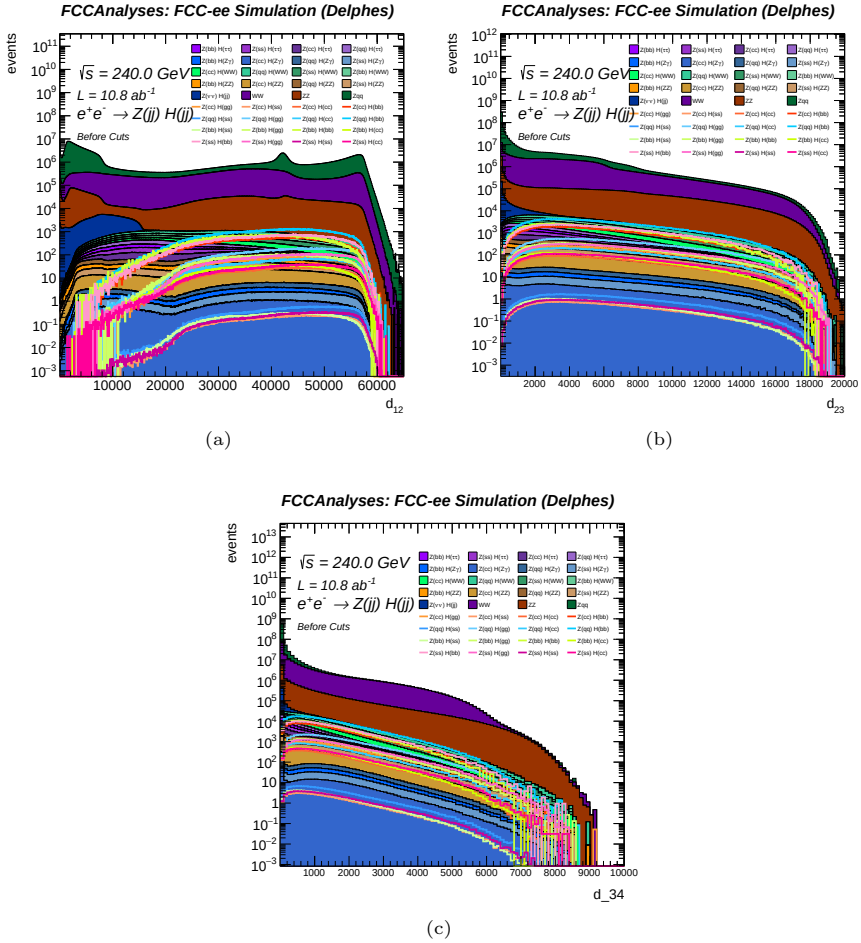


Fig. 25: Distributions for signal and background simulated events of the variables used in the jjH event pre-selection: (a) d_{12}, (b) d_{23}, (c) d_{34}.

8.1.1 Event Categorization

Once an event passes the above selection criteria, it goes through the jet pairing process. During this process, pairs of same flavor jets are formed based on the output

Table 19: Efficiency table

	Lepton cut	M_{vis}, θ_{vis}	d_{ij}	χ^2
$e^+e^- \to Z(cc)H(gg)$	98.7	88.3	87.2	83.5
$e^+e^- \to Z(cc)H(ss)$	99.0	88.4	86.3	82.8
$e^+e^- \to Z(cc)H(cc)$	96.6	88.1	86.1	80.9
$e^+e^- \to Z(cc)H(bb)$	89.7	83.5	81.2	72.2
$e^+e^- \to Z(qq)H(gg)$	99.8	86.2	85.2	83.1
$e^+e^- \to Z(qq)H(ss)$	99.9	86.6	84.6	82.7
$e^+e^- \to Z(qq)H(cc)$	97.8	87.1	85.2	81.6
$e^+e^- \to Z(qq)H(bb)$	91.4	83.8	81.7	73.9
$e^+e^- \to Z(bb)H(gg)$	94.6	87.0	85.9	78.5
$e^+e^- \to Z(bb)H(ss)$	95.0	87.3	85.1	77.5
$e^+e^- \to Z(bb)H(cc)$	92.1	85.7	83.4	74.6
$e^+e^- \to Z(bb)H(bb)$	84.4	79.8	77.3	65.4
$e^+e^- \to Z(ss)H(gg)$	99.8	87.0	85.9	83.7
$e^+e^- \to Z(ss)H(ss)$	99.9	87.2	85.2	83.2
$e^+e^- \to Z(ss)H(cc)$	97.8	87.7	85.7	82.0
$e^+e^- \to Z(ss)H(bb)$	91.3	84.1	82.0	74.0
$e^+e^- \to Z(bb)H(\tau\tau)$	63.7	43.9	32.8	11.7
$e^+e^- \to Z(ss)H(\tau\tau)$	67.1	48.3	36.4	13.6
$e^+e^- \to Z(cc)H(\tau\tau)$	68.0	50.2	38.1	14.2
$e^+e^- \to Z(qq)H(\tau\tau)$	67.9	50.1	38.1	14.2
$e^+e^- \to Z(bb)H(Z\gamma)$	86.5	62.4	61.3	54.9
$e^+e^- \to Z(ss)H(Z\gamma)$	90.5	64.0	62.9	58.6
$e^+e^- \to Z(cc)H(Z\gamma)$	91.7	63.7	62.5	59.1
$e^+e^- \to Z(qq)H(Z\gamma)$	91.6	63.1	61.9	58.7
$e^+e^- \to Z(bb)H(WW)$	64.7	57.4	54.6	39.2
$e^+e^- \to Z(ss)H(WW)$	68.0	59.8	57.0	42.1
$e^+e^- \to Z(cc)H(WW)$	68.7	59.9	57.0	42.4
$e^+e^- \to Z(qq)H(WW)$	68.6	59.4	56.6	42.0
$e^+e^- \to Z(bb)H(ZZ)$	81.8	60.6	57.8	42.2
$e^+e^- \to Z(ss)H(ZZ)$	86.1	63.3	60.5	45.7
$e^+e^- \to Z(cc)H(ZZ)$	87.5	63.9	61.1	46.7
$e^+e^- \to Z(qq)H(ZZ)$	87.5	63.6	60.8	46.4
$e^+e^- \to Z(\nu\nu)H(jj)$	87.5	00.1	00.0	0.0
$e^+e^- \to W^+W^-$	64.1	45.1	37.9	35.7
$e^+e^- \to ZZ$	79.8	43.4	38.1	34.4
$e^+e^- \to Z/\gamma^*(q\bar{q})$	96.5	31.8	07.6	07.1

probability of the tagger for all the 7 scores produced: $K_j = (b_j, c_j, s_j, g_j, \tau_j, u_j, d_j)$, $j = 1, 2, 3, 4$. This results in a 4×7 probability matrix. At first, the max score per jet probabilities is found as $K_j^{\max} = \max\{b_j, c_j, s_j, g_j, \tau_j, u_j, d_j\}$, $j = 1, 2, 3, 4$ and a flavor is assigned to this jet based on this probability. That process leads to fives different cases:

(a) Two jet pairs are formed with same flavor per jet in a pair but the pair flavor is different per pair.

(b) One jet pair is formed with same flavor per jet in this pair but the remaining jets have no equal pair. The first reconstructed pair then is kept and for the remaining jets, the sum of $K_f = K_{f1} + K_{f2}$ where $f = (b, c, s, g, \tau, u, d)$ is calculated. The second pair then is the one with the highest sum of the score: $\max\{K_f\}$.

(c) No pair of same flavor jets is formed. In this case the above mentioned process based on the highest sum of the scores is performed until two pairs with jets of the same flavor per pair are formed

(d) All the four jets are defined of the same flavor

From the above categorization, cases (b) and (c) can end up in either category (a) where two pairs of same flavor jets are formed but the pairs are of different flavor, or category (d) where all the jets are all of the same flavor and therefore pairs of two jets cannot be made directly based on the flavor only. Once two pairs of the same flavor jets are formed, the pairs are identified as the Z or H boson candidates. For events in category (a), $(m_{j_1j_2} - m_Z)^2 + (m_{j_3j_4} - m_H)^2$ is calculated where $m_{j_1j_2}, (m_{j_3j_4}$ are the dijet invariant masses and m_Z, m_H are the Z, H invariant masses fixed to $m_Z = 91.1876\,\text{GeV}, m_H = 125.11\,\text{GeV}$ respectively. If the event ends up in category (d) then the above equation is calculated for all the combinations of jets until the two minimum are found. It is important to note that 100% of the events are reconstructed and no event is lost during the pairing process. The procedure is illustrated schematically in Figure 26 and the confusion matrix in Figure 27.

Once the invariant mass of the two dijet pairs $m_{Z_{jj}}, m_{H_{jj}}$ is calculated, an additional cut is performed: $\sqrt{(m_{Z_{jj}} - m_W)^2 + (m_{H_{jj}} - m_W)^2} > 10$ and $\sqrt{(m_{Z_{jj}} - m_Z)^2 + (m_{H_{jj}} - m_Z)^2} > 10$. The m_Z, m_W, are fixed to $m_Z = 91.1876\,\text{GeV}$ and $m_W = 80.377\,\text{GeV}$ respectively. This additional cut helps to suppress WW, ZZ backgrounds further. Finally, $m_{Z_{jj}}, m_{H_{jj}}$ are required to be in the following ranges: $50 < m_{Z_{jj}} < 125$, and$m_{H_{jj}} > 91$. The $m_{Z_{jj}}, m_{H_{jj}}$ distributions are shown in Figure 28.

The events are further subdivided similar to the $\nu\bar{\nu}$ analysis in Section 7. Three orthogonal categories with varying signal purity are defined: Low (LP), Medium (MP) and High (HP) purity for each of the 16 categories based on the Z, H decay modes. Thus in the full hadronic decays that leads to a total of (BB-like, BC-like, BS-like, BG-like, etc) $\times$ (LP, MP, HP) 48 categories. The boundaries that define LP, MP and HP have been optimized by maximizing the individual significance for each signal in their respective enriched category. The chosen boundaries are summarized in Table 20.

Table 20: Purity based subdivision of events.

	B, C, G	**S**
L	< 1.1	< 0.8
M	$\in [1.1, 1.8]$	$\in [0.8, 1.4]$
H	> 1.8	> 1.4

8.1.2 Fit Results

The fitting strategy is identical, although much more complicated to the one in $\nu\bar{\nu}jj$ describe in Section 7. The signal is extracted by a binned likelihood fit in the m_{rec} and m_{jj} plane. The templates for each process are obtained from simulation. The fitted range for m_{rec} (m_{jj}) is from 40 (60) to 145 (150) GeV, with a bin width of 5

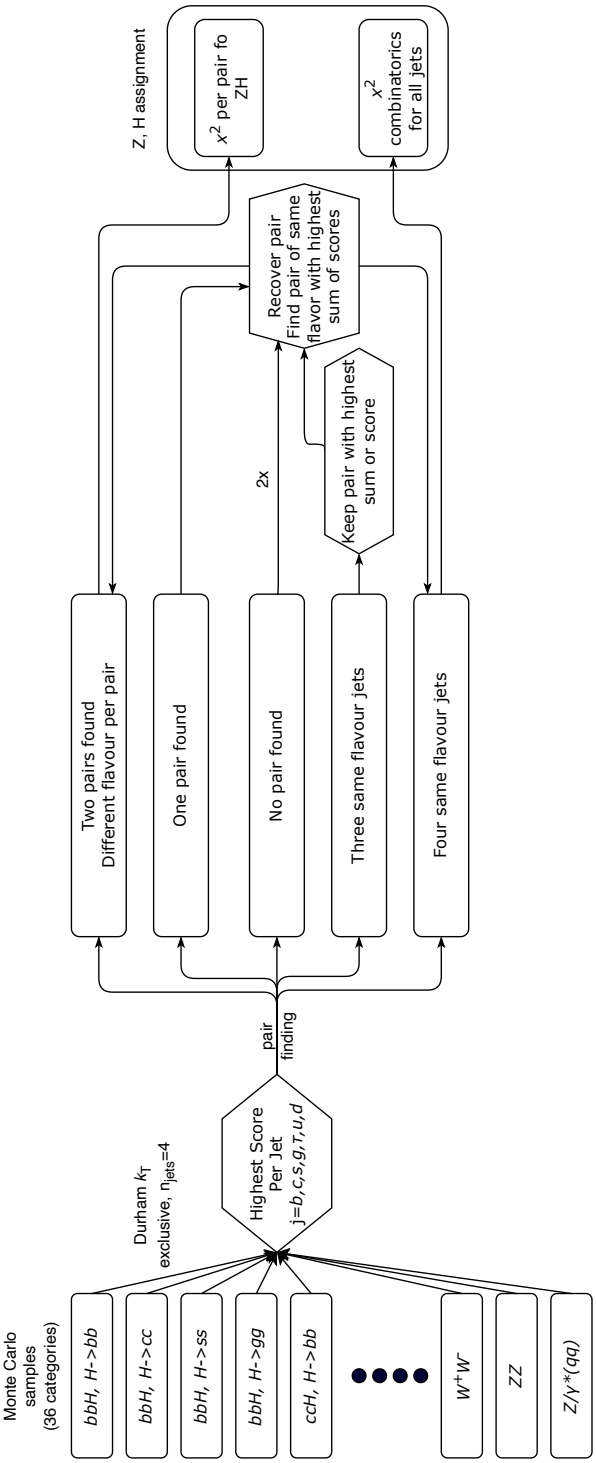


Fig. 26: Illustration of the jet pairing procedure in the fully-hadronic analysis.

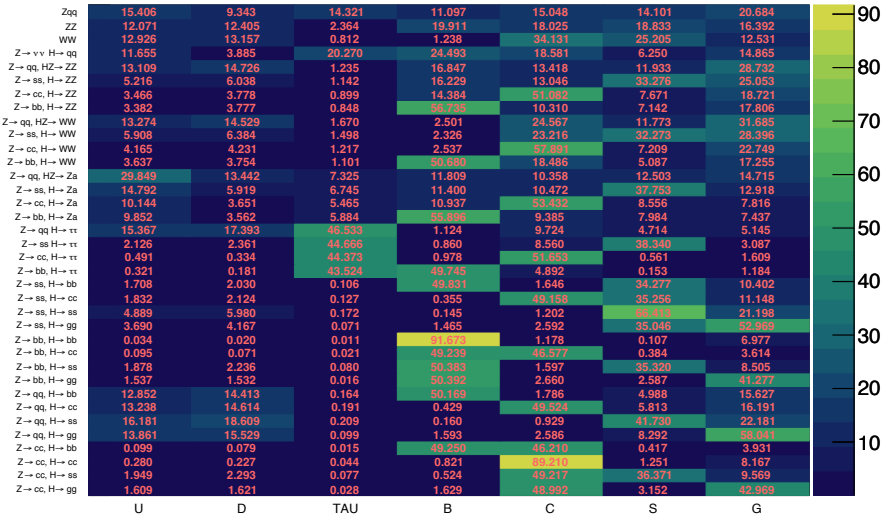


Fig. 27: Confusion matrix. Each row corresponds to a different process. The first seven columns show the fractions of events that have two same-flavour pairs and are labelled as U, D, TAU, B, C, S or G.

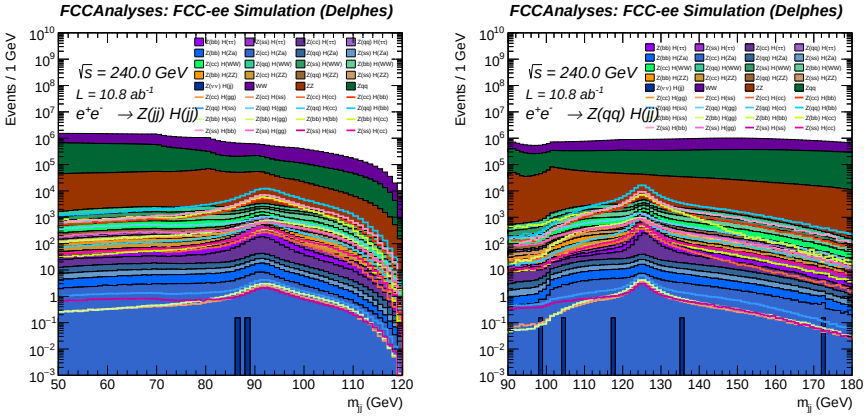


Fig. 28: Distribution of invariant mass M_{jj} of the Z on the left and the H boson on the right.

(1) GeV. The values are chosen to maximize the analysis sensitivity after accounting for the systematic uncertainties discussed next. The fit is performed simultaneously over all analysis categories.

The relative uncertainties on the signal strength of the Higgs boson decays considered in this analysis are summarized in Table 21, and correspond for one interaction point and $10.8\,\mathrm{ab}^{-1}$ of integrated luminosity.

Table 21: Expected Precision for the $\sigma(ZH) \times \mathrm{BR}(H \to jj)$ at 68 % CL

ZH final state		240 (%)	240 & MCStat (%)
$Z \to qq$	$H \to b\bar{b}$	0.21	0.32
	$H \to c\bar{c}$	2.32	3.52
	$H \to gg$	1.74	3.07
	$H \to s\bar{s}$	317.25	480.55
	$H \to \tau\tau$	66.65	110.73
	$H \to WW$	5.15	8.74
	$H \to ZZ$	32.00	52.08

8.2 Event selection at $\sqrt{s} = 365$ GeV

The event selection is similar to the $\sqrt{s} = 240$ GeV presented above except from the $d_{1,2}$, $d_{2,3}$ and $d_{3,4}$ which are tuned for this center of mass energy.

8.2.1 Fit Results

The relative uncertainties on the signal strength of the Higgs boson decays considered in this analysis are summarized in Table 22, and correspond for one interaction point and 3.0 ab^{-1} of integrated luminosity.

Table 22: Expected Precision for the $\sigma(ZH) \times \mathrm{BR}(H \to jj)$ at 68 % CL

ZH final state		365 (%)	365 & MCStat (%)
$Z \to qq$	$H \to b\bar{b}$	0.51	0.56
	$H \to c\bar{c}$	3.87	4.77
	$H \to gg$	3.05	4.45
	$H \to s\bar{s}$	563	718
	$H \to \tau\tau$	258	364
	$H \to WW$	5.72	7.12
	$H \to ZZ$	55.7	68.73

9 Global Fit

A global fit is performed from all the three different Z decay modes described above including either the $\nu\bar{\nu}jj$ analysis from Section 6 or the Section 7. The sensitivity of the Higgs boson in different decay modes is shown in Table 23. In the case of H$\to$ s$\bar{\text{s}}$ we calculated the upper limit at 95% confidence level on $B(H \to \mathrm{s\bar{s}})$ at $1.6 \times$ SM prediction.

Table 23: Expected Precision (%) for the $\sigma(ZH) \times \mathrm{BR}(H \to jj)$ at 68 % CL

	$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	$H \to s\bar{s}$	$H \to \tau\tau$	$H \to ZZ$	$H \to WW$
$Z \to ll$ ($l = \mu, e$)	0.60	3.47	1.93	220	2.54	7.65	1.49
$Z \to qq$	0.32	3.52	3.07	410	110	50	8.74
$Z \to \nu\bar{\nu}$ (I)	0.35	2.06	1.01	100	10.6	11.4	1.28
$Z \to \nu\bar{\nu}$ (II)	0.33	2.27	0.94	140	21.8	19.8	1.89
Combined ($Z \to \nu\nu$ (I))	0.21	1.56	0.85	89	2.46	6.24	0.95
Combined ($Z \to \nu\nu$ (II))	0.21	1.66	0.80	105	3.97	10.1	1.16

Table 24: Relative Uncertainty (%) for the $\sigma_{ZH} \times \mathcal{B}(H \to XX)$ and $\sigma_{\nu_e\bar{\nu}_e} \times \mathcal{B}(H \to XX)$ at 68 % CL

$\sqrt{s}$	240 GeV		365 GeV	
Integrated luminosity	10.8 ab^{-1}		3.0 ab^{-1}	
Channel	ZH	$\nu_e\bar{\nu}_e H$	ZH	$\nu_e\bar{\nu}_e H$
$H \to b\bar{b}$	±0.21	±1.89	±0.41	±0.67
$H \to c\bar{c}$	±1.61	±19.4	±3.13	±3.49
$H \to s\bar{s}$	±120	±990	±360	±290
$H \to gg$	±0.80	±5.50	±2.21	±2.66
$H \to WW$	±1.17	±15.6	±3.18	±5.36
$H \to ZZ$	±9.94	±130	±26.0	±37.1
$H \to \tau^+\tau^-$	±3.67	∞	±11.0	±24.2

10 Conclusion

A study of the sensitivity to the measurement of the signal strengths for Higgs boson decays to $b\bar{b}$, $c\bar{c}$, gg and $s\bar{s}$ in ZH and VBF events with Z boson decays to pairs of neutrinos, electrons or muons or quarks has been presented.

The results assume an integrated luminosity of 10.8 ab^{-1} at $\sqrt{s} = 240$ GeV and 3.0 ab^{-1} at $\sqrt{s} = 365$ GeV of e^+e^- collisions at FCC-ee, reconstructed with a modified version of the IDEA detector concept.

Generator-level events are passed through a simplified simulation and event reconstruction chain, and list of particle candidates are then analysed to identify and properly reconstruct the final states of interest.

In the charged-lepton channel, the ZH signal strengths are expected to be measured with precisions of 0.6%, 3.5%, 1.9% for $H \to b\bar{b}$, $H \to c\bar{c}$, $H \to gg$. The uncertainty on the $H \to s\bar{s}$ signal strength is of the order of 220%.

In the neutrino channel, the ZH signal strengths are expected to be measured with precisions of 0.35%, 2.1%, 1.0% and 100% for $H \to b\bar{b}$, $H \to c\bar{c}$, $H \to gg$ and $H \to s\bar{s}$. An alternative analysis yields similar results (0.33%, 2.3%, 0.9%, 140%) while constraining a bit worse other Higgs boson decays such as those to WW (1.9% vs 1.3%).

In the fully hadronic channel, the ZH signal strengths are expected to be measured with precisions of 0.32%, 3.5%, 3.1% amd 410% for $H \to b\bar{b}$, $H \to c\bar{c}$, $H \to gg$ and $H \to s\bar{s}$. The combination of the three Z boson final states leads to expected uncertainties of 0.21%, 1.6%, 0.8%, 90% on $H \to b\bar{b}$, $H \to c\bar{c}$, $H \to gg$ and $H \to s\bar{s}$, and an uncertainty around 1.0% of $H \to WW$.

Significant contribution to the $\nu\nu H$ final state, especially at $\sqrt{s} = 365$ GeV, is expected from W-boson fusion events where the colliding beam particles emit a neutrino-antineutrino pair and two W bosons that fuse into a Higgs boson. A dedicated analysis that measures separately the $Z(\nu\bar{\nu})H$ and $\nu_e\bar{n}u_eH$ contributions yields an expected precision on the $\nu_e\bar{n}u_eH$ signal strength below %-level for $H \to b\bar{b}$ and few-% level for $H \to c\bar{c}$ and $H \to gg$.

Acknowledgments. We would like to thank the various intern students at the APC laboratory that contributed to various aspects of this study: Paul Paquiez, Mariette Jolly, Paul Guimbard.

This work was funded in part by the U.S. Department of Energy, Office of Science, High Energy Physics under Contracts DE-SC0012704.

Appendix A Technical details about the MC samples

The generator steering files for the MC samples produced with WHIZARD and Pythia 8 are located in https://github.com/HEP-FCC/FCC-config/tree/winter2023/FCCee/Generator.

The Delphes card with the parameterisation of the IDEA detector performance is located at the following url: https://github.com/HEP-FCC/FCC-config/blob/winter2023/FCCee/Delphes/card_IDEA.tcl.

Appendix B Fit Tables

Below are the covariance and correlation matrices as obtained from the global fit.

Correlation matrix for 240 GeV

	zhss_mu	zhbb	zhcc	zhgg	vbfhbb	vbfhcc	vbfhss	vbfhgg
zhss	1.0000	0.0797	0.0623	0.0190	-0.0079	-0.0044	-0.2978	-0.0028
zhbb	0.0797	1.0000	0.0752	0.0238	-0.2159	-0.0125	-0.0131	-0.0037
zhcc	0.0623	0.0752	1.0000	0.0385	-0.0194	-0.2729	-0.0180	-0.0098
zhgg	0.0190	0.0238	0.0385	1.0000	0.0000	0.0033	0.0048	-0.2828
vbfhbb	-0.0079	-0.2159	-0.0194	0.0000	1.0000	0.0853	0.1699	0.0388
vbfhcc	-0.0044	-0.0125	-0.2729	0.0033	0.0853	1.0000	0.1062	0.0554
vbfhss	-0.2978	-0.0131	-0.0180	0.0048	0.1699	0.1062	1.0000	0.0757
vbfhgg	-0.0028	-0.0037	-0.0098	-0.2828	0.0388	0.0554	0.0757	1.0000

Covariance matrix for 240 GeV

	zhss	zhbb	zhcc	zhgg	vbfhbb	vbfhcc	vbfhss	vbfhgg
zhss_mu	1.3734e+00	1.9362e-04	1.1709e-03	1.7923e-04	-1.7602e-04	-1.0043e-03	-3.4694e+00	-1.8149e-04
zhbb	1.9362e-04	4.3020e-06	2.5031e-06	3.9704e-07	-8.4618e-06	-5.0162e-06	-2.6938e-04	-4.2165e-07
zhcc	1.1709e-03	2.5031e-06	2.5725e-04	4.9602e-06	-5.8727e-06	-8.4668e-04	-2.8683e-03	-8.6189e-06
zhgg	1.7923e-04	3.9704e-07	4.9602e-06	6.4607e-05	4.7258e-09	5.1100e-06	3.8184e-04	-1.2509e-04
vbfhbb	-1.7602e-04	-8.4618e-06	-5.8727e-06	4.7258e-09	3.5699e-04	3.1161e-04	3.1918e-02	4.0372e-05
vbfhcc	-1.0043e-03	-5.0162e-06	-8.4668e-04	5.1100e-06	3.1161e-04	3.7426e-02	2.0419e-01	5.8956e-04
vbfhss	-3.4694e+00	-2.6938e-04	-2.8683e-03	3.8184e-04	3.1918e-02	2.0419e-01	9.8823e+01	4.1383e-02
vbfhgg	-1.8149e-04	-4.2165e-07	-8.6189e-06	-1.2509e-04	4.0372e-05	5.8956e-04	4.1383e-02	3.0280e-03

Correlation matrix for 365 GeV

	zhss	zhbb	zhcc	zhgg	vbfhbb	vbfhcc	vbfhss	vbfhgg
zhss	1.0000	0.1959	0.2297	0.0486	-0.0259	-0.0333	-0.1479	-0.0087
zhbb	0.1959	1.0000	0.1683	0.0600	-0.1278	-0.0307	-0.0484	-0.0123
zhcc	0.2297	0.1683	1.0000	0.0940	-0.0294	-0.1509	-0.0522	-0.0185
zhgg	0.0486	0.0600	0.0940	1.0000	-0.0123	-0.0150	-0.0188	-0.1319
vbfhbb	-0.0259	-0.1278	-0.0294	-0.0123	1.0000	0.1172	0.2225	0.0267
vbfhcc	-0.0333	-0.0307	-0.1509	-0.0150	0.1172	1.0000	0.1915	0.0894
vbfhss	-0.1479	-0.0484	-0.0522	-0.0188	0.2225	0.1915	1.0000	0.0729
vbfhgg	-0.0087	-0.0123	-0.0185	-0.1319	0.0267	0.0894	0.0729	1.0000

Covariance matrix for 365 GeV

	zhss	zhbb	zhcc	zhgg	vbfhbb	vbfhcc	vbfhss	vbfhgg
zhss	1.2682e+01	2.8540e-03	2.5633e-02	3.8262e-03	-6.1799e-04	-4.1406e-03	-1.5266e+00	-8.1876e-04
zhbb	2.8540e-03	1.6730e-05	2.1567e-05	5.4225e-06	-3.5021e-06	-4.3823e-06	-5.7339e-04	-1.3351e-06
zhcc	2.5633e-02	2.1567e-05	9.8193e-04	6.5049e-05	-6.1653e-06	-1.6518e-04	-4.7378e-03	-1.5412e-05
zhgg	3.8262e-03	5.4225e-06	6.5049e-05	4.8804e-04	-1.8197e-06	-1.1549e-05	-1.2062e-03	-7.7405e-05
vbfhbb	-6.1799e-04	-3.5021e-06	-6.1653e-06	-1.8197e-06	4.4883e-05	2.7430e-05	4.3193e-03	4.7570e-06
vbfhcc	-4.1406e-03	-4.3823e-06	-1.6518e-04	-1.1549e-05	2.7430e-05	1.2202e-03	1.9387e-02	8.2979e-05
vbfhss	-1.5266e+00	-5.7339e-04	-4.7378e-03	-1.2062e-03	4.3193e-03	1.9387e-02	8.3978e+00	5.6081e-03
vbfhgg	-8.1876e-04	-1.3351e-06	-1.5412e-05	-7.7405e-05	4.7570e-06	8.2979e-05	5.6081e-03	7.0547e-04

References

[1] ATLAS Collaboration: Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B **716**, 1 (2012) arXiv:1207.7214 [hep-ex]. https://doi.org/10.1016/j.physletb.2012.08.020

[2] CMS Collaboration: Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B **716**, 30 (2012) arXiv:1207.7235 [hep-ex]. https://doi.org/10.1016/j.physletb.2012.08.021

[3] Englert, F., Brout, R.: Broken Symmetry and the Mass of Gauge Vector Mesons. Phys. Rev. Lett. **13**, 321–323 (1964). https://doi.org/10.1103/PhysRevLett.13.321

[4] Higgs, P.W.: Broken symmetries, massless particles and gauge fields. Phys. Lett. **12**, 132–133 (1964). https://doi.org/10.1016/0031-9163(64)91136-9

[5] de Florian, D., *et al.*: Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. CERN-2017-002-M (2016) arXiv:1610.07922 [hep-ph]. https://doi.org/10.23731/CYRM-2017-002

[6] Erdelyi, B.A., Gröber, R., Selimovic, N.: How large can the Light Quark Yukawa couplings be? (2024) arXiv:2410.08272 [hep-ph]

[7] ATLAS Collaboration: Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector. Phys. Lett. B **786**, 59 (2018) arXiv:1808.08238 [hep-ex]. https://doi.org/10.1016/j.physletb.2018.09.013

[8] CMS Collaboration: Observation of Higgs Boson Decay to Bottom Quarks. Phys. Rev. Lett. **121**, 121801 (2018) arXiv:1808.08242 [hep-ex]. https://doi.org/10.1103/PhysRevLett.121.121801

[9] ATLAS Collaboration: Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector. CERN-EP-2021-251 (2022) arXiv:2201.11428 [hep-ex]

[10] CMS Collaboration: Search for Higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV. CERN-EP-2022-081 (2022) arXiv:2205.05550 [hep-ex]

[11] Kasieczka, G., Kiefer, N., Plehn, T., Thompson, J.M.: Quark-Gluon Tagging: Machine Learning vs Detector. SciPost Phys. **6**, 69 (2019). https://doi.org/10.21468/SciPostPhys.6.6.069

[12] Cepeda, M., *et al.*: Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. CERN Yellow Rep. Monogr. **7**, 221–584 (2019)

arXiv:1902.00134 [hep-ph]. https://doi.org/10.23731/CYRM-2019-007.221

[13] Expected sensitivity of the ATLAS experiment to $H \to b\bar{b}$ and $H \to c\bar{c}$ decays in the VH production mode at the High Luminosity LHC. Technical report, CERN, Geneva (2025). All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2025-012. https://cds.cern.ch/record/2927842

[14] Abada, A., *et al.*: FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2. Eur. Phys. J. ST **228**(2), 261–623 (2019). https://doi.org/10.1140/epjst/e2019-900045-4

[15] Lucchini, M.T., Chung, W., Eno, S.C., Lai, Y., Lucchini, L., Nguyen, M.-T., Tully, C.G.: New perspectives on segmented crystal calorimeters for future colliders. JINST **15**(11), 11005 (2020) arXiv:2008.00338 [physics.ins-det]. https://doi.org/10.1088/1748-0221/15/11/P11005

[16] Kilian, W., Ohl, T., Reuter, J.: WHIZARD: Simulating Multi-Particle Processes at LHC and ILC. Eur. Phys. J. C **71**, 1742 (2011) arXiv:0708.4233 [hep-ph]. https://doi.org/10.1140/epjc/s10052-011-1742-y

[17] Sjöstrand, T., Ask, S., Christiansen, J.R., Corke, R., Desai, N., Ilten, P., Mrenna, S., Prestel, S., Rasmussen, C.O., Skands, P.Z.: An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159–177 (2015) arXiv:1410.3012 [hep-ph]. https://doi.org/10.1016/j.cpc.2015.01.024

[18] de Favereau, J., Delaere, C., Demin, P., Giammanco, A., Lemaître, V., Mertens, A., Selvaggi, M.: DELPHES 3, A modular framework for fast simulation of a generic collider experiment. JHEP **02**, 057 (2014) arXiv:1307.6346 [hep-ex]. https://doi.org/10.1007/JHEP02(2014)057

[19] Selvaggi, M.: DELPHES 3: A modular framework for fast-simulation of generic collider experiments. J. Phys. Conf. Ser. **523**, 012033 (2014). https://doi.org/10.1088/1742-6596/523/1/012033

[20] Selvaggi, M.: Delphes 3: Latest Developments. J. Phys. Conf. Ser. **762**(1), 012051 (2016). https://doi.org/10.1088/1742-6596/762/1/012051

[21] Sjostrand, T., Mrenna, S., Skands, P.Z.: PYTHIA 6.4 Physics and Manual. JHEP **05**, 026 (2006) arXiv:hep-ph/0603175. https://doi.org/10.1088/1126-6708/2006/05/026

[22] Catani, S., Dokshitzer, Y.L., Olsson, M., Turnock, G., Webber, B.R.:

New clustering algorithm for multi-jet cross-sections in e^+e^- annihilation. Phys. Lett. B **269**, 432–438 (1991). https://doi.org/10.1016/0370-2693(91)90196-W

[23] Bedeschi, F., Gouskos, L., Selvaggi, M.: Jet flavour tagging for future colliders with fast simulation. Eur. Phys. J. C **82**(7), 646 (2022) arXiv:2202.03285 [hep-ex]. https://doi.org/10.1140/epjc/s10052-022-10609-1

[24] Barger, V.D., Cheung, K.-m., Djouadi, A., Kniehl, B.A., Zerwas, P.M.: Higgs bosons: Intermediate mass range at e+ e- colliders. Phys. Rev. D **49**, 79–90 (1994) arXiv:hep-ph/9306270. https://doi.org/10.1103/PhysRevD.49.79

[25] Mahlon, G., Parke, S.J.: Deconstructing angular correlations in Z H, Z Z, and W W production at LEP-2. Phys. Rev. D **58**, 054015 (1998) arXiv:hep-ph/9803410. https://doi.org/10.1103/PhysRevD.58.054015