Sensitivity study of the $Hb\bar{b}$ and $Hc\bar{c}$ couplings for e^+-e^- collisions at FCC

Lea Heckmann
Supervisors: Michele de Gruttola, Gigi Rolandi

September 4, 2017

Abstract

The Future Circular Collider (FCC) is one of the possible future projects proposed at CERN. With a circumference of 80 - 100km it should initially serve as an e^+- e^- collider reaching energies between 90 - 350 GeV. While operating at centre of mass energies of 240 GeV and luminosities of $5*10^{34}$ $cm^{-2}s^{-1}$ FCC should be used as a Higgs factory with the aim of validating standard model predictions. Therefore, it should be able to measure the Higgs couplings with precisions around 1%.

In order to estimate the sensitivity of the branching ratio for $H \rightarrow c\bar{c}$ (BR(Hcc)) a simultaneous simulation and analysis of $Hb\bar{b}$ and $Hc\bar{c}$ is necessary. Therefore, events were generated for both cases and afterwards the CMS detector with an added c-tagging method was used to simulate the detector response. Relevant events were selected and afterwards used for optimising a simultaneous fit determining BR(Hbb) and BR(Hcc) and its uncertainties. The same procedure was applied while changing the efficiencies in the c-tagging algorithm in order to determine the influence of these efficiencies on the sensitivity.

The initial conditions lead to a sensitivity of around 17%. Increasing the c-tagging by 10% results in a precision improvement of 1-1.5% enhancing the precision up to around 13%. However, no change in the precision was observed while decreasing the misidentification efficiencies in the c-tagging method.

1 Introduction

Since decades CERN has been trying to explain the world around us and reveal the fundamental constituents it is made of. Nowadays the standard model is one of the most favoured and proven models to describe these fundamental particles. With the discovery of the Higgs boson it has been confirmed even stronger. However, there are still a large amount of open question, which have not yet been answered. Therefore it is important to measure the standard model parameters precisely and compare them to the theoretical predictions.

One of the possible future project with the aim of constraining standard model predictions is the Future Circular Collider (FCC). It is planned to replace the LHC after its high luminosity phase and will be a circular collider with a circumference between 80 to 100 km. As a first step the FCC should be used as an e^+- e^- collider for the purpose of validating standard model prediction. Afterwards a hadron collider should replace the previous one in order to discover new physics while reaching centre of mass energy of up to 100 TeV.

The FCC e^+- e^- collider (TLEP) is planned to operate at energies between 90 and 350 GeV and should be used to make precision measurements at the Z pole, the WW threshold, the HZ cross section maximum and at the $t\bar{t}$ threshold. Operated at energies of 240 GeV, which represents the maximum of the HZ cross section, with an estimated luminosity of $5*10^{34}$ $cm^{-2}s^{-1}$, TLEP can be used as a Higgs factory helping to measure the couplings of the Higgs boson to its decay products. The couplings can be measured by determining the branching ratios of the different decay channels, which represent the probability of a Higgs decaying in this channel with respect to all possible decay channels. In order to discover new physics the couplings have to be determined with a precision of around 1%. Predictions made for the sensitivity of the Hbb and Hcc branching ratio are inside that range as can be seen in figure 1.Through simulations it has been ascertained that CMS with some improvements would be a good choice for these measurements. [1].

	TLEP 240	ILC 250
σ_{HZ}	**0.4%**	2.5%
$\sigma_{HZ} \times BR(H \to b\bar{b})$	**0.2%**	1.1%
$\sigma_{HZ} \times BR(H \to c\bar{c})$	**1.2%**	7.4%

Figure 1: Predictions made for the branching ratio sensitivity by taking into account all of the possible Z decay channels and 4 detectors at TLEP. The data for TLEP was obtained by extrapolating the predictions made for ILC to the relevant luminosity [1].

In order to be able to measure the coupling of Hcc a c-tagging has to be implemented in the CMS detector. C-tagging describes a method used to identify c-jets in the detector while b-tagging does the same for b-jets. Up to now only b-tagging algorithms are included in CMS since the resolution of the tracking system is not high enough to tag c-jet. But with the assumption that it will be high enough at the time of FCC the c-tagging can be implemented analogously to te recent b-tagging [2]. However, background from $H \to b\bar{b}$ would still absorb most of the $H \to c\bar{c}$ signal due to the much higher branching ratio. Therefore, a simultaneous analysis and fit of Hbb and Hcc is necessary.

Simulations for the determination of the coupling uncertainties are necessary to estimate the importance of TLEP. The following report will outline the applied methods used for the sensitivity studies on $H \to c\bar{c}$.

2 Methods

In order to be able to determine the branching ratio precision, three steps had to be applied.

1. Events resulting from e^+- e^- collisions had to be generated.
2. A simulation for the adapted CMS detector recording these events had to be applied.
3. The output of the simulation had to be analysed and fitted.

This was done by taking the procedures described in [5] as a basic method and extending them corresponding to the new approach. Afterwards the same method was applied for different cases of c-tagging efficiencies with the aim of comparing their influences on the branching ratio precision.

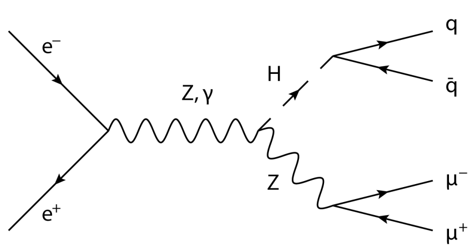
The single steps will be explained in the following sections.

2.1 Event Generation

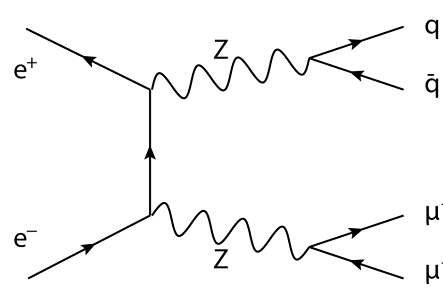
For the event generation Pythia8 was used, which is one of the standard tools for generating high-energy collision events especially concerning hadron processes and interactions [3].

Four samples were generated for each simulation case. Two for the signals producing a Z and a Higgs boson, with the restriction for Higgs decaying either into a $b\bar{b}$ pair in case of the first sample or $c\bar{c}$ in case of the second sample. The same applies to the two background samples, only that two Z bosons are produced with one of them having the same two decay cases. For each of the samples 100,000 e^+- e^- collision were simulated with an initial energy of 240 GeV, which corresponds to the maximum of the ZH cross section. In addition the Higgs mass was fixed to 125 GeV.

The production and decay channels for the background and signal events relevant for our analysis can be seen in figure 2, where the produced quark pair can either be a $b\bar{b}$ pair or a $c\bar{c}$ pair. Therefore, restriction on the decay channel of the second boson into two muons were made in addition two the ones mentioned above.



(a)



(b)

Figure 2: Feynman diagram displaying the relevant production and decay channels for a) the signal and b) the background events.

The resulting output was written to hepmc files which could then be used as an input for the simulation of the CMS detector response.

2.2 Detection

The C++ framework Delphes was used to simulate the CMS detector registering the produced events. Delphes is able to perform fast multipurpose detector response simulations and includes all of the necessary detector systems [4].

The existing CMS structure could be used with small modulations concerning the c-tagging since that has not been included before. The implemented b-tagging is based on the use of efficiency formulas, which describe the probability of a jet being tagged as b-jet, misidentified as a b-jet while being a c-jet or misidentified by default depending on their transverse momentum. The new c-tagging method was implemented analogously to the existing b-tagging algorithm. The misidentification will be also referred to as mistagging in the next sections.

The output of the detector response was saved in root files containing the data of the registered events. Theses could afterwards be used for the analysis.

2.2.1 Tagging Efficiencies

For the ctagging algorithm it the following tagging and mistagging efficiency equations were used.

$$\mathrm{eff(ctag)} = 0.85 * \tanh(0.0025 * \mathrm{p_T}) * \frac{25}{1 + 0.063 * \mathrm{p_T}} \tag{1}$$

$$\mathrm{eff(b, mistag)} = 0.25 * \tanh(0.018 * \mathrm{p_T}) * \frac{1}{1 + 0.0013 * \mathrm{p_T}} \tag{2}$$

$$\mathrm{eff}(\gamma, \mathrm{mistag}) = 0.01 + 0.000038 * \mathrm{p_T} \tag{3}$$

Where eff(ctag) describes the tagging efficiency for a c-jet being identified as a c-jet, eff(b,mistag) the misidentification probability of a b-jet being identified as a c-jet and eff(γ,mistag) the probability of a photon being tagged as a c-jet.

2.3 Data Analysis and Fits

The processing of the data was done in three steps. First the relevant events were selected, afterwards normalised and then fitted simultaneously with the aim of obtaining the precisions of BR(Hbb) and BR(Hcc).

2.3.1 Data Selection

For each of the four generated event files selection criteria were applied in order to distinguish the relevant events from the rest of the data. The first step was to only include events containing at least two muons with a transverse momentum $\mathrm{p_T}$ of more than 15 GeV/c and two b/c-jets with $\mathrm{p_T} > 15$ GeV/c and a pseudo rapidity η of less than 2.4. In addition only muons with a combined energy around the Z mass, so between 60 and 120 GeV, were taken into account. For theses events the missing mass of the corresponding b or c-jets was saved for further analysis. The missing mass $\mathrm{M_{miss}}$ is nothing else than the total initial energy in the collision subtracted by the energy of the two muons:

$$\mathrm{M^2_{miss}} = (\mathrm{s} - \mathrm{E}_{\mu1+\mu2})^2 - (\vec{\mathrm{p}}_{\mu1} + \vec{\mathrm{p}}_{\mu1})^2 \tag{4}$$

s represents the centre of mass energy squared, $\mathrm{E}_{\mu1+\mu2}$ the rest mass of the two muons and $\vec{\mathrm{p}}_{\mu1}$, $\vec{\mathrm{p}}_{\mu2}$ the momenta of the two muons.

2.3.2 Event Weighing

The next step was to weigh the obtained data. For weighing the simulated events the production cross sections as well as the different branching ratios and the integrated luminosity have to be known. These values can be found in table 1.

Table 1: Values used for normalising the event data.

σ_{ZH} [fb]	σ_{ZZ}[fb]	BR Hbb [%]	BR Hcc [%]	BR Zbb [%]	BR Zcc [%]	L [fb^{-1}]
3.962	4.753	58.07	2.883	15.12	12.03	3500

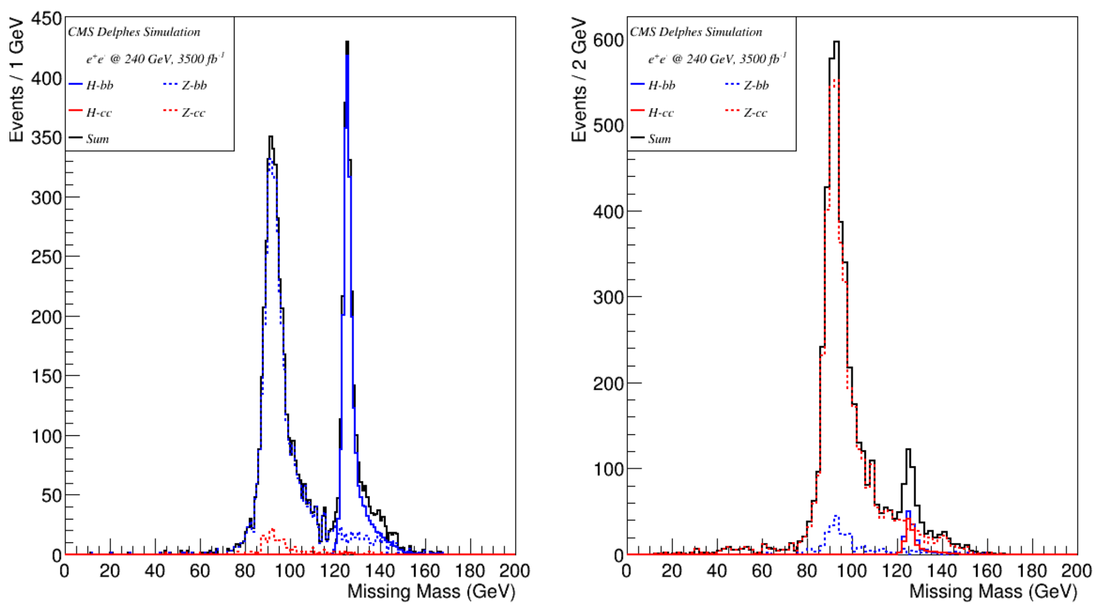


Figure 3: Missing mass histograms after normalisation for the case of 100% c-tag efficiency and 100% b-mistag efficiency. The histograms on the left display the data tag as Z,H-bb and the ones on the right as Z,H-cc.

After applying the normalisation the missing mass histograms were saved in a root file in order to be used for the following fitting procedure. An example can be seen in figure 3. The binning in the displayed range was set two 1 GeV for b-tag histograms and 2 GeV for the c-tag histograms due to conclusion made while optimising the fit described in the next section.

2.3.3 Fit

Afterwards a fit for the obtained data was optimised by using RooFit. This was done by simultaneously fitting for the branching ratios for Hbb and Hcc, since Hbb is a major background overlapping the Hcc signal otherwise.

For the signal events Crystal Ball functions were chosen as probability density functions with a mean value fixed at 125 GeV. The parameters of interest were the corresponding branching ratio, which are included in the number of signal events as can be seen below.

$$N_{sig}^{bb\ btag} = N_{Higgs} * \epsilon^{bb\ btag} * BR(Hbb) \tag{5}$$

$$N_{sig}^{bb\ ctag} = N_{Higgs} * \epsilon^{bb\ ctag} * BR(Hbb) \tag{6}$$

$$N_{sig}^{cc\ btag} = N_{Higgs} * \epsilon^{cc\ btag} * BR(Hcc) \tag{7}$$

$$N_{sig}^{cc\ ctag} = N_{Higgs} * \epsilon^{cc\ ctag} * BR(Hcc) \tag{8}$$

The constant parts of the numbers of signal events were obtained by integrating the ZH data histograms over the region of interest and dividing the result by using the branching ratios given in table 1. Therefore it was possible to only vary the two branching ratios instead of the four numbers of signal events during the fit.

First order polynomials were chosen as probability density functions for the backgrounds events. The number of background events was obtained by again integrating the ZZ data histograms over the region of interest and fixing theses values during the fit.

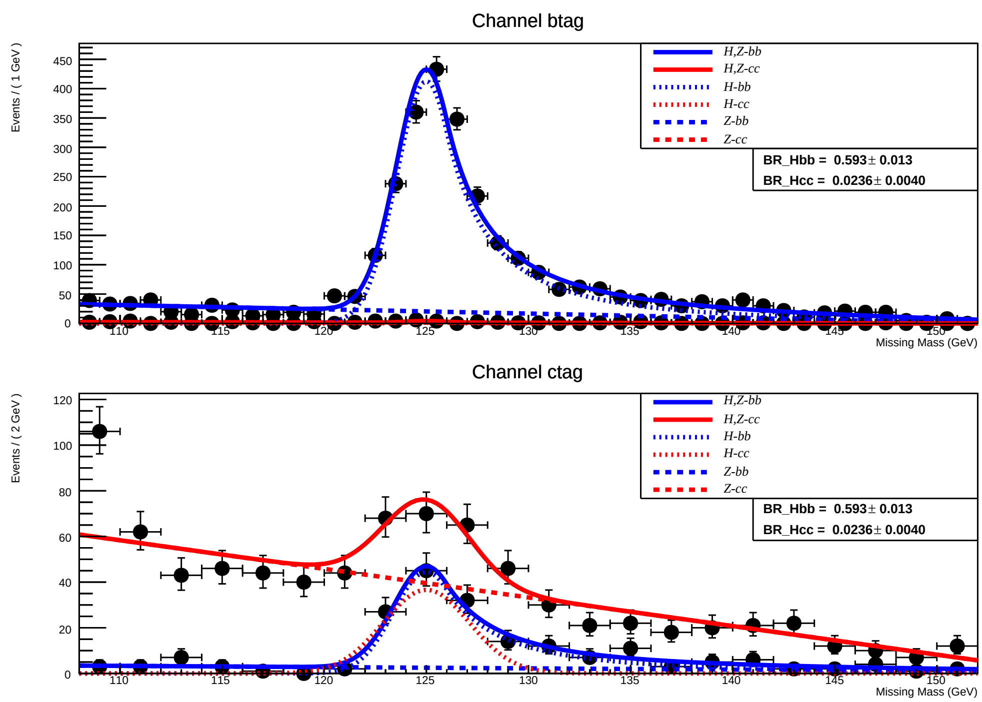
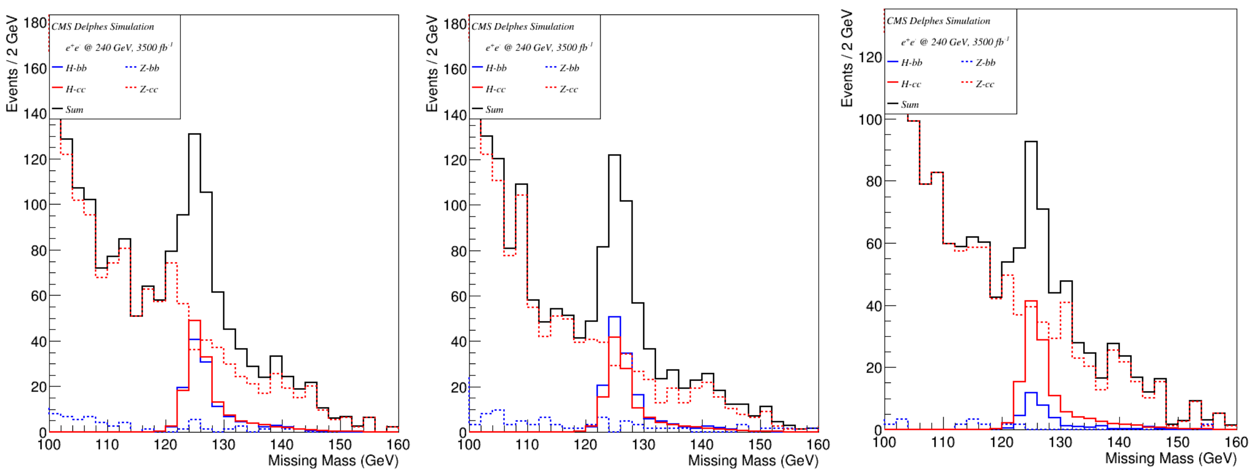


Figure 4: Fit resulting in a precision of 2.19% for BR(Hbb) and 16.95% for BR(Hcc).

The fitting range was chosen around the Higgs mass between 108 and 152 GeV due to the binning in the data histograms. Figure 4 displays the fitting result optimised to the obtained data. A precision of 2% could be reached for BR(Hbb) and 16.95% for BR(Hcc).

2.4 Variation of the Tagging Efficiencies

In order to see the influence of the c-tagging efficiencies on the sensitivity of the branching ratios eff(ctag) and eff(b,mistag) were varied inside the c-tagging. The c-tag identification efficiency was increased by 10%, 20% and 30% while the misidentification efficiency was decrease by 50%, 20% and 10%.



(a) (b) (c)

Figure 5: Comparison of the missing mas histogram for different tagging efficiencies. a) 20% increase on the efficiency of tagging a c-jet as a c-jet b) Initial tagging efficiencies c) 50% decrease on the efficiency of mistagging a b-jet as a c-jet

Figure 5 explains how the c-tagging efficiencies influence the generated events. While comparing a) and b) it can be seen that increasing the c-tagging efficiency leads to more events of the cc channel in the c-tagged data. By decreasing the misidentification efficiency less events of the bb channel are tagged as c-jet as shown by the comparison of b) and c).

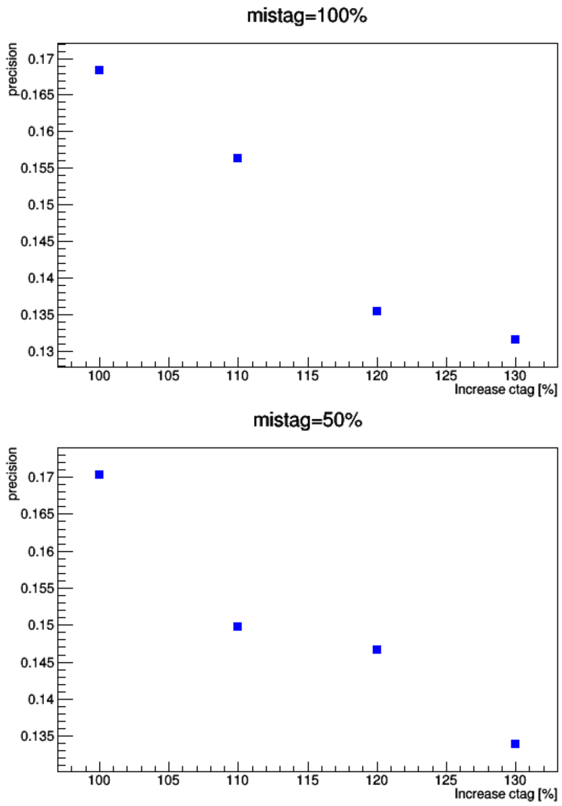
3 Results and Conclusion

By applying the procedure described above for all the different cases of c-tagging and mistagging efficiencies, the dependencies of the precision obtained for the BR(Hcc) on these efficiencies could be determined.

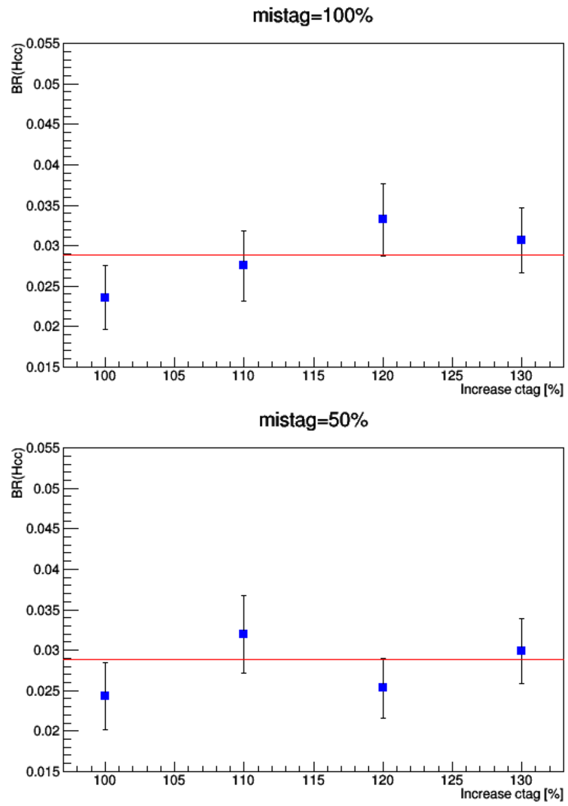
Figure 6 a) shows the dependence of the BR(HCC) sensitivity on the c-tagging efficiency. On the top only the eff(ctag) were varied, which results in an improvement of around 1-1.5 % in the precision while increasing the c-tagging efficiency by 10%. However, decreasing eff(b,mistag) does not seem to have an impact on the sensitivity as can be seen by comparing the two plots above each other. For the one on the bottom the mistagging efficiency has been decreased by 50% and eff(ctag) varied. Since it looks very similar to the plot on top, it can be concluded that the eff(b,mistag) has no influence on the sensitivity.

Figure 6 b) shows the obtained branching ratios and their errors. Except for the case without increasing the c-tagging efficiency all of the results match with the expected value for the branching ratio indicated by the red lines.

However, the precision is still far away from the required 1%, but one should keep in mind, that the applied methods were a fast simulation and a quite simple analysis. In addition only the $Z - \mu^+\mu^-$ channel was used during the simulation. Including $Z - e^+e^-$ should improve



(a)



(b)

Figure 6: a) Comparison of the BR(Hcc) precisions depending on the different increases in the c-tagging efficiency. b) Comparison of the obtained branching ratios and their errors depending on the different increases in the c-tagging efficiencies. The red line indicates the real value of the branching ratio.
Top: Without decreasing the mistagging efficiency. Bottom: While decreasing the mistagging efficiency by 50%.

the sensitivity by a factor of around 4. Furthermore four detectors were taken into account for the predictions while only one was used in this study. Therefore the precision should improve by another factor of 2. By including these factors the results coincide quite good with the predictions made in figure 1 and would allow TLEP to validate standard model predictions.

References

[1] M. Bicer et al., *First Look at the Physics Case of TLEP*, Journal of High Energy Physics, Volume 1, pp 164 (2014)

[2] CMS Collaboration, *Identification of c-quark jets at the CMS experiment*, CMS-PAS-BTV-16-001, `https://cds.cern.ch/record/2205149` (2016)

[3] T. Sjöstrand et al., *A brief introduction to PYTHIA 8.1*, Computer Physics Communications, Volume 178, pp 852-867 (2008)

[4] `https://cp3.irmp.ucl.ac.be/projects/delphes`, accessed 24.08.2017

[5] `https://twiki.cern.ch/twiki/bin/viewauth/CMS/FCCeeZHanalysis`, accessed 24.08.2017