CERN - European Organization for Nuclear Research

LCD-Note-2011-036

Light Higgs Production and Decays to Pairs of Bottom and Charm Quarks at 3 TeV

Tomáš Laštovička*

* *Institute of Physics, Academy of Sciences, Prague, Czech Republic*

September 6, 2012

Abstract

We present a study of the Higgs boson decay to $b\bar{b}$ and $c\bar{c}$ quark pairs. A standard model Higgs boson of 120 GeV mass, produced in the W^+W^- boson fusion process in e^+e^- collisions at $\sqrt{s} = 3$ TeV, was investigated using full detector simulation and reconstruction procedures. The analysis was performed in the framework of the CLIC_SiD detector concept. The considered decay modes contain two heavy flavour jets in the final state and require excellent flavour tagging through precise reconstruction of interaction and decay vertices in the detector. The statistical accuracy on the measurement of the cross section times branching ratio for an integrated luminosity of 2 ab^{-1} was estimated to be 0.23% and 3.1% for $b\bar{b}$ and $c\bar{c}$, respectively.

1. Introduction

A fundamental test of the Standard Model Higgs mechanism is the predicted linear scaling of the Higgs couplings to fermions in proportion to their masses. In e^+e^- annihilations the dominant Standard Model Higgs production mechanisms at 3 TeV are via W^+W^- or Z^0Z^0 fusion, see Figure 1 (left). For this study the Higgs mass is assumed to be 120 GeV and the corresponding production cross sections are 421 fb for the W^+W^- fusion and 42.6 fb for the Z^0Z^0 fusion. Figure 1 (right) illustrates the behavior of cross sections for various Higgs production processes as a function of centre-of-mass energy $\sqrt{s}$. At CLIC, with a total integrated luminosity of 2 ab^{-1}, one expects to produce about 842 thousand $h\nu_e\bar{\nu}_e$ events, allowing an accurate measurement of the branching ratios to heavy quarks as well as the measurement of the branching ratio of rare decays. The b- and c-modes contain two heavy flavour jets in the final state and require high quality flavour tagging in order to separate both jet flavours from each other as well as from light jets.

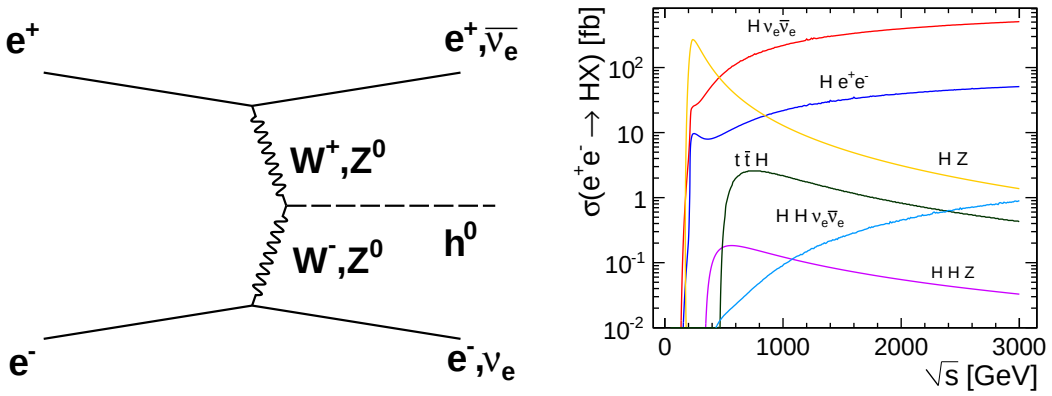

Figure 1: Dominant Higgs production mechanism at 3 TeV through W^+W^- and Z^0Z^0 fusion (left) and the cross sections as a function of $\sqrt{s}$ for $M_h = 120$ GeV (right).

The corresponding Higgs branching ratios in the SM for $h \to b\bar{b}$ and $h \to c\bar{c}$ are 64.7% and 3.3%, respectively [1]. The branching ratios in the e^+e^- collider environment have been studied before [2, 3] in the context of the International Linear Collider (ILC) running at $\sqrt{s} = 250$ GeV.

The paper is organized as follows: the analysis framework is introduced in Section 2; Section 3 describes the details of the jet flavour tagging and Section 4 presents the analysis and its results. The note is summarised in Section 5 and complemented by Appendix A and Appendix B.

2. Analysis Framework

The analysis is performed in the framework of the CLIC_SiD concept [3] using full detector simulation and event reconstruction. The physical processes are generated with the WHIZARD [4, 5] event generator, with fragmentation and hadronization done by PYTHIA [6]. The study assumes

Table 1: Production cross sections and simulated event statistics for the signal processes and for the backgrounds considered for the analysis of $h \to b\bar{b}$ and $h \to c\bar{c}$ decays.

Type	Final state	Cross section σ (fb)	Number of events
Signal	$h\nu_e\bar{\nu}_e$, $h \to b\bar{b}$	272	$1.3 \cdot 10^5$
Signal	$h\nu_e\bar{\nu}_e$, $h \to c\bar{c}$	13.7	$1.3 \cdot 10^5$
Background	$q\bar{q}\nu\bar{\nu}$	1305	$1.7 \cdot 10^5$
Background	$q\bar{q}e\bar{\nu}_e$	5255	$9.1 \cdot 10^4$
Background	$q\bar{q}e^+e^-$	3341	$9 \cdot 10^4$
Background	$q\bar{q}$	3076	$9.6 \cdot 10^4$

an integrated luminosity of $2\,\text{ab}^{-1}$ corresponding to 4 years of data taking at the nominal machine parameters.

The data samples considered for this analysis are listed in Table 1. They include both the Higgs samples and the dominant Standard Model background channels. The detector response to generated events was simulated using the GEANT4 toolkit [7, 8] through a thin layer of SiD specific code, SLIC [9].

Beamstrahlung effects on the luminosity spectrum as well as initial and final state radiation are taken into account. To approximate the CLIC beam structure and background conditions, the equivalent of 60 bunch crossing of $\gamma\gamma \to$ hadrons events were mixed with every simulated event. At 3 TeV, there is an average of 3.2 of these events per bunch crossing. A 10 ns readout window is assumed in the tracking detectors and in all calorimeters except the HCAL barrel, where 100 ns is assumed.

In addition to applying read-out windows off-line, the computation of the cluster time allows to further reduce this background. Assuming ns precision of the calorimeter hit times results in sub-ns precision for the cluster time, which is calculated as a truncated mean of the corresponding hit times. The production time of the reconstructed particle is obtained by correcting the cluster time for its time of flight through the magnetic field. The production time of the particle is required to be consistent with the start of the physics event. Consistency is defined by a time window, whose size depends on the type of particle (hadronic or electromagnetic), its momentum and polar angle θ. This reduces the energy from $\gamma\gamma \to$ hadrons processes to the event further by a factor of 6 or more, while only about 0.5% of the energy from a typical physics event is removed [10].

In the following, fragmentation products of the hadronic systems are forced to two jets using the exclusive k_t algorithm of the FASTJET package [11], where the parameter R is set to 0.7. The LCFIVertex package [12] is used to identify jets according to their quark content as b, c and light quarks and computes the corresponding jet flavour tag values.

The event classification is based on the open source Fast Artificial Neural Network (FANN) package [13]. FANN was modified to account for event weights during the neural network training.

3. Flavour Tagging

The resolution of secondary interactions from the primary vertex is an important criterion in the design of the inner tracking detectors. It is essential for the identification of bottom and charm decays. The flavour identification package LCFIVertex [12], developed by the LCFI collaboration, consists of a topological vertex finder ZVTOP, which reconstructs secondary interactions, and a multivariate classifier which combines several jet-related variables to tag bottom, charm, and light quark jets.

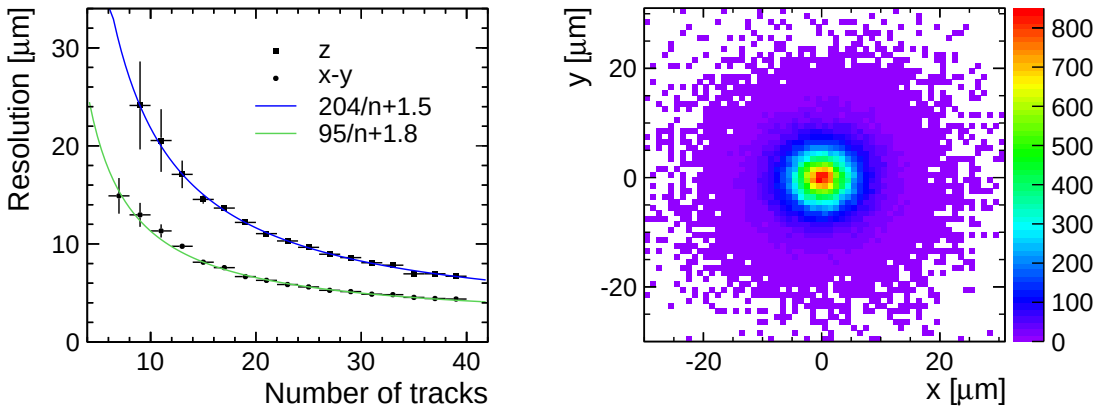

Figure 2: Position resolution of the primary vertex in the $x-y$ (green line) and z (blue line) direction as a function of the number of tracks in the vertex (left). The reconstructed position of primary vertices with more than 20 tracks in the plane transverse to the detector axis (right).

Figure 2 (left) shows the primary vertex reconstruction resolution in xy (green line) and z (blue line) coordinates as a function of the number of tracks in the vertex fit in $e^+e^- \to q\bar{q}\nu\bar{\nu}$ events. The mean jet energy is around 130 GeV. The events were simulated in the CLIC_SiD detector and the equivalent of 60 BX of $\gamma\gamma \to$ hadrons events was overlaid. Appendix A includes more details on how the primary vertex resolution was evaluated. The lines show functions of the form $1/N_{\text{Tracks}}$ which provide an empirical parametrisation of the resolution. The reconstructed primary vertex position in the xy plane transverse to the detector axis for vertices with more than 20 tracks is shown in Figure 2 (right).

Displaced vertices are the most significant characteristic of b quark decays. A combination of several vertex-related variables, complemented by additional track-related variables, form an input for the tagging classifier. The variables and the procedure is described in reference [12] and in Appendix B. In the following, all tagging classifier inputs were considered at the same time - as opposed to the default LCFI approach, where different variables are used for cases with 0, 1 and >1 secondary vertices found in a jet [12]. The rate with which jets of one flavour are tagged as a different flavour, the so called mis-tag efficiency, is used to assess the performance of the package. Figure 3 (left) shows the mis-tag efficiency for c-jets (blue line) and light jets (green

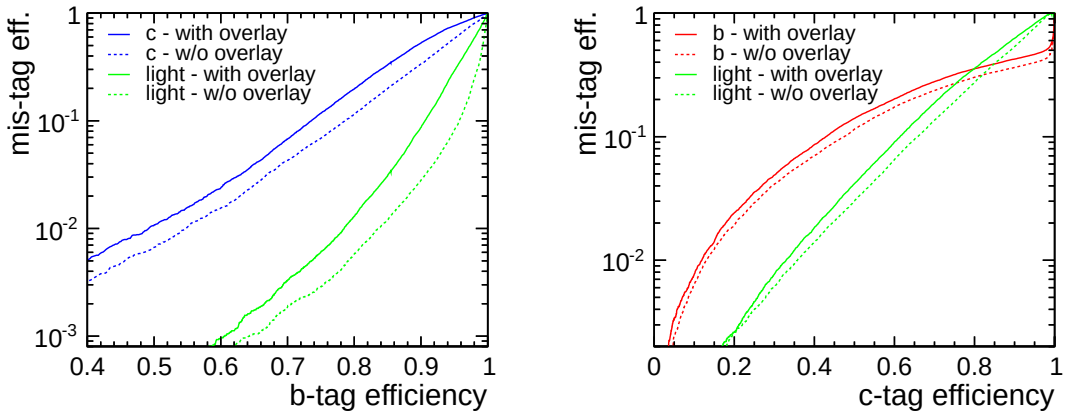

Figure 3: In the left plot, the mis-tag efficiency in the CLIC_SiD detector for charm (blue) and light (green) jets as a function of the b-tag efficiency is shown. The right plot shows the mis-tag efficiency for bottom (red) and light (green) jets as a function of the c-tag efficiency. The mean p_T of the jets is 70 GeV while the mean energy is $\sim$ 130 GeV.

line) as b-jets versus the b-tag efficiency, while Figure 3 (right) shows the mis-tag efficiency for b-jets (red line) and light jets (green line) as c-jets versus the c-tag efficiency. The presence of $\gamma\gamma \rightarrow$ hadrons background is found to reduce the flavour tagging performance, although the effect is not dramatic and, furthermore, it is correlated with significantly degraded jet finding quality. This means that the degradation of the flavour tag performance, shown in Figure 3, has two sources: the flavour tag degradation itself plus a degradation of the jet reconstruction quality due to a more difficult jet finding. For instance, at the b-tag efficiency of 70% the mis-tag efficiency for c-jets (light jets) increases from 4.3% (0.19%) w/o overlay to 6.8% (0.33%) with overlay.

4. Event Selection

The basic event selection requires two jets in each event. Apart from this selection, no further cuts are explicitly imposed and a number of relevant variables is given to neural net for the subsequent multivariate analysis. The $h \rightarrow b\bar{b}$ signal is compared to the background channels in Figure 4 as a distribution in the two most differentiating quantities, the invariant mass of the jets (left) and the sum of neural net jet b-flavour tag classifiers (right). These two quantities were complemented by the following variables to form an input to the signal selection neural net:

- The maximum of the absolute values of jet pseudorapidities;
- The sums of the remaining LCFI jet flavour tag values, i.e. c(udsb)-, c(b)- and b(uds)-tags[1)];

[1]The notion indicates which flavour is tagged against which set of other flavours. For instance, c(b) is the c-flavour tagged against the b-flavour only, while remaining (uds) flavours are not used during the neural net training.

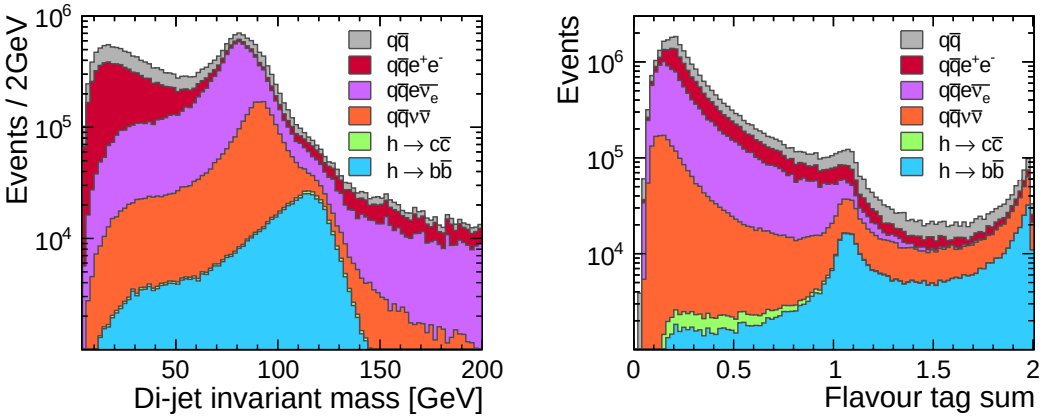

Figure 4: Distributions of the di-jet invariant mass (left) and the sum of the LCFI jet b-flavour tag values (right) for the $h \to b\bar{b}$ and $h \to c\bar{c}$ signals and for the individual backgrounds. All contributions are stacked and scaled to an integrated luminosity of 2 ab^{-1}.

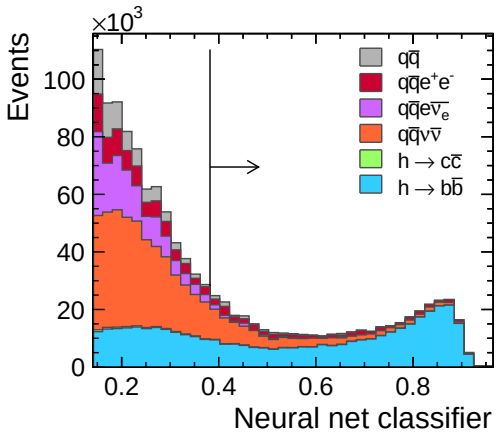

Figure 5: Distribution of the event selection neural network classifier for the $h \to b\bar{b}$ signal and for the individual backgrounds. The arrow illustrates the final event selection cut.

- The distance of jets in the $\eta - \phi$ plane $R_{\eta\phi}$;
- The sum of jet energies;
- The total number of leptons in an event;
- The total number of photons in an event;
- Acoplanarity of jets.

The neural net was trained to separate the $h \to b\bar{b}$ signal accounting for event weights, which are composed from generator level weights and luminosity normalisation weights. The ability of

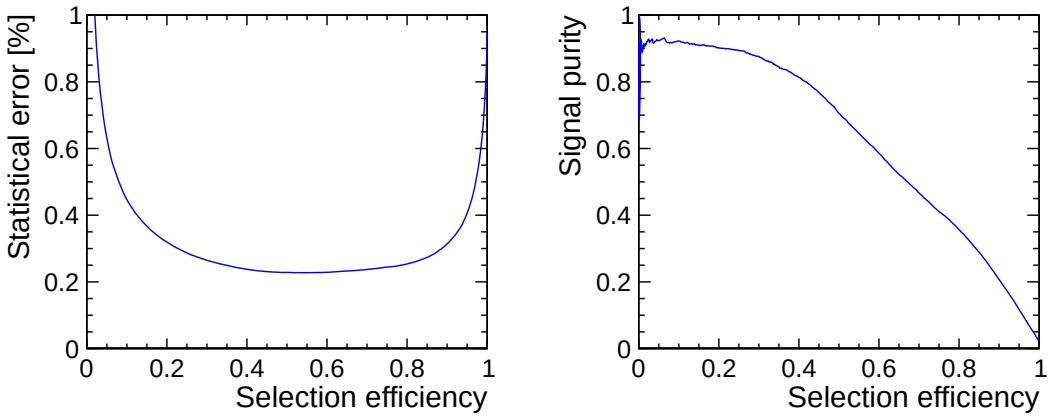

Figure 6: Cross section statistical uncertainty (left) and selected sample signal purity (right) as a function of a signal selection efficiency for the $h \to b\bar{b}$ channel. An integrated luminosity of $2\,ab^{-1}$ was assumed.

the neural net to separate the Higgs signal from the background is illustrated in Figure 5, where the region below 0.15 is not shown due to the very large number of background entries per bin. The arrow indicates the chosen event selection cut. The signal cross section uncertainty and the signal purity depend on the choice of this selection cut.

The performance of the neural net selection can be viewed in an alternative way, shown in Figure 6, where for a particular neural net classifier output cut a corresponding signal cross section measurement uncertainty and the signal purity[2)] is evaluated and plotted as a function of the signal selection efficiency. In the following, the point with the lowest statistical cross section uncertainty was chosen to be the neural net working point. Thus this cross section uncertainty does not include further systematic uncertainties, such as uncertainties of the purity evaluation and of the signal efficiency S/S_{total}, where S and B are the total numbers of selected signal and background events, respectively[3)].

The uncertainty distribution, presented in Figure 6 (left), has a wide flat bottom, thus allowing for the selection of an optimal working point over a wide range of selection efficiencies and signal purities while preserving low uncertainty values.

The results are summarised in Table 2. The statistical uncertainty on the cross section times branching ratio of the decay $h \to b\bar{b}$ is 0.23% with a corresponding signal selection efficiency of 51.8% and the sample purity of 68.4%.

[2]The statistical cross section uncertainty and the signal purity are calculated as $\sqrt{S+B}/S$ and $S/(S+B)$, respectively.

[3]In a real measurement a large sample of inclusive backgrounds would be simulated and normalised to various processes and channels. Its total normalisation uncertainty would then be transferred into the corresponding systematic uncertainty of selected background events. Such a procedure, however, is beyond the scope of this contribution.

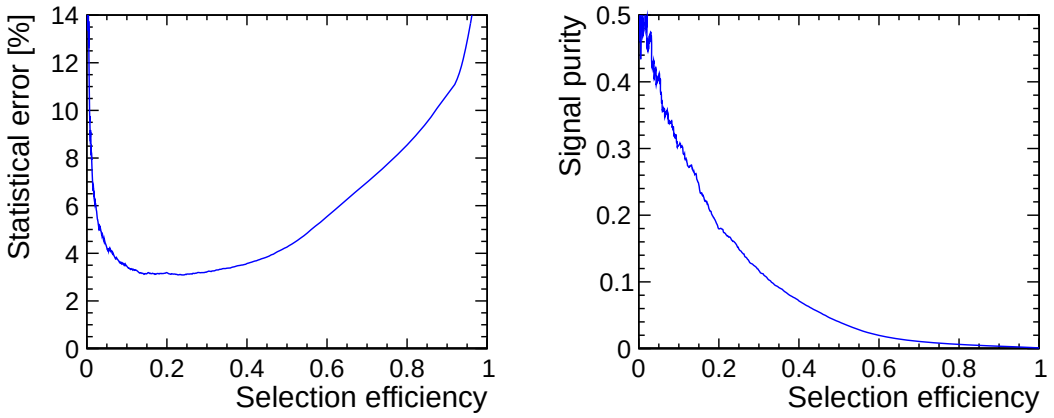

Figure 7: Cross section statistical uncertainty (left) and selected sample signal purity (right) as a function of a signal selection efficiency for the $h \to c\bar{c}$ channel. An integrated luminosity of $2\,ab^{-1}$ was assumed.

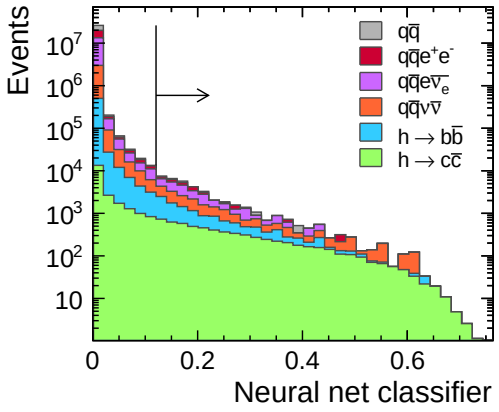

Figure 8: Distribution of the event selection neural network classifier for the $h \to c\bar{c}$ signal and for the individual backgrounds. The arrow illustrates the final event selection cut.

The same analysis procedure [4)] was repeated for the $h \to c\bar{c}$ channel leading to the statistical cross section uncertainty of 3.1 %, with a signal selection efficiency of 24% and the sample purity of 16%. The corresponding cross section statistical uncertainty and selected sample signal purity graphs are presented in Figure 7. The $h \to c\bar{c}$ channel is significantly more difficult to separate from the background and the purity never exceeds 50% for any neural net classifier output cut selection. Compared to $h \to b\bar{b}$ it is caused by a smaller $h \to c\bar{c}$ branching ratio of 3.3% and by a

[4]For the purpose of this analysis, the signal and background definitions in the network were reversed with respect to the $h \to b\bar{b}$ analysis.

Table 2: Signal selection performance and measured cross section × BR statistical uncertainties for $h \to b\bar{b}$ and $h \to c\bar{c}$ decays[5]. All numbers are obtained assuming an integrated luminosity of $2\,ab^{-1}$.

	$h \to b\bar{b}$	$h \to c\bar{c}$
Signal events	282×10^3	660×10^1
Total signal events	544×10^3	274×10^2
Background events	130×10^3	350×10^2
Signal purity	68.4%	16%
Signal efficiency	51.8%	24.1%
Cross section statistical uncertainty	0.23%	3.1%

lower c-flavour tagging performance. The ability of the neural net to separate the $h \to c\bar{c}$ signal from the background is illustrated in Figure 8. The lowest cross section statistical uncertainty is achieved at relatively low signal selection efficiency values of around 20%.

5. Results

The sensitivity to the decay branching ratios of a neutral 120 GeV SM Higgs boson to bottom and charm quarks has been studied at the CLIC centre-of-mass energy of $\sqrt{s} = 3$ TeV and integrated luminosity of $2\,ab^{-1}$. The analysis is based on full simulation and realistic event reconstruction in the CLIC_SiD detector. The statistical accuracy on the cross section times branching ratio measurement were estimated to be 0.23% and 3.1% for $b\bar{b}$ and $c\bar{c}$, respectively. This includes the effect of background from $\gamma\gamma \to$ hadrons on the flavour tagging. Good performance of flavour tagging and the use of neural networks in event selection were critical in obtaining these results.

Given the experience of the LEP experiments [15] in the measurements of hadronic Z decays, with systematic uncertainties between 0.3% - 1.2% for R^0_b and between 1.2% and 10% for R^0_c, one can assume that a systematic uncertainty of around 1% is achievable in $h \to b\bar{b}$ and around 5% in $h \to c\bar{c}$. Even the combination of the statistical uncertainty and a conservative average of the systematic uncertainties from similar measurements at LEP is dominated by the current theoretical uncertainties of 2.8% for $h \to b\bar{b}$ and 12.2% for $h \to c\bar{c}$.

[5]The numbers published in the CLIC Conceptual Design Report [14] differ slightly from the ones presented here, since revised Higgs production cross section and branching ratios were used in this note.

A. Primary vertex reconstruction resolution

The primary vertex reconstruction resolution, presented in Section 3, was studied in bins of a number of tracks used for the vertex fit. The corresponding distributions are shown in Figure 9 and Figure 10 for x and z coordinates, respectively. The distributions are fitted by two Gaussians and their area is used as a weight in order to evaluate a combined width of Gaussians, which is then plotted in Figure 2 (left). The errors are evaluated in a similar manner, i.e. by re-weighting Gaussian fit errors by Gaussian areas.

B. Jet flavour tagging inputs and the LCFIVertex package

The LCFIVertex package [12] implements software for vertex finding, flavour tagging and vertex charge reconstruction. The topological vertex finder ZVTOP, implemented in the package, was originally developed by the SLD experiment [16].

The basic concept of the vertex finding algorithm (ZVRES) is that each track is described by a probability density function in 3-D space which is used to define a vertex function that returns higher values near vertex locations and lower values elsewhere. It also provides criteria for when two vertex candidates are resolved from each other.

The LCFIVertex package also implements an algorithm which extracts MC true jet flavour, quark and leading hadron charge information needed for comparison and performance measure of the flavour tag and vertex/quark charge [17].

The variables most sensitive to jet flavour are defined for jets containing secondary vertices. An important feature of the package is a different treatment of jets with and without the secondary vertices. For the case with no secondary vertices, two tracks (most significant and second-most significant tracks) with the highest impact parameter significance in the $R-\phi$ plane are found in the input jet. Significance is defined as the ratio of the impact parameter to its uncertainty. The impact parameter significances and momenta of the tracks in the $R-\phi$ and $R-z$ planes are used as flavour tag inputs. The measured quantities from the second-most significant track help distinguish between c and b jets, which are more likely to have two tracks with high impact parameter significance, with one coming from leading hadron decay and the other from the charmed hadron produced in that decay.

Another important input variable is the *joint probability*, a quantity introduced by ALEPH [18]. The joint probability is calculated [12, 17] separately in $R-\phi$ and $R-z$ from the impact parameter significances of tracks in the jet. A distribution of unsigned impact parameter significances for IP tracks is considered known.

If there are secondary vertices found within a jet six more variables can be calculated that help in further differentiation of jet flavours. The variables are:

- The 3D decay length and decay length significance of the vertex;
- The total momentum |p| of the set of tracks assigned to the decay chain;
- The p_T corrected vertex mass, M_{P_T};
- The number of all non-IP tracks, $N_{trk,vtx}$;

– The secondary vertex probability of the tracks assigned to decay chain; vertex fitting is performed using these tracks and probability calculated from the fit χ^2.

The momentum |p| is the modulus of the vector sum of all decay chain track momenta. Fitting a common vertex to these tracks and calculating the probability from the χ^2 value yields the secondary vertex probability. The transverse momentum-corrected mass is then defined as $M_{P_T} = \sqrt{M_{Vtx}^2 + \mid p_T^{Vtx} \mid^2} + \mid p_T^{Vtx} \mid$.

The flavour tag input distributions are shown in Figure 11. Only the significances of the tracks with the highest significance are shown. In the number of vertices in a jet input the primary vertex is always counted, thus the distribution starts at 1.

The number of leptons in a jet, shown in Figure 11, was not used in the LCFIVertex package. It was added in this analysis to aid improving the jet flavour tag quality.

References

[1] A. Denner et al. Standard Model Higgs-Boson Branching Ratios with Uncertainties. *Eur.Phys.J.*, vol. C71 p. 1753, 2011.

[2] Y. Banda, T. Lastovicka, and A. Nomerotski. Higgs boson hadronic branching ratios at the ILC. *Phys. Rev. D*, vol. 82 p. 033013, 2010.

[3] H. Aihara, P. Burrows, and M. Oreglia, eds. *SiD Letter of Intent*. 2009. ArXiv:0911.0006v1.

[4] W. Kilian, T. Ohl, and J. Reuter. WHIZARD: Simulating multi-particle processes at LHC and ILC, 2007. arXiv:0708.4233v1 .

[5] M. Moretti, T. Ohl, and J. Reuter. O'Mega: An optimizing matrix element generator, 2001. arXiv:hep-ph/0102195v1.

[6] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, vol. 05 p. 026, 2006. hep-ph/0603175.

[7] S. Agostinelli and *et al*. Geant4: a simulation toolkit. *Nucl. Instrum. Methods Phys. Res.*, vol. A 506 pp. 250–303, 2003.

[8] J. Allison and *et al*. Geant4, developments and applications. *IEEE Trans. Nucl. Sci.*, vol. 53(1) pp. 270–278, 2006.

[9] SLIC. http://www.lcsim.org/software/slic/doxygen/html/.

[10] J. Marshall, A. Munnich, and M. A. Thomson. PFA: Particle flow performance at CLIC. vol. LCD-Note-2011-028, CERN, 2011.

[11] M. Cacciari and G. Salam. Dispelling the N3 myth for the Kt jet-finder. *Phys. Lett.*, 2006. B641 [hep-ph/0512210].

[12] A. Bailey and *et al.* (LCFI Collaboration). LCFIVertex package: vertexing, flavour tagging and vertex charge reconstruction with an ILC vertex detector. *Nucl. Instrum. Methods Phys. Res.*, vol. A 610 pp. 573–589, 2009.

[13] Fast Artificial Neural Network Library (FANN). http://leenissen.dk/fann/html/files/fann-h.html.

[14] L. Linssen et al., eds. *CLIC Conceptual Design Report: Physics and Detectors at CLIC*. CERN-2012-003. Geneva, 2012.

[15] ALEPH Collaboration et al. Precision electroweak measurements on the Z resonance. *Phys.Rept.*, vol. 427 pp. 257–454, 2006.

[16] D. J. Jackson. A topological vertex reconstruction algorithm for hadronic jets. *Nuclear Instruments and Methods in Physics Research*, vol. A(388) pp. 247–253, 1997.

[17] E. Devetak. *Measurement of the Top Quark Mass, Cross Section and Anomalous Couplings at the International Linear Collider*. Ph.D. thesis, University of Oxford, 2009.

[18] ALEPH Collaboration. *Physics Letters*, vol. B(313) p. 535, 1993.

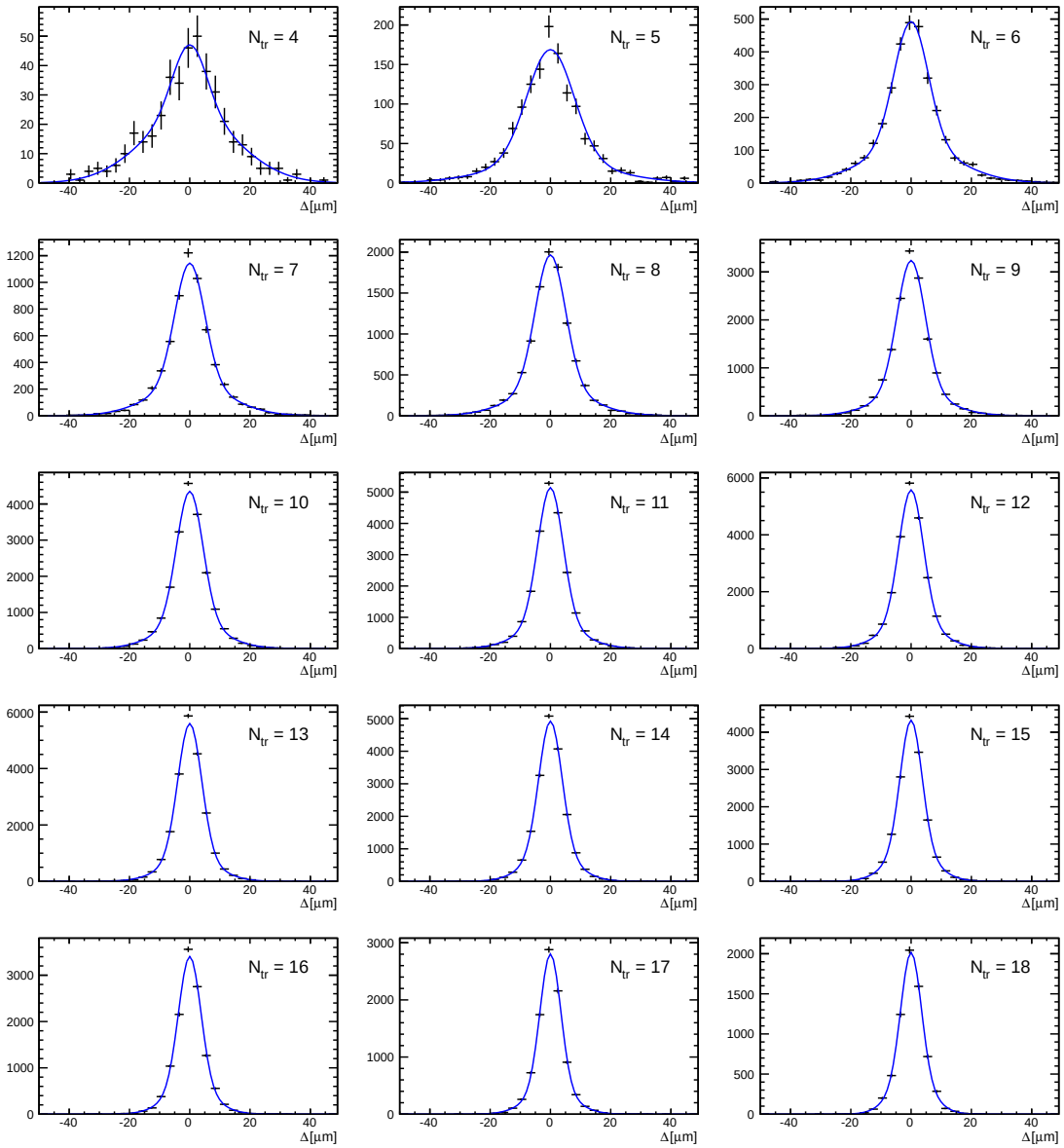

Figure 9: A difference between generated and reconstructed primary vertex x position in bins of number of tracks used for the primary vertex fit. Lines correspond to double-Gaussian fits.

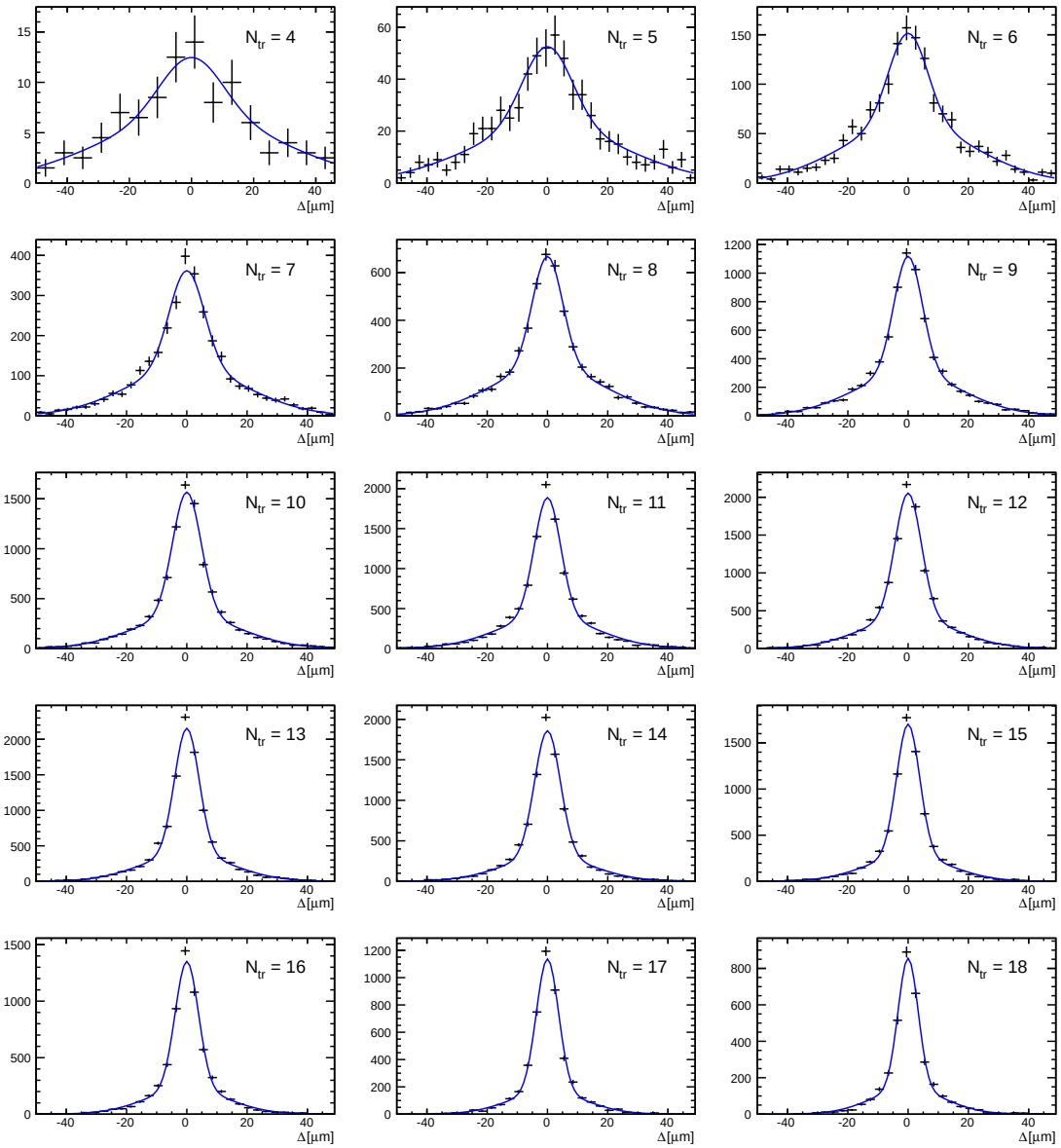

Figure 10: A difference between generated and reconstructed primary vertex z position in bins of number of tracks used for the primary vertex fit. Lines correspond to double-Gaussian fits.

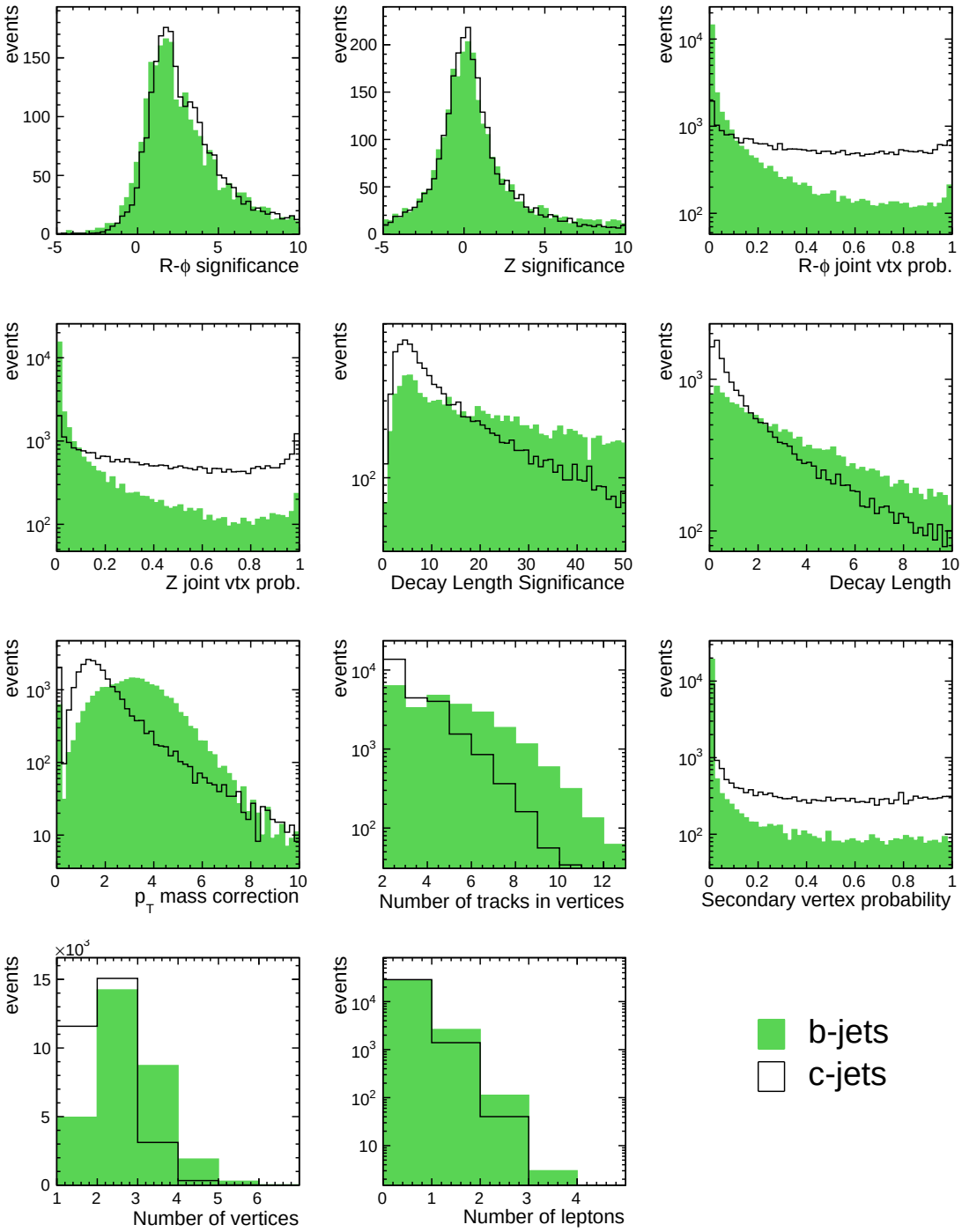

Figure 11: LCFIVertex flavour tagging inputs used in the flavour tagging neural nets. The plot is based on fully simulated and reconstructed $h\nu_e\bar{\nu}_e$ samples with 60 $\gamma\gamma \to$ hadrons events overlaid. The events are clustered into two jets using the k_t algorithm implemented in the FASTJET package with parameter $R = 0.7$.