Flavour Tagging at CLIC

Tomas Lastovicka[1] Katja Seidel[2,3] Jan Strube[4]

1- Institute of Physics, Academy of Sciences of the Czech Republic
Prague - Czech Republic

2- Max-Planck-Institut für Physik
Munich - Germany

3- Excellence Cluster 'Universe', TU München
Garching - Germany

4- CERN
Meyrin - Switzerland

We present the performance of the LCFI flavour tagging package in a realistic CLIC environment. The application is demonstrated on the examples of the measurement of the cross section times branching ratio of light Higgs decays to b and c quarks at 3 TeV, a study of heavy Higgs decays at 3 TeV and of top pair production at 500 GeV. All studies are based on full detector simulation with a realistic account of the machine-induced background at CLIC.

1 Introduction

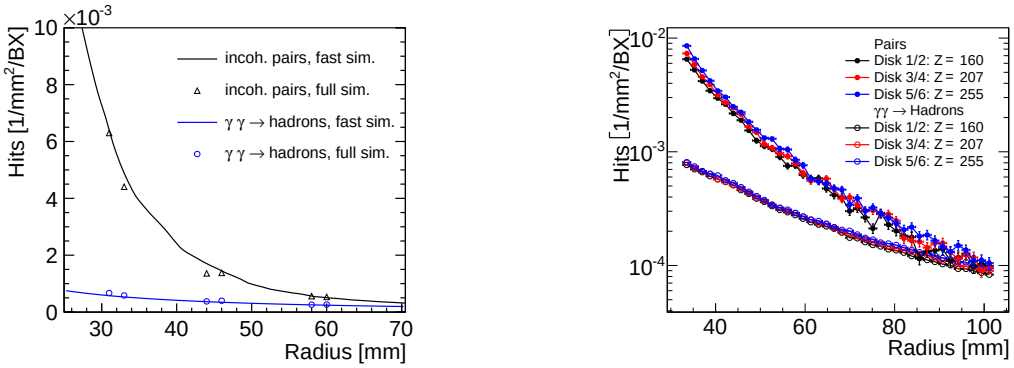
The CLIC accelerator provides unique opportunities for precision physics, due to the large production cross section of Higgs bosons at higher energies, and the increased energy reach compared to the ILC could prove to be an essential advantage in studying Supersymmetry. At the same time, the difficult machine environment bears its own challenges for precision physics. We demonstrate that the approach to flavour tagging that was developed for the ILC provides excellent results also at a CLIC machine at 500 GeV and at 3 TeV.

The LCFI flavour tagging package [1] has been developed for the ILC and is based on the ZVTOP algorithm invented at SLD. We will give here only a brief recap and focus on the differences to the ILC. After clustering the reconstructed event into jets, the ZVTOP algorithm finds secondary vertices from the tracks in a given jet. In contrast to LEP and ILC analyses, where the Durham jet clustering algorithm is employed, it is found that the CLIC environment requires the use of k_t-style algorithms developed for hadronic environments. The performance of the vertex fitting routines on primary vertices is demonstrated in the CLIC_SiD detector concept.

Information from the secondary vertices is used together with other jet-based variables as the input to neural networks for the flavour tagging. Accounts of the performance of the flavour tagging in the measurement of the cross section times branching ratio of light Higgs decays to b and c quarks at 3 TeV [6], a study of heavy Higgs decays at 3 TeV [2] and of top pair production at 500 GeV [7] are given in Section 5. All analyses use a full GEANT4 detector simulation and take into account realistic CLIC backgrounds.

2 Vertex Occupancies

The high fields of the CLIC beams produce several hadronic events in photon collisions per bunch crossing [3]. This background from $\gamma\gamma \to$ hadrons and incoherent electron-positron



(a) barrel

(b) forward disks

Figure 1: Average hit densities in the CLIC_ILD barrel (left) and forward (right) vertex detectors for particles originating from incoherent electron-positron pairs and from $\gamma\gamma \rightarrow$ hadrons. For the barrel region, the full simulation of the detector response is compared to a fast parametric tracking of the primary particles in the magnetic field. For the forward region, the results are shown for the full simulation only. Safety factors for the simulation uncertainties and cluster formation are not included.

pairs leads to high occupancies of the inner layers of the vertex detectors, potentially posing challenges to identifying secondary decay vertices close to the interaction point (IP). Figure 1 shows the occupancies from these two background sources in the barrel and forward regions of the vertex detector of the CLIC_ILD concept at 3 TeV. The position of the vertex detector layers is indicated by the x-coordinate of the black empty triangles or the blue circles. To reduce this background the detector concepts at CLIC make extensive use of time stamping, assuming 10 ns in the silicon detectors. While the occupancies in the inner double layer are several times higher than in the other layers, mainly due to background from incoherent electron-positron pairs, these hits have negligible impact on the track reconstruction efficiencies [4, 5].

3 Jet Clustering

The LCFI package uses the ZVTOP vertex finding algorithm to identify regions where multiple tracks originate from the decay of a long-lived secondary particle. The seed for such a vertex candidate is found by pairwise overlapping all tracks within a jet. The performance of finding vertices with only one track is greatly enhanced by using the jet axis. However, the beam-induced background from $\gamma\gamma \rightarrow$ hadrons is causing LEP-style clustering algorithms to fail. At a 3 TeV CLIC machine, k_t algorithms developed for hadron machines are more appropriate. Figure 2 shows the energy that is clustered in jets for a Durham-style algorithm used at LEP (ee_kt), and for a k_t algorithm as it is applied in the LHC experiments. The different colors show how the jet energy changes with the different timing cuts on the reconstructed particles. The effect of the $\gamma\gamma \rightarrow$ hadrons background is clearly visible. As this background is more dominant in the forward regions of the detector, this clearly has

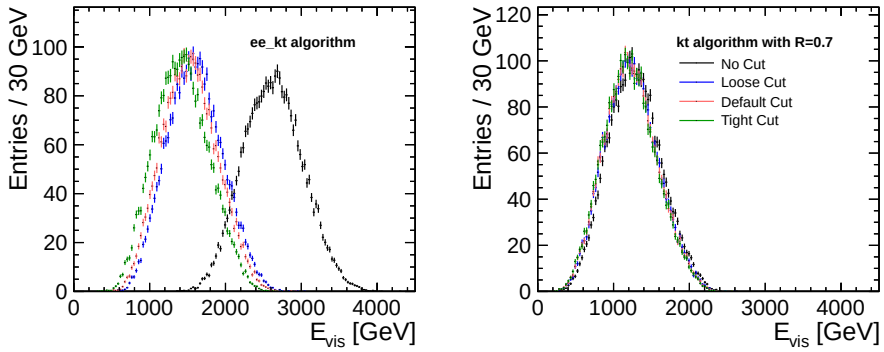


Figure 2: The visible energy in events with two jets and missing energy for different timing cuts (`tight`, `default` and `loose`) and without timing cuts on the reconstructed particles for the Durham (ee_kt) algorithm (left) and the k_t-algorithm (right).

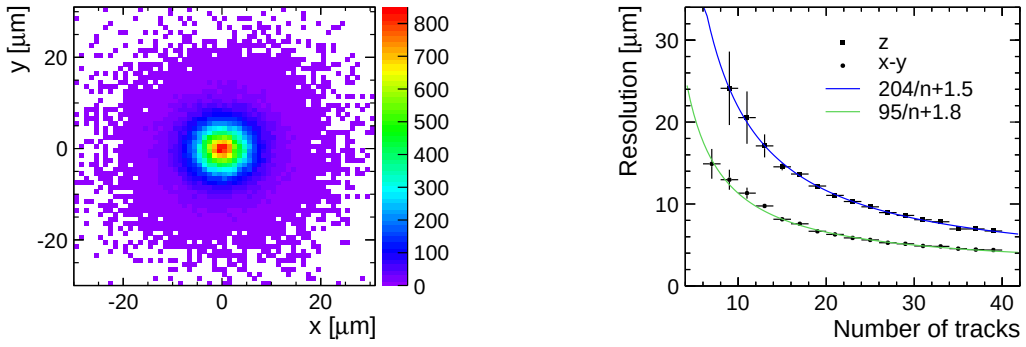


Figure 3: Location in the x-y plane of the primary vertices found in an event (left) and resolution of the primary vertices versus the number of tracks attached to the vertex (right).

an effect on the found jet axis. Events were clustered into two jets, and while the Durham algorithm clusters all particles into the jets, the k_t algorithm clusters particles in the far forward region into an additional beam-jet that contains most of the background. This reduces a potential bias on the jet axis, which is used in the jet flavour tagging.

4 Primary Vertex Resolution

The secondary vertex decay length is one of the most discriminating variables between b-, c- and light jets. We use the known primary vertex position at CLIC to judge the performance of the vertex reconstruction. Figure 3 (left) shows the position of primary vertices with more than 20 tracks in di-jet events in the x-y plane. The primary vertex in the simulation is always at the origin of the detector coordinate system. Figure 3 (right) shows the vertex resolution of the primary vertex in the x-y plane (solid points) and along the z axis (solid

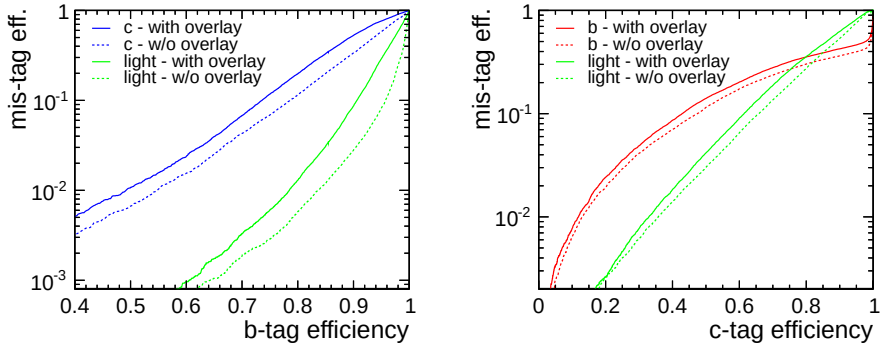


Figure 4: Purity versus efficiency curves for the LCFI flavour tagging networks in a mixed sample of $h\nu\nu$ events and $qq\nu\nu$ events. Performance for tagging $h \to b\bar{b}$ events (left). Performance for tagging $h \to c\bar{c}$ events (right).

squares) versus the number of tracks in the vertex. The distribution of the reconstructed vertex position was fitted with two Gaussian functions and the shown resolution is the weighted mean of the two fitted functions. The resolution of secondary vertices can be deduced from this figure, as it depends primarily on the number of tracks.

5 Flavour Tagging Performance on Physics Samples

5.1 Performance at 3 TeV

The performance of the LCFI flavour tagging package has been evaluated on the sample of light Higgs decays produced for the CLIC CDR. The flavour tagging uses three neural networks to distinguish light, c- and b-jets. The networks are trained to distinguish b-jets from c- and light jets, c-jets from b- and light jets and c-jets from b-jets, respectively. The nets were trained on samples with di-jets and missing energy. Figure 4 shows the mis-tag efficiency for b- and c-jets of the flavour tagging network trained on b- (left) and c-jets (right) versus the tagging efficiency, where the mis-tag efficiency is the fraction of incorrectly identified jets.

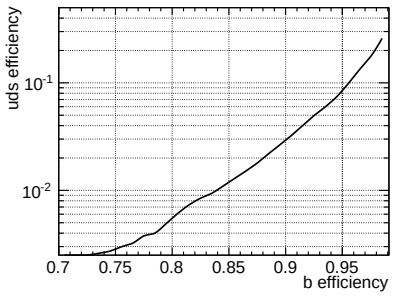


Figure 5: The performance in heavy Higgs events at a 3 TeV CLIC

Flavour tagging in events with heavy Higgs decays bears its own sets of challenges, due to the boost of the b hadrons, which results in 30 % of the tracks from secondary vertices originating beyond the innermost layer of the vertex detector, resulting in fewer hits. The default LCFI flavour tagging networks were augmented with track-based variables and particle identification to aid the reconstruction of B decays without a reconstructed secondary vertex. They were trained on events with the same kinematic properties as the signal events. The efficiency of tagging a light jet as a b-jet versus the efficiency of correctly identifying b-jets is shown in

Figure 5. In spite of the cited difficulties, the performance in these events is comparable to that in light Higgs events.

5.2 Performance at 500 GeV

In addition to the analysis of light Higgs decays to b and c quarks, and the analysis of heavy Higgs decays, the package was used in the study of the top pair production at 500 GeV. The purity versus the b-tag efficiency is shown in Figure 6. The purity is the fraction of correctly identified b-jets over all tagged b-jets, evaluated in a sample of signal decays and the Standard Model background after pre-selection cuts.The analysis was carried out in the CLIC_ILD detector concept, but due to the lower machine-induced background, the vertex detector was moved inward from its position in the detector for a 3 TeV machine. The flavour tagging networks were trained independently for this analysis. The analysis of $t\bar{t}$ events is challenging due to the high multiplicity of the final state, where the jet finding performance impacts on the ability to find secondary vertices. Additionally, the effect of the pile-up background from $\gamma\gamma \to$ hadrons events (Overlay) is shown for these events. The deterioration from this source is visible, but small.

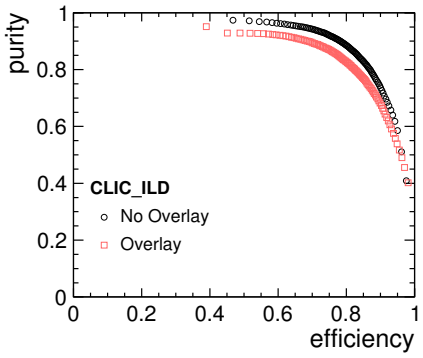


Figure 6: Flavour tagging performance in the top pair production analysis at a 500 GeV CLIC; the black circles show the performance without background from $\gamma\gamma \to$ hadrons events, while the red circles show the performance with this background.

6 Summary

We have given a brief overview over the challenges to flavour tagging at a CLIC machine and how they were addressed in the studies for the Physics and Detectors volume of the CLIC CDR. Especially the background from $\gamma\gamma \to$ hadrons events poses challenges to the event reconstruction. Jet clustering techniques developed for hadron machines are relatively insensitive to this background. The resulting effect on the flavour tagging performance in the presented analyses is small.

References

[1] 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. A*, A 610:573–589, 2009.

[2] M. Battaglia and P. Ferrari. A study of $e^+e^- \to H^0A^0 \to b\bar{b}b\bar{b}$ at 3 TeV at CLIC, 2010. CERN LCD-Note-2010-006.

[3] D. Dannheim and A. Sailer. Beam-induced backgrounds in the CLIC detectors, 2011. CERN LCD-Note-2011-021 .

[4] C. Grefe. Tracking performance in CLIC_SiD_CDR, 2011. CERN LCD-Note-2011-034.

[5] M. Killenberg and J. Nardulli. Tracking performance and momentum resolution of the CLIC_ILD detector for single muons, 2011. CERN LCD-Note-2011-013.

[6] T. Lastovicka. Light Higgs boson production and hadronic decays at 3 TeV, 2011. CERN LCD-Note-2011-036 .

[7] K. Seidel, S. Poss, and F. Simon. Top quark pair production at a 500 GeV CLIC collider, 2011. CERN LCD-2011-026.