PREPARED FOR SUBMISSION TO JINST

N[TH] WORKSHOP ON X
WHEN
WHERE

Reconstruction and classification of tau lepton decays with a future Linear Collider

B. Xu,[a,1] **S. Green,**[a] **J. S. Marshall,**[a] **M. A. Thomson**[a]

[a] *Cavendish Laboratory,*
JJ Thomson Avenue, Cambridge, CB3 0HE, UK

E-mail: xu@hep.phy.cam.ac.uk

ABSTRACT: Seven tau lepton decay final states, $e^- \bar{\nu}_e \nu_\tau$, $\pi^- \bar{\nu}_\mu \nu_\tau$, $\pi^- \nu_\tau$, $\rho\left(\pi^-\pi^0\right) \nu_\tau$, $a_1\left(\pi^-\pi^0\pi^0\right) \nu_\tau$, $a_1\left(\pi^-\pi^-\pi^+\right) \nu_\tau$ and $\pi^- \pi^- \pi^+ \nu_\tau$ were studied at the future $e^- e^+$ Linear Collider. The selection efficiency for each final states were compared for the centre of mass ($\sqrt{s}$) $e^- e^+$ collision energies of 100, 200, 500 and 1000 GeV and for the silicon-tungsten electromagnetic calorimeter (ECal) cell sizes from 3 to 20 mm. The difficulty of separating the final states lies in the reconstruction of multiple nearby photons. The overall hadronic decay selection efficiency is over 90% for c.o.m. $\sqrt{s}$= 100 GeV for the range of the ECal cell sizes, whilst the selection efficiency degrades significantly from 3 mm to 20 mm ECal cell size for $\sqrt{s}$= 500 and 1000 GeV.

KEYWORDS: Only keywords from JINST's keywords list please

ARXIV EPRINT: 1234.56789

[1] Corresponding author.

Contents

1 Introduction

Many experiments, including the Large Electron Positron Collider (LEP), has studied the tau lepton to a great details [1]. The total tau lepton hadronic decay width depends on the strong coupling constant. The branch ratio of tau decay tau hence provides a precision test of the Standard Model and models beyond the Standard Model. The spin state of the tau lepton could be inferred from the decay product and can be used to measure the CP(the product of charge conjugation and parity symmetries) of the Higgs with a Higgs decaying to a tau pair channel.

Final state separation of tau decay provides a good benchmark of the detector performance. The tau lepton has a very short life time and it will decay before reaching the calorimeter. The final states of the tau lepton decay mainly consist of charged particles and multiple photons. Separating different charged particle relies on the performance of the tracking system, whilst separating multiple nearby photons requires an excellent electromagnetic calorimeter (ECal) resolution.

A previous study with the International Large Detector (ILD) in the context of the International Linear Collider (ILC) was performed [2], where the impact of the varying the magnetic field and the size of the ECal were discussed. It was shown that about 95 % $\pi^- \nu_\tau$ and 90 % $\rho\left(\pi^-\pi^0\right) \nu_\tau$ and $a_1\left(\pi^-\pi^0\pi^0\right) \nu_\tau$ final states were correctly reconstructed.

The study presented in this paper was done using the CLIC_ILD detector concept with the PandoraPFA software package . The CLIC_ILD detector concept [3] is designed for the Compact LInear Collider (CLIC) based on the ILD detector [4], with a Time Projection Chamber, and a Silicon and Tungsten fine granularity ECal designed for the approach of the Particle flow calorimetry [5].

The difficulty of the τ decay mode separation is to correctly reconstruct photons in the final states. Two main features of the reconstruction software, PandoraPFA [6], help to separate the final states. Firstly, the iterative track cluster association algorithms connecting reconstructed tracks to the cluster showers in the calorimeters, providing a good identification of the charged particles and leave a cleaner environment for the neutral particles. Secondly, a transverse calorimeter shower profile based photon reconstruction algorithm carefully identifies and separates nearby photons, using a likelihood photon identification algorithm. Along side with other reconstruction algorithms

in the PandoraPFA software, charged particles and photons are well reconstructed and used as inputs for τ decay mode separation.

In this paper, we present a study for the separation of tau lepton decay final states, as a benchmark for the CLIC_ILD detector optimisation, by varying the size of the ECal cells and the centre of mass energy ($\sqrt{s}$) of the $e^- e^+ \rightarrow \tau^- \tau^+$ interaction.

2 Simulation and Reconstruction

To obtain a clean environment to separate the tau final state, we used the $e^- e^+ \rightarrow \tau^- \tau^+$ channel. The main mechanism is the pair production of the τ pair, via s channel.

Simulated Monte Carlo (MC) samples were generated with the generator software WHIZARD 1.95 [7]. PYTHIA 6.4 [8] is used for the hadronisation and is tuned to the LEP results []. The interface to TAUOLA [9] is used to describe the τ lepton decays. The initial state radiation (ISR) and the beam induced background were not simulated, but final state radiation (FSR) was simulated. This was because the study was perform as a benchmark study for the detector optimisation, and a clean environment is preferably to study the impact of the change of the design of the detector.

Around two millions events per ECal cell size and per c.o.m. energy were simulated before any generator level cuts. An event was considered if the event passes a set of cuts at generator level. The cuts are

- the final state photons not converting to electron pair in the tracking system,
- the tau leptons decaying in certain half polar angle region and
- the visible energy of the tau lepton decay more than 5 GeV.

The half polar angle acceptance is 0.3 to 0.6 rad and 0.8 to 1.57 rad which cover the barrel and the end cap region excluding the barrel-end cap transitional region. The visible energy of the tau lepton decay is defined as the energy of the tau minus the energy of the tau neutrino. Again the cuts were chosen to obtain a clean environment of the interaction to study the effect of the different detector models on the τ final state separation.

Events were simulated with software MOKKA [10] with the CLIC_ILD detector geometry description, based on the GEANT 4 package [11]. Events were reconstructed with ilcsoft version v01-17-07 [12] and PandoraPFA version v02-02-00 [6], where the photon reconstruction is described in [13].

The events were simulated at $\sqrt{s}$= 100, 200, 500 and 1000 GeV, with different ECal square cell sizes of 3, 5, 7, 10, 15 and 20 mm.

3 Analysis strategy

Seven decay final states of the tau lepton shown in table 1 were studied, which cover 92.58 % of all tau decays. The decay modes not covered have branching fractions lower than 1% each. These final states can be classified into three categories: leptonic decays, one-prong with photons and three-prong with photons.

Table 1. Branching fractions of the seven τ^- decays in this study, taken from [14]. τ^+ decays similarly to τ^-.

Decay mode	Branching fraction / %
$e^- \bar{\nu}_e \nu_\tau$	17.83±0.04
$\pi^- \bar{\nu}_\mu \nu_\tau$	17.41±0.04
$\pi^- \nu_\tau$	10.83±0.06
$\rho\left(\pi^-\pi^0\right) \nu_\tau$	25.52±0.09
$a_1\left(\pi^-\pi^0\pi^0\right) \nu_\tau$	9.30±0.11
$a_1\left(\pi^-\pi^-\pi^+\right) \nu_\tau$	8.99±0.06
$\pi^- \pi^- \pi^+ \nu_\tau$	2.70±0.08

Separating charged particles is mainly testing the performance of the tracking system as well as the ECal design, which provides information of association of the charged tracks to the clusters in the ECal. However, the difficulty of separating the hadronic final states mostly comes from correctly separating nearby photons. An excellent ECal spatial resolution is required for reconstructing multiple nearby photons.

The analysis strategy is outlined in the following. The detector space is divided into two halves using the thrust axis. Thrust is defined as $T = max_{\hat{n}} \frac{\sum_i |p_i.\hat{n}|}{\sum_i |p_i|}$, where p_i is the momentum three-vector of a Particle Flow Object (PFO), $\hat{n}$ is the thrust axis, a unit 3-vector that maximise the thrust, T. PFOs were then separated into two sets based on the sign of the dot product between the momentum three-vector and the thrust axis three-vector.

A set of discriminative variables were calculated for multivariate analysis.

The first of variables separate final states into three broad categories: leptonic decays, one-prong with photons and three-prong with photons. The variables are the number of reconstructed PFOs of $\mu^\pm$, $e^\pm$, γ, $\pi^\pm$ and charged PFOs. As shown in figure 1, there is a clear distinction between different final states. However, there is still a lot of confusion between final states with multiple photons.

The next set of varibles are designed to separate leptonic final states from the hadronic final states. Although leptonic final states have very different topologies to the hadronic state, i.e. $\mu^\pm$ deposits most energy in the muon chamber and some energy in the ECal and $e^\pm$ deposited electromagnetic shower profile in the ECal, the difficulty is to correctly separate $e^\pm$ from $\pi^\pm$, where $\pi^\pm$ could start showering early in the ECal and be similar to a electromagnetic shower.

Varibles of interests are $\frac{\sum_i E_{i,ECal}}{\sum_i E_{i,tot}}$, $\frac{\sum_c E_{c,ECal}}{\sum_c E_{c,tot}}$, $\langle E_{calo} \rangle$, $\langle d_T \rangle$, $\langle Layer_{L,start} \rangle$, $\langle \Delta Profile_L \rangle$, $\frac{N_{MIP}}{N_{calo}}$, $\langle \frac{E_c}{P_c} \rangle$ and $\frac{E_\mu}{E_\tau}$, where E_{ECal} is the energy deposited in the ECal, E_{tot} is the total energy deposited in the calorimeter, i is summing over all PFOs, c is summing over charged PFOs, $\langle E_{calo} \rangle$ is the the average energy of a calorimeter hit, $\langle d_T \rangle$ is the average transverse width of a cluster shower, $\langle Layer_{L,start} \rangle$ is the average longitudinal start layer of a cluster shower, $\langle \Delta Profile_L \rangle$ is the average discrepancy of a cluster longitudinal shower profile to an electromagnetic shower profile, $\frac{N_{MIP}}{N_{calo}}$ is the fraction of calorimeter hits profiled as minimum ionising particles, $\langle \frac{E_c}{P_c} \rangle$ is the average ratio of the energy and the momentum of charged particles. E_μ is the energy of the reconstructed $\mu^\pm$.

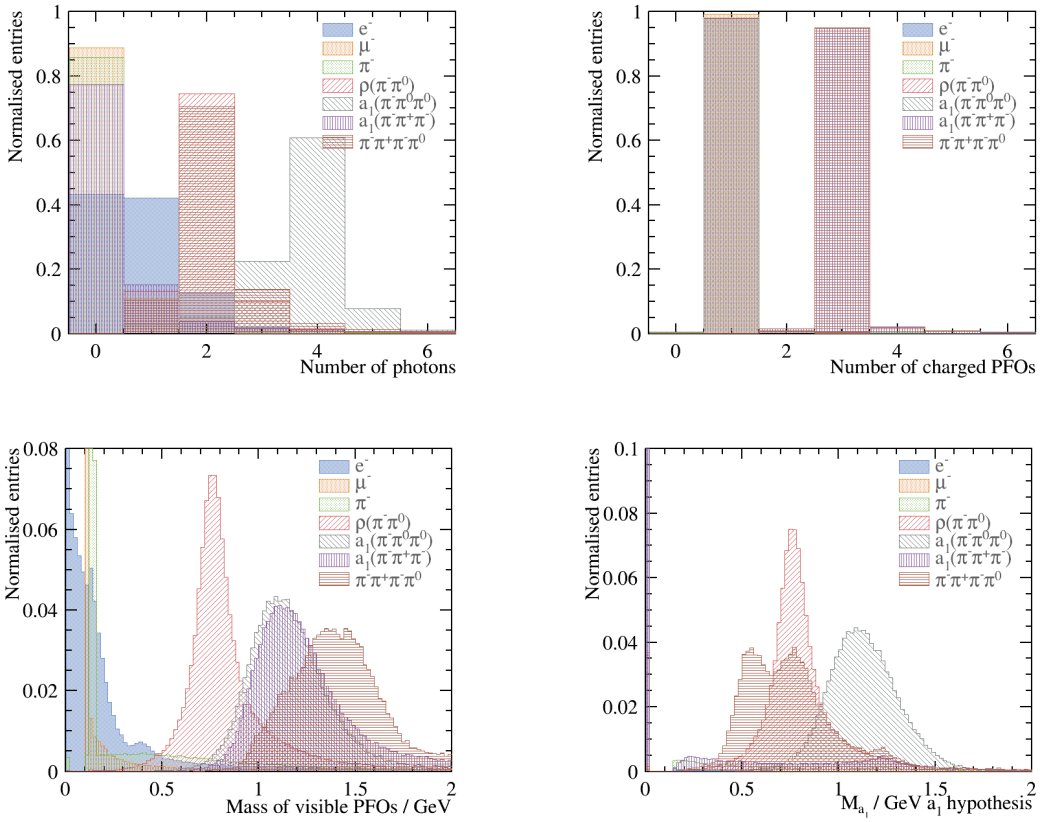

Figure 1. The example normalised distribution for discriminative variables for seven final states, $e^-\ \overline{\nu}_e\ \nu_\tau$, $\pi^-\ \overline{\nu}_\mu\ \nu_\tau$, $\pi^-\ \nu_\tau$, $\rho\left(\pi^-\pi^0\right)\ \nu_\tau$, $a_1\left(\pi^-\pi^0\pi^0\right)\ \nu_\tau$, $a_1\left(\pi^-\pi^-\pi^+\right)\ \nu_\tau$ and $\pi^-\ \pi^-\ \pi^+\ \nu_\tau$, separated with truth information, with $\sqrt{s}$= 100 GeV for nominal CLIC_ILD detector model. The top left and top right, bottom left and bottom right plots are the normalised entries against the number of photons, number of charged PFOs, invariant mass of visible PFOs and the invariant mass of $a_1\left(\pi^-\pi^0\pi^0\right)$ for $a_1\left(\pi^-\pi^0\pi^0\right)$ hypothesis, respectively. There is a clear distinction between different final states in each plot.

For the hadronic decay final states, variables are M_{PFOs}, M_γ, $M_{\pi^\pm}$, M_c, M_n, $\frac{E_\gamma}{E_\tau}$, $\frac{E_{\pi^\pm}}{E_\tau}$ and $\frac{E_c}{E_\tau}$, where M_{PFOs}, M_γ, $M_{\pi^\pm}$, M_c, M_n are the invariant masses of all PFOs, photons, charged pions, charged PFOs and neutral PFOs respectively, and E_γ, $E_{\pi^\pm}$, E_c, E_τ are the energy of the γ, $\pi^\pm$, charged particle and τ respectively. For final states with ρ or a resonance, the hypothesis tests are used. For example, for the $a_1\left(\pi^-\pi^0\pi^0\right)\ \nu_\tau$ final state, the χ^2_a to minimise is,

$$\chi^2_a = \left(\frac{m_{a,fit} - m_a}{\sigma_a}\right)^2 + \left(\frac{m_{\pi^0,fit} - m_{\pi^0}}{\sigma_{\pi^0}}\right)^2 + \left(\frac{m_{\pi^{0*},fit} - m_{\pi^0}}{\sigma_{\pi^0}}\right)^2, \tag{3.1}$$

where $m_{\pi^0,fit}$ and $m_{\pi^{0*},fit}$ are the invariant masses of all possible two photons combinations, σ_a and σ_{π^0} are the half width of the invariant mass distribution of reconstructed a and π^0 using the truth information, and m_a and m_π are the masses of a and π^0, taken from [14]. If there are two or

three photons, the χ^2_{a} expression will be reduced and not including $m_{\pi^{0^*},fit}$ term. If there are fewer than two photons, the χ^2_{a} expression would only contain $m_{\mathrm{a},fit}$ term.

For the $\rho\left(\pi^-\pi^0\right)\,\nu_\tau$ final state, a similar χ^2_ρ test for ρ hypothesis is used to extract $m_{\rho,fit}$ and $m_{\pi^0,fit}$ variables. χ^2_ρ is similar to χ^2_{a} with ρ replacing a and only one $m_{\pi^0,fit}$ term.

Figure 1 shows the $m_{\mathrm{a},fit}$ where $\rho\left(\pi^-\pi^0\right)\,\nu_\tau$, $\mathrm{a}_1\left(\pi^-\pi^0\pi^0\right)\,\nu_\tau$ and $\pi^-\,\pi^-\,\pi^+\,\nu_\tau$ final states contribute to the a resonance, although only $\mathrm{a}_1\left(\pi^-\pi^0\pi^0\right)\,\nu_\tau$ final has a real a resonance. This is due to the structure of the χ^2_{a} minimisation function allowing final states with more than two photons and one $\pi^\pm$ to contribute.

Energy of the τ is assume to be the same as the energy of $\mathrm{e}^\pm$ beam, which is half of the $\sqrt{s}$energy. Recoil momenta were calculated assuming the $\mathrm{e}^-\,\mathrm{e}^+$ collision happened at the centre of mass energy. Both assumptions are largely valid when there is no ISR contribution.

For the multivariate analysis, the multiclass class of the TMVA package [15] was used to train the seven final states simultaneously. The multiclass class is an extension of the standard signal-background classifier. For each final state, the multiclass classifier will train the final state as the signal against all other final states as the background. This process is repeated for each final state. The classifier output for a single event is a normalised number for each final state, where the sum is one. The number of a final state of a event can be used as the probability. The event is classified into a particular final state if the final state has the highest classifier output number. The advantage of using the multiclass is that the correction between different final states are accounted for and the classifier output are correctly adjusted for multiple final states, hence one event can only be classified into one final state.

Half of the randomly selected samples were used in the training process and the other half were used for testing.

The TMVA multiclass classifier used is boosted decision tree with gradient boosting (BDTG), as it was found to give for the best performance. The MVA classifier is trained and optimised to give the best overall separation across all final states.

4 Results and discussion

The reconstruction efficiencies for the seven final state of the tau decaying with c.o.m. energy of 100 GeV for the nominal CLIC_ILD detector are shown in the table 2.

The study was repeated with $\sqrt{s}$= 100, 200, 500, 1000 GeV. The ECal square cell sizes were also varied at 3, 5, 7, 10, 15 and 20 mm, whilst keeping the the total ECal size the same. The results table were are in the appendix X.

To compare the impact of the ECAL cell sizes and the $\sqrt{s}$energies on the separation of tau final states, the selection efficiencies were plotted in the figure 2. The leptonic decay selection efficiencies are not shown as they are similar across different ECal cell sizes. This is because the $\mathrm{e}^\pm$ and $\mu^\pm$ identifications mostly rely on the tracking system, which was not varied in this study. The energy deposited in the calorimeter are used for the association to the tracks but it has a small impact on the lepton identification.

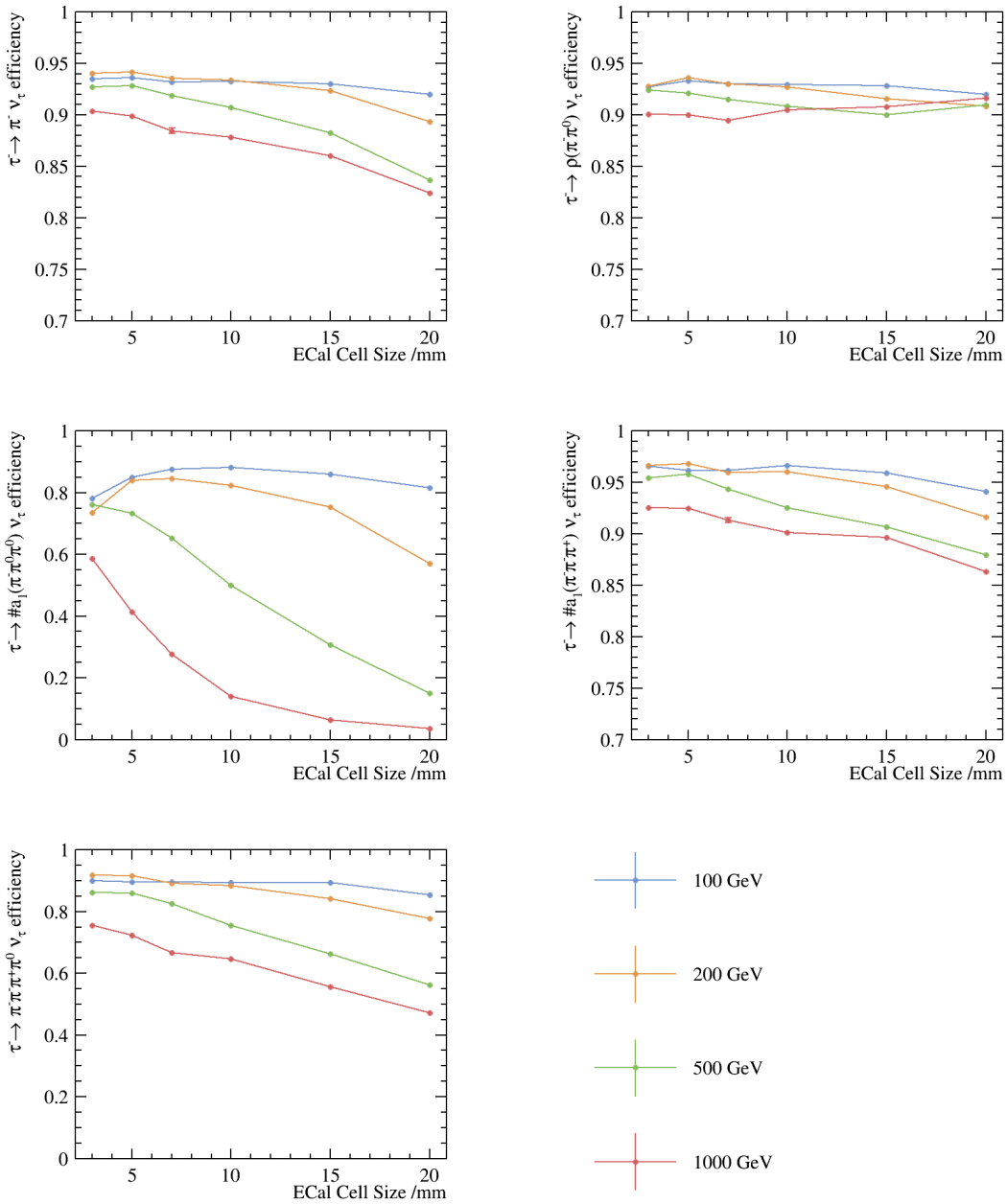

Figure 2. The selection efficiencies for various final states against the ECal cell size for different c.o.m. energies with the nominal CLIC_ILD detector model are shown. The top left, top right, middle left, middle right and bottom left plots are for the $\pi^-\ \nu_\tau$, $\rho\left(\pi^-\pi^0\right)\ \nu_\tau$, $a_1\left(\pi^-\pi^0\pi^0\right)\ \nu_\tau$, $a_1\left(\pi^-\pi^-\pi^+\right)\ \nu_\tau$ and $\pi^-\ \pi^-\ \pi^+\ \nu_\tau$ final states respectively. From the top to the bottom, blue, orange, green and red lines are representing the $\sqrt{s}$= 100, 200, 500 and 1000 GeV respectively. Note that the y axis are not the same for displaying purpose.

Table 2. The probability of reconstruction of true decay modes in columns in percent, with $\sqrt{s}$= 100 GeV for nominal CLIC_ILD detector model. Bold numbers show the correctly reconstructed terms. Numbers less than 0.25% are not shown. Statistical uncertainties are less than 0.25%. Final states include ν_τ, which is not shown.

Reco ↓ True →	$e^- \bar{\nu}_e$	$\pi^- \bar{\nu}_\mu$	π^-	$\rho\left(\pi^-\pi^0\right)$	$a_1\left(\pi^-\pi^0\pi^0\right)$	$a_1\left(\pi^-\pi^-\pi^+\right)$	$\pi^-\,\pi^-\,\pi^+$
$e^- \bar{\nu}_e$	99.8	-	0.9	1.1	0.8	-	-
$\pi^- \bar{\nu}_\mu$	-	99.5	0.5	-	-	-	-
π^-	-	0.3	93.2	0.9	-	0.4	-
$\rho\left(\pi^-\pi^0\right)$	-	-	4.1	93.0	10.5	0.6	2.8
$a_1\left(\pi^-\pi^0\pi^0\right)$	-	-	-	4.3	88.2	-	1.0
$a_1\left(\pi^-\pi^-\pi^+\right)$	-	-	1.0	0.3	-	96.6	6.9
$\pi^-\,\pi^-\,\pi^+$	-	-	-	0.4	0.4	2.4	89.3

Overall, the hadronic decay selection efficiency decreases as the $\sqrt{s}$energy increases. This is due to the fact that when τs are boosted at higher $\sqrt{s}$energies, the separation between decay products is smaller. Hence it is more difficult to reconstruct multi-photon final states correctly.

As the ECal cell sizes increase, the reconstruction efficiencies generally decrease. Larger cell sizes have lower spatial resolutions, making the separating of nearby photons more difficult.

For the $a_1\left(\pi^-\pi^0\pi^0\right)$ ν_τ final state, the selection efficiency for 500 GeV rises from ECal cell sizes 15 mm to 20 mm and the one for 1000 GeV rises from 7 , to 20 mm actually goes up as cell size increases. This is because when the algorithm can not reconstruct four photons in the $a_1\left(\pi^-\pi^0\pi^0\right)$ ν_τ final state, and the event topology would be very similar to the $\rho\left(\pi^-\pi^0\right)$ ν_τ final states.

For the $\sqrt{s}$= 100 and 200 GeV, the selection efficiency of the 5 mm ECal cell size is better than that of the 3 mm. One possible explanation is that the and the PandoraPFA have been optimised for the nominal ILD detector with the 5 mm ECal cell size, which shares the same ECal structure with the nominal CLIC_ILD detector.

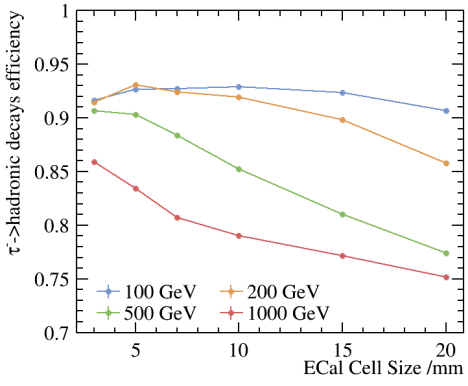

Figure 3. The τ hadronic decay selection efficiency against the ECal cell size for different $\sqrt{s}$energies with the nominal CLIC_ILD detector model are shown. The blue, orange, green and red lines are representing the $\sqrt{s}$= 100, 200, 500 and 1000 GeV respectively.

In order to compare the overall separation power of all the final states across c.o.m. energy and the ECal cell sizes, we constructed a single parameter function, the τ hadronic decay final state efficiency function,

$$\varepsilon_{had} = \frac{\left(\Sigma_i Br_i \varepsilon_i\right)}{\Sigma_i Br_i}, \tag{4.1}$$

where Br_i is the branching fraction of a hadronic final state after the generator level cut, ε_i is the selection efficiency of the final state and the i is summing over five hadronic decay final state of τ. Leptonic decays, $e^- \bar{\nu}_e \nu_\tau$ and $\pi^- \bar{\nu}_\mu \nu_\tau$, were not included as the variation of the leptonic decay selection efficiency is small.

In the figure 3, τ hadronic decay final state efficiency, ε_{had}, against the ECal cell size with different $\sqrt{s}$is shown. ε_{had} decreases when cell sizes increases and when $\sqrt{s}$increases. Again, ε_{had} of the 5 mm ECal cell size is better than that of the 3 mm for 100 and 200 GeV lines possibly due the optimisation of the software fro the nominal ILD 5 mm cell size.

The ε_{had} is above 90% for the ECal cell size from 3 to 20 mm for the $\sqrt{s}$= 100 GeV. For $\sqrt{s}$= 200 GeV, the ε_{had} decreases from over 90% to 86% for the ECal cell size from 3 to 20 mm. The degradation of the ε_{had} is significant for the 500 and 1000 GeV lines, where the ε_{had} drops from over 90% to 77% and from 86% to 75% respectively.

For low $\sqrt{s}$, namely 100 and 200 GeV, up to 15 mm cell sizes of ECal will give a good performance for τ hadronic decay modes separation, and the ε_{had} is above 90%. For the high c$\sqrt{s}$, namely 500 and 1000 GeV, it is preferential to have a small ECal cell size for τ hadronic decay modes separation. There is about 15% degradation of ε_{had} for ECal cell size from 3 to 20 mm.

The paper illustrated the usage of reconstruction of the tau decay modes as a benchmark for the detector optimisation.

Acknowledgments

The authors would like to thank P. G. Roloff for helping to generate the simulated samples.

References

[1] ALEPH, S. Schael *et al.*, Phys. Rept. **421**, 191 (2005), hep-ex/0506072.

[2] T. H. Tran, V. Balagura, V. Boudry, J.-C. Brient, and H. Videau, The European Physical Journal C **76**, 468 (2016).

[3] L. Linssen, A. Miyamoto, M. Stanitzki, and H. Weerts, (2012), 1202.5940.

[4] ILD Concept Group - Linear Collider Collaboration, T. Abe *et al.*, (2010), 1006.3396.

[5] J. S. Marshall, A. MÃijnnich, and M. A. Thomson, Nucl. Instrum. Meth. **A700**, 153 (2013), 1209.4039.

[6] J. S. Marshall and M. A. Thomson, Eur. Phys. J. **C75**, 439 (2015), 1506.05348.

[7] W. Kilian, T. Ohl, and J. Reuter, European Physical Journal C **71** (2011).

[8] T. Sjostrand, (1995), hep-ph/9508391.

[9] S. Jadach, Z. Was, R. Decker, and J. H. Kuhn, Comput. Phys. Commun. **76**, 361 (1993).

[10] P. Mora de Freitas and H. Videau, p. 623 (2002).

[11] GEANT4, S. Agostinelli *et al.*, Nucl.Instrum.Meth. **A506**, 250 (2003).

[12] F. Gaede and J. Engels, EUDET Report (2007).

[13] B. Xu, Improvement of photon reconstruction in PandoraPFA, in *Proceedings, International Workshop on Future Linear Colliders (LCWS15): Whistler, B.C., Canada, November 02-06, 2015*, 2016, 1603.00013.

[14] Particle Data Group, K. A. Olive *et al.*, Chin. Phys. **C38**, 090001 (2014).

[15] TMVA Core Developer Team, J. Therhaag, AIP Conf.Proc. **1504**, 1013 (2009).