Prepared for submission to JINST

N[TH] Workshop on X
when
where

Reconstruction and classification of tau lepton decays with future e^- e^+ linear collider

B. Xu,[a,1] John? Mark? Steve? Who to put?[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, $e^- \overline{\nu}_e \nu$, $\mu^- \overline{\nu}_\mu \nu_\tau$, $\pi^- \nu_\tau$, $\pi^- 2\gamma \nu_\tau$, $\pi^- 4\gamma \nu_\tau$, $\pi^+ 2\pi^- \nu_\tau$ and $\pi^+ 2\pi^- 2\gamma \nu_\tau$ were studied at the future e^- e^+ Compact LInear Collider. The selection efficiency for each final states were compared for the centre of mass (c.o.m.) 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 overall hadronic decay selection efficiency is over 90% for c.o.m. collision energy of 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 c.o.m. collision energy of 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], as the decay product of the tau provides a precision test of the Standard Model and a measurement of the spin state of the tau lepton. The spin state 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. As many of the final states of the tau decay consist of boosted charged particles with different numbers of photons and the ECal provides important calorimetric information for correctly reconstructing and separating nearby photons, this makes tau lepton decay final state separation suitable for the electromagnetic calorimeter (ECal) optimisation.

A previous study with the International Large Detector (ILD) in the context of the International Linear Collider (ILC) was performed [2]. Photons were reconstructed with GARLIC software package [3, 4] 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 % $\rho^-\ \nu_\tau$ and $a_1(1260)\ \nu_\tau$ 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 [5] is designed for the Compact LInear Collider (CLIC) based on the ILD detector [6], with a Time Projection Chamber, and a Silicon and Tungsten fine granularity ECal designed for the approach of the Particle flow calorimetry [7].

In this paper, we present a study for the separation of tau lepton decay final states, as a benchmark for CLIC_ILD detector optimisation, by varying the size of the ECal cells and the centre of mass (c.o.m.) energy 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, shown in the figure 1.

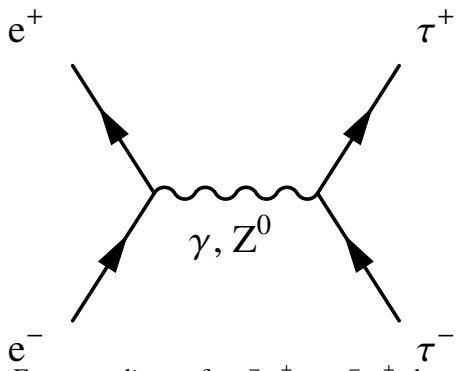

Figure 1. The Feynman digram for $e^- e^+ \rightarrow \tau^- \tau^+$ channel via γ or Z^0.

Simulated Monte Carlo (MC) samples were generated with the generator software WHIZARD 1.95 [8]. PYTHIA 6.4 [9] is used for the hadronisation and is tuned to the LEP results []. The interface to TAUOLA [10] is used to describe the τ lepton decays. The initial state radiation (ISR) and the beam induced background were not simulated in samples used in this study, but final state radiation (FSR) was simulated.

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 designed to obtain a clean environment. 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.

Events were simulated with software MOKKA [11] with the CLIC_ILD detector geometry description, based on the GEANT 4 package [12]. Events were reconstructed with ilcsoft version v01-17-07 [13] and PandoraPFA version v02-02-00. Note that this version of PandoraPFA comes with an improved photon reconstruction [14].

The c.o.m. energy of the $e^- e^+ \rightarrow \tau^- \tau^+$ channel were simulated at 100, 200, 500 and 1000 GeV. The same event were simulated with different ECal cell sizes of 3×3, 5×5, 7×7, 10×10, 15×15 and 20×20 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. These final states can be classified into three categories: leptonic decays, one-prong with photons and three-prong with photons. The difficulty of separating the final states mainly comes from the correctly separating final states within each category, in particular the separation of nearby photons as a boosted neutral pion decays to two boosted spatially close photons.

The analysis strategy is outlined in the following. First 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

Table 1. Branching fractions of the seven τ^- decays in this study, taken from [15]. τ^+ decays similarly to τ^-.

Decay Chain	Final Product	Branching fraction / %
$\tau^- \rightarrow$	$e^- \overline{\nu}_e \nu_\tau$	17.83±0.04
$\tau^- \rightarrow$	$\mu^- \overline{\nu}_\mu \nu_\tau$	17.41±0.04
$\tau^- \rightarrow$	$\pi^- \nu_\tau$	10.83±0.06
$\tau^- \rightarrow \rho^- \nu_\tau \rightarrow \pi^- \pi^0 \nu_\tau \rightarrow$	$\pi^- 2\gamma \nu_\tau$	25.52±0.09
$\tau^- \rightarrow a_1(1260) \nu_\tau \rightarrow \pi^- 2\pi^0 \nu_\tau \rightarrow$	$\pi^- 4\gamma \nu_\tau$	9.30±0.11
$\tau^- \rightarrow a_1(1260) \nu_\tau \rightarrow$	$\pi^+ 2\pi^- \nu_\tau$	8.99±0.06
$\tau^- \rightarrow \pi^+ 2\pi^- \pi^0 \nu_\tau \rightarrow$	$\pi^+ 2\pi^- 2\gamma \nu_\tau$	2.70±0.08

three-vector of a Particle Flow Object (PFO), $\hat{n}$ is the thrust axis, a unit 3-vector that maximise the thrust, T. The thrust was calculated with the Thrust Marlin Processor. 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 multi variant analysis, listed below,

- the ratio of the total energy deposited in the ECal and the total energy of all charged PFOs,
- the ratio of the total energy deposited in the ECal and the total energy of all PFOs,
- the invariant masses of all PFOs, photons, charged pions, charged PFOs and neutral PFOs,
- the ratios of the tau lepton energy and the total energy of the tau decay product, where decay products are muons, photons, charged pions and charged PFOs,
- the number of PFOs of muons, electrons, photons, charged pions and charged PFOs,
- the invariant masses of ρ^- and π^0 for ρ^- hypothesis test,
- the invariant masses of $a_1(1260)$ and two π^0 for π^0 hypothesis test,
- the average energy of a calorimeter hit cell,
- the average transverse width of a cluster shower,
- the average longitudinal start layer of a cluster shower,
- the average discrepancy of a cluster longitudinal shower to an electromagnetic shower profile,
- the fraction of calorimeter hits profiled as minimum ionising particles,
- the average ratio of the energy and the momentum of charged particles.

The particle identifications and all quantities were computed by the PandoraPFA.

The ρ hypothesis test is to find the best ρ decay candidates by minimising chi squared according to $\chi^2_\rho = (m_{\rho,fit} - m_\rho/\sigma_\rho)^2 + (m_{\gamma_1\gamma_2} - m_{\pi^0}/\sigma_{\pi^0})^2$, where $m_{\gamma_1\gamma_2}$ is the invariant mass from all

possible two photons combinations, $m_{\rho,fit}$ is the invariant mass of $m_{\gamma_1\gamma_2}$ with all possible $\pi^{\pm}$ combinations, σ_{ρ} and σ_{π^0} are the half width of the invariant mass distribution of reconstructed ρ and π^0 using the truth information. The χ^2 expression will reduce to $\chi^2_{\rho} = (m_{\rho,fit} - m_{\rho}/\sigma_{\rho})^2$ if there is only one photon. In which case two photons from a π^0 are assumed to be reconstructed as one photon.

Similarly the $a_1(1260)$ hypothesis test is to find the best $a_1(1260)$ decay candidates by minimising chi squared according to $\chi^2_{a_1(1260)} = (m_{a_1(1260),fit} - m_{a_1(1260)}/\sigma_{a_1(1260)})^2 + (m_{\gamma 1\gamma 2} - m_{\pi^0}/\sigma_{\pi^0})^2 + (m_{\gamma 3\gamma 4} - m_{\pi^0}/\sigma_{\pi^0})^2$, where $m_{\gamma_1\gamma_2}$ and $m_{\gamma_3\gamma_4}$ are the invariant masses from all possible two photons combinations, $m_{a_1(1260),fit}$ is the invariant mass of $m_{\gamma_1\gamma_2}$ nd $m_{\gamma_3\gamma_4}$ with all possible $\pi^{\pm}$ combinations, $\sigma_{a_1(1260)}$ and σ_{π^0} are the half width of the invariant mass distribution of reconstructed $a_1(1260)$ and π^0 using the truth information. If there are two or three photons, the χ^2 expression will reduce to $\chi^2_{a_1(1260)} = (m_{a_1(1260),fit} - m_{a_1(1260)}/\sigma_{a_1(1260)})^2 + (m_{\gamma 1\gamma 2} - m_{\pi^0}/\sigma_{\pi^0})^2$ assuming that two photons from one π^0 are reconstructed as one photon. If there is one photon, the χ^2 expression will reduce to $\chi^2_{a_1(1260)} = (m_{a_1(1260),fit} - m_{a_1(1260)}/\sigma_{a_1(1260)})^2$ assuming that four photons from two π^0 are reconstructed as one photon.

The last six listed variables are specialised in separating a e^- from a π^0. The e^- will deposit an distinctive electromagnetic shower profile in the ECal, which will be different than that of a π^0.

Recoil momenta were calculated assuming the e^- e^+ collision happened at the centre of mass energy, which is largely valid with no ISR contribution.

For the multivariate analysis, the multiclass class of the TMVA package [16] 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 signal against all other final states as 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 adds to 1. 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 was 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 c.o.m. energy of 100, 200, 500, 1000 GeV. The ECal cell sizes were also varied at 3 x 3, 5 x 5, 7 x7, 10 x 10, 15 x 15 and 20 x 20 mm, whilst keeping the the total

Table 2. The probability of reconstruction of true decay modes in columns in percent, with c.o.m. energy of 100 GeV for nominal CLIC_ILD detector model. Statistical uncertainties are shown.

Reco ↓ True →	e^-	μ^-	π^-	$\pi^- 2\gamma$	$\pi^- 4\gamma$	$\pi^+ 2\pi^-$	$\pi^+ 2\pi^- 2\gamma$
e^-	$99.76_{\pm 0.02}$	$0.02_{\pm 0.01}$	$0.86_{\pm 0.04}$	$1.08_{\pm 0.03}$	$0.76_{\pm 0.05}$	$0.03_{\pm 0.01}$	$0.01_{\pm 0.01}$
μ^-	$0.01_{\pm 0.00}$	$99.51_{\pm 0.03}$	$0.50_{\pm 0.03}$	$0.10_{\pm 0.01}$	$0.01_{\pm 0.01}$	$0.01_{\pm 0.01}$	$0.00_{\pm 0.00}$
π^-	$0.08_{\pm 0.01}$	$0.33_{\pm 0.02}$	$93.24_{\pm 0.12}$	$0.86_{\pm 0.03}$	$0.05_{\pm 0.01}$	$0.36_{\pm 0.03}$	$0.04_{\pm 0.02}$
$\pi^- 2\gamma$	$0.13_{\pm 0.01}$	$0.12_{\pm 0.01}$	$4.06_{\pm 0.09}$	$92.97_{\pm 0.08}$	$10.45_{\pm 0.19}$	$0.56_{\pm 0.04}$	$2.75_{\pm 0.13}$
$\pi^- 4\gamma$	$0.02_{\pm 0.01}$	$0.01_{\pm 0.00}$	$0.09_{\pm 0.01}$	$4.29_{\pm 0.07}$	$88.21_{\pm 0.20}$	$0.03_{\pm 0.01}$	$1.00_{\pm 0.08}$
$\pi^+ 2\pi^-$	$0.01_{\pm 0.00}$	$0.02_{\pm 0.01}$	$1.03_{\pm 0.05}$	$0.27_{\pm 0.02}$	$0.14_{\pm 0.02}$	$96.62_{\pm 0.09}$	$6.86_{\pm 0.20}$
$\pi^+ 2\pi^- 2\gamma$	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.23_{\pm 0.02}$	$0.44_{\pm 0.02}$	$0.39_{\pm 0.04}$	$2.38_{\pm 0.07}$	$89.34_{\pm 0.25}$

ECal size the same. The results table were are in the appendix X.

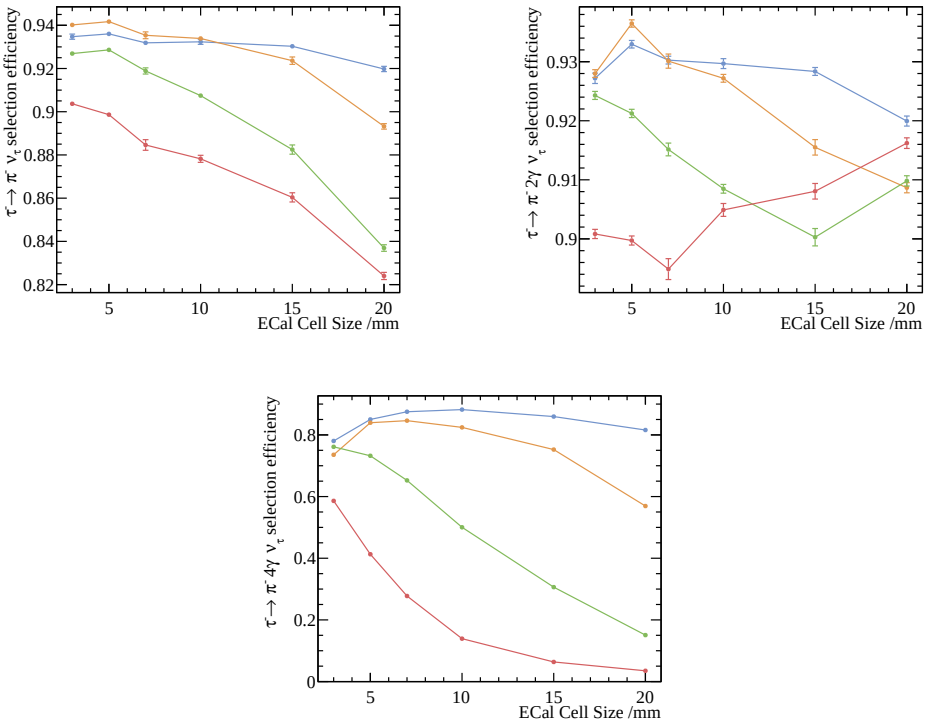

Figure 2. The selection efficiencies for various final states against the ECal cell size for different c.o.m. energies with the nominal CLIC_ILD detector model are shown. The top left, top right and the bottom are for the $\pi^- \nu_\tau$, $\pi^- 2\gamma\ \nu_\tau$ and $\pi^- 4\gamma\ \nu_\tau$ final states respectively. From the top to the bottom, blue, orange, green and red lines are representing the c.o.m. $e^- e^+$ collision energies of 100, 200, 500 and 1000 GeV respectively.

To compare the impact of the ECAL cell sizes and the c.o.m. $e^- e^+$ collision energies on the separation of tau final states, the selection efficiencies for various final states against the ECal cell size for different c.o.m. energies are shown in the figure 2.

Overall, the selection efficiency decreases as the c.o.m. energy increases. This is due to the fact that when τs are boosted at higher 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 $\pi^-\ 2\gamma\ \nu_\tau$ 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 $\pi^-\ 4\gamma\ \nu_\tau$ final state, the 4γ are often merged and the event topology would be very similar to the $\pi^-\ 2\gamma\ \nu_\tau$ final states. Hence more reconstructed events are identified as the $\pi^-\ 2\gamma\ \nu_\tau$ final state.

For the c.o.m. energy of 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.

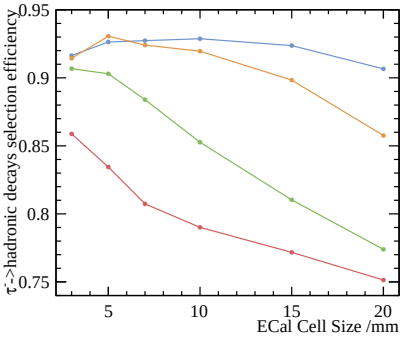

Figure 3. The τ hadronic decay selection efficiency against the ECal cell size for different c.o.m. energies with the nominal CLIC_ILD detector model are shown. The blue, orange, green and red lines are representing the c.o.m. $e^-\ e^+$ collision energies of 100, 200, 500 and 1000 GeV respectively.

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, $\varepsilon_{hadronic} = \Sigma_i Br_i \varepsilon_i / \Sigma_i Br_i$, where Br_i is the branching fraction of a 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^- and μ^-, were not included as the variation of the leptonic decay selection efficiency is small.

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

The $\varepsilon_{hadronic}$ is above 90% for the ECal cell size from 3 to 20 mm for the c.o.m. energy of 100 GeV. For 200 GeV, the $\varepsilon_{hadronic}$ decreases from over 90% to 86% for the ECal cell size from 3 to 20 mm. The degradation of the $\varepsilon_{hadronic}$ is significant for the 500 and 1000 GeV lines, where the $\varepsilon_{hadronic}$ drops from over 90% to 77% and from 86% to 75% respectively.

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

Acknowledgments

The authors would like thank P. G. Roloff for helping generating 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] J.-C. B. M. Reinhard, (2009).

[4] D. Jeans, J. C. Brient, and M. Reinhard, JINST **7**, P06003 (2012), 1203.0774.

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

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

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

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

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

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

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

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

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

[14] 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.

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

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