CERN - European Organization for Nuclear Research

LCD-Note-2012-004

Measurement of $\tilde{\tau}_1$ pair production at CLIC

A. Münnich*

* *CERN, CH-1211 Geneva 23, Switzerland*

July 13, 2012

Abstract

We present a study performed for the CLIC Conceptual Design Report Volume 3 on the measurement of $\tilde{\tau}_1$ pair production at $\sqrt{s} = 1.4$ TeV. Only the hadronic decay of τs is considered. Results obtained using full detector simulation and including beam-induced backgrounds for the mass and for the production cross section of the $\tilde{\tau}_1$ are discussed.

1 Introduction

For the 3rd volume of the CLIC Conceptual Design Report (CDR) [1], several physics processes were selected [2] to benchmark the performance of general purpose detectors at different centre-of-mass energies. The study reported in this note is carried out for a centre-of-mass energy of 1.4 TeV. The goal is to demonstrate the reconstruction of τ leptons at high energies and the presence of machine-induced backgrounds.

The relevant GUT scale parameters of the chosen SUSY model are given in [2].

The final state considered for this study is the decay of the $\tilde{\tau}_1$ into $\tau\tilde{\chi}_1^0$ and the hadronic decays of the τ and hence the investigated final state has two τ jets and missing energy. The corresponding branching ratios and cross section are listed in Table 1.

Table 1: The particle masses, final states, branching ratios and cross section used in the $\tilde{\tau}_1$ production study.

Process:	$e^+e^- \to \tilde{\tau}_1\tilde{\tau}_1$
Particle masses:	$m_{\tilde{\tau}_1} = 517$ GeV
	$m_{\tilde{\chi}_1^0} = 357$ GeV
Final state:	$\tau\tau\tilde{\chi}_1^0\tilde{\chi}_1^0$
Branching ratios:	$\tilde{\tau}_1 \to \tau\tilde{\chi}_1^0$ (99%)
	$\tau \to$ hadrons (64.8%), $\tau \to e\nu\nu$ (17.8%), $\tau \to \mu\nu\nu$ (17.4%)
Cross section:	$\sigma = 2.4$ fb

2 Monte Carlo production

The physics events used for the study presented here were produced with the same procedures as used for the CLIC CDR [1]. Events were generated using the `WHIZARD 1.95` [3] program. Initial and final state radiation (ISR and FSR) were enabled during the event generation. The luminosity spectrum expected at a 1.4 TeV CLIC machine was used during the event generation [4]. The hadronization of final state partons was simulated using `Pythia` [5]. The generated events were subsequently passed through the detector simulation program Mokka [6] which is based on the Geant4 [7] package. The CLIC_ILD [8] detector geometry model was used.

Events were overlayed with pileup from $\gamma\gamma \to$ hadrons interactions corresponding to 60 bunch crossings [9]. The reconstruction chain included an improved version [10] of the PandoraPFA [11] algorithm to reconstruct particle flow objects.

An overview of all produced Monte Carlo (MC) samples is given in Tab. 2.

Table 2: Cross sections and integrated luminosities of the available Monte Carlo samples for the $\tilde{\tau}_1$ pair production and the relevant backgrounds.

Process	Cross section [fb]	Luminosity [ab^{-1}]
$ee \to \tilde{\tau}_1\tilde{\tau}_1$	2.4	21.1
$ee \to \tau\tau\tilde{\chi}_1^0\tilde{\chi}_1^0$ (excl. signal)	0.03	666.8
$ee \to \tau\tau\tilde{\chi}_1^0\tilde{\chi}_1^0\nu\nu$	$3.4{\cdot}10^{-5}$	348371
$ee \to \tau\tau$	5.3	5.7
$ee \to \tau\tau\nu\nu$	38.5	4.6
$ee \to ee\tau\tau$	67.6	1.3
$ee \to \mu\mu\tau\tau$	2.0	10.9
$\gamma\gamma \to \tau\tau$	404.3	0.7
$\gamma\gamma \to \tau\tau\nu\nu$	84.3	2.4
$\gamma\gamma \to ee\tau\tau$	2.6	5.9
$\gamma\gamma \to \mu\mu\tau\tau$	10.6	2.4
$ee \to qq\nu\nu$	648.3	0.5

3 Event reconstruction

The steps to reconstruct events with two τs from particle flow objects (PFOs) are described in this section. The presence of pileup from the process $\gamma\gamma \to$ hadrons increases the number of reconstructed PFOs in typical signal events by a factor 10 and the total visible momentum by a factor four. On the other hand, the background particles are emitted mostly in the forward direction.

A large fraction of the background can be rejected using combined timing and transverse momentum cuts [1]. The effects of different variants of these timing cuts referred to as "loose selected PFOs", "selected PFOs" and "tight selected PFOs" are illustrated in Fig.1. The shape of the reconstructed τ energy distribution remains robust against the background. The τ reconstruction efficiency however, benefits from choosing the "tight selected PFOs" because it reduces the number of background particles remaining in the event. If a τ candidate contains a background particle it will most likely not pass the quality criteria to be accepted as a reconstructed τ.

The reconstruction of τ leptons is done with the `TauFinder` [12] which is essentially a seeded cone based jet clustering algorithm. Via steering parameters like the opening angle of the search and isolation cones the reconstruction algorithm can be optimized for a given τ signature. For this study a scan of the algorithm's parameters was carried out on events with quarks ($ee \to qq\nu\nu$) to find a setting that minimizes the fake rate, meaning mistaking a quark jet for a τ jet.

The following parameter set was chosen for the reconstruction of τ leptons:

- Minimum p_T for τ seed: 15 GeV
- Maximum for invariant mass of τ candidate: 2.5 GeV

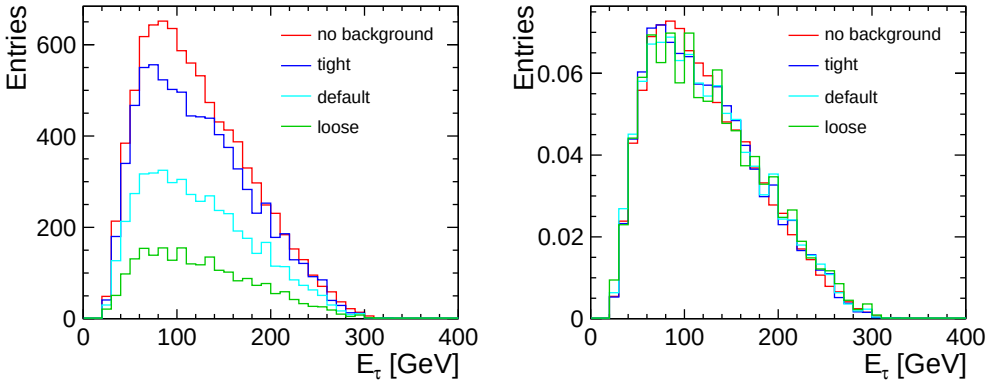

Figure 1: Reconstructed τ energy in signal events without background and for the different variants of the combined timing and momentum cuts to remove pileup from beam-induced backgrounds. The right plot shows the same distributions normalized illustrating that there is no effect on the shape of the distribution.

- Opening angle of search cone: 0.07 rad
- Opening angle of isolation cone (relative to search cone): 0.3 rad
- Maximum energy allowed in isolation cone: 2.0 GeV

With these settings a fake rate of 0.6% to mistake a quark for a τ is achieved. The efficiency to reconstruct a τ from the signal is 60%.

4 Event selection

The selection of $\tilde{\tau}_1^+\tilde{\tau}_1^-$ pair production events is performed in two steps. First, a cut-based pre-selection is applied. The remaining backgrounds are suppressed further using boosted decision trees in a second step. These two steps are described in the following two subsections.

4.1 Pre-Selection cuts

Due to the high cross section of some of the background channels it is necessary to apply some cuts already before entering the full simulation. As some of the variables used in these cuts are sensitive to FSR the cuts can not be applied in WIZARD but only after the fragmentation in `Pythia`. This means the cuts are applied on the stdhep files before entering the Geant4 simulation. The following cuts are applied for both τ candidates on almost all background channels:

- $10 < \theta_\tau < 170$ deg, where θ_τ is the polar angle of the τ candidate

Table 3: Efficiency flow for the different steps of the analysis.

Step	Efficiency or Branching Ratio [%]	Events
Produced	BR($\tilde{\tau}_1 \rightarrow \tau\tilde{\chi}_1^0$) = 0.99	50949
τ reconstruction	Eff(τ) = 0.6	18440
hadronic τ decay	BR($\tau \rightarrow$ hadrons) = 0.6	6694
pre-selection cut	Eff(PreSel) = 0.8	5399

- $p_T > 20$ GeV
- $\Delta\phi < 178$ deg
- angle of τ system > 0.4 rad (23 deg)
- $40 <$ invariant mass of τ system < 650 GeV

For one background channel ($ee \rightarrow \tau\tau$) the full simulation was carried out on a data set with and without these cuts applied. A comparison of the number of events remaining in the data set after applying the pre-selection cuts confirmed that no bias is introduced by applying these cuts at this level. The cuts themselves are based on a study of signal and background channels at generator level.

The pre-selection cuts used in the analysis have to be stronger than the cuts already applied at generator level to avoid a bias. In addition a cut on the leptonic energy content in the τ candidate is applied to select only hadronic τ decays. Further cuts are applied to eliminate background dominated areas of the phase space.

- no leptons in τ
- $15 < \theta_\tau < 165$ deg
- $p_T > 25$ GeV
- $\Delta\phi < 177$ deg
- angle of τ system > 0.5
- $45 <$ invariant mass of τ system < 600 GeV
- Thrust < 0.99
- $80 <$ transverse mass of τ system < 500 GeV
- Number of tracks in τ candidate either 1 or 3

The signal efficiency for the different steps is summarized in Table 3.

After the pre-selection 10% of the produced signal statistics is left for the analysis. Already 64% of the statistics is lost by requiring both τs to decay hadronically. If we define only those events as signal, then the reconstruction and pre-selection combined have an efficiency of 30%.

4.2 Event selection using boosted decision trees

To distinguish between signal and background events further, the Toolkit for Multivariate Analysis (TMVA) [13] is used. Boosted decision trees (BDT) proved to be the most efficient classifiers for this analysis. For training purposes, 30% of the available events for each process are used. These events are not considered in the analysis to measure masses or cross sections.

Event classification

The boosted decision trees are trained using 16 variables describing the event topology and kinematic quantities of the reconstructed τ candidates:

- Missing transverse momentum $p_{\mathrm{T,miss}}$
- Thrust of the τ system and the full event
- Oblateness of the τ system and the full event
- Sum of the transverse momenta of both τ candidates
- $\cos\theta_{\tau,1}$ and $\cos\theta_{\tau,2}$, where θ_τ is the polar angle of the τ candidate
- Invariant mass of the τ system
- Transverse mass of the τ system
- Angle between the two τ candidates
- θ^{miss}, where θ^{miss} is the polar angle of the missing momentum
- $\Delta\phi$ between the two τs
- Highest momentum of a particle in the event
- Visible energy in the event
- Energy of the rest group (particles not belonging to a τ candidate)

As examples for the input variables, the polar angle of a τ candidate and the missing momentum in the event for the signal and all backgrounds are shown in Fig. 2. The signal candidates are more central than the backgrounds and have larger missing momentum.

Using the input variables described above, the classifier response for each event is computed which is referred to as BDT in the following. Figure 3 shows the distribution of the BDT for the signal and the different backgrounds. Signal events tend to be at higher BDT values than the backgrounds. Additionally, the signal efficiency, purity and significance for events passing the pre-selection as a function of the BDT cut value are shown. The highest significance is reached for a BDT of 0.08 which is then chosen for the analysis giving an efficiency of 57% and a purity

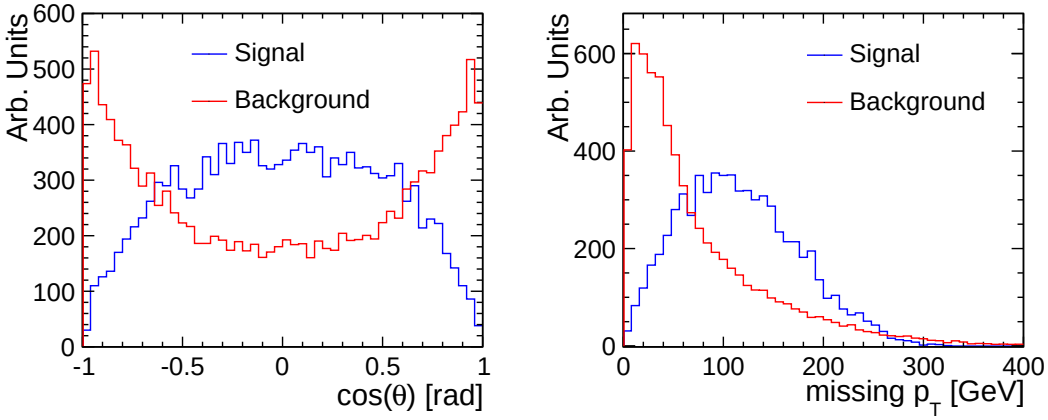

Figure 2: Polar angle (left) of the τ candidate and missing momentum in the event (right) for the signal and all backgrounds. The normalization of all distributions is arbitrary to illustrate the different shapes.

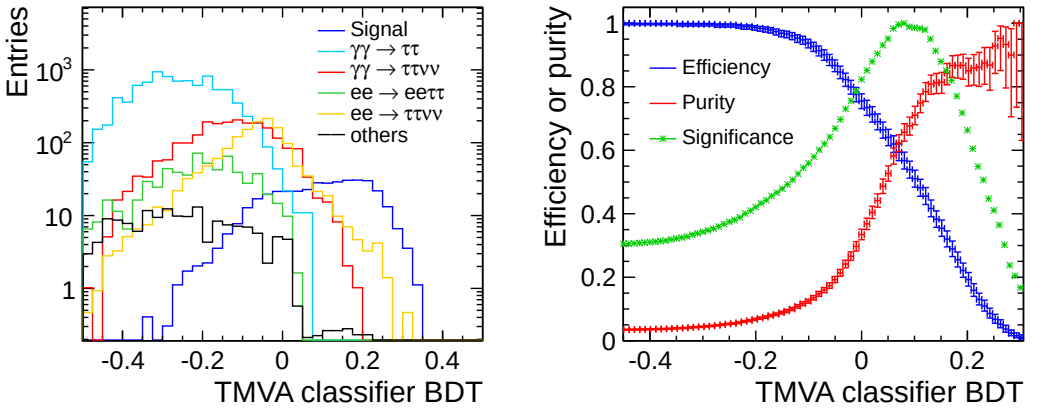

Figure 3: Distribution of the BDT values for the signal and the backgrounds for 1.5 ab^{-1} (left) and the selection efficiency, purity and significance in dependence on the chosen BDT cut value (right).

of 70% for this selection. Combined with the efficiency from the reconstruction and the preselection this amounts to a signal efficiency of 5% of the produced data set or 15% of hadronic τ decays.

The energy of the reconstructed τ candidates before and after the selection is shown in Fig. 4. The signal and all backgrounds were scaled to an integrated luminosity of 1.5 ab^{-1}. The plots demonstrate the significant improvement of the signal-to-background ratio due to the event se-

lection. Most of the backgrounds are completely suppressed after the BDT cut. The two remaining backgrounds are $ee \to \tau\tau\nu\nu$ and $\gamma\gamma \to \tau\tau\nu\nu$

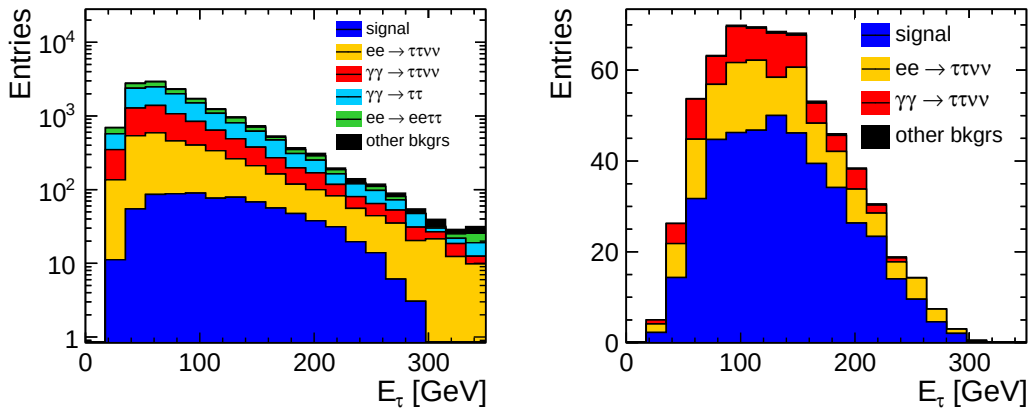

Figure 4: Reconstructed τ energy before (left) and after (right) the application of the BDT cut for the signal and for the backgrounds. The histograms are stacked on top of each other. All distributions are scaled to an integrated luminosity of 1.5 ab^{-1}.

Figure 5 shows the efficiency and purity in dependence on the reconstructed τ energy. The low energy region is dominated by background as can be seen from Figure 4 and therefore a drop in efficiency and purity is observed for τ energies below 100 GeV.

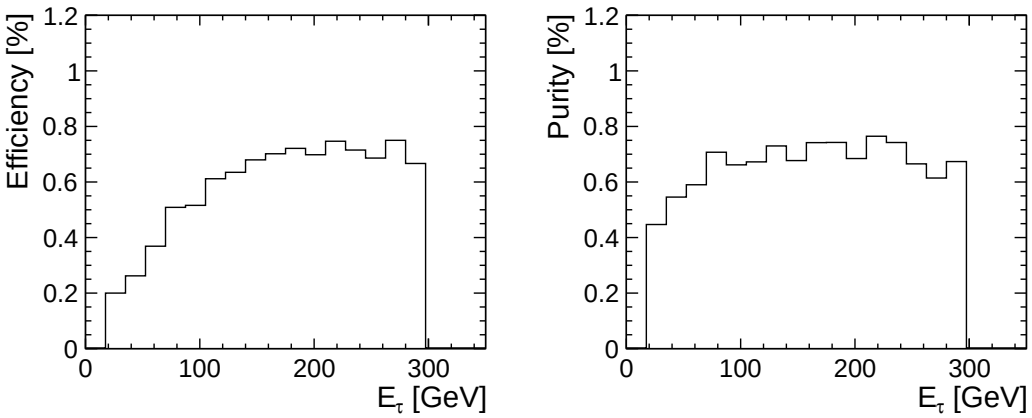

Figure 5: Efficiency (left) and purity (right) of the event selection in dependence on the reconstructed τ energy.

5 Signal extraction with template fit

The cross section and mass of the $\tilde{\tau}_1$ is determined using the template method where signal Monte Carlo samples for different mass hypotheses were produced with full simulation and considering pileup from $\gamma\gamma \rightarrow$ hadrons. The energy distribution of the reconstructed τs is shown in Fig. 6 on the left for different mass templates. To avoid large uncertainties due to the limited statistics of the available Monte Carlo samples for the background processes, the sum of all backgrounds is parametrized using a smooth function. The obtained functions are compared to the sum of all background contributions for an integrated luminosity of 1.5 ab^{-1} in Fig. 6 on the right.

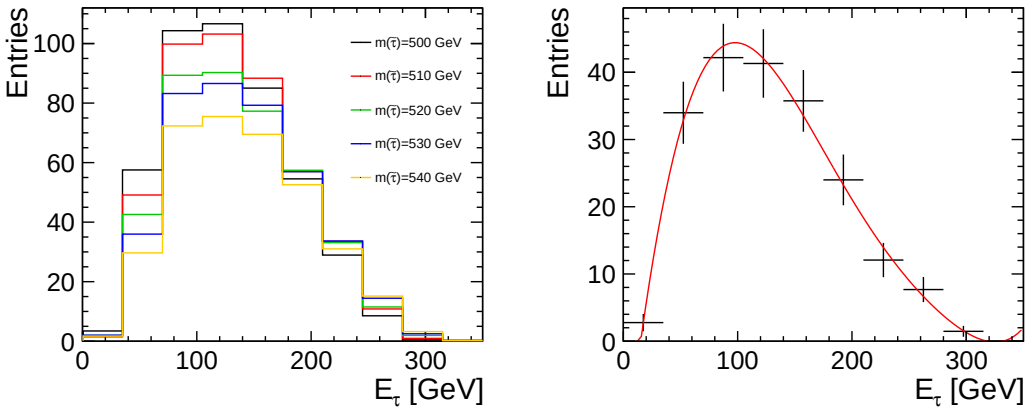

Figure 6: Different mass templates for the χ^2 calculation (left). Parametrization of the background for the selected data set (right).

The fits are restricted to the energy range $50 < E(\tau) < 260$ GeV to exclude bins with low statistics or regions where the parametrization does not describe the backgrounds well. Two-dimensional fits are performed simultaneously to the mass and production cross section to account for the correlation between both quantities. The following function is minimized:

$$\chi^2_j = \sum_{i=1}^{N} \frac{(n_{\text{data},i} - c_j \cdot n_{\text{template}_j,i} - n_{\text{background},i})^2}{\sigma^2_{\text{data},i} + \sigma^2_{\text{template}_j,i} + \sigma^2_{\text{background}_j,i}},$$

where the sum runs over all τ energy bins i. The optimal normalization c_j is obtained for every tested mass value j. The number of entries in a given energy bin for the templates and for the measured distribution are referred to as $n_{\text{template}_j,i}$ and $n_{\text{data},i}$, respectively, while the corresponding statistical uncertainties are given by $\sigma_{\text{template}_j,i}$ and $\sigma_{\text{data},i}$. The expected number of background events is referred to as $n_{\text{background}_j,i}$ with a statistical uncertainty $\sigma_{\text{background},i}$. The obtained minimal χ^2 values for different mass values is shown in Fig. 7.

The optimal value of the $\tilde{\tau}_1$ mass is obtained from the minima of a parabola fitted to the χ^2 distributions. The toy MC method is used to estimate the statistical uncertainties of the measured

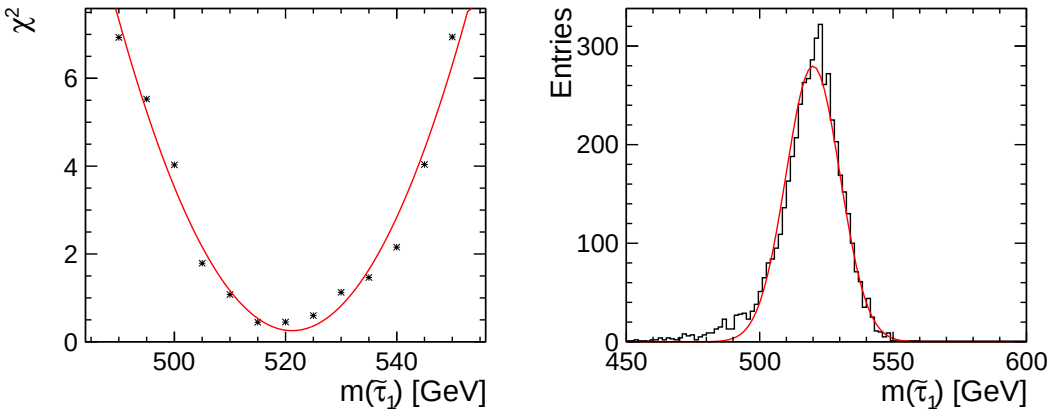

Figure 7: Minimal χ^2 values as a function of $M(\tilde{\tau}_1)$ (left) and the distribution of $M(\tilde{\tau}_1)$ obtained using a toy MC with 5000 trials (right).

mass. For this purpose, the data points in the measured τ energy distributions are smeared using a Gaussian distribution of width $\sqrt{n_{\text{data},i}}$, where $n_{\text{data},i}$ is the number of entries in a given bin. This step is repeated 5000 times and the optimal mass value is extracted for each iteration. The resulting distribution of optimal mass values is shown in Fig. 7. The width of this distribution taken from a Gaussian fit represents the statistical uncertainty of the measured $\tilde{\tau}_1$ mass.

The cross section and its statistical uncertainty is obtained in the same way. The results are given in Tab. 4. Additionally, the fit results obtained when the number of bins in the reconstructed τ energy distribution was increased are given. The results demonstrate the stability of the fit procedure. Given the limited statistics the number of bins can not be extended further while still providing enough entries per bin for the χ^2 fit. The extracted $\tilde{\tau}_1$ mass value of 521 GeV for 10 bins and 519 GeV for 15 bins is compatible with the input value of 517 GeV within the statistical uncertainty.

Table 4: Statistical uncertainty of the $\tilde{\tau}_1$ mass and pair production cross section obtained from two parameter template fits. The number of bins between 0 and 350 GeV used for the histograms of the reconstructed τ energy distributions in the fit is increased from 10 to 15. An integrated luminosity of 1.5 ab^{-1} is assumed.

Parameter	Binning	Uncertainty
$M(\tilde{\tau}_1)$	10	10.2 GeV
	15	11.3 GeV
$\sigma(\tilde{\tau}_1\tilde{\tau}_1)$	10	8.1%
	15	7.5%

5.1 Systematic Uncertainties due to Event Selection

Systematic uncertainties from the event selection process were studied in two different ways. First the influence of the choosen BDT value was studied by choosing two different BDT values instead of 0.08. The event selection and subsequent template fit were carried out with a BDT value of 0.05 as well as 0.1. The resulting variation in mass value and error listed in Table 5 is well within the statistical precision.

Table 5: Sensitivity of extracted mass values and errors to the chosen BDT value for the event selection.

BDT value	$M(\tilde{\tau}_1)$ [GeV]	Uncertainty [GeV]
0.05	519.8	10.8
0.08	521.2	10.2
0.1	521.0	10.2

The second evalution is adressing the systematics of training the BDT with a signal using the exact stau mass which in reality might not be known precisely. Therefore two more BDTs were trained using the lowest ($M(\tilde{\tau}_1)$=490 GeV) and the highest ($M(\tilde{\tau}_1)$=550 GeV) mass template as signal input. The event selection and template fit were repeated using these two additional BDTs. The observed shift in mass is in the same order as previously seen from variation of the binning. The results are summarized in Table 6.

Table 6: Sensitivity of extracted mass values and errors to the mass of the signal used to train the BDT.

trained BDT	$M(\tilde{\tau}_1)$ [GeV]	Uncertainty [GeV]
low mass (490 GeV)	516.2	8.7
nominal mass (517 GeV)	521.2	10.2
high mass (550 GeV)	521.5	11.7

For the template fit method the mass of the $\tilde{\chi}_1^0$ was not varied and the exact mass of 356.57 GeV from the SUSY model was used. From a corresponding study of sleptons as well as gauginos the mass uncertainty for the $\tilde{\chi}_1^0$ was found to be 0.5 GeV. In order to estimate the effect of the mass error on $\tilde{\chi}_1^0$ on this study two templates were generated, one with increasing and one with decreasing the $\tilde{\chi}_1^0$ mass by 1 GeV. The difference of these two templates is consistent with zero given the statistical errors as shown in Fig 8. Therefore the influence of the uncertainty of the $\tilde{\chi}_1^0$ mass is negligible for this study.

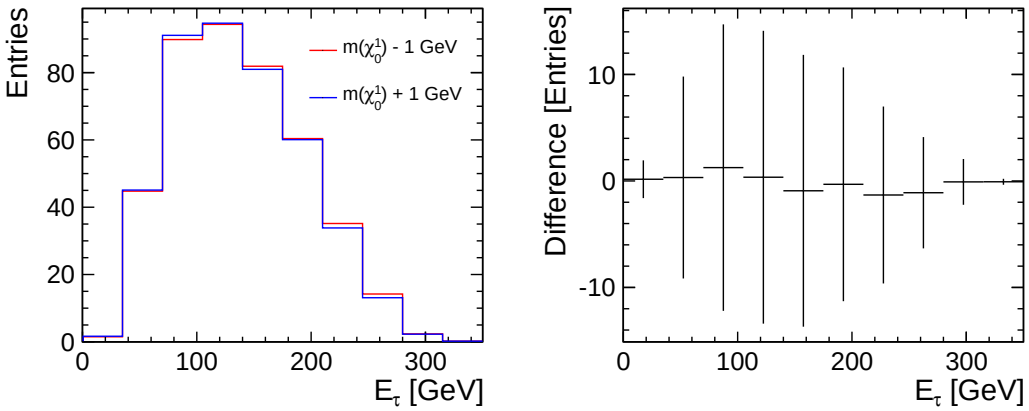

Figure 8: Templates with 1 GeV variation on the $\tilde{\chi}_1^0$ mass (left) and their difference (right). The statistical fluctuations are larger than the impact of a change in the $\tilde{\chi}_1^0$ mass.

6 Conclusion and Summary

The signal from $\tilde{\tau}_1$ pair production is extracted from hadronic decays of the two τs. The study was performed using full simulation and considering pileup from $\gamma\gamma \to$ hadrons. A center-of-mass energy of 1.4 TeV and an integrated luminosity of 1.5 ab^{-1} were used. The signal was extracted using a template fit. The pair production cross section is extracted with a statistical precision of 8% while the mass of the $\tilde{\tau}_1$ particle is determined with a statistical accuracy of about 11 GeV (2%).

References

[1] L. Linssen et al., *Physics and Detectors at CLIC: CLIC Conceptual Design Report*, CERN-2012-03

[2] B. Allamach et al., *The physics benchmark processes for the detector performance studies of the CLIC CDR Volume 3*, LCD-Note-2012-003 (2012).

[3] W. Kilian, T. Ohl and J. Reuter, *WHIZARD: Simulating Multi-Particle Processes at LHC and ILC*, arXiv:0708.4233 (2007); M. Moretti et al., *O'Mega: An Optimizing Matrix Element Generator*, LC-TOOL-2001-040, also arXiv:hep-ph/0102195 (2001).

[4] B. Dalena, J. Esberg and D. Schulte, *Beam-induced backgrounds in the CLIC 3 TeV CM energy interaction region*, Proceedings of LCWS11, arXiv:1202.0563 (2012).

[5] T. Sjörstrand, S. Mrenna and P. Skands, *PYTHIA 6.4 physics and manual* , JHEP 05 (2006) 026.

[6] Simulator for the Linear Collider (MOKKA), `http://www.lcsim.org/software/slic/`.

[7] S. Agostinelli et al., *GEANT4 - a simulation toolkit*, Nucl. Instr. Meth. A 506 (2003) 250; J. Allison et al., *Geant4 Developments and Applications*, IEEE Trans. Nucl. Sci. 53 (2006) 270.

[8] A. Münnich and A. Sailer, *The CLIC ILD CDR Geometry for the CDR Monte Carlo Mass Production*, LCD-Note-2011-002 (2011).

[9] P. Schade, A. Lucaci-Timoce, *Description of the signal and background event mixing as implemented in the Marlin processor OverlayTiming*, LCD-Note-2011-006 (2011).

[10] J. Marshall and M.A. Thomson, *Redesign of the Pandora Particle Flow algorithm*, Report at IWLC2010 (2010).

[11] M.A. Thomson, *Particle Flow Calorimetry and the PandoraPFA Algorithm*, Nucl. Inst. Meth. A 611 (2009) 25.

[12] A. Münnich, *TauFinder: A Reconstruction Algorithm for τ Leptons at Linear Colliders*, , LCD-Note-2010-009 (2010).

[13] A. Hoecker et al., *TMVA 4 - Toolkit for Multivariate Data Analysis with ROOT*, arXiv:physics/0703039, also CERN-OPEN-2007-007 (2009).