CERN - European Organization for Nuclear Research

LCD-Note-2010-009

`TauFinder`: A Reconstruction Algorithm for τ Leptons at Linear Colliders

A. Muennich*

* *CERN, Switzerland*

October 11, 2010

Abstract

An algorithm to find and reconstruct τ leptons was developed, which targets τs that produce high energetic, low multiplicity jets as can be observed at multi TeV e^+e^- collisions. However, it makes no assumption about the decay of the τ candidate thus finding hadronic as well as leptonic decays. The algorithm delivers a reconstructed τ as seen by the detector. This note provides an overview of the algorithm, the cuts used and gives some evaluation of the performance.

A first implementation is available within the ILC software framework as a MARLIN processor . Appendix A is intended as a short user manual.

Contents

1. The `TauFinder` Algorithm

The proposed algorithm resembles a jet finder cone algorithm using the 4 vectors of all detectable charged and neutral particles with some specific criteria and cuts. The method for the algorithm is the following:

1. Starting with the highest energy, each **charged particle** is tested as a **seed** for the τ candidate based on transverse momentum and impact parameter.

2. Once a seed is found, the remaining charged particles present within the **search cone** around the seed are added to the τ candidate adjusting the direction of the cone for the new combined momentum. The search cone is defined by the opening angle between the momenta of the two particles.

3. After that, **neutral particles** are added to the τ candidate in the same fashion.

4. The steps 1 through 3 are repeated until no further seed is found.

5. The momenta and energies of all particles associated to one τ candidate are **combined** into a reconstructed τ.

6. Finally, once all τ candidates in the event are found a check is performed to see whether one candidate was erroneously split up by the algorithm. This can happen in cases where one or more decay products with lower momentum are just outside of the search cone. If the angle between two reconstructed τ candidates is smaller than the opening angle of the search cone they are **merged**.

Whether the reconstructed τ candidate is accepted is evaluated based on a few selection cuts discussed in section 1.2.

1.1. Data Sets

The data sets and their statistics used to evaluate the algorithm are listed in table 1. All processes were simulated at a CLIC[1] energy of 3 TeV including initial state radiation and in the case of the SUSY processes also beamstrahlung. The parameters for the SUSY processes are according to Benchmark Point K'[2]. The different topologies are not weighted to the same luminosity, so that the contribution of τs from $\tilde{\tau}$s dominates the distribution illustrating properties of the τs. The performance of the algorithm however is of course evaluated separately for the different topologies.

Process	Events	True τs
$e^+e^- \to W^+W^-$	672	154
$e^+e^- \to t\bar{t}$	836	265
$e^+e^- \to \tilde{\tau}_1^+\tilde{\tau}_1^-$ ($m_{\tilde{\tau}_1}$=896 GeV)	5000	10000
$e^+e^- \to \chi_1^+\chi_1^-$ (m_{χ_1}=1064 GeV)	1000	1324
$e^+e^- \to H^0A^0$ (m_{H^0}=1161 GeV,m_{A^0}=1153 GeV)	762	541

Table 1: Physics processes used to study τ properties and evaluate the algorithm. All SUSY parameters are chosen according to Benchmark Point K'[2].

1.2. The Selection Cuts

There are a couple of cuts to influence the algorithm and to select "good" τs from the candidates. Some are fixed and others can be changed by the user:

Fixed Quality Cuts

- The multiplicity of tracks within the τ-jet is low, therefore the number of charged tracks must be larger than zero but smaller than four.
- The total number of charged and neutral particles combined to a τ has to be below 10.
- The charge of the τ has to add up to 1 or -1.

These cuts are based on studying the τ decay products based on Monte Carlo truth (MC). Figure 1 shows the distribution of the number of charged and the sum of charged and neutral decay products based on the processes listed in Table 1.

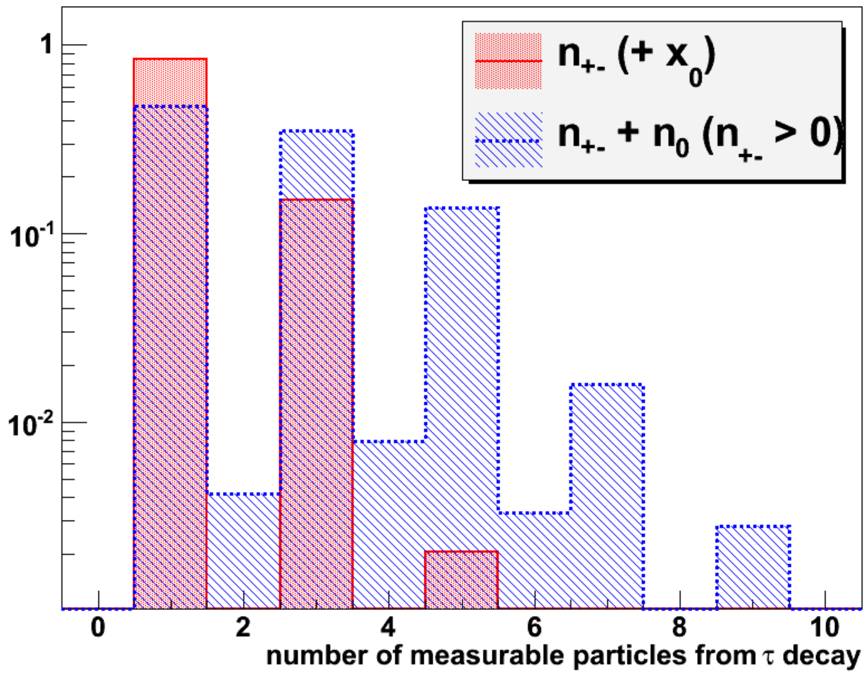

Figure 1: Number of charged particles ($n_\pm$) plus any number of neutral particles (x_0) and the sum of charged and neutral particles ($n_\pm + n_0$) in the detector from one τ decay based on MC truth from the processes listed in Table 1. A τ has at least one charged decay product, therefore $n_\pm > 0$.

User Parameters

Other selection cuts can be set by the user. These user parameters are listed here with the default values given in brackets.

- Reconstruction Cuts:
  - A general cut to suppress background by requiring a minimum transverse momentum for a particle to be considered in the algorithm ($p_T > 1$ GeV/c).
  - A minimum transverse momentum for the τ seed ($p_T > 5$ GeV/c).
  - A lower limit on the impact parameter D0 for the τ seed ($D0 > 10^{-5}$ mm) .

  - An upper limit on the impact parameter D0 for the τ seed (D0 < 0.5 mm).
  - The opening angle of the search cone (0.05 rad).

- Quality Cuts:
  - The isolation criterion consist of two parameters:
    1. The opening angle of the isolation cone given relative to the search cone (+0.02 rad). Since τs are mostly isolated jets this second cone defines an area around the search cone which is used to evaluate the energy content of the surroundings.
    2. A limit on the energy of the most energetic particle that is allowed within the isolation cone (< 5 GeV).

Figure 2 shows the true full energy of the τ and the part visible in the detector for three of the processes in Table 1.

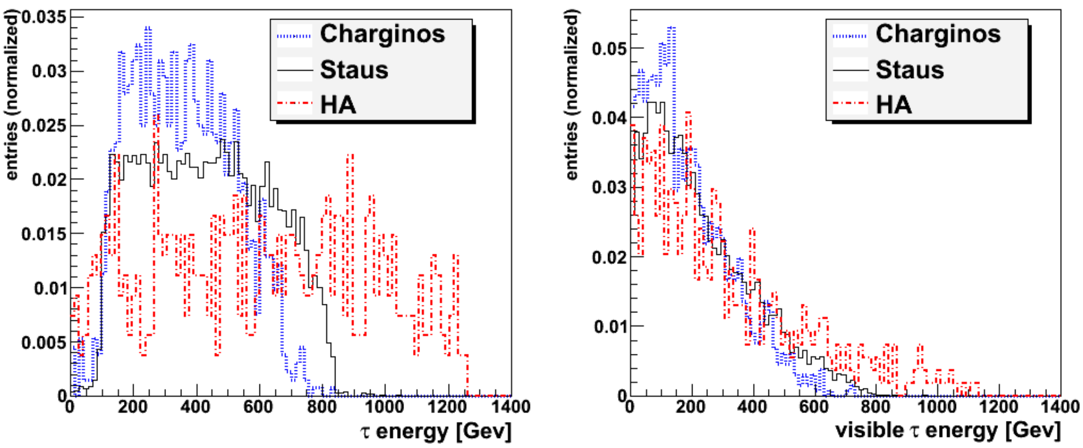

Figure 2: Full and visible energy of the τ based on Monte Carlo truth for three different processes from Table 1.

Figure 3 gives an example of the distribution of the impact parameter and the opening angle of the τ jet based on Monte Carlo information from the processes listed in Table 1. The choice of the selection cuts will depend on the event topology in question and the background conditions.

2. Evaluation of the Algorithm

2.1. Evaluation Criteria

In order to evaluate the algorithm the following variables are used:

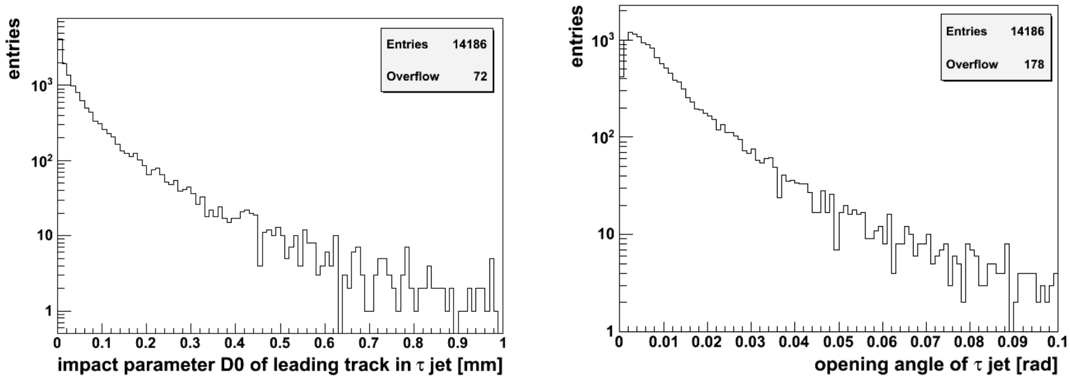

Figure 3: Impact parameter of the leading track and opening angle of the τ jet based on MC truth from the processes listed in Table 1.

- N_τ: Number of τs in the MC truth.
- *Missed*: Number of τs not recognized, e. g. seed not found, or rejected by selection cuts.
- *Reconstructed*: Number of τs reconstructed.
- *Matched*: Number of reconstructed τs where at least one of the particles used to form the τ links back to a τ in the MC truth.
- *Fake*: Number of reconstructed τs where none of the particles used to form the τ links back to a τ in the MC truth.
- *Clean*: Number of reconstructed τs where all the particles used to form the τ link back to a τ in the MC truth.
- *Contaminated*: Difference between *Matched* and *Clean*.

Figure 4 illustrates as an example how a data sample of Charginos splits into the different contributions.
In order to define the efficiency and purity the important variables are N_τ, *Matched* and *Reconstructed*:

$$\text{Efficiency}: \quad E = \frac{Matched}{N_\tau} = 94.5\% \pm 0.5\%$$

$$\text{Purity}: \quad P = \frac{Matched}{Reconstructed} = 97.3\% \pm 0.6\%$$

The numbers given correspond to the example illustrated in Figure 4 and the errors are calculated using a probability density function to derive the variance according to [3].

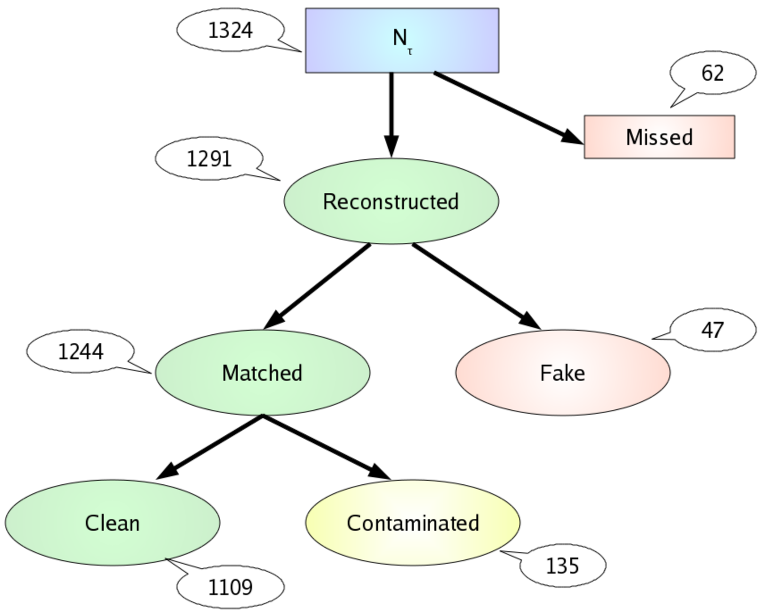

Figure 4: Illustration of nomenclature within a data sample. The numbers are an example of the different contributions when running `TauFinder` on a data sample of Charginos.

2.2. Influence of the Selection Cuts

To study the influence of the selection cuts on efficiency and purity a parameter scan was carried out with the following cut values:

- p_T (PT) $>$ [0, 1] GeV/c
- p_T of seed (PTS) $>$ [0, 5, 10] GeV/c
- $10^{-5} < \mathrm{D0} <$ [0.3, 0.5, 0.7] mm
- Search cone (SC): [0.03, 0.05, 0.07] rad
- Isolation cone (IC): [0.02, 0.04] rad
- Isolation energy (IE) $<$ [3, 5, 10] GeV

on three different data sets from Table 1 without background: $e^+e^- \rightarrow \tilde{\tau}_1^+\tilde{\tau}_1^-$, $e^+e^- \rightarrow \chi_1^+\chi_1^-$ and $e^+e^- \rightarrow H^0A^0$.
These results are displayed in Figure 5 which shows a comparison of the performance of `TauFinder` on the three different processes without background. Depending on the combination chosen for the cuts the performance can be optimized for either efficiency or purity.

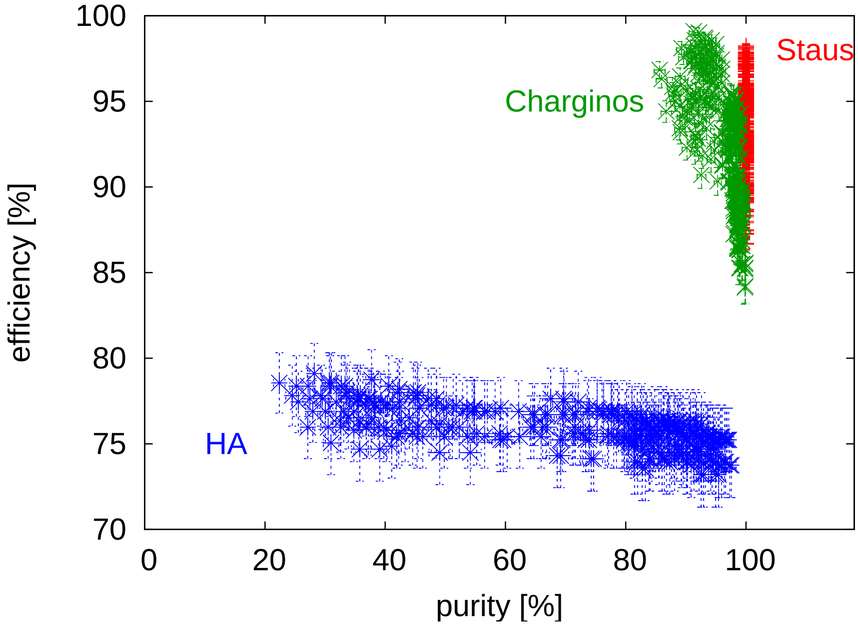

Figure 5: Performance of `TauFinder` for three different data sets without background for a parameter scan of selection cut values.

The reconstruction of τs from $\tilde{\tau}$s decays can be compromised by a high energetic photon radiated off earlier in the process chain. Therefore the efficiency is not 100% because these contaminated τ candidates will fail the isolation cut. The purity however is always 100% since there are only τs and neutralinos in the sample.
The Chargino decays to 60% into $\tilde{\tau}$s and to 40% into Ws. The slight drop in efficiency and purity compared to the pure $\tilde{\tau}$ sample is caused by the W decay into quarks. The jets produced by quarks are very similar to the jets from τs which increases the number of falsely reconstructed τs (*Fake*).
In the case of HA both decay into Ws and many quark jets are present which leads to a higher number of *Fake* τs. Hence the algorithm is less efficient. Figure 6 shows the different contributions to mis-identification for τs in the process $e^+e^- \rightarrow H^0A^0$. The efficiency for a b-quark to be tagged as a τ can reach up to 40% depending on the cut selection.
Table 2 gives an overview about the effect of the different cuts on efficiency and purity. The listed cut selection for the different event topologies is once optimized for efficiency and the other time for purity. In the case of the HA the trade off between the two is rather large. Furthermore the sacrifice of purity does not gain as much in efficiency, which is also evident in Figure 5 where the spread in purity is large but the range in efficiency is limited.
Starting with the cut selections in Table 2 and varying just one cut at a time an estimate of the power of the cut can be obtained. The most influential cut is the selection of a minimum transverse momentum for the seed. This cut improves the purity in the HA sample by 60%.

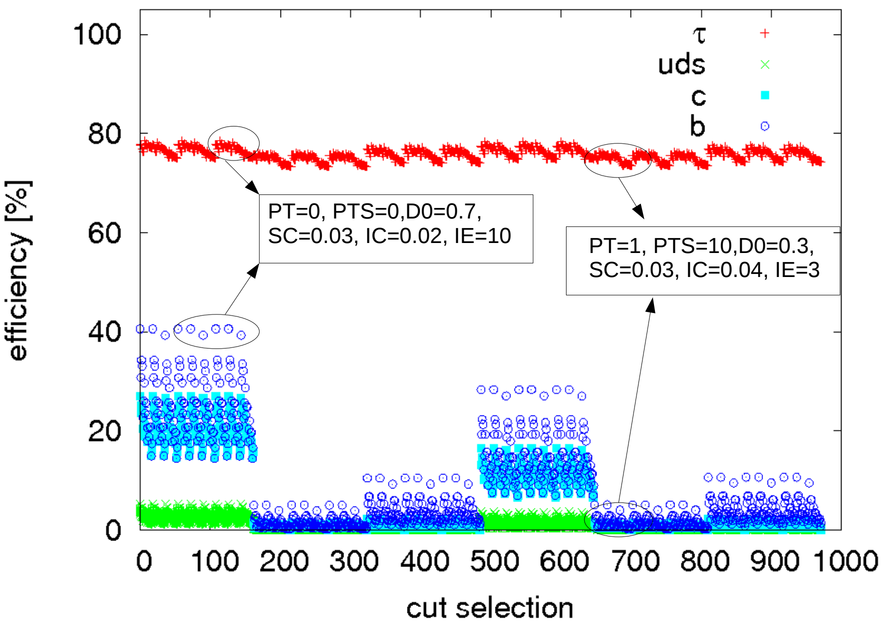

Figure 6: Tagging efficiencies for τ_s and background from quarks. The x axis represents the variation of the selection cuts and the y axis the τ tagging efficiency for the τ itself and the mis-tagging of quarks for the process $e^+e^- \to H^0A^0$.

Other cuts can change efficiency and purity in the order of a few percent.

Process	optim.	PT	PTS	D0	SC	IC	IE	E	P
general	E	-	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$	$\uparrow$	$\uparrow$	$\downarrow$
	P	-	$\uparrow$	$\downarrow$	$\downarrow$	$\uparrow$	$\downarrow$	$\downarrow$	$\uparrow$
$\tilde{\tau}_s$	E	0	0	0.7	0.07	0.02	10	98.2 ± 0.1	100
$\tilde{\chi}_{1s}$	E	0	0	0.7	0.07	0.02	10	99.1 ± 0.3	91.3 ± 0.7
	P	1	10	0.3	0.03	0.04	3	84.2 ± 1.0	99.9 ± 0.1
HA	E	0	0	0.7	0.05	0.02	10	79.1 ± 1.7	28.2 ± 1.2
	P	0	10	0.3	0.05	0.04	3	73.8 ± 1.9	97.6 ± 0.8

Table 2: Cut selection optimized for either efficiency (E [%]) or purity (P [%]) for different processes. The arrows in the first two rows indicate the trend of each cut to optimize E or P.

2.3. Influence of $\gamma\gamma$ background

Processes in the barrel are mostly unaffected by the forward peaked $\gamma\gamma$ background. Therefore the process $e^+e^- \to W^+W^-$ was chosen to study the influence of this background on the algorithm. Based on 4549 events with 1011 τs different levels of background (0, 20 and 40 bunch crossings (BX)) per event were generated. Efficiency and purity were evaluated for the same cut parameter scan as in section 2.2. The influence of the overlaid $\gamma\gamma$ background is shown in Figure 7. In the case of no background there is a parameter space were higher purity can be achieved without sacrificing efficiency. This can not be reached in the presence of background. The difference between 20 and 40 bunch crossings is however minimal.

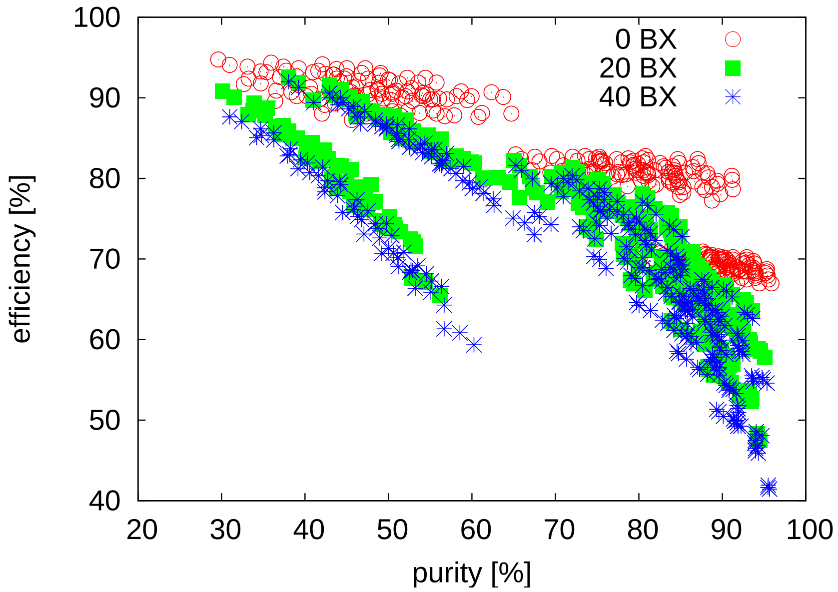

Figure 7: Performance of TauFinder for three different levels of $\gamma\gamma$ background (0, 20 and 40 bunch crossings) for a parameter scan of selection cut values.

2.4. Influence of τ decay channel

TauFinder is generic and reconstructs hadronic and leptonic τ decays. However the algorithm is more geared towards jets and does not use any information about lepton ID. In order to study whether the decay channel of the τ has an impact on the performance of the algorithm the main decay channels were evaluated separately. Table 3 lists the decay channels responsible for about 94% of all τ decays.

For each process in Table 1 the efficiency of TauFinder for every of the main decay channels in Table 3 is determined. The results are given in Table 4. There is no significant difference between the performance of finding a τ decaying into leptons or hadrons. The production process of the τ and the selection cuts are the dominant factors.

Decay	Occurrence [%]
$\tau \rightarrow \mu$ + missing energy	18
$\tau \rightarrow$ e + missing energy	17
$\tau \rightarrow \pi$ + missing energy	37
$\tau \rightarrow \pi\pi\pi$ + missing energy	12
$\tau \rightarrow \pi\pi^0$ + missing energy	10

Table 3: Main decay channels of a τ as seen in the detector.

Process	Decay	Eff. [%]	tot. Eff.[%]	Purity [%]
$e^+e^- \rightarrow W^+W^-$	$\tau \rightarrow \mu$	77.1 ± 3.1	80.3 ± 1.3	91.1 ± 1.0
	$\tau \rightarrow$ e	81.1 ± 3.0		
	$\tau \rightarrow \pi$	82.9 ± 2.0		
	$\tau \rightarrow \pi\pi\pi$	84.2 ± 3.3		
	$\tau \rightarrow \pi\pi^0$	70.6 ± 4.3		
$e^+e^- \rightarrow tt$	$\tau \rightarrow \mu$	42.9 ± 7.4	49.1 ± 3.1	69.5 ± 3.3
	$\tau \rightarrow$ e	52.0 ± 6.8		
	$\tau \rightarrow \pi$	56.0 ± 5.1		
	$\tau \rightarrow \pi\pi\pi$	45.9 ± 7.9		
	$\tau \rightarrow \pi\pi^0$	36.0 ± 9.1		
$e^+e^- \rightarrow \tilde{\tau}_1^+\tilde{\tau}_1^-$	$\tau \rightarrow \mu$	98.6 ± 0.3	98.2 ± 0.1	100
	$\tau \rightarrow$ e	97.5 ± 0.4		
	$\tau \rightarrow \pi$	98.6 ± 0.5		
	$\tau \rightarrow \pi\pi\pi$	98.4 ± 0.4		
	$\tau \rightarrow \pi\pi^0$	98.2 ± 0.4		
$e^+e^- \rightarrow \chi_1^+\chi_1^-$	$\tau \rightarrow \mu$	97.0 ± 1.2	98.3 ± 0.4	95.0 ± 0.6
	$\tau \rightarrow$ e	98.0 ± 1.0		
	$\tau \rightarrow \pi$	99.2 ± 0.4		
	$\tau \rightarrow \pi\pi\pi$	100		
	$\tau \rightarrow \pi\pi^0$	96.3 ± 2.0		
$e^+e^- \rightarrow H^0A^0$	$\tau \rightarrow \mu$	80.2 ± 3.7	75.2 ± 1.8	97.1 ± 0.8
	$\tau \rightarrow$ e	80.4 ± 3.8		
	$\tau \rightarrow \pi$	69.2 ± 3.4		
	$\tau \rightarrow \pi\pi\pi$	75.8 ± 5.2		
	$\tau \rightarrow \pi\pi^0$	77.5 ± 5.9		

Table 4: Efficiency of the algorithm separated for different processes and τ decay channels. The same selection cuts were used for all processes.

3. Performance including Detector Resolution

In order to study the performance of the algorithm under more realistic conditions the detector resolution needs to be taken into account. The three measured quantities which are important for the `TauFinder` are the impact parameter, the momentum and the energy of the particles. The impact parameter is the most influential because it determines the seed of the τ.

3.1. Impact Parameter Resolution

A Gaussian smearing is applied to the impact parameter D0 with a width of

$$\sigma_{D0} = \sqrt{a^2 + \frac{b^2}{p^2(\sin\theta)^3}}.$$

The parameters a and b are given by the resolution of the vertex detector and are expected to be in the range of 2 μm $< a <$ 6 μm and 10 μm GeV $< b <$ 20 μm GeV at a Linear Collider.
With the introduction of the impact parameter resolution the lower cut on D0 is no longer efficient since now all particles have a significant value for D0. Increasing this cut drastically lowers the efficiency because most τs do have a small D0 as could be seen in Figure 3. Increasing the cut on the transverse momentum of the seed however does improve the performance. In the following efficiency and purity for three different processes will be compared. All cuts remain the same (PT=0, D0=0.3, IM=2, SC=0.03, IC=0.04, IE=3) and p_T of the seed (PTS) is varied. The impact parameter resolution is set to a=2 μm and b=20μm GeV.
Figure 8 shows the efficiency and purity of the algorithm for the process $e^+e^- \rightarrow H^0A^0$ with different cut values of the p_T of the seed comparing the effect of the impact parameter resolution with no smearing. The efficiency drops with increasing p_T cut value but is not effected by the impact parameter resolution. Without a cut on p_T of the seed the purity is almost zero. With p_T>30 GeV the purity can almost be recovered at little loss to the efficiency.
In the case of $e^+e^- \rightarrow \chi_1^+\chi_1^-$ the result is shown in Figure 9. Here the efficiency is also unaffected by the resolution but drops faster with increasing p_T. The Purity can not be fully recovered and remains constant once p_T reaches 20 GeV.
For the process $e^+e^- \rightarrow WW$ depicted in Figure 10 the situation is more challenging. The efficiency is again independent of the impact parameter resolution but drops rapidly when the p_T cut is increased. The loss of purity is severe and can not be recovered. The reason for this is the Standard Model background the process itself generates when the W decays leptonically. The electron and the muon from the W decay look like the ones from the τ decay thus generating many fake τ candidates.
The exact value of the impact parameter resolution has not much influence on the performance of the `TauFinder`. Efficiency and purity are stable when varying a between 2 μm and 6 μm and b between 6 μm GeV and 10 μm GeV.

3.2. Momentum, Energy Resolution and Background

Analog to the impact parameter resolution a Gaussian smearing is applied to the momentum of a particle and to its energy. The goal for the transverse momentum resolution at a Linear Collider

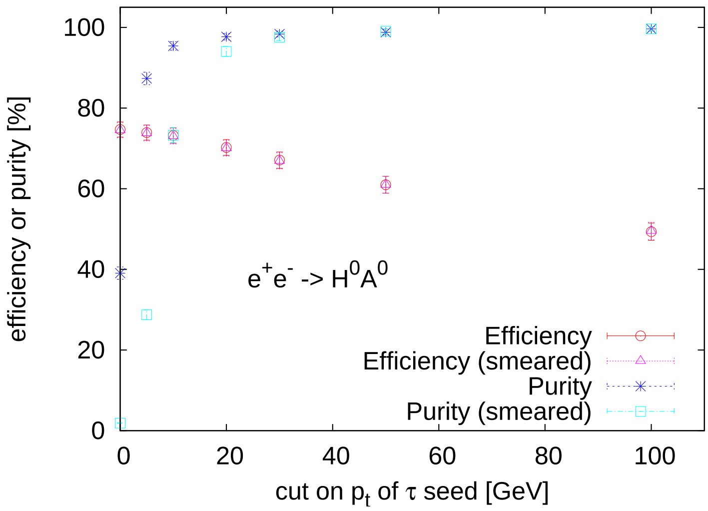

Figure 8: Performance of `TauFinder` without and with detector resolution in dependence on the cut on p_T of the seed for the process $e^+e^- \to H^0A^0$.

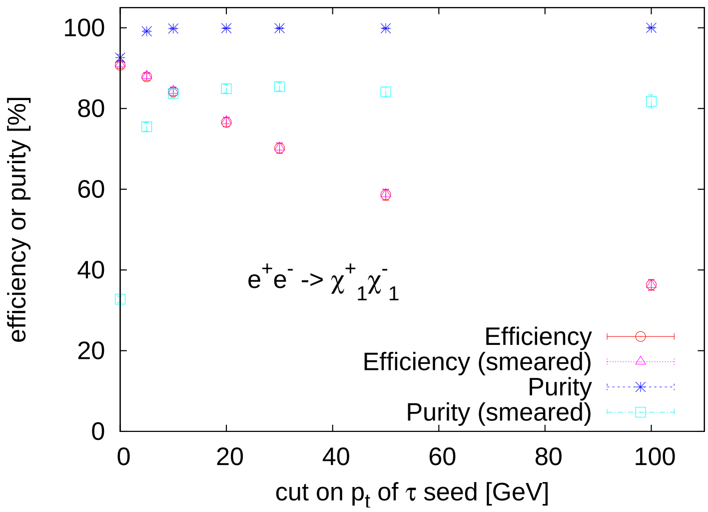

Figure 9: Performance of `TauFinder` without and with detector resolution in dependence on the cut on p_T of the seed for the process $e^+e^- \to \chi_1^+\chi_1^-$.

is $\sigma_{p_t} = 5 \cdot 10^{-5} p_t^2$. The energy resolution depends on the charge of the particle. If the particle is charged it is measured with the tracking system and the energy resolution is then given by the momentum resolution. Neglecting the mass of the particle the same value is used for both momentum and energy resolution smearing. For neutral particle the energy resolution is given

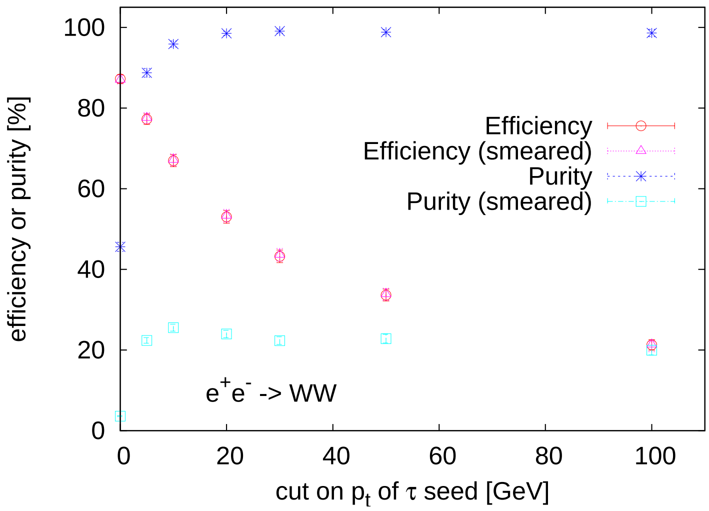

Figure 10: Performance of `TauFinder` without and with detector resolution in dependence on the cut on p_T of the seed for the process $e^+e^- \rightarrow$ WW.

by the calorimeter. In the worst case the neutral particle is measured in the hadronic calorimeter that is estimated to have an energy resolution of 60%. The effect of energy and momentum resolution on the `TauFinder` performance is negligible. The changes in efficiency and purity for a=30% or a=60% and a momentum resolution of $5 \cdot 10^{-5}$/GeV or $2 \cdot 10^{-4}$/GeV are within the error bars.

The introduction of background in combination with detector resolution has the same effect as for a perfect resolution.

4. Conclusion and Outlook

An algorithm to reconstruct τ leptons has been developed and evaluated based on MC information. The performance is very promising but depends strongly on the physics process. The main background arises from quark jets. The `TauFinder` has also been tested on reconstructed information obtained with PandoraPFA[4] but in this case the performance was not as good, due to problems in the reconstruction to correctly identify particles and assign the correct energy and charge. Once these issues have been improved and are more realistic in terms of performance of the reconstruction the algorithm will be evaluated based on the full detector simulation and reconstruction taking into account detector effects and reconstruction capabilities.

When available the information of a vertex for the τ jet can also be helpful to reject background and clean up the τ candidate. Furthermore a flight distance and therefore lifetime could be calculated possibly allowing to improve the distinction between jets from quarks and τs. Further improvements could be obtained when ckecking the τ candidate content and trying to reconstruct Π_0 from the photons in the neutral contribution.

A. User Manual

This is the more technical part, explaining how to set up and run the MARLIN processor. A working installation of the ILC software framework[5] containing MARLIN[6] and LCIO[7] is necessary to use `TauFinder`.

A.1. The Package

The source code can be obtained from the web[8] and consist of three MARLIN processors:

- `PrepareRECParticles`:
  Example how to prepare `Tracks` and `MCParticles` as input for the `TauFinder` by filling them into an `LCCollection` of `ReconstructedParticle`.
- `TauFinder`:
  The main part of the package performing the search for τs.
- `EvaluateTauFinder`:
  This is a processor to evaluate the performance of the `TauFinder` and also illustrates how to refer back to the MC truth or the objects combined into the τ.

A.2. Preparing the Input

`TauFinder` runs on an `LCCollection` containing the LCIO objects of type `Reconstructed-Particle`. In order to run the processor on Monte Carlo truth or just tracks or a combination of reconstructed objects a pre-processor called `PrepareRECParticles` has to be executed. By default it fills `Tracks` and `MCParticles` into a new collection of `ReconstructedParticles`. This processor can be extended to convert any object or combination of objects the user wants to run `TauFinder` on.
The following functions of `ReconstructedParticle` will be called in `TauFinder` and have to be set in the conversion in order to provide `TauFinder` with the necessary information:

- `getMomentum()`
- `getCharge()`
- `getEnergy()`
- `getTracks()`

Theses items are essential for the computation of the impact parameter and the angle between the seed and the particle. Charged particles need to have at least one track assigned to the `ReconstructedParticle` which is used to compute the impact parameter for the seed. In order to do that the function `getReferencePoint()` of the `Track` is used and has to return a point along the particle track. If the model to describe a track in LCIO changes and the reference point is no longer on the helix this part will have to change accordingly. In the current helix track model that is used to compute the impact parameter the vertex is assumed to be at the origin (x=0, y=0, z=0).
In addition a value for the magnetic field has to be supplied via the GEAR file. This is also needed for the computation of the impact parameter.

A.3. The Output

TauFinder will write a new collection with the τs as ReconstructedParticles. The processor EvaluateTauFinder gives an example on how to read it and find the link to tyhe MC truth.

A.4. Running the Processor

A complete example steering file to run TauFinder that first uses the processor PrepareRECParticles to provide the input for TauFinder based on Tracks and MCParticles will be available with the processor. Here, the main part to configure TauFinder to run on MCParticles is listed:

```
<processor name="MyPrepareRECParticles" type="PrepareRECParticles">
   <parameter name="outputColMC" value="MCParticles_tau"/>
</processor>

<processor name="MyTauFinder_MC" type="TauFinder">
   <parameter name="inputCol" value="MCParticles_tau"/>
   <parameter name="outputCol" value="TauRec_MC"/>
   <parameter name="pt_cut" value="1"/>
   <parameter name="D0seedmax" value="0.5"/>
   <parameter name="ptseed" value="5"/>
   <parameter name="searchConeAngle" value="0.07"/>
   <parameter name="isolationConeAngle" value="0.03"/>
   <parameter name="isolationEnergy" value="5.0"/>
</processor>
```

References

[1] H. Braun *et al.*, *CLIC 2008 Parameters*, CLIC-NOTE-764 (2008)

[2] M. Battaglia *et al.*, *Updated post - WMAP benchmarks for supersymmetry*, Eur.Phys.J. **C33** (2004), 273

[3] T. Ullrich, Z. Xu, *Treatment of Errors in Efficiency Calculations*, arXiv:physics/0701199v1

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

[5] Homepage: http://ilcsoft.desy.de/portal/software_packages/

[6] F. Gaede, J. Engels, *Marlin et al - A Software Framework for ILC detector R&D*, EUDET Report 2007-11

[7] F. Gaede *et al.*, *LCIO - A persistency framework for linear collider simulation studies*, CHEP03 March 2003, La Jolla, USA
Homepage: http://lcio.desy.de/

[8] https://twiki.cern.ch/twiki/bin/view/CLIC/Software