CERN - European Organization for Nuclear Research

Draft Version 1.0

Documentation and Manual for `TauFinder` implemented as a MARLIN processor

A. Muennich*

* *CERN, Switzerland*

May 27, 2010

Abstract

Within the ILC software framework a MARLIN processor to find and reconstruct τ leptons was developed. The algorithm targets τs that produce narrow, low multiplicity jets. However, it makes no assumption about the decay of the τ thus finding hadronic as well as leptonic decays. The algorithm delivers a reconstructed τ as seen by the detector which is not corrected for missing energy. This note provides an overview of the implemented algorithm, the cuts used and gives some evaluation of the performance. Chapter A is intended as a short user manual.

Contents

1. The TauFinder Algorithm **2**
1.1. Data Sets 3
1.2. The Selection Cuts 3

2. Evaluation of the Algorithm **6**
2.1. Evaluation Criteria 6
2.2. Influence of the Selection Cuts 6
2.3. Influence $\gamma\gamma$ background 8
2.4. Influence of τ decay channel 10

3. Conclusion and Outlook **10**

A. User Manual **12**
A.1. The Package 12
A.2. Preparing the Input 12
A.3. The Output 13
A.4. Running the Processor 13

1. The `TauFinder` Algorithm

The proposed algorithm resembles a jet finder cone algorithm 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 in $[\theta\phi]$.
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 is accepted is evaluated based on a few selection cuts discussed in section 1.2.

1.1. Data Sets

The data sets and their statistic used to evaluate the algorithm are listed in table 1. All processes were simulated at 3 TeV including initial state radiation and in the case of the SUSY processes also beam strahlung. The parameters for the SUSY processes are according to Benchmark Point K'[1]. 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
$e^+e^- \rightarrow W^+W^-$	672	154
$e^+e^- \rightarrow tt$	836	265
$e^+e^- \rightarrow \tilde{\tau}_1^+\tilde{\tau}_1^-$	5000	10000
$e^+e^- \rightarrow \chi_1^+\chi_1^-$	1000	1324
$e^+e^- \rightarrow H^0A^0$	762	541
$e^+e^- \rightarrow$ Full SUSY Spectrum	2000	1903
Total	10270	14187

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

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 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 six.
- The total number of charged and neutral particles combined to a τ has to be below 10.
- The impact parameter can not be exactly zero.

These cuts are based on studying the τ decay products based on Monte Carlo (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.

User Parameters

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

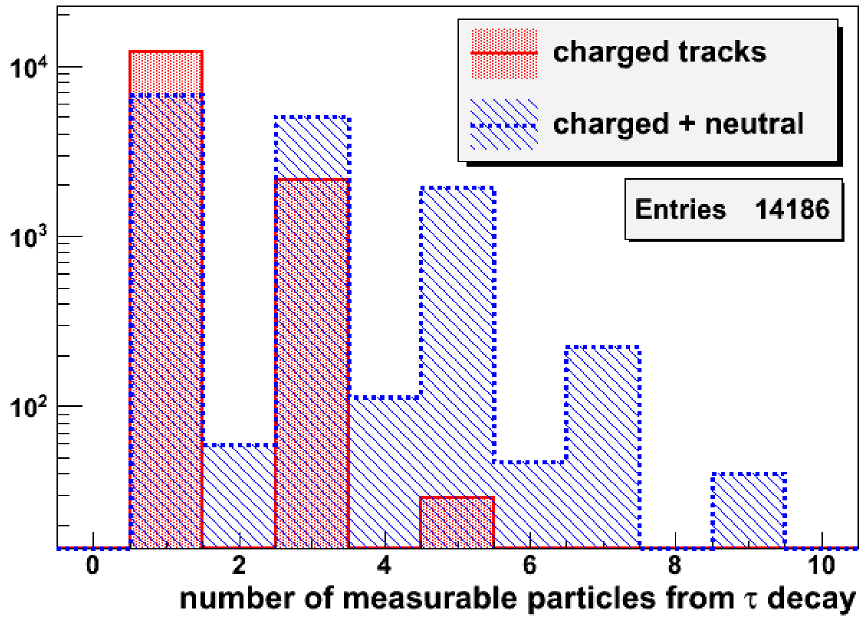

Figure 1: Number of charged tracks and the sum of charged and neutral particles in the detector from one τ decay.

- A general cut to suppress background by requiring a minimum transverse momentum for a particle to be considered in the algorithm ($p_T > 0$ GeV/c).
- A limit on the impact parameter D0 for the τ seed (D0 < 0.5 mm).
- A minimum transverse momentum for the τ seed ($p_T > 5$ GeV/c).
- A limit on the invariant mass of the τ candidate (< 2 GeV/c^2). This allows for some contamination, meaning particles within the τ candidate that in reality do not belong to it, which is likely to happen in the case of background.
- The opening angle of the search cone (0.05 rad).
- 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 sum of all particles that is allowed within the isolation cone (< 5 GeV).

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

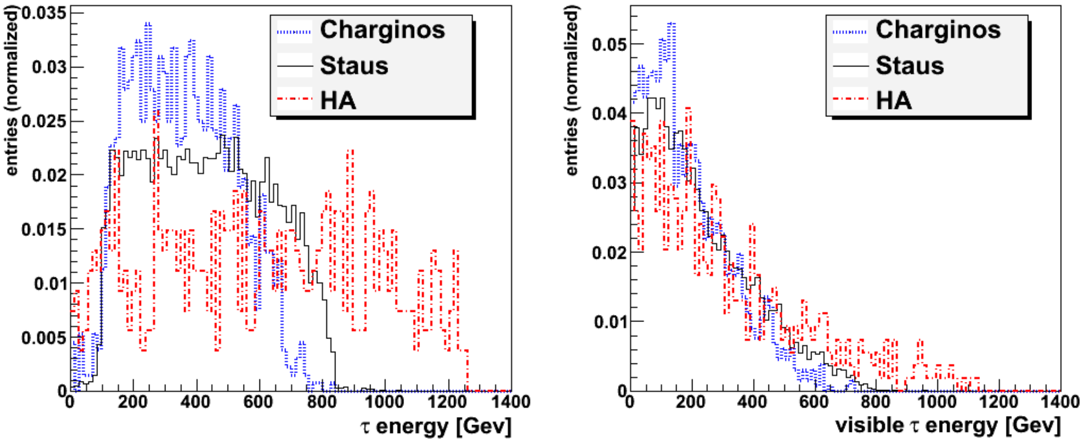

Figure 2: Full and visible energy of the τ based on Monte Carlo truth for three different processes from Table 1.

Figure 3 gives an example of the distribution of the impact parameter and the opening angle of the τ jet based on MC τs 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.

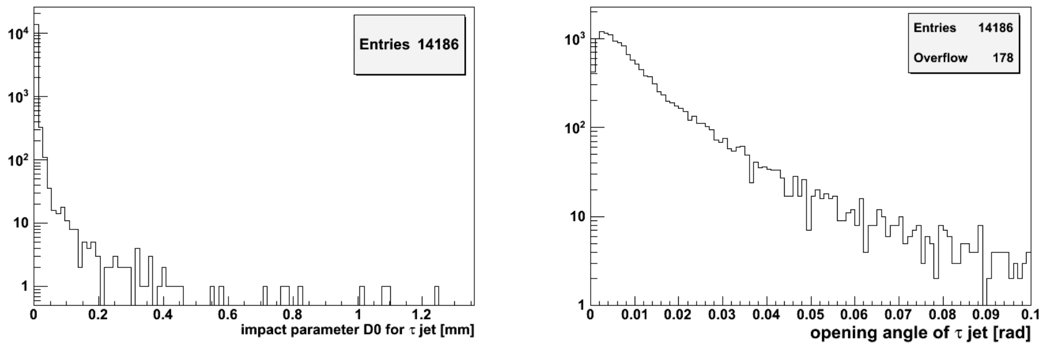

Figure 3: Impact parameter and opening angle of the τ jet based on MC truth from the processes listed in Table 1.

2. Evaluation of the Algorithm

2.1. Evaluation Criteria

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

- 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*.
- $\Delta E<1\%$: Number of reconstructed τs where the deviation of the reconstructed energy from the visible MC energy is less than 1%.

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*:

$$\begin{aligned}
\text{Efficiency}: \quad E &= \frac{Matched}{N_\tau} = 94.5\% \pm 0.5\% \\
\text{Purity}: \quad P &= \frac{Matched}{Reconstructed} = 97.3\% \pm 0.6\% \\
\text{Efficiency } \Delta\text{E} < 1\%: \quad E_E &= \frac{\Delta E < 1\%}{Matched} = 94.2\% \pm 0.7\%
\end{aligned}$$

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 [2].

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
- D0 $<$ [0.3, 0.5, 0.7] mm

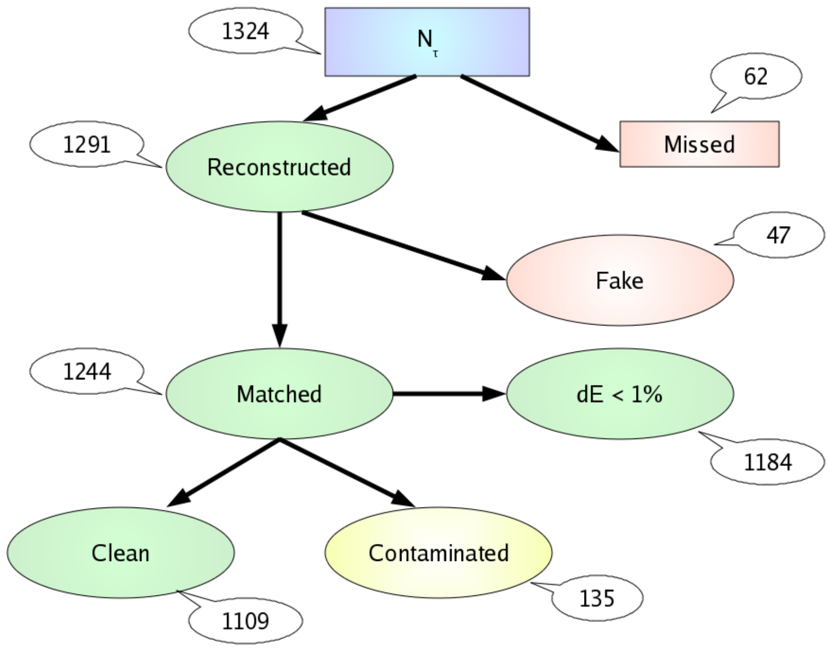

Figure 4: Illustration of nomenclature within a data sample. The numbers are an example of the different contributions when running `TauFinder` on a data sample of Charginos.

- Invariant mass (IM) $<$ [2, 10, 100] GeV/c^2
- 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:

1. $e^+e^- \to \tilde{\tau}_1^+\tilde{\tau}_1^-$
2. $e^+e^- \to \chi_1^+\chi_1^-$
3. $e^+e^- \to H^0A^0$

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.
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 cut on the invariant mass. The purity however is always 100%.

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 light quarks. The jets produced by light 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 light quarks are present which leads to a higher number of *Fake* τs. Hence the algorithm is less efficient.

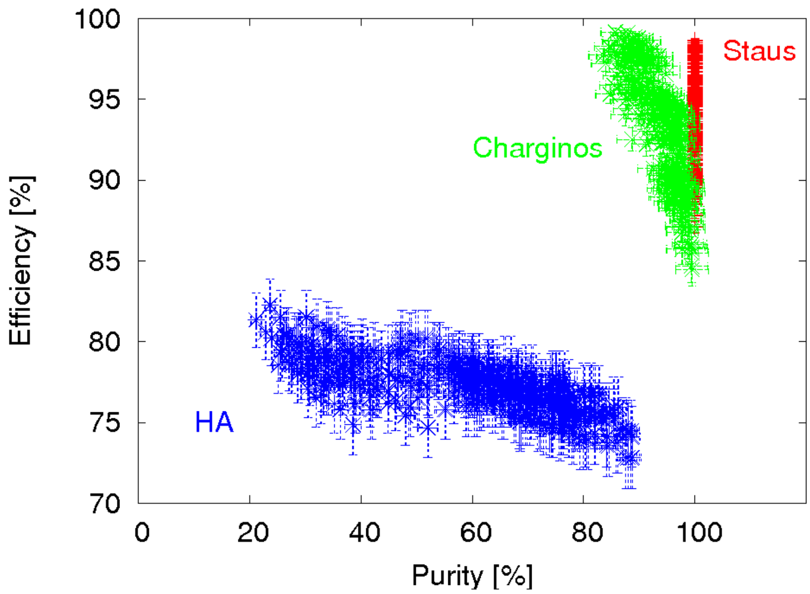

Figure 5: Performance of `TauFinder` for three different data sets without background for a parameter scan of selection cut values.

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

2.3. Influence $\gamma\gamma$ 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

Process	optim.	PT	PTS	D0	IM	SC	IC	IE	E	P
general	E	-	↓	↑	-	↑	↓	↑	↑	↓
	P	-	↑	↓	-	↓	↑	↓	↓	↑
$\tilde{\tau}_s$	E	0	0	0.7	10	0.05	0.02	10	98.7 ± 0.1	100
$\tilde{\chi}_{1s}$	E	0	0	0.7	2	0.07	0.02	10	99.2 ± 0.3	85.8 ± 0.9
	P	0	10	0.5	2	0.03	0.04	3	85.5 ± 1.0	99.5 ± 0.2
HA	E	0	0	0.7	2	0.05	0.02	10	82.3 ± 1.6	23.6 ± 1.0
	P	0	10	0.3	2	0.03	0.04	3	72.8 ± 1.9	88.7 ± 1.5

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.

cut parameter scan as in section 2.2. The influence of the overlaid $\gamma\gamma$ background is minimal as shown in Figure 6. In the case of no background there is a parameter space were higher purity can be achieved without sacrificing efficiency.

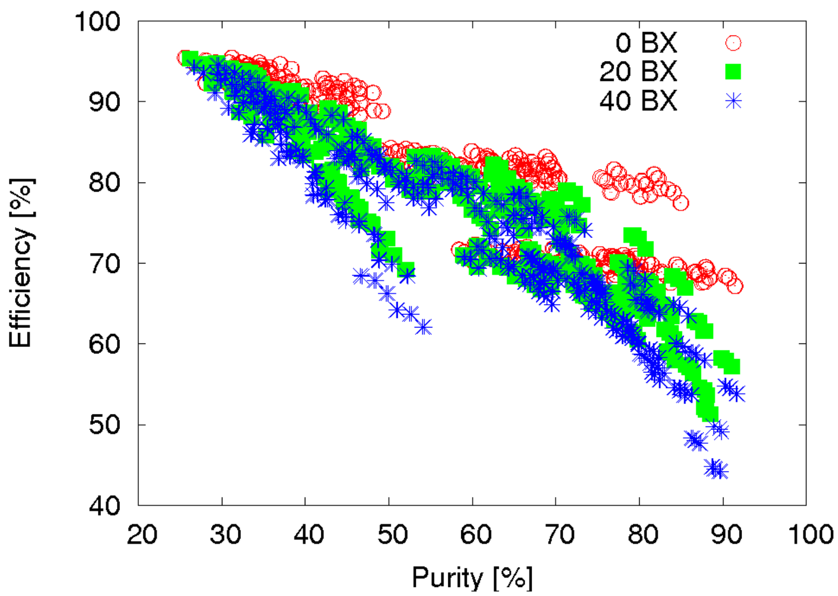

Figure 6: Performance of `TauFinder` for three different levels of $\gamma\gamma$ background for a parameter scan of selection cut values.

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.

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

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

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.

3. Conclusion and Outlook

So far the algorithm has been evaluated based on MC information. It has also been tested on reconstructed information obtained with PandoraPFA[3] 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 `TauFinder` will be evaluated based on the full detector simulation and reconstruction taking into account resolution 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 light quarks and τs.

Process	Decay	Eff. [%]	tot. Eff.[%]	Purity [%]
$e^+e^- \rightarrow W^+W^-$	$\tau \rightarrow \mu$	69.7 ± 7.7	81.2 ± 3.1	70.2 ± 3.4
	$\tau \rightarrow e$	83.3 ± 7.6		
	$\tau \rightarrow \pi$	85.4 ± 4.8		
	$\tau \rightarrow \pi\pi\pi$	90.9 ± 6.6		
	$\tau \rightarrow \pi\pi^0$	77.8 ± 12.9		
$e^+e^- \rightarrow tt$	$\tau \rightarrow \mu$	47.6 ± 7.4	56.6 ± 3.0	35.9 ± 2.3
	$\tau \rightarrow e$	66.0 ± 6.5		
	$\tau \rightarrow \pi$	60.0 ± 5.1		
	$\tau \rightarrow \pi\pi\pi$	64.9 ± 7.6		
	$\tau \rightarrow \pi\pi^0$	36.0 ± 9.1		
$e^+e^- \rightarrow \tilde{\tau}_1^+\tilde{\tau}_1^-$	$\tau \rightarrow \mu$	95.7 ± 0.5	95.4 ± 0.5	100
	$\tau \rightarrow e$	94.3 ± 0.5		
	$\tau \rightarrow \pi$	95.5 ± 0.3		
	$\tau \rightarrow \pi\pi\pi$	97.7 ± 0.4		
	$\tau \rightarrow \pi\pi^0$	92.8 ± 0.8		
$e^+e^- \rightarrow \chi_1^+\chi_1^-$	$\tau \rightarrow \mu$	94.9 ± 1.5	94.0 ± 0.7	96.4 ± 0.5
	$\tau \rightarrow e$	92.7 ± 1.7		
	$\tau \rightarrow \pi$	94.3 ± 1.0		
	$\tau \rightarrow \pi\pi\pi$	94.8 ± 1.9		
	$\tau \rightarrow \pi\pi^0$	91.7 ± 2.7		
$e^+e^- \rightarrow H^0A^0$	$\tau \rightarrow \mu$	83.6 ± 3.4	77.3 ± 1.8	66.2 ± 1.9
	$\tau \rightarrow e$	83.2 ± 3.6		
	$\tau \rightarrow \pi$	69.8 ± 3.4		
	$\tau \rightarrow \pi\pi\pi$	75.8 ± 5.2		
	$\tau \rightarrow \pi\pi^0$	79.6 ± 5.7		
$e^+e^- \rightarrow$ Full SUSY Spectrum	$\tau \rightarrow \mu$	91.9 ± 1.5	91.6 ± 0.6	91.4 ± 0.6
	$\tau \rightarrow e$	87.2 ± 1.8		
	$\tau \rightarrow \pi$	92.8 ± 1.0		
	$\tau \rightarrow \pi\pi\pi$	96.2 ± 1.3		
	$\tau \rightarrow \pi\pi^0$	86.7 ± 2.5		
All	$\tau \rightarrow \mu$	93.6 ± 0.5	93.5 ± 0.2	95.3 ± 0.2
	$\tau \rightarrow e$	92.4 ± 0.5		
	$\tau \rightarrow \pi$	93.7 ± 0.3		
	$\tau \rightarrow \pi\pi\pi$	95.9 ± 0.4		
	$\tau \rightarrow \pi\pi^0$	90.6 ± 0.7		

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

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[4] containing MARLIN[5] and LCIO[6] is necessary to use `TauFinder`.

A.1. The Package

The source code can be obtained from the web[7] 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 use the `LCRelations` 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. In addition it will provide LCRelations that allow to trace the τ back to the original input that was provided. The processor EvaluateTauFinder gives an example on how this is done.

A.4. Running the Processor

A complete example steering file to run TauFinder that first uses the processor PrepareREC-Particles 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="relCol" value="TauRecLink_MC"/>
   <parameter name="pt_cut" value="1"/>
   <parameter name="D0seed" value="0.5"/>
   <parameter name="ptseed" value="5"/>
   <parameter name="invariant_mass" value="2"/>
   <parameter name="searchConeAngle" value="0.07"/>
   <parameter name="isolationConeAngle" value="0.03"/>
   <parameter name="isolationEnergy" value="5.0"/>
</processor>
```

References

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

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

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

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

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

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

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