Università di Pisa

Facoltà di Scienze Matematiche, Fisiche e Naturali

Corso di Laurea in Fisica

Anno Accademico 2006/2007

Elaborato Finale

A NEW ALGORITHM FOR ELECTRON REJECTION IN TAU IDENTIFICATION FOR CDF EXPERIMENT AT THE FERMILAB TEVATRON

Relatore
Prof. Giorgio Bellettini

Correlatore
Dott. Fabrizio Scuri

Candidata
Chiara Debenedetti

29 Gennaio 2008

Contents

1 Introduction **1**

2 The Fermi National Laboratory and the CDF detector **4**
2.1 Fermilab 4
2.2 The Fermilab accelerator facilities 4
2.2.1 Proton source 4
2.2.2 Main Injector 5
2.2.3 Antiproton source 5
2.2.4 Recycler 6
2.2.5 Tevatron 6
2.3 The CDF-II detector 7
2.3.1 Tracking system 8
2.3.2 TOF (Time Of Flight detector) 10
2.3.3 Calorimeters 10
2.3.4 Muon detectors 11
2.3.5 Trigger 11

3 Phenomenology **13**
3.1 Tau lepton 13

4 Algorithms for electron and tau identification **16**
4.1 Electron identification 16
4.2 Tau identification 18
4.3 Main differences between the two identification algorithms . . 19
4.4 Electron rejection in tau identification: default method 20
4.5 Electron rejection by tau to electron matching method 22
4.5.1 Treatment of particular cases of misidentified electrons 22
4.5.2 Electrons not reconstructed as electromagnetic objects 23
4.6 The new function named "*tauelectron filter*" 23

5 Methods for procedure validation and preliminary results **25**
5.1 Data samples used in the evaluation of electron misidentification probability 25
5.2 Background composition studies 26

5.3 Preliminary Results 28
5.3.1 Monte Carlo 28
5.3.2 Data 28
5.3.3 Scale factors 33
5.3.4 Further efficiency studies 33

6 Conclusions **34**

Bibliography **36**

Chapter 1

Introduction

The work presented in this thesis is part of my training plan in the INFN-DoE[1] program "Summer Student 2007" at Fermilab, Batavia, Illiniois. The main activity to accomplish the training plan was a contribution to update the software code used to reconstruct and to identify leptons in the CDF experiment running at the Tevatron.

Event reconstruction described in this thesis proceeds in two main steps.

In the first step, CDF code applies a confirmation of trigger selection cuts providing lists of "objects" to be analyzed in the second which adopts suitable algorithms to identify particles, roughly reconstructed in the first step from detector information.

The argument focused here is the lepton identification starting from electromagnetic, muon, and tau objects; they correspond to three independent sets of selection criteria, described in paragraph 4 and applied in the temporal order listed above.

The work done for this thesis is part of a bigger project to improve lepton identification algorithms, aiming to increase both efficiency and purity. The general project was developed following three possible schemes, whose schematic view is shown in the figures below. Figure 1.1 refers to the standard procedure for electron-tau separation based on a cut on the electro-magnetic fraction as a function of the ratio E/p and related to the parameter ξ defined in the following sections.

Scheme in figure 1.2 shows a completely different approach where electron rejection is made by matching tau objects to electromagnetic ones.

Figure 1.3 refers to the third approach based on a new function making electron candidates, to be rejected, from tau candidates.

This thesis is organized as follows; in chapter 2 a short description of Fermilab and of the CDF detector is given; a short review of the Tau phenomenology to be studied in CDF is given in chapter 3; a more detailed description of the different algorithms proposed in the lepton identification

[1]Department of Energy of US

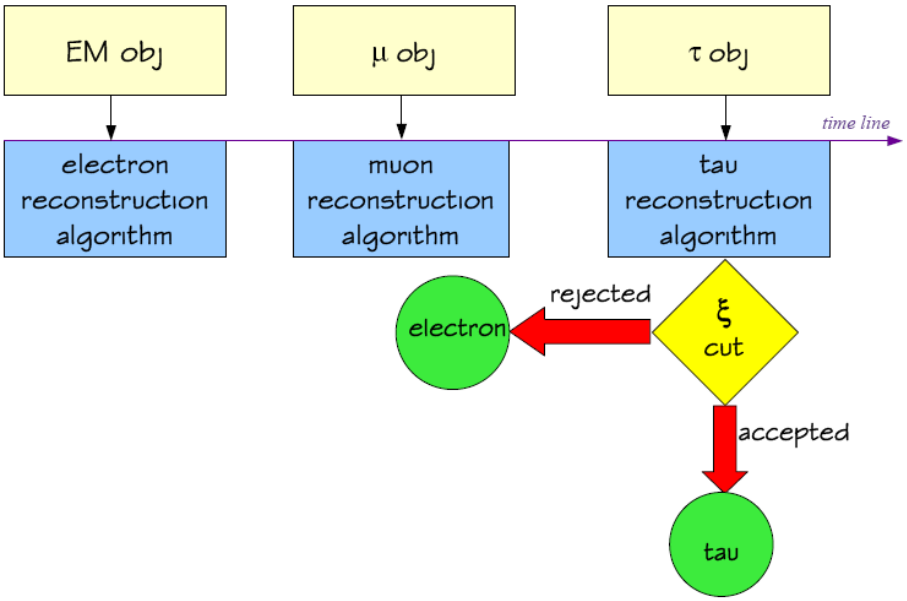

Figure 1.1: Standard electron rejection scheme, based on ξ parameter (see section 4.4)

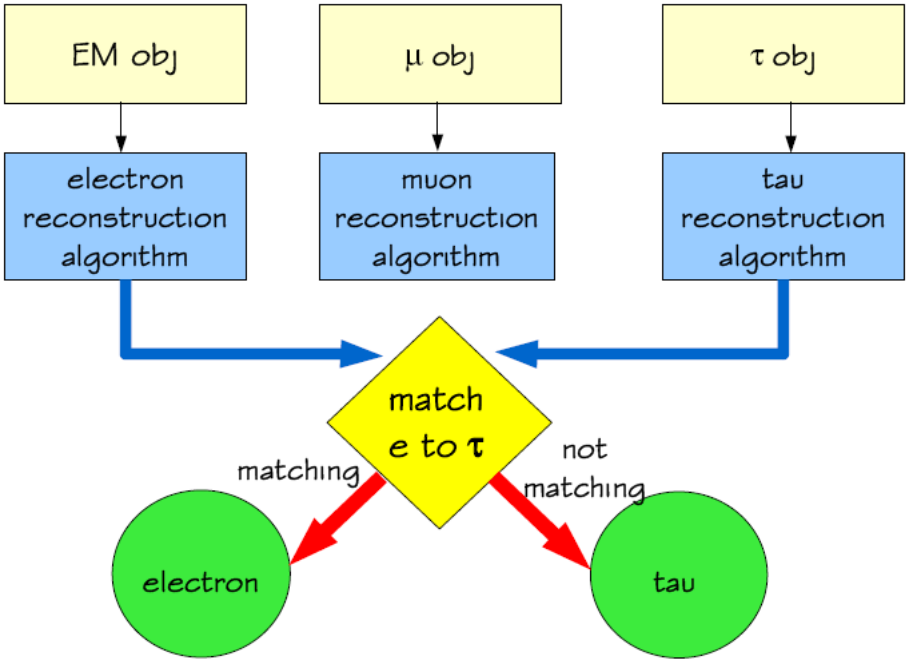

Figure 1.2: Scheme of electron rejection based on matching algorithm (see section 4.5)

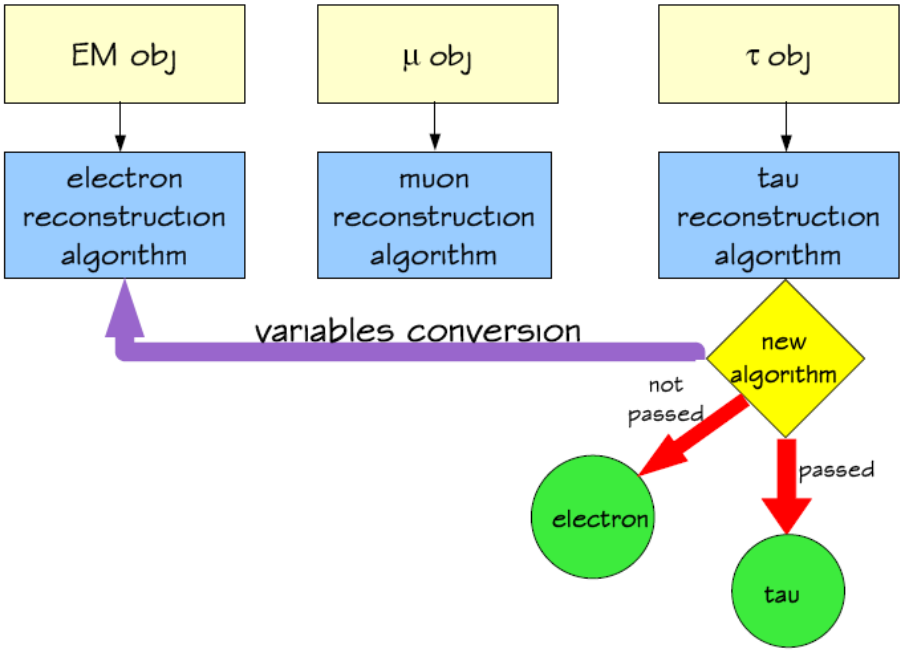

Figure 1.3: *Tauelectron filter* based electron rejection scheme (see section 4.6)

project is presented in chapter 4; method to validate the new procedure and preliminary results are presented in chapter 5; conclusions are summarized in chapter 6.

Chapter 2

The Fermi National Laboratory and the CDF detector

2.1 Fermilab

The work described in this thesis was done for the CDF experiment, which is taking data (winter 2007) on the Fermilab Tevatron collider. Fermilab is located about 60 km west of Chicago, in the area comprising the towns of Naperville, Batavia and Aurora. The Tevatron is a $p\bar{p}$ collider, with two running experiments, CDF and D0. These experiments took data until 1997 during the so-called Tevatron Run I. From 1997 to 2001 the Tevatron, CDF and D0 were upgraded for the Run II at higher collision centre of mass energy $\sqrt{s} = 1.96\ GeV$.

Particles colliding inside Tevatron are produced and pre-accelerated in a large accelerator complex, which is briefly described in the following. The layout of the Fermilab accelerator facilities is shown in figure 2.1

2.2 The Fermilab accelerator facilities

2.2.1 Proton source

Negative hydrogen ions are accelerated to 750 KeV in an electrostatic Cockroft-Walton accelerator. H^- ions are transferred to a linear accelerator and accelerated to 400 MeV. Ions are stripped through a thin Carbon target at linac exit and protons are injected into a booster synchrotron and accelerated to 8 GeV. 8 GeV protons bunches are sent to the Main Injector.

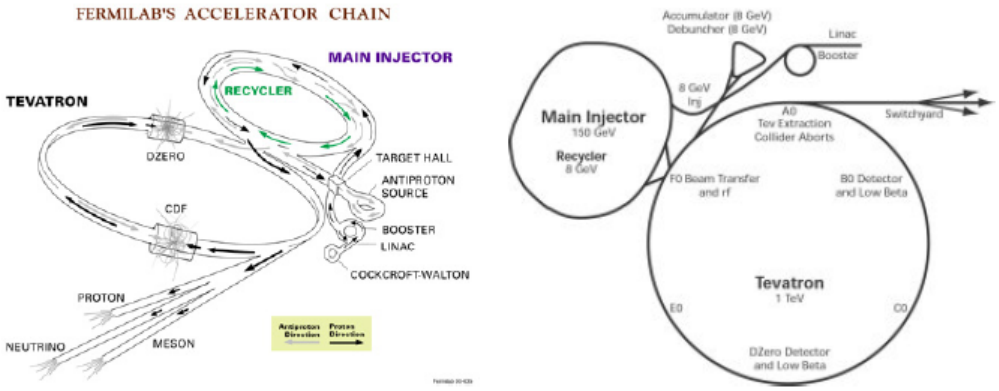

Figure 2.1: Layout of the accelerator chain at Fermilab Tevatron, from the source to the collision, courtesy of M. Giunta [5]

2.2.2 Main Injector

Placed in a new tunnel[1], the Main Injector is the major upgrade built for Run II. It is a synchrotron of 150 GeV maximum energy which serves a number of purposes:

- high intensity secondary hadron and neutrino beam production for fixed target experiments;
- antiproton production by a 120 GeV proton beam on a source target;
- acceleration from 8 to 150 GeV of protons and antiprotons to be injected into the Tevatron (Collider Mode);

2.2.3 Antiproton source

Once reached 120 GeV kinetic energy in the Main Injector the proton beam is extracted and directed onto a 7 cm Nikel target. About 20 antiprotons over 1 million interacting beam protons exit the target with about 8 GeV kinetic energy and a wide energy spread and angular aperture and enter a forward magnetic channel. A focusing and bending magnetic system separates antiprotons . Antiprotons produced in a single Main Injector burst are further stochastically[2] cooled inside the Debuncher to reduce their wide angle and energy spread while diluting their radiofrequency bunch structure, and sent to the Accumulator, a 8 GeV synchrotron. Here, subsequent

[1] in Run I final pre-acceleration and injection into the Tevatron was made by the so-called Main Ring hosted in the Tevatron tunnel

[2] stochastic cooling is a system by which beam fluctuations of beam emittance are sensed and damped by feedbacks along the orbit

source bursts are accumulated to an approximately continuous current of up to about 150 mA. When the Accumulator acceptance is saturated antiprotons are transferred into a larger acceptance constant field 8 GeV storage ring named Recycler, housed in the Main injector tunnel, where currents up to about 350 mA can be accumulated. By radiofrequency manipulations a large fraction of the Recycler beam is transferred into the Main Injector, where they are accelerated to 150 GeV for final injection into the Tevatron.

2.2.4 Recycler

Originally designed to recycle antiprotons left unused at the and of each Tevatron store, the Recycler is a constant field ring housed in the same tunnel with the Main Injector. Recycler operates as a second-stage accumulator allowing storage of the very large number of antiprotons as required for the Tevatron high luminosity program.

2.2.5 Tevatron

It is a 1 km radius, superconducting magnet, synchrotron for $p\bar{p}$ collisions. Protons and antiprotons move on opposite circular trajectories, being kept separate from each other by electrostatic separators all along the orbit except at two interaction regions. Beams are accelerated by radiofrequency cavities to 0.98 TeV in energy. Typical beam intensities are obtained with 36 bunches of 30×10^{10} protons and 36 of 3×10^{10} antiprotons. Beams are brought to collide at two points along the ring where they are squeezed to minimum transverse size for maximum luminosity by a focusing system with magnetic quadrupoles. The CDF and D0 detectors are located at the two interaction regions.

Tevatron characteristics

- Centre of mass energy
$$\sqrt{s} = 2E_p \tag{2.1}$$
where s is the squared invariant collision energy.

- Instantaneous luminosity
$$L = \frac{f \cdot N_B \cdot N_p N_{\bar{p}}}{2\pi(\sigma_p^2 + \sigma_{\bar{p}}^2)} \tag{2.2}$$
where f is the revolution frequency, N_B the number of bunches inside the ring (36), N_p ($N_{\bar{p}}$) is the number of protons (antiprotons) in each bunch, and $\sigma^2_{p(\bar{p})}$ is the p ($\bar{p}$) bunch width[3]. In stable operating conditions to maximize the integrated luminosity, Tevatron colliding beam

[3]in a gaussian approximation of the bunch density distributions in two orthogonal transverse directions, defined as follows: $\sigma^2 = \sigma_x^2 + \sigma_y^2$

runs are stopped when the luminosity has decreased to about 10% of the initial value, and setup operations of a new run are started. Runs may last up to 20 h, run setup may take a few hours.

- Interaction rate
$$\frac{dn_{int}(t)}{dt} = L\sigma \tag{2.3}$$
where σ is the total $p\bar{p}$ cross section, an increasing function of energy.

- The total number $n(T)$ of interactions in a period T is obtained by the integration of equation 2.3:
$$n(T) = \sigma \int_0^T L dt \tag{2.4}$$

The typical values for last year Tevatron operation are shown in figure 2.2.

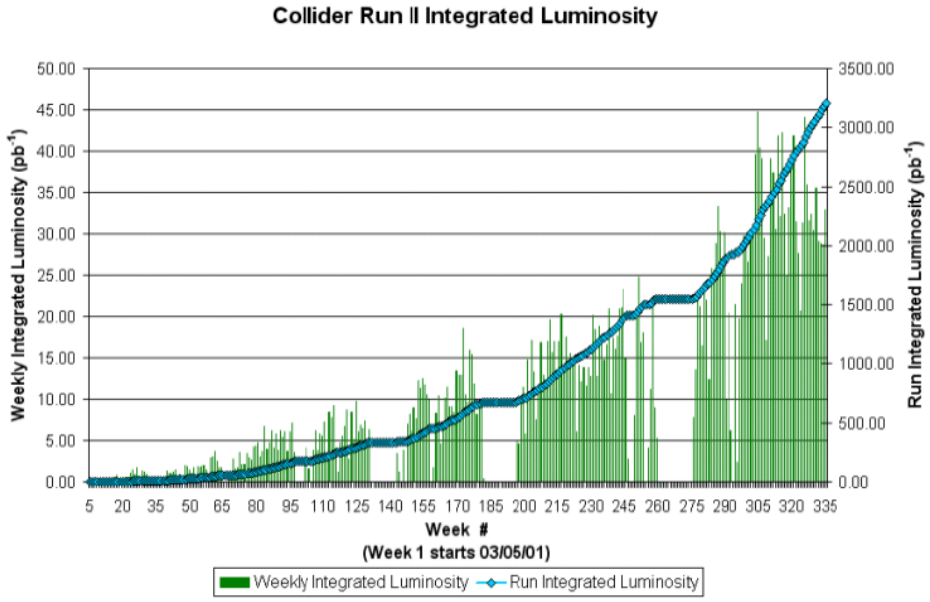

Figure 2.2: Weekly and integrated luminosity

2.3 The CDF-II detector

CDF is a multipurpose experiment. Highlights on its wide physics program are on Higgs boson searches, top quark mass and decay modes measurements, heavy quark physics, high precision electroweak parameter measurements;

to accomplish this ambitious program, CDF detector must have an excellent tracking system, high precision calorimeters, a good lepton and hadron identification system. A more detailed description of the detector can be found in ref [1]. A short description of the main components is given here.

The detector has cylindrical symmetry (figure 2.3) : the origin of coor-

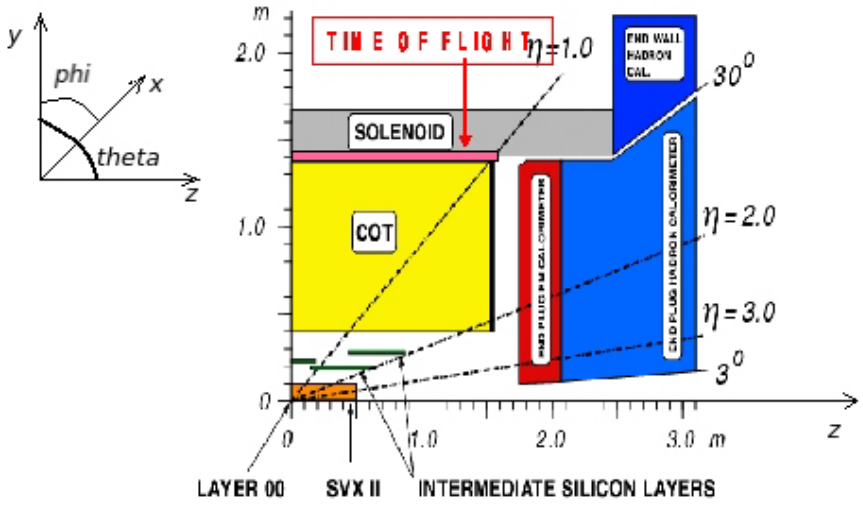

Figure 2.3: Detector coordinate system

dinates system is the collision point (primary vertex) and z axis along the proton beam direction. In addition to the polar angle θ, the polar direction is described using also the variable η (pseudorapidity), defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.5}$$

$\Delta\eta$ intervals are invariant for z direction boosts. Azimuthal angle φ is defined in the plane orthogonal to the z axis.

The detector base structure, from the inner to the outer part, is the following: tracking system, Time-of-Flight detectors, solenoid, calorimeters, muon detectors. Electronics for trigger, slow control and data acquisition systems complete the detector components.

2.3.1 Tracking system

The volume of the tracking system is roughly a cylinder 2.8 m in diameter and 3.5 m in length.

Charged particles, moving in the 1.4 T axial magnetic field produced by a superconducting solenoid, ionize the tracker sensitive layers and generate

signals associated to given positions and named 'hits': helix segments are fitted to the hits and the track kinematical parameters extracted.

The tracking system consists of different sub-detectors: up to 8 layers of silicon detectors (L00, SVX II, ISL), and an outer drift chamber (COT), covering different pseudorapidity regions.

A more detailed view of the tracking system is given in figure 2.4.

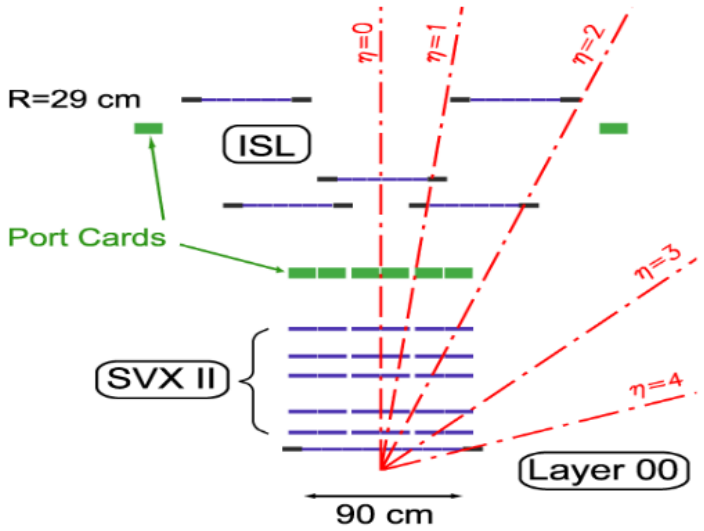

Figure 2.4: Tracking system scheme

The silicon detector has three main components:

Layer 00 (L00)

It is the innermost detector, directly surrounding the beam pipe for optimum resolution on charged tracks impact parameter[4]. It consists of single side silicon strip detectors, 6 μm resolution.

SVX II (Silicon VerteX detector)

Made of 5 layers of double side silicon detectors, it measures (r,φ) plane coordinates[5], and z coordinates[6]. Its pseudorapidity coverage is: $|\eta| < 2$. It is a key detector for primary and secondary vertex reconstruction.

[4]minimum distance from beam axis of extrapolated track in the transverse plane

[5]axial strips; 12 μm resolution

[6]stereo strips, either wide stripes perpendicular to z or standard thin stripes rotated by 1.2^o with respect to the axial ones, providing a z-resolution of about 70 μm

ISL (Intermediate Silicon Layer)

It is the outer silicon detector, just before the COT. The ISL central region ($|\eta| < 1$) has one layer, helping in track reconstruction. The outer region ($1 < |\eta| < 2$) has two layers, covering a pseudorapidity region where COT has poor acceptance. Both parts have axial and stereo (1.2^o rotated) strips.

COT (Central Outer Tracker)

It is a cylindric drift chamber surrounding the silicon tracker and filled with a gas mixture (Ar, Ethane, CF_4). COT is divided into 8 superlayers (with 12 sense wires each), 4 axial and 4 stereo (forming a $\pm 2^o$ angle with respect to z axis). Each superlayer is divided azimuthally into super cells (with 35^o inclination with respect to radial direction), where sense wires are alternated to field wires. Ionization electrons drift towards sensitive wires along directions perpendicular to the radius under the combined action of the electric field in the cell and of the Lorentz force.

COT acceptance relative to the nominal interaction point is $|\eta| < 1$. The luminous region is approximately Gaussian with a standard half-width of about 15 cm. The chamber expands to $|z| < 155$ cm; hit spatial resolution is 150 μm. Specific energy loss (dE/dx) and momentum are measured contemporarily in the fitted trajectories. Momentum resolution for fully contained tracks ($p_t > 400$ MeV/c, $|\eta| < 1$) is $\frac{\Delta p}{p} \approx \frac{0.01\%}{p^2}$ (p in GeV/c).

2.3.2 TOF (Time Of Flight detector)

A layer of 4 m long, 2 cm thick trapezoidal bars scintillator tiles surrounding the external COT wall; it is able to separate protons, kaons, pions up to about 1.5 GeV/c in momentum. Time resolution is about 130 ps. Fair particle identification is possible in CDF in this momentum range by combining TOF and dE/dx information.

2.3.3 Calorimeters

The calorimeter system absorbs photons, electrons and hadrons and measures their energy.

CDF calorimeters have a front electromagnetic compartment and a rear hadronic compartment, both with tower structure pointing to the interaction region; calorimeters and are divided into 3 different regions, depending on their η acceptances: central (CEM, CHA), wall (WHA) and plug (PEM, PHA).

Central calorimeters have full azimuthal coverage and $|\eta| < 1.1$ for CEM, $|\eta| < 1.3$ for CHA. Tower angular segmentation is defined by: $\Delta\eta = 0.1$, $\Delta\varphi = 15^o$.

Plug calorimeters cover the forward region up to $|\eta| < 3.6$ with a projective tower structure of approximately the same φ and η widths[7].

EM calorimeter towers are made of Pb foils (absorber layers) alternated to scintillator plates (sampling layers) read-out with phototubes. Inside the EM compartments, additional detector layers are placed at the depths where the shower is expected to be maximum ($6X_0$[8]). At large angles, CES[9] is a gas chamber with about 0.2 mm resolution, while PES[10] is made of scintillators with 1 mm resolution. Before entering CEM[11], particles pass through chambers called Central Preshower, which provide information on em particles showering early inside the solenoid.

Hadron calorimeters are made of sandwiched iron and scintillator layers seen by phototubes. They are about 4.7 interaction length thick and provide an energy resolution of about $\frac{75\%}{\sqrt{E_t}} \oplus 3\%$ (E_t in GeV).

2.3.4 Muon detectors

They are sandwiched plastic scintillator and proportional chamber layers in the external part of the detector, covering a region $|\eta| < 2$. There are four different muons detection systems, named: CMU[12], CMP[13], CMX[14], IMU[15]. Additional material is placed before some muon chambers in order to reject neutral particles, such as π^0's, which could have reached the external zone of the detector.

2.3.5 Trigger

Tevatron bunch crossing frequency is 2.5 MHz (one crossing every 396 ns). Since the data acquisition rate is limited to 100 Hz at 20 MB/s throughput, an efficient online event filter (trigger) is necessary to select events of interest having typical cross section much lower than hadronic cross sections[16].

The CDF trigger is a three level system, briefly described in the following.

Level 1

Its activation time is 5.5 μs (from collision instant); it analyzes information from detectors with faster response, from calorimeters, muon chambers, and

[7] a coarser segmentation in often chosen online with a tower clusterization depending on physics to be studied

[8] where X_0 is the radiation length for electron in the material of the EM calorimeters

[9] Central Electromagnetic Shower maximum detector

[10] Plug Electromagnetic Shower maximum detector

[11] energy resolution: $\frac{13.5\%}{\sqrt{E_t}} \oplus 1.7\%$

[12] Central MUon chambers

[13] Central Muon uPgrade

[14] Central Muon eXtension

[15] Intermediate MUon extensions

[16] for instance 0.1 mb for $p\bar{p} \to b\bar{b}$ processes against 100 mb for $p\bar{p} \to q\bar{q}$

COT. Level 1 decisions are based on some acceptance criteria to make fast track reconstruction and extrapolation. Events matching selection criteria according to different trigger paths are passed to level 2.

Level 2

It takes fast decisions within 20 μs. Secondary vertices candidates are reconstructed from the track impact parameter computed from the Secondary Vertex Tagger (SVT), a smart system, based on specialized electronics, which uses SVX information to match silicon detector track elements to L1 track candidates. CDF capability to detect displaced vertices and to make good heavy flavor physics is mainly due to the excellent performances of SVT (40 μm resolution on the primary vertex in the transverse plane).

Level 3

Based on an online processor farm, it completes event selection, with algorithms using all the information from the detector; software used at L3 has the same structure of the one performing offline reconstruction. L3 output rate is about 75 Hz.

Chapter 3

Phenomenology

A short overview of the tau physics that can be studied at CDF is given in the following.

3.1 Tau lepton

Tau is a pointlike particle of the third generation of leptons, counterpart of electron and muon. It is a fermion (intrinsic spin $\frac{1}{2}$) and has the same electric charge of the electron. It is a heavy particle, with a mass of 1776.99 ± 0.29 MeV/c^2, and a mean life of $2.90 \times 10^{-13}s$. It interacts both weakly and electromagnetically. Because of its short mean life, it is not possible to observe it directly, but it has to be reconstructed from its decay products. Tau is the only lepton which can decay into hadrons, because of its relative large mass. Weak decay implies the conservation of leptonic number, i.e. the presence of a tau neutrino in the final state[1]. Most frequent tau decays are:

$\tau^- \to e^- \bar{\nu}_e \nu_\tau$
$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$
$\tau^- \to \pi^- \nu_\tau$
$\tau^- \to \pi^- \pi^+ \pi^- \nu_\tau$
$\tau^- \to \pi^- \pi^0 \pi^0 \nu_\tau$

Tau identification in a multipurpose experiment like CDF Run II is very important to deepen studies of the Standard Model parameters and, at the same time, possibly, to measure processes to be included in the new physics framework. The Standard Model is a theoretical description of fundamental interactions, with except of gravity, using quantum field theory, consistently with quantum mechanics and special relativity. It essentially consists in a grouping of quantum electroweak and chromodynamics. It describes particles using the mathematical framework of quantum fields and it can be

[1] in case of antitau decay, in the final state leptonic number conservation requires the presence of a tau antineutrino

divided into three main parts: matter particles, force mediating particles, Higgs boson.

Inside the Standard Model, taus identification is relevant to study some decay modes of intermediate bosons ($W^{\pm}$, Z^0). The study of these modes improve the knowledge on electroweak parameters. $W^{\pm}$ and Z^0, mediating together with the photon (γ) the electroweak force, can decay weakly in the following leptonic final states:

$W^+ \rightarrow l^+ \nu_l$

$W^- \rightarrow l^- \bar{\nu}_l$

$Z^0 \rightarrow l^+ l^-$

where l can be e, μ or τ.

Z^0 decay modes into electron and tau pairs are widely studied in electron misidentification analysis (section 5.1). Monte Carlo samples of the simulated "clean" mode $Z^0 \rightarrow e^+ e^-$, with a background involving hadronic jets, produced in association with Z^0 boson in $p\bar{p}$ interactions, and Drell-Yan processes are used for that purpose. The Drell-Yan processes are typical of hadronic collisions and consist in the annihilation of a quark of the hadron and an antiquark of the antihadron, with the production of a pair of leptons ($l^+ l^-$).

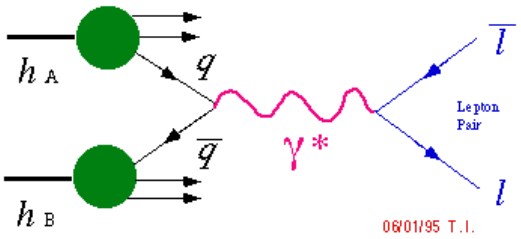

Figure 3.1: Feynman diagram of Drell-Yan process

This interaction can be mediated by either a virtual photon or a Z boson.

CDF accuracy in secondary vertex reconstruction and the good resolution of the tracking system and of the hadron calorimeter permits a more detailed study of τ hadronic decays[2], which branching ratios are: $(10.90 \pm 0.07)\%$ for one-pion mode, $(9.25 \pm 0.12)\%$ and $(9.33 \pm 0.8)\%$ respectively for tau decaying into one charged and two neutral pions and three charged pions.

Taus are also very important in Higgs boson studies, which is a not observed massive scalar particle whose intrinsic spin value is 0, according to the Standard Model prediction. It is supposed to be the key for the theoretical quantum field description, because it would give an explanation

[2] section 3.1

to high mass differences between elementary particles. Tau lepton can be a "signature" for eventual Higgs discovery because:

- it is one of the possible final states of the boson decay: $H \to \tau^+\tau^-$

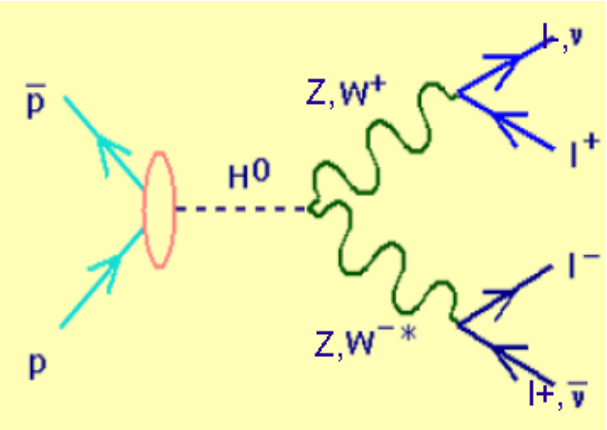

Figure 3.2: Feynman diagram of Higgs decay with leptonic final states

- it is also present in the decay of the intermediate boson which can be produced in association with Higgs in $p\bar{p}$ events.

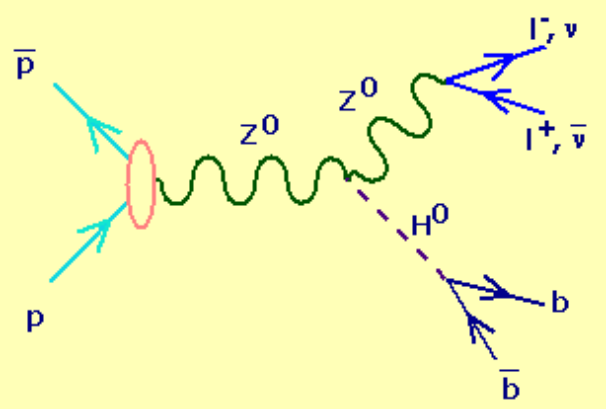

Figure 3.3: Feynman diagram of associated production of a Higgs boson decaying in a quark-antiquark pair and a Z^0 boson decaying in a lepton anti-lepton pair

Chapter 4

Algorithms for electron and tau identification

In CDF Run II, offline reconstruction code runs different algorithms for leptons identification. Electron and tau identification algorithms are discussed in this thesis. The whole procedure for reconstruction and identification consists of several steps. All data collected with the CDF detector, selected by the trigger, and grouped in data streams according to the specific trigger paths are processed by the algorithms for electromagnetic and tau "object" reconstruction. Objects are not real particles, rather they are a collection of track characteristics to define candidates further processed by the identification algorithms. Central electrons are generally reconstructed by both electron and tau reconstruction algorithms. This results in a higher efficiency and lower purity for tau reconstruction with respect to the electron one. The work discussed here is dealing directly with this evidence.

4.1 Electron identification

The electron identification algorithm begins considering an energy deposit over threshold inside the calorimeter, having or not an association to a track in the list produced by the default tracking algorithm. In case of no association, a particular algorithm is used to recover tracks which can be associated to the energy deposit. At this point the real electron identification begins.

Electrons are required to be central[1] ($|\eta| < 1$) and to satisfy all cuts listed in table 4.1; additonal cuts are applied on the transverse energy E_t , defined as follows:

$$E_t = E \sin\theta \tag{4.1}$$

and on the ratio R:

$$R = \frac{E_{CHA}}{E_{CEM}} \tag{4.2}$$

[1] pseudorapidity η is defined in section 2.3

central electron	$\eta < 1$
transverse energy	$E_t > 20\ GeV$
energy fraction	$R < 0.05$
relative isolation	$Iso_{rel} < 0.1$
beam constrained transverse momentum	$P_t^{bc} > 10\ GeV/c$
number of hits on axial wires	$n_{hit}^{ax} > 3$
number of hits on stereo wires	$n_{hit}^{st} > 3$
ratio between energy and momentum	$0.5 \leq \frac{E}{p(GeV)} \leq 2$
primary vertex z coordinate	$\lvert z_{pv}\rvert < 60\ cm$
CES fiducial along z	$9\ cm \leq z_{track} \leq 230\ cm$
CES fiducial along x	$\lvert x_{track}\rvert < 21.5\ cm$

Table 4.1: Cut values in electron identification algorithm

where E_{CHA} and E_{CEM} are the fractions of the total energy deposited respectively in the hadron and in the electromagnetic calorimeter. Isolation[2] is also required for electrons. A minimum χ^2 test is used to reject as photons e.m. objects associated to tracks whose parameters are found to be not compatible with an electron.

At this point, the algorithm makes a more refined analysis of the information from the COT, using the following variables:

- beam constrained transverse momentum (p_t), defined as follows:

$$p_t = p \sin\theta \tag{4.4}$$

  where p is the module of momentum $\mathbf{p}$

- number of hits in stereo and axial wires. This information is used to remove cosmic rays, having less hits inside the chamber, with respect to a particle coming from the interaction vertex

- the ratio between track energy and momentum; it must be $\frac{E}{p} = 1$ (natural units, $c = 1$) for electrons.

The primary vertex is required to be in the collision region.

Other variables from CES (subsection 2.3.3) are analysed to check the compatibility between track parameters and energy deposit.

[2] relative isolation is defined as

$$Iso_{rel} = \frac{\sum^{annular} E_t}{\sum^{isocone} E_t} \tag{4.3}$$

where isolation annular and cone are defined in section 4.2

4.2 Tau identification

Identification algorithm requires taus to be in the central region ($|\eta| < 1$). Additional cuts are applied on the following variables:

- total "visible" transverse energy associated to any cluster of calorimeter cluster over threshold, defined as follows:

$$E_t^{vis} = \frac{p_t \cdot E_{tot}}{p_{tot}} \tag{4.5}$$

  where p_t is the transverse momentum of the cluster, defined in equation 4.4 of section 4.1, E and p are respectively total energy measured by the tower cluster and total momentum ;

- seed tower transverse energy (E_t^{st}) the tower with the highest fraction of energy in the cluster ;

- seed track transverse momentum (p_t^{st}).

All applied cuts are listed in table 4.2. The further steps are the π^0's removal and the τ isolation analysis. π_0's are removed applying a cut on the number of tau objects tracks inside the so-called *isolation annular*, consisting in a zone delimited by two cones whose apertures, defined with respect to the track direction, are $\alpha_{sign} = 10^o$ (defining the "signal" cone) and $\alpha_{iso} = 30^o$ (defining the "isolation" cone). Taus are supposed to be inside the signal cone; therefore, the tracks inside the annular are removed because they are π^0's faking taus. Isolation requirement is the same of the electron identification code. A procedure similar to that described for the electrons uses COT information in a χ^2 test for best track association to tau. Cosmic rays background is rejected by requiring impact parameter (d_0) to be small. Fiducial cuts on CES variables are also applied to electromagnetic shower clusters. Finally, cuts are applied by the identification algorithm on different mass variables, defined as follows:

1. calorimeter mass $m_{cal} = M$, where M is computed from the cluster four momentum: $P = (E, \mathbf{p})$, $M = (E^2_{cluster} - \mathbf{p}^2_{cluster})^{1/2}$

2. seed track mass m_{st}; same definition of visible mass, but considering seed tower and track four-momentum;

3. visible mass m_{vis}, defined as:

$$m_{vis} = M_{st} + M_{cone} \tag{4.6}$$

   where with M_{cone} is the mass calculated from the total four-momentum associated to the π^0's eventually present in the anular described above.

The algorithm is completed with the part relative to the electron rejection, which will be discussed more in detail in the following sections. In this it is useful to remind the main differences between the two described algorithms.

central tau	$\eta < 1$
visible transverse energy	$E_t^{vis} > 20\ GeV$
seed tower transverse energy	$E_t^{st} > 10\ GeV$
seed track transverse momentum	$p_t^{st} > 4.5\ GeV/c$
number of tracks inside isolation anular	$n_{tr}^{an} < 0$
relative isolation	$iso_{rel} < 0.1$
number of hits on axial wires	$n_{hit}^{ax} > 3$
number of hits on stereo wires	$n_{hit}^{st} > 3$
COT fiducial along z axis	$\|z_{COT}\| < 155\ cm$
vertex z coordinate in COT	$z_{pv} < 60$
CES fiducial along z	$9\ cm \leq z_{track} \leq 230\ cm$
CES fiducial along x	$\|x_{track}\| < 21.5\ cm$
calorimeter mass	$m_{cal} < 4\ GeV/c^2$
seed track mass	$m_{st} < 2\ GeV/c^2$
visible mass	$m_{vis} < 1.8\ GeV/c^2$

Table 4.2: Cut values in tau identification algorithm

4.3 Main differences between the two identification algorithms

The most striking effect of the difference between the two identification algorithms is that tau reconstruction recovers some cases of true electrons discarded by the electron one and, therefore, it can also be interpreted as an improvement for electron reconstruction code.

A limit of the electron reconstruction algorithm is the fact that it works on only one calorimeter wedge at a time. This way, the reconstruction process of electrons radiating high energy photons or entering the electromagnetic calorimeter close to a tower border splits their energy across different wedges and identifies two different objects which can be either both rejected because of their low energy or misreconstructed with wrong energy values, because only the fraction deposited inside one wedge is taken into account. These cases are recovered by additional special algorithms, merging electromagnetic objects depositing their energy in two adjacent wedges. The limit of this procedure is the fact that it can possibly result in some misreconstruction problems for the electron to be merged or due to the loss of low energy electromagnetic objects spreading their energy deposit in different wedges, all under threshold.

Standard electron reconstruction is not really efficient also in case of electrons with retarded shower. Sometimes, electromagnetic shower happens to begin later than the expected time, so a large fraction of energy is deposited inside the hadron calorimeter. This results in a high value of the

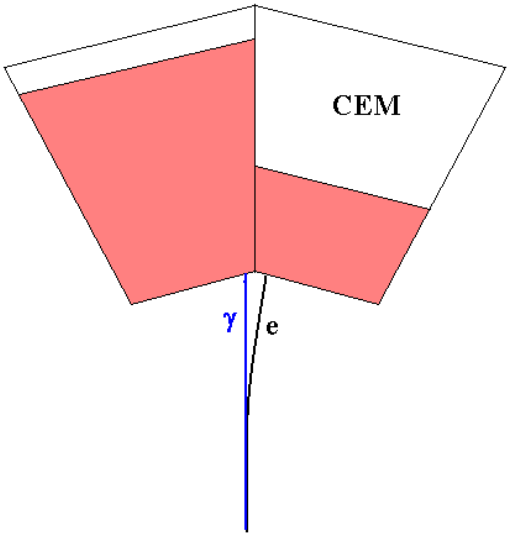

Figure 4.1: Example of "split" electrons

ratio $R = \frac{E_{had}}{E_{cal}}$ (equation 4.2).

On the other hand, tau reconstruction algorithm was originally designed to work across different wedges at the same time in order to reconstruct objects coming from τ decays, such as jets, which can spread their energy across different wedges.

4.4 Electron rejection in tau identification: default method

Standard procedure in CDF for electron rejection from τ's is based on the so called "ξ-cut" method. The ξ parameter is defined as follows:

$$\xi = \frac{E}{p}(1 - f_{em}) \tag{4.7}$$

where f_{em}, the object electromagnetic fraction:

$$f_{em} = \frac{E_{em}}{E_{tot}}. \tag{4.8}$$

E_{em} is the fraction of energy deposited inside the electromagnetic calorimeter. The curve in the plane f_{em} vs E/p of figure 4.2 is obtained with the value $\xi = 0.2$ in equation 4.7. The requirement for taus to lie below the curve in figure 4.2 introduces a very low probability for electrons to be misidentified as taus ($P_{\xi}^{misid}(e) \simeq 0.3\%$), but a high inefficiency in tau identification ($I_{\xi}^{id}(\tau) \simeq 10\%$).

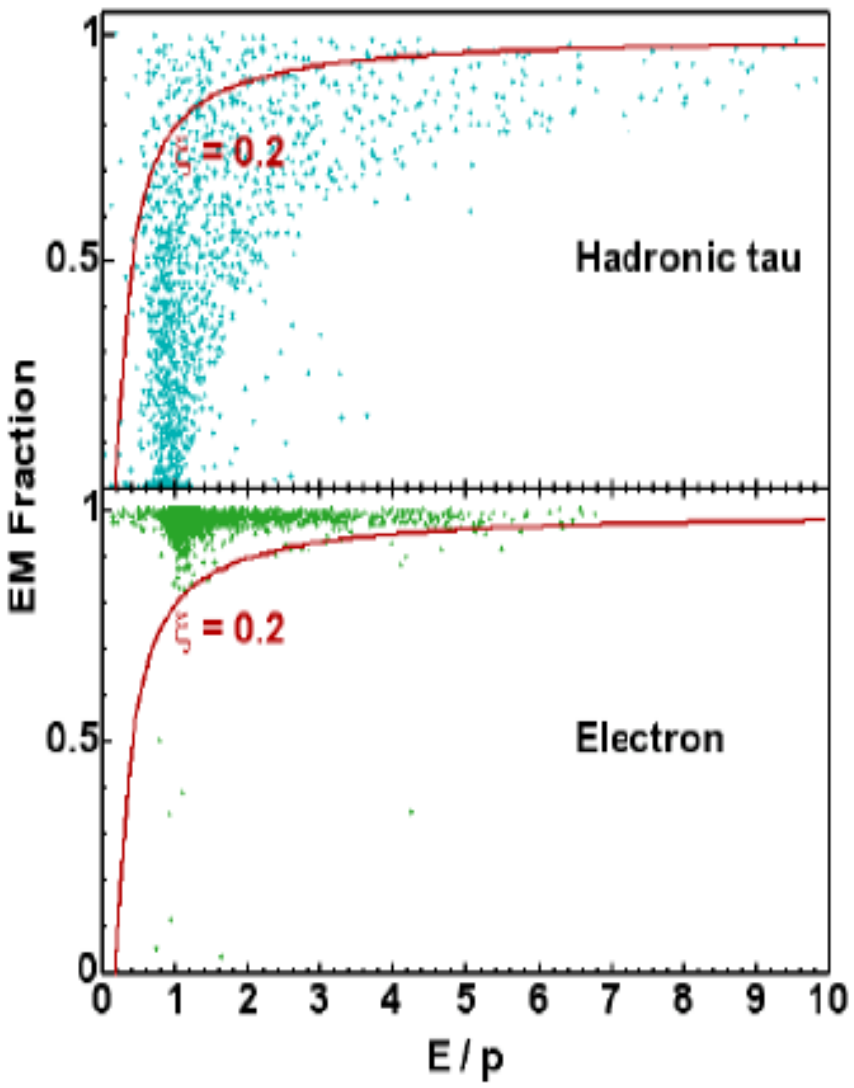

Figure 4.2: Distributions of taus (up) and electrons (down) in ($\frac{E}{p}$,f_{em}) plane. In red the curve $f_{em}(\frac{E}{p})$ where $\xi = 0.2$

In fact, a not negligible number of τ's is above the curve, because of the fluctuation of the electromagnetic fraction at high values. In order to increase the tau identification efficiency with a small loss in electron rejection, the cut value on the parameter ξ was moved from 0.2 to 0.15.

4.5 Electron rejection by tau to electron matching method

A different approach was proposed to increase the efficiency in tau identification keeping the same electron misidentification probability of the default algorithm.

At the end of τ's reconstruction algorithm the cut on the electromagnetic fraction based on the ξ parameter is removed and tau candidates are compared to the elements in the list of previously reconstructed electrons. This comparison is based on ΔR variable, defined as follows:

$$\Delta R = \sqrt{\Delta\varphi^2 + \Delta\eta^2}, \tag{4.9}$$

where $\Delta\varphi$ is the difference in azimuthal angle and $\Delta\eta$ is the difference in pseudorapidity. If a tau is close to an electron, i.e. $\Delta R < 0.2$, tau is removed from the list and rejected as an electron.

With this requirement, the inefficiency in pure tau identification decreases from 10% of the standard algorithm to 3%.

Another new of this method is the use of loosen cuts for identified electrons in this specific comparison. This was done in order to increase the electron sample for matching procedure, however some cases were still not recovered.

Analysis of simulated data spot the characteristics of these left cases. We can divide them into two main families.

4.5.1 Treatment of particular cases of misidentified electrons

Two examples about treatment of electrons not recovered even in case of loose identification cuts are discussed in the following:

1. low energy electrons which share their energy deposit across different wedges, because they shower close to the tower borders. In this case, generally originating two e.m. objects, the modified multi-wedge identification algorithm doesn't work because of the cut on the minimum energy deposit; the only solution to recover this kind of events is the re-introduction of the merging algorithm in the post identification.

2. electrons showering late and depositing more energy then expected inside the hadron calorimeter. The treatment of these candidates, failing

only the cut on R (equation 4.2) is based on a new discriminant variable, including calorimeter resolution, not discussed in this thesis.

4.5.2 Electrons not reconstructed as electromagnetic objects

Electrons are not reconstructed as electromagnetic objects, but they are reconstructed as tau objects, when the seed tower fraction R is larger than 0.125, and E_t is smaller than 100 GeV.

It is clear that the matching algorithm, described in section 4.5, does not work in this case. Only a small fraction of these events are recovered by the standard method based on the ξ cut.

The proposed solution to gain efficiency by recovering all cases which can introduce misidentification problems is the new approach described in next section.

4.6 The new function named "*tauelectron filter*"

The new approach, implemented by a totally new function to be included in the CDF reconstruction code, is based on an algorithm which makes electromagnetic objects out of reconstructed taus. This way, all objects which have electromagnetic characteristics, and possibly faking taus, are considered.

After passing all identification cuts of the standard algorithm, but the ξ one, reconstructed τ's are sent to loose electron identification algorithm. If they pass all electron selection criteria, they are removed from the tau list, and recognized as electrons.

The work to write the new function, the most relevant personal contribution to the new procedure, was implemented in different steps:

- *variable conversion.*

  This part of the work consisted in the analysis of both electron and τ reconstruction algorithms in order to check which variables of the tau algorithm had to be converted in order to use the existing functions for electron reconstruction; for example: E/p is now calculated using the energy deposit of the τ cluster in the electromagnetic calorimeter. The variable conversion is implemented at the end of tau reconstruction algorithm, when τ's have passed all the other selection criteria, but electron rejection one.

- *electron identification* with loose cuts.

  After variable conversion, tau candidates are processed with the electron identification algorithm, using loose cuts with the goal to efficiently reject objects which are not real τ's.

- *electron rejection.*

If tau candidates pass all loose cuts of the electron selection, they are tagged as electron candidates, and removed from τ list. Tighter cuts are applied for ultimate identification as electrons of rejected taus.

This new approach seems adequate to solve, at the same time, the problem of electrons misidentified as taus, because not reconstructed as electromagnetic objects, and the need to gain efficiency in electron identification.

The logic of this new approach also allows to recover the case of "split" electrons, i.e. electrons spreading their energy across different wedges (subsection 4.5.1), reconstructed by the tau algorithm as only one object. This results in an improvement in electron identification because the code becomes simpler and avoids merging algorithms.

Chapter 5

Methods for procedure validation and preliminary results

Real data samples corresponding to $\sim 1\ fb$ of integrated luminosity and simulated data samples from the official CDF Monte Carlo production were used to study a validation method for the new procedure, based on the analysis of di-lepton invariant mass distributions.

Full Monte Carlo simulation include event generation Pythia[1], detector[2] and trigger[3] simulation.

5.1 Data samples used in the evaluation of electron misidentification probability

First controls were made on a $Z \rightarrow ee$ (figure 5.1) Monte Carlo sample containing also background events of jets faking tau hadronic decays and of electrons from the Drell-Yan process (Feynman diagram in figure 3.1, chapter 3).

The method consists in the study of the invariant mass spectrum obtained by assigning the electron mass to all tagged electromagnetic object in the event and the tau mass to all objects identified as tau ("probe" electron, figure 5.2). The purpose is to evaluate the number of electrons which are misidentified as taus by counting the number of events falling in the Z^0 mass region of the spectrum (figures 5.3 and 5.4, for MC and data). Z^0 signal and background events are evaluated by fitting to the histograms a Gaussian function (signal) superimposed to a polynomial (background).

[1] reference [2]
[2] reference [3]
[3] reference [4]

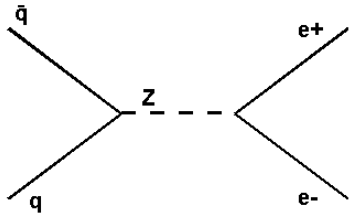

Figure 5.1: Feynman diagram of $Z \rightarrow e^+e^-$ event in $p\bar{p}$ collisions

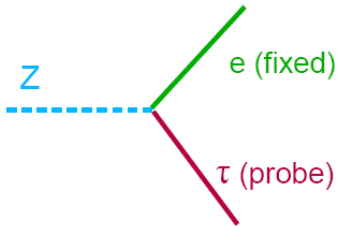

Figure 5.2: Mass assignments for di-lepton invariant mass study

Events associated to the Z^0 signal count the number of electrons identified as taus. At this point, the uncorrected probability of electron misidentification of the new algorithm can be evaluated with equation 5.1:

$$P_{misid} = \left(\frac{N(Z \rightarrow e\tau_{id})}{N(Z \rightarrow e\tau_{norej})} \right)_{Z^0 region} \tag{5.1}$$

where the numerator (figure 5.3) is the number of $e\tau$ pairs in the Z^0 region after tau identification and the denominator (figure 5.4) is the number of $e\tau$ pairs before electron rejection. Values of both quantities of ratio 5.1 are computed from the parameters of the gaussian fit. The plot in figure 5.3 refers to the MC sample where the new *tauelectron filter* function is used. Electrons are identified with the tight cut set.

To correct for unsubtracted background effects the misidentification probability calculated equation 5.1 only true events $Z^0 \rightarrow e^+e^-$ must be considered; therefore, understanding of the nature of the background in the region of the invariant mass under the Z peak is needed.

5.2 Background composition studies

Monte Carlo data samples are used to study the background composition in different regions of the invariant mass spectrum; the method is based on the

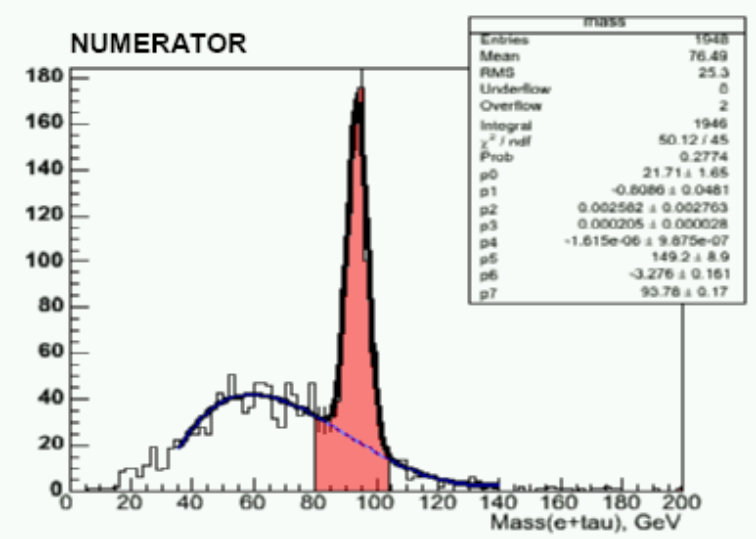

Figure 5.3: Invariant mass of $e\tau$ pairs with identified τ's

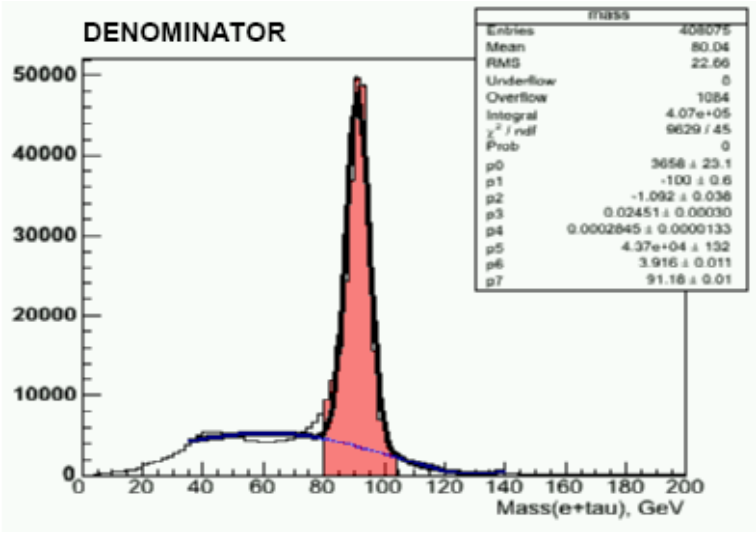

Figure 5.4: Invariant mass of $e\tau$ pairs before electron rejection

matching of reconstructed taus, including electron rejection cut, to electrons listed at the event generation level. This way, any particle misidentified as a τ is known to be originated by an electron or a jet.

This analysis revealed that, under the Z mass peak region, almost 80% of the background events are coming from electrons and 20% from jets faking taus. In the region outside the Z mass window most of events are coming from jets, apart from a few events involving low energy electrons ($M < 80$ GeV/c^2). These events have been identified as Drell-Yan processes with a detailed analysis of the Monte Carlo information at generation level.

With the background fractions quoted above for the Z^0 signal region, the corrected misidentification probability is computed applying background subtraction.

A relative systematic error of 5% on the corrected electron misidentification probability, due to the method and to the fit procedure, is estimated by varying the width of the Z^0 signal region within its natural width and the background parametrization.

The same procedure is applied to the data sample, assuming the same background composition as in the simulated data; the study of the systematic uncertainty due to this assumption is in progress. Results are summarized in the next section.

5.3 Preliminary Results

5.3.1 Monte Carlo

The probability of electron misidentification in tau identification was measured for the Monte Carlo sample described in section 5.1. Results in case of the *tauelectron filter* function is: $P^{\tau e}_{misid} = (0.29 \pm 0.02)\%$, to be compared with $P^{\xi}_{misid} = (1.29 \pm 0.13)\%$, obtained with the default "ξ-cut" method. Errors include statistical contribution and systematics coming from the fit procedure. The huge difference between the two values is the result of the comparison between the histograms shown in figures 5.5 and 5.6.

5.3.2 Data

A completely different result is obtained when comparison of the two methods is done on real data; values of electron misidentification probability are: $P^{\tau e}_{misid} = (0.24 \pm 0.06)\%$ in the case of *tauelectron filter* function and $P^{\xi}_{misid} = (0.37 \pm 0.08)\%$ in the case of the default method. Values are compatible within the errors. Histograms of the τ-e invariant mass are shown in figures 5.7 and 5.8.

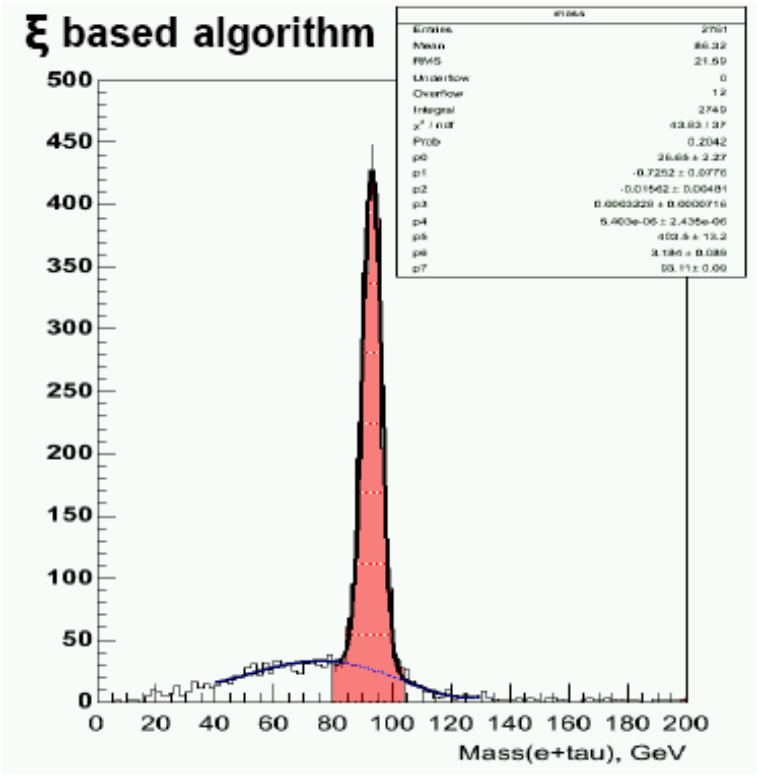

Figure 5.5: Monte Carlo e-τ invariant mass using standard algorithm with $\xi = 0.15$

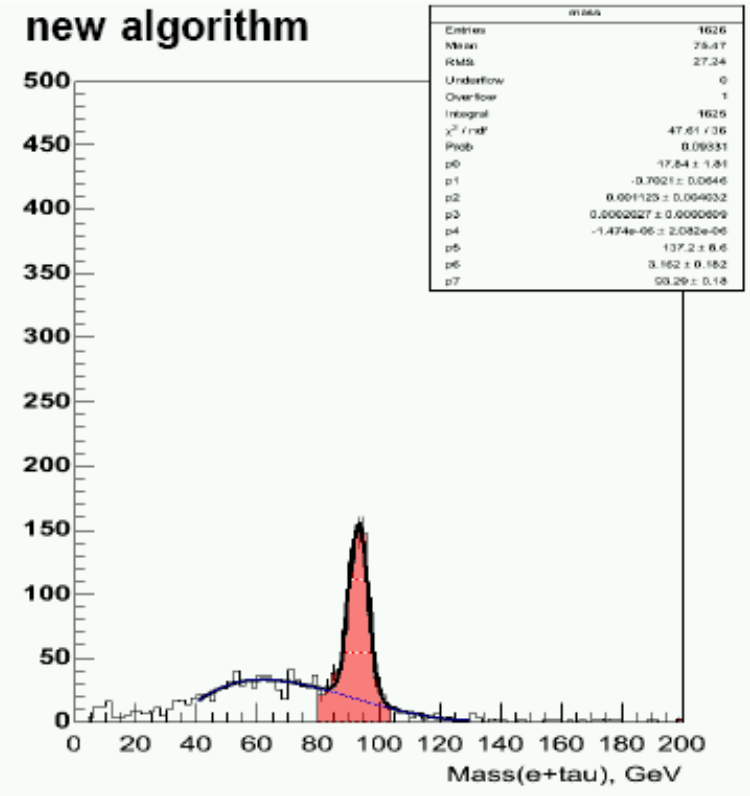

Figure 5.6: Monte Carlo invariant mass using new *tauelectron filter* function-based algorithm

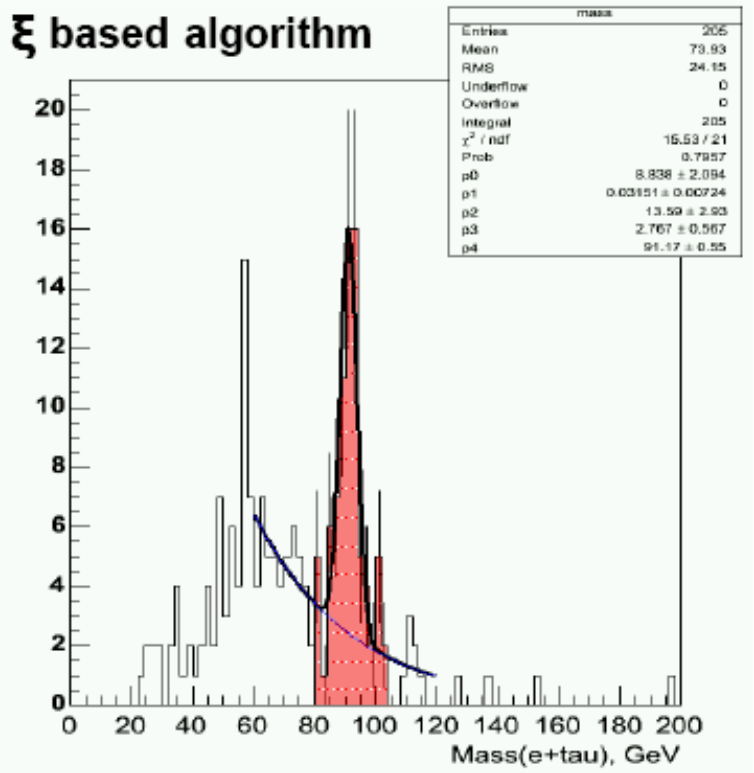

Figure 5.7: τ-e invariant mass using standard algorithm with $\xi = 0.15$ in real data

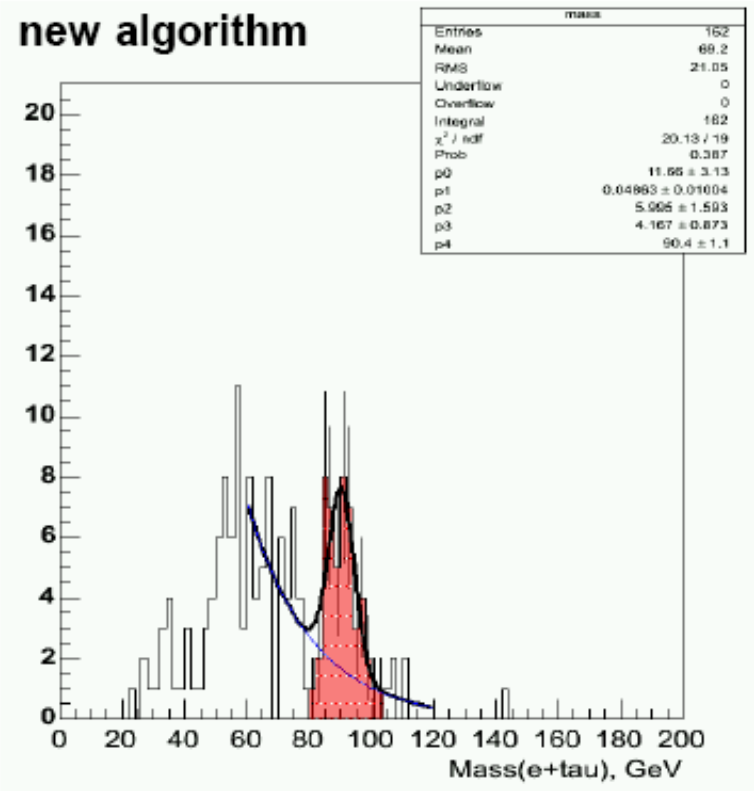

Figure 5.8: τ-e invariant mass using new *tauelectron filter* function-based algorithm in real data

5.3.3 Scale factors

Data to Monte Carlo scale factor is the ratio between misidentification probability values calculated in data and Monte Carlo samples respectively:

$$S = \frac{P_{misid}^{Data}}{P_{misid}^{MC}} \tag{5.2}$$

A complete agreement between the real data and Monte Carlo predictionswould result in a S value close to 1. For the new algorithm the following value: $S_{\tau e} = (0.83 \pm 0.21)$ is obtained; the error is dominated by the error on misidentification probability on real data. In the case of the "ξ-cut" ($\xi = 0.15$) the result: $S_{\xi} = (0.29 \pm 0.07)$ is obtained.

Data to Monte Carlo scale factor is closer to one in the case of new algorithm, where there is fair agreement between Monte Carlo and the data. In case of the "ξ-cut" the scale factor value is very low, indicating an overestimate of misidentification rate in Monte Carlo likely due to a not perfect description of the energy deposit in the calorimeter in Monte Carlo.

5.3.4 Further efficiency studies

Further efficiency and purity studies of tau sample using $W \to \tau\nu$ Monte Carlo are in progress.The possibility to have an E_t dependance of the efficiency of the new algorithm is under investigation.

Chapter 6

Conclusions

Study of the *tauelectron filter* method presented in this thesis is the last step of a sequence of modifications proposed to improve the procedure for tau identification in CDF.

In the first step, an optimization of the ξ parameter value (section 4.4) was tried ($\xi = 0.15$); the parameter ξ determines the contour of the electron-tau separation region in the electromagnetic fraction versus E/p plane. Actions on the ξ value represented a minor change in the default τ identification block of the CDF code, whose different lepton identification algorithms run independently. This had the advantage to use a single parameter to tune the electron rejection criteria, but also the limit to inhibit simultaneous increase of efficiency in both tau identification and electron rejection.

In the second step, a new logic was introduced in the tau identification procedure. Taus identified without any other electron rejection are removed if their track overlaps with a pre-identified electron one. Electrons considered in this process are identified with looser cuts with respect to standard electron reconstruction algorithm.

This procedure requires the interaction of two blocks of the code which in standard approach work separately and has the advantage to increase the tau identification efficiency but it has the intrinsic limit of loosing all cases of not reconstructed electrons.

In the third and last step of the sequence was introduced the new *tauelectron filter*; in this case, the new code widely mixes the existing electron and tau blocks to produce e.m. objects out of tau ones in addition to the e.m. objects reconstructed by the standard algorithm, resulting in an increase of the code complexity. However, several benefits are obtained as discussed below.

The full lepton identification procedure in the new approach has a cyclic structure; first runs the e.m. object block, then the muon block, the tau block and, finally, part of the e.m. object block runs again applying looser cuts to the tau candidates. A similar cyclic structure was already present in

the logic of the default procedure, since the cut on the e.m. fraction of tau candidates was implemented with observables typical of e.m. objects.

A relevant result obtained with the new *tauelectron filter* is a sizable improvement in electron rejection from taus. Moreover, this approach introduces news in different parts of the CDF software environment.

First of all, the logic in lepton identification is changed with the following prompt results:

1. electron contamination in the identified tau sample is decreased without loosing in efficiency of tau identification with respect to the separation method based on the e.m. fraction cut (ξ parameter) to separate electrons and taus

2. the electron reconstruction efficiency can be improved by combining standard electron reconstruction algorithm and new approach, which recovers almost all cases lost in the standard procedure by making electrons out of tau objects.

3. finally, comparison of results on simulated data obtained with the new and the standard ways to reject electron from taus spots that Monte Carlo overestimates the probability for electrons to be misidentified as taus when the cut on the e.m. fraction is applied. This is a direct evidence that Monte Carlo doesn't reproduce electromagnetic fraction distributions as a function of E/p and it represents the most striking result of the work done in this thesis.

Possible future developments towards a further improvement in tau reconstruction efficiency could come from studies in progress on the dependence of reconstruction efficiency from lepton E_t.

Acknowledgements A special thank to doctor Pavel Murat, Fermilab, promotor of my training plan as summer student, and to Yanjun Tu, for their assistance during my job in CDF. I would like to thank also professor Giorgio Bellettini and doctor Fabrizio Scuri for their help in writing this thesis.

Bibliography

[1] The CDF II Collaboration. *The CDF II Detector Technical Design Report, 1996.* FERMILAB-Pub-96/390-E, http://www-cdf.fnal.gov/upgrades/tdr/tdr.html.

[2] T. Sjostrand et al. *Pythia 6.2 Physics and Manual.* hep-ph/0108264.

[3] Fermilab public note *CD-doc-2276-V0*

[4] S. Rolli et al. *Hardware trigger simulation at CDF*, proceedings of the International Conference on Computing in High Energy Physics (CHEP 2000). Padova, Italy, 7-11 Feb 2000.

[5] M. Giunta. *Precision Measurement of the Top Quark Mass at CDF.* PhD thesis.