CDF/ANAL/TOP/CDFR/8627
Draft 0.0
October 2, 2007

Search for the ttbar tau dilepton signal and measurement of $r_\tau = BR(t \to \tau\nu b)/BR(t \to l\nu b)$ with 1 fb^{-1}

Stephane Tourneur, Aurore Savoy-Navarro

LPNHE - University of Paris 6, France

Abstract

The search for top quark pairs implying tau leptons in their decay products at the CDF detector, has two main goals: To observe one of the last ttbar chanels yet to be observed, and to check for any disagreement with the standard model expectation. The signal is tiny and hard to extract witin a background dominated by Z-¿TauTau+2 jets events, and W+3 jets events in which one jet is misidentified as a hadronic tau decay. One of the main aspects of this thesis is the development of a rigorous method to estimate the leading background of W+jets within an environment with high jet multiplicity, high missing transverse energy, one electron or muon and one hadronic tau. The analysis was performed using two data samples totalizing integrated luminosities of 350/pb and 1/fb. After having carefully checked the coherence of the method used, a 76% evidence has been found for the top in tau signal. Also, the relative excess R of top signal as compared with the standard model value has been measured to be less than 1.5 at 95% confidence level.

DRAFT OCTOBER 2, 2007

Contents

1 Theoretical motivations

1.1 Observation of the top pair production in the lepton+τ channel

The goal of this study is to extract the top pair events in which one top produces an electron or a muon, and the other top produces a τ lepton decaying into hadrons, from the mass of events produced by the 1.96 TeV $p\bar{p}$ collisions at Tevatron. With the increase of the luminosity and the upgrading of the CDF detector compared to the Run I period (1993-1994), it will be hopefully possible during the Run II to establish for the first time a clear evidence of this still poorly known top decay channel.

The signature of this process: $p\bar{p} \to t\bar{t} + X \to e, \mu + \nu_{e,\mu} + \tau_{had} + \nu_\tau + b\bar{b} + X$ is thus characterized by one electron or one muon, a τ jet, two b-quark jets and the total missing transverse energy (referred below as $\not\!\!E_T$) resulting from the ν's emitted in the decays of the W's into leptons and consequently in the decays of the τ lepton(s). Therefore this analysis requires to master the identification of all the following

DRAFT OCTOBER 2, 2007

Integrated Luminosity	$t\bar{t}$ $\rightarrow lepton + \tau_{had}$ produced	passing geometry and kinematics requirements
350 $pb{-}1$	96	12.2
1 $fb{-}1$	276	34
4 $fb{-}1$	1102	138

Table 1: *Monte Carlo estimations of top pair events produced in the lepton + tau channels, assuming a cross section of 7.3 pb for the whole top pair production*

fundamental objects: Electrons, muons and hadronic decays of the τ lepton and eventually of the b-quark jets. A good measurement of the total missing transverse energy and of the transverse energy (E_T) of the two b-jets is critical too. Understanding the most important backgrounds, i.e. the processes: $Z \rightarrow \tau_l\tau_{had}$, $W \rightarrow l\nu + jets$, QCD with jets faking a τ hadronic decay, and diboson production, in a high jet multiplicity environment is a prerequisite.

In Run I, the analysis of the 109 pb^{-1} of 1.8 GeV $p\bar{p}$ collisions concluded on the observation of 4 $e\tau$ and $\mu\tau$ candidate events, where 2.5 ± 0.4 background events and 1.1 ± 0.4 top events were expected (assuming $\sigma_{t\bar{t}} = 7.7 \pm 2pb$, Run I combined result). Three of the 4 candidate events also had b-tagged jets, where 0.28 background events (and 0.63 top events) were expected. This was formally giving a 3 σ significance for the presence of non background events, but because only 0.63 top events were expected, the nature of these 3 events was not clear [2].

In Run II, the increase from 1.8 to 1.96 GeV in the center of mass energy has yielded an increase in the cross section for the top pair production estimated by the theoretical calculations to grow from 5 pb to 6.5 pb, i.e. by on the order of 30% . Furthermore, the higher luminosity produced by the Tevatron at Run II allowed CDF to collect about 350 pb^{-1} data, by September 2004 , and about 1 fb^{-1} data by the end of 2005, data that are available for this analysis. A total of at least $4fb^{-1}$ data are expected to be recorded by 2008, if the machine continues to run as presently. There are good expectations for getting up to 6 fb^{-1} or even 8 fb^{-1} by 2009, i.e. before the LHC is running at full speed. The Table 1 gathers the number of signal events that are expected to be produced in the τ dilepton channels both in the electron and muon cases, taking 7.3 pb as the total top pair production cross section [1], for the total luminosity taken into account for this thesis (i.e. 350 pb^{-1} and 1 fb^{-1}) and extrapolating directly these numbers to a case of 4 fb^{-1} total integrated luminosity.

A clear 3 σ observation of the top in τ dilepton channels should be possible soon with an optimization of $S/\sqrt{B}$.

DRAFT OCTOBER 2, 2007

1.2 A key-tool for beyond the Standard Model searches

Apart from the observation of this top decay channel into dileptons with at least one τ lepton, this study aims to the measurement of the cross section of this decay process. This will not really help improving the precision of the overall top pair production cross section, as the studied decay process cannot add significantly to the precision. The main point here is to check if as expected from the Standard Model (SM) the top decays only into Wb or if there is room for other decay channels as in particular, the decay of the top into a charged Higgs plus a bottom quark. This charged Higgs would then decay into a τ lepton and its corresponding neutrino, leading to an enhancement of the ratio, $r_\tau = \frac{t\to\tau\nu_\tau b}{t\to l\nu_l b}$ (l = e or μ), in contradiction with the value of 1 predicted by the Standard Model. The optimization of the r_τ value should be done using $\frac{S}{\sqrt{S+B}}$. Indeed any value greater than 1 would thus be an indication of physics beyond the Standard Model. For instance if there is a charged Higgs of mass lower than the top mass, it would preferently couple with the top quark because of its high mass. For the same reason, it would couple much more to the τ lepton than to lighter leptons. The decay chain: $t \to bH^+ \to b\tau\nu_\tau$ would yield to values of r_τ greater than 1. A total integrated luminosity of at least $4fb^{-1}$ or more is needed to establish a value of r_τ different from 1. While waiting for more data, at least one can improve the first upper limit of 5.2 achieved by CDF with 195 pb^{-1} of data [**?**]. Moreover a crucial aspect of this study is that understanding Physics signatures that involve missing energy, several jets, 2 or 3 leptons and especially τ leptons is instrumental for the search of New Physics. Indeed, these are typical SUSY signatures, and if $tan\ \beta$ is high enough, the rate of decays into τ leptons as compared to other processes are predicted to become predominant. Therefore, being able to handle high multiplicity signatures including τ leptons is a new important achievement in $p\bar{p}$ colliders.

2 Data selection and analysis flow

2.1 Introduction

This analysis makes use of two data sets which were taken during two different run periods, before and after the September 2004 Tevatron shutdown. The first period corresponds to about 350 pb^{-1} data, while the second period allowed to reach more than 1 fb^{-1} data for this analysis (until February 2006).

The data are triggered by high p_T electron and muon triggers.

As described in the introduction of the first section, the goals of this analysis are twofold: The extraction of the top pair signal through the lepton+τ decay channel, and the measurement of the ratio $r_\tau = \frac{BR(t\to\tau\nu b)}{BR(t\to l\nu b)}$. Therefore, two different selections are performed in order to optimize each measurement.

This work follows a *blind* analysis method and therefore the observation in the data of the signal event candidates is done at the very end, after the consistency of the analysis method, the precision of our predictions, have been thoroughly checked with

DRAFT OCTOBER 23, 2007

the use of control samples.

Another important point of the work presented here is the study of a crucial aspect of this search, namely the jet to tau fake rate, as the tau lepton is identified in this environment by its hadronic decay (so-called hadronic tau or τ_h).

This chapter gives in details the various steps in the data selection in order to extract the signal. Likewise, it describes the estimate of all the dominant backgrounds.

2.2 Monte Carlo simulations

Monte Carlo (referred below as MC) simulated samples are used in this analysis to get an estimate of the number of signal events and physics background (Z, WW) events passing our analysis selection.

The simulation follows a run-dependent scheme: For instance, a sample used to reproduce the 2002-2004 period corresponding to the first 350 pb^{-1} mimics the detector and beam conditions of the runs recorded during this period. This way of doing enables a better matching between the simulated and real events within the course of each run. Moreover, the generated events are overlayed with minimum bias events from multiple interactions, with a weight proportional to the instantaneous run luminosity.

A particular decay package, named Tauola [11], is systematically used to handle in a proper way the tau lepton decays, taking into account the tau polarization.

The prediction of a number of events is never based on the Monte Carlo simulation only, but on the result of a mixture between real and simulated data. In order to attempt measuring the properties of the dramatically small scales of the top quark one needs a well marked out path to guide the experimentalist from the already known regions to the distant ones. It is like exploring the extremely far universe: we need "standard candles". Today, W and Z bosons have become the standard candles of many analyses in very high energy physics. In this analysis, we used the W and Z mass peaks as beacons for a lot of things: Z cross section for integrated luminosity measurement and to scale the lepton identification efficiency in the simulation, W and Z masses (80.4 and 91.2 GeV) for lepton energy tuning in data and simulations. Also, we used the data consisting of a Z boson decaying into two leptons (electrons or muons) accompanied with two jets to scale our Monte Carlo predictions for the number of Z+2 jets events. Thus, to sum up the idea behind the use of Monte Carlo simulations, the strategy is to perform Monte Carlo simulations, scale them so that they agree with W and Z candles, and then use the simulation to extrapolate the result into less known regions.

We only use Monte Carlo generators at leading order:

- Pythia [9] is used for the simulation of the top pair events, for W and Z candles. Pythia is parametrized to reproduce the CDF minimum bias events (tune A [12]).

- The leading-order matrix elements generator Alpgen [10], interfaced with Pythia or Herwig [8] in order to handle the parton shower and hadronization processes,

DRAFT OCTOBER 2, 2007

is used to bring us from the observed number of $Z \rightarrow ll+ \geq 2jets$ (l=e, μ) events to an estimate of the number of $Z \rightarrow \tau\tau$ events surviving our analysis selection.[1]

DRAFT OCTOBER 2, 2007

[1]Since the normalization of the number of $Z+ \geq 2jets$ events is obtained from the data itself, we do not rely on complicated matching schemes between partons and jets and combinations of Z+0, 1, 2, 3 and more partons. However, we only need one Alpgen sample, namely Z + 2 partons. The interface with Pythia is tested (see section 3.7.1.2) to well reproduce the number of extra jets and thus to ensure that the Z + 2 partons sample is indeed an inclusive $Z+ \geq 2jets$ sample.

2.3 Data preselection

Two analyses were successively performed : the first one uses 350 pb^{-1} of data recorded until September 2004, and the second one uses the overall 1 fb^{-1} data that were recorded on tape until the shutdown of February 2006.

The data used for the signal measurement was triggered by the inclusive high P_t electron (or muon) trigger.

The high P_t electron trigger looks for a cluster in the central electromagnetic calorimeter with a transverse energy greater than 18 GeV and with less than one tenth of energy deposited in the hadronic calorimeter.

The high P_t muon trigger requires stubs in the CMUP or CMX sub-detector systems. The stubs must also match a track in the COT with a transverse momentum greater than 18 GeV/c.

In order to make the preselected samples, the data selected by these triggers is used and the trigger requirement is reasserted at the offline reconstruction level. Furthermore, eventual double events are removed.

We also take care that the events belong to the official CDF good run lists meeting these requirements : "good electrons" for the electron sample and "good electrons, good muons" for the muon sample. This good run list ensures the high-quality of the data finally saved for the analyses.

2.3.1 The electron identification

Only electrons produced in the central barrel are used in this analysis. These electrons are produced in the central part of the detector, flying through all the layers of the COT to the central electromagnetic calorimeter. Their pseudorapidities are thus comprised between -1 and 1. Their identification makes use of the central calorimeter, the CES, and the COT.

The identification of $high - P_t$ central electrons includes two parts:

1. First form basic loose objects using calorimeter information only. These are calorimeter clusters with a large electromagnetic fraction. These are called "CdfEmObject" and are defined this way:

   - look for CEM clusters seeded by a tower with an electromagnetic transverse energy greater than 2 GeV
   - the cluster must have $E_{had}/E_{em} < 0.125$ or a transverse energy greater than 100 GeV

2. CdfEmObject must pass the series of cuts defined in Table 2.

The *Lshr* variable is described in details in [13]. The purpose of this variable is to provide some discrimination of electrons and photons from hadronic showers faking these particles in the central electromagnetic calorimeter (CEM), by comparing the

DRAFT OCTOBER 2, 2007

observed sharing of energy deposition between towers in the CEM to that expected for a "true" electromagnetic shower. It is defined as

$$Lshr = 0.14 \frac{\sum_i (M_i - P_i)}{\sqrt{(0.14\sqrt{E_{EM}})^2 + \sum_i (\Delta P_i)^2}}$$

where the sums are over the one or two towers in the electromagnetic calorimeter cluster adjacent to the seed tower and in the same Φ-wedge as the seed tower. M_i is the measured energy in an adjacent tower, P_i is the predicted energy deposit in the adjacent tower, known from test beam data. E_{EM} is the total electromagnetic energy in the calorimeter cluster, and ΔP_i is an estimate of the uncertainty in P_i. All energies are in units of GeV.

The *CES strip* χ^2 variable is decribed in [14]. The CES (central electromagnetic calorimeter "shower max" strips and wires) strip cluster is fit to a standard set of electron shower profiles from test beam data. The *CES strip* χ^2 corresponds to the "goodness of fit" χ^2:

$$\chi^2 = \frac{1}{4} \sum_{i=1}^{N} \frac{y_i^2 - y^2(x_i)}{\sigma_i^2}$$

where y_i is the measured fraction of energy for channel i, $y(x_i)$ is the fraction expected from the standard profile in channel i and σ_i is the RMS fluctuations in channel i measured from 10 GeV electron test beam.

2.3.2 The muon identification

The high pt muon identification follows the one usually applied in CDF [32]. It is used to search for the unique muon present in the $t\bar{t} \to \mu\tau\nu\nu b\bar{b}$ channel. Only central muons are used here in order to benefit from the whole CDF tracking system.

The general strategy for the muon identification is to search for minimum ionizing tracks that leave a track in the tracker, and then go through the calorimeter with a minimal energy deposit. The track is finally reconstructed as a so-called *stub* in the muon drift chambers. High pt muon identification has to fight against backgrounds such as:

- muons produced inside a heavy quark jet: these are likely to be surrounded by other hadrons and thus can be reduced by imposing the muon to be isolated. The isolation is imposed at the calorimeter level in order to count also the energy of neutral particles in the isolation cone.
- Charged kaons or pions decaying while flying through the Central Outer Tracker. These mesons decay into a muon and a neutrino about 100% of the time for charged pions and 63% for charged kaons. The meson and muon are then sometimes reconstructed as a unique track with an artificially low curvature and thus

DRAFT OCTOBER 2, 2007

Variable	Cut
Region	CEM
Track Z_0	< 60 cm
Electromagnetic calorimeter cluster E_T	> 20 GeV
Track p_T	> 10 GeV/c
is a $\gamma \rightarrow ee$ conversion?	no
is fiducial?	yes
N COT axial Super Layers	≥ 3
N COT stereo Super Layers	≥ 2
E_{Had}/E_{Em}	$< 0.055 + 0.00045$ E[GeV]
Relative calorimeter isolation	< 0.1
Lateral Shower Profile (Lshr)	< 0.2
E/P	$< 2.$ unless $p_T > 50$ GeV/c
CES Δz track-cluster matching	< 3 cm
signed CES Δx	-3 cm $< q * \Delta x <$ 1.5 cm
CES strip χ^2	< 10
Efficiency	0.812 ± 0.004

Table 2: Central electron identification cuts and efficiency. The efficiency is the ratio of electrons passing the cuts of the upper part of the table (above the double line) that also pass the cuts below.

appear as a high pt muon. Against this high background, on top of the isolation requirement, a tight cut on the ΔX distance between the COT track extrapolated to the muon detector and the muon chamber stub (see ΔX cut below) can be efficient.

- *punchthrough hadrons* are hadrons that enter the calorimeter and produce hits in the muon system. Most punchthroughs are due to tertiary pions or kaons within the hadron shower which decay to muons. These are reduced by the isolation, the minimum ionization requirement and the cut on the ΔX track-stub distance.

The search for the muon starts online at the trigger level with the identification of the relevant stubs in the CMUP and in the CMX devices.

Different selections are applied depending on which region of the central muon detector the muon candidate points to. The set of sequential cuts that make up the central muon identification are summarized in Table 3.

In addition to these cuts, the muon track extrapolated up to the muon chamber radius is required to fall inside the fiducial volume of the muon subdetectors CMUP or CMX (Fig.1). Also, any event in which a cosmic ray is found is discarded.

However, the different devices of the central muon system were not operational at the same time. Indeed, the CMP so-called "bluebeam" region (the region on the top of the CMP comprised in the angles $45^o < \Phi < 52.3^o$) has been operating stably since

DRAFT OCTOBER 2, 2007

Track p_T	> 20 GeV
Electromagnetic Calorimeter Energy deposit	$< 2 + \max(0, 0.0115(p - 100))$ GeV
Hadronic Calorimeter Energy deposit	$< 6 + \max(0, 0.028(p - 100))$ GeV
E_T in cone of R = 0.4 around muon tower	$< 0.1 p_T$
Number of axial SL with $\geq$5 hits forming the track	≥ 3
Number of stereo SL with $\geq$5 hits forming the track	≥ 3
Track $\lvert z_0 \rvert$	< 60 cm
Tracks with no silicon hits attached: d_0	< 0.2 cm
Tracks with silicon hits attached: d_0	< 0.02 cm
muon chamber stub-track matching:	
For CMUP muons: ΔX_{CMU}	< 3 cm
ΔX_{CMP}	< 5 cm
For CMX muons: ΔX_{CMX}	< 6 cm

Table 3: The selection cuts applied to muons, where p is the track momentum ; SL stands for "COT super-layer" ; the track z_0 is the z-coordonate of the point of closest approach between the COT track and the beam axis ; d_0 is the impact parameter ; ΔX is the distance, in the $r - \Phi$ view, between the reconstructed stub and the expected position from the muon near the stub as obtained by extrapolating the track to a position close to the stub.

run number 154449 (after 2002). For earlier runs the bluebeam region was either noisy or turned off. Thus any muon with stubs in the bluebeam region for the runs < 154449 (ie. before 20/11/2002) are rejected.

Likewise, any muon with stubs in the CMX devices are rejected for the runs $<$ 150145 (ie. before 20/08/2002).

Also, the new keystone and miniskirt regions inside the CMX chamber was not operational before the Automn 2004 Tevatron shutdown, for runs < 190697. For these runs that correspond to the first 350 pb^{-1} of data, the CMX muons were thus restricted to the arches (Fig. **??**). Finally, the CMX wedge 14 on West side (Fig. **??**) presents a big drop in the online trigger efficiency at level 1 starting from the runs > 190697 ; muons with stubs found inside it are thus excluded accordingly.

DRAFT OCTOBER 2, 2007

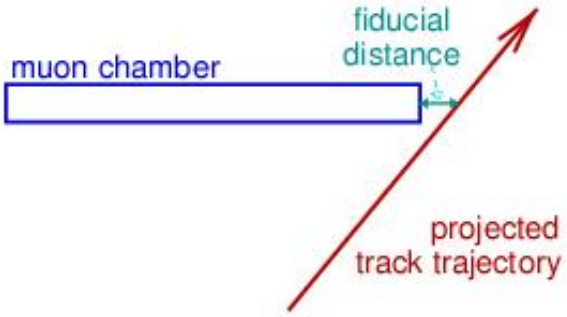


Figure 1: The so-called fiducial distance between the muon track extrapolated to the muon chamber radius and the edge of the muon chamber is required to be <0.

DRAFT OCTOBER 2, 2007

2.3.3 The tau lepton identification

Tau leptons are much harder to identify than the other charged leptons, especially in the case of a $\bar{p}p$ collider. The only way to identify them in this collider environment is by their hadronic decays into one or more charged pions, and in addition eventual neutral pions. They form a so-called tau-jet accompanied by a neutrino. This jet is often difficult to distinguish from other jets that are therefore a serious physics background to overcome. Tau leptons decay hadronically about 65 % of the times.

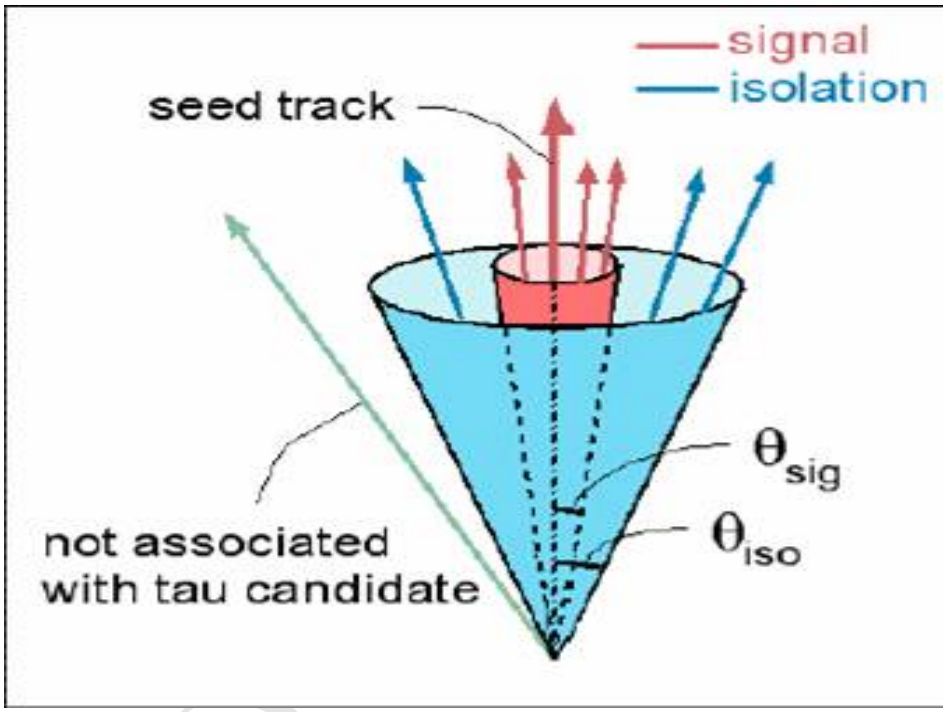


Figure 2: The so-called tau cone (in red) and tau isolation cone (in blue) containing tau tracks and isolation tracks. The tau isolation cone opening angle θ_{iso} is fixed at 40^o. The tau cone starts with an opening of 10^o and decreases with the tau energy as described in Fig. 5.

The identification of the tau object has led to a lot of experimental work which make use at best of all the subtleties of the detector and profit especially from the tracking device to well identify the special 1 or 3 track feature of this jet. This analysis profits from the work achieved for years in the CDF collaboration in order to identify and select at best these tricky objects. Basically, the tau identification algorithm searches for narrow isolated jets(Fig. 2), taking profit of the high tau lepton boost due to the fact that the tau lepton mass is low compared to its kinetic energy[2].

[2]The tau mass is $m_\tau = 1776.99^{+0.29}_{-0.26} MeV/c^2$ [38] and the tau identification requires the reconstructed tau transverse energy to be greater than 15 GeV.

DRAFT 2007

The tau selection takes also advantage of the shower maximum detector (CES) to attempt to reconstruct the photons inside jets [15], that are often produced inside the hadronic tau-jets because of the π^0 decays. Indeed, these photons are observable via their energy deposits in the CES that appear as strip/wire clusters. Clusters composed of 5 strips/wires contain about 95% of the energy of electrons/photons. The two photons from neutral pion decays will be merged and appear as a single cluster most of the time. But, for π^0 energies smaller than 10 GeV, the photons can be resolved in some cases [15]. Inside the tau identification algorithm, CES clusters unmatched with COT tracks are called π^0's (see Fig.3).

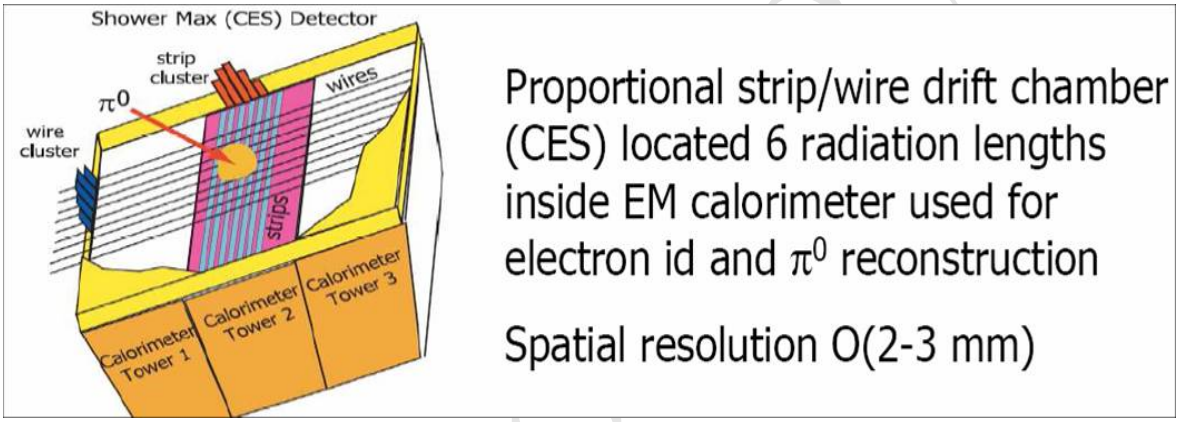


Figure 3: Neutral pions or photon reconstruction with the CES subdetector, used for the tau identification.

The tau lepton identification is described in more details here below and leads to an efficiency that varies from 35 to 45% depending on the transverse energy of the tau lepton. It is found that about 1% of central jets are mistagged, meaning misidentified as tau-jet (Fig.4).

The series of cuts applied for identifying a tau lepton decaying hadronically is divided into two parts:

- The so-called TauFinder algorithm requires:
  - A seed tower with $E_T > 6$ GeV
  - A "seed" track pointing to the seed tower with $p_T > 4.5$ GeV/c
  - < 6 neighbouring towers, each with $E_T > 1$ GeV
  - A cluster in $|\eta| < 1.1$
- The tau identification cuts are then applied to those objects passing the previous TauFinder requirements an consists of this set of cuts:

DRAFT OCTOBER 2007

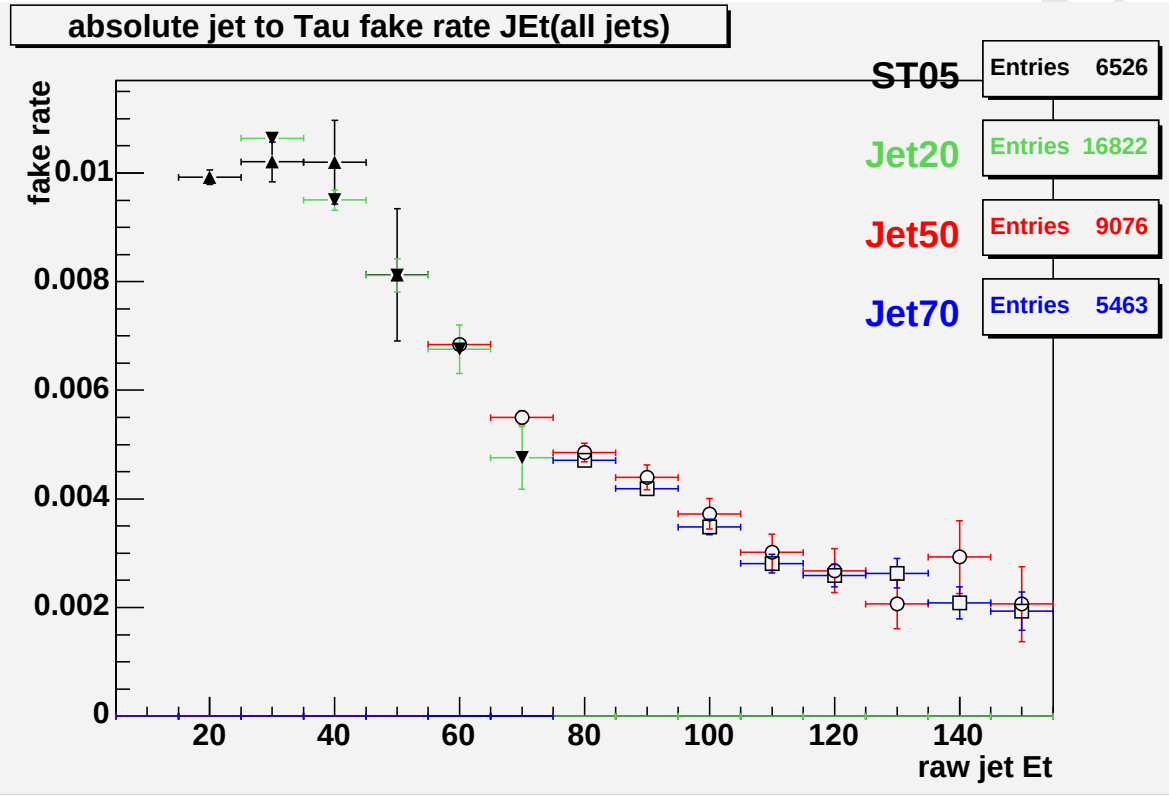


Figure 4: The probability for a central jet to be misidentified as a hadronic tau, depending on the jet transverse energy. This fake rate was calculated by us using the four QCD triggers Single Tower 5 GeV, Jet 20 GeV, Jet 50 GeV and Jet 70 GeV, restricting oneselves to the triggers unbiased energy regions. These triggers are described in the note 8208 devoted to the jet to tau fake rate.

- The number of tracks with p_T above 1 GeV/c found in the τ shrinking cone (a cone with a radius of less than 10 degrees, see Fig.5), must be equal to 1 or 3.
- The absolute value of the tau lepton electrical charge $|Q|$ must be equal to 1.
- An electron veto defined by: $\frac{E_{had}}{\sum_{tracks} P_{tracks}} > 0.15$.
- The sum of the transverse momentums of all tracks and π^0's reconstructed in the τ cone must be greater than 15 GeV/c.
- The z-coordinate, along the beam axis, of the tau lepton ($\tau|z_0|$) must be less than 60 cm.
- The impact parameter of the tau seed track ($\tau|d_0|$) must be smaller than 0.2 cm.

DRAFT OCT 0

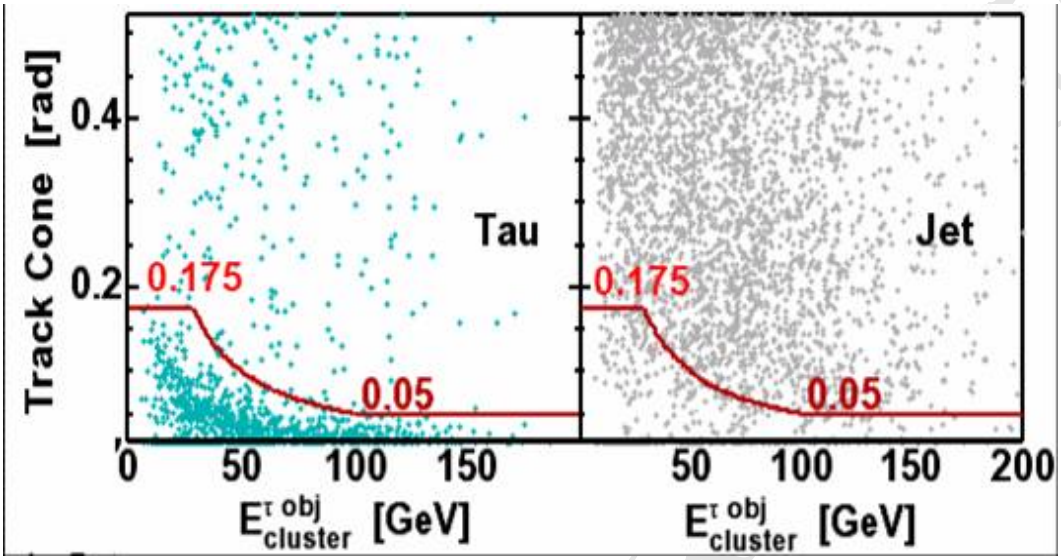


Figure 5: The tau cone opening is shrinking with the tau-jet energy, from 0.175 rad down to 0.05. This varying cone shape provides a good acceptance for "true" taus (left plot) and a good rejection against jets faking taus (right plot).

- The invariant mass of system composed of the tracks and reconstructed π^0's in the τ cone must be smaller than the tau lepton mass, i.e. than 1.8 GeV/c^2.
- The energy deposited outside the τ cone in the isolation cone of $\Delta R = 0.4$ cannot exceed 10% of the τ energy.
- The number of tracks inside the isolation cone with p_T above 1 GeV/c must be equal to zero (track isolation cut).
- The number of π^0 with transverse energy greater than 0.5 GeV in the isolation cone must be equal to zero (π^0 isolation cut).
- The ratio cluster E_T / seed track p_T must be greater than 0.5 (muon veto cut).
- The seed track quality defined by at least 3 stereo and axial superlayers with at least 5 hits must be fulfilled
- The fiduciality defined by the condition: 9 cm $<$ seed track $|z_{CES}| < 216$ cm, must be fulfilled.

DRAFT OCTOBER 7

2.4 The energy corrections and validation of the fundamental objects

In this subsection, the computation of the energy scale factors and the integrated luminosity measurement using the benchmark processes $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$, are presented. The extraction of the Z signal is also a way to check that the identification of the electron, muon and tau leptons is well understood in the real data and in the simulated samples. Finally, the $W \rightarrow \mu\nu$ process will be used to check that the measured transverse energy due to neutrinos is under control, and a description of the way the jet energy is calibrated at CDF will be summarized.

2.4.1 Electrons

The validation of the electron identification In order to validate the central electron identification in our analysis, we check that we are able to reproduce the efficiencies for the electron identification published by the CDF electroweak group [23]. We take the data sample triggered by the high p_T electron trigger and a $Z \rightarrow ee$ Monte Carlo sample generated with Pythia. Two electron types are defined in order to calculate the identification efficiency: A "tight electron" is an object that is identified as an electron following the criteria defined in the subsection 3.3.1. A "loose electron" is an electromagnetic cluster in the central calorimeter("CdfEmObject") with a transverse energy greater than 20 GeV and associated with a COT track with $z_0 < 60$ cm and $p_T > 10$ GeV/c. Then, two numbers are defined:

- NTT is the number of events containing two tight electrons. The two electrons must have opposite electrical charges and an invariant mass comprised between 76 GeV/c^2 and 106 GeV/c^2.

- NTL is the number of events containing one tight electron and one loose electron[3]. They also must have opposite charges and an invariant mass between 76 GeV/c^2 and 106 GeV/c^2.

Both data samples of sizes NTT and NTL are composed mostly of $Z \rightarrow ee$ events. The background contamination was found to be on the order of 0.15% for NTT and 1.90% for NTL [23]. Our numbers NTT and NTL are thus scaled accordingly, respectively by 0.9985 and 0.9810 in order to count only the number of $Z \rightarrow ee$ events.

Since either of the two electron candidates could be chosen as the loose electron if both objects pass tight criteria, the formula used for the electron identification efficiency is:

$$\epsilon_{CEM} = \frac{2 \times NTT}{NTT + NTL}.$$

[3]Note that, by construction, a tight electron is also a loose electron, so that NTT is always smaller than NTL

DRAFT OCTOBER 2, 2007

Cut	sewk7d	ztop2i
conversion	0.9693	0.9703
Fiducial	0.9831	0.9878
COT ax.seg.	0.9990	0.9993
COT st.seg.	0.9998	1.0000
Had/Em	0.9929	0.9881
Cal. isolation	0.9731	0.9766
Lshr	0.9918	0.9848
E/P	0.9337	0.9283
CES dZ	0.9976	0.9988
signed CES dX	0.9983	0.9991
CES strip chi2	0.9667	0.9810

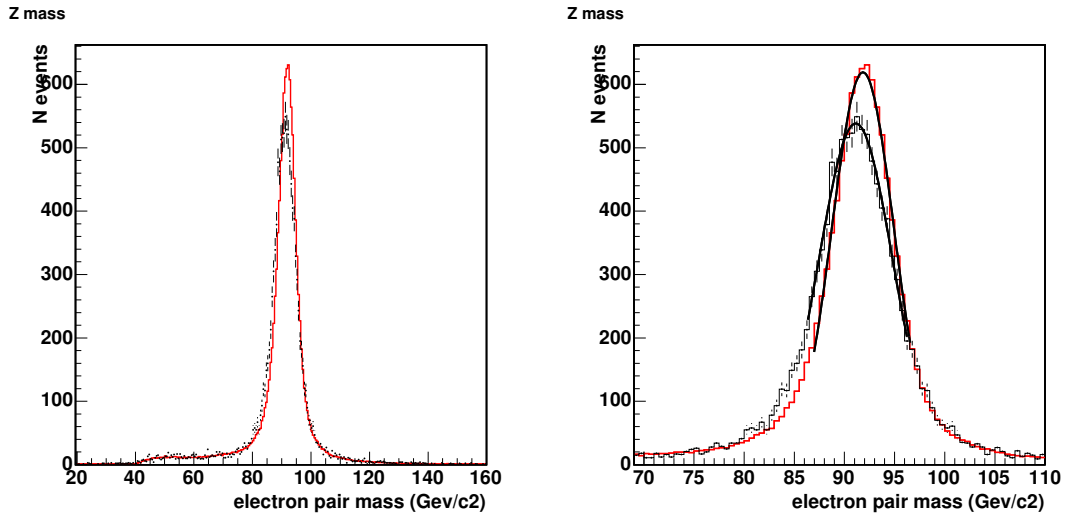
Table 4: CEM N-1 efficiencies in the data (sewk7d) and in $Z \rightarrow ee$ Monte Carlo (ztop2i) for the 350 pb^{-1} analysis.

An efficiency of $\epsilon_{CEM}^{data} = 0.812$ is found for data and $\epsilon_{CEM}^{MC} = 0.815$ for the Monte Carlo sample [23]. The Monte Carlo acceptances are thus scaled by a factor of $\frac{\epsilon_{CEM}^{data}}{\epsilon_{CEM}^{MC}} = 0.996 \pm 0.005$ for electrons (350 pb^{-1}), in agreement with [23].

We also checked the so-called $N-1$ efficiencies. Our results are gathered in Table 4. The $N-1$ efficiencies correspond to the probability for a *true electron* that successfully passes all the electron identification cuts but one, to furthermore pass this last cut. We use the same formula as the one used for the total efficiency, except that the *loose* electron must pass all *tight* selection criteria but the one cut in question. These numbers agree with the standard CDF numbers for the 350 pb^{-1} analysis [23]. This makes us confident that our identified central electrons are the same objects as the electrons selected in other official high p_T analyses with central *tight* electrons.

Electron energy tuning with $Z \rightarrow ee$ events The Z mass has been a well known standard model parameter since the LEP experiments. Its value was measured to be $M_Z = 91.1876 \pm 0.0021$ GeV/c^2 [38]. The hard work from the collaboration to calibrate the electromagnetic calorimeter energy enables to get a sharp Z lineshape when asking for two tight central electrons in the event. This is drawn in Fig.6(a) for the data sample recorded between 2005 and 2006 (black points) and the corresponding Pythia simulation (red histogram). A gaussian fit gives the position of the maxima that are close to the expected value (M_Z- 34 MeV/c^2) but still a little bit displaced by an order of 1% (see Figs.6(b) and 7). The electron energy scale factors are then calculated in order to make the data and Monte Carlo Z lineshapes peak at the right energy value. The obtained scale factors are gathered in Table 5 for data and Monte Carlo and for both analyses, the first one dealing with the years 2002-2004 data taking period ($\sim$ 350 pb^{-1}), and the second one with both 2002-2004 and 2005-2006 data taking periods ($\sim$ 1 fb^{-1}).

DRAFT OCTOBER 2, 2007



(a) *Full energy range*

(b) *A gaussian fit is applied to the Z lineshapes to calculate the Z peak position*

Figure 6: Z lineshapes for the 2005-2006 electron data and corresponding Pythia simulation. The black dotted histogram is for real data and the red histogram is the result of a Pythia simulation.

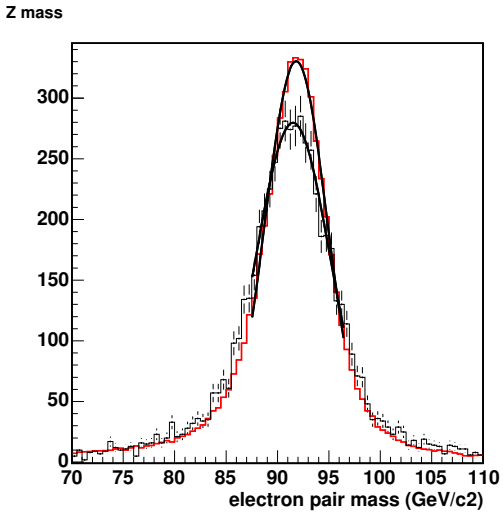


Figure 7: Z lineshapes for the 2002-2004 electron data and corresponding Pythia simulation. The black dotted histogram is for real data and the red histogram is the result of a Pythia simulation.

DRAFT

DRAFT OCTOBER 2, 2007

	First analysis (2002-2004 data)		Second analysis (2002-2006 data)			
			2002-2004 runs		2005-2006 runs	
	data	MC	data	MC	data	MC
Energy scale factor	1.000	0.996	0.996	0.988	1.000	0.988

Table 5: Energy scale factors to be applied to the electron energy for all Monte Carlo (MC) and real data samples, and for the two analyses.

Calculation of the integrated luminosity with $Z \to ee$ events A very important part of the CDF detector is the Cherenkov Luminosity Counters (CLC) installed in the very forward regions around the beam tube. This equipment allows the experiment to measure the luminosity with an uncertainty of 6%, of which 4.4% comes from the acceptance and operation of the luminosity monitor and 4.0% comes from the CDF I and E811 $p\bar{p}$ inelastic cross section measurements at $\sqrt{s} = 1.8\ TeV$ [40] [41]($\sigma_{in} = 59.3 \pm 2.3mb$ [39]) extrapolated to 1.96 TeV ($\sigma_{in} = 60.7 \pm 2.4mb$)[4].

This feature is extensively used by many analyses. The value of the $Z/\gamma^* \to ll$ cross section at the Tevatron, measured from CDF using the CLC and with only 72 pb^{-1} of Run II data, is:

$$\sigma(p\bar{p} \to Z/\gamma^*) \times BR(Z/\gamma^* \to ll) = 254.9 \pm 3.3(stat.) \pm 4.6(syst.) \pm 15.2(lum.)pb \text{ [44]},$$

in good agreement with NNLO theoretical calculations (see Fig. 8), which have uncertainties of 2% ($\sigma_{Z/\gamma^*} \times BR(p\bar{p} \to Z \to ll) = 251.3 \pm 5.0$ pb, following [45]).

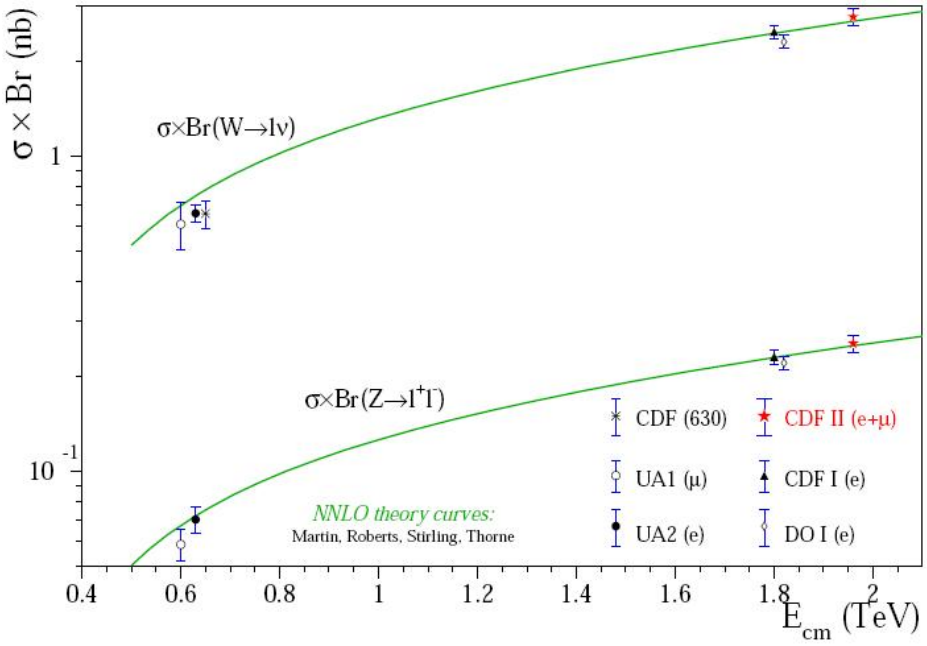


Figure 8: $W \to l\nu$ and $Z \to ll$ cross section measurements as a function of the $p\bar{p}$ center-of-mass energy, E_{cm}. The solid lines correspond to the theoretical NNLO Standard Model calculations from [45]

In the analysis of the 2002-2006 sample, we use the $Z \to ee$ events as a candle to

[4]E811 and CDF I used a luminosity independent method to measure the $p\bar{p}$ inelastic cross section in the 1.8 TeV collisions of the Tevatron: $\sigma_{in}\alpha\frac{N_{el}N_{in}}{(N_{el}+N_{in})^2}$, where N_{el} and N_{in} are the observed numbers of elastic and inelastic events. The extrapolation to 1.96 TeV follows theoretical predictions [42] [43], according to which the inelastic cross section increases with energy as $ln^2 s$.

DRAFT - 12, 2007

compute the integrated luminosity of our data samples [5].

We take a Pythia Monte Carlo sample of $Z/\gamma^* \rightarrow e^+e^-$ events simulated with the restriction that the invariant mass of the virtual boson is more than 20 GeV/c^2. The real data sample is obtained from the High p_T electron trigger, and corresponds to the two data taking periods, 2002-2004 and 2005-2006.

In order to select $Z \rightarrow ee$ events with a very low background contamination, we select the events according to the following requirements:

1. Two central electrons are found and identified as *tight* electrons, as defined in the subsection 3.3.1.
2. The two electrons must have opposite electrical charges
3. Their invariant mass is required to lie between 76 GeV/c^2 and 106 GeV/c^2

The data sample integrated luminosity (Lum) is then calculated using the following formula:

$$N^{obs} - N^{bkg} = Lum \times \sigma_{[20,\infty]}(p\bar{p} \rightarrow Z/\gamma^* \rightarrow ee) \times \epsilon_{MC} \times (1 - (1 - \epsilon_{trigger})^2) \times (\frac{\epsilon_{cem}^{data}}{\epsilon_{cem}^{mc}})^2$$

where:

- N_{obs} is the number of events passing the event selection observed in the real data sample, 1780 events in the 2002-2006 sample.
- N_{bkg} is the number of background events expected to contaminate the signal selection. This is 0.15% of the signal, that is around 25 events in the whole sample.
- $\sigma_{[20,\infty]}(p\bar{p} \rightarrow Z/\gamma^* \rightarrow ee) = 355 \times 1.4 = 497$ pb, is the cross section of the $Z/\gamma^* \rightarrow ee$ process generated with a virtual boson mass greater than 20 GeV/c^2 by Pythia. Indeed, Pythia evaluates the leading order cross section associated to this process at 355 ± 3 pb, and the corresponding K factor[6] is 1.4.
- ϵ_{MC} is the rate of $Z \rightarrow ee$ events passing the event selection in the Monte Carlo sample. We compute its value: $0.0346 \pm 0.0001_{stat.}$
- $\epsilon_{trigger}$ is the probability for a tagged electron (central with E_T >20 GeV) in a selected $Z/\gamma^* \rightarrow ee$ event to have fired the high p_T electron trigger. We use the electron trigger efficiency calculated in [47] and [46] for the top dilepton analyses, that is $\epsilon_{trigger} = 0.962 \pm 0.006$ for the 2002-2004 period and $\epsilon_{trigger} = 0.977 \pm 0.004$ for the 2004-2006 period.

[5] For the first analysis with 350 pb^{-1}, we used the CLC for the luminosity measurement.

[6] the so-called *K factor* is the factor by which a cross section calculated at the leading order needs to be multiplied in order to take into account the radiative corrections.

DRAFT OCTOBER 2, 2007

- $\frac{\epsilon_{cem}^{data}}{\epsilon_{cem}^{mc}}$ is the efficiency scale factor for the central tight electron identification, introduced in the subsection 3.4.1.1. The factor used is 0.996 ± 0.005 for the first analysis. For the second analysis, the values measured in [48] are used, that is 0.986 ± 0.004 for the 2002-2004 period and 0.977 ± 0.004 for the last period of runs.

With this method, we calculate an integrated luminosity of **1048 pb^{-1}** for our whole 2002-2006 high p_T electron sample, divided into 355 pb^{-1} for the 2002-2004 data sample and 693 pb^{-1} for the 2005-2006 data sample. This compares very well with the integrated luminosities obtained from the CLC measurements quoted here [50]: 375 ± 22 pb^{-1} and 727 ± 44 pb^{-1} for both periods respectively[7].

As the systematic uncertainty goes, we can get a rough estimate of it by adding in quadrature the systematic uncertainty associated with the $\sigma(p\bar{p} \rightarrow Z/\gamma^*) \times BR(Z/\gamma^* \rightarrow ll)$ cross section measurement quoted above (CLC luminosity uncertainty excluded) and the uncertainty on the NNLO theoretical cross section. Doing this way, we get 3%. This is thus certainly smaller than the 6% uncertainty associated to the CLC luminosity. Our analysis is still limited by statistics and not sensitive to the systematic error. For coherence with the luminosity uncertainties quoted at CDF, we choose to affect the CLC luminosity uncertainty to our luminosity measurement from the Z signal. Thus, the value taken for the integrated luminosity is 1048 ± 63 pb^{-1}

2.4.2 The muon validation with $Z \rightarrow \mu\mu$ events

In order to validate the muon identification used in this analysis, the well known $Z \rightarrow \mu\mu$ signal is used.

It is checked that a correction factor of 0.997 must be applied to the muon simulated energy in order to reproduce the Z peak in data.

Figure 9 shows the muon p_T distribution for events with two tight central muons identified in the event. The event is required to pass the cosmic veto, a requirement that is common to any selection of events relying on muons.

In order to increase statistics and to check an eventual $Z \rightarrow \mu\mu$ veto, the distributions of the muon p_T (10) is plotted for events with one tight central muon and one isolated track in the following pseudorapidity region:$(-2 < |\eta| < 2)$.

[7]Different analyses using the same good run list do not systematically get the same integrated luminosities for their samples because some files can be lost or corrupted during the data processing or the ntuple making. That is the reason why each analysis is always required to compute its own integrated luminosity.

DRAFT OCTOBER 2, 2007

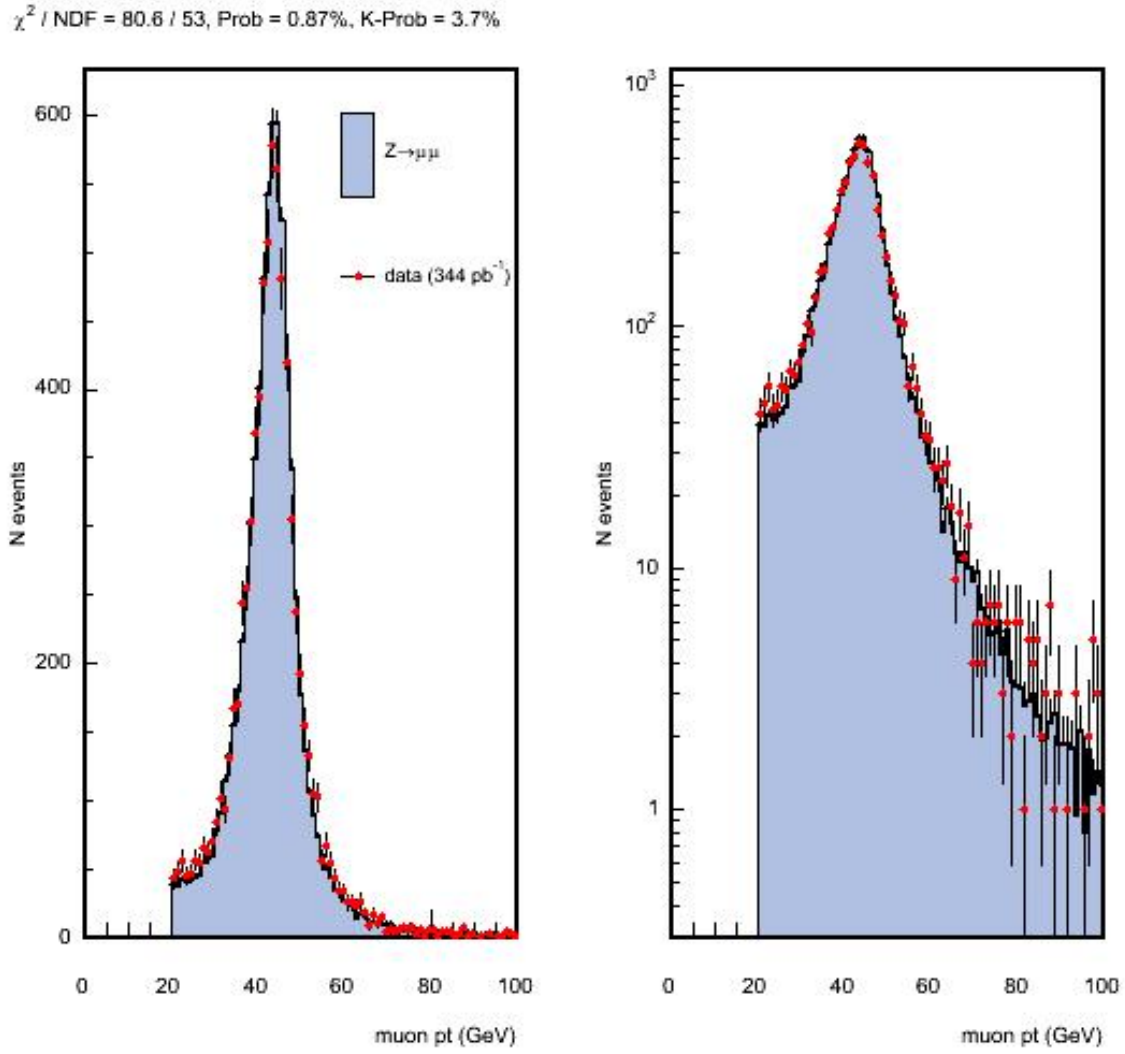


Figure 9: The muon p_T distribution in $Z \to \mu\mu$ events. Left: linear scale ; Right: log scale

DRAFT 2007

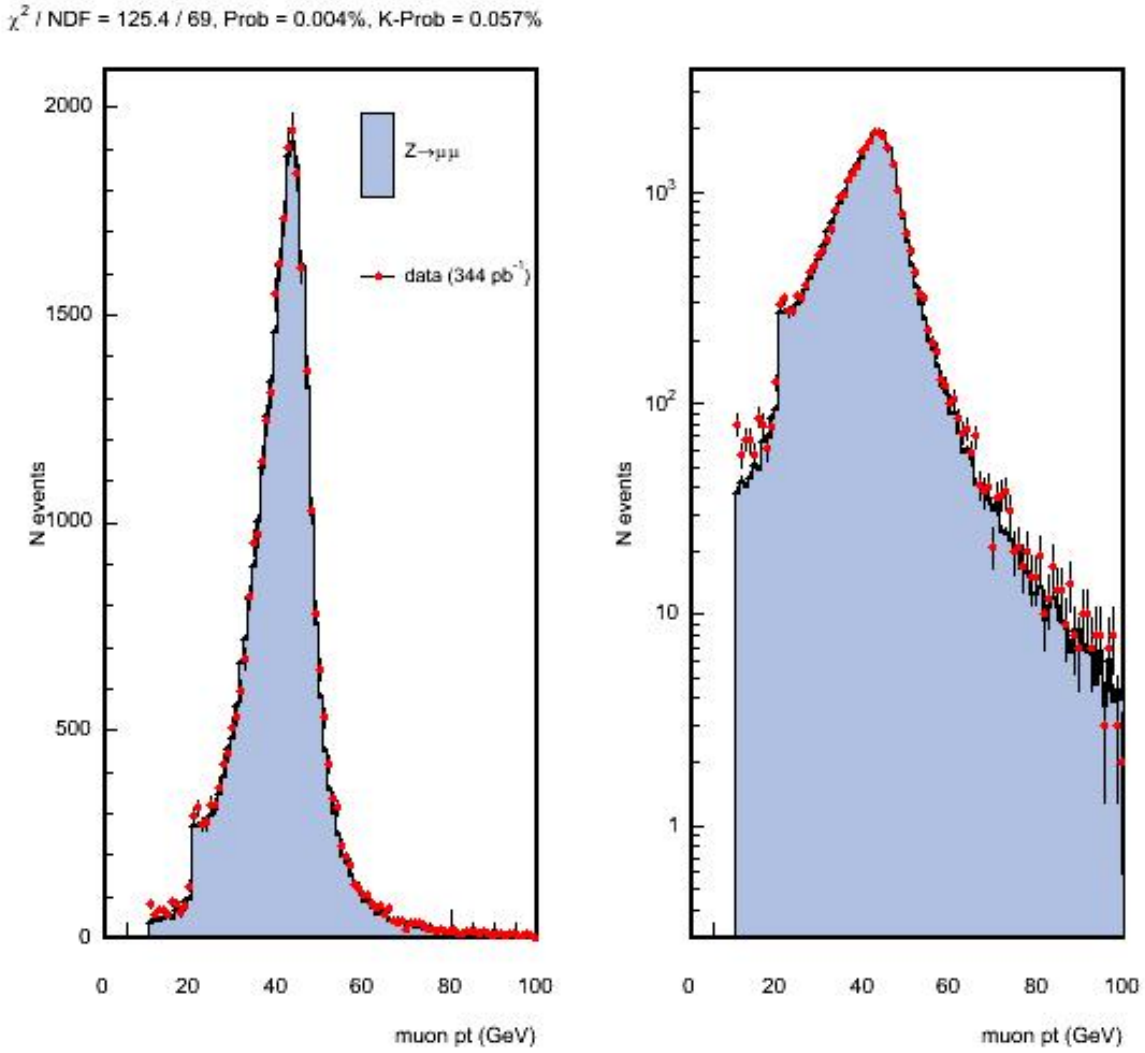


Figure 10: The muon p_T distribution in $Z \rightarrow CMUP/X + track$ events. Left: linear scale ; Right: log scale

DRAFT 07

2.4.3 The tau lepton validation with $Z \rightarrow \tau\tau$ events

In order to check that the identification of the hadronically decaying tau leptons are under control in the analysis, the $Z \rightarrow \tau\tau$ signal is extracted from the 1 fb^{-1} data sample. Since the Z decays into two hadronically decaying tau leptons are impossible to extract from the overwhelming dijet QCD background, one of the two tau leptons from the Z decay is required to decay into hadrons while the other tau lepton is required to decay leptonically.

A $Z \rightarrow \tau\tau$ analysis, to which we participated, has already been performed at CDF [52] [53]. It uses a specific trigger that has been developed at CDF, the so-called lepton+track trigger [51], which allows to select events with lower p_T electrons or muons, namely with p_T as low as 10 GeV/c. However, the tool that we developed to estimate the background of jets faking tau leptons in our top analysis (see subsection 3.8) cannot be used on data selected with this trigger because of the bias on the tau leg introduced by triggering on "tau-like" isolated tracks. Thus, instead, the $Z \rightarrow \tau\tau$ analysis described in this subsection makes use of the high p_T electron trigger.

The event selection applies the following series of cuts:

- One central *tight* electron is required to be identified in the event. The electron p_T threshold is set at 20 GeV/c in order to satisfy the trigger requirement.
- One identified central tau lepton decaying hadronically is found
- The ΔR angle difference in the $\eta - \Phi$ space between the tau and the electron is greater that 2.4
- The electron and the tau lepton must have opposite electrical charges
- $Z \rightarrow ee$ veto: No *loose* electron (cf 3.4.1.1) can be found in the event so that the invariant mass of the two electrons is comprised between 66 GeV/c^2 and 116 GeV/c^2.
- $W \rightarrow e\nu$ veto (see Fig.11 extracted from [53]): $W_{p_T} > 24$ GeV/c or $W_{M_T} > (50 - 1.25 \times W_{p_T})$, where $W_{p_T} = |\vec{p_T^e} + \vec{\not{E}_T}|$, and $W_{M_T} = \sqrt{2 \times p_T^e \not{E}_T \times (1 - cos(\Delta\phi))}$. $\Delta\phi$ is the 2D angle in the $r - \Phi$ plane between the electron track and the missing transverse energy vector.
- $W_{M_T} < 50$ GeV/c^2

We observe 583 events in the real data sample, while the sum of the predictions for the Z signal and the backgrounds (jets and electrons faking hadronic tau decays) is 610 events. These 610 expected events divide into 173 events from jets faking taus and 437 events from $Z \rightarrow \tau\tau$. We find no contribution from $Z \rightarrow ee$ events with one electron faking a tau lepton ; the reason for this is our choice of a very tight Z veto. Fig.12 compares the observed and predicted tau lepton p_T distributions. The $Z \rightarrow \tau\tau$ signal prediction is obtained from a Pythia sample and the background from jets faking tau

DRAFT OCTOBER 2, 2007

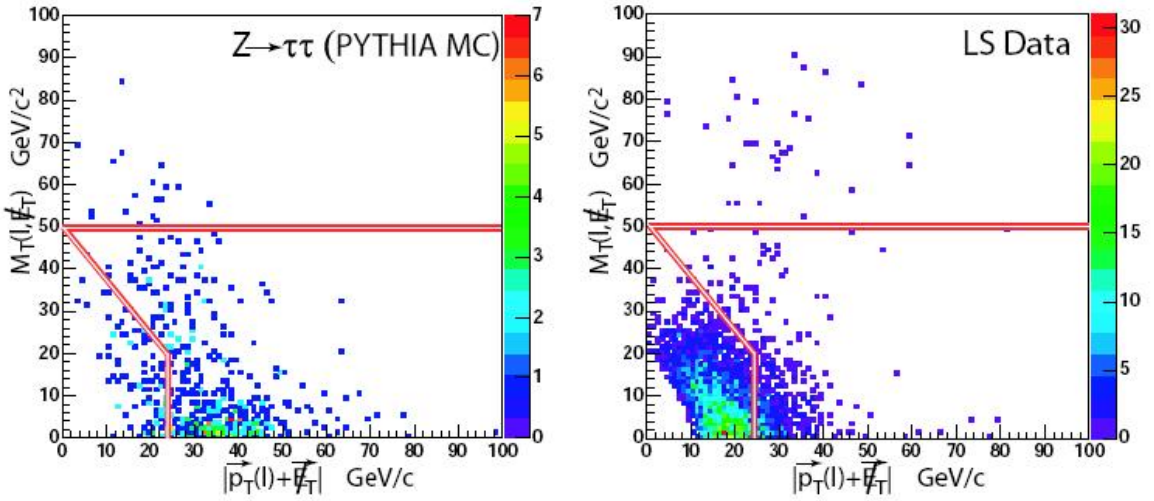


Figure 11: Distribution of M_T versus p_T for: left)$Z \rightarrow \tau\tau$ signal events from Pythia ; right) Real data events passing all cuts except that the two leptons have the same electric charge (dominated by QCD, $\gamma + jets$ and $W + jets$ backgrounds). [53]

leptons is the output of our jet to tau fake rate decribed in details in the subsection 3.8. The uncertainty associated to it is 16%. After having subtracted the background, the "observed" number of $Z \rightarrow \tau\tau$ events is thus $583 \pm 24_{stat.} - 173 \pm 28_{sys.} = 410 \pm 24_{stat.} \pm 28_{sys.}$. This is in good agreement with our prediction of 437 $Z \rightarrow \tau\tau$ events within the systematic error of the fake rate. Thus we consider that the Monte Carlo efficiency tor the tau identification agrees with the one observed in data, that is to say that the tau identification scale factor is given a central value of 1. As the systematic error goes, we need to take into account the high statistical error associated to this measurement because of the rather low number of events observed. As a conclusion, the scale factor for the tau lepton identification is calculated to be $\epsilon_{\tau_{ID}} = 1.0 \pm 0.07_{stat.} \pm 0.06_{sys.}$. The uncertainty associated to the tau lepton identification efficiency is thus 9%.

We conclude that we are able to extract the $Z \rightarrow \tau\tau$ signal and estimate its amplitude with Pythia simulations.

DRAFT OCTOBER

2007

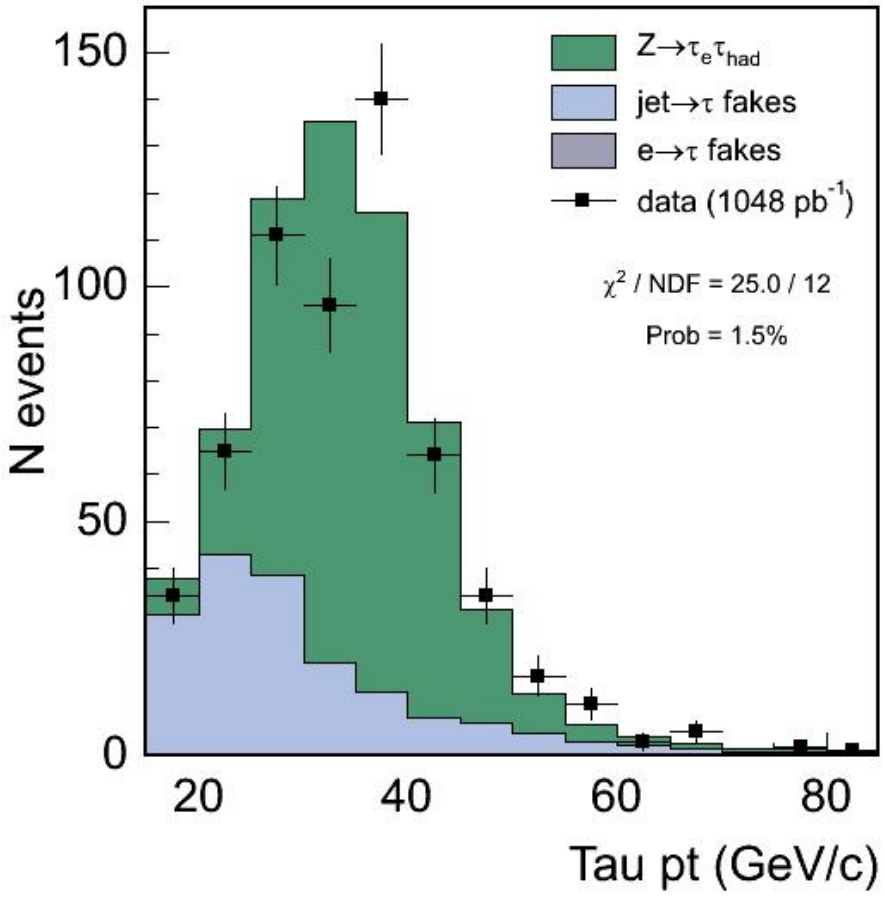


Figure 12: The tau p_T distribution in $Z \to \tau\tau$ candidate events

DRAFT

2.4.4 Validation of the missing transverse energy with $W \to \mu\nu$ events

Once the simulated muons are tuned for energy and identification efficiency, an extraction of the $W \to \mu\nu$ signal can be performed in order to make sure that the presence of neutrinos in an event is well taken into account by the missing transverse energy (MEt) calculated in the muon channels. In fact, all Monte Carlo samples used in this analysis have some true missing E_T, namely: $Z \to \tau\tau$, diboson and top signal sample.

The missing transverse energy is computed by summing the transverse energy vectors of all the towers in the calorimeter and using the highest p_T vertex as the origin.

The missing transverse energy is corrected for the mismeasurement of jets in the calorimeter using the standard jet energy correction algorithm at level 5 (cf 3.4.5). Are excluded from the list of jets all the objects with energies lower than 8 GeV or the ones that match an identified muon or tau, or the ones with an electromagnetic fraction higher than 90%.

The missing transverse energy is also corrected for the presence of muons in the event. For every central muon passing the tight muon identification cuts, the calorimeter energy is replaced by the muon track energy. Note that this implies that MEt calculation can be flawed by the presence of non central or of any minimum ionizing particle which is not identified as a tight muon in the event. The reason for not correcting for these objects is that the simulation framework used here is not able to reproduce well enough the rate of fake muons, especially in the forward regions of the detector where the track density is higher and where there is only the silicon standalone tracking.

The $W \to \mu\nu$ selection requires the presence of a tight central CMUP muon, a missing $\mathrm{E}_T > 20$ GeV, and no other CdfMuon in the event (Z veto).

Figure 13 shows the comparison between the missing E_T as extracted from data and the simulated one.

The prediction made up by summing the simulations of $W \to \mu\nu$, $W \to \tau\nu$ and $Z \to \mu\mu$ events reproduces the data. The low energy tail was shown only to underline the QCD contribution.

DRAFT OCTOBER 2, 2007

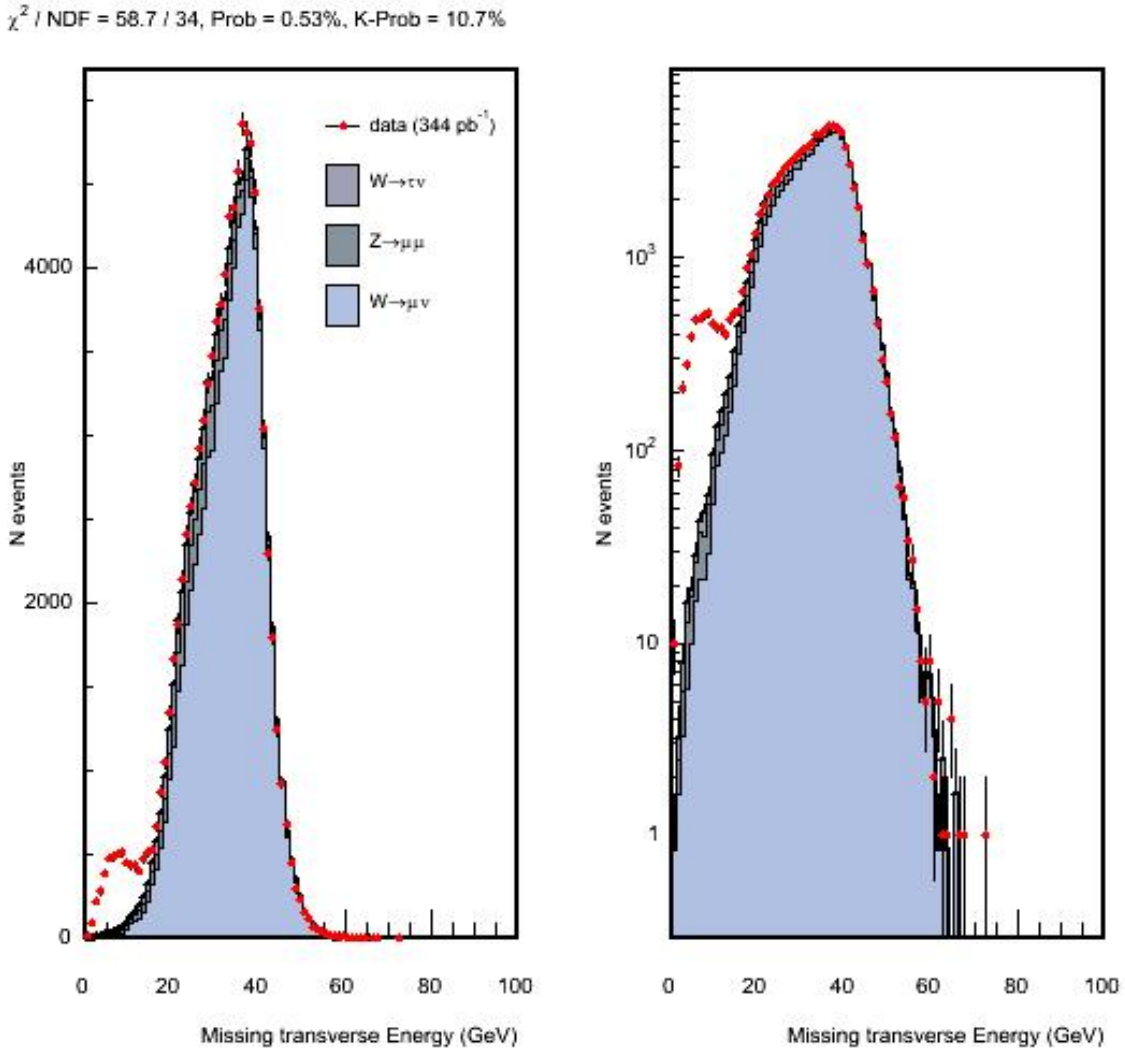


Figure 13: The MEt distribution from $W \to \mu\nu$ events. Left: linear scale ; Right: log scale

DRAFT 2007

2.4.5 The jet energy correction factors

The jet energy is computed by summing the transverse energy of all the calorimeter towers calculated at the primary vertex of the event. These towers are the ones included in the jet cluster based in this analysis on the so-called *cone algorithm*[8], although it might be noted that CDF has also performed detailed analyses on jet properties based on the kT algorithm[9] [36]. Whereas the end plug calorimetry was rebuilt for Run II, the barrel calorimetry remains unchanged from the point of view of the detector itself. However all the front-end and readout electronics have been changed as well as all the calibrations and the monitoring systems that survey and calibrate each component of this detector. A system of on-line calibrations and corrections is applied to correct for the variations in the functioning of the detector and of its associated electronics (pedestal, dead or noisy channels etc..). Apart from these on-line tasks, there is a constant work to correct the reconstructed energy and validate the corresponding data at the data handling and processing levels.
Another important issue for all the physics analyses that include jets as in particular our analysis, is to get a good agreement between the Monte Carlo simulations and the data for what concerns the reconstructed jet energy. In fact the Monte Carlo simulations start from the parton level whereas the jet reconstructed from the data start from just the opposite side, i.e. the calorimeter towers. How to correctly link these two sides is a main issue. The path to follow from the parton jet to the calorimeter jet is schematized in Fig.14.

The jet energies computed by the Monte Carlo are tuned to agree with the jet energies in jets from reference samples for some well-known physics events, such as: J/Ψ, Z peaks, minimum bias events, etc. Correction factors are then computed in order to cope with the discrepancies between those real data and the corresponding Monte Carlo data.
A task force was conducted more than a year ago to reinforce the work on this issue. It was mainly driven by the important physics goal to achieve the best possible estimate on the top mass (see Section 1). The main parameter damaging this estimate was recognized to be the so-called jet energy scale. This is the factor that allows adjusting at best the reconstructed jet energy and ensures a good agreement between data and Monte Carlo. The result of this work [37] is summarized in the Fig.16. It shows that, at this stage, the correction is applied following four steps to the jets :

1. The Eta-dependent corrections: This scales jets outside the $0.2 < |\eta| < 0.6$ region to jets inside the region, depending on η and p_T of the jet. This eta range is chosen since it is far away the cracks or non-instrumented regions.

[8] The cone algorithm forms jets by associating calorimeter towers centered within a radius of 0.4 in the $\eta \times \phi$ space. The algorithm starts with a trial on the cone geometrical center that can be any tower with a transverse energy greater than 1 GeV, and it computes the cone centroid. If the calculated centroid is aligned with the geometrical center of the cone, the cone is labelled as stable, and is kept in the list of jets. The algorithm continues to run until a stable solution is found.

[9] The approach of the kT algorithm [35] consists in merging pairs of towers following an increasing order in transverse momentum.

DRAFT OCTOBER 2, 2007

2. Multiple interactions correction: This correction (UEM) subtracts the energy contribution in the jet cone from eventual additional $p\bar{p}$ interactions in average. This is particularly useful for high energy runs when the average number of interactions per event becomes higher (see Fig. 15).

3. *Absolute*: This is the name given to the correction to the jet energy measured in the calorimeter for any non-linearity and energy loss in the un-instrumented regions of each calorimeter.

4. Underlying event correction: This correction subtracts the energy associated with the spectator partons that falls inside the jet cone.

The jet correction applying these four steps sequentially is called "level 5" correction. The total systematic uncertainties for jets with corrected transverse energies above 15 GeV are found to lie between 3 and 8%, depending on the energy (see Fig.16).

DRAFT OCTOBER 2, 2007

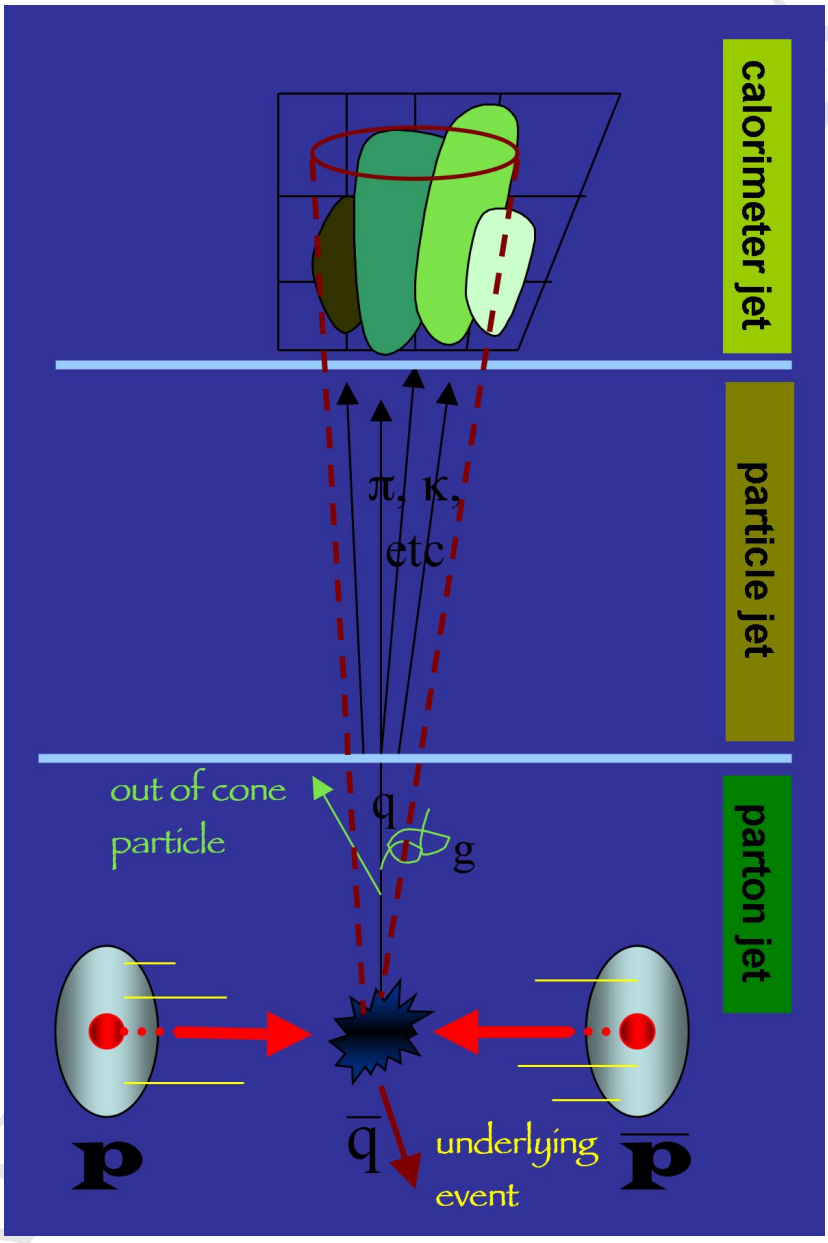
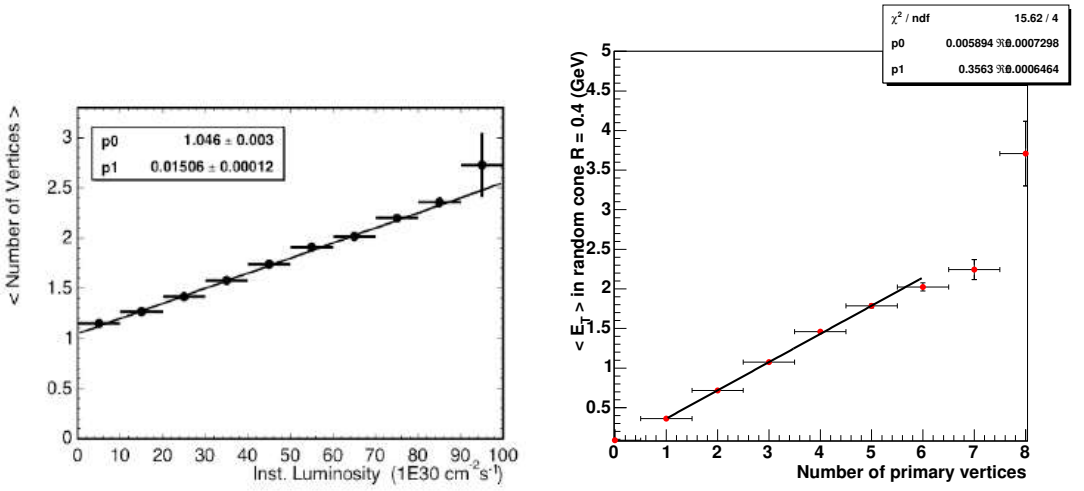


Figure 14: The schematical path from a parton jet to a calorimeter jet

DRAFT 2007



(a) *The increase in the number of event vertices with respect to the instantaneous luminosity increase*

(b) *Average energy to be subtracted from the jet cone as a function of the number of event vertices*

Figure 15: *Jet energy correction due to multiple interactions*

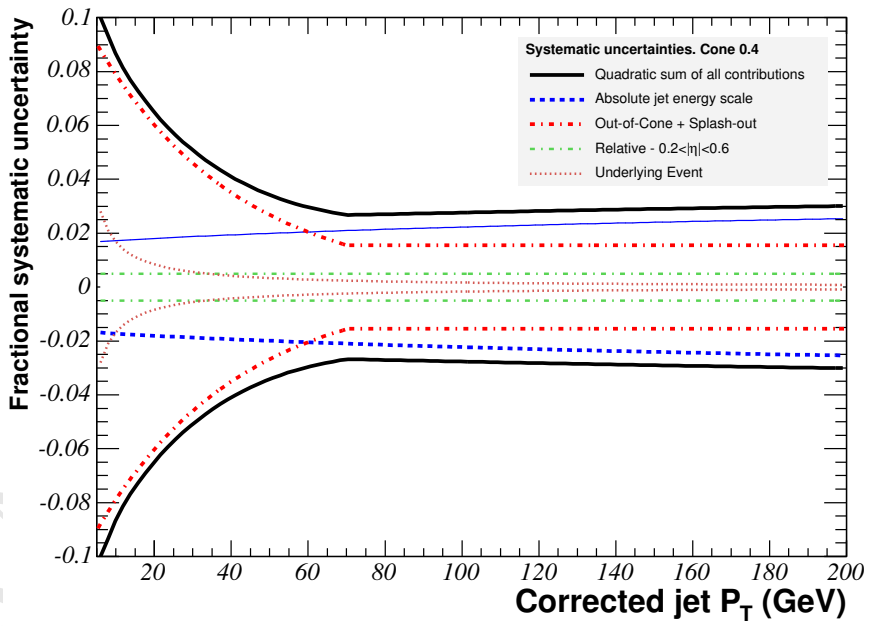


Figure 16: Systematic uncertainties on the jet energy correction

2.5 The event selection strategy and the acceptance applied for the 1 fb^{-1} analysis

2.5.1 Method used to reduce the $Z \to \tau\tau+$ 2 jets background

Here we describe the way our Z veto is built. This has not changed since the 350 pb^{-1} analysis, except in the way this Z veto is used in the event selection. Whereas in the previous analysis, events failing the Z veto were cut out, in the 1 fb^{-1} analysis, the Z veto is used as a variable inside a likelihood method described later.

An efficient cut must be applied to the $Z \to \tau\tau$ background, which would be otherwise the highest background in the analysis. It is based on the kinematical characteristics of these events. The goal is to build a Z veto that harms the top signal as little as possible. This part of the analysis is only based on Monte Carlo simulated events, as the kinematics of these events is well reproduced by the Monte Carlo.

Angular configuration of $Z \to \tau\tau+$ 2 jets events The requirements of at least two extra jets, a high missing transverse energy ($\not\!\!E_T > 20$ GeV) and a high scalar sum of transverse energy (Ht), select events with a high Z boost, where the tau leptons are emitted close to each other in the laboratory frame. Conversely, the $t\bar{t}$ dilepton events rather favour two back to back leptons. Therefore the angle difference between the two leptons serves here as a discriminant variable.

The $Z \to \tau\tau$ events considered in this analysis have one tau decaying into an electron or a muon with $E_T > 20$ GeV accompanied with two neutrinos, and the other tau decaying into one or more hadrons with a total $E_T > 15$ GeV along with one neutrino. Because of the high boost of the two tau leptons, the hadronical decay of the tau produces so-called narrow jets. For the same reason, the neutrinos from the tau decay are colinear to the tau.

The kinematical event topologies depend on the spins of the tau leptons coming from the Z boson decay. Let us label Jz as the projection of the spin J on the Z boson direction. In what follows: R means right-handed, L means left-handed and tau decays into X+ν_τ):

- Case 1: Z Jz=+1 OR -1, therefore (tau+ is R and tau- is L) OR (tau- is R and tau+ is L)
- Case 2: Z Jz=0, therefore (tau+ is L and tau- is L) OR (tau- is R and tau+ is R)

This leads to three different kinematical configurations, that can be qualitatively described as follows:

- Configuration 1 corresponds to a right-handed tau+ and a left-handed tau-, and is characterized by:
  1. the two tau neutrinos are emitted along the same directions of the two taus

DRAFT OCTOBER 2, 2007

2. therefore giving a high true missing E_T between the two taus (Fig 18)

- Configuration 2 corresponds to a right-handed tau- and a left-handed tau+ and is characterized by:

  1. the two tau neutrinos are emitted in the direction opposite to the two taus
  2. because of the tau boost, the energy of the tau neutrinos is small, leading thus to a small true missing E_T
  3. therefore the total missing E_T which passes the cut of 20 GeV is due to mismeasurements of the two jet energies, adding one to each other
  4. therefore MEt is sitting between the 2 jets (if the jet E_T is underestimated) or opposite to the two jets (if the jet E_T is overestimated).
  5. Since the two jets are back to back with the Z boost, MEt is thus opposite to the 2 taus (fig.17) or in between (fig.18)

- Configuration 3 corresponds to the case where the two taus have the same helicity (Jz=0) and is characterized by:

  1. one tau will emit its neutrino frontwards and the other one will emit its neutrino backwards
  2. therefore leading to some true missing E_T between the two taus

It should be noted that the leptonic decay of the second tau leads to some sort of corrections to the simplified picture described here above. However they do not significantly modify the main conclusion which is that if the missing energy is high enough (ie $\not\!\!E_T > 20$ GeV), it most likely points on the same direction as the two taus (most probable case) (fig.18), or opposite to them(fig.17).

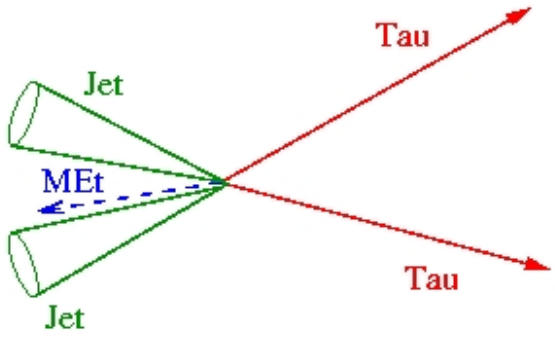


Figure 17: Configuration with missing E_T opposite to the two tau leptons

In order to define this in a quantitative way, two Monte Carlo samples are used. One is the $Z \rightarrow \tau\tau + 2$ jets sample, simulated with AlpGen plus Herwig and the other

DRAFT OCTOBER 2, 2007

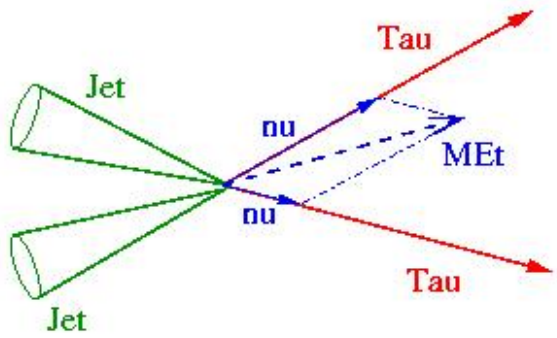


Figure 18: Configuration with missing E_T between the two tau leptons

one is the top signal sample simulated with Pythia. Figures 19 and 20 are based on these two samples respectively. They show the observed angular connections between the two leptons and the $\not\!\!E_T$ directions in the Z and top events.

In order to isolate the $Z \rightarrow \tau\tau + 2$ jets events from the top events, it is first required that the identified tau and lepton (electron or muon) are close to each other. Because of a lack of statistics of our Monte Carlo samples, the cut applied on the difference in azimuthal angle ($d\phi$) cannot be optimized. It is simply defined by the condition: $d\phi(\tau$, lepton) < 1.9 rad. This condition is derived from the results plotted in figures 19 and 20. Moreover, from the discussion about the missing tranverse energy direction just above, a second condition is imposed, namely: the total missing transverse energy $\not\!\!E_T$ must point between the tau and the lepton (see the so-called angular sector A in the Figures 19 and 20) or must be opposite to them (see the so-called angular sector B of Figures 19 and 20).

The Z mass cut on the subsample of selected events This subsample is defined as made of the subsamples A and B (as shown in Figures 19 and 20). The cut of all these events would be too harmful for the top signal. We try instead to reconstruct the Z mass in order to better separate this background from the signal. As shown in Fig.18, the missing E_T can be projected onto the τ-lepton axes in order to approximate the neutrinos transverse energies. This can be done for the events in which the missing transverse energy points between the two τ-leptons. Once done, the mass of the τ lepton pair can be reconstructed. It is expected to peak around the Z-mass in the $Z \rightarrow \tau\tau$ events.

The Fig.21 shows the reconstructed Z-mass for the Z and top events. A cut is applied at 115 GeV. It cannot be much improved because of a lack of statistics of our Monte Carlo sample. Cutting events with a mass smaller than 115 GeV leaves 10% of Z-events from subsamples A and B.

To summarize, the events that fulfill the conditions listed here below don't pass the condition of the Z veto:

DRAFT OCTOBER 2, 2007

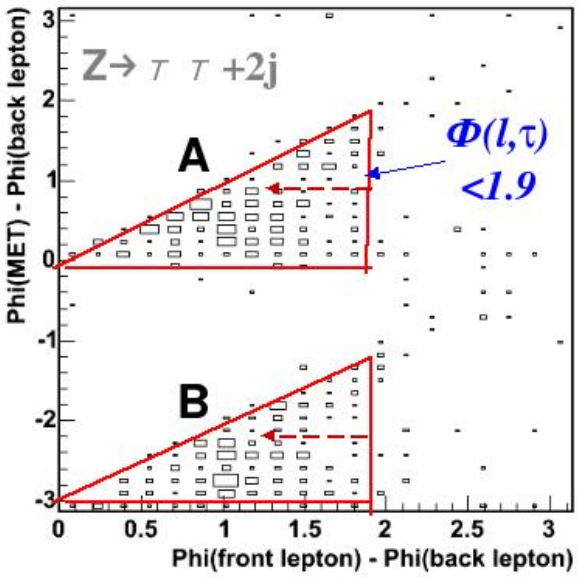


Figure 19: Angular relations between the two leptons and missing E_T : Z background case. The triangle A corresponds to $\not\!\!E_T$ pointing between the tau and the lepton, and the triangle B corresponds to $\not\!\!E_T$ back to back with the tau and the lepton. The *front* and *back* adjectives refer to the relative angular direction of the two leptons in the directly oriented $r - \Phi$ plane.

- The difference in azimuthal angle between the tau and the other lepton (electron or muon), dPhi(tau, lepton), must be less than 1.9 rad.
- The total missing transverse energy, $\not\!\!E_T$, must stay in between or opposite to the two identified leptons.
- The transverse mass of the tau plus the other lepton and $\not\!\!E_T$ must be less than 115 GeV.

To get rid of events failing the Z veto (as was done in the 350 pb^{-1}) would eliminate 70% of the Z-background and only 8% of the top into tau dilepton signal.

DRAFT OCTOBER 2, 2007

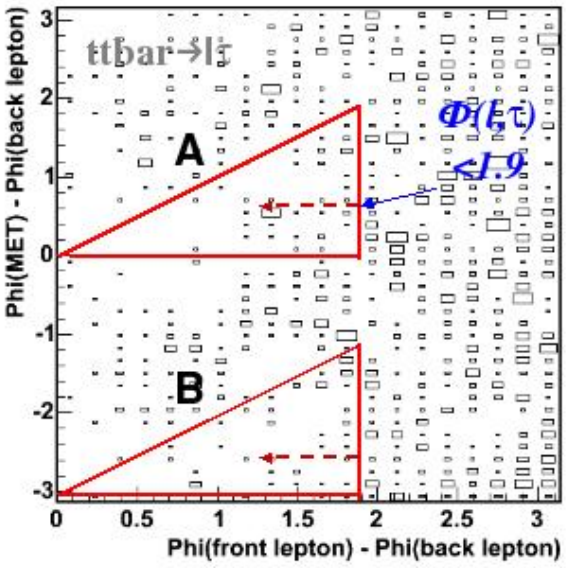


Figure 20: Angular relations between the two leptons and the missing E_T : $t\bar{t}$ case. The triangle A corresponds to $\not\!\!E_T$ pointing between the tau and the lepton, and the triangle B corresponds to $\not\!\!E_T$ back to back with the tau and the lepton

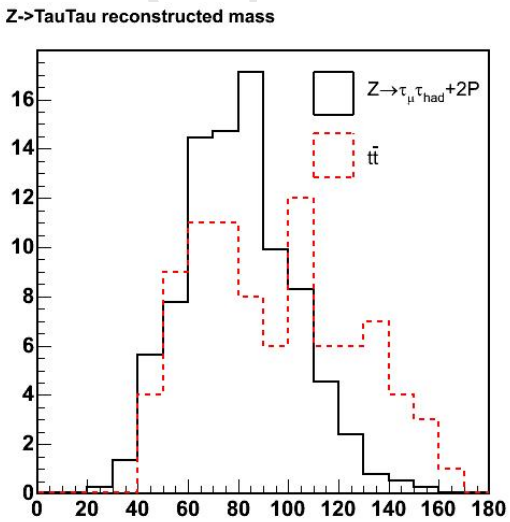


Figure 21: The reconstructed mass of the Z-boson in $Z \rightarrow \tau\tau$ + jets. Comparison with top into tau dilepton events for events passing the A-B preselection. Note that this plot gathers A+B regions, and thus is worse for $Z \rightarrow \tau\tau$ than it would be for the region A alone.

DRAFT - OBER , 2007

2.5.2 Discussion about the possible use of b-tagging in this analysis

The presence of two b-quark jets in the signal events is accounted for in our selection by the requirement of two jets with pseudorapidities between -2 and 2 and transverse energies greater than 15 GeV. We do not profit from the b-flavour of the quarks that can be tagged using the relatively high lifetime of the produced B meson ($c\tau \sim 500\mu m$) inside the jet. Thanks to a sophisticated vertex detector and the associated trigger system the CDF detector is able to tag b-quarks by measuring impact parameters with a high precision and in realtime with a sophisticated second level triggering system SVT. Displaced vertex from the primary vertex of the event are thus identified already at this early stage in the trigger system. Several b-tagged jet algorithms were worked out in CDF profiting from all these detector unique capabilities.

A requirement of at least one b-tagged jet has been applied in the case of a top dilepton cross section measurement at CDF with 750 pb^{-1} of data [30] ("dilepton" = ee, $e\mu$, $\mu\mu$) . The same selection as for the CDF top dilepton cross section measurement described in [31] was used, before applying b-tagging. It is shown that with an identification efficiency of at least one b-tagged jet in top dilepton events of 55%, a rejection factor of around 92% of the non-top background can be obtained. For a $t\bar{t}$ cross section of 6.7 pb, the numbers of signal (S) and background (B) events expected in 750 $pb^{-}1$ vary from (S=36 ; B=19) to (S=20 ; B=1.5).

A background drop of 92% at a cost of 45% for the signal increases the significance $\frac{S}{\sqrt{B}}$ for the generic $t\bar{t}$ dilepton signal (ee, $e\mu$, $\mu\mu$) by 77%. Unfortunately, this cannot be extrapolated to the $t\bar{t}$ tau dilepton signal for several reasons. The main reason is the high probability for a central jet to be wrongly identified as a central tau lepton decaying hadronically. Such a jet that passes the tau identification is called a "fake" tau, and the probability for this to happen is called the "jet to tau fake rate". The jet to tau fake rate has been proven to lie between 0.5 and 1% for a generic central jet, depending on the jet E_T and the energy density in the event. This is around ten times bigger than the probability for a central jet to be identified as an electron. Because of this high jet to tau fake rate, the $t\bar{t} \rightarrow l\nu jjbb$ becomes a dominant background in the top in tau analysis whereas this background was small in the top dilepton analysis with electrons and muons only, and this background is not decreased by the requirement of a b-tagged jet in the event. The $t\bar{t} \rightarrow l\nu jjbb$ background accounts for around one fourth of all backgrounds present after the 350 pb^{-1} analysis selection is applied. Indeed, instead of a 92% background decrease, the use of b-tagging would yield to a less efficient decrease of around 80% at a cost of 45% for the signal acceptance, resulting to a smaller increase of the significance $\frac{S}{\sqrt{B}}$ by about 20%.

This eventual gain of 20% would be obtained at a high cost; the simultaneous treatment of jets faking taus and of jets mistagged as b-quark jets would result to an important complication of the background estimation. We would be enforced to rely on not well known quantities such as heavy flavour fractions in the events with a high jet multiplicity and a high activity Ht, and this would increase the dependence on the behaviour of the jet to tau fake rate in $t\bar{t} \rightarrow l\nu jjbb$ events. However, the jet to tau

DRAFT OCTOBER 2, 2007

fake rate can be validated in W+jets events (see subsection 3.8), but, at least with the present integrated luminosity, it cannot be validated in $t\bar{t} \rightarrow l\nu jjbb$ events. It could be argued that we arbitrarily chose not to use b-flavour tagging of one of the two b-quark jets in the signal events, although we use a tight identification of the tau lepton decaying hadronically. The answer to this is that our goal is not to measure the top cross section but to observe the top decay into $\tau\nu q$ in order to eventually test the branching ratio ($t \rightarrow \tau\nu q$). This imposes to well identify the tau leptons.

Finally, let us emphasize again that the major motivation of this analysis is to achieve the best possible measurement of $r_\tau = \frac{t\rightarrow\tau\nu_\tau b}{t\rightarrow l\nu_l b}$, and to validate it. The sensitivity to new physics and the interest for measuring the ratio $r_\tau = \frac{t\rightarrow\tau\nu_\tau b}{t\rightarrow l\nu_l b}$ (l = e or μ) leads to the fact that the figure of merit is given here by $\frac{S}{\sqrt{S+B}}$. For this issue, the use of b-tagging does not really help.

2.5.3 The event selection

In order to gain in acceptance, some cuts are relaxed as compared to the 350 pb^{-1} analysis. The requirement for the highest E_T jet to have a transverse momentum greater than 25 GeV/c is not applied anymore and the Ht cut is decreased from 205 GeV down to 160 GeV. A new discriminant variable is built to overcome the resulting increase in the background. It is based on a likelihood method.

The 1 fb^{-1} event selection (defined in **??**) is based on the following set of requirements:

1. One central isolated electron or muon with a transverse momentum greater than 20 GeV/c.

2. One central isolated tau lepton with a transverse energy greater than 15 GeV, opposite in charge to the first lepton.

3. At least two jets, each with a pseudorapidity between -2 and 2, and with a transverse energy greater than 15 GeV

4. A missing transverse energy greater than 20 GeV

5. An activity characterized by Ht greater than 160 GeV

6. A Likelihood-based discrimination

2.5.4 Construction of a discriminant likelihood variable

The idea is to combine several variables to make up a unique variable with a better discriminating power. The likelihood variable made up from n individual variables is defined as the product Π Sn/Bn, where Sn is the signal distribution of the n^{th} variable and Bn is the background distribution of the n^{th} variable.

Four Monte Carlo samples are used to make predictions, namely:

DRAFT OCTOBER 2007

- Pythia is used for the signal and the $t\bar{t}$ →l+jets events. The top mass is fixed at 175 GeV.
- A $W \rightarrow e\nu + 3P$ Alpgen+Pythia sample is used to reproduce W+jets events.
- An Alpgen+Herwig $Z \rightarrow \tau\tau + 2P$ sample is used to reproduce the Z+jets events.

Each Monte Carlo sample uses the same run-dependent scheme as already used in the 350 pb^{-1} analysis, with minimum bias events added to each event according to the known luminosity of the corresponding event run. The Monte Carlo samples were run by the top and the electroweak working groups in CDF.

In order to calculate the number of events with an identified tau lepton in the $t\bar{t}$→l+jets and in the W+jets samples, the same jet to tau fake rate matrix as the one developed for the 350 pb^{-1} analysis is used to weight the Monte Carlo events. The final number of events predicted for the sum of $t\bar{t}$→l+jets and of W+jets is then scaled to be identical to the number of fake tau events computed in the 1 fb^{-1} data sample. This last step is necessary because the Monte Carlo does not well estimate the rate of jets passing the tau denominator selection as used in the jet to tau fake rate definition.

Selection of variables A set of ten kinematical variables are selected. They are chosen for their abilities to discriminate the top in tau signal against the three major backgrounds which are the Z, the W and the $t\bar{t} \rightarrow$ l+jets physics processes. Here below is the list and definition of these ten parameters:

1. The event mass: This is the invariant mass of the 4-vector made of the sum of the energy-momentums of the electron, the tau, the 2 jets, and the transverse missing energy. Ideally, this should be close to the double of the top mass in the case of top events, and much lower for the Z and W backgrounds. This discriminates against Z and W.
2. The Ht parameter: This variable was defined in the 350 pb^{-1} analysis. It is correlated with the event mass. This discriminates against Z and W events.
3. The ratio $\frac{\sum p_T}{\sum Ez}$: This is the sum of the transverse momentums of the light lepton, the tau, the two jets, divided by the sum of the z-components of the energies of the same objects. This is a measure of the centrality of the event. Top events are known to be more central because of the very high top mass. This discriminates against Z and W.
4. The missing transverse energy ($\not\!\!E_T$): Top events have large missing transverse energies whereas $Z \rightarrow \tau\tau$ events have a small missing transverse energy. This discriminates against Z.
5. The lepton-tau azimuthal angle difference: This variable was studied for the 350 pb^{-1} analysis's Z-veto. This discriminates against Z.

DRAFT OCTOBER 2, 2007

6. The Z-veto summary: This variable summarizes the result of the Z-veto procedure already described for the 350 pb^{-1} selection. This has four possible output values:

   - If the event doesn't pass the Z-veto, the Z-veto summary is given the value 1 (respectively -1) if the $\not\!\!E_T$ points between (respectively opposite to) the two tau lepton diections.
   - If the $\not\!\!E_T$ points neither between nor opposite to the two tau leptons or the angle between the two tau leptons is larger than 1.9 rad, then this variable has the value 0.
   - If the event corresponds to the configuration A or B of the figure 19 and the invariant mass $M(e, \tau, \not\!\!E_T)$ is larger than 115 Gev/c^2, then the variable is given the value 2.

   The Z-veto summary variable discriminates against the Z background.

7. The sum of the transverse energies of the two highest E_T jets: This discriminates against Z and W.

8. The transverse mass of the (lepton, $\not\!\!E_T$) system: This is an attempt to reconstruct the W mass in the W and $t\bar{t}$ →l+jets events, where the missing E_T is due to a unique neutrino. $Z \rightarrow \tau_e\tau_{had}+jets$ events often have their neutrinos emitted close to the light lepton as it was explained in the Z-veto section. Thus Z background events usually have small (lepton, $\not\!\!E_T$) transverse mass. This variable efficiently discriminates against all three backgrounds.

9. The total number of jets: Selected W, Z and top in tau events have most of the time a minimum number of jets required, i.e. two, whereas $t\bar{t}$ →l+jets usually have three jets in addition to the fake tau. This discriminates against $t\bar{t}$ →l+jets.

10. Number of tracks in the tau-jet: The number of tracks in the tau-jet is required to be either 1 or 3 by the tau identification algorithm. It is mostly 1 if the tau is true, whereas it is most of the time 3 when the tau is false. This discriminates against fake taus, mainly due to W and $t\bar{t}$ →l+jets backgrounds.

Fig.22 presents the 10 distributions for the signal and the three backgrounds.

Likelihood ratio For each of the ten variables k, from $k = 1$ to 10, a likelihood ratio distribution L_k is defined as the ratio of the variable probability density distribution for the signal over the distribution for the sum of the backgrounds against which the variable k is meant to discriminate, as indicated explicitely for each of the ten variables described above. For instance, in the case of the missing E_T variable, the sum of the backgrounds is made of the Z and W distributions, each weighted by their corresponding cross section. This way of computing the sum of the backgrounds allows to decrease in a more even way all backgrounds, no matter how important their contribution is

DRAFT OCTOBER 21 2001

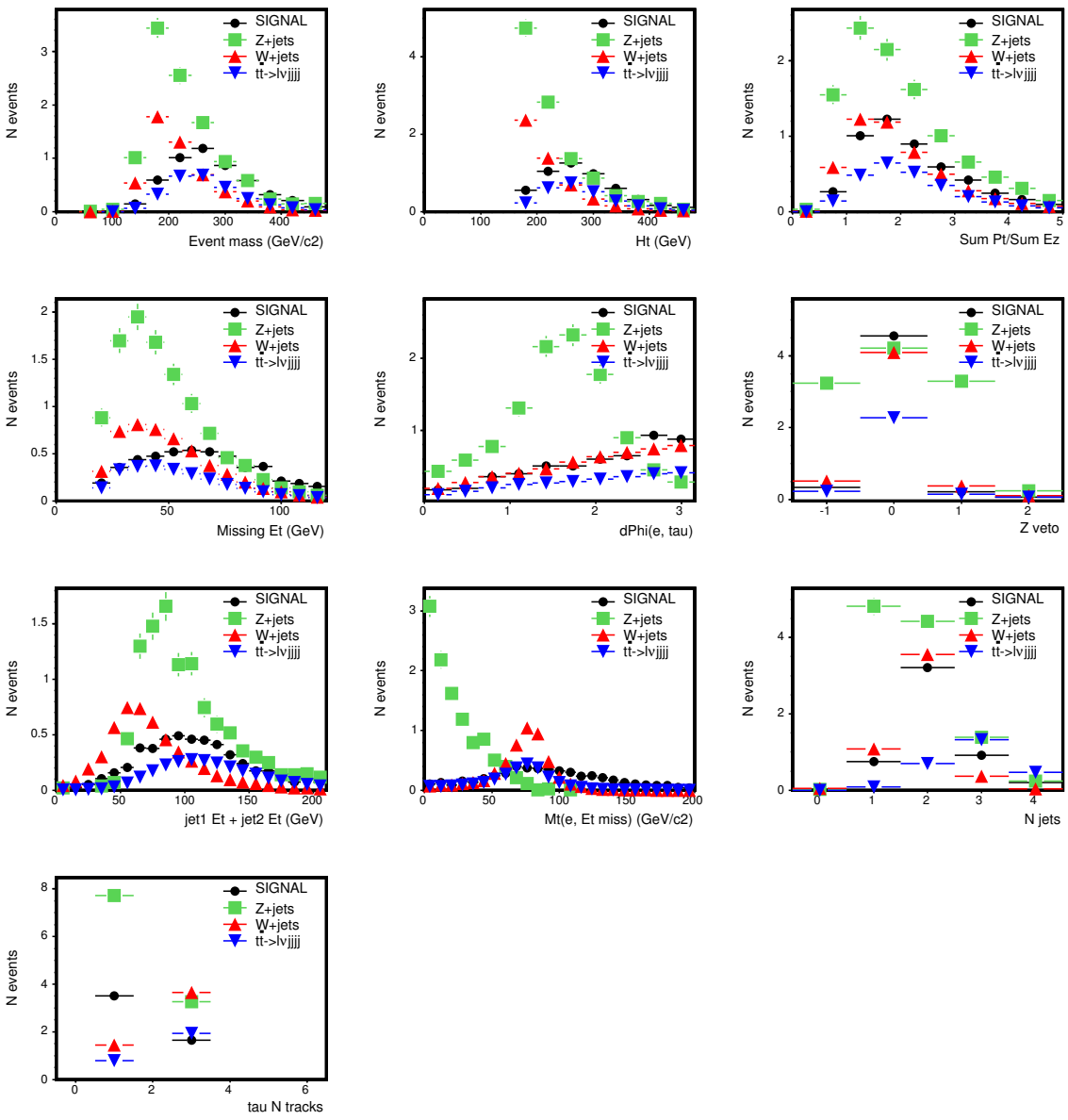


Figure 22: Superposition of the distributions for the 10 variables selected to be part of the likelihood ratio.

at the beginning. In our case, adding all the backgrounds in the denominator would basically only have decreased the Z background, because this is by far the predominant one after all event selection cuts but the likelihood one have been applied.

The final likelihood $L0_n$ variable is defined as the product of the n first likelihood ratios Lk:

$$L0_n = \prod_{k=1}^{n} L_k$$

Fig.23 shows the ten $L0_n$ likelihood ratio distributions obtained for the signal, compared with the same distributions for the three backgrounds. The way the likelihood method is separating the backgrounds from the signal after each of the ten steps is clearly noticeable in this figure. The $L0_{10}$ likelihood variable is as expected the most discriminant one. This $L0_{10}$ variable is the one used for the 1fb^{-1} analysis likelihood discrimination.

DRAFT OCTOBER 2, 2007

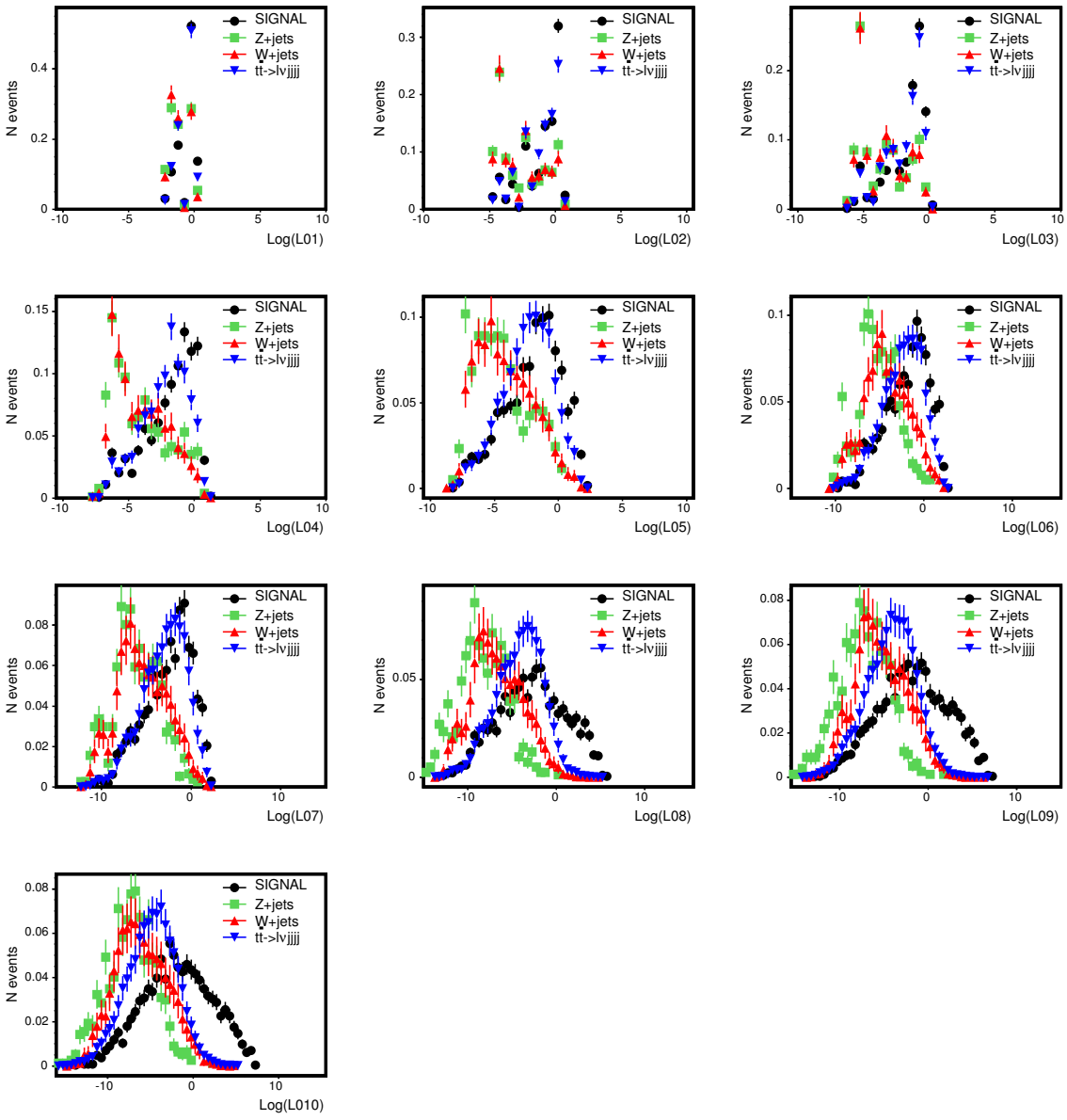


Figure 23: Superposition of the distributions for the 10 likelihood variables $L0_n$.

DRAFT 2007

An attempt to build another likelihood function Compared with other statistical methods for signal-background discrimination (neural networks, decision trees, support vector machines, etc), the likelihood method has at least three advantages : it is simple; the use of more discriminant variables in the likelihood function does not systematically require an increase in the statistics of the event samples; last but not least, the likelihood method is downright unbeatable if the variables used are strictly uncorrelated. In case a given variable is correlated with many of the other variables, this variable will impose the shape of the final likelihood variable, and therefore eventually degrade the contribution of some other less correlated variable. An attempt was made to build another likelihood method that would be less sensitive to the correlations between the variables. It is based on the same variables as the ones used for L0. They were ranked following the same order in which they were presented earlier. This order takes care of putting close to each other correlated variables. The method is iterative and proceeds as follows:

1. Start from the first variable (here: event mass) and set $L1_1 = L0_1$. $B1_1$ is defined as the event mass distribution for the sum of the W and Z backgrounds (weighted by cross sections), and $S1_1$ as the event mass distribution for the signal, weighted by the signal cross section.

2. For each variable 'k'[10] (k running from 2 to 10), for each background, build the background $b1_k$ distribution by looping over the events and weighting them by the product $\prod_{i=1}^{k-1} \frac{S1_i}{B1_i}$. Compute the weighted background distribution

$$B1_k = \sum_i Xs_i b1_i,$$

   where Xs_i is the cross section of the background i and the sum is made over the only backgrounds against which the variable k is meant to discriminate.

3. For each variable 'k' (k running from 2 to 10), compute the signal distribution $S1_k$ as the unweighted distribution for the variable k.

4. For each variable 'n' (n running from 2 to 10), for each background and for the signal, compute the L1 likelihood

$$L1_n = \prod_{k=1}^{n} \frac{S1_k}{B1_k}$$

Fig.24 shows all the reweighted $b1_k$ distributions for backgrounds and $S1_k$ distribution distribution for the top in tau signal.

Fig.25 shows all the ten $L1_n$ likelihood ratio distributions for the signal, compared with those of the three backgrounds.

[10] From now on, by variable 'k', we mean the k^{th} variable in the ordered list of the ten discriminant variables. By background i, we mean one of the three backgrounds, Z, W or top(lepton+jets)

DRAFT OCTOBER 23, 2001

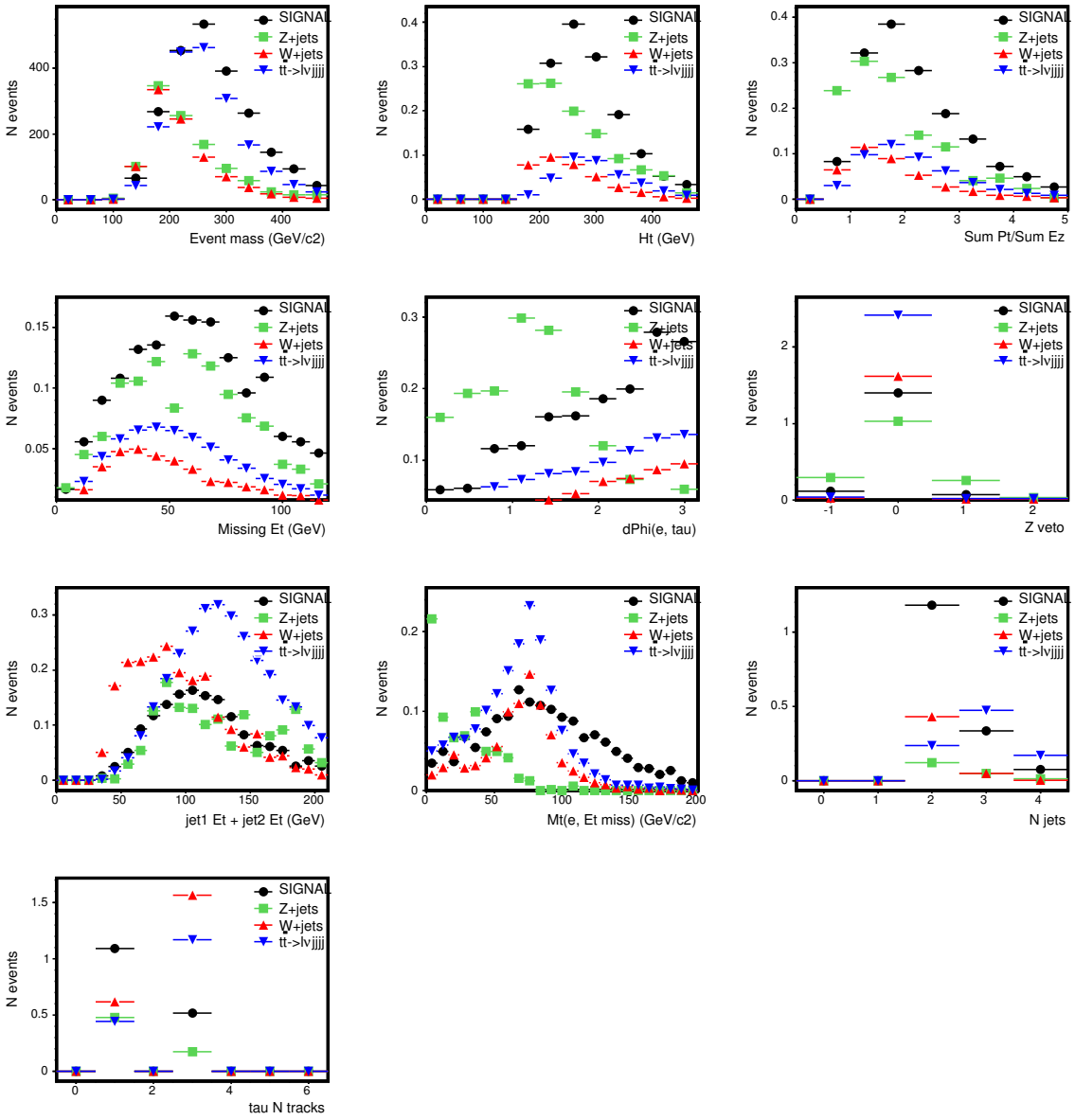


Figure 24: Superposition of the reweighted $b1_k$ and $S1_k$ distributions for the 10 variables (k=1,...,10) .

DRAFT 2007

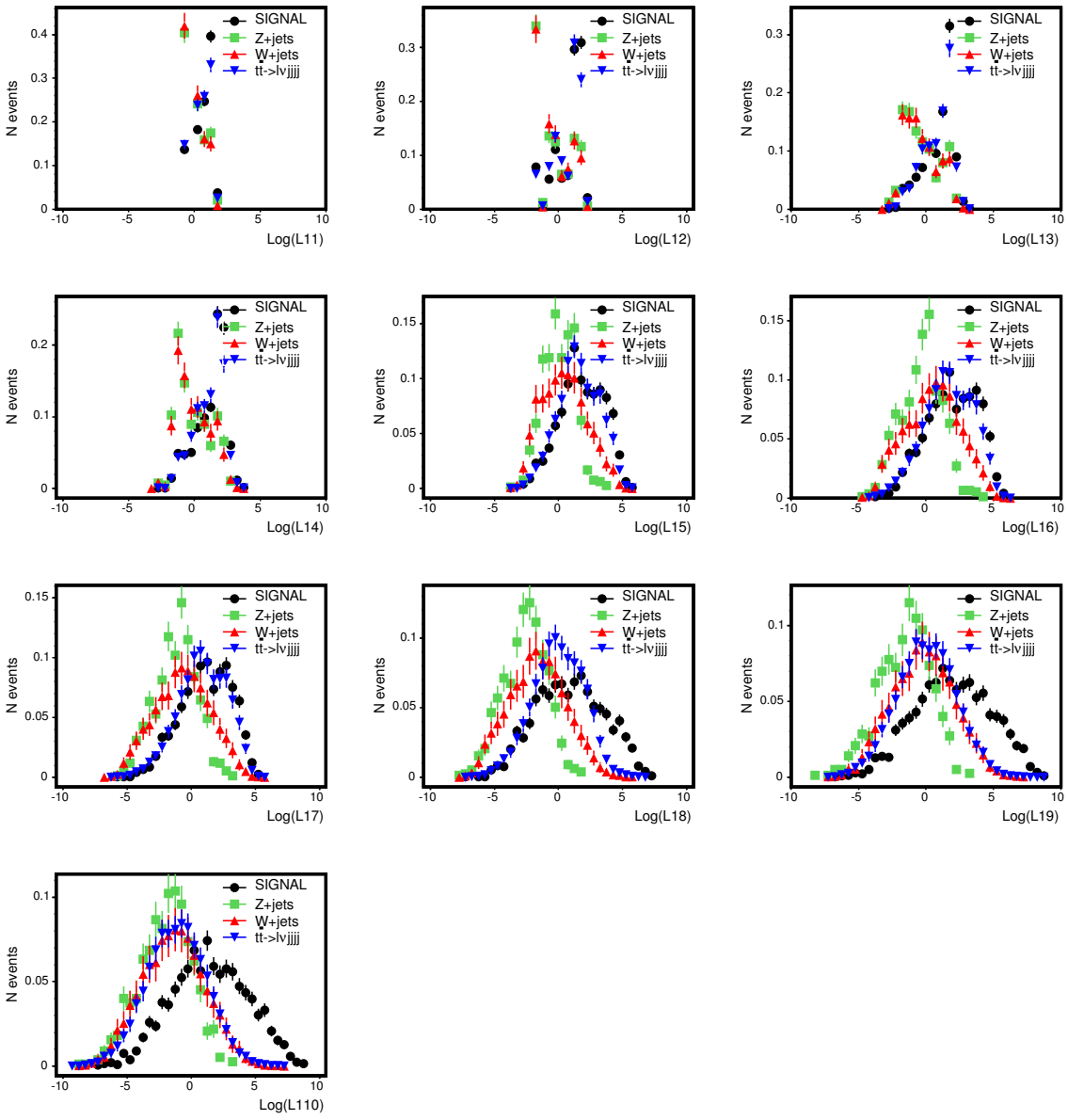


Figure 25: Superposition of the distributions for the 10 likelihood variables $L1_n$.

DRAFT 2007

Choice of the final likelihood cut for the 1fb^{-1} selection Two likelihood variables are therefore at disposal, namely: L0 and L1. For both of them, a cut value, c, is defined that only keeps the events with $L0 > c$ or $L1 > c$. The goal being the observation of the top in tau signal over the background, if the expected number of background events was high enough (say greater than 10), then the cut c would be chosen to optimize $S/\sqrt{B}$, where S is the number of signal events left and B the total number of background events left. However, it was found that, even with 1 fb^{-1} of data, the optimal number of background events B is smaller than 1 and thus $\sqrt{B}$ largely underestimates the standard deviation of the Poisson distribution of mean B. Thus, instead, the optimal cut c is chosen to minimize the expected $p-value$ of the null hypothesis for the *non-existence of the signal*. The way this expected p-value p_{exp} is computed, knowing S and B, is the following:

$$p_{exp} = \frac{\int_{S+B}^{\infty} dx B^x/\Gamma(x+1)}{\int_0^{\infty} dx B^x/\Gamma(x+1)}$$

where Γ is the Euler function extending the factorial to real values.

Fig. 26 shows the variation of S as a function of B for various choices of the cut c. This is done for each of the ten likelihood variables $L0_n$ (black triangles) and $L1_n$ (red points). The fixed black point represents the result if the same cut based selection as was developed for the 350pb^{-1} analysis is applied to the 1fb^{-1} sample.

Fig. 27 shows the sensitivity of each likelihood method, depending on the value given to the likelihood cut c. This figure shows that both likelihood methods achieve much better sensitivities than the cut based algorithm defined for the 350pb^{-1} analysis. Furthermore, the L0 likelihood looks like the better choice in this particular case. The function $L0_{10}$ exhibits the smaller expected p-values for values of c comprised around 0. These likelihood cuts leave less than 0.4 events for the Z and fake tau backgrounds and less that 2.5 events for the signal. Thus, the cut c is chosen to be 0.

As the search for deviations from the standard model expectation for the ratio $r_\tau = \frac{BR(t\to\tau\nu b)}{BR(t\to l\nu b)}$ goes, the quantity that needs to be optimized is different, since this becomes $\frac{S}{\sqrt{S+B}}$. Thus, we can define an alternative cut on L0, c', which is optimized for the branching ratio measurement. Fig. 28 shows the sensitivities reached for the two likelihood methods (L0 and L1), depending on where the likelihood cut c' is placed. This is shown for each of the ten likelihood variables $L0_n$ and $L1_n$, n varying from 1 to 10. Here the conclusion is that the likelihood method L0 does not do much better but as well as the cut-based selection. The value of the cut c' chosen to be the best is $c' = -5$. This leaves 6.7 top into tau signal events for 8.5 Z and fake taus background events. This is very close to the performance achieved by the sequential cuts selection that leaves 6.5 top into tau signal events against 7.5 background events.

DRAFT OCTOBER 21 2007

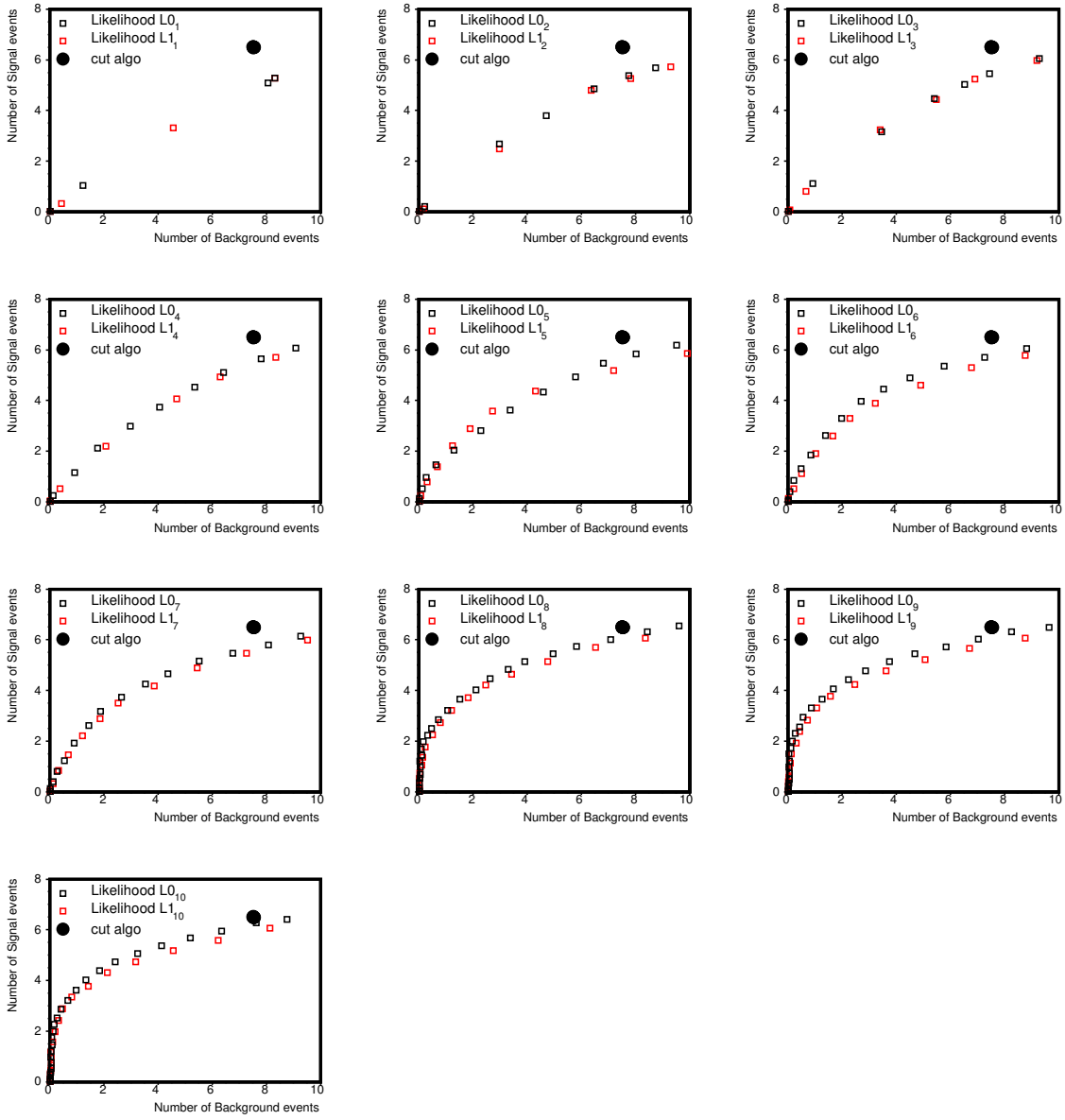


Figure 26: Relations between the remaining number of signal and background events for the L0, L1 and cut algorithm.

DRAFT 2007

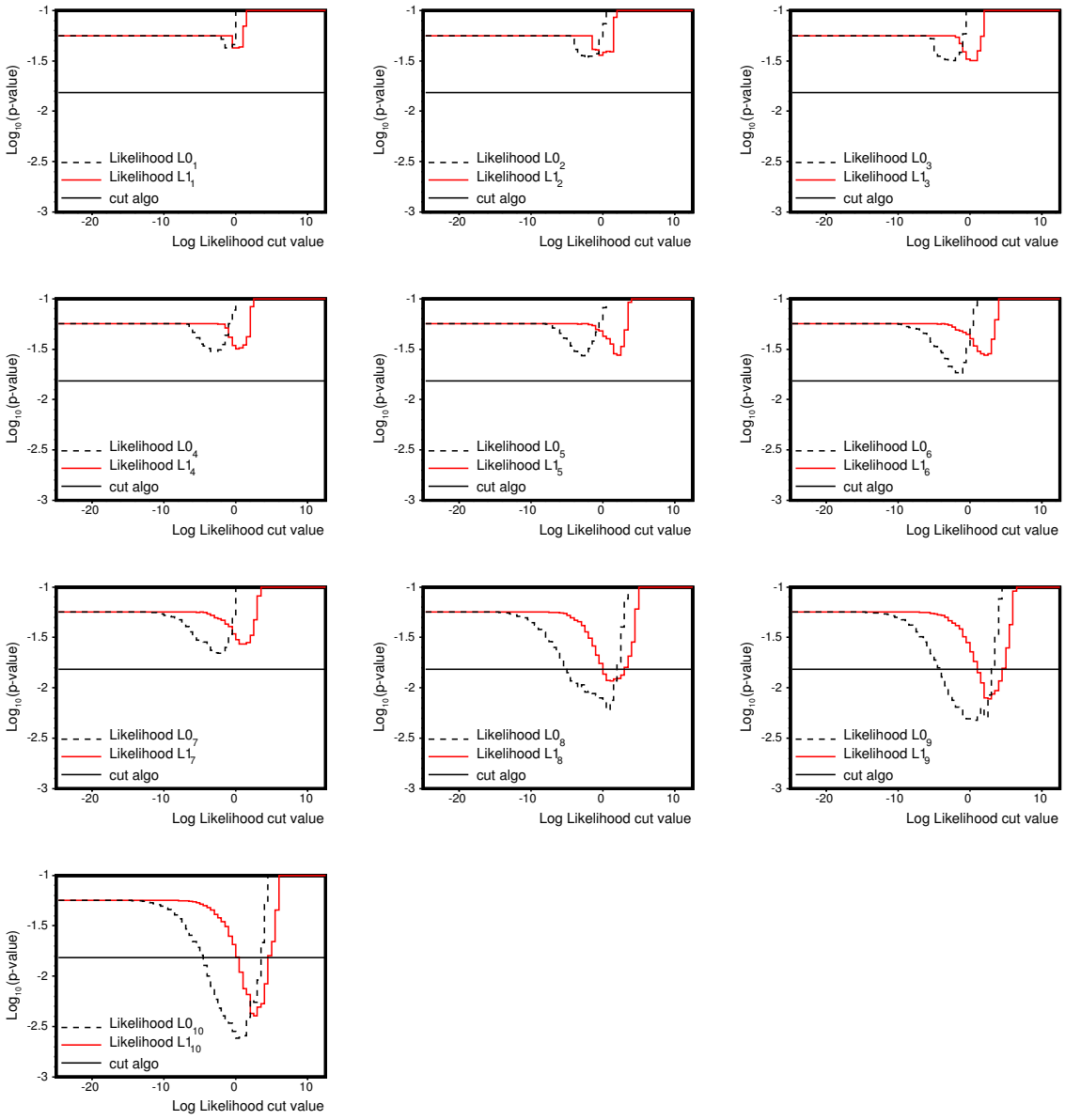


Figure 27: Relations between the logarithm of the expected p-value and the c cut on the log likelihood number, for L0 (black) and L1 (red) likelihood methods. The horizontal line stands for the sensitivity achieved with the cut based selection of the 350 pb^{-1} analysis if applied to the 1 fb^{-1} sample.

DRAFT 07

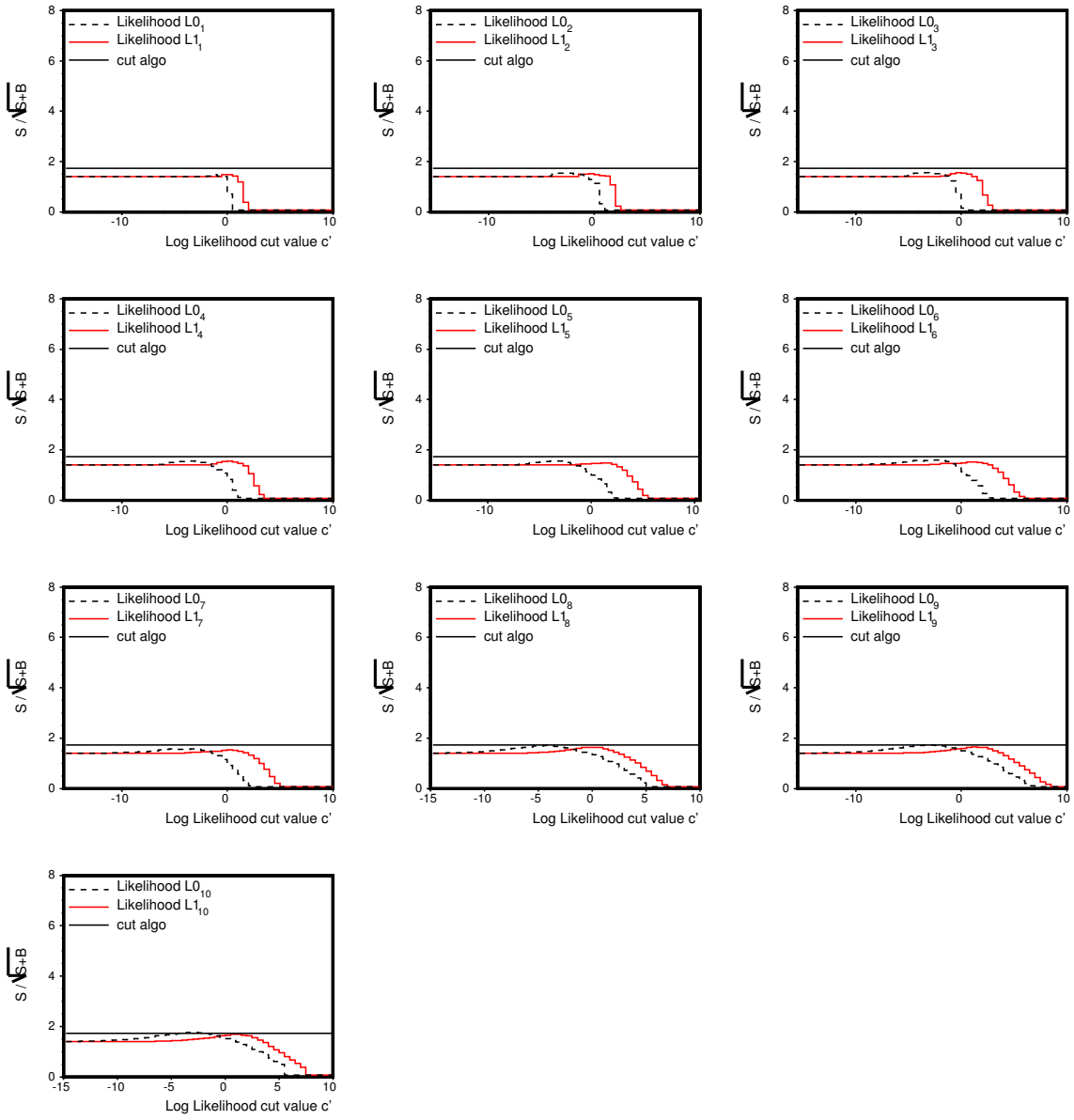


Figure 28: Relations between the logarithm of the expected p-value and the c' cut on the log likelihood number, for L0 (black) and L1 (red) likelihood methods. The horizontal line stands for the sensitivity achieved with the cut based selection of the 350 pb^{-1} analysis if applied to the 1 fb^{-1} sample.

DRAFT 07

2.5.5 The event acceptance

The estimate of the signal acceptance is based on the Pythia Monte Carlo generator, tuned with data for lepton identification efficiencies. We use a $t\bar{t}$ Monte Carlo sample made by the electroweak working group (sample labelled as "tewk0z") and we apply the event selection as defined in subsection 3.5.1. The top mass is set at 175 GeV/c^2. To ensure that there is no double counting of events due to fake reconstructed leptons, the sample is filtered to keep only events where a W decays into an electron or a muon and the other one into a tau, itself decaying hadronically. Then, each reconstructed object is required to match it's generator level parent particle. This is done by requiring that the distance $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ be less than 0.04 for electrons and muons and 0.2 for the tau.

The raw signal efficiency estimate obtained from the Monte Carlo sample is summarized in the Table 6. Most variables are common to the 350 pb^{-1} analysis and were explained in the subsection 3.4.3. They are reminded here again with N(cut) as the number of events found passing this cut and all previous ones.

N($e\mu_{ID}$): the electron (muon) passes all identification cuts.

N(τ_{ID}): the tau passes all the identification cuts.

N(Opp Sign): the electron (muon) and the tau lepton have opposite electric charges.

N(2 jets>15GeV): Two jets are found between -2 and 2 in pseudorapidity and with a transverse energy for each, greater than 15 GeV.

N($\not\!\!E_T$>20GeV): The corrected missing transverse energy is required to be more than 20 GeV.

N(H_t>160GeV): the sum of E_T of the two jets, of the electron (muon), of the tau lepton and of $\not\!\!E_T$ is greater than 160 GeV.

N(L0 > -5): the L0 log likelihood ratio, as defined in the subsection 3.5.2 is greater than -5.

N(L0 > 0): the L0 log likelihood ratio, as defined in the subsection 3.5.2 is greater than 0.

The last step in the selection is to correct the MC acceptance for discrepancies with the real data, using the correction factors to the Monte Carlo, listed in Table 7 [32].

As a result, after having combined the Monte Carlo acceptances of Table 6 and the scale factors of Table 7, the following acceptances are obtained:

- $L0 > 0$ (for the top into tau signal observation) :
  - $e + \tau$: $7.7 \pm 0.5_{stat.} \times 10^{-5}$
  - $\mu + \tau$: $4.6 \pm 0.4_{stat.} \times 10^{-5}$
- $L0 > -5$ (for the measurement of $r_\tau = \frac{t\to\tau\nu_\tau b}{t\to l\nu_l b}$):
  - $e + \tau$: $5.0 \pm 0.1_{stat.} \times 10^{-4}$
  - $\mu + \tau$: $3.4 \pm 0.1_{stat.} \times 10^{-4}$

DRAFT OCTOBER 29 2007

	Number of MC events			
Cut	$e\tau_{had}$	$\tau_e\tau_{had}$	$\mu\tau_{had}$	$\tau_\mu\tau_{had}$
N(channel)	62385	8903	46653	8186
N($e\mu_{ID}$)	20635	1343	13116	1026
N(τ_{ID})	3242	233	2065	148
N(Opp Sign)	3217	233	2053	147
N(2 jets>15GeV)	2704	192	1664	124
N($\not\!\!E_T$>20GeV)	2516	173	1533	112
N(H_t>160GeV)	2480	169	1514	107
N(L0>-5)	1991	116	1212	76
N(L0>0)	882	34	502	21

Table 6: *Acceptance table: Number of events in Pythia $t\bar{t}$ passing each individual cut of the event selection.*

type	scale factor value	
	2002-2004 runs	2005-2006 runs
$\epsilon_{trigger}$: high p_T CEM	0.962 ± 0.006	0.9773 ± 0.0044 [46]
$\epsilon_{trigger}$: high p_T CMUP	0.8890 ± 0.0043	0.9187 ± 0.0028 [49]
$\epsilon_{trigger}$: high p_T CMX	0.9675 ± 0.0033	0.9508 ± 0.0029 [49]
ϵ_{ID}^{e}	0.986 ± 0.004	0.975 ± 0.004 [48]
ϵ_{ID}^{CMUP}	0.9285 ± 0.0051	0.9242 ± 0.0037 [49]
ϵ_{ID}^{CMX}	0.9988 ± 0.0055	0.9760 ± 0.0039 [49]
ϵ_{ID}^{τ}	1.0 ± 0.09	1.0 ± 0.09

Table 7: *Scale factors by which to multiply the Monte Carlo acceptance.*

2.5.6 Expected number of signal events observed in 1 fb^{-1} data

For the $t\bar{t}$ cross section, the last CDF combined result is used [21], namely: 7.3 pb. Assuming this cross section, using the signal acceptance, here are the expectations for the numbers of signal events for both electron and muon channels and both choices of the likelihood cut:

- $L0 > 0$ (for the top into tau signal observation) :
  - $e + \tau$: $1.7 \pm 0.1_{stat.}$
  - $\mu + \tau$: $1.1 \pm 0.1_{stat.}$
- $L0 > -5$ (for the measurement of $r_\tau = \frac{t \to \tau \nu_\tau b}{t \to l \nu_l b}$):
  - $e + \tau$: $3.8 \pm 0.1_{stat.}$
  - $\mu + \tau$: $2.6 \pm 0.1_{stat.}$

DRAFT OCTOBER 2, 2007

Thus, we expect a total of 2.8 signal events after the tight likelihood cut and around 6.4 events after the loose likelihood cut.

DRAFT OCTOBER 2, 2007

2.6 Background estimation

The signal region, characterized by one central high p_T lepton (electron or muon), one central tau-jet, missing E_T and 2 high E_T jets, is populated with backgrounds that can be divided into two categories:

1. physics backgrounds whose estimation relies on Monte Carlo tuned with data. The non negligible ones are:
   - $Z \rightarrow \tau\tau$ + jets
   - $WW \rightarrow \tau\nu_\tau l\nu_l$ + jets
2. Backgrounds due to the misidentification of taus. These are derived from data only because the Monte Carlo is not tuned to reproduce the correct jet shapes and overestimates the ability of jets to mimic a tau like narrow jet:
   - jets faking taus : this category gathers mostly W+jets events, but also QCD, $t\bar{t} \rightarrow l + jets$ and all physics processes generating a lepton accompagnied at least by three high E_T jets.
   - electrons faking taus : this is mainly a background for the electron channel, that contains $Z \rightarrow ee$ + jets

2.6.1 Monte Carlo based backgrounds

A prerequisite to any analysis with leptons is to check that the lepton identification and energy measurement are well under control, both in the data and in the Monte Carlo samples.

Fake missing E_T Whereas the diboson and signal samples contain true missing E_T and thus do not rely on the ability of the simulation framework to reproduce well any fake missing E_T contribution, the $Z \rightarrow \tau_l\tau_{had}$ background has both a true and fake missing E_T component. The simulation of fake missing E_T must thus be compared to data.

$Z \rightarrow \mu\mu$ is used for this analysis because these are well mastered clean events without any true missing E_T contribution.

Figure 29 compares Monte Carlo and data for the missing E_T calculated in $Z \rightarrow \mu\mu$ events. This comparison is not enough, as fake missing E_T in $Z \rightarrow \tau_l\tau_{had}$ + jets will essentially come from jet mismeasurement, rather that from the lepton.

A way to probe fake missing Et due to the extra jets or underlying event, is to look at the missing Et component orthogonal to the Z boson in the transverse plan. This is sure to have a zero true value, and to be due to underlying event and extra jets. The figure 30 shows this as a function of the scalar sum of transverse calorimeter energies in the event (Sum E_T), and compare Pythia to data. This sum E_T doesn't take account of the muon energies since muons do not deposit much in the calorimeter. The curve

DRAFT OCTOBER 2, 2007

of fig 30 is thus a good snapshot of the behaviour of the fake missing Et due to jet mismeasurement, a good measure of the missing E_T resolution in events characterized by a high sum of jet transverse energies.

Stating the apparent good agreement between Monte Carlo and data, no correction to the simulated missing Et is performed through the analysis.

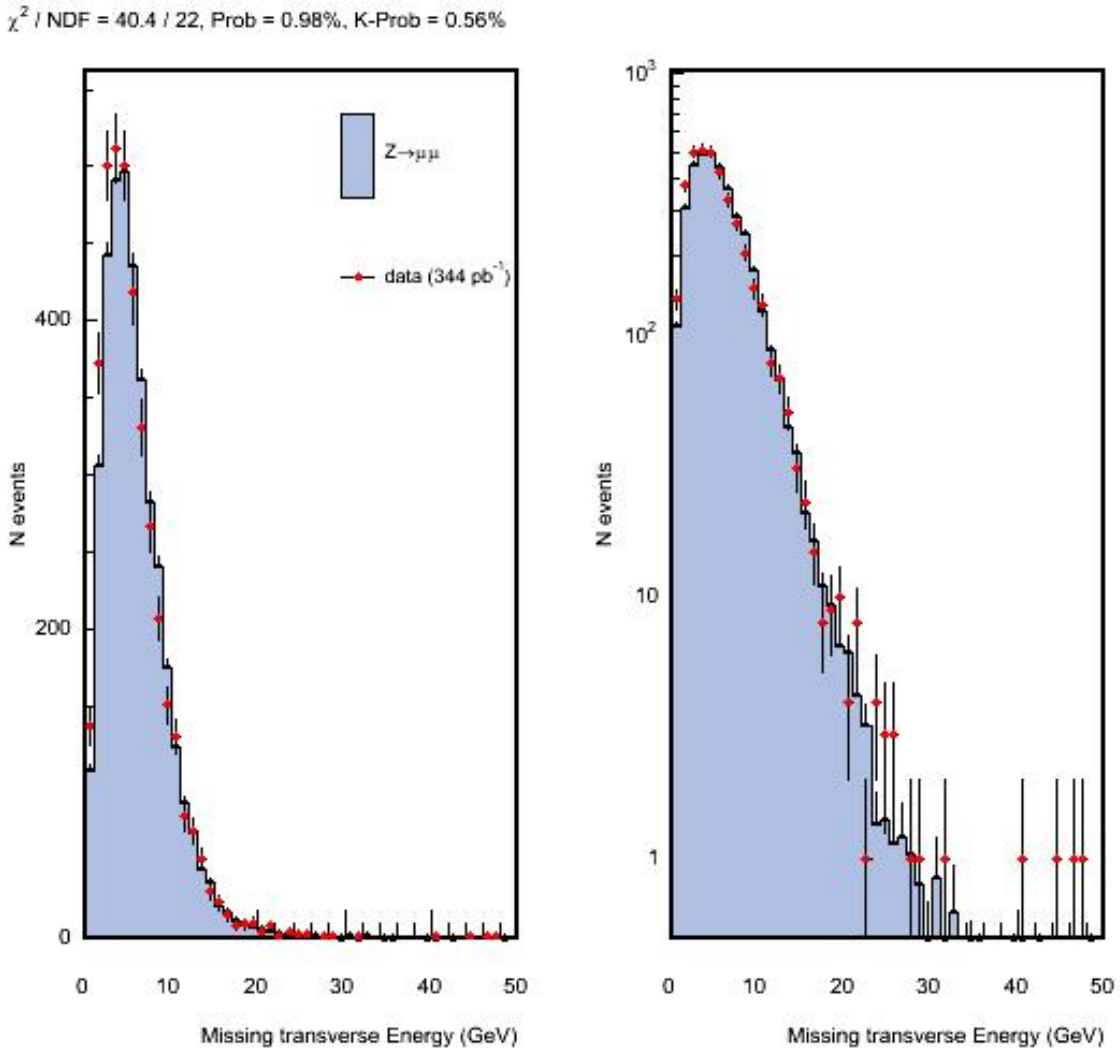


Figure 29: *Fake missing E_T present in $Z \rightarrow \mu\mu$ events. Left: linear scale ; Right: log scale.*

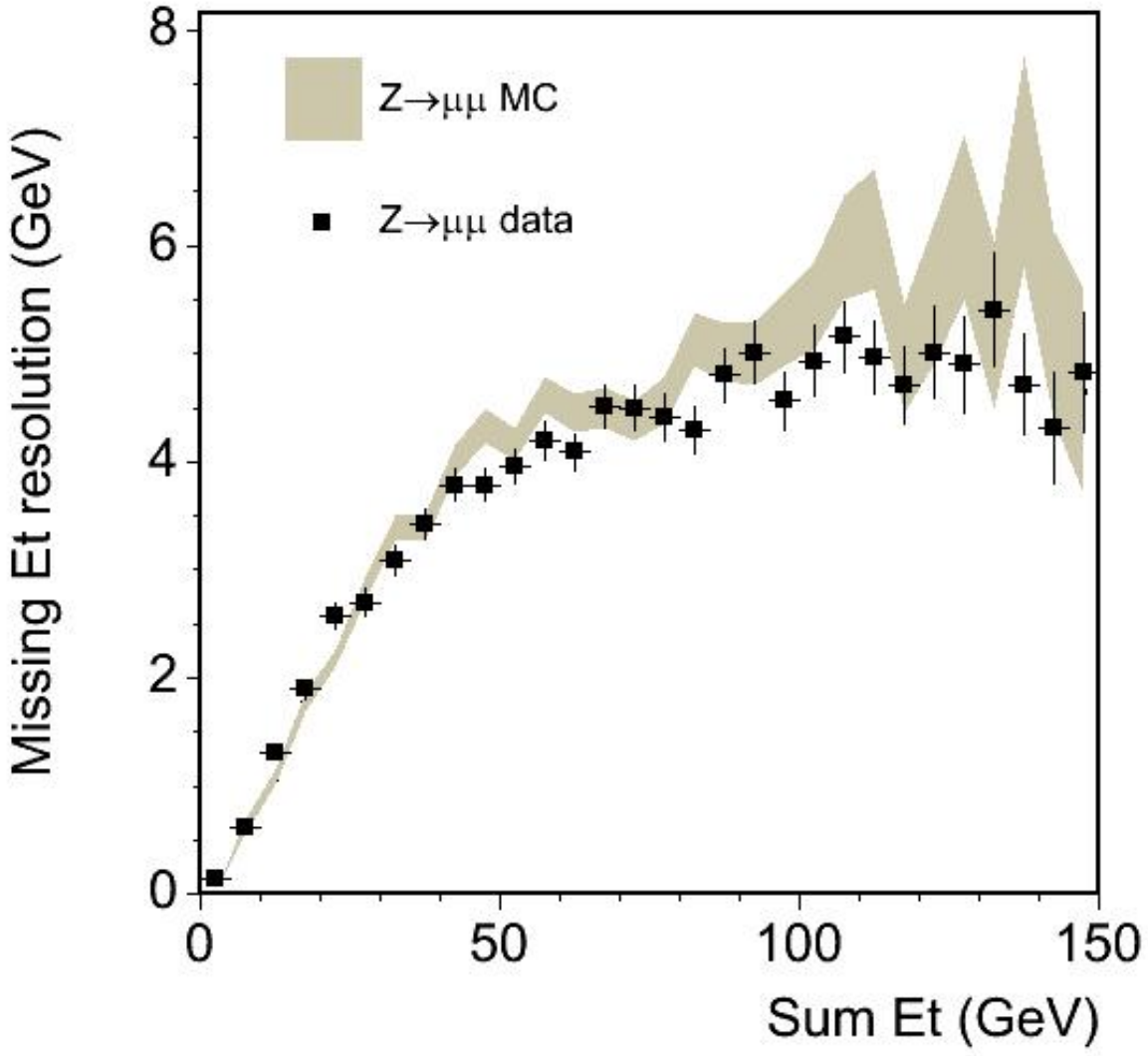


Figure 30: *Missing E_T resolution orthogonal to the Z boson direction as a function of Sum E_T.*

DRAFT 2007

N jets factors The three former paragraphs show evidence that muons and missing E_T are under control in data as well as in the simulation framework around Pythia. This is not enough for our needs because the backgrounds include 2 extra jets, which adds a further difficulty. Mastering the simulation of events with several radiated jets is a high challenge for Monte Carlo generators. In addition to 'traditional' generators like Pythia or Herwig, that were essentially developed to reproduce leading order $2 \to 2$ processes (eg. $q\bar{q} \to Z \to ll$, $gg \to g \to t\bar{t}$,etc), some generators based on the S matrix elements (ME generators), like Madgraph or AlpGen, are available. These are supposed to better reproduce the energy spectrum of the higher order jets. Moreover, a more straightforward reason for using ME generators is that it saves us the generation of millions of useless events that present a too low number of extra jets. Thus the generation of $Z \to \tau\tau + 2$ jets events will make use of AlpGen sample of $Z \to \tau\tau + 2$ partons, but the question is now: How to know the cross section of such a sample to normalize it? Regardless of the availability of precise theoretical predictions for such cross sections at next to leading or higher orders, it is essential to check it in data. Although it is not possible to get a solid data sample of $Z \to \tau\tau + 2$ jets, this is possible for $Z \to ee$ or $Z \to \mu\mu + \geq 2$ jets. If the cross section of $Z \to \mu\mu + \geq 2$ jets can be obtained from data, this can be assumed to be valid also for $Z \to \tau\tau + \geq 2$ jets, thus providing the needed normalization for our Alpgen sample. Once normalized on Z+2 jets, the Alpgen sample is used to make predictions into more restricted regions of the phase space, with a statical power hardly accessible to Pythia.

Figure 31 compares the number of extra jets with corrected energies greater than 15 GeV, and pseudorapidities between -2 and 2 in $Z \to \mu\mu$ data and Pythia simulation(zewk6m). The simulation agrees with data on the 2 jets bin but, the number of events being low, the cut on jets are relaxed to increase statistics and get a more robust comparison (Fig 32, 33). The agreement gets better in the 2 jet bins as statistics increases. Same histograms, with a cut on missing $E_T < 20$ GeV added, were drawn to test if backgrounds with high missing E_T, possibly present in the 2 jets bin, such as $t\bar{t}$, were relevant. The observed independance of the number of events with 2 jets with the missing Et cut proves us that backgrounds with high missing Et are negligible in that region. This study gives us confidence in the capacity of Pythia to make solid predictions in the number of extra jets, up to N jets = 2. Stating the good level of agreement in the 2 jets bin, no correction is made to the Pythia prediction and a systematics error of 5% is introduced for the number of events from the MC based backgrounds.

The number of jets observed in $Z \to \mu\mu$ will be used to normalize the prediction from Alpgen+herwig used in the analysis for the $Z \to \tau\tau$ background. Furthermore, it is claimed that this validates the use of Pythia for the estimate of the small diboson background.

DRAFT OCTOBER 2, 2007

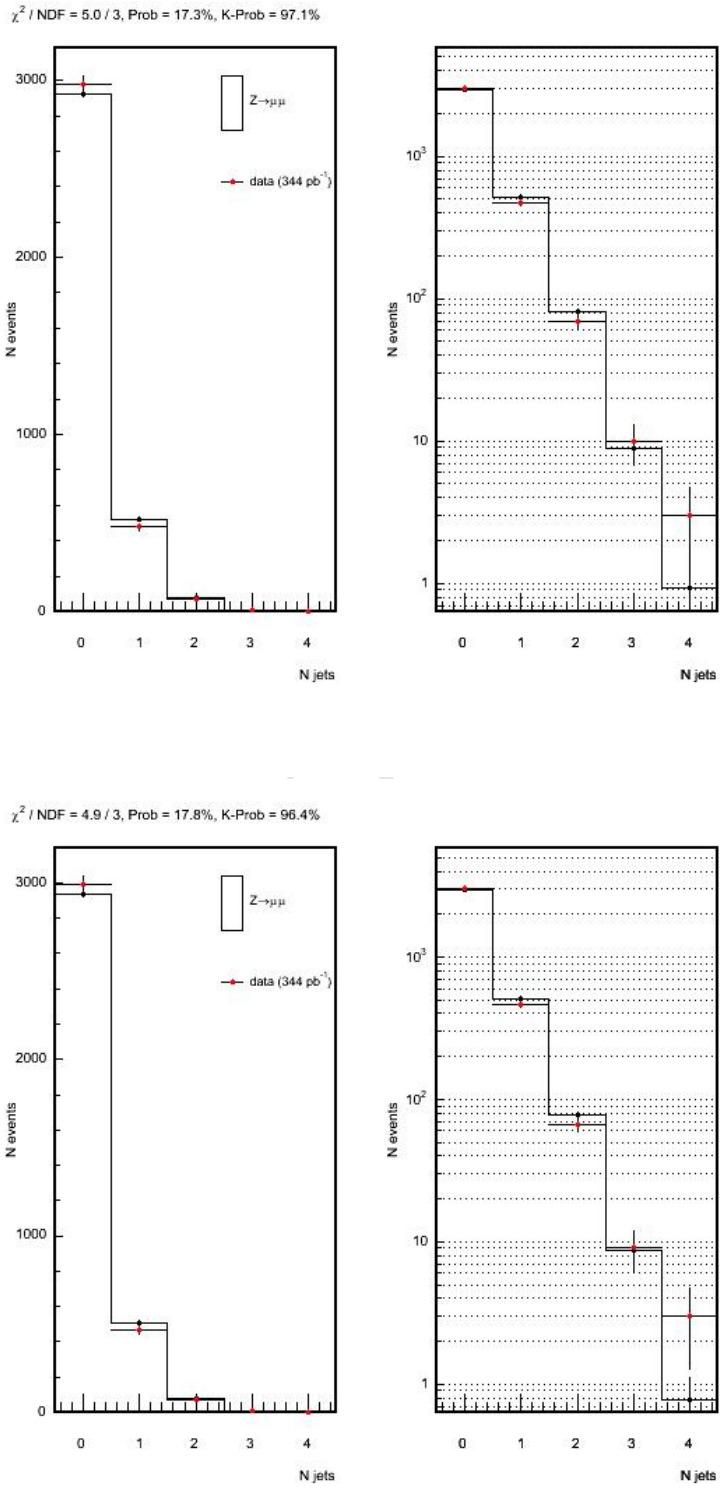


Figure 31: Number of extra jets in $Z \to \mu\mu$. Jet selection: $E_T^{cor} > 15$ GeV, $|\eta| < 2$. Top: two central stub muons, bottom: two central stub muons, MET<20 GeV.

DRAFT 2007

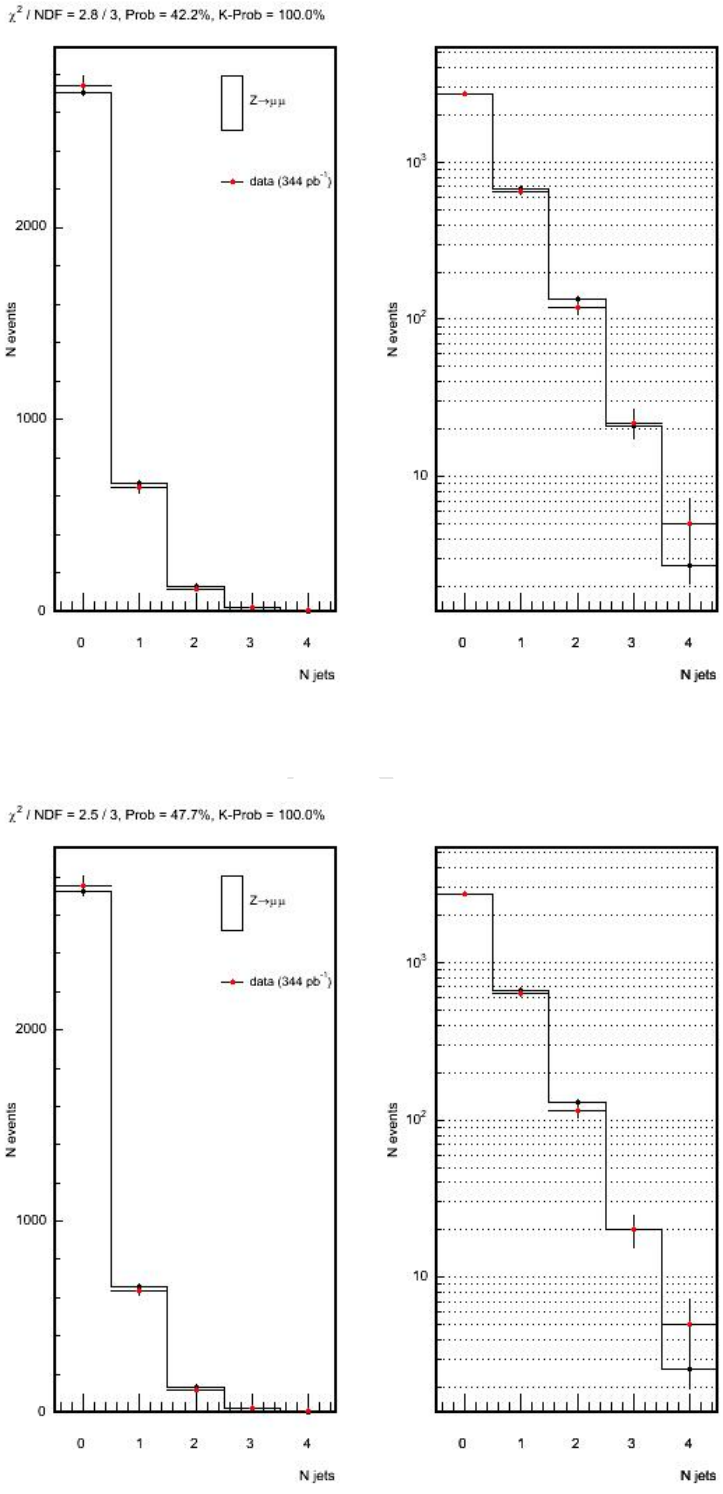


Figure 32: Number of extra jets in $Z \to \mu\mu$. Jet selection: $E_T^{cor} > 10$ GeV, $|\eta| < 2$. Top: two central stub muons, bottom: two central stub muons, MET<20 GeV.

DRAFT 2007

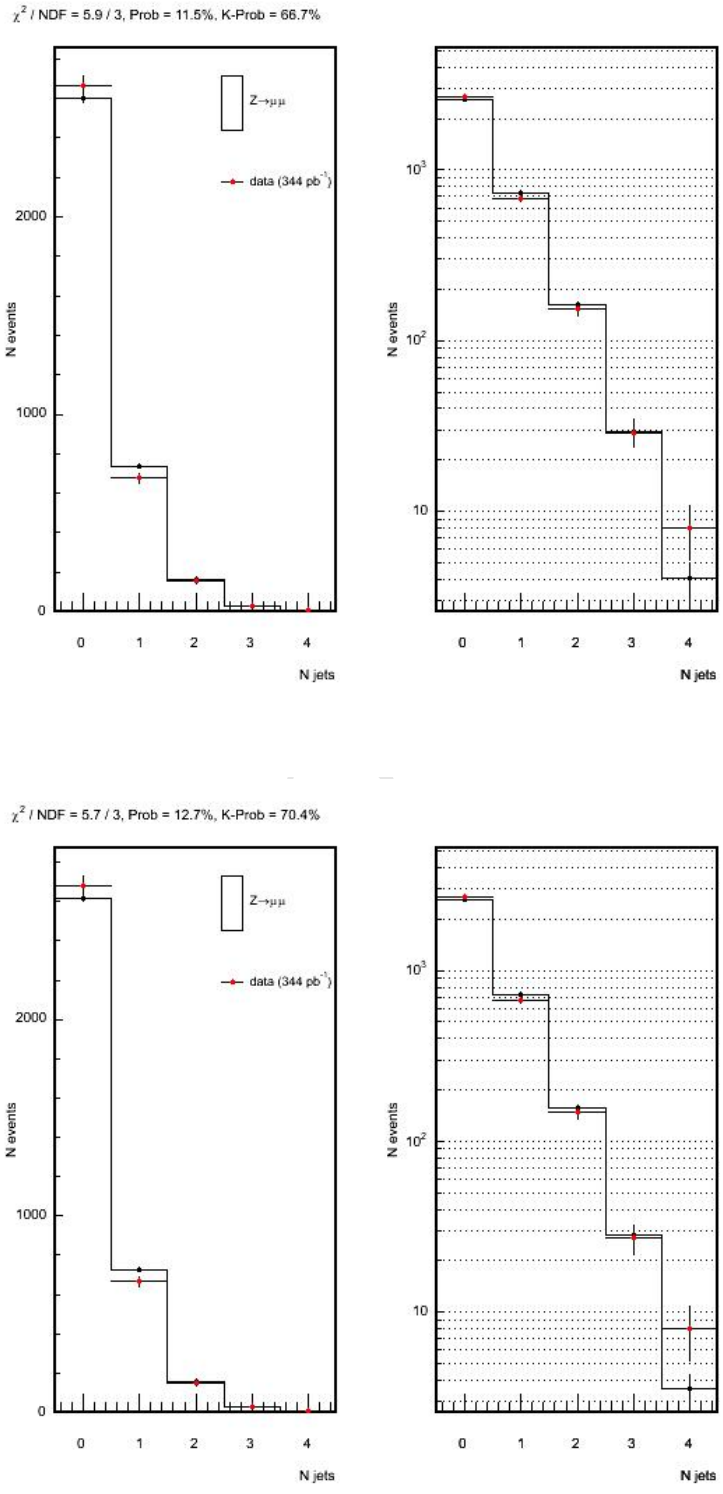


Figure 33: Number of extra jets in $Z \to \mu\mu$. Jet selection: $E_T^{cor} > 12$ GeV, $|\eta| < 2.5$. Top: two central stub muons, bottom: two central stub muons, MET<20 GeV.

DRAFT 2007

2.6.2 data-driven backgrounds: electrons and jets faking tau lepton hadronic decays

Electrons faking tau leptons The main physical mechanism leading to electrons being misidentified as τ leptons is strong Bremsstrahlung. Indeed, if the electron emits a high p_T photon, the electron track p_T has lower momentum and it can pass the electron veto cut $\xi = E_{had}/P_{trk} > 0.1$.

It was shown in the note 6308 [17] that the probability for an electron passing all tau requirements to survive to the electron veto was $3.6 \pm 0.6\%$ for $\frac{E_{had}}{\sum P} > 0.1$ and $1.2 \pm 0.3\%$ for $\frac{E_{had}}{\sum P} > 0.15$. We use the second harder cut and fold this probability value in our data to estimate this background.

Jets faking tau leptons The probability for a jet passing a subsample of the tau identification cuts (labelled as *denominator* cuts and to be defined in section 3.8.2) to further pass successfully the rest of the tau selection cuts is called the *jet to tau fake rate*.

The method consists in determining step by step the jet to tau fake rate from data, starting from events with two jets, then including events with higher jet multiplicity, and finally adding the case of jets as produced in $W \rightarrow e\nu$ + jets, which are a major source of jets faking taus in this analysis.

The jet to tau fake rate formula for the $W \rightarrow e\nu$ + jets events is shown to be given by : FkR(Jet E_T, Sum E_T)*f(N jets) , where FkR is a probability matrix (Fig. **??**) and f(Njets) are correction factors. All this was the subject of the CDF note 8208.

DRAFT OCTOBER 2, 2007

2.7 Checks and N jets control regions

Control regions are defined as the group of data events containing one tightly identified lepton (muon or electron), one identified central tau, missing transverse energy bigger than 20 GeV, and ≥ 0, ≥ 1 extra jets.

The figures 34, 35, 36, 37 show in red points the observed distributions of events in the ≥ 0 and ≥ 1 jet bin (beware that the tau-jet is not counted as a jet), superimposed on the predictions made up by summing the only three non negligible backgrounds (fake taus, $Z \rightarrow \tau\tau$ and, in a lesser extent, WW). The distributions are not normalized, they are raw predictions and dominated by the jet to tau fake rate.

The figures 38, 39, 40, 41 show the same control regions distributions in the case of the muon+tau channel.

The check through all the control regions is a great success in both the muon and electron channels. A quantification of the agreement is given by the χ^2 derived probability of consistency between the observed and predicted distributions, and printed on the top of each control histogram.

DRAFT OCTOBER 29, 2007

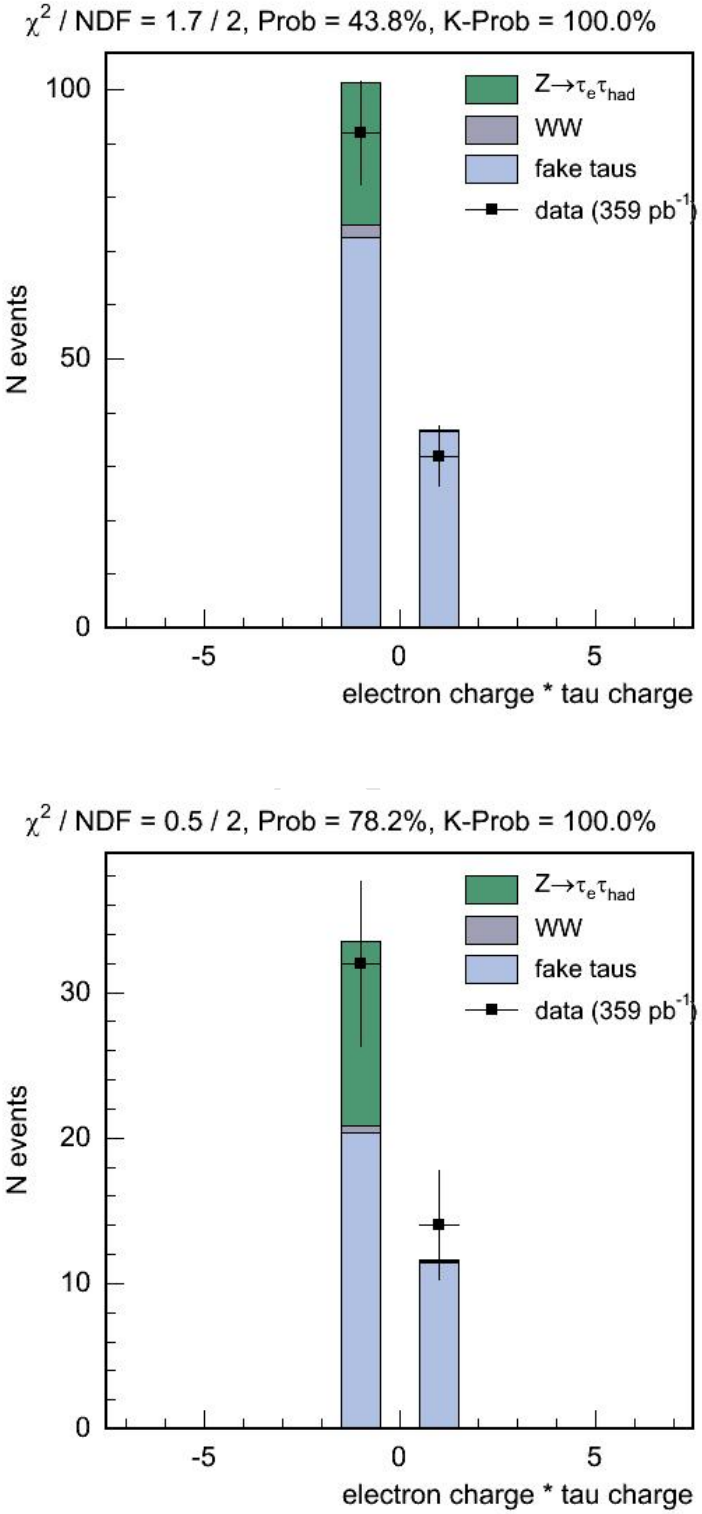


Figure 34: Charge product of the electron and the tau for events from the high p_T electron trigger sample, having 1 identified electron, 1 identified tau and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

DRAFT 2007

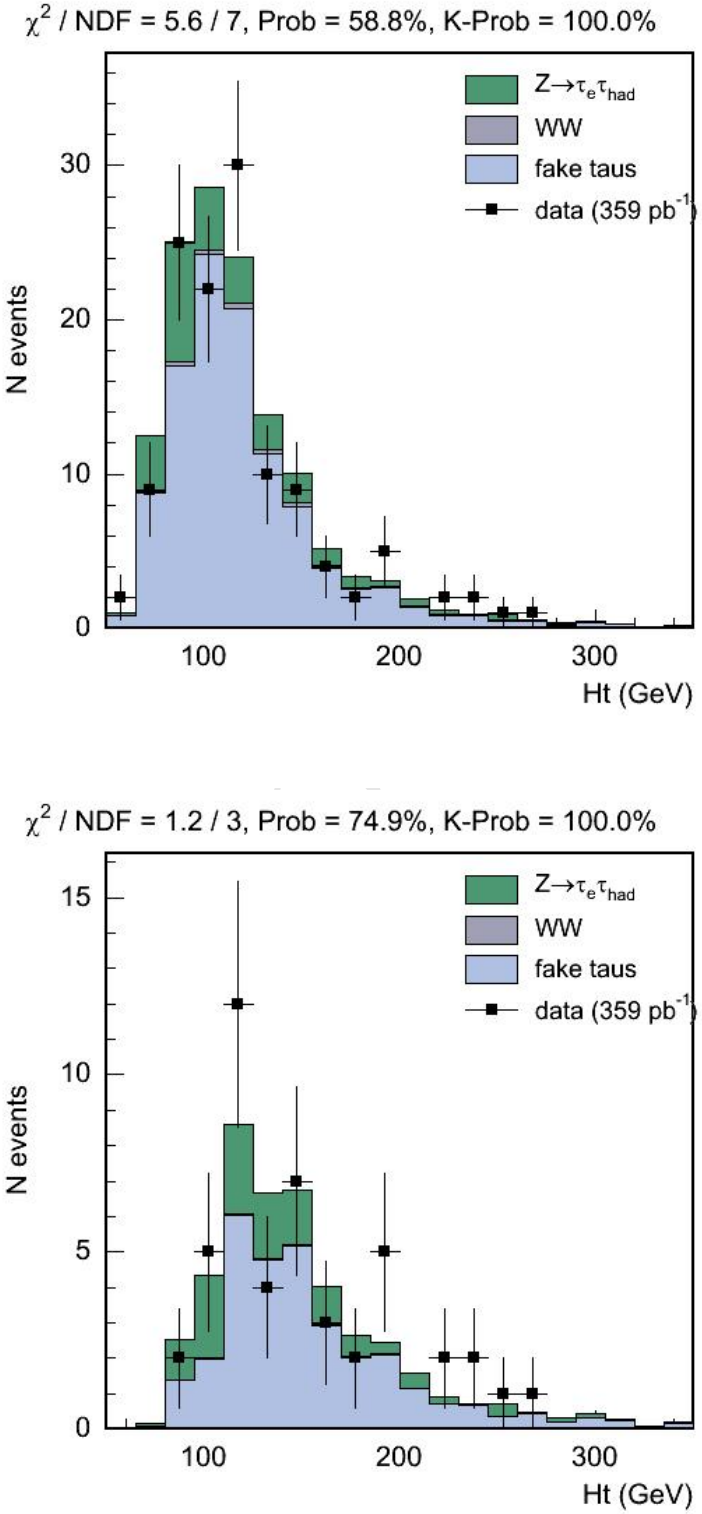


Figure 35: Event activity Ht for events from the high p_T electron trigger sample, having 1 identified electron, 1 identified tau and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

DRAFT 2007

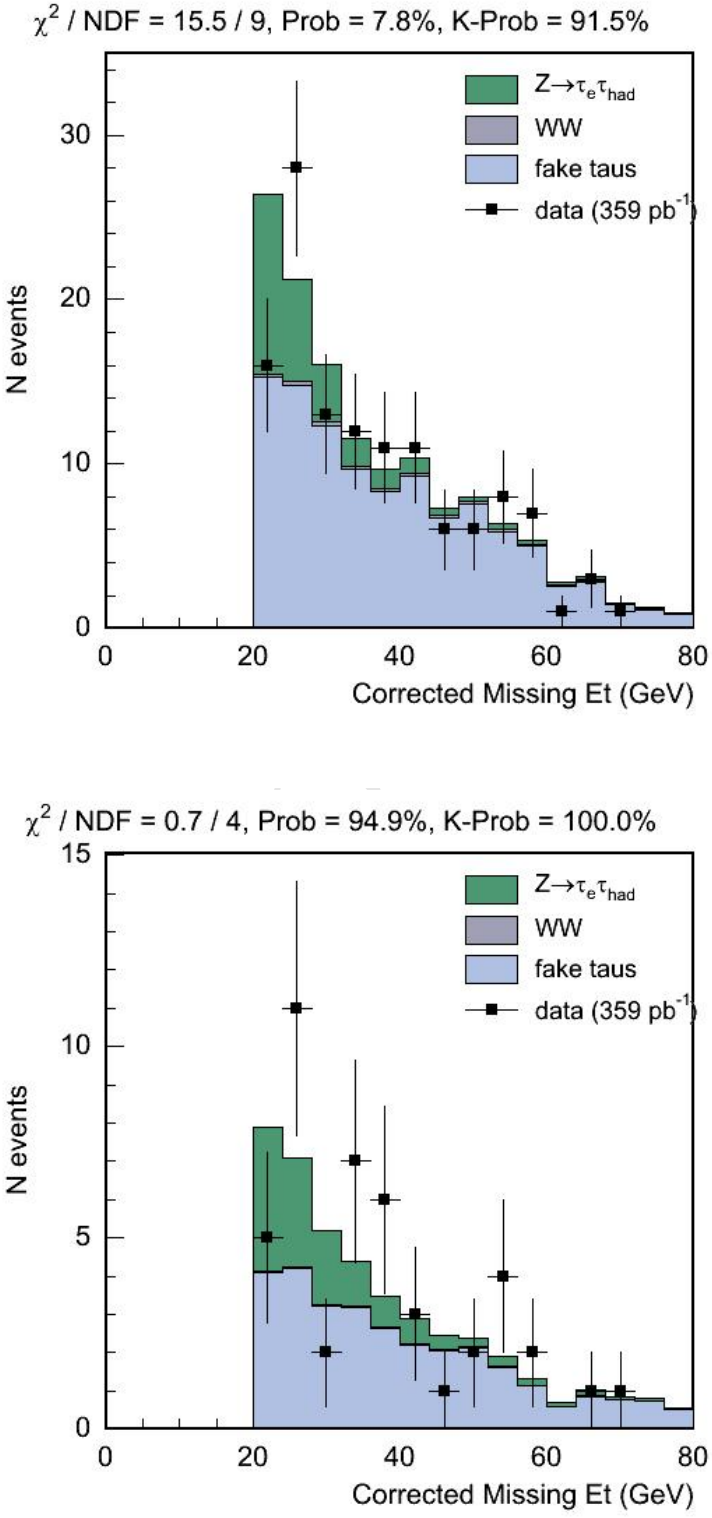


Figure 36: Event transverse Missing Energy for events from the high p_T electron trigger sample, having 1 identified μ, 1 identified τ and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

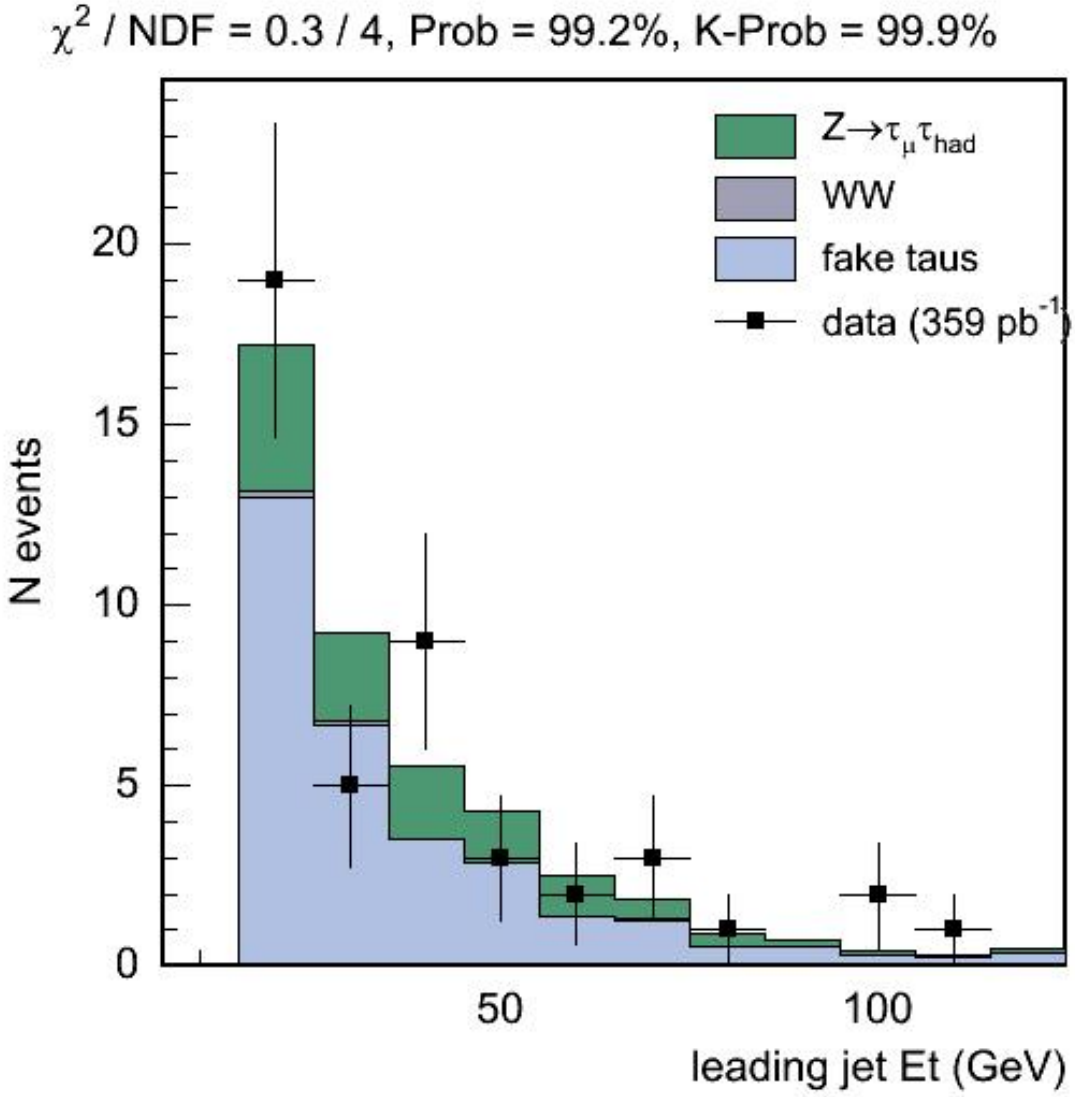


Figure 37: Leading jet E_T for events from the high p_T electron trigger sample, having 1 identified μ, 1 identified τ and MET>20 GeV and at least 1 other jet.

DRAFT 2007

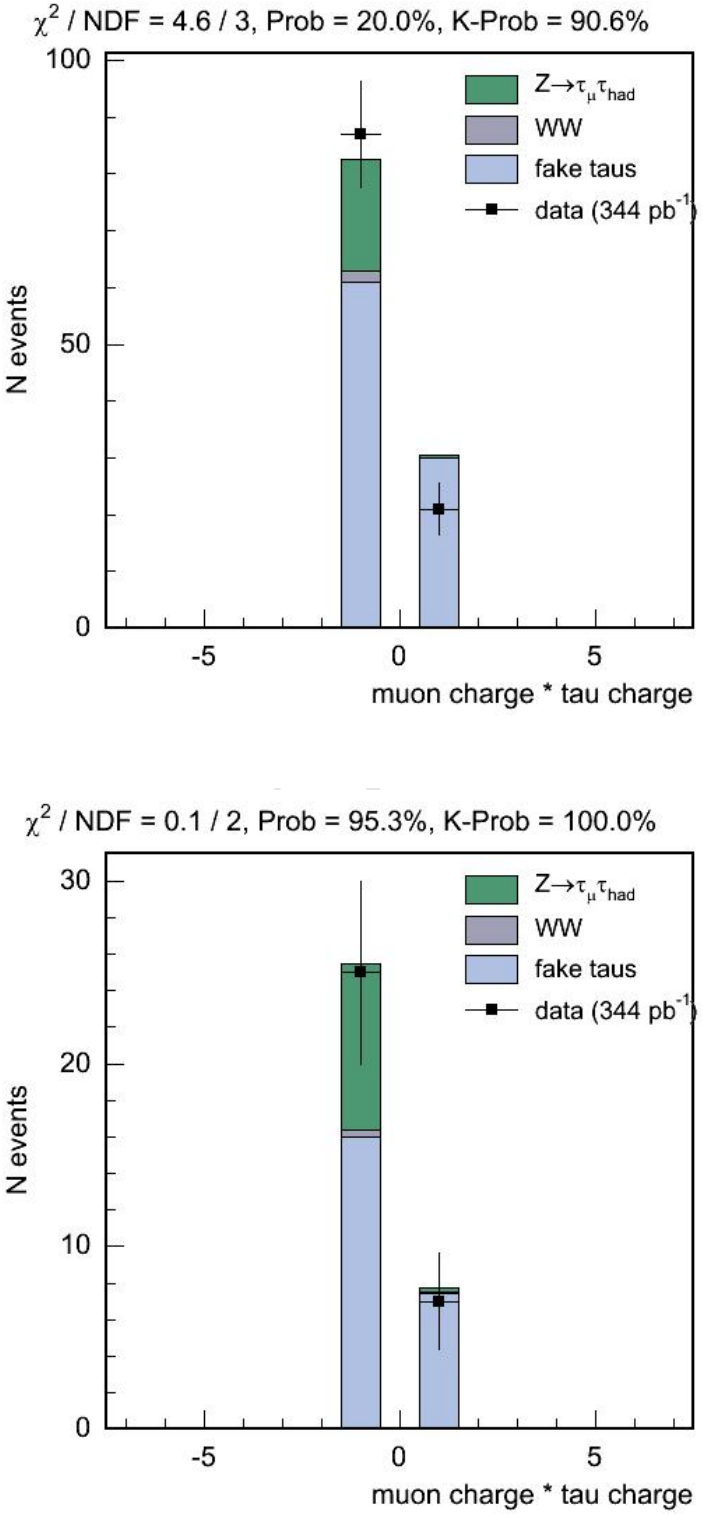


Figure 38: Charge product of the muon and the tau for events from the high p_T muon trigger sample, having 1 identified muon, 1 identified tau and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

DRAFT 2007

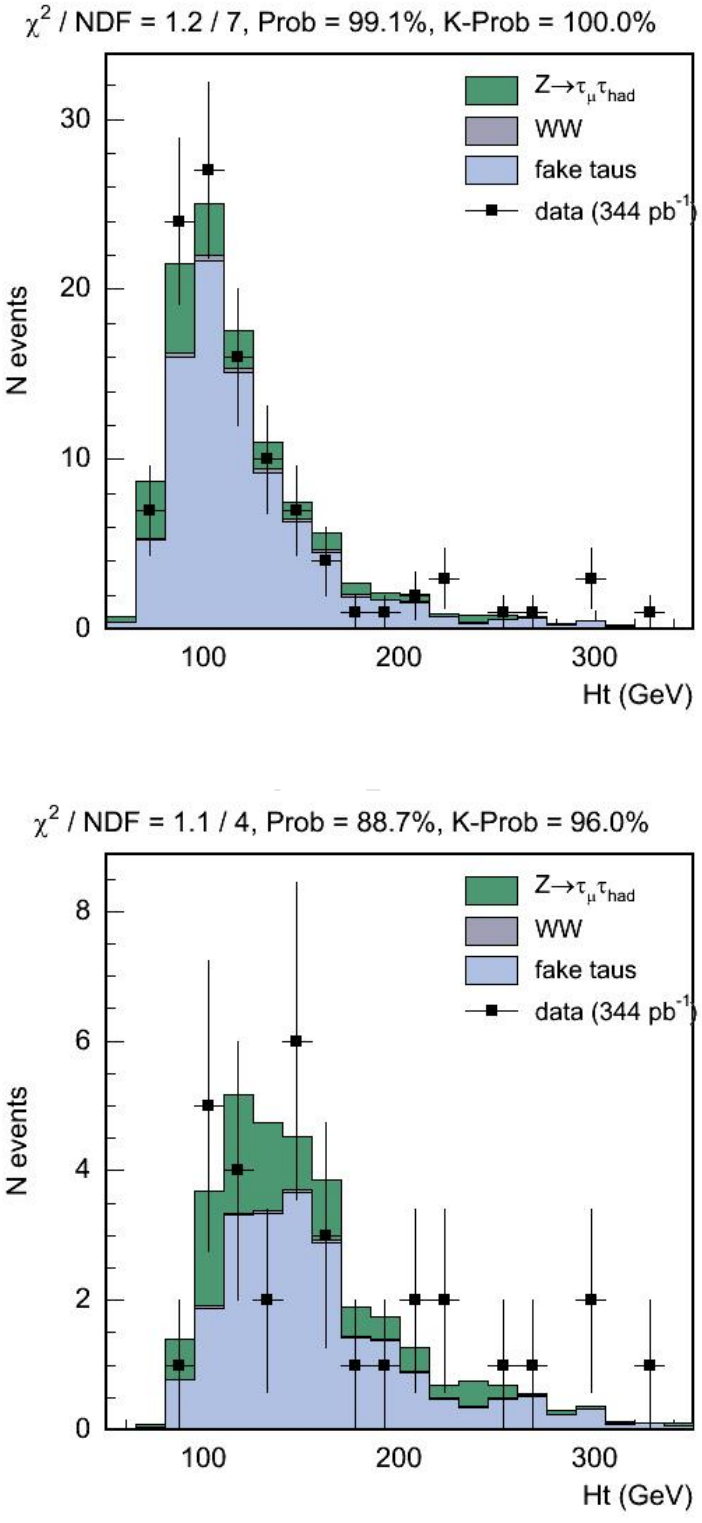


Figure 39: Event activity Ht for events from the high p_T muon trigger sample, having 1 identified muon, 1 identified tau and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

DRAFT 2007

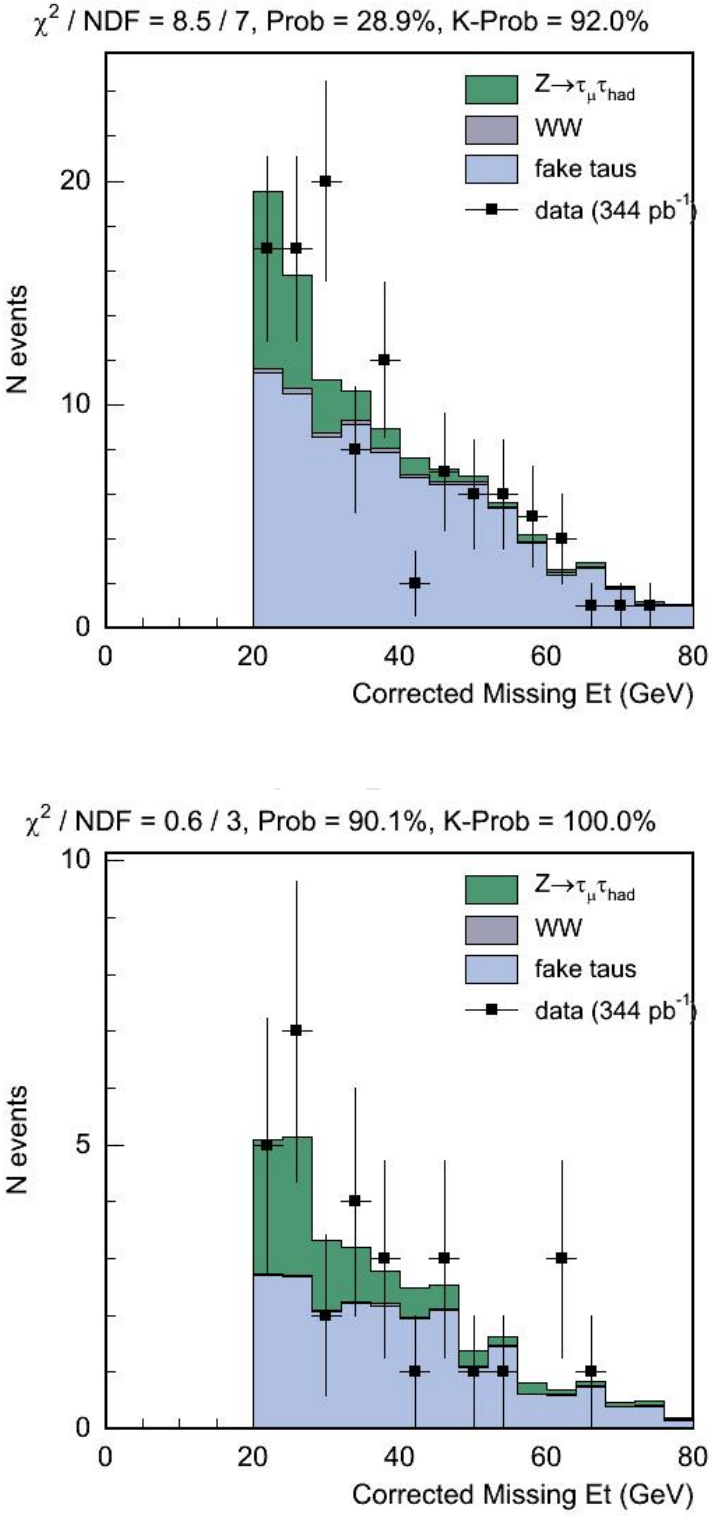


Figure 40: Event transverse Missing Energy for events from the high p_T muon trigger sample, having 1 identified μ, 1 identified τ and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

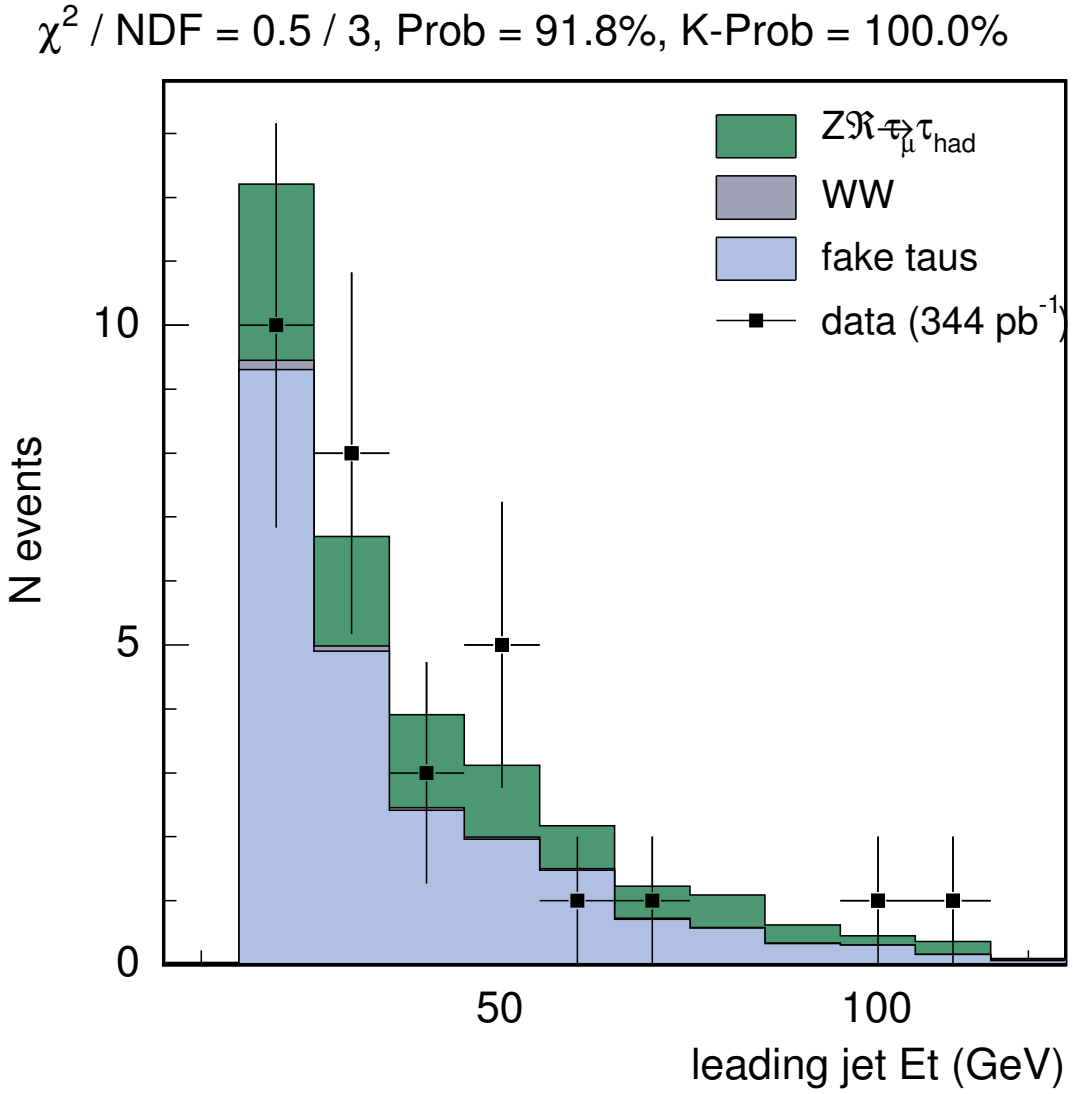


Figure 41: Leading jet E_T for events from the high p_T muon trigger sample, having 1 identified μ, 1 identified τ and MET>20 GeV and at least 1 other jet.

DRAFT 07

3 Sensitivities and final results

3.1 Systematic uncertainties

The obtention of the systematic error of 30% for the jet to tau fake rate has been described here in the subsection 3.7.

The systematic error of 25% taken for the electron faking taus background has been described in the note [22].

For the two Monte Carlo based backgrounds (Z and WW), the systematic error is the sum of the error due to the limited size of the Monte Carlo sample used and of the uncertainty inherent to the Monte Carlo. A 5% error obtained from the N jets study (subsection 3.6.1.2) summarizes the systematic error on the number of events passing the first part of the event selection (1 lepton, 1 ID tau, 2 jets with energies higher than 15 GeV). The subsequent cuts concern the missing E_T, the energy of the leading jet, the Ht cut, the product of lepton charges and the Z rejection. The efficiency for passing these last cuts are estimated by AlpGen+Herwig (Z case), and by Pythia (WW). The study of missing E_T and $Z \rightarrow \mu\mu + X$ (cf subsections 3.4.4 and 3.7.1.1) suggests that the part of systematic error due to $\not\!\!E_T$ is small. The very good χ^2/d.o.f. obtained in the N jets control regions for the Ht, charge product and leading jet distributions (cf subsection 3.9) favor a reasonable value for this systematic error. A systematics concerning the probability of passing the Z veto could be evaluated by comparing lepton angles and MEt direction in Z events with high Ht. Noticing the low value of these two backgrounds compared to the data-driven backgounds, such a study is not considered a priority and a conservative assumption of a systematic error of 30% is rather made for the two MC based backgrounds.

The knowledge of the error on the signal acceptance is not needed to achieve the first goal of this analysis, that is the mere establishment of the existence of the top tau dilepton signal. Indeed, in order to compute a $p-value$, only the expected number of background events and the number of observed events are used. However, it is needed for the measurement of the ratio $r_\tau = \frac{BR(t\rightarrow\tau\nu b)}{BR(t\rightarrow l\nu b)}$. Since the same simulation package and same definition of electrons and muons have been used in this analysis as in the $t\bar{t}$ dilepton analyses, we inherit from the systematic uncertainty calculated in [54] for the non-tau part. We have an additional 9% uncertainty for the hadronic tau identification efficiency, as explained in the subsection 3.3.2.3 about the $Z \rightarrow \tau\tau$ signal extraction. The uncertainties on the signal acceptance are gathered in Table 8.

DRAFT OCTOBER 2, 2007

Source	Systematic Error (%)
Monte Carlo Generator	2.4
ISR/FSR	4.4
PDF's	0.8
Jet Energy Scale	3.1
Multiple Interactions	1.7
Electron and muon identification	4.0
Hadronic tau identification	9.0
Total	11.7

Table 8: Uncertainties affecting the $t\bar{t}$ acceptance. The total error is the sum in quadrature of each contribution.

DRAFT OCTOBER 2, 2007

3.2 Sensitivities and results with 1 fb^{-1}

3.2.1 1 fb^{-1} analysis expectations and sensitivity

The loose likelihood selection In the table 10, are summarized the predicted signal and background events for the 1 fb^{-1} analysis, using the loose likelihood cut (Log likelihood > -5).

The $Z \rightarrow ee$ background is the result of the electron to tau fake rate folded in the electron sample only. The $Z \rightarrow \mu\mu$ background is estimated from Monte Carlo simulations. Using ten millions events simulated with Pythia, we do not find any event passing the signal selection. One event found in the simulation would correspond to a 0.03 event prediction in the data. Thus we quote in the acceptance table 0.0 ± 0.03 for this small background.

The most probable value for S+B is 14. This is the most probable number of events to be observed. Following the same method as explained above, we calculate the expected p-value to be 4.9%. This is the probability for the background alone to fluctuate up to 14 events or more.

Also, we calculate the 95% Confidence Level (CL) upper limit we can expect to set on the ratio $r_\tau = \frac{t\rightarrow\tau\nu q}{t\rightarrow l\nu q}$ (l=e or μ). We use a frequentist approach for this calculation. We need to answer the following question: Assuming the true ratio r_τ is equal to the measured value ($r_\tau^{meas.} = \frac{N_{obs}}{N_{SM}}$), where N_{obs} is the number of observed events minus the number of background events predicted, and N_{SM} is the standard model expectation (ie. the value S), what is the value of r_τ^{sup} so that 95% of the experiments performed in the same condition (same analysis, same detector, same luminosity) would measure a ratio r_τ smaller than r_τ^{sup}? Under the assumption that 14 events are actually observed, taking our predictions for S and B (see Table 10), a value or $r_\tau^{sup} = 2.1$ is found, meaning that we expect to exclude r_τ values higher than 2.1 at 95% CL. This is a measure of the sensitivity of this analysis to the r_τ measurement, but, as for the p-value measurement, the result can be greatly better or worse depending on the actual number of events observed in data.

The tight likelihood selection In the table 11, are summarized the predicted signal and background events for the 1 fb^{-1} analysis, using the tight likelihood cut (Log likelihood > 0).

At this level of background rejection, a subtlety about the use of the jet to tau fake rate in the data becomes apparent. Indeed, the jet to tau fake rate can not be folded directly into the data because of some true *tau* contamination in the denominator events sample. To get round this problem, the fake rate is only applied to those denominator tau-jets that fail at least one of the tau identification numerator requirements.

The most probable value for S+B is 3. This is the most probable result for the observed number of events. Following the same method as explained above, we calculate the expected p-value for the backgound alone (0.44 ± 0.14) to fluctuate up to 3 or more events to be 1.2%. In the table 9, we gather the different p-values that can be got, depending on the number of events really observed in data, with the correspond-

DRAFT OCTOBER 2, 2007

Number of actually observed events	Probability of this observation	p-value	evidence for signal
0	0.04	100%	5.7%
1	0.13	35%	32%
2	0.20	7.6%	76%(1.17 σ)
3	0.22	1.2%	95.3%(1.99 σ)
4	0.18	0.16%	99.2%(2.65 σ)
5	0.12	0.018%	99.86%(3.19 σ)
6	0.06	0.0018%	99.97%(3.62 σ)
7	0.03	0.00016%	99.995%(4.06 σ)
8	0.01	0.000012%	99.9991%(4.44 σ)

Table 9: *Summary of expectations for the search for top tau dilepton signal with* $Log(L0_{10} > 0)$ *and 1.05* fb^{-1}. *The probability of each observation is calculated with the assumption that the standard model is valid, ie the top tau dilepton signal exists with the expected cross section. P-values are the probabilities for the background alone to have fluctuated up to the number of observed events or more. Signal evidences are the probabilities for the signal existence, knowing the number of actually observed events, and assuming a prior probability of 0.5/0.5 for the existence or non-existence of the top tau dilepton signal.*

ing probability for it to happen. The result shows a very strong dependence with the number of events actually observed.

DRAFT OCTOBER 2 2007

	Electron + tau $1.05\ fb^{-1}$	Muon + tau $1.05\ fb^{-1}$
jet$\rightarrow\tau$ fakes	$3.4 \pm 0.4_{stat.} \pm 1.1_{sys.}$	$1.0 \pm 0.2_{stat.} \pm 0.3_{sys.}$
$Z \rightarrow ee$, $Z \rightarrow \mu\mu$	$0.24 \pm 0.08_{stat.} \pm 0.6_{sys.}$	0.0 ± 0.03
$Z \rightarrow \tau_{\rightarrow l}\tau_{\rightarrow had}$+jets	$1.3 \pm 0.1_{stat.} \pm 0.4_{sys.}$	$1.1 \pm 0.1_{stat.} \pm 0.3_{sys.}$
$WW \rightarrow \tau\nu_\tau l\nu_l + jets$	$0.13 \pm 0.01_{stat.} \pm 0.04_{sys.}$	$0.12 \pm 0.01_{stat.} \pm 0.04_{sys.}$
Total Background	**5.1**$\pm 0.4_{stat.} \pm 1.3_{sys.}$	**2.2**$\pm 0.2_{stat.} \pm 0.4_{sys.}$
	7.7$\pm$1.7	
SIGNAL ($t\bar{t} \rightarrow l + \tau$) (assuming $\sigma(t\bar{t}) = 7.3$ pb and $m_{top} = 175$ GeV)	$3.8 \pm 0.1_{stat.} \pm 0.4_{sys.}$	$2.6 \pm 0.1_{stat.} \pm 0.3_{sys.}$
	6.4 ± 0.7	

Table 10: *Summary of predicted signal and backgrounds in 1 fb^{-1} with the loose likelihood cut Log($L0_{10}$) > -5.*

	Electron + tau $1.05\ fb^{-1}$	Muon + tau $1.05\ fb^{-1}$
jet$\rightarrow\tau$ fakes	$0.18 \pm 0.09_{stat.} \pm 0.05_{sys.}$	$0.05 \pm 0.05_{stat.} \pm 0.02_{sys.}$
$Z \rightarrow ee$, $Z \rightarrow \mu\mu$	$0.06 \pm 0.03_{stat.} \pm 0.015_{sys.}$	0.0 ± 0.03
$Z \rightarrow \tau_{\rightarrow l}\tau_{\rightarrow had}$+jets	0.01 ± 0.01	0.01 ± 0.01
$WW \rightarrow \tau\nu_\tau l\nu_l + jets$	$0.05 \pm 0.01_{stat.} \pm 0.02_{sys.}$	$0.05 \pm 0.01_{stat.} \pm 0.02_{sys.}$
Total Background	$0.30 \pm 0.10_{stat.} \pm 0.06_{sys.}$	$0.11 \pm 0.06_{stat.} \pm 0.03_{sys.}$
	0.44$\pm$0.14	
SIGNAL ($t\bar{t} \rightarrow l + \tau$) (assuming $\sigma(t\bar{t}) = 7.3$ pb and $m_{top} = 175$ GeV)	$1.7 \pm 0.06_{stat.} \pm 0.16_{sys.}$	$1.1 \pm 0.04_{stat.} \pm 0.10_{sys.}$
	2.8 ± 0.3	

Table 11: *Summary of predicted signal and backgrounds in 1 fb$-$1 with the tight likelihood cut Log($L0_{10}$) > 0.*

3.2.2 Observation in data: 1 fb^{-1} result

Measurement of the r_τ ratio In order to measure the r_τ ratio, we open the box in the 1050 pb^{-1} data, looking at the number of events surviving the loose likelihood selection (Log(L0) > -5). We observe **11** events compatible with tau dilepton events, summarized in the table **??**. There are 8 events in the electron channel and 3 events in the muon channel.

This is in good agreement with the standard model expectations.

Since the prediction for the number of background events is $B = 7.7 \pm 1.7$, these 11 events give a measured number of signal events of $S_{meas.} = 3.3 \pm 1.7$. Since the predicted number of signal events from the standard model is $S = 6.4 \pm 0.7$, we measure a ratio $r_\tau = \frac{3.3 \pm 1.7}{6.4 \pm 0.7}$, ie. $r_\tau = 0.52 \pm 0.49_{stat.} \pm 0.29_{sys.}$.

Taking the observed number of events (11) as a central value for the "true" S+B, and

DRAFT OCTOBER 2, 2007

taking onto account the uncertainties on the signal and the background predictions, we find that 95% of experiments will observe 17 events or less. This results to a measured superior limit on the ratio r_τ:

$$\mathbf{r_\tau < 1.5} \text{ at } 95\% \text{ CL.}$$

The eleven events were scrutinized. The figures 42 and 43 present the distributions for seven kinematical variables, H_T, $\not\!\!E_T$, τ-jet E_T, lepton (electron or muon) E_T, the two leading jet E_T, and the L0 likelihood distribution. These are superimposed on the standard model predictions that are the sum of the distributions for the signal and the background, in order to contol that nothing is abnormal. The distributions don't show any obvious disagreement.

Search for the top tau dilepton signal We apply the tight likelihood selection to the 1 fb^{-1} sample and find **2 events**. The two events are found in the electron channel.

We conclude (see Table 9) that the measured p-value is **7.6%**, and that we got a **1.17 sigma evidence** for the signal.

4 Conclusion and perspectives

This analysis is very near to reach the sensitivity for a 3 σ evidence for the top tau dilepton signal. It already gives an interesting information about the r_τ ratio : $r_\tau < 1.5$ at 95% CL.

It is quite clear that it is worth to pursue this analysis with the integral luminosity increase already achieved and which should at least double the statistics for summer 2007.

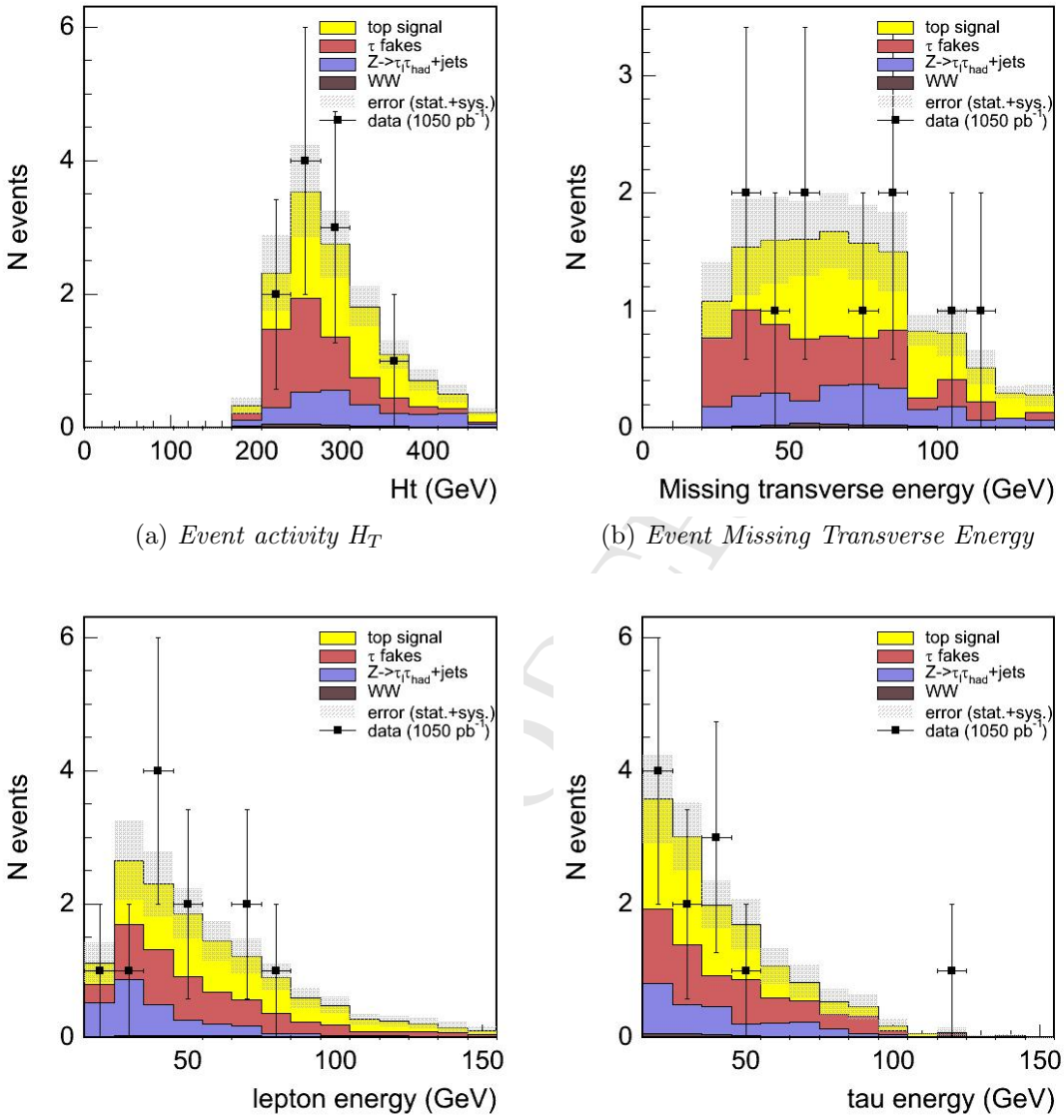
For the longer term, a factor at least of 4 in integrated luminosity will provide on the order of 60 signal plus background events and therefore allow to clarify the situation before the start of LHC, maybe giving sure indication of a possible excess for the r_τ ratio value.

It is also quite important to use this analysis for searching for a charged Higgs in the top decay channel $t \to H^+ b$. The charged Higgs decay into $\tau + \nu$ will be the dominant contribution in this key analysis that will thus use our analysis result at first.

It would be important at this stage to better tune the Monte Carlo. A Monte Carlo providing good jet shapes, and good shower shapes in the calorimeter, could open the way to a new estimation of backgrounds containing fake taus, and thus maybe make it possible to use very discriminant variables in the likelihood method, like the tau-jet isolation and mass, that are in our approach not usable because of their use in the tau lepton identification, and especially in the numerator of the jet to tau fake rate.

It would also be interesting at this stage to compare our results with a very different and complementary analysis performed in parrallel that made use of B tagging and looser likelihood-based tau identification.

DRAFT OCTOBER 2, 2007



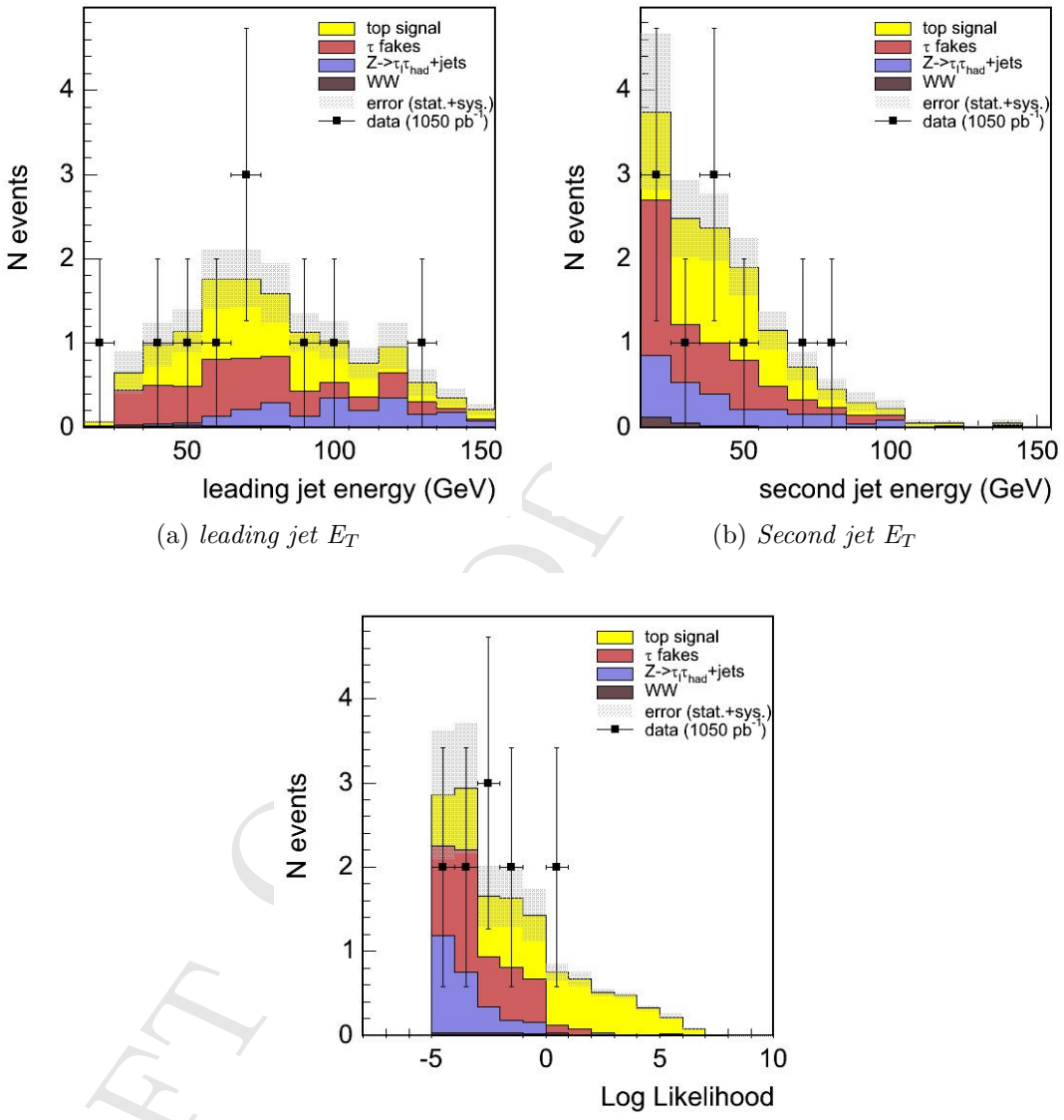
(a) *Event activity* H_T (b) *Event Missing Transverse Energy*

(c) *electron or muon* E_T (d) τ*-jet* E_T

Figure 42: Comparison of the 5 observed events distributions with the standard model expectations.

Finally, LHC will hopefully be the place to discover a charged Higgs (if any). In any case, it will be a top factory allowing to study in details this tricky top decay channel.

DRAFT



(a) *leading jet* E_T

(b) *Second jet* E_T

(c) *L0 likelihood*

Figure 43: Comparison of the 11 observed events distributions with the standard model expectations.

DRAFT 07

References

[1] M. Cacciari et al., The $t\bar{t}$ Cross-Section at 1.8 and 1.96 TeV: A Study of the Systematics due to Parton Densities and Scale Dependence, JHEP 0404:068 (2004)

[2] M. Hohlmann, Observation of top quark pairs in the dilepton channel using electrons, muons and taus, Ph.D. Thesis, University of Chicago, August 1997

[3] http://www-ecool.fnal.gov

[4] The CDF II Detector - Technical Design Report, FERMILAB-Pub-96/390-E

[5] L. Balka et al., Nucl.Instr.Meth. A267 (1988) 272

[6] M. Albrow et al., Nucl.Instr.Meth. A480 (2002) 524

[7] Bertolucci et al., Nucl.Instr.Meth. A267, 301

[8] G. Corcella et al., HERWIG 6: An Event Generator for Hadron Emission Reactions with Interfering Gluons (including supersymmetric processes), JHEP **01**, 10 (2001)

[9] T.Sjostrand et al., High-Energy-Physics Event Generation with PYTHIA 6.1, Comput. Phys. Commun. **135**, 238 (2001)

[10] M.Mangano et al., ALPGEN, a generator for hard multiparton processes in hadronic collisions, JHEP **0307:001** (2002)

[11] S. Jadach et al., TAUOLA 2.5 Preprint CERN-TH-6793 (1992)

[12] R. Field et al. (for the CDF collaboration), PYTHIA Tune A, HERWIG, and JIMMY in Run 2 at CDF, Preprint hep-ph/0510198, to appear in the proceedings of HERA-LHC workshops (2005)

[13] R. Wagner, Electron Identification in Run II: Understanding and Using *Lshr*, CDF/DOC/ELECTRON/6249 (2003)

[14] R. Harris, R. Blair, S. Kuhlmann, CES Response and χ^2 for 1990 Testbeam Electrons, CDF/ANAL/ELECTRON/CDFR/1432 (1991)

[15] A. Anastassov, Non-isolated π^0/γ reconstruction, /CDF/ANAL/EXOTIC/CDF/6688 (2003)

[16] S. Demers et al., Event Selection and Acceptance for Top to Tau Dilepton Analysis, internal CDF note 6921, March 2004

[17] J. Insler et al., Determining the Electron Fake Rate to Hadronic taus from the Data, CDF note 6408, May 2003

DRAFT OCTOBER 2, 2007

[18] CDF collaboration, A search for $t \to \tau\nu q$ in t anti-t productionFERMILAB-PUB-05-484-E, Submitted to Phys.Rev.Lett.

[19] S. Demers Konezny, A Measurement of BR(top-¿tau+neutrino+quark), Phd Thesis, University of Rochester, 2004

[20] S. Demers et al., A Z mass cut to reduce the $Z\gamma^* \to \tau\tau$ background in the $t\bar{t}$ tau dilepton analysis, CDF internal note CDF/ANAL/TOP/CDFR/6922, March 2004

[21] CDF note 8148, Feb.23 2006

[22] S. Demers et al., Determining the Electron Fake Rate to Hadronic taus from the Data, internal CDF note 6408, April 2003

[23] C. Hill, J. Incandela and C. Mills, Electron Identification in Offline Release 5.3, internal CDF note 7309, February 2005

[24] S. Tourneur and A. Savoy-Navarro, jet to tau fake rate, cdf note 8208

[25] S.Tourneur et al., Tau fake rate, Tau meeting, 12/01/2005

[26] S.Tourneur et al., Tau fake rate, Tau meeting, 12/08/2005

[27] T. Vaiciulis et al., Determining the Jet Fake Rate to Hadronic taus from the Data, CDF note 6784, May 2003

[28] http://www-cdf.fnal.gov/internal/people/links/StephaneTourneur/tauIDplots/list.html

[29] S.Tourneur, Top Properties Workshop, 11-10-2005

[30] M. Tecchio et al., Measurement of $t\bar{t}$ cross section in Dilepton plus SecVtx tagged events, internal CDF note 8157, March 2006.

[31] CDF II Collaboration, D.Acosta et al., Measurement of the $t\bar{t}$ Cross Section in $p\bar{p}$ Collisions in $\sqrt{s} = 1.96$ TeV using Dilepton Events, Phys.Rev.Lett.93:142001, 2004.

[32] V. Martin, High p_T Muon ID Cuts and Efficiencies for use with 5.3.1 DATA and 5.3.3 MC, internal CDF note CDF/DOC/MUON/CDFR/7367 (2005).

[33] Y. Dokshitzer et al., Basics of perturbative QCD, ed. J.Tran Thanh Van, Editions Fontières, Gif-sur-Yvette (1991).

[34] Ya.I. Azimov et al., Z.Phys. C **27** (1985) 65; Z.Phys. C **31** (1986) 213.

[35] S.Catani et al., Longitudinalli Invariant K(T) Clustering Algorithms for Hadron-Hadron Collisions, Nucl.Phys.B406:187-224 (1993)

DRAFT OCTOBER 2, 2007

[36] CDFII Collaboration, A.Abulencia et al., Measurement of the Inclusive Jet Cross Section Using the kT Algorithm in $p\bar{p}$ Collisions at $\sqrt{s}$=1.96 TeV , Phys. Rev. Lett. **96**, 122001 (2006)

[37] A. Bhatti et al., Determination of the jet energy scale at the collider detector at Fermilab, Nucl.Instr.Meth.A566:375-412 (2006)

[38] K. Hagirawa et al., Physical Review **D 66**, 010001 (2002)

[39] S. Klimenko, J. Konigsberg, and T. M. Liss, Averaging of the Inelastic Cross Sections Measured by the CDF and the E811 Experiments, FERMILAB-FN-0741 (2003).

[40] CDF I Collaboration, F. Abe et al., Phys. Rev. **D50**, 5550 (1994).

[41] C. Avila et al., Phys. Lett. **B445**, 419 (1999).

[42] A. Capella et al., Zeit. Phys. C3 (1980) 329.

[43] A. Capella et al., Phys. Reports 236 (1994) 225.

[44] CDF II Collaboration, D. Acosta et al., First Measurements of Inclusive W and Z Cross Sections from Run II of the Tevatron Collider, Phys. Rev. Lett. 94 (2005) 091803

[45] P. Sutton, A. Martin, R. Roberts, and W. Stirling, Phys. Rev. **D45**, 2349 (1992).

[46] V. Boisvert, Trigger Efficiencies for the High E_T Central Electrons in the Gen6 data, CDF internal note CDF/DOC/ELECTRON/CDFR/7939 (2006)

[47] Y. Ishizawa and J. Nielsen, Trigger Efficiencies for High E_T Electrons, CDF internal note CDF/DOC/ELECTRON/CDFR/7401 (2005)

[48] T. Spreitzer et al., Electron Identification in Offline Release 6.1.2, CDF internal note CDF/DOC/ELECTRON/CDFR/7950 (2006)

[49] U. Grundler et al., High-P_T muons recommended cuts and efficiencies for Summer 2006, CDF internal note CDF/ANAL/TOP/CDFR/8262 (2006)

[50] http://www-cdf.fnal.gov/internal/physics/top/RunIITopProp/gen6Sum06/lumi.html

[51] S. Baroiant et al., Selection of tau leptons with the CDF Run II trigger system, Nucl.Instr.Meth.A518:609-611 (2004)

[52] A. Anastassov et al., Extraction of $Z \rightarrow \tau\tau$ Signal using Run II Lepton+Track Trigger, Electron Channel, CDF internal note, CDF/ANAL/EXOTIC/CDFR/6402 (2003)

DRAFT OCTOBER 2, 2007

[53] S. Baroiant et al., Cross Section Measurement for Z to tau tau in pp-bar Collisions at $\sqrt{s}$=1.96 TeV, CDF/PUB/ELECTROWEAK/CDFR/8132, to be submitted to Phys. Rv. D (2006)

[54] The CDF Collaboration, A Measurement of the $t\bar{t}$ Production Cross Section Using Dileptons, CDF public conference note, CDF/TOP/8103 (2006)

[55] J. Rosner, M.P. Worah, and T. Takeuchi, Phys. Rev. D 49, 1363 (1994).

DRAFT OCTOBER 2, 2007