UNIVERSITE PIERRE & MARIE CURIE
LA SCIENCE A PARIS

Thèse de doctorat de l'Université Paris 6

spécialité
Champs Particules Matières

présentée par
Stéphane TOURNEUR

pour obtenir le grade de
Docteur de l'Université Paris 6

Search for top pair production in the dilepton decay channel with tau lepton at the CDF experiment

Recherche de la production de paires de quarks top dans le canal de désintégration dileptonique avec lepton tau dans l'expérience CDF

soutenue le 17 novembre 2006 devant le jury composé de :

M. John Conway (président)
M. Pascal Debu
M. Juan Fuster
M. Michael Joyce
M. Luigi Rolandi (rapporteur)
Mme Aurore Savoy-Navarro (directeur de thèse)
M. Guillaume Unal (rapporteur)

Contents

Résumé de thèse

Le sujet de cette thèse est la recherche des désintégrations du quark top impliquant des leptons tau au détecteur CDF, dans les collisions proton-antiproton à 1.96 TeV du Tevatron.

Cette analyse nécessite la maîtrise de l'identification de tous les objets fondamentaux (électrons, muons, désintégrations hadroniques du tau) et une bonne reconstruction de l'énergie des jets et de l'énergie transverse manquante, dans un contexte difficile qui est celui des collisions proton-antiproton. Une bonne compréhension des fonds principaux ($Z \rightarrow \tau_e \tau_{had}$, événéments avec un jet imitant un tau, diboson WW) est un prérequis.

Les buts d'une telle étude sont triples : est visée tout d'abord une mise en évidence de l'un des derniers canaux de désintégration du top non encore observé, à savoir la désintégration en lepton τ accompagné d'un quark de beauté et d'un neutrino. L'autre objectif est d'obtenir une estimation du rapport $r_\tau = \frac{BR(t \rightarrow \tau \nu q)}{BR(t \rightarrow l \nu q)}$; si une déviation significative par rapport à 1 de ce rapport s'avérait être à la portée avec la luminosité disponible, cela serait une indication de physique au-delà du modèle standard. Un autre aspect important de cette analyse vient de ce qu'une bonne compréhension des signatures comprenant plusieurs jets, de l'énergie manquante, des leptons, en particulier des leptons tau, est cruciale pour la recherche de la supersymétrie. En particulier, dans le cas de modèles SUSY à grand $tan\beta$, le taux de désintégrations en leptons tau est prédit devenir dominant. Cette analyse requiert aussi un taux élevé de données pour pouvoir extraire le signal recherché d'un fond d'autres processus physiques très élevé.

L'analyse des premières données collectées lors du Run II de CDF s'est faite sur le sous-ensemble des événements comprenant un électron ou un muon de plus de 20 GeV dans la partie centrale du détecteur, pour laquelle la reconstruction des traces bénéficie du trajectographe central de haute précision dans la détection des traces et de leur impulsion. Ce trajectographe constitue un élément essentiel pour identifier les leptons tau dans leur mode de désintégration en un ou plusieurs pions chargés, dite désintégration hadronique. L'accent a d'abord été mis sur la détermination de l'acceptance du signal top à l'aide des données et de simulations Pythia et Herwig, et sur le calcul du fond électrofaible avec des échantillons simulés par Pythia et par le générateur Monte-Carlo à éléments de matrices AlpGen. Ensuite, une grande partie du travail a été dédiée à la recherche de nouvelles méthodes améliorant la détermination des difficiles fonds QCD et W+jets où un jet est faussement reconnu comme une désintégration hadronique du lepton tau. Ce fond constitue la première source d'erreur systématique de la mesure. Une méthode a pu être mise au point qui s'affranchit de conjectures hasardeuses utilisées dans les mesures précédentes et améliore la précision sur l'estimation des fonds QCD et W+jets. La probabilité pour un jet d'être identifié par erreur comme un lepton tau est extraite d'échantillons de données

réelles dominées par des événements comprenant deux jets opposés. La validité de ce "taux de jet en tau" dans des ensembles de données dominés par des événements de type W+jets est ensuite démontrée jusqu'à une multiplicité de jets égale à trois, avec une précision supérieure à 30%.

Une première analyse, effectuée sur les premiers 350 pb^{-1} de données collectées jusqu'à septembre 2004 a été officialisée par l'expérience CDF et ainsi rendue publique. Cette analyse utilise une méthode de discrimination du signal par rapport au fond au moyen de coupures séquentielles. Il s'agit d'une analyse non biaisée, effectuée en "aveugle", c'est-à-dire que l'on s'interdit de regarder les données dans la zone de signal avant d'avoir complètement validé et entériné notre méthode de prédiction du nombre d'événements observés. La même méthode de sélection du signal est ensuite appliquée au lot de données, résultant sur une comparaison du nombre d'événements effectivement observé avec les prédictions. Par cette méthode, nous avons pu faire une prédiction de 2.7 ± 0.7 événements de fonds pour 2.2 ± 0.3 événements dus au signal de top en tau. Nous avons ensuite observé l'existence de 5 événements candidats dans ce premier lot de données. La probabilité pour le fond seul de fluctuer à 5 événements ou plus est de 16% (p-valeur), ce qui constitue une évidence de 1 sigma pour le signal de top en tau.

Cette étude s'est poursuivie et a donné lieu à une nouvelle analyse effectuée sur les 1 fb^{-1} de données collectées jusqu'à février 2006 par l'expérience CDF. La méthode d'estimation du bruit de fond et du signal reste inchangée. La sélection discriminante du signal par coupures séquentielles est quant à elle remplacée avantageusement par une méthode statistique de rapport de vraisemblance. La construction de cette nouvelle variable de vraisemblance utilise dix variables cinématiques choisies pour leur pouvoir discriminant entre les principaux fonds ($Z \rightarrow \tau\tau$ +jets, W+jets et $t\bar{t} \rightarrow l\nu q\bar{q}b\bar{b}$) et le signal. Par cette méthode, nous parvenons à une sensibilité s'approchant des 3 sigmas pour la mise en évidence du signal de top en tau. La prédiction sur le nombre d'événements de signal est de 3.2 ± 0.3 et celle sur le fond de 0.4 ± 0.1 dans l'ensemble du lot de données de 1 fb^{-1}. Deux événements étant observés dans les données, une p-valeur de 8% est obtenue, correspondant à une mise en évidence du signal de 1.2 sigma. Une mesure du rapport $r_\tau = \frac{BR(t\rightarrow\tau\nu q)}{BR(t\rightarrow l\nu q)}$ a également été conduite dans cet échantillon en relaxant la sélection sur le rapport de vraisemblance. Quatorze événements sont attendus et onze événements sont effectivement observés, rendant possible l'établissement d'une nouvelle limite sur le rapport : $r_\tau < 1.5$ à 95% de niveau de confiance. Ce résultat améliore la précédente mesure également réalisée par CDF ($r_\tau < 5.2$ à 95% de niveau de confiance, résultat obtenu avec 195 pb^{-1} de données).

Summary

This thesis presents a search for quark to decays involving tau leptons with the CDF detector, among the proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV from the Tevatron.

The analysis needs to control the identification of many fundamental objects (electrons, muons, tau hadronic decays) and a good reconstruction of the jet energy and missing transverse energy, in the difficult environment of the proton-antiproton collisions. A very good understanding of the main backgrounds ($Z \rightarrow \tau_e \tau_{had}$, events with jets faking a tau, WW diboson) is a prerequisite.

The scientific goals of this study are threefold. At first, it aims at demonstrating the existence of one of the last top quark decays yet to be observed, that is the channel where the top decays into a tau lepton accompanied by a bottom quark and a neutrino. The other goal is to compute the ratio $r_\tau = \frac{BR(t \rightarrow \tau \nu q)}{BR(t \rightarrow l \nu q)}$; If a significant deviation from 1 happened to be observable with the available integrated luminosity, this could be an indication of physics beyond the standard model. Another important aspect of this work is that a good understanding of signatures involving several jets, large missing transverse energy, leptons, in particular tau leptons, is crucial for searches for supersymmetry. In particular, in the frame of SUSY models with high tan β, the branching ratios into tau leptons os predicted to become predominant. Also, this analysis requires a large integrated luminosity in order to be able to extract the searched signal from a very high background of other physics processes.

The analysis of the first data recorded during the CDF Run II was performed on a subset of events containing an electron or a muon in the detector central part with a transverse energy exceeding 20 GeV, for which the track reconstruction takes advantage of the whole depth of the central tracker. This central tracker is essential fot the tau lepton identification through its decay into one or several charged pions. The work has been firstly focused on the computation of the top signal acceptance, using real data and Pythia and herwig simulations, and on the $Z \rightarrow \tau\tau$ electroweak background estimation using the matrix elements Monte-carlo generator Alpgen.

Then, an important part of this thesis has been dedicated to the development of improved methods to estimate the large QCD and W+jets backgrounds that enter the signal sample by the way of a jet being mistaken as a tau hadronic decay. This background estimation accounts for a major fraction of the systematic error of this measurement. The developed method demonstrates the validity of unsure extrapolations usually used and improves the precision on the QCD and W+jets backgrounds estimation. The probability for a jet to be mistakenly identified as a tau lepton is measured in real data samples dominated by events with two back-to-back jets. This "jet to tau fake rate" is then proven to be applicable into data samples dominated by W+jets events up to a jet multiplicity

equal to three. This is established with an uncertainty smaller than 30%.

The first analysis, making use of the 350 pb^{-1} data recorded until September 2004 was endorsed by the CDF collaboration and thus made public. This analysis uses sequential cuts in order to extract the signal from the background. This is an unbiased analysis, performed following the "blind" analysis method, that is to say that any observation of the data in the signal region is forbidden before a thorough validation of the number of observed events prediction method has been made, and approved by the collaboration. While 2.7 ± 0.7 background events and 2.2 ± 0.3 signal events predicted, 5 events were actually observed in the data. The probability for the background alone (2.7 ± 0.7) to have fluctuated up to 5 events or more is 16% (p-value). This is equivament to 1-sigma evidence for the top into tau signal.

This study was followed up by the analysis of the whole 1 fb^{-1} data sample recorded until February 2006 by the CDF experiment. The method used to estimate the signal and the background remained unchanged. The cut based signal selection has been advantageously replaced by likelihood ratio method. This new likelihood variable construction used ten kinematical variables that were chosen because of their high discriminant power between the main backgrounds ($Z \rightarrow \tau\tau$ +jets, W+jets et $t\bar{t} \rightarrow l\nu q\bar{q}b\bar{b}$) and the signal. This method allows for a sensitivity nearing 3 sigma for the top tau dilepton signal evidence. The predictions are 3.2 ± 0.3 for signal events and 0.4 ± 0.1 for background events in the whole 1 fb^{-1} sample. Two events are actually observed, leading to a new p-value of 8%, corresponding to a 1.2 σ evidence for the signal. A measurement of the ratio $r_\tau = \frac{BR(t\rightarrow\tau\nu q)}{BR(t\rightarrow l\nu q)}$ is also performed in this sample by loosening the likelihood selection. Eleven events were observed in the data against 14 expected, making it possible to establish a new limit on the ratio: $r_\tau < 1.5$ at 95% Confidence Level. This result improves the previous measurement also done by CDF ($r_\tau < 5.2$ at 95% CL, using 195 $pb^{-}1$).

Chapter 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: $p\bar{p} \rightarrow t\bar{t} + X \rightarrow e, \mu + \nu_{e,\mu} + \tau_{had} + \nu_\tau + b\bar{b} + X$ 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 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

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 1.1: *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*

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 1.1 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 $S/\sqrt{B}$.

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, $r_\tau = \frac{t\rightarrow\tau\nu_\tau b}{t\rightarrow l\nu_l b}$ (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: $t \rightarrow bH^+ \rightarrow b\tau\nu_\tau$ 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.

Chapter 2

The Experimental complex

2.1 Introductory remarks

The data that are used for this research work were collected by the CDF experiment running at the Tevatron $p/\bar{p}$ collider at Fermi National Laboratory during the period from January 2002 to February 2006. This period corresponds to the first running period of the so-called Run II of the Tevatron. The Tevatron accelerator went through a dramatic upgrade for the Run II. It resulted in a slight but useful increase in the center of mass energy from 1.8 TeV up to 1.96 GeV, an order higher in luminosity expected to reach the nominal value of 2×10^{32} cm^{-2} s$^{-}1$ with a decrease in the bunch crossing time from 3 μs down to 396 ns.

To accomodate with this new accelerator conditions and with the new challenging physics goals, mainly in the B and top physics and in the exploration of the Higgs sector and New Physics, the CDF detector has undergone through an agressive upgrade program. This upgrade started at the end of Run I in 1995 and was completed in October 2001, right at the beginning of Run I. The upgraded CDF detector, so called CDF II included options that were foreseen to be only added in the course of Run II and indeed were ready at the start of Run II, namely, the Time Of Flight (TOF) and the innermost Silicon layer L00, which was added to the new vertex detector. It allowed the CDF physicists to achieve the commissioning of this almost entirely new detector at the same time the machine was doing its own commissioning run.

These upgrades of both the machine and the CDF detector are crucial to permit the research reported here. Indeed the search for the last dilepton channel of the $t\bar{t}$ pair production, namely the one including at least one tau lepton requires to fully exploit all these detector and machine improvements. This is essentially due to the fact that tau leptons are harder to identify and accurately measure especially in the environmental conditions of a $p/\bar{p}$ collider. The much lower detection efficiency of such objects thus requests both a highly performing machine (higher center of mass energy and/or higher luminosities will increase the production rate) and a highly performing detector. Indeed the process, $p\bar{p} \rightarrow t\bar{t} + X$, where at least one of the produced W decays into a tau and the associated neutrino ends up in signatures characterized by an electron or a muon, a tau lepton, a certain amount of missing transverse energy and of jet activity (at least the two b-jets associated to the W bosons from the two top decays). Besides the identification

of the tau lepton in a $p/\bar{p}$ collider environment can only be done through its so-called hadronic decay, i.e. when it decays into one or three charged pions and some neutral pions. The tau lepton is therefore materialized in a so-called tau-jet. From the list of elementary objects that are included in the signature of this process it is quite clear that the related analysis work will have to make the best profit of all the detector components. In counterpart the detector as well as the machine performances will be instrumental for the success of this research. For all these reasons it is indeed the first time that this study can be tackled with some good chances of success.

In the description of the various components of the CDF II detector we will point out its importance with respect to this analysis.

2.2 Accelerator complex

The accelerator complex is briefly described in this section. The main challenging upgrades it underwent for Run II are the construction of the new Main Ring and later of the electron cooling which is a real premiere and a key element in the success of the Run II. Without it, it would have not been possible to reach and perhaps even supersede the nominal value expected for this run period. It should be pointed out again that the increase in the center of mass energy although slight is already significant for this particular search.

The Tevatron at the Fermi National Accelerator Laboratory (FNAL) is a $p\bar{p}$ collider with a center of mass energy of $\sqrt{s} = 1.96$ TeV. It is located 30 miles west of Chicago.

The Tevatron is a multi-staged accelerator comprised of the following elements:

1. Proton source
    - Cockcroft-Walton pre-accelerator (preacc)
    - Linear accelerator (linac)
    - Booster
2. Antiproton source
    - Target
    - Debuncher
    - Accumulator
    - Recycler
3. Main Injector
4. Tevatron Colliding Ring

Figure 2.1 shows the components of the Tevatron. We now consider each element in turn. For a detailed explanation of the Tevatron, see Reference [3].

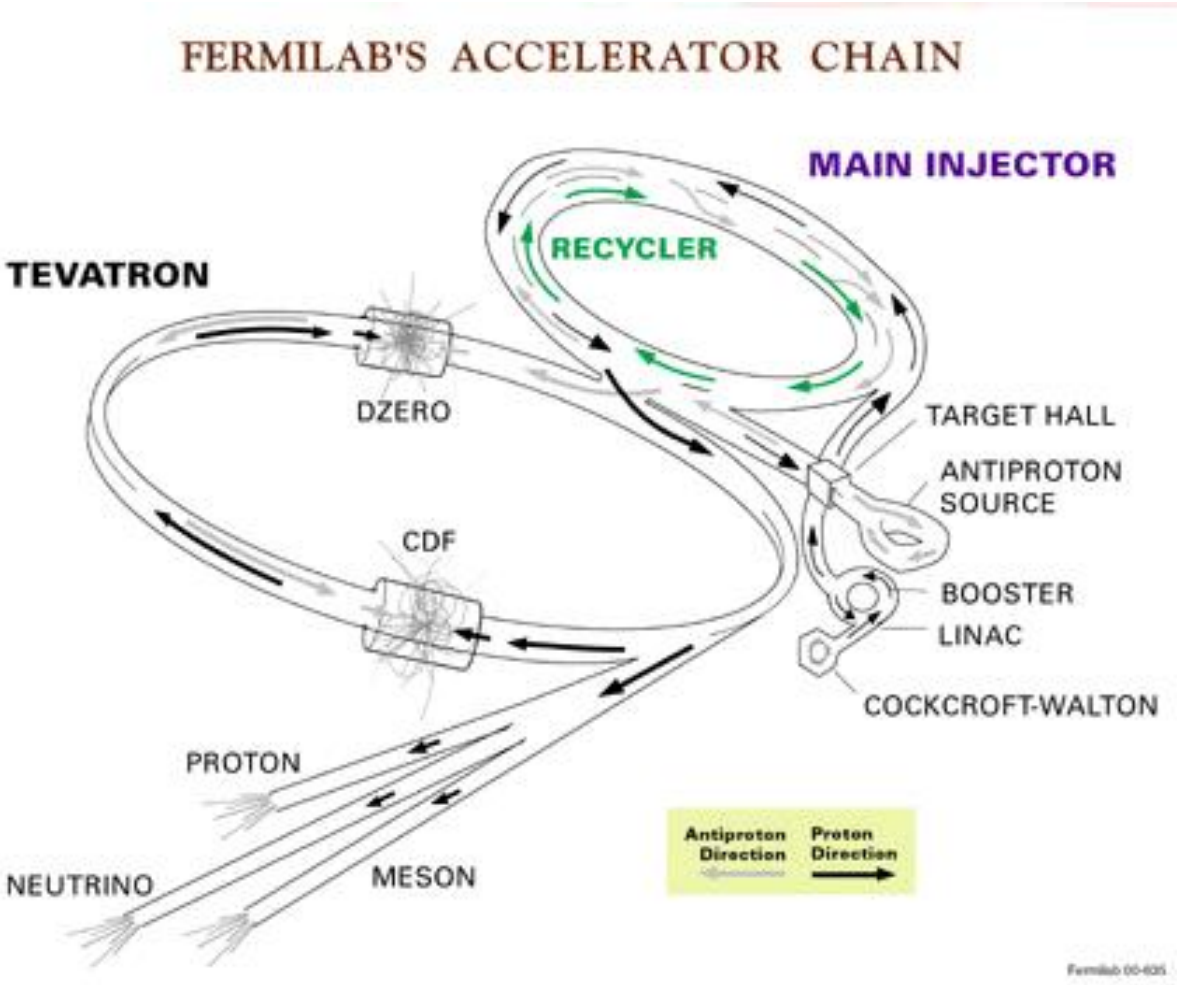


Figure 2.1: the accelerator chain of the Tevatron

2.2.1 Proton source

The proton source is made up of three subsystems, the pre-accelerator (preacc), the linear accelerator (linac), and the booster ring. Together, these subsystems are capable of producing 8 GeV protons to be injected into the Main Injector.

Cockcroft-Walton pre-accelerator The Cockcroft-Walton pre-accelerator (preacc) is a source of negatively charged, accelerated hydrogen ions. It takes hydrogen ions off a cesium target (made from hydrogen gas). The pre-accelerator produces 750-KeV hydrogen ions every 66 milliseconds. It then transfers the hydrogen ions into the linear accelerator (linac).

Linear accelerator The linear accelerator (linac) receives negatively charged hydrogen ions at 750 KeV and accelerates them to an energy of 400 MeV. This is accomplished using radiofrequency-pulses of magnetic fields, carrying the particles along on a "wave" of electromagnetic radiation. This is also done once every 66 milliseconds. The beam is also focused at this point using quadrupole magnets. This cooled beam of H^- ions is then sent to the booster.

Booster The booster receives the cooled beam of H^- ions from the linac, and strips the electrons off, leaving bare protons (H^+ ions). It then accelerates the protons to 8 GeV. This is the first synchrotron (circular accelerator) in the Tevatron complex. It is composed of a series of 75 magnets arranged around a 75 meter radius circle, with 18 RF cavities inside. This stage of production is also operated at 66 milliseconds, with sufficient phase offsets to catch the ions from the linac.

At this point, the beam is sent to a transfer line, which then sends the proton beam to the main injector, and then on to the Tevatron for colliding physics.

2.2.2 Main injector

The main injector is a circular synchrotron about half a kilometer in radius. It is composed of 18 accelerating RF cavities, and can accelerate 8 GeV protons from the booster to 150 GeV every 2.2 seconds, which are then passed to the Tevatron. It can also accelerate protons to 120 GeV which are then used to strike the antiproton source and create antiprotons.

2.2.3 Antiproton source

The antiproton source is made up of one target and three subsystems, the debuncher, accumulator, and the recycler. Together, the antiproton source is capable of producing antiprotons at 8 GeV, to be injected into the main injector. In principle, another source of antiprotons is "recycled" antiprotons from the last stack that are stored in the recycler. This system would significantly decrease the time in between stores, so it would be possible to integrate more luminosity. However, this functionality is not yet commissioned, and there is little hope that the antiprotons will ever be recycled.

Target The antiproton source is a nickel target. Protons from the main injector at 120 GeV strike the target, creating a spray of particles. The particles are then sent through a magnetic field, and the particles with different masses and charges will curve at different radii. The antiprotons are then selected from this spray. This is called magnetic spectroscopy. The antiprotons selected have a wide range of momenta, averaging 8 GeV.

Debuncher The debuncher is a rounded triangular synchrotron with a mean radius of 90 meters. It can accept 8 GeV protons from the main injector for studies, or 8 GeV antiprotons from the target station. It can take the antiprotons from the target, which have a large momentum spread, using a process called "bunch rotation" which is an RF-manipulation. It is also where stochastic cooling is performed on the antiprotons to "cool" the beam's transverse energy, making it more linear. Indeed, the antiprotons leave the target at a wide range of energies, positions, and angles, so that the antiproton beam has there a high temperature, it is "hot". The stochastic cooling technique[1] enables to focus the antiproton beam on a particle by particle basis. It corrects each antiproton trajectory: a pick-up electrode records the deviation from the ideal orbit and communicates with a "kicker" that adjusts an electromagnetic field to push the antiproton towards the right direction. The cooled antiprotons are then injected into the accumulator at 8 GeV with small momentum spread.

Accumulator The accumulator is also a rounded triangular synchrotron, in the same tunnel as the debuncher. The antiprotons are then accumulated and cooled here until they are sent into the recycler via the main injector.

2.2.4 The Electron Cooling in the Recycler

Recycler The recycler ring and the main injector share the same tunnel. Its purpose is to store antiprotons, during many hours, which is crucial because the number of antiprotons available has always been a major limiting factor in producing high luminosities. The Recycler also maintains the antiprotons' momenta at 8.9 GeV/c. It can then transfer these antiprotons into the Tevatron for shots.

A major and challenging upgrade of the machine is the electron cooling within the recycler ring. It was achieved in 2005 and allowed a significant increase in the performances of the accelerator [2].The goal of electron cooling is to cool the 8.9 GeV/c antiprotons by mixing them in a 20 m long cooling section with a cold beam of 4.8 MeV/c electrons. The heat exchange cools the antiproton beam and thus reduces its emittance. The technique of electron cooling had been proposed for the first time around 1966. It took ten years to manage to make it work at low energy accelerators like the one in Novosibirsk. The Tevatron Recycler group needed ten years of hard work before achieving this first successful implementation of electron cooling at very high energies and in a large accelerator. For high intensity beams the antiproton beam cooling with cold electron beams is more advantageous than the already used stochastic cooling described above. This very

[1]The invention of this technique and its implementation at the $SP\bar{P}S$ accelerator at CERN where W and Z particles were eventually discovered owed a Nobel Price to Simon van der Meer in 1984.

important achievement from the beam division allowed Tevatron to reach instantaneous peak luminosities beyond $2.10^{32}\ cm^{-2}s^{-1}$. After the Spring 2006 shutdown, CDF has been recorded data at high rates averaging 30 pb^{-1} per week and still ever increasing. This is three times higher than the performances of the year 2004, before the 350 pb^{-1} sample was recorded and permits high hopes for the interest of the Tevatron Physics program in the next few years before the LHC runs at full speed.

2.2.5 Tevatron

The Tevatron is actually the name of the main circular ring at Fermilab, although we colloquially call the entire chain the "Tevatron". It is a circular synchrotron with a 1 km radius. It is composed of eight accelerating cavities, quadrupole and dipole focusing magnets. The Tevatron is also cryogenically cooled to 4K, and the accelerating cavities are made of superconducting niobium/titanium alloy. It is desirable to use superconducting magnets because the very large fields necessary to maintain TeV-scale energies would require currents so large that it is more cost effective to use superconducting magnets than ordinary resistive magnets.

The Tevatron accepts protons and antiprotons from the Main Injector or the Recycler (for antiprotons) at 150 GeV. They are then accelerated ("ramped") to 980 GeV, which takes 85 seconds. Since the antiprotons and the protons are oppositely charged, they circle in opposite directions in the magnetic field, and are housed in the same ring. The Tevatron can then sustain both beams for hours at a time (called a "store"). The number of collisions per second is described by the "luminosity" (which will be described below). When the antiproton beam is used up by interactions, scraping, and losses, the luminosity drops below the point of utility for the colliding physics experiments, and the store is ended.

2.2.6 Luminosity Measurements

The beam flux at colliders is measured in terms of luminosity. It has units of 1/Area. The expression for the luminosity is

$$L = \frac{fnN_pN_{\bar{p}}}{2\pi(\sigma_p^2+\sigma_{\bar{p}}^2)} \times F(\frac{\sigma_l}{\beta^*})$$

where f is the revolution frequency in Hertz, n is the number of bunches, $N_{p/\bar{p}}$ is the number of protons/antiprotons per bunch, and $\sigma_{p/\bar{p}}$ is the RMS beam size at the interaction point. This is multiplied by a form factor that depends on the bunch length σ_l and β^*, which is a measure of the transverse beam width.

Table 2.1 shows the accelerator parameters for Run I and Run II. Figure 2.2 shows the total luminosity collected by CDF from 2002 to 2006.

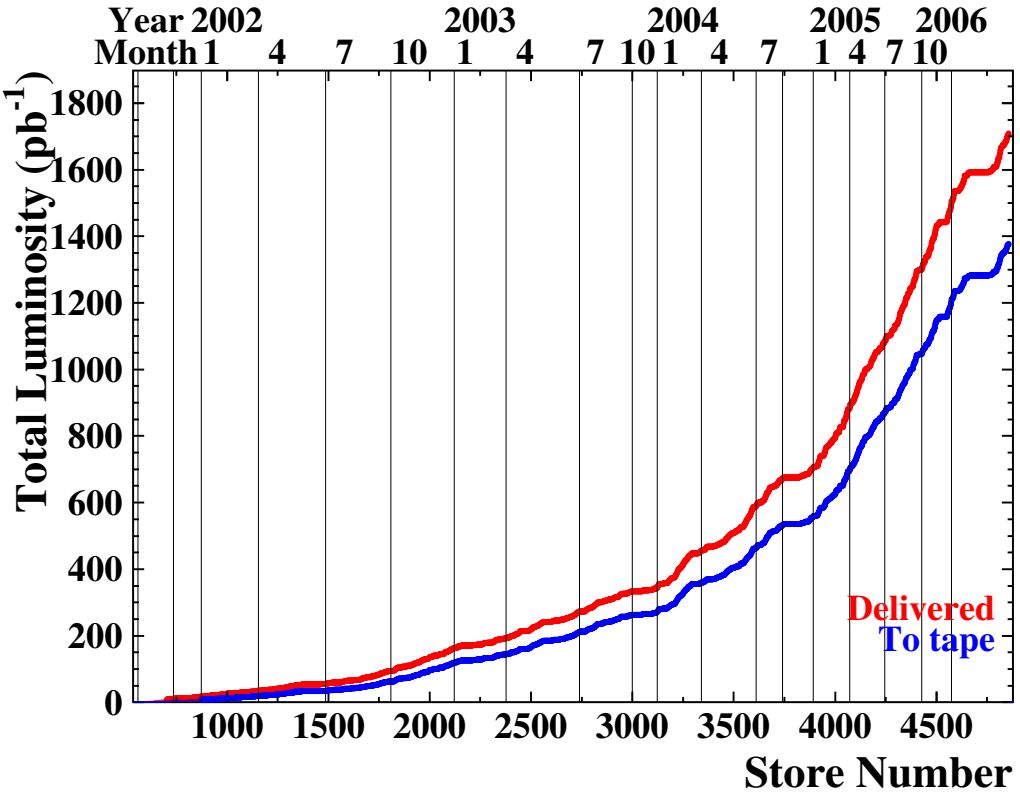


Figure 2.2: Recorded and delivered integrated luminosity

Bunches ($p \times \bar{p}$)	Run Ib (6x6)	Run IIa (36x36)
Protons/bunch (N_p)	2.3×10^{11}	2.7×10^{11}
Antiprotons/bunch ($N_{\bar{p}}$)	5.5×10^{10}	$3.0 - 7.0 \times 10^{10}$
Energy (GeV)	900	980
Bunch spacing (ns)	3500	396
Typical Luminosity ($\mathrm{cm}^{-2}\mathrm{s}^{-1}$)	1.6×10^{31}	$0.8 - 2.0 \times 10^{32}$

Table 2.1: Accelerator parameters in Run I and Run II

2.3 The CDF II detector

The data used for the measurement reported in this thesis were collected with the upgraded Collider Detector at Fermilab (CDF II) [5]. The detector upgrades were made to accommodate the higher luminosities and new beam conditions resulting from the upgrades to the Tevatron accelerator complex from Run I. It makes use of all the expertise acquired in the previous runs of this detector. Indeed CDF which started to run with the Tevatron in 1986, had already undergone through series of upgrades for Run 0, Run I (a and b) where the major upgrade was to include for the first time in such an environment a vertex detector. However the upgrade for Run II has been by far much more important and consisted in almost a full reconstruction of the CDF I detector.

CDF is an azimuthally and forward-backward symmetric apparatus designed to study $p\bar{p}$ collisions at the Tevatron. It is a general-purpose solenoidal detector that combines precision charged particle tracking with fast projective calorimetry and fine grained muon detection. A schematic view of the detector is shown in Figure 2.3. CDF has a cylindrical layout centered on the accelerator beam-line. Tracking detectors are installed in the region directly around the interaction point to measure charged particle trajectories inside a uniform magnetic field. The field is produced by a 5m long superconducting solenoid located at the outer radius of the tracking region. It is aligned along the proton beam direction and has a strength of 1.4 Tesla. Calorimeter modules are arranged in a projective tower geometry around the outside of the solenoid to provide energy measurements for both charged and neutral particles. The outermost part of the detector consists of drift chambers used to detect muons which typically pass through the calorimeter.

CDF uses a polar coordinate system to describe the geometry of the detector and the trajectories of the particles. The positive z-axis is defined along the direction of the protons. A particle trajectory is then describes by an azimuthal angle ϕ around the beam axis and by the polar angle θ between the proton beam. Due to the hadronic nature of the collision the longitudinal (z) component of an event energy is not fixed, contrary to e^+e^- colliders. Therefore, this is convenient to replace the polar angle θ by the pseudorapidity $\eta = ln(tan(\theta/2))$, since dη is invariant under boosts in the z-direction; more precisely, the rapidity y is the real invariant and the pseudorapidity η is a good approximation of y. Perpendicular to the beam, $|\eta|$ becomes 0 and extends to approximately 3.6 at the most forward part of the detector.

2.3.1 The Tracking System

The precise measurement of the tracks is mandatory from various points of view in this analysis. For instance it is a crucial ingredient for the tagging of the tau lepton and the identification and accurate measurement of the tau-jet in which it materializes. Furthermore, track isolation requirements will prove to be efficient in various aspects of this study. The Run II trigger includes a major novel feature namely the tracks that are reconstructed at the first level of triggering and combined with the calorimeter, the muon detector or the vertex detector; it is therefore very important to better identify and measure the electrons and muons and their transverse momentum already at this stage. The new vertex Silicon system is essential for the accurate determination of the vertex of the

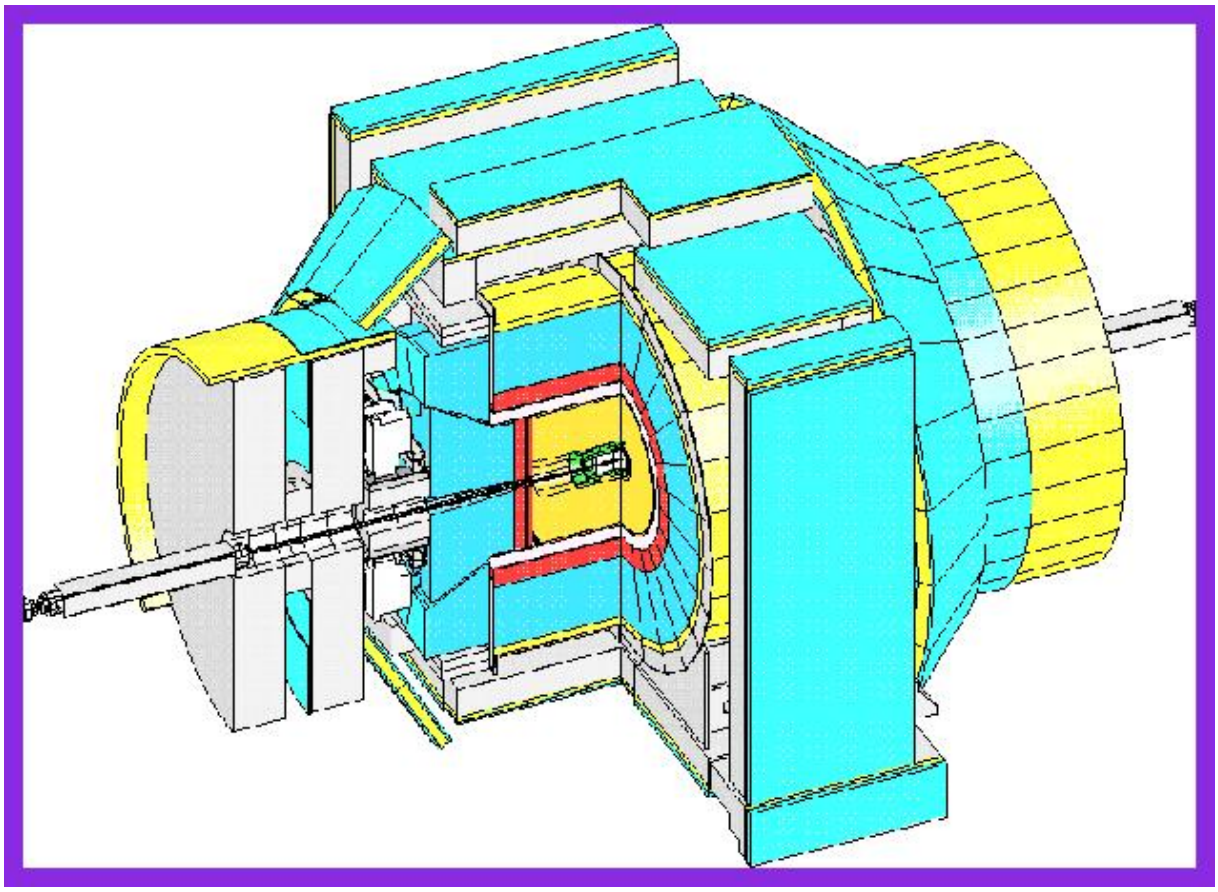


Figure 2.3: CDF detector

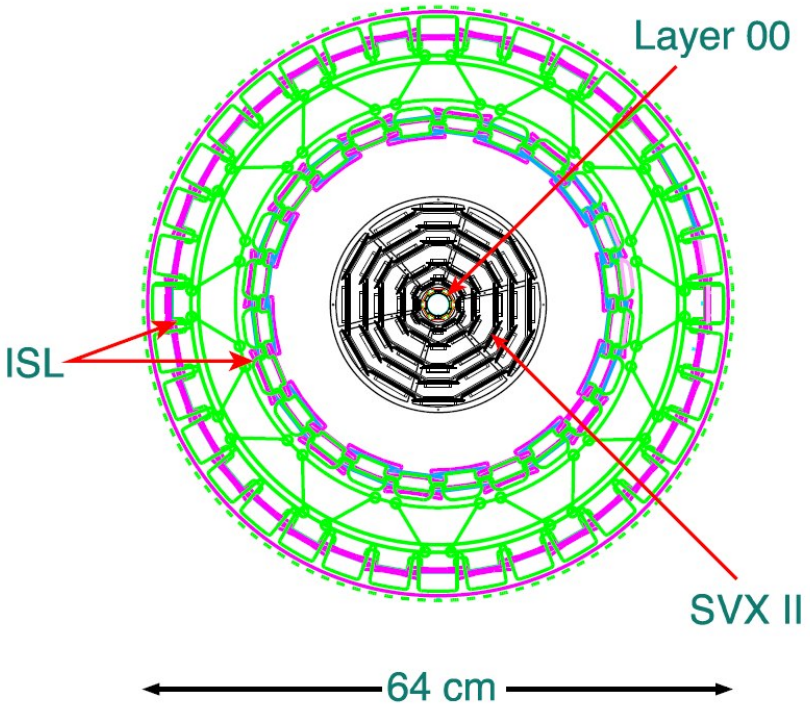


Figure 2.4: Frontal view of the silicon detector

event. It should be noted that as explained later in this report, the b-tagging facilities that mainly rely on the Silicon tracking system will not be used in this analysis and also not the possibility of standalone Silicon tracking to go beyond the central region defined by $|\eta| \leq 1$. In all this analysis the leptons will be central objects.

The tracking system consists of a silicon micro-strip detector and an open-cell wire drift chamber that surrounds the silicon detector. When a charged particle passes through the tracking system it ionizes the material/gas around it. A dedicated tracking algorithm searches for a continuous pattern of localized charge depositions (hits) and reconstructs the path of the charged particle, called the particle 'track'. The silicon micro-strip detector consists of three sub-detectors shown in Figure 2.4.

Starting from the beampipe the first Silicon layer, labelled as Layer00 is located at 1.4 cm in radius. This is a layer of radiation hard silicon wafers which is directly mounted around the beam pipe. It constitutes the first sub-system of Silicon layers. The second sub-system is the Silicon VerteX detector (SVX II), which consists of 5 layers of double sided silicon wafers extending from a radius of 2.4 to 10.7 cm from the beam. One side of the wafer has silicon strips aligned in the r-plane (axial), while the other side has strips that are either parallel to the beam (stereo) or at a 2 degree angle (small angle stereo). The Intermediate Silicon Layer provides 2 more layers of double sided silicon wafers and extends from radii 20 to 28 cm. The entire system allows track reconstruction in three dimensions with precision track information at the smallest possible radii. The impact parameter resolution of the silicon detector is on the order of 40 μm. The silicon detector can provide tracking information out to a pseudo-rapidity region of $|\eta| < 2.0$.

The new open-cell drift chamber, referred to as Central Outer Tracker (COT), is located directly outside of the silicon tracking detectors in the radial direction. The chamber consists of eight super-layers (SL) of 310 cm length cells and radii between 40 and 132 cm from the beam axis. Each super-layer contains 12 layers of sense wires strung between alternated layers of potential wires. The wires in four of the super-layers lie parallel to the beam axis to provide particle track reconstruction in the transverse plane (axial layers). The other 4 super-layers are present a 2 degree angle with respect to the beam axis to allow for track reconstruction in the z-direction (stereo layers). The two types of super-layers are alternated in the chamber starting with a stereo layer at the innermost radius. The COT is filled with a gas mixture of about 60% argon and 40% ethane. The mixture was chosen to ensure a fast drift velocity on the order of 100 μm/ns. Since the maximum drift distance in the chamber is 0.88 cm, it is compatible with the shorter interval between beam bunch crossings in Run II. The upgrade of this gaseous detector pushes at the limits the jet cell chamber technology, in order to be able to cope with the increasingly more and more difficult environment. It needs a constant monitoring of its functioning and the novel first level trigger system associated to this detector (eXtremely Fast Tracker, XFT) must undergo various and very innovative upgrades along with the luminosity increase.

The charge and momentum of a particle-track are determined by measuring the curvature of the track in the magnetic field. The solenoid produces a 1.4 T magnetic field inside the tracking volume that is uniform to 0.1 % in the acceptance region. The transverse momentum of a reconstructed track is determined from $p_T = Bqr$, where B is the strength of the magnetic field, q is the charge of the particle and r is the measured radius of curvature. The momentum resolution of the COT is given by $\sigma_{p_T}/p_T^2 = 1.7\ 10^{-3}[GeV/c]^{-1}$.

The present work relies on criteria of quality track insurance (see selection criteria of the tracks in the tau jet for instance), and many different capabilities of this highly performing tracking system without which such a study would not have been possible.

2.3.2 The Calorimeter System

The overall calorimeter system is another key element in this work. It allows to identify and accurately measure various objects that are essential ingredients in this analysis: Jets (recognition, triggering, energy reconstruction), electrons, charged and neutral pions that enter in the tau identification, total transverse energy and missing transverse energy of the event. Furthermore the calorimeter isolation criteria is widely used in this analysis. Although only electrons or tau leptons produced in the central barrel are considered in this analysis, the end-plug calorimetry is well used for instance for the Ht or missing transverse energy.

The central barrel CDF calorimeter is a sampling scintillator calorimeter and measures the energy of particles produced in the $p\bar{p}$ collision. The calorimeter is located behind the tracking volume and consists of layers of active (scintillator) material sandwiched between sheets of absorbing material such as lead and steel. As particles pass through the calorimeter, they interact with the layers of material and produce 'showers' of secondary particles. Light guides collect photons produced in the scintillator layers and direct them to the photomultiplier-tubes. Due to the specific nature of the interactions of particles

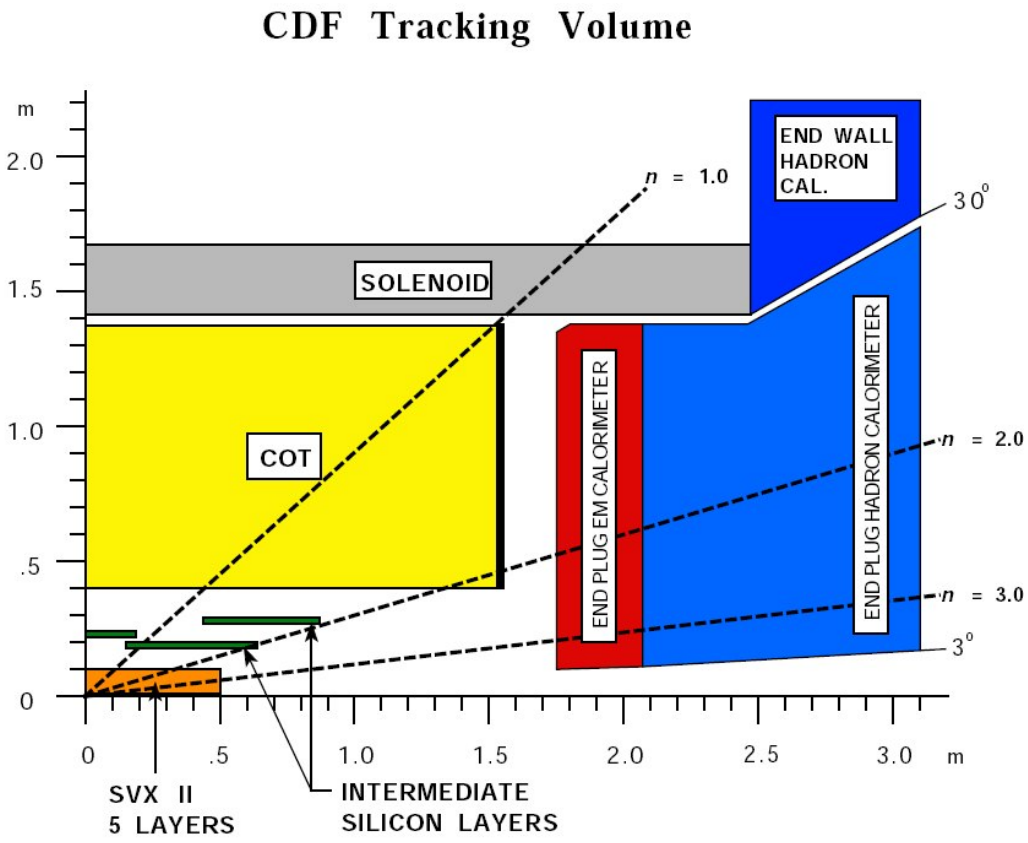


Figure 2.5: Schematic view of the CDF tracking volume

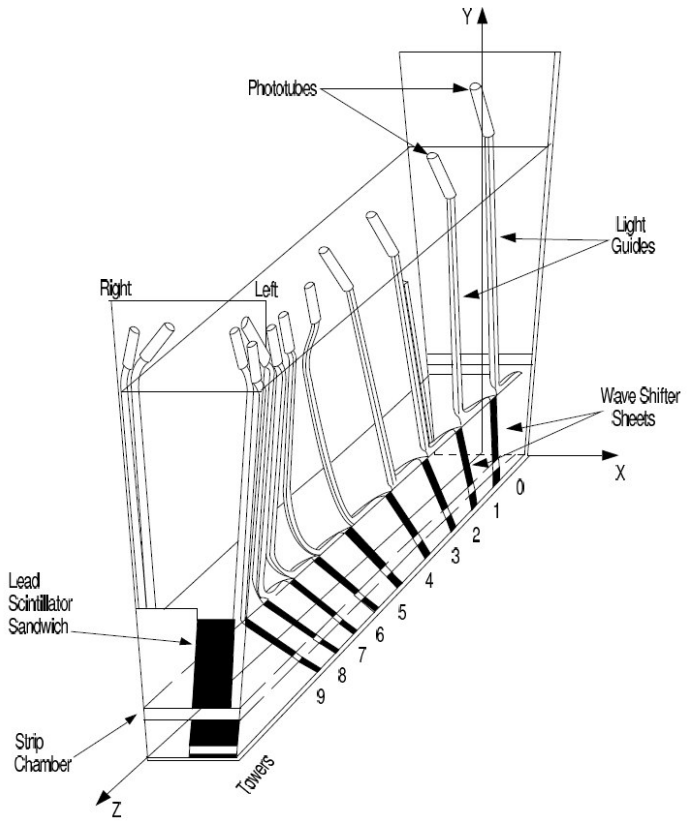


Figure 2.6: A wedge in the central calorimeter

with the material, electrons and photons shower over short distances, depositing the majority of their associated energy in the electromagnetic section of the calorimeter. Hadrons (strongly interacting particles), on the other hand, shower over longer distances in the calorimeter and deposit the most significant fraction of their associated energy in the hadronic section of the calorimeter. In the CDF detector, the electromagnetic sections of the calorimeter are immediately followed by the hadronic sections. The calorimeter is divided into a central calorimeter (and endwall part) covering $0 < |\eta| < 1.1$ (1.3), and a forward plug calorimeter covering $1.1 < |\eta| < 3.6$.

The central calorimeter as in Run I, consists into the same 48 azimuthal wedges of 15 inches. Each wedge is grouped into ten readout towers (often called calorimeter towers) with a projective geometry, as shown in Figure 2.6. The calorimeter contains an inner electromagnetic section (CEM) and an outer hadronic (CHA/WHA) section. The CEM contains 31 layers of 0.125 cm of lead interleaved with 5.0 mm of polystyrene scintillator. The CHA is made of 32 layers of 2.5 cm steel interleaved with 1.0 cm scintillator. The WHA was constructed with 15 layers of 5.0 cm of steel and 1.0 cm of scintillator.
The forward plug calorimeter has been completely rebuilt for Run II. The original gas calorimeter was replaced with scintillator plate calorimetry using scintillator tiles readout by embedded wave-length shifting (WLS) optical fibers. Both the new plug electromagnetic calorimeter (PEM) and the new plug hadronic calorimeter (PHA) use the same polystyrene based scintillator and photomultiplier tubes used in the CEM. The PEM

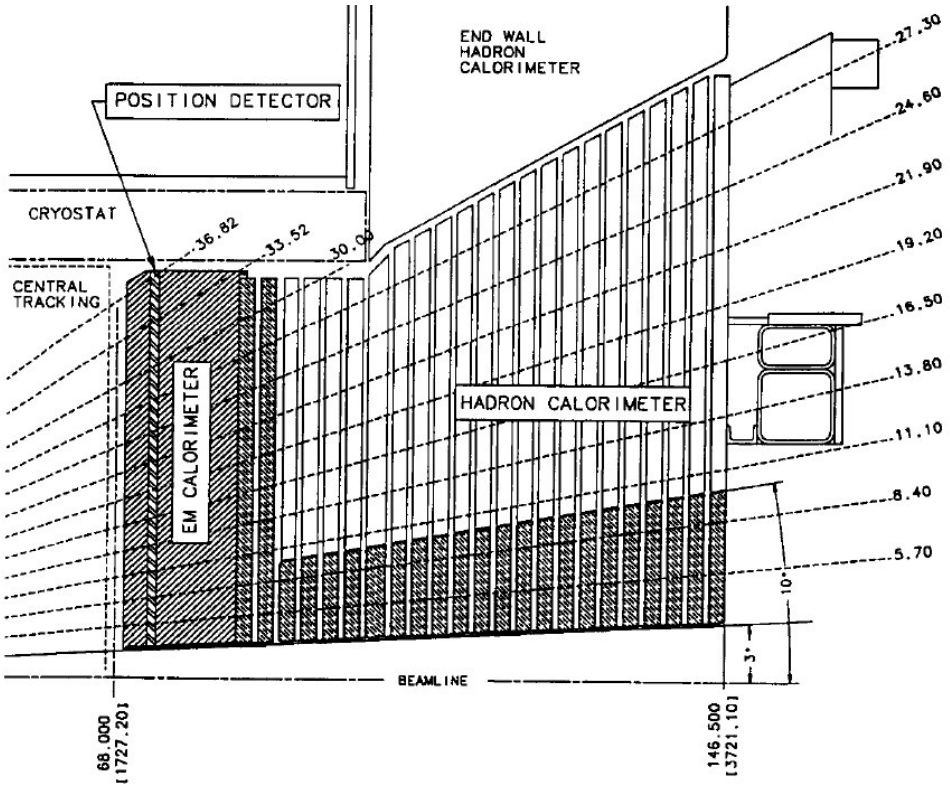


Figure 2.7: Transverse view of the plug calorimeter

contains 23 layers of 0.45 cm of lead interleaved with 4.0 mm of scintillator while the PHA is made of 23 layers of 2.5 cm steel interleaved with 1.0 cm of scintillator. The energy deposited in all layers of the calorimeter is summed together for each calorimeter tower separately. The central calorimeter provides about 480 readout towers each covering 0.1×15^o in $\eta \times \phi$ space. The plug calorimeter, shown in Figure 2.7, provides another 480 readout towers, each covering either 7.5^o or 15^o in ϕ and a variable range in η.

The calorimeter energy resolution was measured using test-beam data. The measured energy resolution for electrons in the electromagnetic calorimeters was found to be $14\%/\sqrt{E_T}$?? for the CEM and $14.4\%/\sqrt{E} + 0.7\%$?? for the PEM where the units of energy are GeV. The energy resolution for single pions in the hadronic calorimeter was found to be $50\%/\sqrt{E}$ for the CHA, $80\%/\sqrt{E}$ for the WHA, and $80\%/\sqrt{E} + 5\%$ for the PHA??.

Proportional chambers (CES in the central part and PES in the plug) are embedded in the region of the electromagnetic calorimeter where electrons produce the maximum shower intensity. More precisely, the CES that is much used in this thesis lies at depth $\sim$ 6 radiation length, corresponding to the shower maximum development of 10 GeV electrons. The CES modules have got anode wires running parallel to the z axis and orthogonal cathod strips. Particle hits in CES appear as wire and strip clusters. These shower maximum detectors are used to measure the profile of a shower and to extract the location of the incident particle within a given tower. The increased shower position resolution provides additional selection criteria for the electron identification.

Last but not least the Front end and readout electronics associated to these detectors have been completely rebuilt. This was of course in order to cope with the new bunch crossing features, namely the decrease by an order of magnitude of the bunch crossing time. Moreover profiting from this upgrade, the electronics went from being still only analog at the detector stage to become digital. This was a novelty that various calorimeters adopted in running experiments or at CMS experiment at the LHC at the end of the 90's.It was a real improvement in the overall detector performances.
A very common measure in the physics of hadronic collisions is the transverse component of momentum and energy which we define as $p_T = psin(\theta)$ and $E_T = Esin(\theta)$, respectively. In addition, the missing transverse energy, $\not\!\!E_T$ in an event is nominally defined as $\not\!\!E_T = -|\Sigma E_T^i n_i|$, where the n_i are unit vectors in the azimuthal plane pointing from the reconstructed event vertex to the i^{th} calorimeter tower.

2.3.3 The Muon System

Muons are another key object of this work that indeed divides into two dilepton channels, the so-called electron plus tau channel and the muon plus tau lepton channel. Both are equivalenty important and bring comparable amounts of statistics. Muons also enter in the evaluation of the total missing transverse energy. The muon detectors were covering a rather poor domain in pseudorapidity in previous CDF runs. For Run II they went under a large upgrade in order to extend their coverage and also in what concerns their associated electronics and the associated triggering system that also include as for the electrons the XFT trigger.

The muon detectors are located behind the calorimeter modules. Muons typically pass completely through the calorimeter modules leaving only a small fraction of their energy in the calorimeter (minimum ionizing). As a result, a muon must have a minimum p_T of 1.4-2.2 GeV/c (depending on the amount of material along its path) to reach the muon detectors. The CDF muon system consists of four independent systems of proportional wire chambers and scintillators: The Central MUon Detector (CMU) and the Central Muon UPgrade (CMP) both cover $0 < |\eta| < 0.6$; the Central Muon extension (CMX) which covers $0.6 < |\eta| < 1.0$; and the Barrel Muon Upgrade Detector (BMU) which covers $1.0 < |\eta| < 1.5$. The BMU is a new detector system commissioned in the initial part of Run II and is not used in this measurement. The CMU sub-detector consists of four layers of drift chambers as shown in Figure 2.8. The rectangular chambers have a 50 m sense wire running parallel to the z-axis through the center. The chambers are filled with a mixture of argon/ethane gas. As a muon candidate passes through the muon chambers, it ionizes the gas in the drift chambers. The hit position in the drift cells are determined from the drift time of the ions to the wire.

The CMP sub-detector consists of four layers of wire drift chambers of identical design to the CMU, but they are staggered by half cell per layer. The chambers are located behind 60 cm of steel and form a rectangular box around the detector. The CMP is mainly used in combination with the CMU detector to further improve the purity in muon identification and the combined system is referred to as the CMUP sub-detector. Located on top of the outermost layer of the CMP is the Central Scintillator Upgrade (CSP). The CSP system consists of a single layer of scintillator plates to provide additional timing

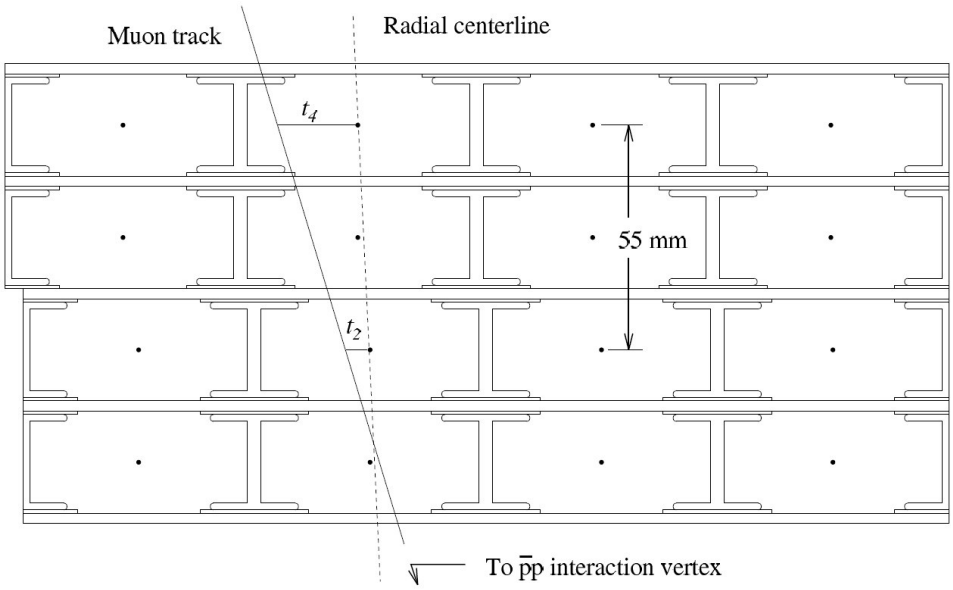


Figure 2.8: Schematic drawing of the CDF muon drift chambers (CMU)

information. The Central Muon Extension sub-detector (CMX) is a conical arrangement of drift chambers similar to those of the CMP. Different to the CMU/CMP, the system consists of eight layers of rectangular chambers which are grouped in pairs to form four continuous layers. A matching layer of scintillator plates (CSX) are mounted on both sides of the CMX system which is used in coincidence with the wire chambers to further improve the timing of the system and reduce the fake rate. Fig.2.9 present a schematic view of the phi wedges disposition in the CMX, that will be referred to in the section 3.3.2 about the muon identification. Each of the three regions, referred as CMX arches, keystone, and miniskirts were fully operational at different times in the Run 2 data taking. The CMX sub-detector angular acceptance was increased by 45% compared with the Run I when only the CMX arches were at disposal.

2.3.4 The Trigger and Data Acquisition Systems

A major upgrade was performed on the triggering system in CDF which was indeed totally rebuilt. Among the crucial new pieces, there are two novelties that represent indeed technological prowess. Both are related to the tracking system, namely the XFT already mentioned that is built from the information provided by the COT and the SVT (Silicon Vertex Trigger) which uses the information of both the COT and the vertex detector. The XFT is associated at the electron calorimeter information and refines already at Level-1 the electron selection. Likewise the XFT is associated to the muon detector information and ensures the quality of the muon triggering at Level 1. This study makes use of various of those triggers as for instance the so-called jet triggers that select events with at least a reconstructed jet with an energy above a certain threshold and the electron and muon high p_T triggers which constitutes the basis of the data used for extracting the signals we are looking for.

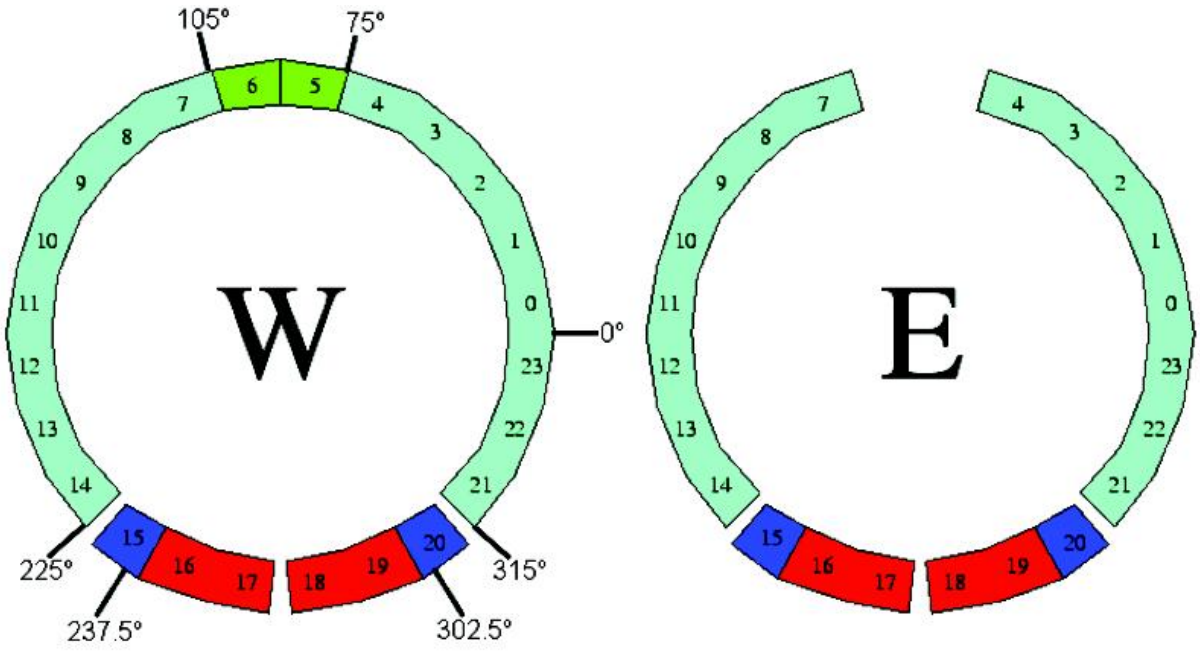


Figure 2.9: Schematic drawing of the East ($\eta > 0$) and West parts of the CMX muon system. The numbers refer to the phi wedge numbering. Wedges 5-6 are called "Keystone" regions, blue and red wedges ($225^o < \Phi < 315^o$) are called the "miniskirts", and light green wedges are the CMX arches. Note that most of wedges 15 and 20 actualy do not exist. The Φ angles at 237.5^o and 302.5^o reflect their actual edges.

Currently, collisions in the Tevatron occur at a rate of 2.5 MHz and the average event size, representing information from all detector channels, is about 250 KBytes/event. If each event were to be read out, we would need to write 630 GBytes/s to disk - clearly a very challenging task. Luckily not each event contains physics of interest and CDF utilizes a 'trigger system' to select events for future analysis. The upgraded trigger system accepts events at a rate of 80 Hz and writes them to disk. This corresponds to an event rejection factor of roughly 30000 and is obtained by using a three-level system where each level is designed to provide sufficient rejection to allow for a processing with a minimal dead-time at each subsequent level. The first level of the trigger (Level-1) has to be the fastest and utilizes custom hardware to select events based on preliminary information from the calorimeters, tracking chamber and muon detectors. All detector data are fed into a 6 μs pipeline to provide time for processing required at Level-1. The global Level-1 decision must be made before the corresponding collision data reach the end of the pipeline. The Level-1 trigger reduces the data rate from 2.5 MHz to less than 20 kHz. The Level-2 system is a set of programmable logic devices which has access to more refined information including the silicon tracking system. The decision time is 20 s and dedicated hardware is used to reconstruct clusters in the calorimeter for electron and jet reconstruction as well as simple track reconstruction. Level-2 provides a factor of 100 reduction over Level-1 passing a 300 Hz data rate to Level-3. Events which pass Level-2 are analyzed by a farm of approximately 300 computers, each fully analyzing and reconstructing events. Events which pass this last level of the trigger are delivered to the data-logger system which transfers the data to the storage as well as to the monitoring system. Monitoring ensures that the entire detector and trigger system were working properly during data taking. A maximum rate of 20 MBytes/s can be written to mass storage which corresponds to an event rate of about 80 Hz that can pass Level-3. Recent developments aim for an event-size reduction to increase the Level-3 accept rate to about 100 Hz.

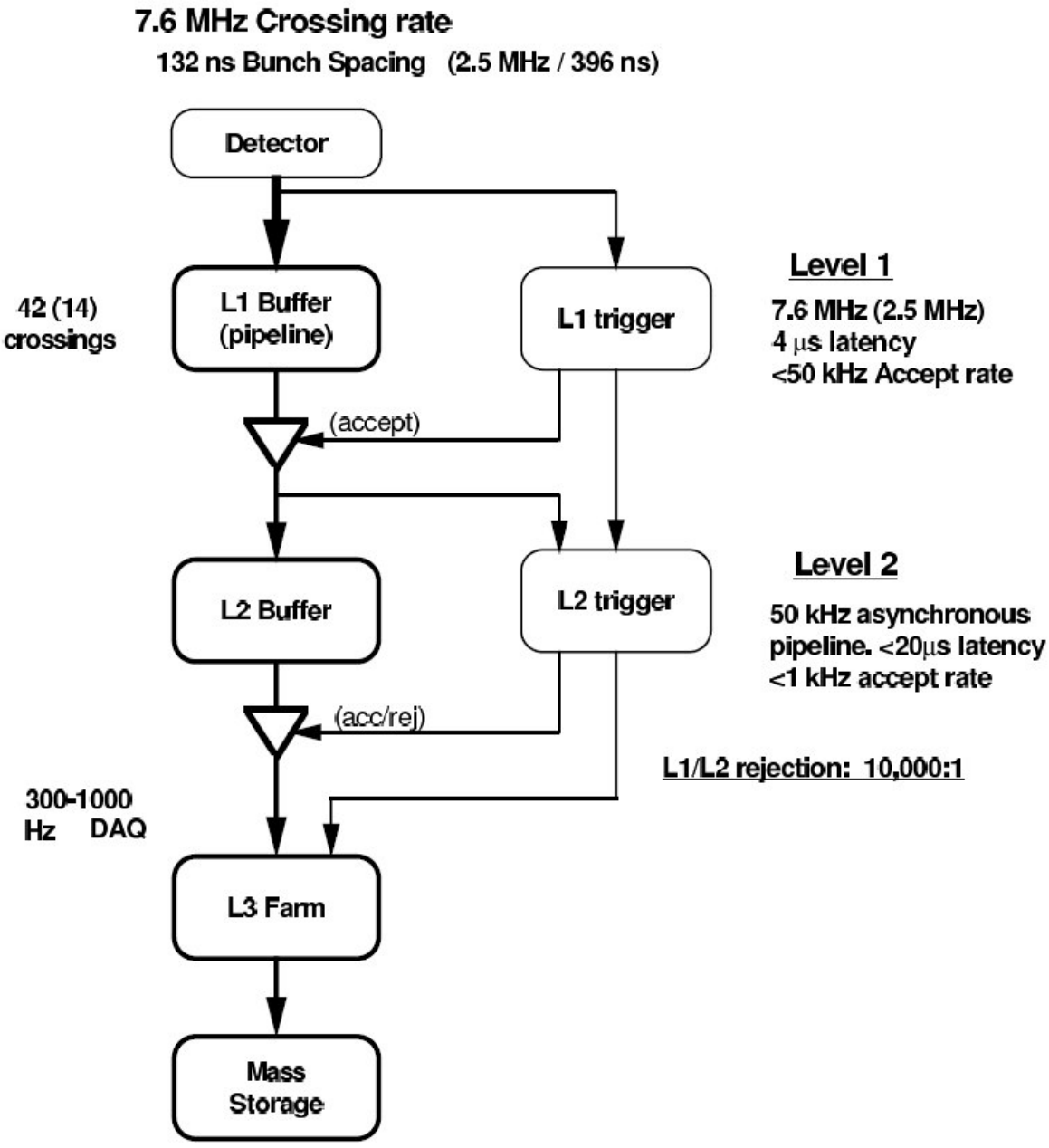


Figure 2.10: Schematic view of the 3-level CDF trigger system

Chapter 3

Data selection and analysis flow

3.1 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 chapter, 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 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.

3.2 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 [12], 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 [10] 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 [13]).
- The leading-order matrix elements generator Alpgen [11], interfaced with Pythia or Herwig [9] in order to handle the parton shower and hadronization processes, 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]

[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.

3.3 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.

In the case of the here so-called first analysis, we are left with a sample of about 344 pb^{-1} in the muon case and of 359 pb^{-1} for the electron case.

3.3.1 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 (see Chapter 2).

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 3.1.

The *Lshr* variable is described in details in [14]. 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 observed

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 3.1: 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.

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 [15]. The CES (central electromagnetic calorimeter "shower max" strips and wires, see chapter 2) 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.

3.3.2 The muon identification

The high pt muon identification follows the one usually applied in CDF [31]. It is used to search for the unique muon present in the $t\bar{t} \rightarrow \mu\tau\nu\nu b\bar{b}$ 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 seach 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 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 (see the description of CDF muon chambers in Chapter 2). The set of sequential cuts that make up the central muon identification are summarized in Table 3.2.

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.3.1). 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 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

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 3.2: The selection cuts applied to muons, where p is the track momentum ; SL stands for "COT super-layer" (see section 2); 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.

that correspond to the first 350 pb^{-1} of data, the CMX muons were thus restricted to the arches (Fig. 2.9). Finally, the CMX wedge 14 on West side (Fig. 2.9) 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.

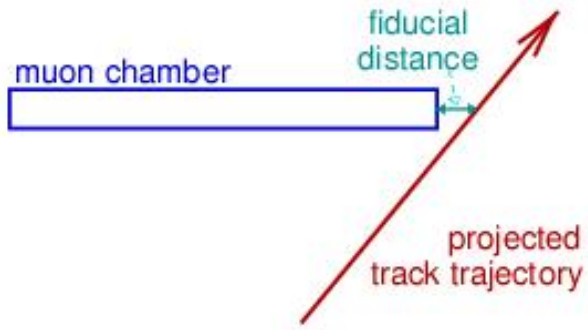


Figure 3.1: The so-called fiducial distance between the muon track extrapolated to the muon chamber radius and the edge of the muon chamber is required to be <0.

3.3.3 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.

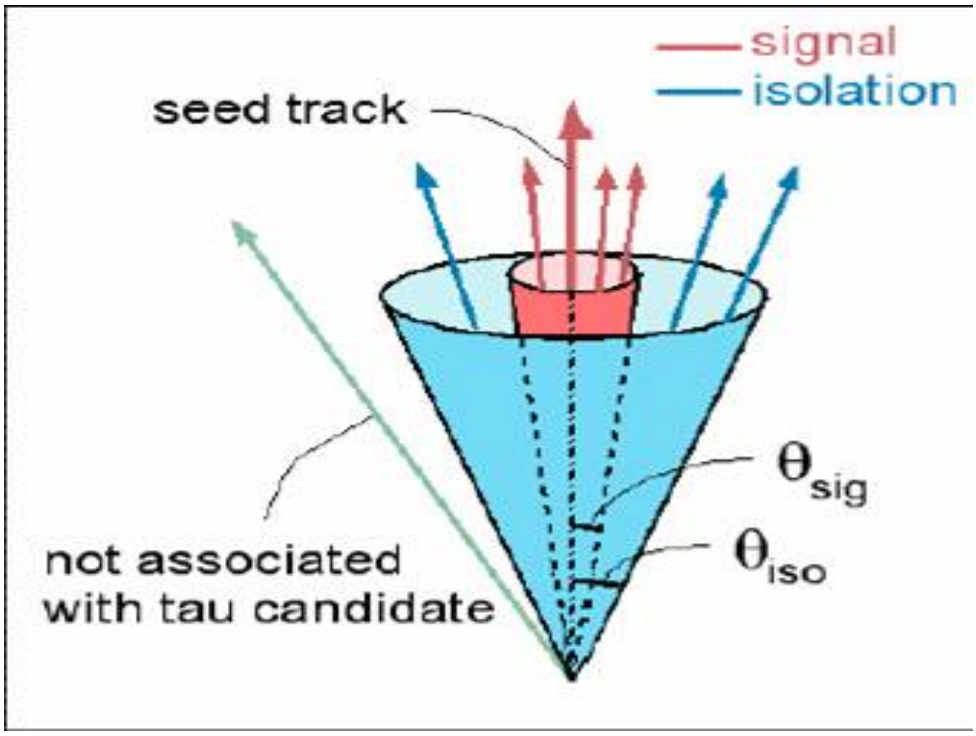


Figure 3.2: The so-called tau cone (in red) and tau isolation cone (in blue) containing tau tracks and isolation tracks. The tau isolation cone opening angle θ_{iso} is fixed at 40^o. The tau cone starts with an opening of 10^o and decreases with the tau energy as described in Fig. 3.5.

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. 3.2), 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].

The tau selection takes also advantage of the shower maximum detector (CES) to attempt to reconstruct the photons inside jets [16], that are often produced inside the

[2]The tau mass is $m_\tau = 1776.99^{+0.29}_{-0.26} MeV/c^2$ [37] and the tau identification requires the reconstructed tau transverse energy to be greater than 15 GeV.

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 [16]. Inside the tau identification algorithm, CES clusters unmatched with COT tracks are called π^0's (see Fig.3.3).

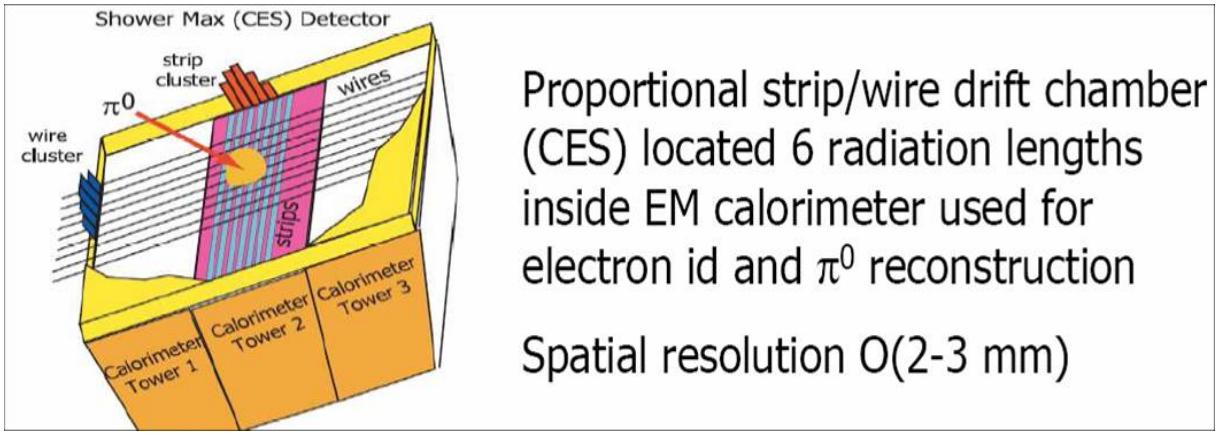


Figure 3.3: Neutral pions or photon reconstruction with the CES subdetector, used for the tau identification.

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.3.4).

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:
  - 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.3.5), must be equal to 1 or 3.
  - The absolute value of the tau lepton electrical charge $|Q|$ must be equal to 1.

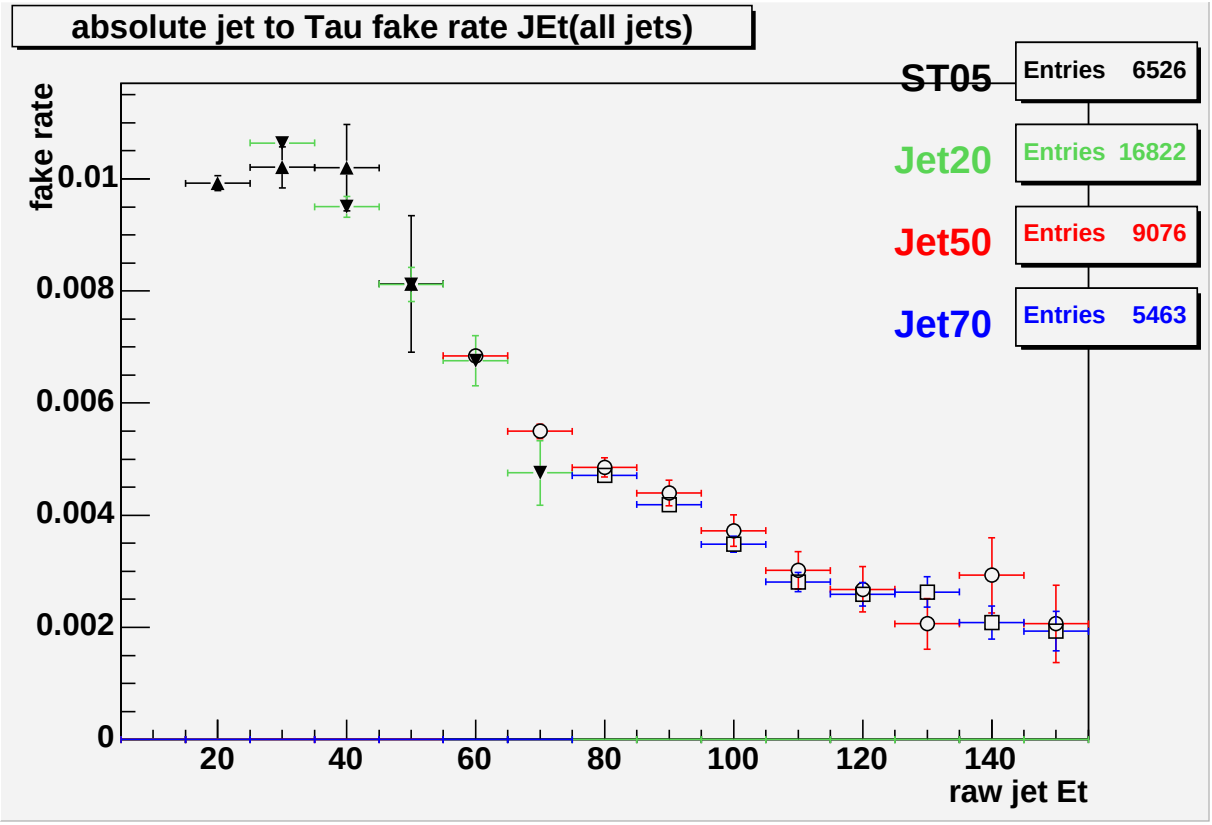


Figure 3.4: The probability for a central jet to be misidentified as a hadronic tau, depending on the jet transverse energy. This fake rate was calculated by us using the four QCD triggers Single Tower 5 GeV, Jet 20 GeV, Jet 50 GeV and Jet 70 GeV, restricting oneselves to the triggers unbiased energy regions. These triggers are described in the chapter devoted to the jet to tau fake rate.

- 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.
- 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).

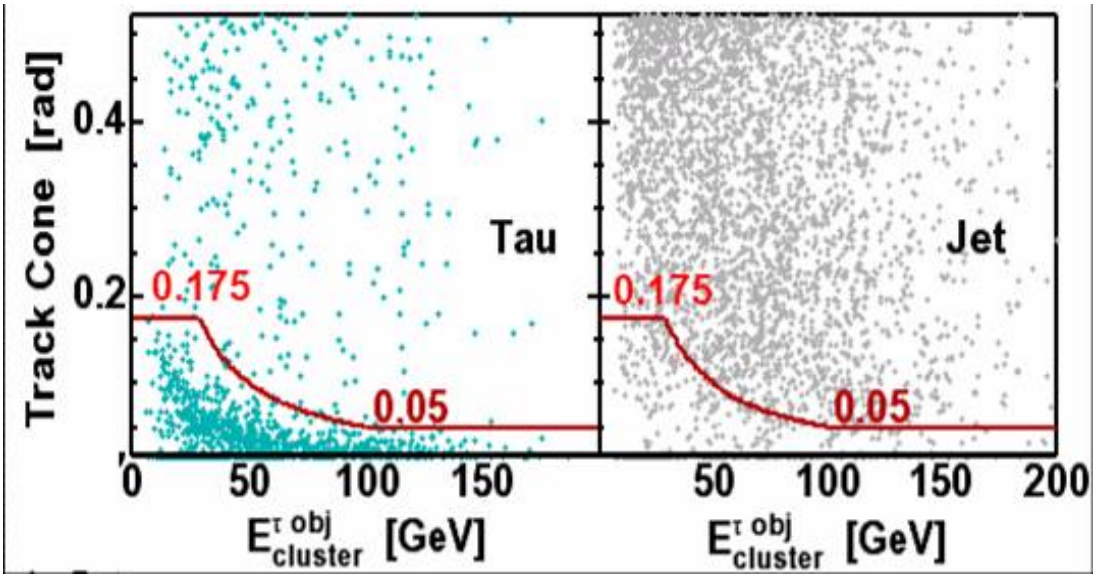


Figure 3.5: The tau cone opening is shrinking with the tau-jet energy, from 0.175 rad down to 0.05. This varying cone shape provides a good acceptance for "true" taus (left plot) and a good rejection against jets faking taus (right plot).

- The 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.

3.4 The energy corrections and validation of the fundamental objects

In this section, 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.

3.4.1 Electrons

3.4.1.1 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 [24]. 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 [24]. 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

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

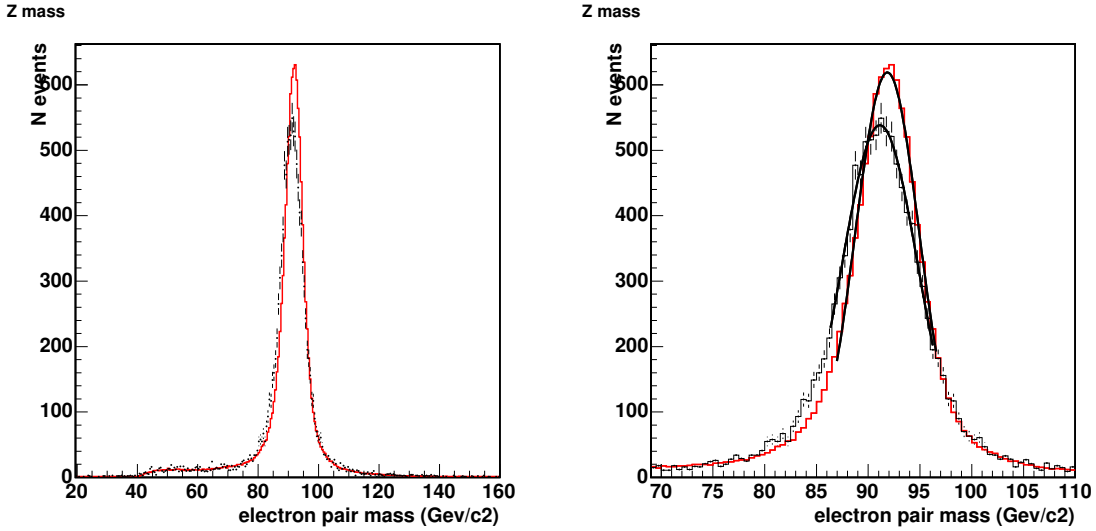
Table 3.3: CEM N-1 efficiencies in the data (sewk7d) and in $Z \rightarrow ee$ Monte Carlo (ztop2i) for the 350 pb^{-1} analysis.

An efficiency of $\epsilon_{CEM}^{data} = 0.812$ is found for data and $\epsilon_{CEM}^{MC} = 0.815$ for the Monte Carlo sample [24]. 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 [24].

We also checked the so-called $N-1$ efficiencies. Our results are gathered in Table 3.3. 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 [24]. 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.

3.4.1.2 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 [37]. 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.3.6(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.3.6(b) and 3.7). 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 3.4 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}).



(a) *Full energy range*

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

Figure 3.6: 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.

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

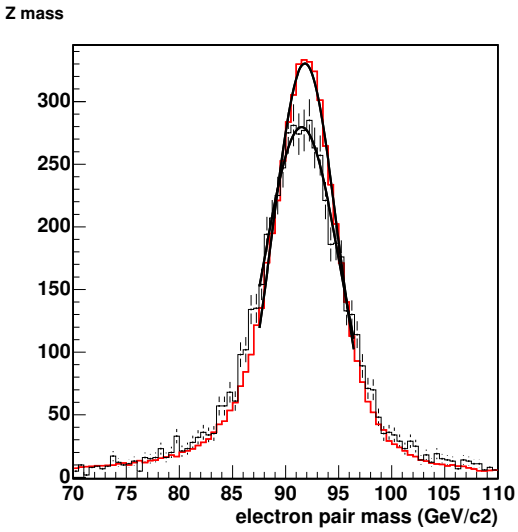


Figure 3.7: Z lineshapes for the 2002-2004 electron data and corresponding Pythia simulation. The black dotted histogram is for real data and the red histogram is the result of a Pythia simulation.

3.4.1.3 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 $\sqrt{s} = 1.8\ TeV$ [39] [40]($\sigma_{in} = 59.3 \pm 2.3mb$ [38]) 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:

$$\sigma(p\bar{p} \to Z/\gamma^*) \times BR(Z/\gamma^* \to ll) = 254.9 \pm 3.3(stat.) \pm 4.6(syst.) \pm 15.2(lum.)pb \text{ [43]},$$

in good agreement with NNLO theoretical calculations (see Fig. 3.8), which have uncertainties of 2% ($\sigma_{Z/\gamma^*} \times BR(p\bar{p} \to Z \to ll) = 251.3 \pm 5.0$ pb, following [44]).

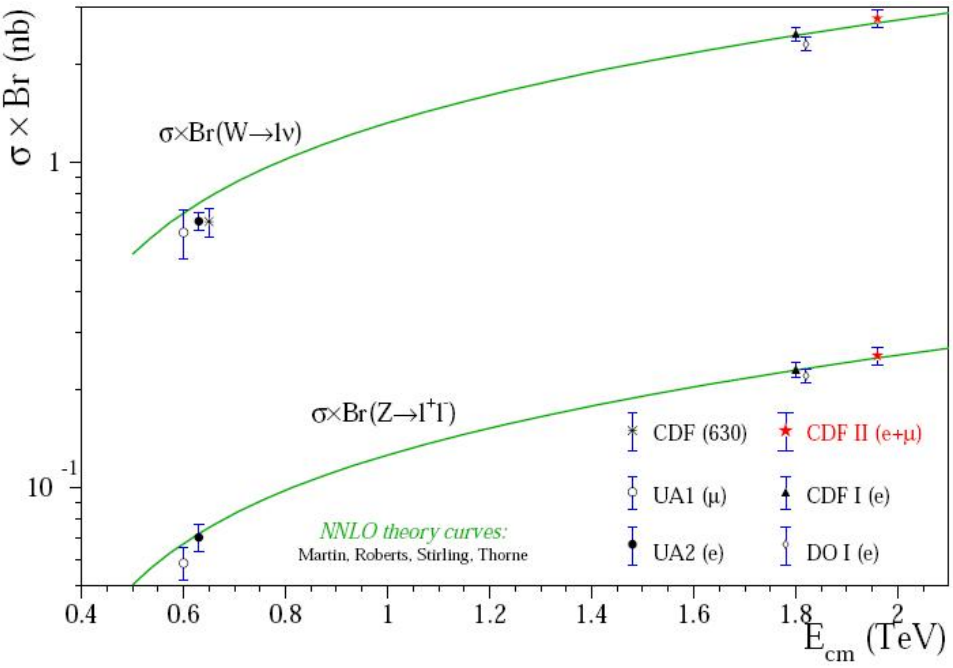


Figure 3.8: $W \to l\nu$ and $Z \to ll$ cross section measurements as a function of the $p\bar{p}$ center-of-mass energy, E_{cm}. The solid lines correspond to the theoretical NNLO Standard Model calculations from [44]

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 [41] [42], according to which the inelastic cross section increases with energy as $ln^2 s$.

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 [46] and [45] 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.
- $\frac{\epsilon_{cem}^{data}}{\epsilon_{cem}^{mc}}$ 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 [47] are used, that is 0.986 ± 0.004 for the 2002-2004 period and 0.977 ± 0.004 for the last period of runs.

[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.

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 [49]: 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}

3.4.2 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 3.9 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 (3.10) 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.

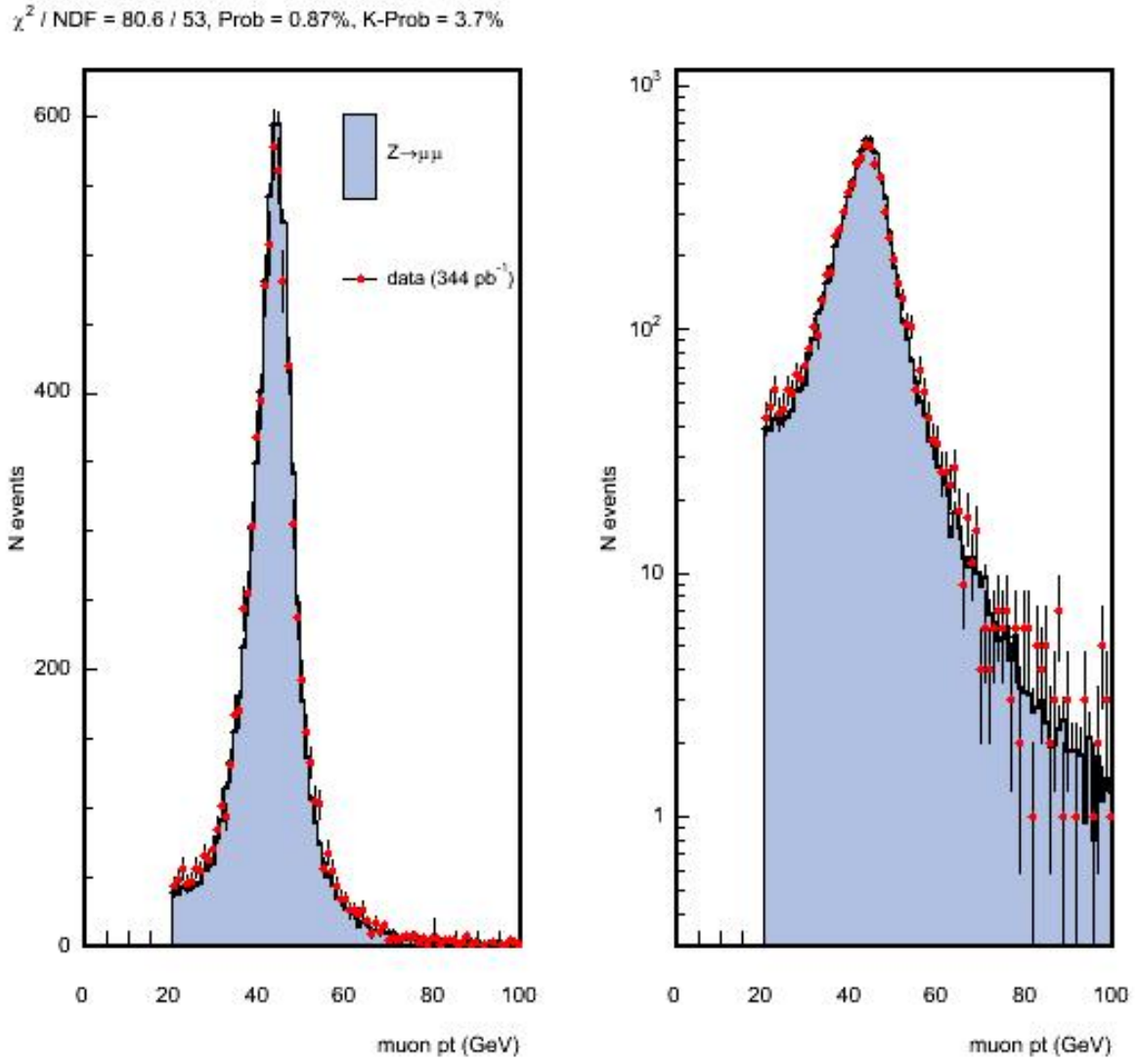


Figure 3.9: The muon p_T distribution in $Z \to \mu\mu$ events. Left: linear scale ; Right: log scale

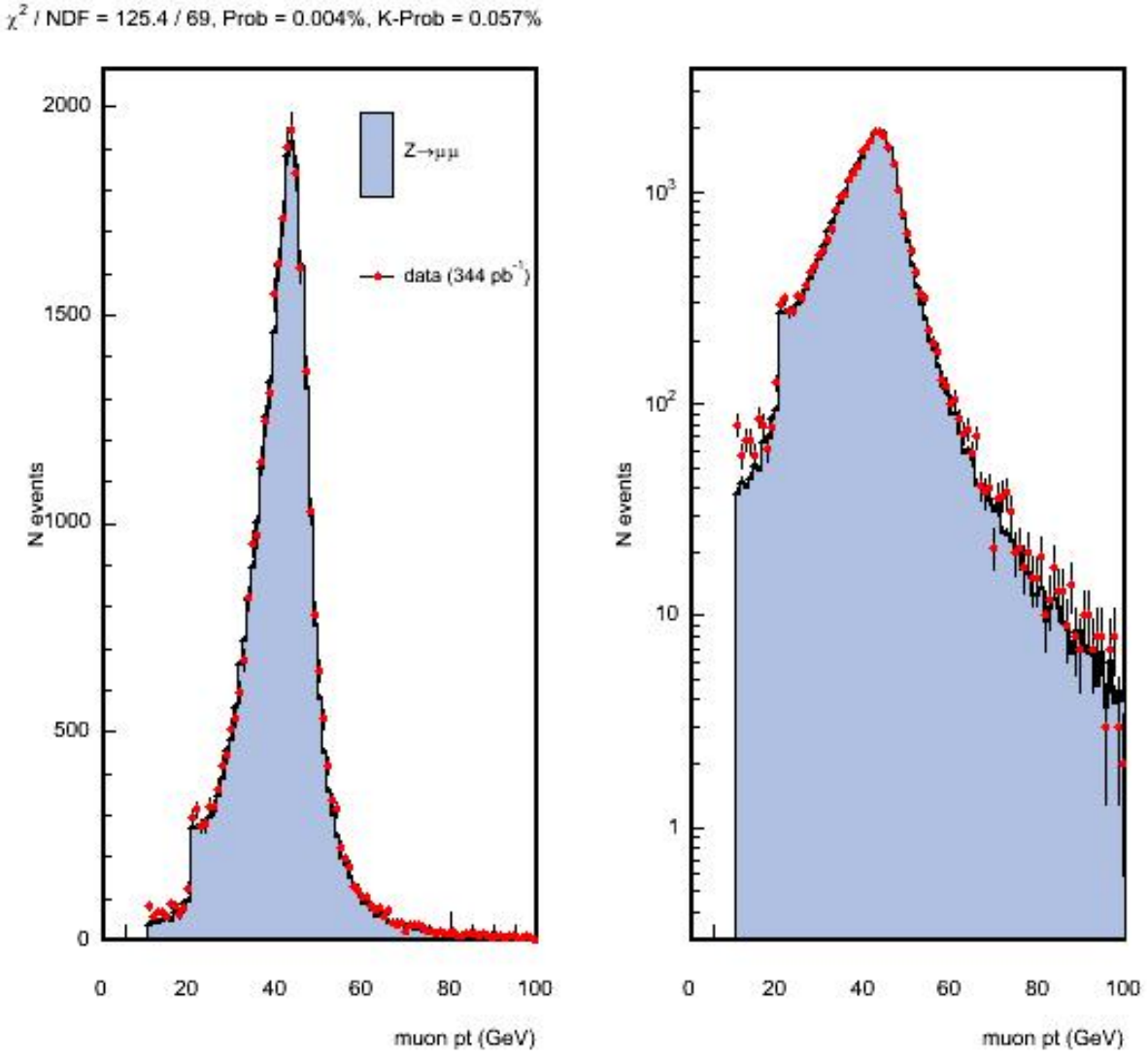


Figure 3.10: The muon p_T distribution in $Z \rightarrow CMUP/X + track$ events. Left: linear scale ; Right: log scale

3.4.3 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 [51] [52]. It uses a specific trigger that has been developed at CDF, the so-called lepton+track trigger [50], 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 section 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.3.11 extracted from [52]): $\mathrm{W}_{p_T} > 24$ GeV/c or $\mathrm{W}_{M_T} > (50 - 1.25 \times \mathrm{W}_{p_T})$, 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.
- $\mathrm{W}_{M_T} < 50$ 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.3.12 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 leptons is the output of our jet to tau fake rate decribed in details in the section 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

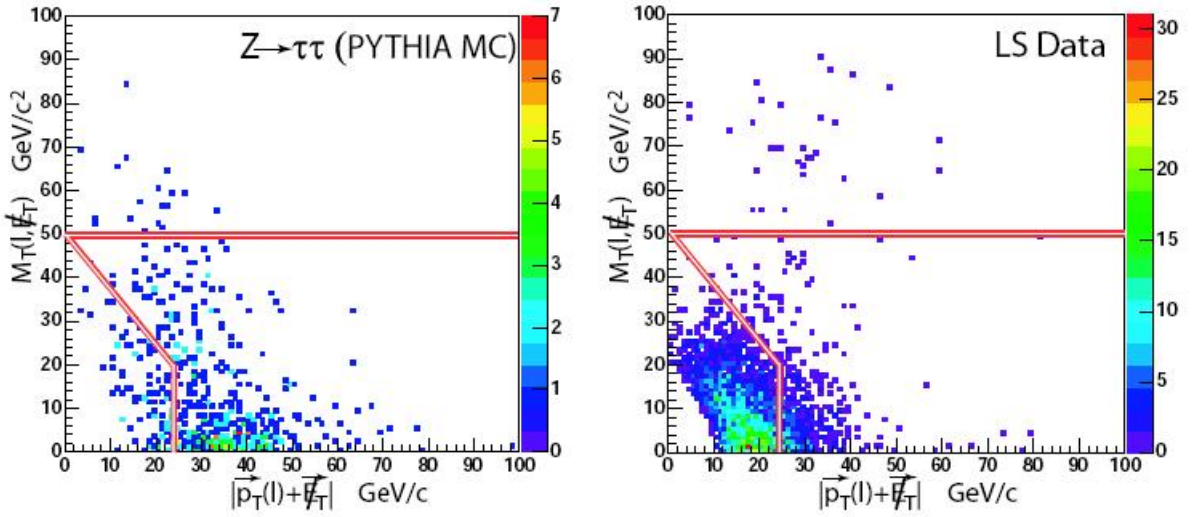


Figure 3.11: Distribution of M_T versus p_T for: left)$Z \rightarrow \tau\tau$ signal events from Pythia ; right) Real data events passing all cuts except that the two leptons have the same electric charge (dominated by QCD, $\gamma + jets$ and $W + jets$ backgrounds). [52]

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.

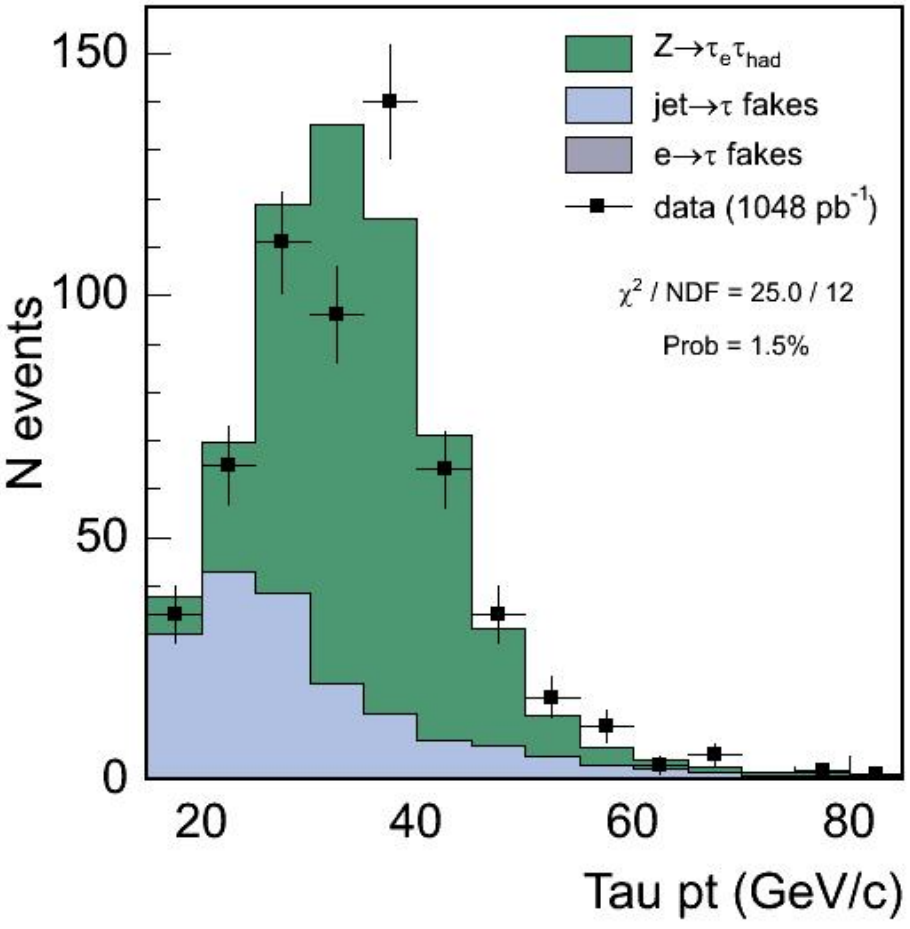


Figure 3.12: The tau p_T distribution in $Z \rightarrow \tau\tau$ candidate events

3.4.4 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 3.13 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.

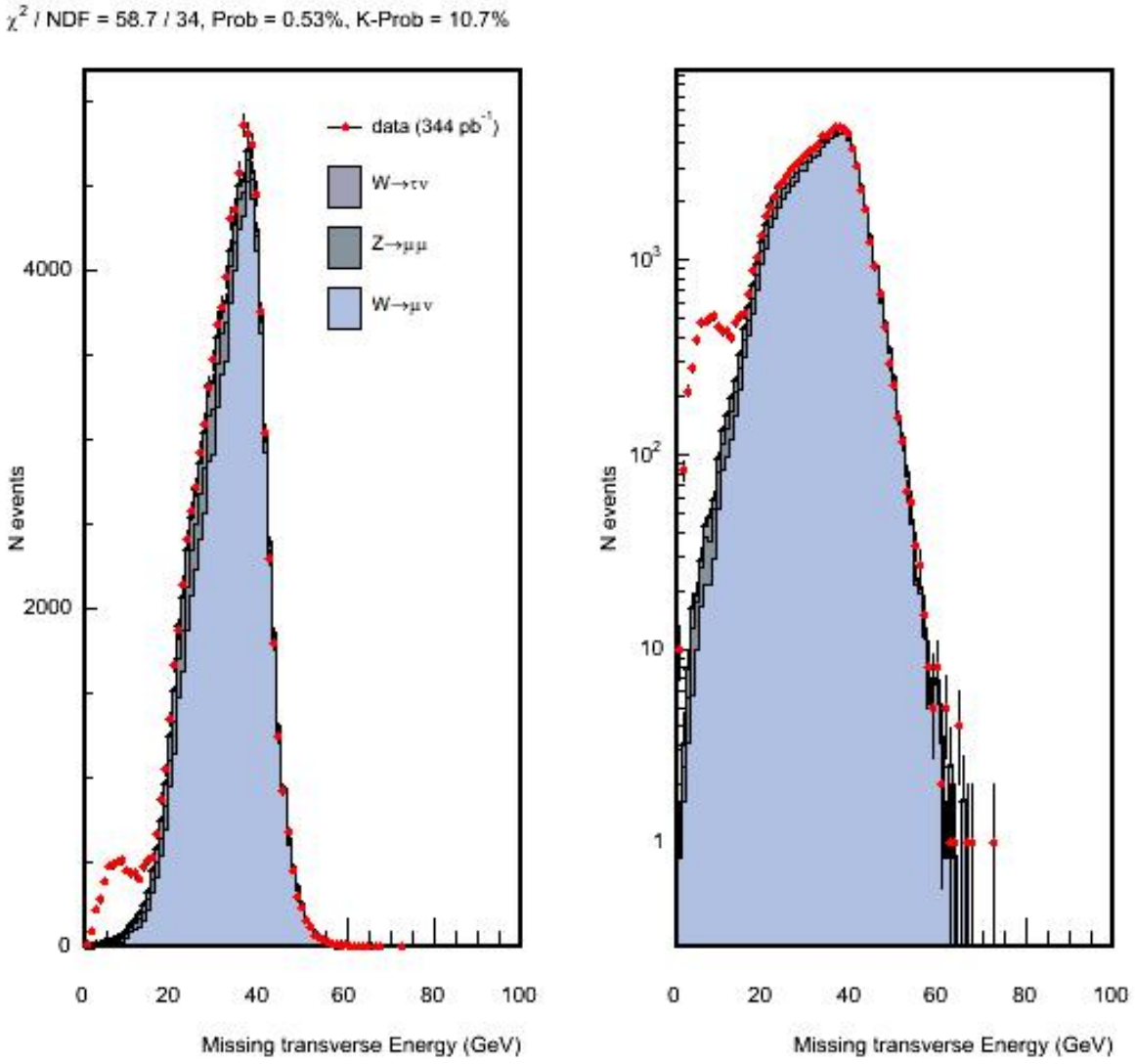


Figure 3.13: The MEt distribution from $W \rightarrow \mu\nu$ events. Left: linear scale ; Right: log scale

3.4.5 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] [35]. As explained in Chapter 2, 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.3.14.

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 Chapter 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 [36] is summarized in the Fig.3.16. 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.

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

[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 [34] consists in merging pairs of towers following an increasing order in transverse momentum.

is particularly useful for high energy runs when the average number of interactions per event becomes higher (see Fig. 3.15).

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.3.16).

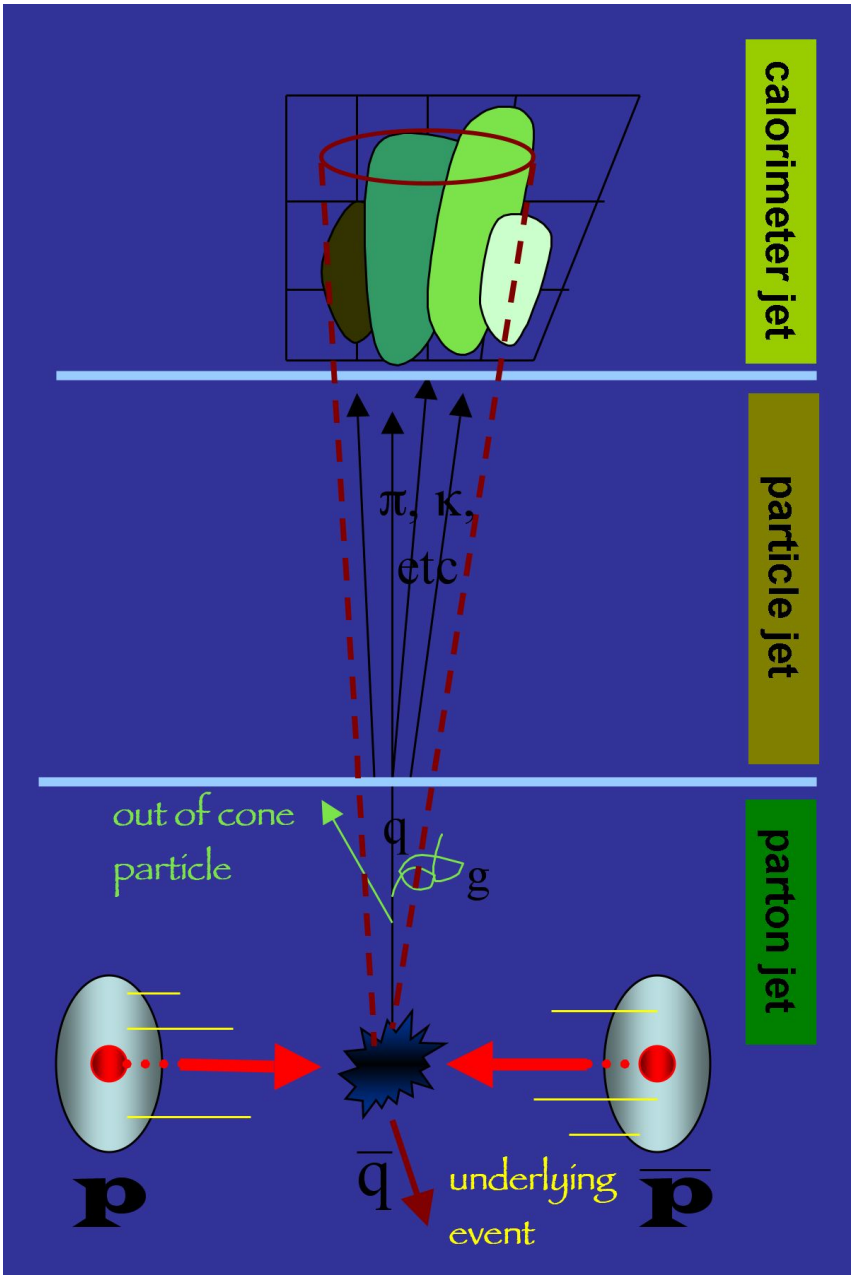
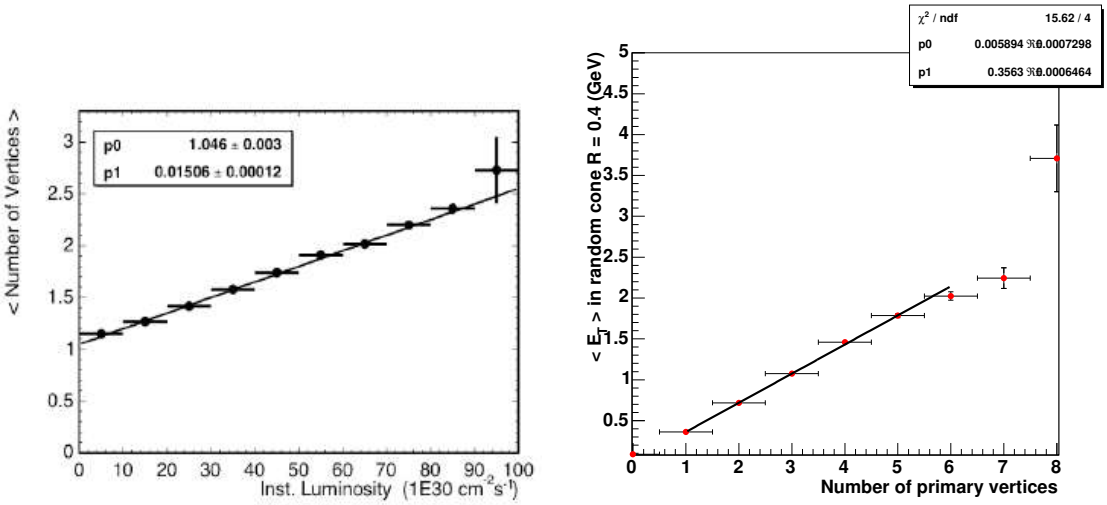


Figure 3.14: The schematical path from a parton jet to a calorimeter jet



(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 3.15: *Jet energy correction due to multiple interactions*

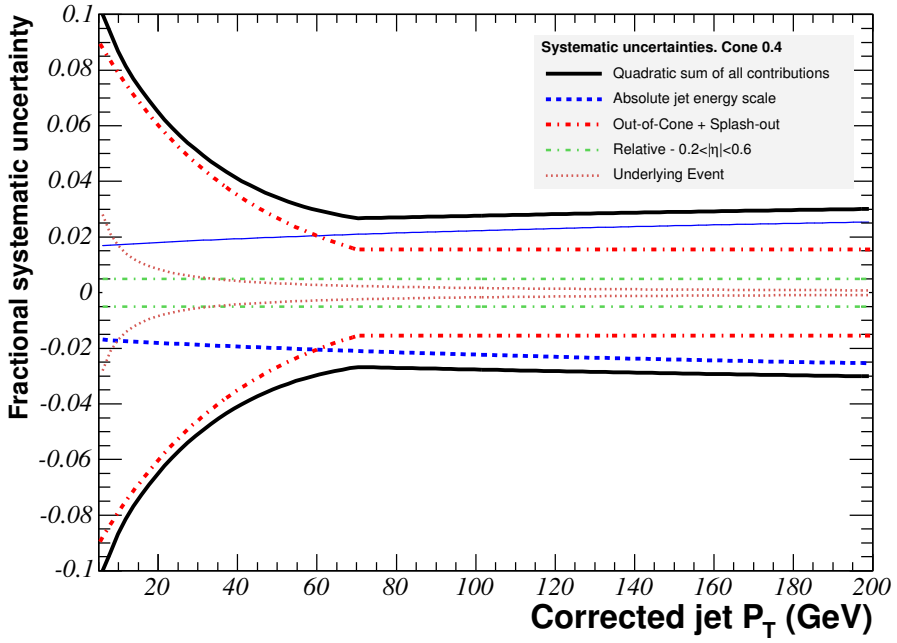


Figure 3.16: Systematic uncertainties on the jet energy correction

3.5 The event selection strategy and the acceptance applied for the 350 pb^{-1} analysis

3.5.1 The event selection

The event selection used in the analysis of the first 350 pb^{-1} data is rather similar to the one applied for the analysis of the first data sample in Run II with 195 pb^{-1} data ([20] and [19]). It mainly differs on the Z veto. This selection is based on:

1. One central isolated electron or muon with a transverse momentum greater than 20 GeV/c.
2. One central isolated tau with a transverse energy greater than 15 GeV, opposite in charge to the first lepton.
3. At least 2 jets with pseudorapidities between -2 and 2, and with transverse energies greater than 25 and 15 GeV respectively.
4. A total missing transverse energy greater than 20 GeV.
5. An activity in the event defined by: Ht $>$ 205 GeV (defined below)
6. A veto against the $Z \rightarrow \tau\tau$ + jets events.

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

An efficient cut must be applied to the $Z \rightarrow \tau\tau$ 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.

3.5.2.1 Angular configuration of $Z \rightarrow \tau\tau$+ 2 jets events

The requirements of at least two extra jets, a high missing trasverse energy ($\not\!\!E_T > 20$ GeV) and a high scalar sum of transverse energy (Ht) greater than 205 GeV, 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 $Z \rightarrow \tau\tau$ 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
  2. therefore giving a high true missing E_T between the two taus (Fig 3.18)
- 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.3.17) or in between (fig.3.18)
- 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.3.18), or opposite to them(fig.3.17).

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 one is the top signal sample simulated with Pythia. Figures 3.19 and 3.20 are based on these

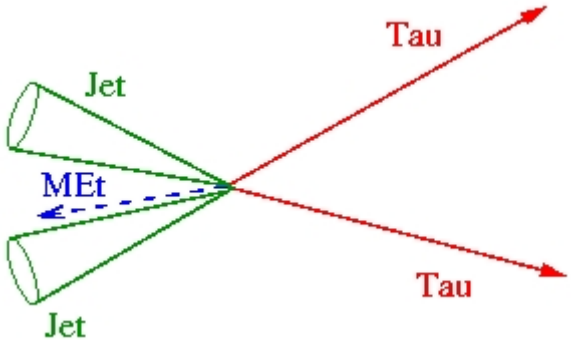


Figure 3.17: Configuration with missing E_T opposite to the two tau leptons

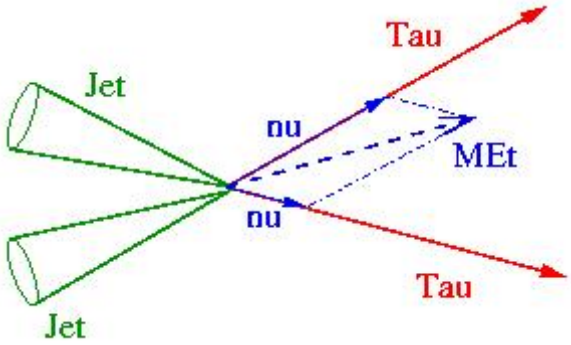


Figure 3.18: Configuration with missing E_T between the two tau leptons

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 3.19 and 3.20. 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 3.19 and 3.20) or must be opposite to them (see the so-called angular sector B of Figures 3.19 and 3.20).

3.5.2.2 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 3.19 and 3.20). 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

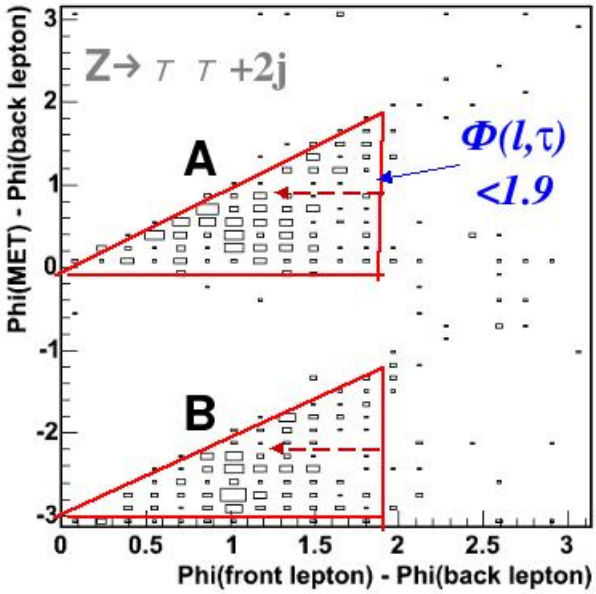


Figure 3.19: Angular relations between the two leptons and missing E_T : Z background case. The triangle A corresponds to $\not\!\!E_T$ pointing between the tau and the lepton, and the triangle B corresponds to $\not\!\!E_T$ back to back with the tau and the lepton. The *front* and *back* adjectives refer to the relative angular direction of the two leptons in the directly oriented $r - \Phi$ plane.

signal. As shown in Fig.3.18, 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.3.21 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 are cut out:

- 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.

This selection eliminates 70% of the Z-background and only 8% of the top into tau dilepton signal.

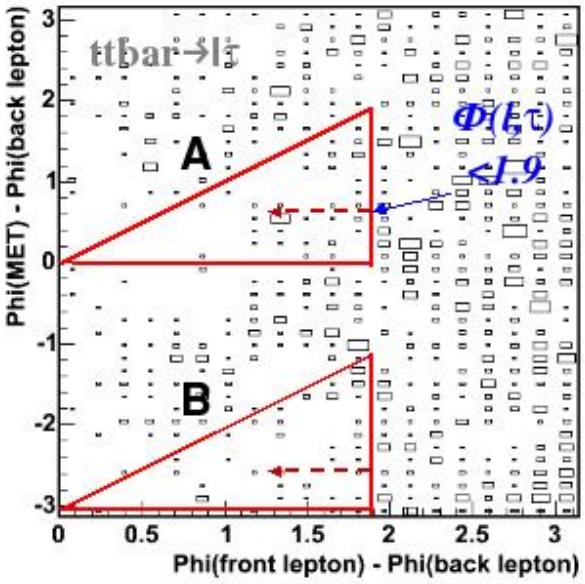


Figure 3.20: Angular relations between the two leptons and the missing E_T : $t\bar{t}$ case. The triangle A corresponds to $\not\!\!E_T$ pointing between the tau and the lepton, and the triangle B corresponds to $\not\!\!E_T$ back to back with the tau and the lepton

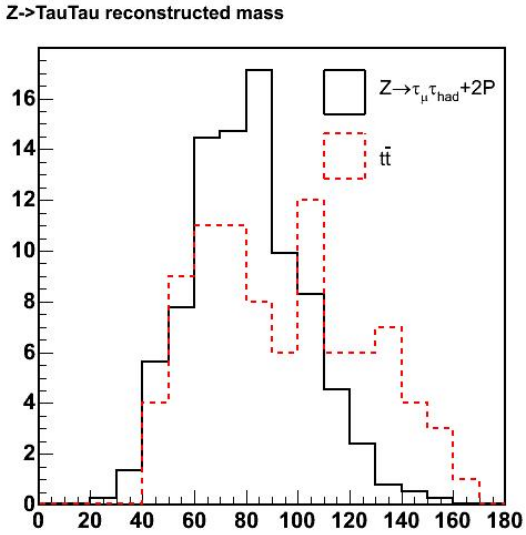


Figure 3.21: The reconstructed mass of the Z-boson in $Z \rightarrow \tau\tau$ + jets. Comparison with top into tau dilepton events for events passing the A-B preselection. Note that this plot gathers A+B regions, and thus is worse for $Z \rightarrow \tau\tau$ than it would be for the region A alone.

3.5.3 The event acceptance

The computation of the signal acceptance relies on the Pythia Monte Carlo generator, tuned with data for what concerns the lepton identification efficiencies. A Monte Carlo sample made by the top working group (sample labelled as "ttopkl") is used here. The top mass is set at 175 GeV/c^2. In order to ensure that there is no double counting of events due to fake reconstructed leptons, the sample is filtered such as to keep only events where the W decays into an electron or a muon and the other one into a tau lepton, which then decays hadronically. Each reconstructed object is then required to match its generator level parent particle. This is achieved by asking that the distance parameter defined as $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, be less than 0.04 for electrons and muons. As the direction of the reconstructed hadronic tau goes, it is a little bit shifted from its parent τ lepton because of the undetected τ neutrino emitted during the decay. The angle ΔR between the reconstructed τ lepton and its parent particle has ben calculated with Pythia and is shown in Fig.3.22. A choice of $\Delta R \leq 0.2$ is made for the tau matching cut.

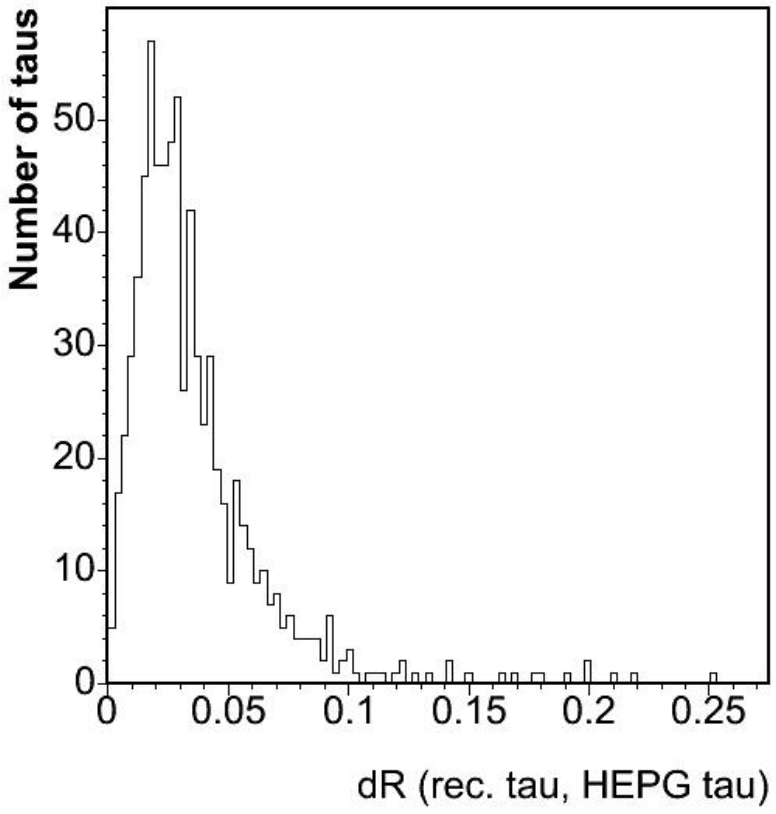


Figure 3.22: Angular difference in the $\eta - \Phi$ space between the generator level τ parton and the reconstructed hadronically decaying τ, calculated with the Pythia generator.

The estimate of the efficiency of the raw signal as obtained with the Monte Carlo

sample is given in Table 3.5. The geometrical and kinematical acceptances are computed applying the series of requirements here below:

N(channel) is the number of generated events passing a particular decay channel.

N($e\mu_{geom}$) is the number of events in which a reconstructed electron (muon) passing a very basic selection (CdfEmObject or CdfMuon) is found and matches a generated electron(muon). It must hit an active zone of the central calorimeter and of the COT (requirement of "$fiduciality''$), and the track $z0$ must be less than 60 cm.

N($e\mu_{kine}$): The electron (muon) must pass the kinematical requirements: $E_T > 20$ GeV and $track\ Pt > 10$ GeV/c for electrons, and $track\ Pt > 20$ GeV/c for muons.

N(τ_{geom}): A reconstructed tau passing a basic selection (TauFinder) is found and matches a generated tau. It must be fiducial in the central calorimeter, and the leading track must lie within 60 cm in z0 and 200 mm in d0 of the interaction point.

N(τ_{kine}): The p_T of the tau lepton, computed using tracks and pi0s, must exceed 15 GeV/c

N(1 jet): A jet (cone 0.4) $E_T > 15\ GeV$ (reclustered with the highest p_T vertex as a reference point, and corrected to the hadron level at level 5) and between -2 and +2 in pseudorapidity is found.

N(2 jet): A second jet (cone 0.4) with corrected $E_T > 15\ GeV$ and between -2 and +2 in pseudorapidity is found

Geometrical and kinematical acceptances of 0.25% for e+τ_{had}, and of 0.02% for $\tau_e\tau_{had}$ are found. To compare with previous studies performed in the past at CDF, we relax the requirement of two jets. This yields to 0.31% for e+τ_{had} and 0.03% for $\tau_e\tau_{had}$, in good agreement with [2].

In order to obtain the total Monte Carlo event acceptance, we apply the requirements of our event selection in that order, where N(cut) means the number of events found passing this cut and all former ones.

N($e\mu_{ID}$): The electron (muon) passes all identification cuts but isolation

N(τ_{ID}): The tau lepton passes all identification cuts but the isolation one.

N($e\mu_{iso}$): The electron (muon) is isolated.

N(τ_{ID}): The tau is isolated.

N(Opp Sign): The electron (muon) and tau lepton have opposite charge.

N(1 jet > 25 GeV): One of the jets has more than 25 GeV in E_T

N($\not\!\!E_T > 20$ GeV): The missing transverse energy, computed from the sum of calorimeter tower energies (taking the highest p_T vertex as a reference point and giving each tower an artificial zero mass to get a momentum vector from each tower energy), corrected for isolated central identified muons and for jet corrections, is required to be more than 20 GeV

N($H_t > 205$ GeV): The sum of E_T of the two jets, the electron (muon), the tau lepton and $\not\!\!E_T$ is greater than 205 GeV

N (Z veto): Number of events passing the $Z \rightarrow \tau\tau + 2$ jets veto, described in section 3.5.2.

The last step is to correct the Monte Carlo acceptance for discrepancies with the real data, using the correction factors to the Monte Carlo, gathered in Table 3.8 [31].

Finally, 673 events are found in the $e + \tau_{had}$ channel and 60 in the $\tau_e\tau_{had}$ channel out of 20937 (corresponding to a branching ratio, BR, equal to 0.65*2/81) and 3771

	Number of MC events			
Cut	$e\tau_{had}$	$\tau_e\tau_{had}$	$\mu\tau_{had}$	$\tau_\mu\tau_{had}$
N(channel)	20937	3771	20588	3685
N($e\mu_{geom}$)	11178	1605	11148	1609
N($e\mu_{kine}$)	9471	830	9637	828
N(τ_{geom})	3990	383	4018	334
N(τ_{kine})	3577	343	3585	295
N(1 jet)	3536	335	3538	291
N(2 jets)	2840	263	2793	233
N($e\mu_{ID}$)	2319	218	1906	142
N(τ_{ID})	1284	131	1107	81
N($e\mu_{iso}$)	1216	117	1033	75
N(τ_{iso})	887	93	768	56
N(Opp Sign)	884	92	763	55
N(2 jets>15GeV)	884	92	763	55
N(1 jet>25GeV)	874	91	754	55
N($\not\!\!E_T$>20GeV)	805	86	688	41
N(H_t>205GeV)	729	70	610	31
N(Z veto)	673	60	568	31

Table 3.5: *Acceptance table: Number of events in Pythia $t\bar{t}$ passing each individual cut of the event selection*

type	scale factor value
$\epsilon_{trigger}$: high p_T CEM	0.962 ± 0.006 [46]
$\epsilon_{trigger}$: high p_T CMUP	0.9078 ± 0.0047 [31]
$\epsilon_{trigger}$: high p_T CMX	0.9649 ± 0.0040 [31]
ϵ^{e}_{ID}	0.996 ± 0.004 [24]
ϵ^{CMUP}_{ID}	0.8738 ± 0.0086 [31]
ϵ^{CMX}_{ID}	0.9889 ± 0.0063 [31]
ϵ^{τ}_{ID}	0.95 (350 pb^{-1}) or 1.0 ± 0.09 (1 fb^{-1})

Table 3.6: *Trigger efficiencies, and scale factors (ratio of efficiencies for data and Monte-Carlo)*

(corresponding to a branching ratio, BR, equal to 2*0.175*0.65*0.108*0.108) events. This yields to a Monte Carlo acceptance of $0.049 \pm 0.002\%$ and $0.0043 \pm 0.0006\%$ respectively. After multiplying with the relevant scale factors given in the Table 3.8, the following acceptance values are obtained: $0.043 \pm 0.002\%$ for the $e + \tau_{had}$ case and $0.0039 \pm 0.0006\%$ for the $\tau_e \tau_{had}$ case. It gives a total of **0.047 ± 0.002%** relative to all $t\bar{t}$ acceptance. The quoted uncertainties for the acceptances are statistical only.

For the muon channels, after separating the CMUP and the CMX contributions, acceptance values of 0.026 ± 0.001% for the CMUP muons and of 0.0083 ± 0.0017% for the CMX muons are obtained as an average through the $344 pb^{-1}$ of good muon list data.

3.5.4 Expected number of signal events observed in the data

For the $t\bar{t}$ cross section, the last CDF combined 2006 result is used [22]: 7.3 pb. Assuming this cross section, using the signal acceptance, 1.32±0.05 signal events are expected in the electron channel $359 pb^{-1}$ good electron run list data. 0.92±0.05 signal events are expected in the muon channel $344 pb^{-1}$ good muon run list data. This makes a total of **2.2±0.1** tau dilepton events expected in the pre-september 2004 shutdown data. In comparison, the former analysis made with 195 pb^{-1} had an expectation of 1.0±0.2 signal events. [19]

3.5.5 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 as decribed in Chapter 2. 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 [29] ("dilepton" = ee, $e\mu$, $\mu\mu$) . The same selection as for the CDF top dilepton cross section measurement described in [30] 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 fake rate can be validated in W+jets events (see section 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.

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

3.6.1 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 3.5) 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

3.6.2 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 nth variable and Bn is the background distribution of the nth variable.

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

- Pythia is used for the signal and the $t\bar{t}$ $\rightarrow$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}$$\rightarrow$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.

3.6.2.1 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.

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 3.19 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.3.23 presents the 10 distributions for the signal and the three backgrounds.

3.6.2.2 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 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.3.24 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.

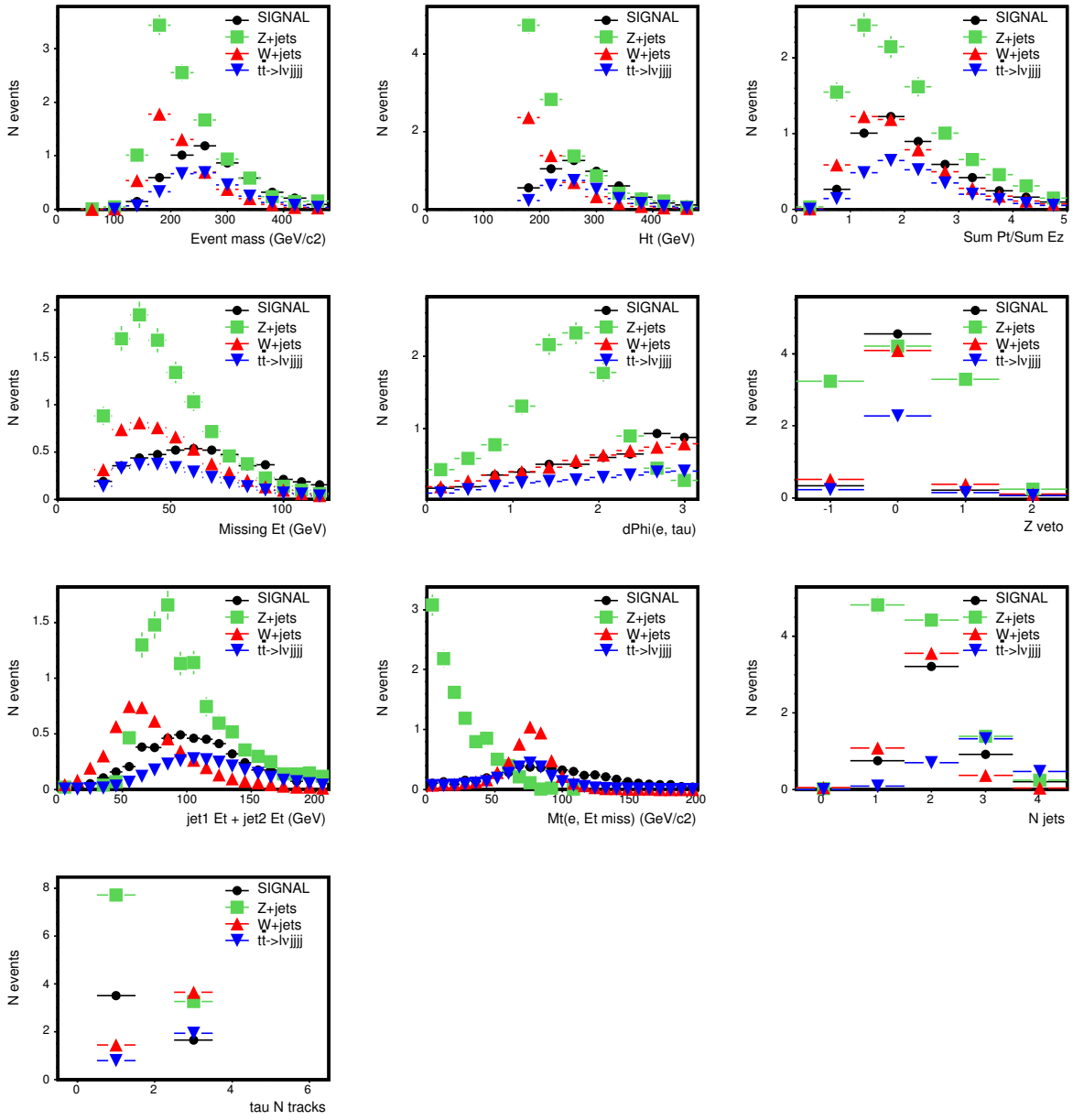


Figure 3.23: Superposition of the distributions for the 10 variables selected to be part of the likelihood ratio.

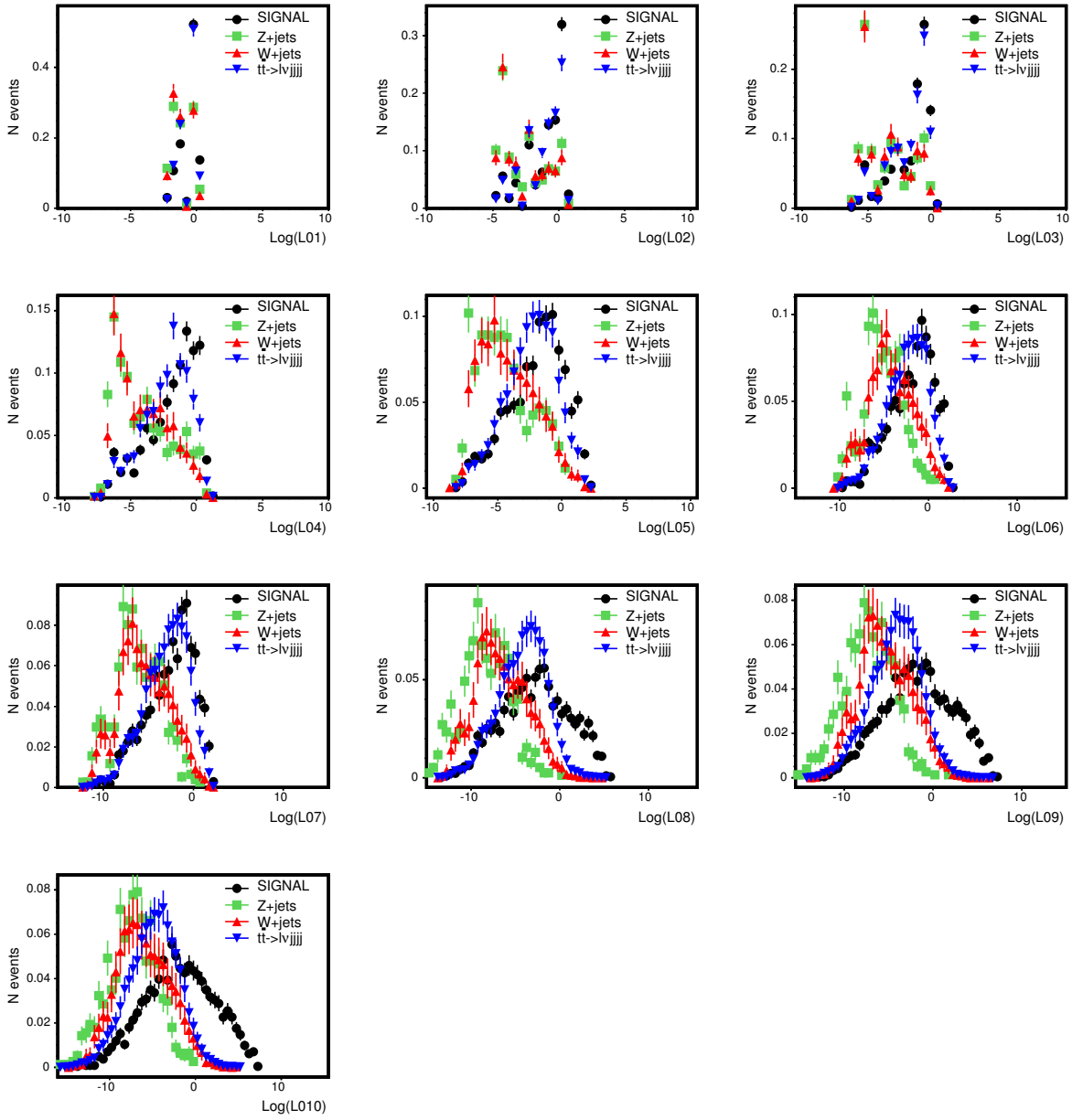


Figure 3.24: Superposition of the distributions for the 10 likelihood variables $L0_n$.

3.6.2.3 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.3.25 shows all the reweighted $b1_k$ distributions for backgrounds and $S1_k$ distribution distribution for the top in tau signal.

Fig.3.26 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)

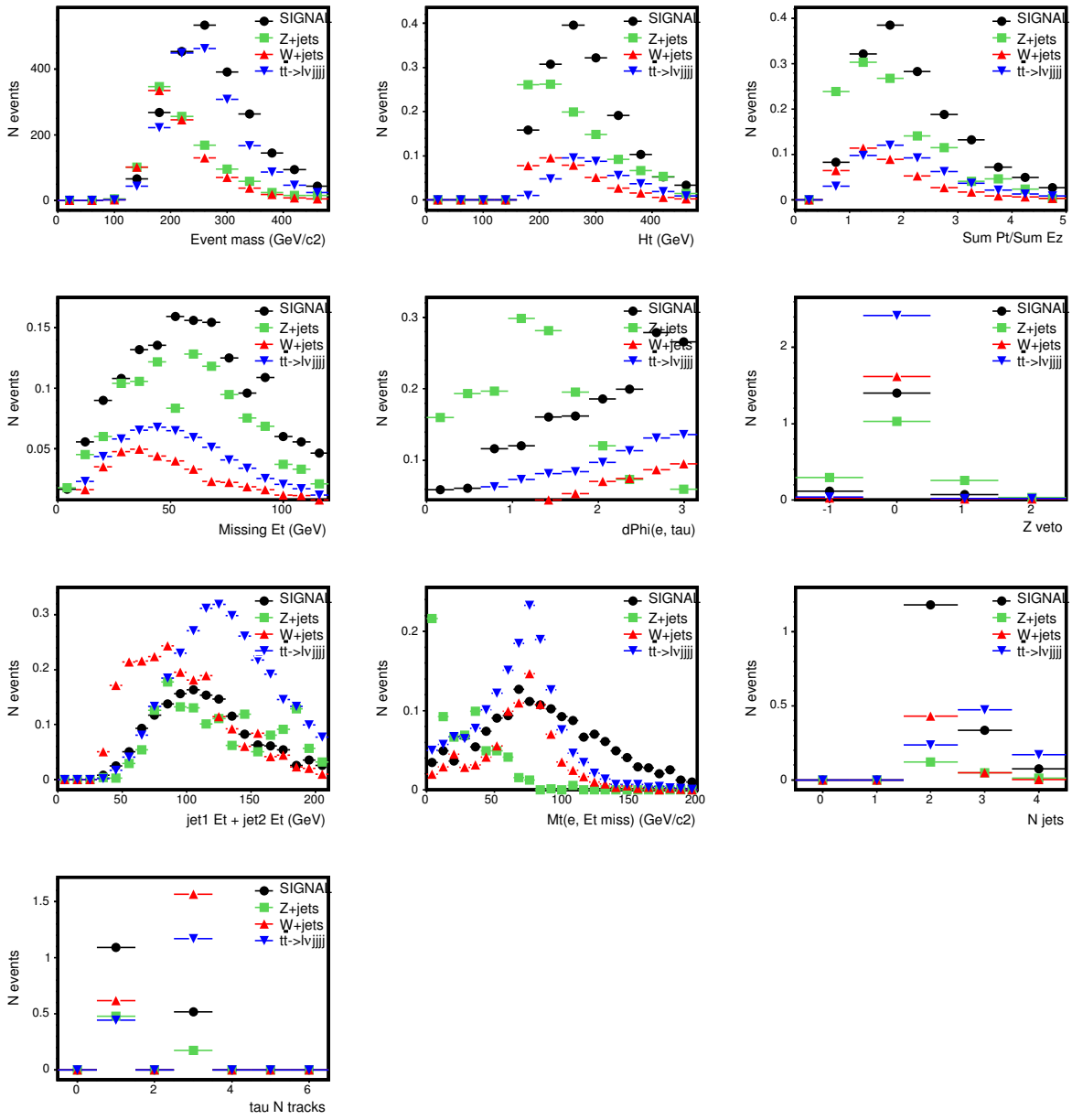


Figure 3.25: Superposition of the reweighted $b1_k$ and $S1_k$ distributions for the 10 variables (k=1,...,10) .

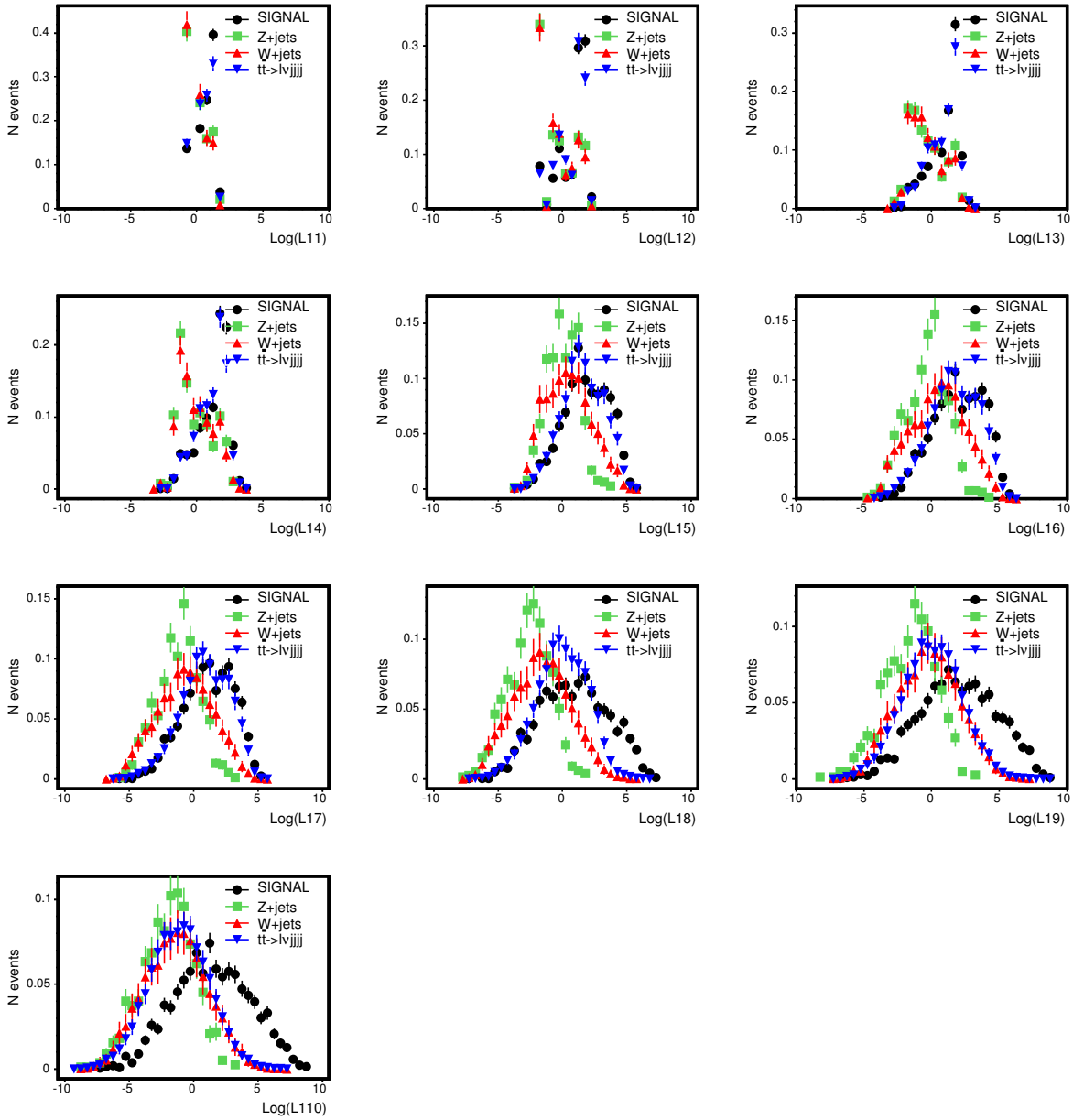


Figure 3.26: Superposition of the distributions for the 10 likelihood variables $L1_n$.

3.6.2.4 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. 3.27 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. 3.28 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. 3.29 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.

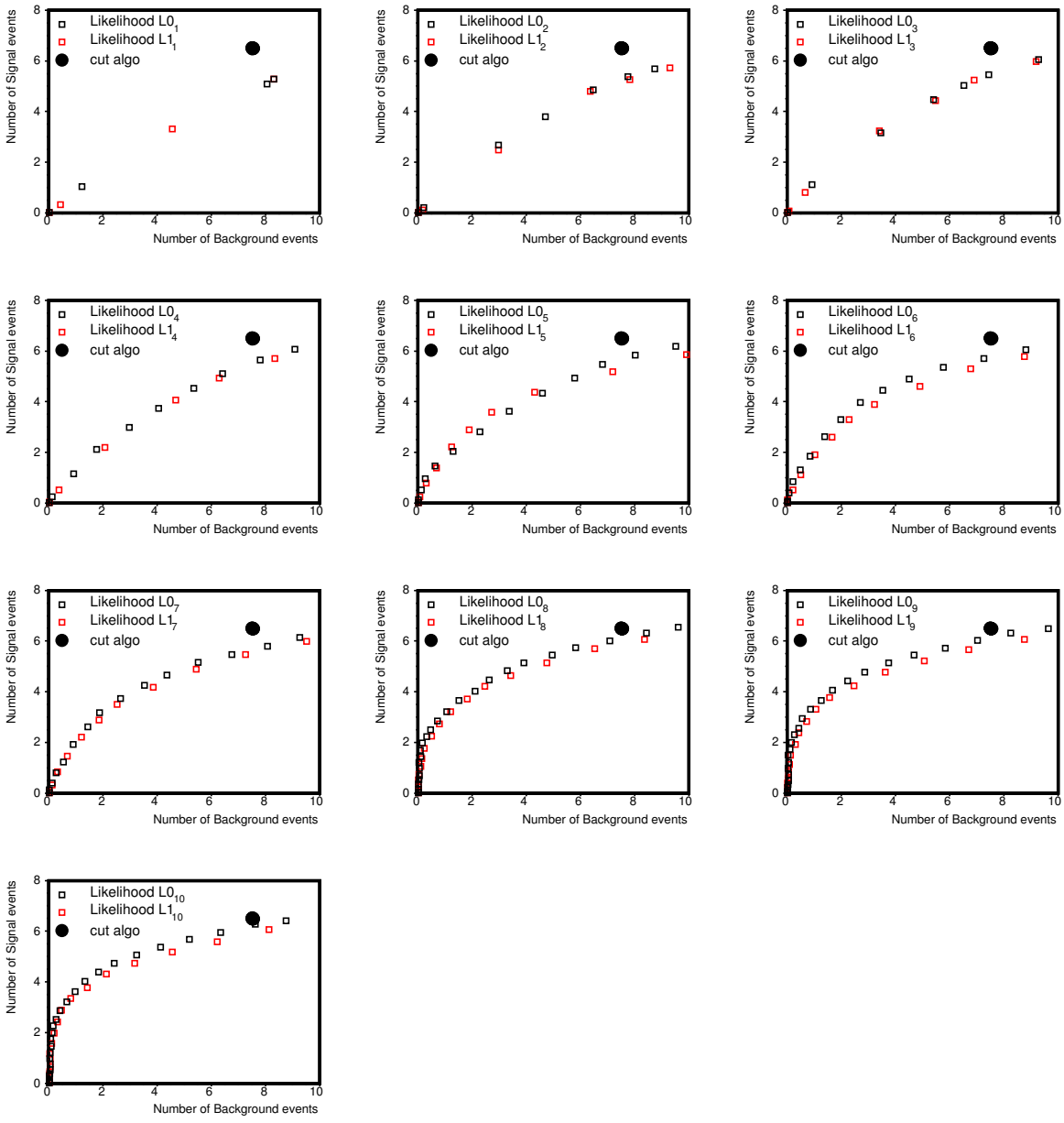


Figure 3.27: Relations between the remaining number of signal and background events for the L0, L1 and cut algorithm.

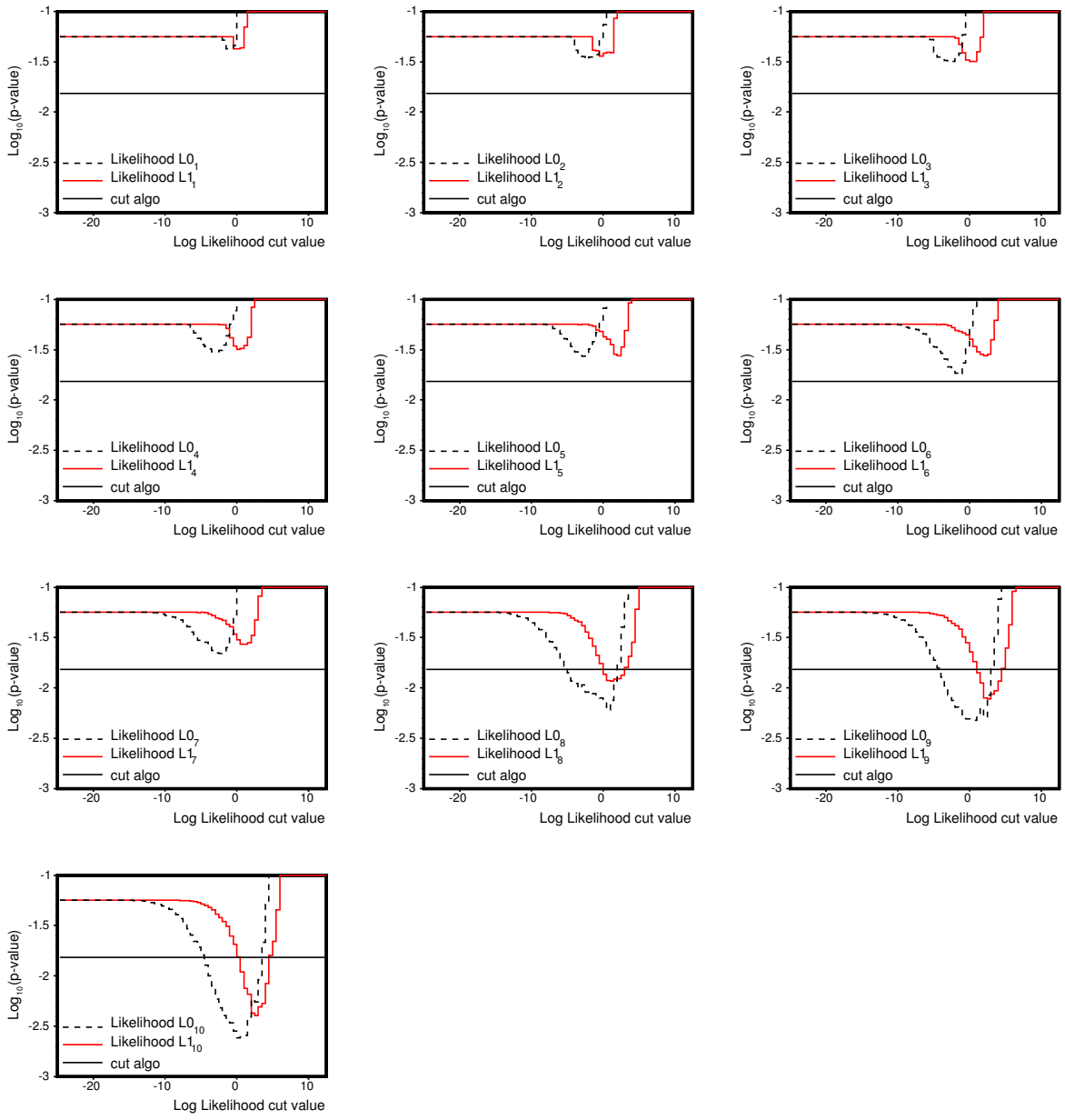


Figure 3.28: Relations between the logarithm of the expected p-value and the c cut on the log likelihood number, for L0 (black) and L1 (red) likelihood methods. The horizontal line stands for the sensitivity achieved with the cut based selection of the 350 pb^{-1} analysis if applied to the 1 fb^{-1} sample.

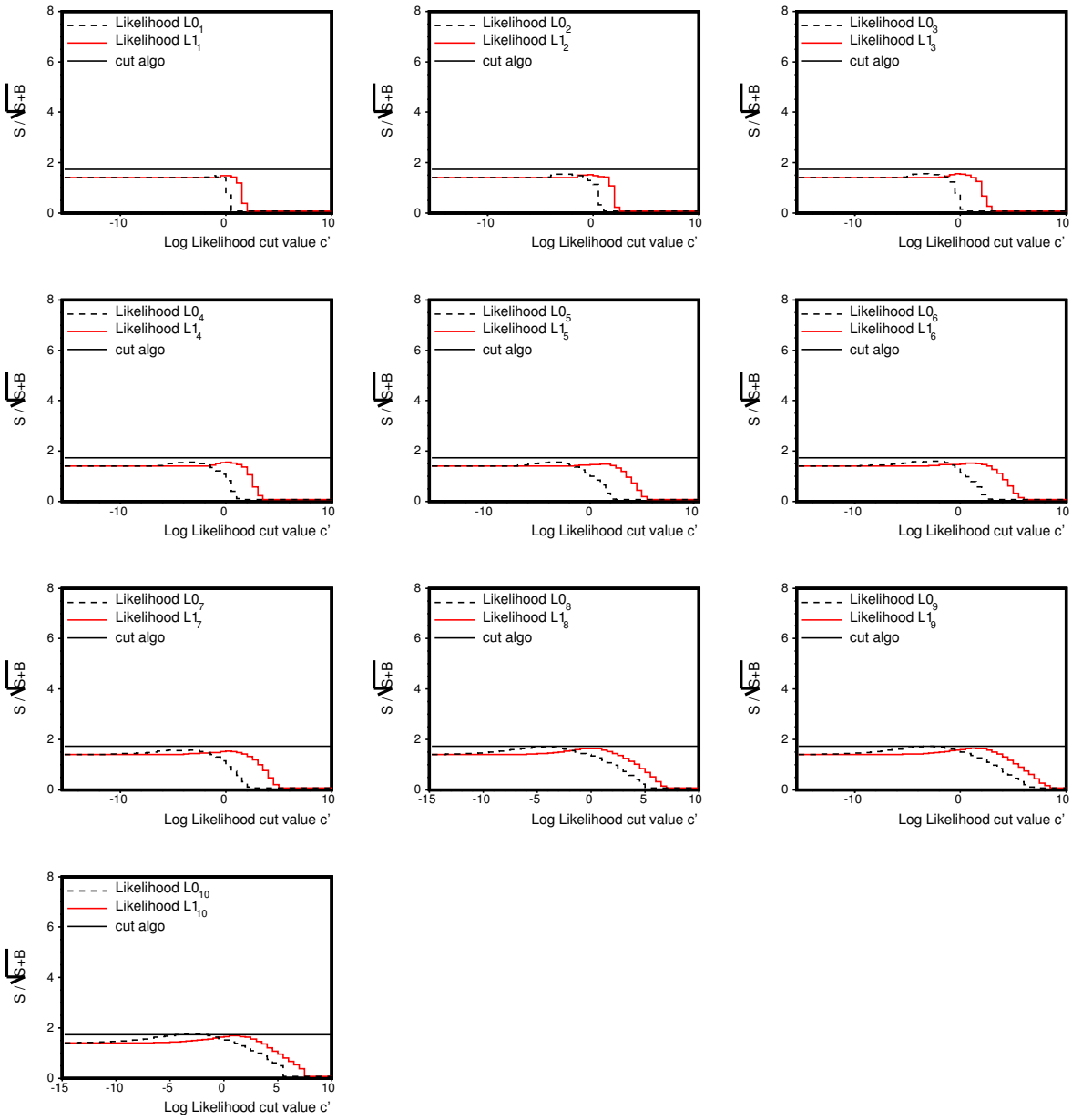


Figure 3.29: Relations between the logarithm of the expected p-value and the c' cut on the log likelihood number, for L0 (black) and L1 (red) likelihood methods. The horizontal line stands for the sensitivity achieved with the cut based selection of the 350 pb^{-1} analysis if applied to the 1 fb^{-1} sample.

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

3.6.3 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 section 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 3.7. Most variables are common to the 350 pb^{-1} analysis and were explained in the section 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 section 3.5.2 is greater than -5.

N(L0 > 0): the L0 log likelihood ratio, as defined in the section 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 3.8 [31].

type	scale factor value	
	2002-2004 runs	2005-2006 runs
$\epsilon_{trigger}$: high p_T CEM	0.962 ± 0.006	0.9773 ± 0.0044 [45]
$\epsilon_{trigger}$: high p_T CMUP	0.8890 ± 0.0043	0.9187 ± 0.0028 [48]
$\epsilon_{trigger}$: high p_T CMX	0.9675 ± 0.0033	0.9508 ± 0.0029 [48]
ϵ_{ID}^{e}	0.986 ± 0.004	0.975 ± 0.004 [47]
ϵ_{ID}^{CMUP}	0.9285 ± 0.0051	0.9242 ± 0.0037 [48]
ϵ_{ID}^{CMX}	0.9988 ± 0.0055	0.9760 ± 0.0039 [48]
ϵ_{ID}^{τ}	1.0 ± 0.09	1.0 ± 0.09

Table 3.8: *Scale factors by which to multiply the Monte Carlo acceptance.*

As a result, after having combined the Monte Carlo acceptances of Table 3.7 and the scale factors of Table 3.8, 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 $r_\tau = \frac{t \to \tau \nu_\tau b}{t \to l \nu_l b}$):
  - $e + \tau$: $5.0 \pm 0.1_{stat.} \times 10^{-4}$
  - $\mu + \tau$: $3.4 \pm 0.1_{stat.} \times 10^{-4}$

3.6.4 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 [22], 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 $r_\tau = \frac{t \to \tau \nu_\tau b}{t \to l \nu_l b}$):
  - $e + \tau$: $3.8 \pm 0.1_{stat.}$
  - $\mu + \tau$: $2.6 \pm 0.1_{stat.}$

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.

3.7 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

3.7.1 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.

3.7.1.1 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 3.30 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 3.31 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 of fig 3.31 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.

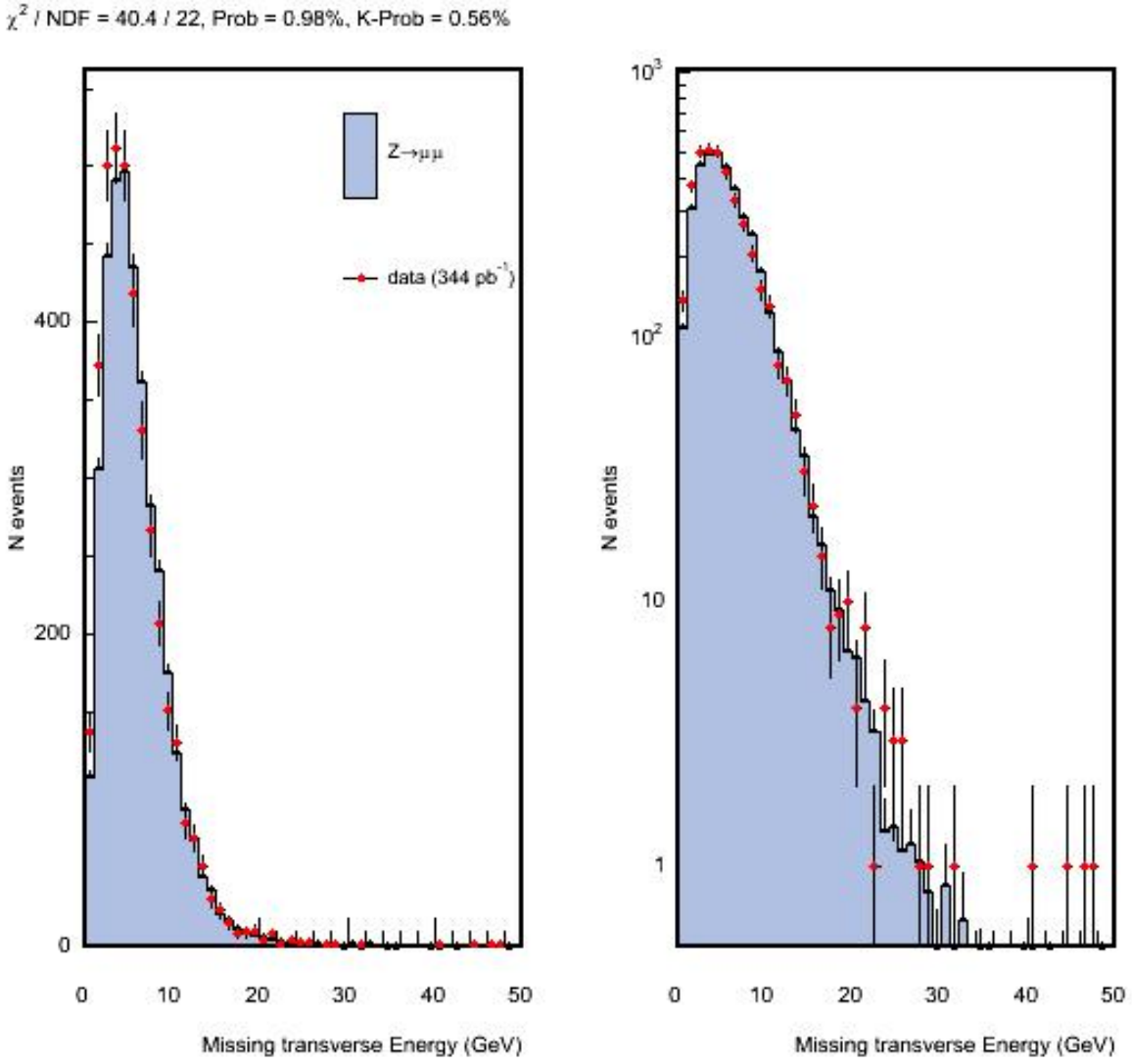


Figure 3.30: *Fake missing E_T present in $Z \to \mu\mu$ events. Left: linear scale ; Right: log scale.*

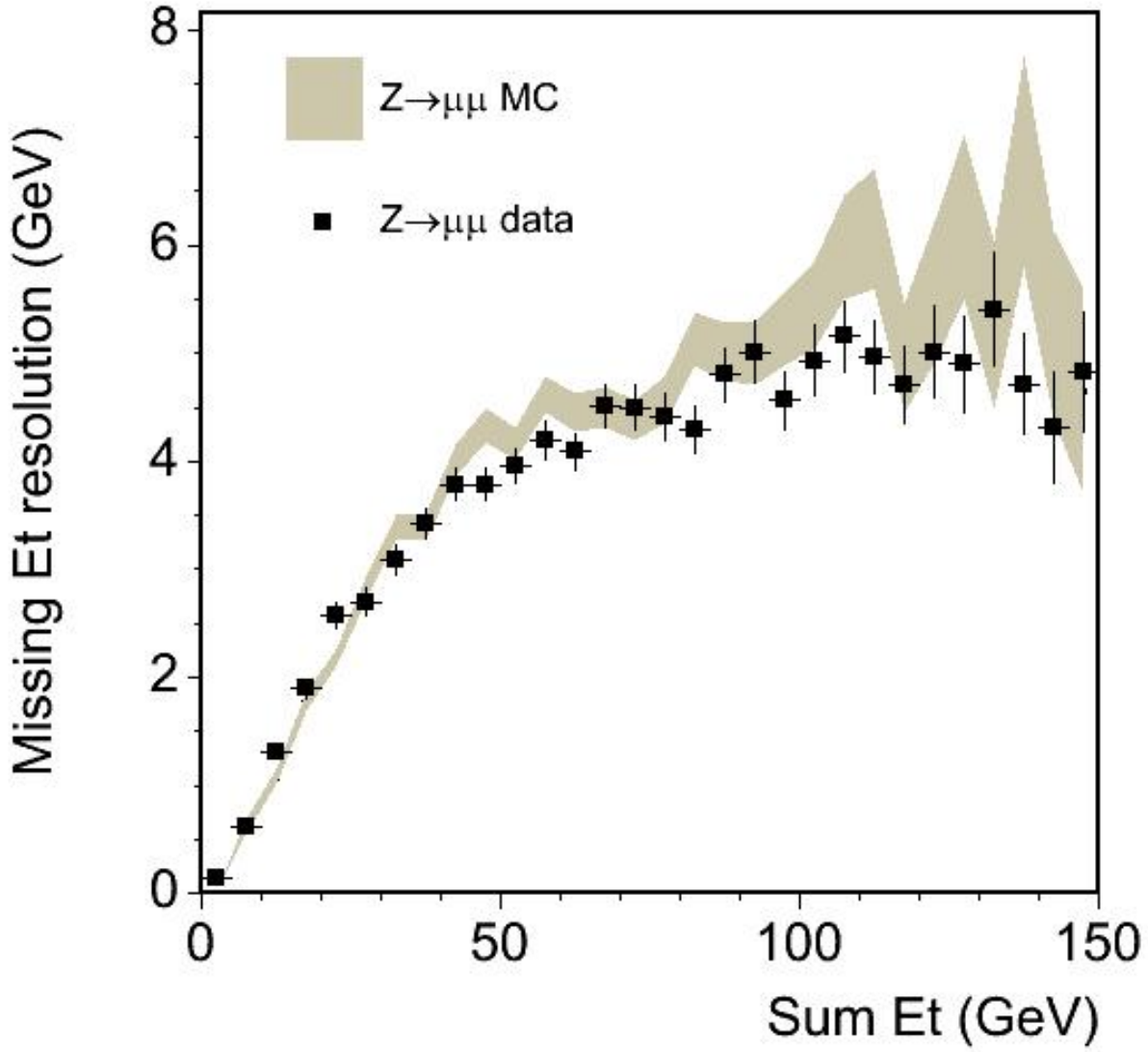


Figure 3.31: *Missing E_T resolution orthogonal to the Z boson direction as a function of Sum E_T.*

3.7.1.2 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 $2 \rightarrow 2$ processes (eg. $q\bar{q} \rightarrow Z \rightarrow ll$, $gg \rightarrow g \rightarrow t\bar{t}$,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 $Z \rightarrow \tau\tau + 2$ jets events will make use of AlpGen sample of $Z \rightarrow \tau\tau + 2$ 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 $Z \rightarrow \tau\tau + 2$ jets, this is possible for $Z \rightarrow ee$ or $Z \rightarrow \mu\mu + \geq 2$ jets. If the cross section of $Z \rightarrow \mu\mu + \geq 2$ jets can be obtained from data, this can be assumed to be valid also for $Z \rightarrow \tau\tau + \geq 2$ 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 3.32 compares the number of extra jets with corrected energies greater than 15 GeV, and pseudorapidities between -2 and 2 in $Z \rightarrow \mu\mu$ 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 3.33, 3.34). 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 $Z \rightarrow \mu\mu$ will be used to normalize the prediction from Alpgen+herwig used in the analysis for the $Z \rightarrow \tau\tau$ background. Furthermore, it is claimed that this validates the use of Pythia for the estimate of the small diboson background.

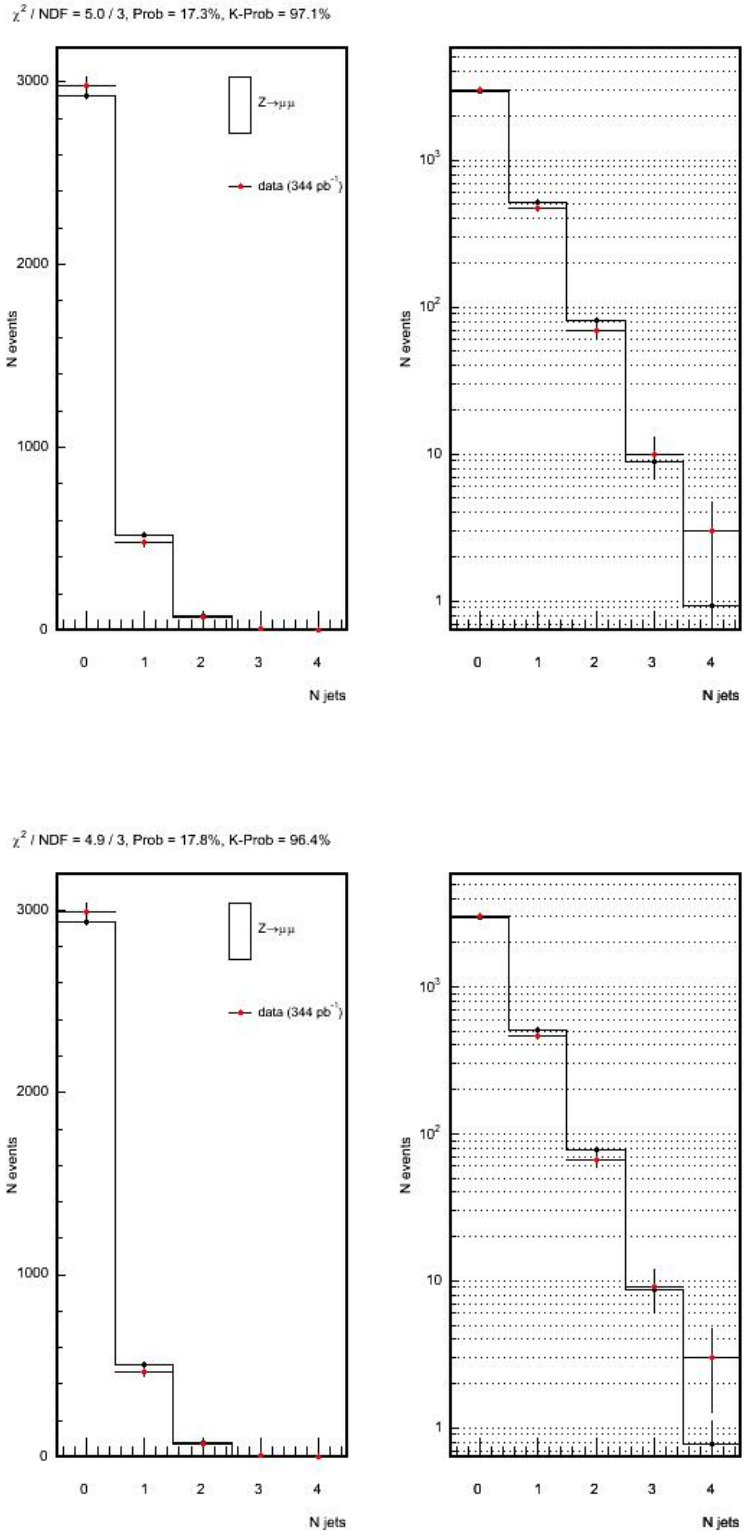


Figure 3.32: Number of extra jets in $Z \to \mu\mu$. Jet selection: $E_T^{cor} > 15$ GeV, $|\eta| < 2$. Top: two central stub muons, bottom: two central stub muons, MET<20 GeV.

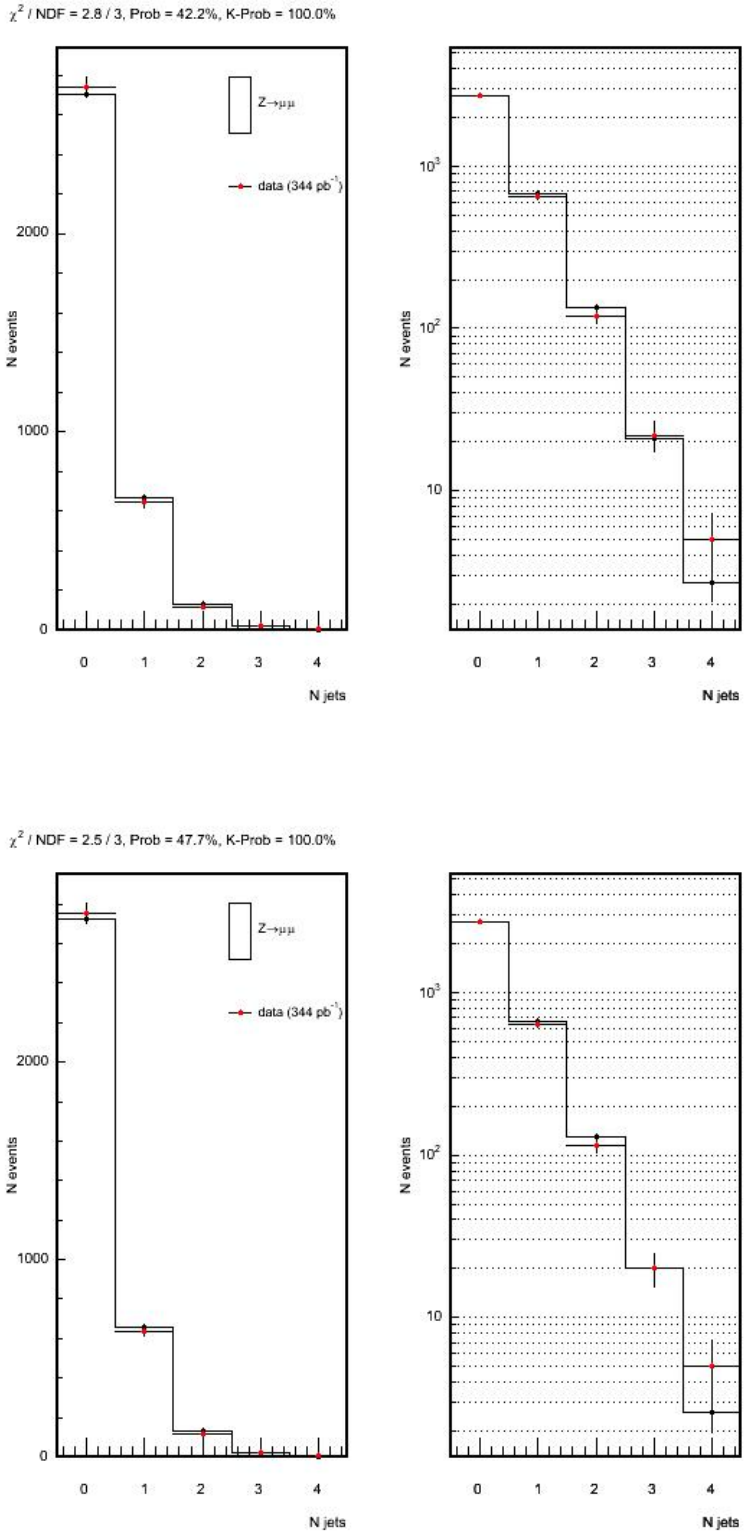


Figure 3.33: Number of extra jets in $Z \rightarrow \mu\mu$. Jet selection: $E_T^{cor} > 10$ GeV, $|\eta| < 2$. Top: two central stub muons, bottom: two central stub muons, MET<20 GeV.

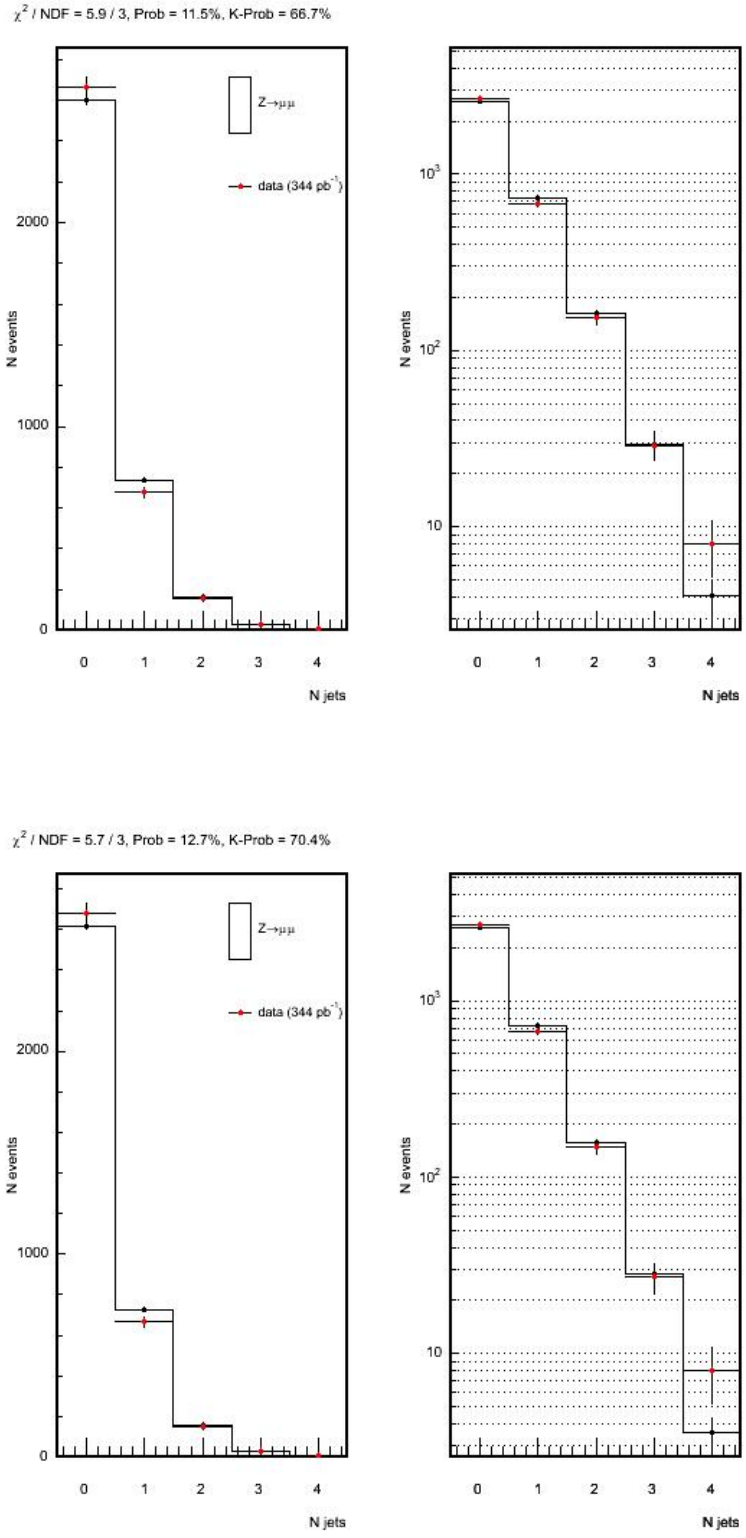


Figure 3.34: Number of extra jets in $Z \to \mu\mu$. Jet selection: $E_T^{cor} > 12$ GeV, $|\eta| < 2.5$. Top: two central stub muons, bottom: two central stub muons, MET<20 GeV.

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

3.7.2.1 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 [18] 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.

3.7.2.2 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. 3.37) and f(Njets) are correction factors. All this is the subject of the section 3.8.

3.8 Estimation of the background due to jets faking taus

3.8.1 Introduction

The goal was to estimate in this analysis, the number of jets that are misidentified as tau leptons (labelled as **jets faking taus**), especially those from the dominant background, i.e. the process: $W \rightarrow l\nu + \geq 3\ jets$. The Monte Carlo (Pythia and Herwig) overestimates the jet to tau fake rate by factors less than 2 [26, 27]. In the absence of a tuned MC tool able to reproduce the behaviour of the jets faking taus in different phase space regions of interest, the attempt was made to estimate it from the data only.

Indeed the inherent difficulty in this study, is that there is not enough data available that describe the jets in the needed signal phase space region . The signature of this signal is characterized by one lepton (electron or muon), a missing transverse energy ($\not\!\!E_T$) larger than 20 GeV, a high-activity in the detector summarized by ($Ht > 205GeV$) and at least three jets (of which one is identified as a lepton tau decaying hadronically).

Our method consists in estimating the jet to tau fake rate due to the jets produced in the process $W \rightarrow l\nu + \geq 3\ jets$ from the jet to tau fake rate computed on several sets of QCD dijet events. This is achieved following the three steps:

1. Estimate of a tau fake rate in the *unbiased* QCD dijet events.
2. Extrapolation to the QCD events with higher jet multiplicities and transverse energy, using the Top Multijet (SumEt) sample.
3. Verification that the obtained tau fake rate also applies to the jets produced in the process W+jets and thus validates the method.

3.8.2 Definition of the jet to tau fake rate

The jet to tau fake rate is defined as the average probability for a jet which passes the set of denominator cuts, which are enumerated below, to pass the remaining numerator cuts (also enumerated here below) and thus to be misidentified as a tau lepton.

The definition of the denominator is based on a high-quality selection, namely: The muon and the electron vetos are applied, the jet is required to be in the central region, the tau track must be of good quality and compatible with the interaction point. Furthermore it is based on the TauFinder algorithm in order to be able to use the tau variables at the output of the jet to tau fake rate. Otherwise, we would end up with some jets which have a non zero probability to be a tau, and thus cannot be rejected although they do not have a TauFinder object associated to them.

1. The denominator is defined with the following series of conditions:
   - The TauFinder algorithm requests:
     - A seed tower of > 6 GeV transverse energy
     - A seed track with transverse momentum p_T of > 4.5 GeV/c

  - $\leq$ six neighbouring towers with $>$ 1 GeV each
  - A cluster with $|\eta| < 1.1$
- The fiducial requirements are: 9 cm $<$ seed track $|z_{CES}| <$ 216 cm
- The transverse momentum of the tracks and neutral pions contained in the tau jet must be less than 15 GeV/c
- The z-distance to the vertex of the tau lepton, $\tau|z_0|$, must be less than 60 cm.
- The impact parameter of the tau lepton, $\tau|d_0|$, must be less than 0.2 cm.
- The electron veto defined by: $\frac{E_{had}}{\sum P} > 0.15$ must be applied.
- The muon veto, defined by cluster E_T / seed track p_T, must be less than 0.5.
- The seed track quality defined as at least 3 stereo and axial superlayers with $\geq$ 5 hits must be applied.

2. The numerator is defined with the following series of requirements

- The transverse mass of the tracks plus the neutral pions must be less than 1.8 GeV/c^2
- The relative calorimeter isolation computed in a cone in ΔR of 0.4 must be less than 0.1
- The track isolation, defined as the number of tracks with a transverse momentum of at least 1 GeV/c in a conical region between 10 and 30 degrees must be equal to zero.
- The Pi0 isolation defined as the number of Pi0 in a conical region between 10 and 30 degrees must be equal to zero

3.8.3 Estimate of the jet to tau fake rate in the dijet events

Several triggers selecting so-called QCD events, i.e. based on events with jets are at disposal. The samples used for estimating the jet to tau fake rate in the dijet data are:

- The trigger **ST05** selects events with a single calorimeter tower collecting more than 5 GeV.

- The trigger **Jet20** selects events with at least one jet defined in a cone in ΔR of 0.7, with more than 20 GeV transverse energy and with a single tower of more than 5 GeV transverse energy at level 1.

- The trigger **Jet50** selects events with at least one jet defined in a cone in ΔR of 0.7, with more than 50 GeV transverse energy and with a single tower of more than 5 GeV transverse energy at level 1.

- The trigger **Jet70** selects events with at least one jet defined in a cone in ΔR of 0.7, with more than 70 GeV transverse energy and with a single tower of more than 10 GeV transverse energy at level 1.

The data samples are dominated by events with two transversely back to back jets with transverse energies close to each other, i.e. well balanced in energy.

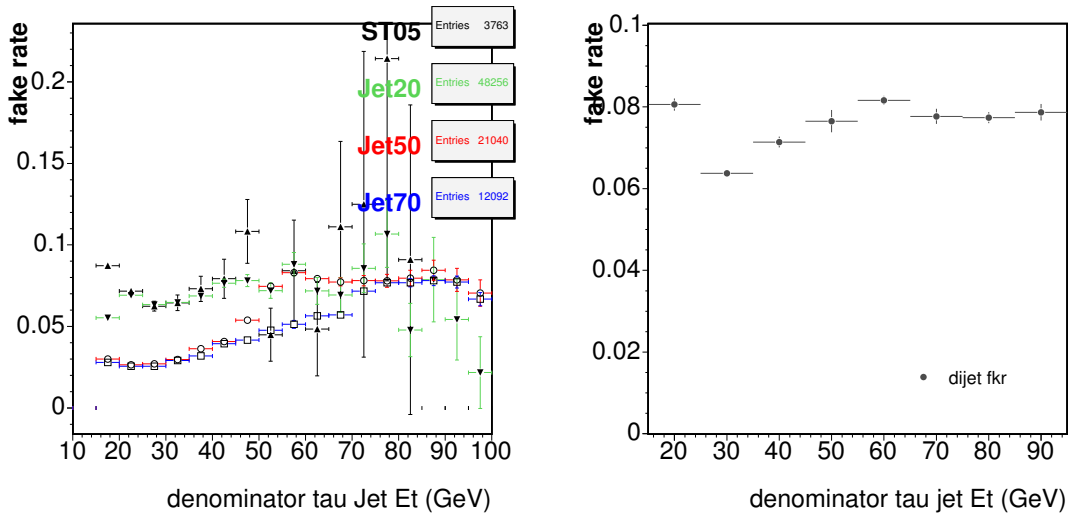
It should be noted that the triggers are based on jets with a cone in $\Delta R = 0.7$, whereas the jets used in this analysis are selected with the standard ClusterModule 0.4 cone algorithm: A trigger jet energy is thus always greater than the matching jet energy at the analysis level. Everytime a jet transverse energy (jet E_T) will be mentioned, it refers to a jet with a cone in ΔR of 0.4 and an uncorrected jet E_T, re-clustered at the highest p_T vertex in the event.

Using the ST05 sample, it was checked that the amount of jets with $E_T > 75$ GeV and with no tower with an energy exceeding 10 GeV is very small. Furthermore, the efficiency for a jet passing the tau identification denominator cuts, and with Et above 25 GeV, to be selected by the **Jet20** trigger, is on the order of 100%. The trigger efficiency plateau is indeed reached around this energy of 25 GeV. Likewise, all tau denominator jets with E_T greater than 55 and 75 GeV will be selected by the triggers **Jet50** and **Jet70** respectively. The denominator selection includes a cut of 6 GeV on the tau seed tower E_T; it is assumed that the **ST05** trigger is almost 100% efficient in selecting any events containing such a denominator jet. If 100% of the denominator jets are selected by these triggers it means that all these jets are *unbiased* by this trigger selection. The sample of denominator jets, in a given E_T range, extracted from any of the four triggers is thus well representative of the set of denominator jets of the same E_T range produced at Run II, or at least this is a good approximation. This gives us the means to calculate the jet to tau fake rate for unbiased QCD data as a function of the jet transverse energy.

The figure 3.35(a) shows the jet to tau fake rate as a function of the jet E_T. As expected, the jet to tau fake rates from the four samples agree remarkably well in the unbiased E_T regions, above the trigger thresholds. However, the jet to tau fake rate calculated for jets with E_T below the trigger threshold, shows some discrepancy as expected. Indeed the typical event provided by the jet triggers is a transversely back-to-back di-jet event with the two jets of approximately the same E_T and slightly above the trigger threshold. If a jet with E_T below the trigger threshold is selected, chances are high that this jet is coming from the gluon splitting produced by one of the original jets. The splitting leads to two lower energy jets too close to pass the numerator cuts, as these cuts are mainly based on isolation requirements. A lower jet to tau fake rate is thus expected for these *biased* jets and this is observed as shown in fig.3.35.

After having got rid of the biased jets below the energy thresholds, a much better estimate of the jet to tau fake rate is obtained by adding the contributions of the high energy portions of the four QCD triggers (see fig.3.35(b)). This is the $jettotaufakerate$ for generic dijet data. The jet to tau fake rate varies between 6 and 8%, depending on jet E_T.

This fake rate is not directly used in the top analysis but is perfectly suited to inclusive $Z \rightarrow \tau\tau$ or $H \rightarrow \tau\tau$ background estimates. It serves as a basis for estimating the jet to tau fake rates in more and more complicated QCD scenarios.



(a) *4 QCD triggers tau fake rates*

(b) *Estimated tau fake rate in unbiased dijet events*

Figure 3.35: *Jet to tau fake rate as a function of the uncorrected jet E_T. A jet is matched to every denominator tau to get the jet E_T.*

3.8.4 Estimate of the jet to tau fake rate in the multijet events

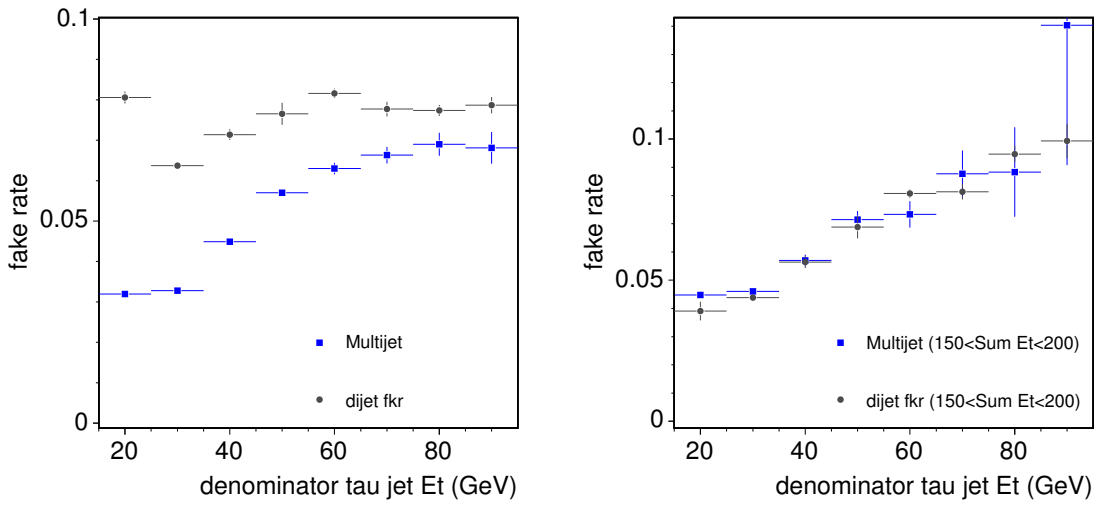
3.8.4.1 The jet to tau fake rate matrix FkR(jet E_T, Sum E_T)

Estimating the jet to tau fake rate for events with higher E_T activity and higher number of jets is much more difficult than in the case of inclusive dijet events.

The reference trigger here is the **Top Multijet** trigger which was developed and built for selecting the pair of top quarks events where the two produced W bosons decay into quarks. This gives typically six jets events and in any case it ensures a high activity and high jet multiplicity in the events. The trigger selection is made on events with 4 jets with E_T above 15 GeV and a sum of calorimeter transverse energies above 125 GeV. This is also called the **SumEt** trigger.

The data sample provided by the **SumEt** trigger allows to check the jet to tau fake rate, obtained before for dijet events, in event samples with a higher jet multiciplity and higher activity. Fig. 3.36(a) shows a comparison between the jet to tau fake rate computed in dijet samples with the jet to tau fake rate computed in the **SumEt** multijet sample. A discrepancy between the jet to tau fake rates in the two samples is observed.

Indeed, in order to reproduce the jet to tau fake rate in high jet multiplicity regions, it is required to take into account the higher value of Sum E_T in these events. A new parameter is thus added, namely the sum of the transverse energies of all the calorimeter towers in the event. It is computed at the highest p_T vertex of the event and corrected for the presence of standard central tight muons from the CMUP and CMX muon chambers (as defined in section 2.3.1.2). Fig. 3.36(b) shows as an example the nice agreement achieved between the jet to tau fake rate from dijet events recalculated from each trigger sample (ST05, Jet20, Jet50, Jet70) when restricting ourselves to events with SumE_T



(a) *Discrepancy between the jet to tau fake rate from* **SumEt** *events and the one from the dijet events.*

(b) *Agreement between the jet to tau fake rate from* **SumEt** *events and from the dijet events if 150 GeV < SumEt < 200 GeV.*

Figure 3.36: *Comparison between the jet to tau fake rates from dijet vs SumEt Multijet events.*

between 150 and 200 GeV.

The SumE_T parameter is thus introduced as a second parameter to compute the jet to tau fake rate so that it is also applicable to high jet multiplicity QCD events. As a result, a 2-dimensional matrix in(jet E_T, Sum E_T) is obtained (see Fig.3). It includes 16×8 bins, with the jet E_T varying from 15 to 95 GeV and the SumE_T varying from 0 to 400 GeV. The numbers quoted in this matrix are obtained from the four JetE_T triggers (always restricted to unbiased energy regions, above trigger thresholds) and from the multijet trigger. It is labelled as the *jet to tau fake rate matrix*. Note that the matrix top left portion is left empty because any event necessarily has a sum Et greater than the Et of any jet found in the event. The variations in the rates readable in the matrix top left part are thus simply due to some high statistical fluctuations of small numbers of events.

This matrix makes good predictions for the number of jets faking taus in the high p_T electron trigger (table 3.9), where no selection except the trigger requirement and the tau lepton identification are applied. Any jet matching the electromagnetic cluster passing the electron trigger requirement is removed from the list of the tau candidates.

3.8.4.2 Corrections depending on the number of jets and uncertainties on the estimate of the jet to tau fake rate in QCD events

The ability of the jet to tau fake rate matrix to correctly predict the number of jets faking tau leptons in event samples with given jet multiplicities, is now tested. To do this, we take events from the five QCD triggers used to build the matrix, split them according to their event jet multiplicities, and compare the number of identified tau lepton candidates really observed with the number predicted from the jet to tau fake rate matrix. The jet

	elec_25	*elec_35*
pred	9436	4675
obs	8588	4310
obs/pred	0.91±0.01	0.92±0.02

Table 3.9: *Predictions and observations of the number of jets faking tau leptons in the high p_T electron trigger. elec_y refers to the set of events in the electron trigger where the denominator jet $E_T > y$ GeV.*

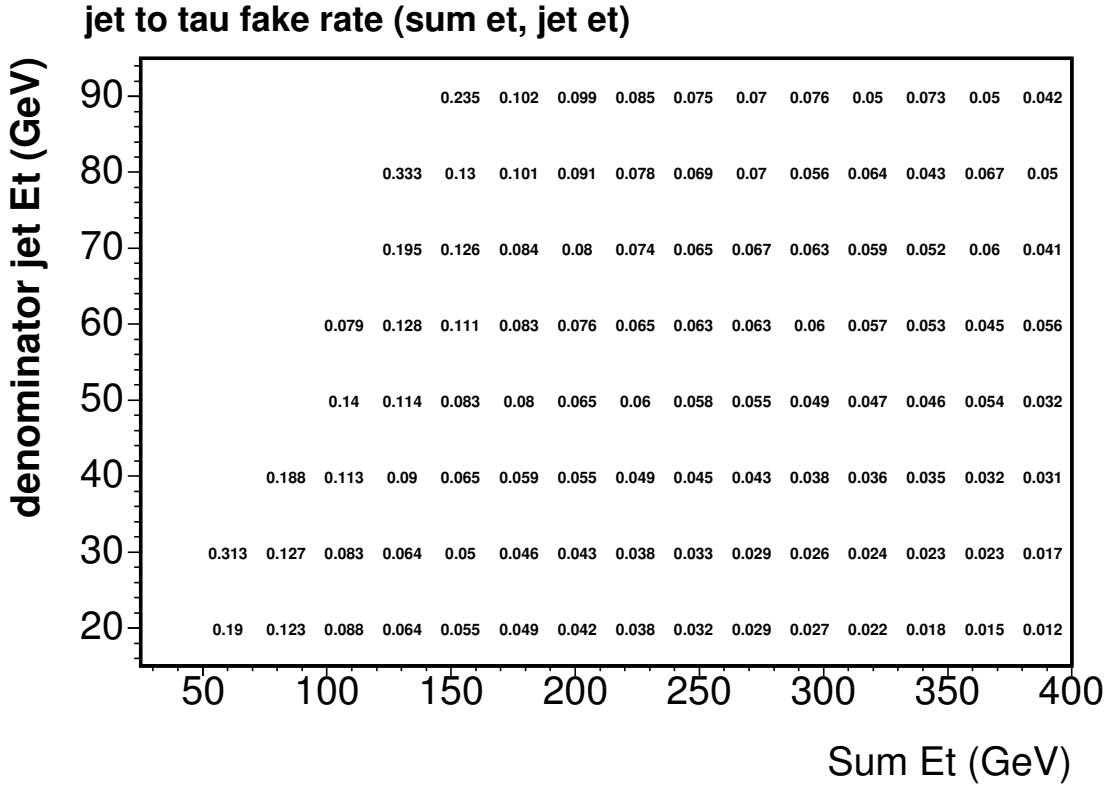


Figure 3.37: The jet to tau fake rate matrix used in the analysis.

to tau fake rate matrix is then to be considered successful if these two numbers are found equal. Let's remark at this point that this is true only because the number of true tau leptons is negligible compared with the number of jets faking tau leptons in the samples considered for the tests. The reason for this is that the probability for a jet to fake a hadronic tau decay is high (on the order of 1%)[11].

From the Jet20 trigger sample, we make two samples of jets, namely the *Jet20_25* that collects the jets with E_T greater than 25 GeV and that pass the denominator tau cuts, and *Jet20_35* that includes jets with E_T greater than 35 GeV and that pass the denominator tau cuts. Likewise, we define the jet samples *Jet50_55* and *Jet70_75*, from the Jet50 and Jet70 samples respectively. Then, *elec_25* and *elec_35* are defined in the same way from the *elec* high p_T electron trigger sample.

The table 3.10 tests the jet to tau fake rate prediction accuracy inside the six samples ST05, Jet20_25, Jet20_35, Jet50_55, Jet70_75 and SumEt. It shows three numbers, once for each sample: The observed number of jets identified as tau leptons, the predicted number of jets passing the tau identification selection given by the fake rate matrix, and the ratio between these last two numbers. The results are split into three categories following the event jet multiplicity. The table 3.11 performs the same tests for the jets in the elec_25 and elec_35 samples.

The analysis of these results shows that the number of jets faking tau leptons is still overestimated in events with high jet multiplicity and underestimated in those with low jet multiplicity. A new correction factor depending on the total number of jets in the event, labelled as *f(N jets)* is thus applied in order to get more valid predictions for events with more than two jets. The f(N jets) values must mirror the average of the ratios of observed over predicted numbers of tau leptons measured in the tests. The sample of jets from *Jet*20_25 has nice characteristics because it has high statistics in the different jet multiplicities considered and, furthermore, it shows ratios compatible with the average of the ratios calculated in all the other samples. That is why we choose it as our reference sample to decide the values for the f(N jets) correction factors.

As far as the determination of the systematic error associated with the fake rate goes, we compute it by taking the largest difference between all the ratios $\frac{Number\ observed\ taus}{Number\ predicted\ taus}$ measured in the eight samples and the ratio measured within Jet20_25.

One gets:

- f(1 jet) = 1.06±0.06
- f(2 jets) = 0.98±0.10
- f(≥3 jets) = 0.87±0.14

3.8.5 The jet to tau fake rate related distributions

Any estimated jet to lepton fake rate should correctly predict the number of misidentified leptons in any sample. It is however not supposed to be predictive with respect to the

[11] If we were considering other lepton fake rates (like electron), the contamination from true leptons should be taken into account carefully.

	$ST05$	$Jet20\_25$	$Jet20\_35$
1 jet (= the tau)	2026/1814 = 1.12±0.02	4967/4674 = 1.06±0.01	613/601 = 1.02±0.04
2 jets	1540/1748 = 0.88±0.03	10035/10230 = 0.98±0.01	2516/2506 = 1.00±0.02
≥3 jets	196/239 = 0.82±0.08	1946/2226 = 0.87±0.02	774/874 = 0.89±0.03
	$Jet50\_55$	$Jet70\_75$	$SumEt$
1 jet (= the tau)	578/537 = 1.08±0.04	214/199 = 1.08±0.08	42/51 = 0.82±0.15
2 jets	4502/4556 = 0.99±0.02	2521/2550 = 0.99±0.02	1002/951 = 1.05±0.03
≥3 jets	2832/2822 = 1.00±0.02	2031/2111 = 0.96±0.02	24535/24370 = 1.01±0.01

Table 3.10: *Predictions and observation of the number of jet to tau fakes within the 5 triggers. Jetx_y means a fake rate applied in the Jetx trigger with only denominator jets with $E_T > y$ GeV*

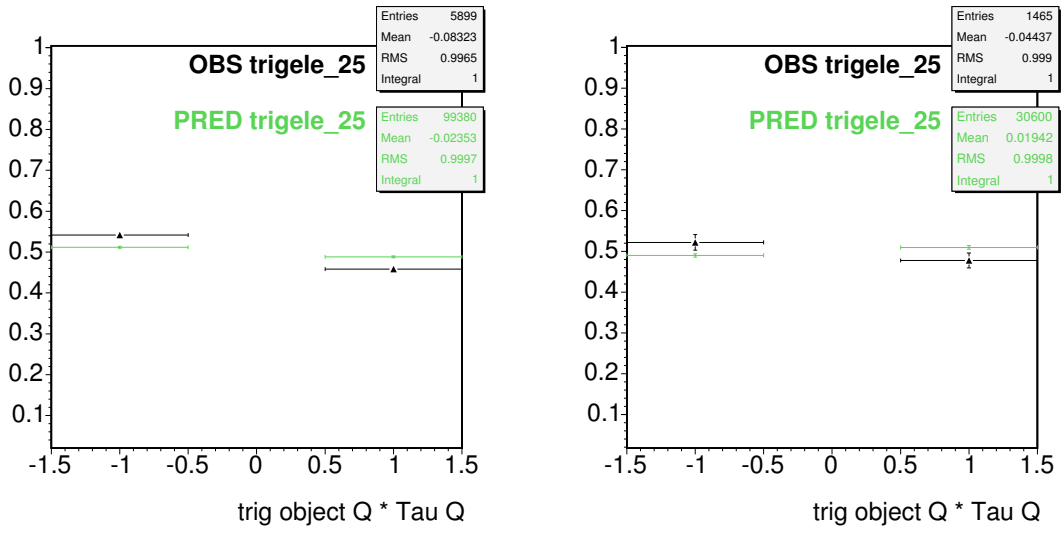
	$elec\_25$	$elec\_35$
2 jets (=the tau+the trigger object)	6419/6915 = 0.93±0.01	3132/3252 = 0.96±0.02
≥ 3 jets	2169/2521 = 0.86±0.02	1178/1423 = 0.83±0.02

Table 3.11: *Observations/Predictions of number of jet to tau fakes in the high p_T electron trigger. elec_y refers to the set of events in the electron trigger in which the denominator jet $E_T > y$ GeV*

distributions of any variable correlated to the variables used for the lepton identification. The case of the tau track isolation variable can be used as an example: Each jet passing the tau identification verifies the track isolation cut and thus has no track with transverse momentum higher than 1 GeV/c, within the tau isolation cone; however, the jet to tau fake rate applied to the denominator jets will sometimes include tracks in the isolation cone and it cannot therefore reproduce the distribution of the number of tracks in the isolation cone. The only way to overcome this issue is to correct the predicted distributions. In the case of the 350 pb^{-1} top dilepton analysis, the variables used for the signal event selection in addition to the number of jets are: The product of the tau lepton charge with the charge of the other lepton, the activity Ht in the event, defined as the sum of the transverse momentum of the tau, of the other lepton and of the other jets in the event plus the total transverse missing energy ($\not\!\!E_T$), and the leading jet transverse energy.

The fig.3.38 shows the predicted and observed distributions for the product of the leptons charges in the case of the sample selected with the high p_T electron trigger. Fig.3.39 shows the comparison of the measured and predicted Ht and leading jet E_T variables with the same sample of data.

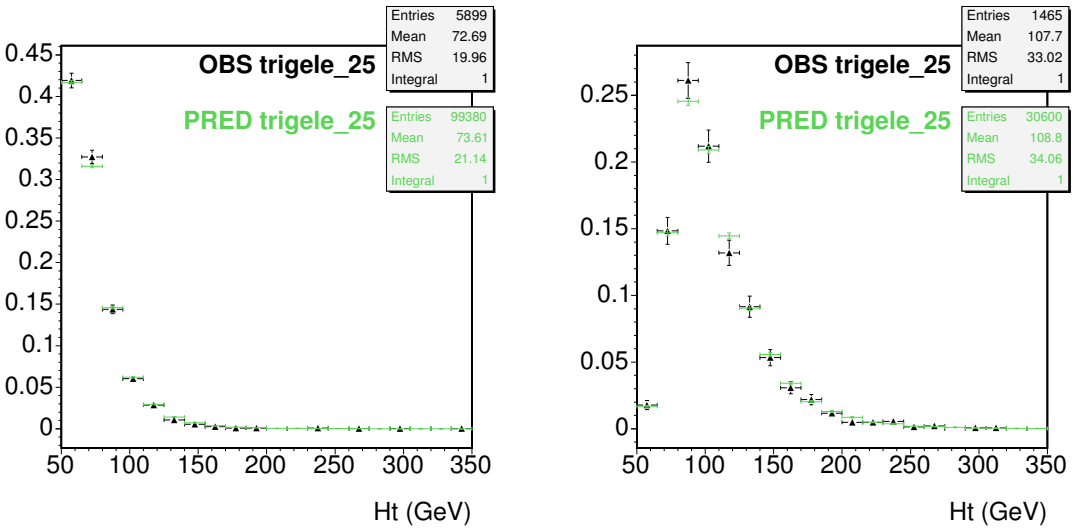
The agreement is rather good. The small discrepancy observed in the lepton charge product must be however corrected. Indeed each time the jet to tau fake rate is applied to the prediction of the lepton charge product, the predicted number of events with opposite charge must be multiplied by a factor 1.07, taking care to keep the total number of predicted events unchanged as this must be a correction to a predicted distribution, not to the total number of events.



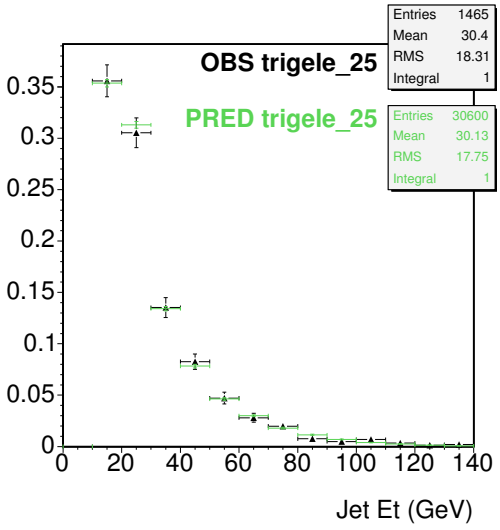
(a) *Case with 1 jet in the event (=the tau jet)*

(b) *Case with 2 jets in the event*

Figure 3.38: The predicted lepton charge product distributions obtained from the jet to tau fake rate compared with the observed ones in the elec_25 sample.



(a) *Ht 0j* (b) *Ht 1j*



(c) *leading jet* E_T

Figure 3.39: distributions from fake rate

3.8.6 The jet to tau fake rate for $W \rightarrow l\nu + jets$ events

The jet to tau fake rate obtained is defined as *fkr(jet E_T, sum E_T)× f(N jets).*

As previously discussed it applies to QCD backgrounds with both low or high activity and a number of jets up to 3. Can one apply it to the case of jets from $W \rightarrow l\nu + jets$ events? The presence of the W decaying into a lepton should not modify the shape of the jets. The fact that these events have a rather high activity and a large number of jets is taken into account by the SumE_T parameter and the N jets factor.

However, is the proportion of quark to gluon jets the same in W+jets than in QCD events? This is an important question because, in QCD, gluons have higher coupling strengths than quarks to emit extra gluons and, therefore, gluon jets tend to be broader than the quark jets [32] [33]. Because of this, one could expect a lower value for the jet to tau fake rate in the case of gluon jets than in the case of quark jets.

How to verify this possible feature of W+jets events in data? To do so, it is needed to select regions dominated by W+jets events. This is achieved by requiring that the events fulfill the following criteria: $\not\!\!E_T > 20\ GeV$, one central electron with $E_T > 20\ GeV$, and extra jets. But the additional requirement of one well-identified tau lepton results in a dramatic drop in the statistics and therefore it is not anymore possible to perform this test on the remaining sample (see first line of table 4). The requirements on $\not\!\!E_T$ and on the electron can hardly be loosened if one wants to stick in regions dominated by W+jets events.

We made an attempt to compute the ratios of jet to tau fake rates obtained in Pythia dijet events and Pythia inclusive W events [27] and to use these ratios to correct the jet to tau fake rate computed on QCD data. But this did not work properly.

Using data, the only way is to loosen the tau identification.To achieve this, seven selections based on the same tau denominators are defined. Firstly, in order to keep the tests independant from the original jet to tau fake rate numbers and to stay far from the top signal region in the 3 jets bins, they all require the denominator tau not to pass successfully the whole tau identification. The simple criteria applied are gathered in the first column of Table 3.12.

For each of these seven selections, the exact same method as the one previously described is applied to define a new jet to tau fake rate matrix and the new f(N jets factors) from the ST05, Jet20, Jet50, Jet70 and SumEt samples.

The results of the seven tests are shown in Table 3.12. The numbers quoted in the Table show the comparison between the observed number of identified tau leptons with the predicted number of jets faking taus obtained from our jet to tau fake rate. Backgrounds such as $Z \rightarrow \tau\tau$, WW, electrons faking taus and $t\bar{t}$ have been subtracted from the number of observed events because these events contain true taus or sources of fake taus other than jets.Here below are listed the conclusions:

- In all jet bins, the results in the ratio of predicted to observed taus are compatible with 1; it means that the jet to tau fake rate defined for the QCD samples can be applied to the regions dominated by W+jets events.

- Half the difference observed between the largest and the smallest ratio is taken as a systematic error for each jet multiplicity:

$\tau^i ID$	e+$\not\!\!E_T$+1 jet	e+$\not\!\!E_T$+2 jets	e+$\not\!\!E_T$+ $\geq$3 jets
1 complete τ ID	61/72 = 0.85±0.12	28/24 = 1.17±0.20	-
1 denom. τ w M<1.8 GeV	389/392 = 0.99±0.06	137/144 = 0.95±0.08	57/48 = 1.19±0.14
1 denom. τ w cal.iso<0.1	431/420 = 1.03±0.05	189/195 = 0.07±0.07	69/66 = 1.05±0.12
1 denom. τ w N π^0 iso=0	673/642 = 1.05±0.04	266/269 = 0.99±0.06	95/94 = 1.01±1.03
1 denom. τ w N trk iso = 0	191/217 = 0.88±0.07	80/81 = 0.99±0.11	31/26 = 1.19±0.20
1 τ ID w cal.iso>0.1	68/64 = 1.06±0.12	17/19 = 0.89±0.23	8.1/5.1 = 1.59±0.44
1 τ ID w M>1.8 GeV	143/127 = 1.13±0.09	55/53 = 1.04±0.14	18/17 = 1.06±0.24
1 τ ID w N trk iso>0	44/51 = 0.86±0.14	27/24 = 1.12±0.20	11.1/7.1 = 1.56±0.38

Table 3.12: Predictions of number of jet to tau fakes with the standard Tau ID (upper line) and 7 selections of denominator taus failing to pass the tau selection

- 1 jet: (1.13-0.85)/2 = 0.14
- 2 jets: (1.17-0.89)/2 = 0.16
- $\geq$3 jets: (1.59-1.01) = 0.29

These systematic errors are then added quadratically with the systematic errors of the jet to tau fake rate applicable to the non W+jets QCD backgrounds.

Finally, the jet to tau fake rate obtained for the W+jets events is found to be the same as the one for the other QCD events, apart from a larger systematic error. This jet to tau fake rate for W+jets events, expressed as a function FkR(jet E_T, sum E_T)×f(N jets), with FkR defined by the matrix of Fig.3.37, and with updated values for the function f(N jets), namely:

- f(1 jet)=1.06±0.15
- f(2 jets)=0.98±0.19
- f($\geq$3 jets)=0.87±0.28

3.8.7 Conclusion

A method has been established to estimate the jet to tau fake rate in the dijet events, the events with a high jet multiplicity, and the W+jets events. A different jet to tau fake rate has been defined and can be applied for each of the three following analysis cases:

- Jets faking taus as backgrounds for $Z \rightarrow \tau\tau$ or $H \rightarrow \tau\tau$ signals: The jet to tau fake rate is given by the function of jet E_T plotted in Fig.3.35.
- Jets faking taus as backgrounds for $Z \rightarrow \tau\tau$ + jets signal: The jet to tau fake rate is given by FkR(jet E_T, sum E_T)×f(N jets), where FkR is the matrix reproduced in Fig.3.37 and f(N jets) has the following values:
  - f(1 jet)=1.06±0.06
  - f(2 jets)=0.98±0.10
  - f($\geq$3 jets)=0.87±0.14

- W + jets process where a jet fakes a tau lepton: The jet to tau fake rate is given by FkR(jet E_T, sum E_T)×f(N jets), where f(N jets) has the following values:
  - f(1 jet)=1.06±0.15
  - f(2 jets)=0.98±0.19
  - f(≥3 jets)=0.87±0.28

The later jet to tau fake rate is the one to be applied to predict the number of expected background events due to jets faking taus in the present $t\bar{t} \rightarrow l\tau\nu\nu qq$ analysis. This is given by

$$FkR(jetE_T, sumE_T) \times 0.87 \pm 0.28,$$

where FkR is the matrix of Fig.3.37. This gives a conservative uncertainty of 30%.

3.9 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 3.40, 3.41, 3.42, 3.43 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 3.44, 3.45, 3.46, 3.47 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.

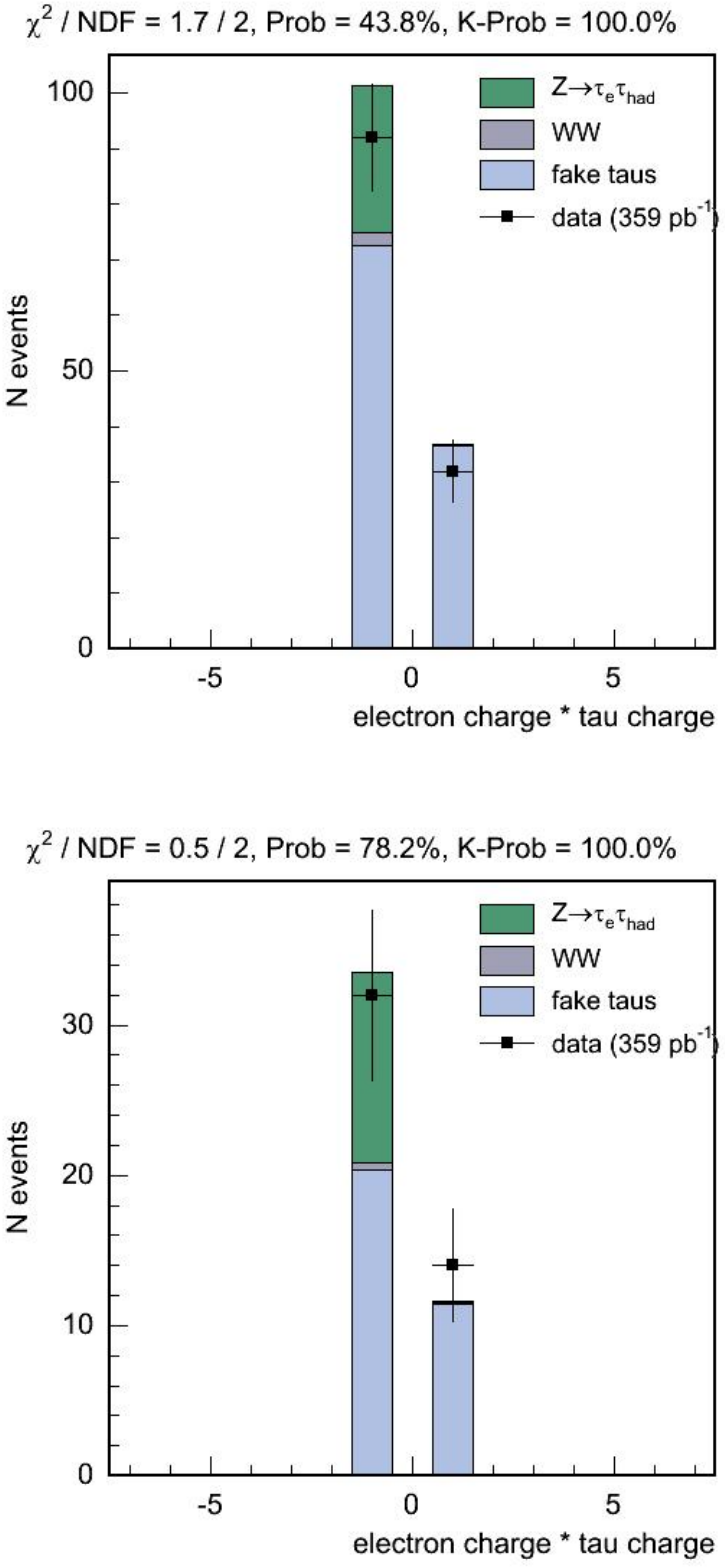


Figure 3.40: Charge product of the electron and the tau for events from the high p_T electron trigger sample, having 1 identified electron, 1 identified tau and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

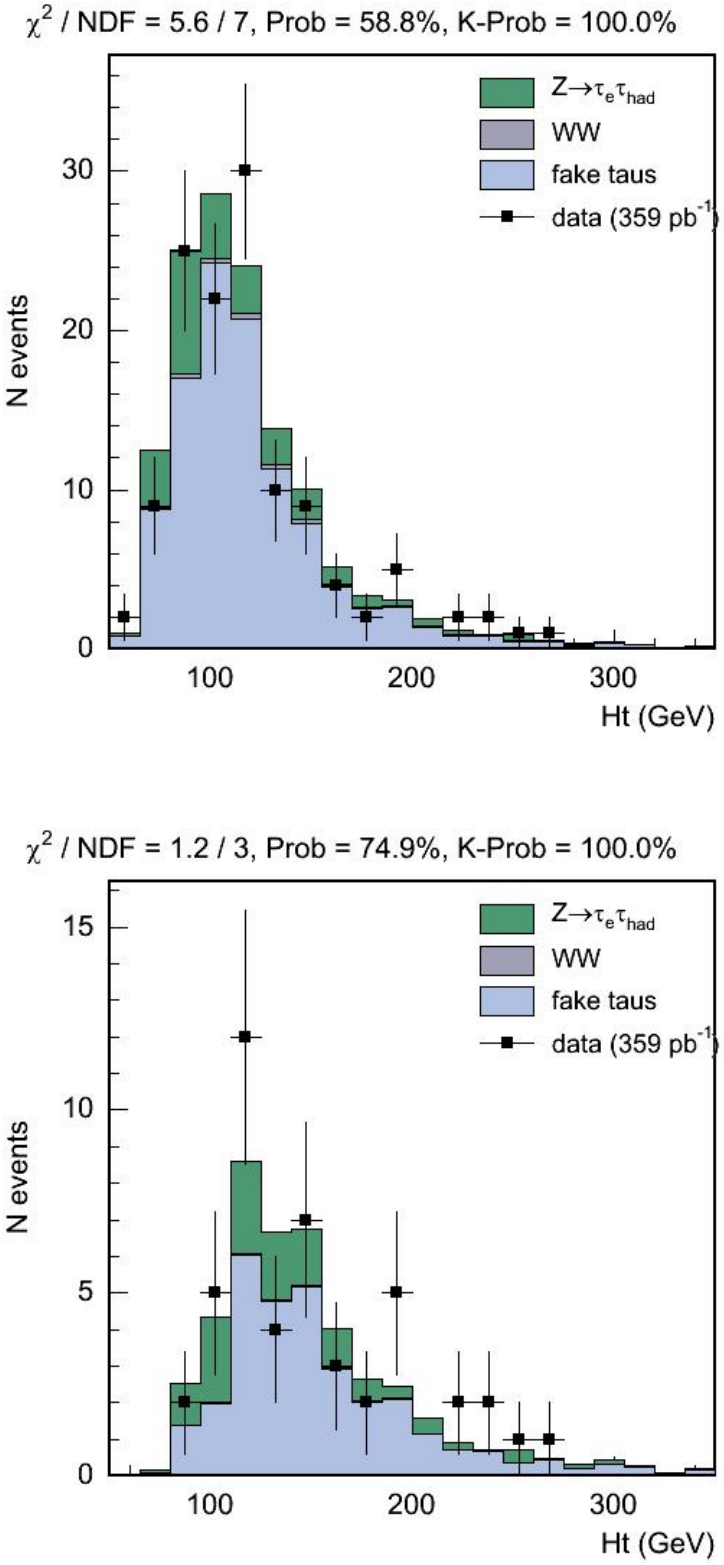


Figure 3.41: Event activity Ht for events from the high p_T electron trigger sample, having 1 identified electron, 1 identified tau and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

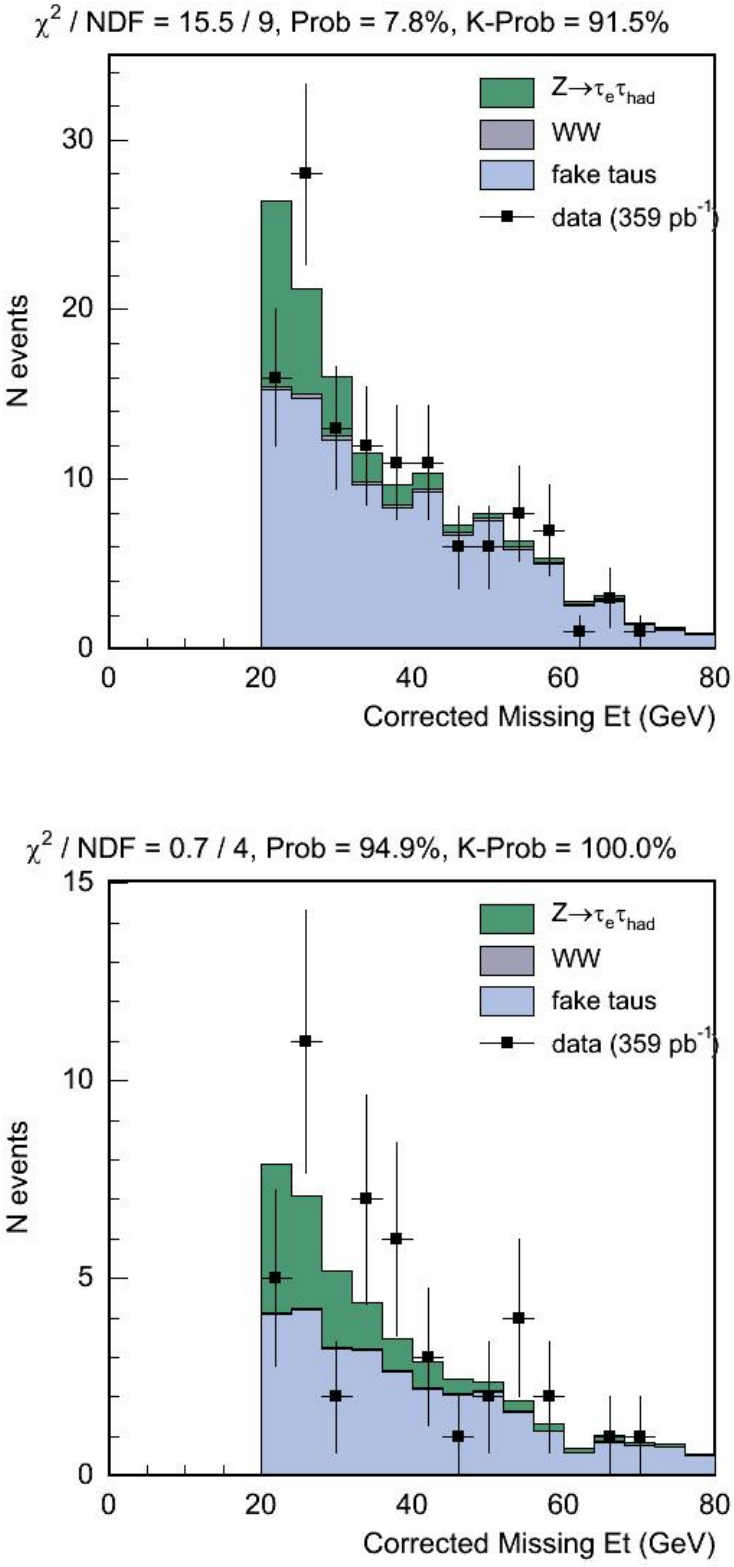


Figure 3.42: Event transverse Missing Energy for events from the high p_T electron trigger sample, having 1 identified μ, 1 identified τ and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

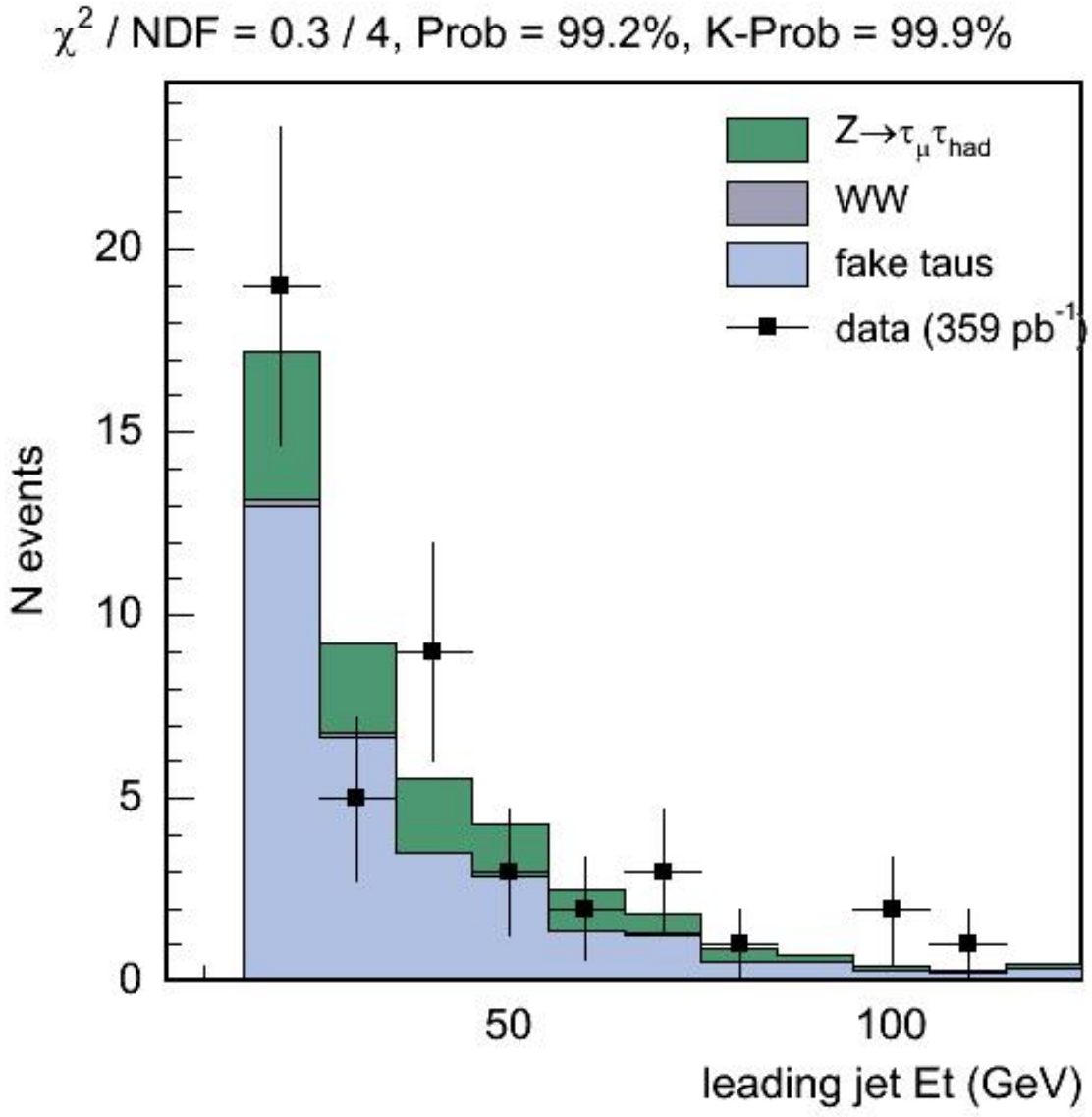


Figure 3.43: Leading jet E_T for events from the high p_T electron trigger sample, having 1 identified μ, 1 identified τ and MET>20 GeV and at least 1 other jet.

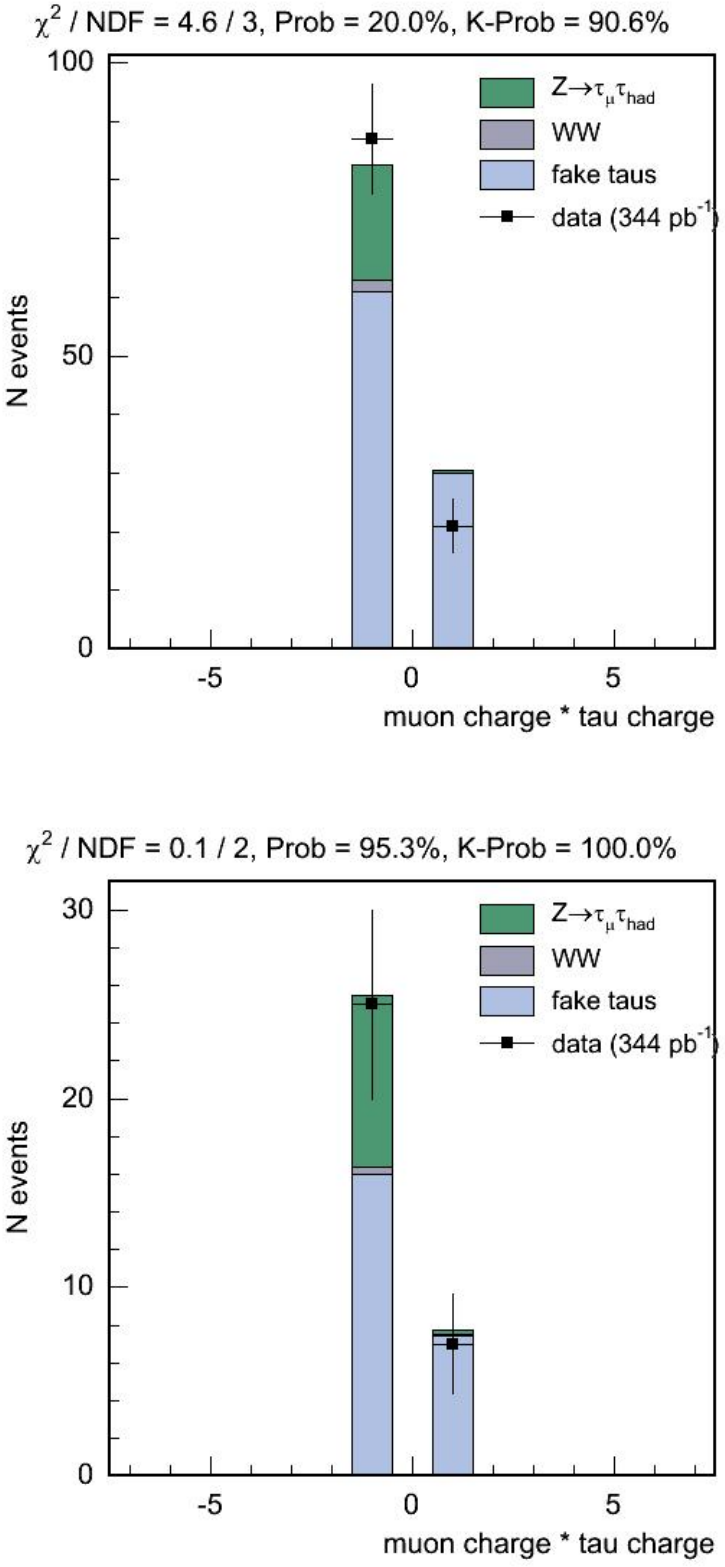


Figure 3.44: Charge product of the muon and the tau for events from the high p_T muon trigger sample, having 1 identified muon, 1 identified tau and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

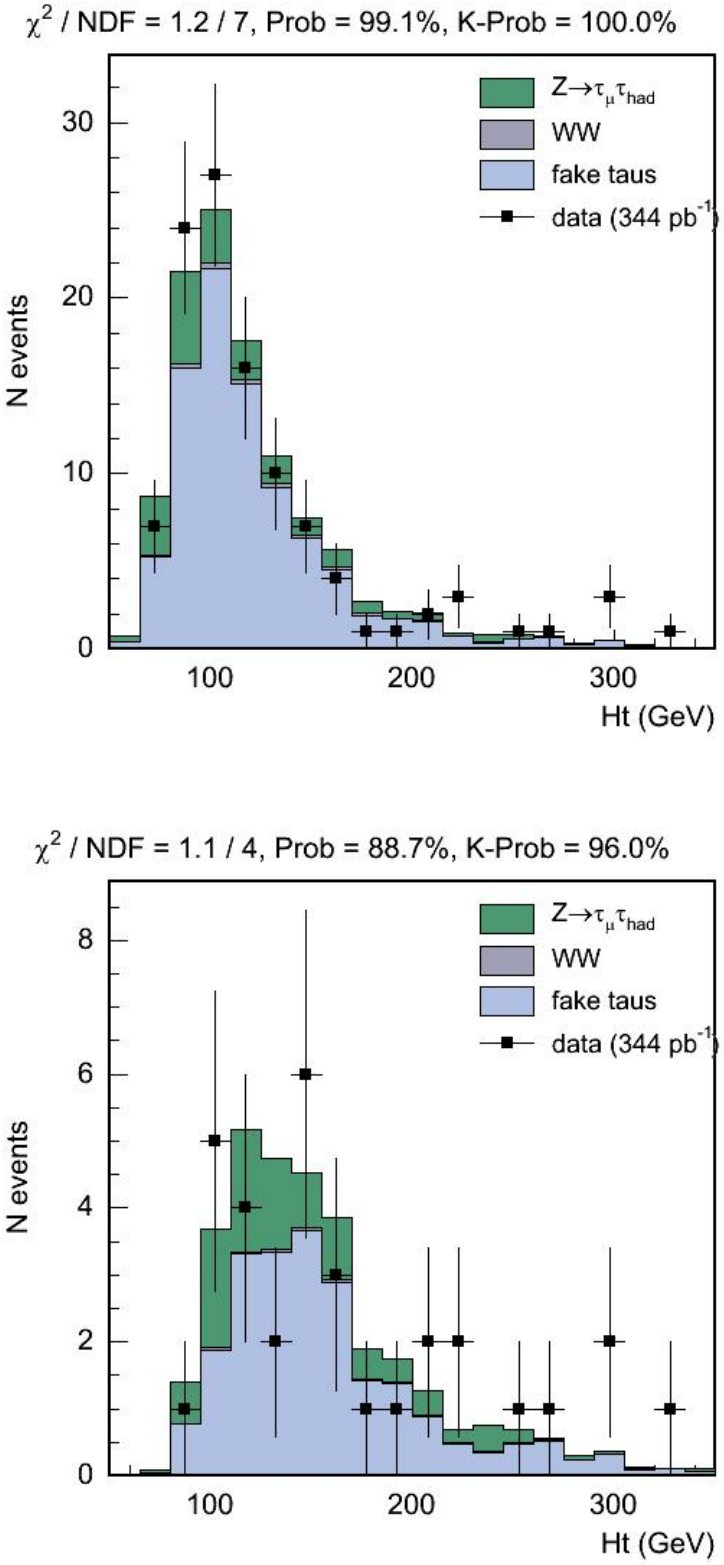


Figure 3.45: Event activity Ht for events from the high p_T muon trigger sample, having 1 identified muon, 1 identified tau and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

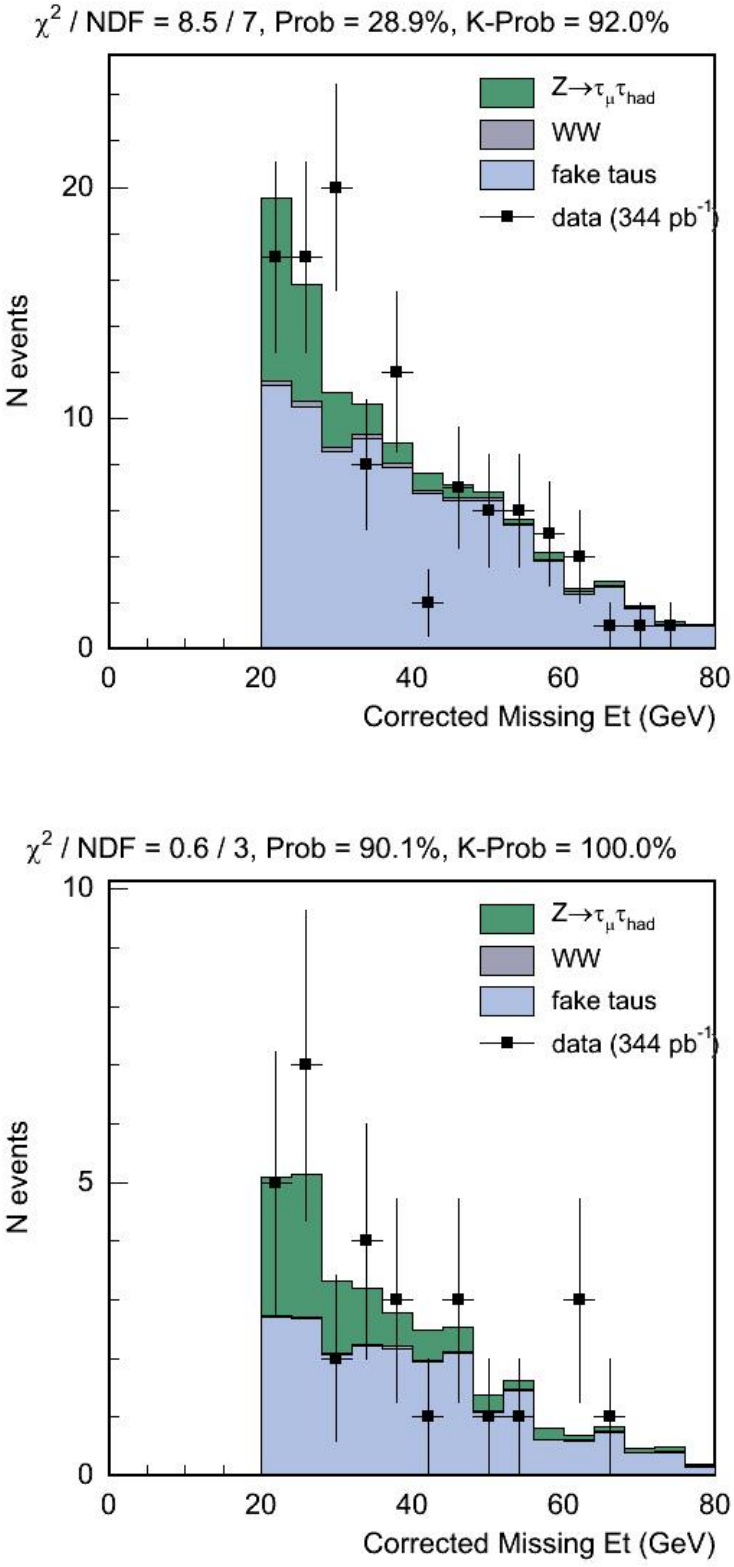


Figure 3.46: Event transverse Missing Energy for events from the high p_T muon trigger sample, having 1 identified μ, 1 identified τ and MET>20 GeV. Top : ≥ 0 jet control region ; bottom : ≥ 1 jet.

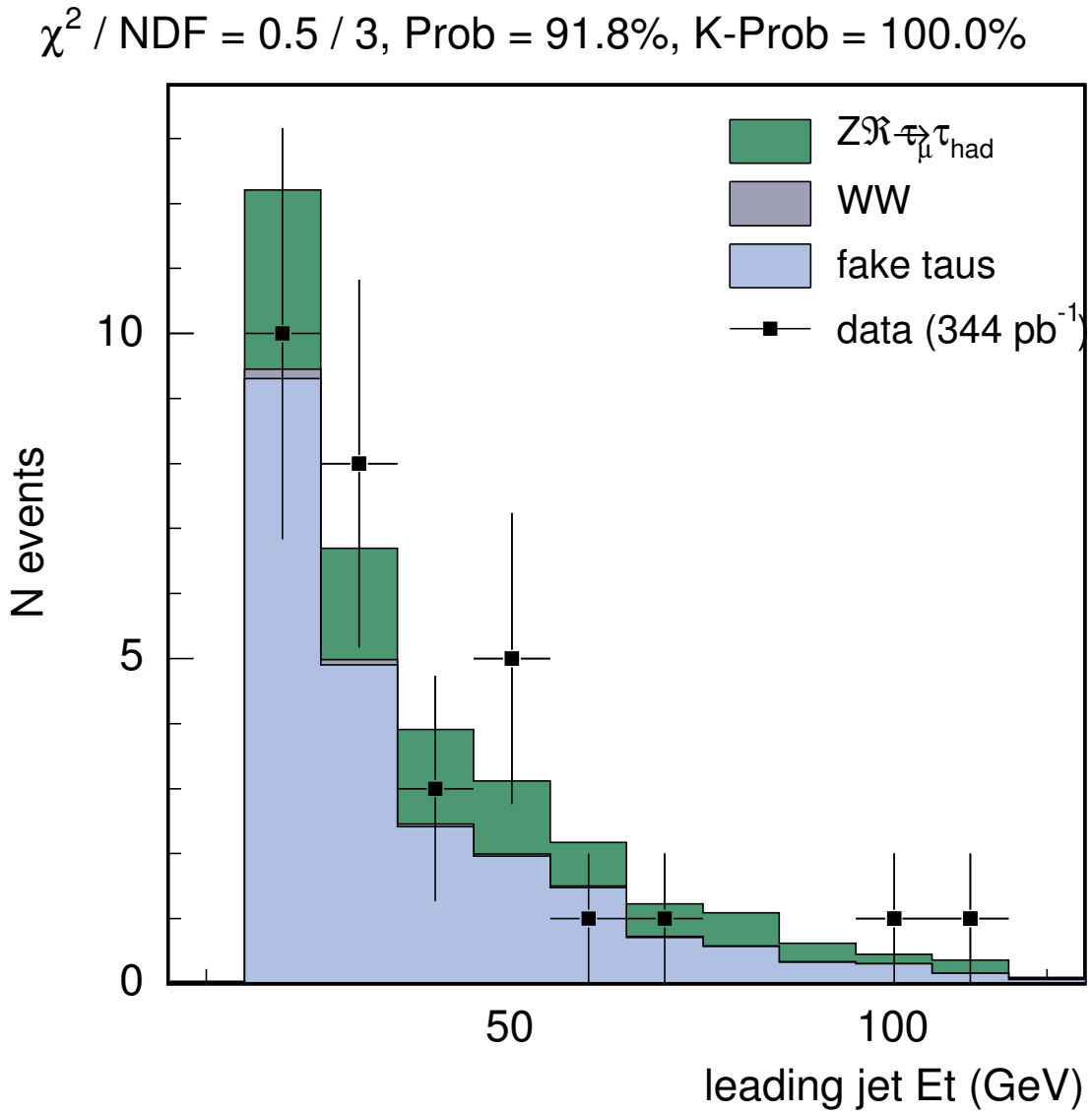


Figure 3.47: Leading jet E_T for events from the high p_T muon trigger sample, having 1 identified μ, 1 identified τ and MET>20 GeV and at least 1 other jet.

Chapter 4

Sensitivities and final results

4.1 Systematic uncertainties

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

The systematic error of 25% taken for the electron faking taus background has been described in the note [23].

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 sections 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 section 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 [53] 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

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

uncertainties on the signal acceptance are gathered in Table 4.1.

4.2 Sensitivities and results with 350 pb^{-1}

4.2.1 350 pb^{-1} analysis expectations and sensitivity

In the table 4.2, are summarized the predicted signal and background events. For the signal, the top pair production cross-section is assumed to be the central value given by the last CDF combined result : 7.3 pb. This result assuming that the top mass is 175 GeV, this assumption is made here too.

The background coming from electrons faking hadronic taus is the result of the folding of the fake rate cited in the subsection 3.7.2.1. In the 350 pb^{-1} analysis, this was folded both in the electron and muon sample, although this contribution is expected to come primarily from $Z \rightarrow ee$+jets events. For the 1 fb^{-1} result presented here after, the Z origin will be clearly assumed and the electron to tau fake rate will only be applied to the electron channel.

The "statistical" errors quoted in table 4.2 correspond to the error due to the limited statistics of the Monte Carlo sample for Z, WW backgrounds and for the signal. As the backgrounds based on fake rates go, the "statistical error" corresponds to the limited size of the denominator sample on which the fake rate is applied. In the case of the jet to tau fake rate, around 10 events are found in the electron data sample before the "numerator" tau identification cuts are applied, which gives a 30% ($1/\sqrt{10}$) "statistical" uncertainty. The equality with the systematic uncertainty associated with the jet to tau fake rate is accidental. These "statistical" errors are all uncorrelated and thus add in quadrature to get the total "statistical" error. Note that this so-called "statistical" error is actually part of the total systematic error of the measurement.

The quoted systematic errors corespond to the systematic uncertainties associated to the estimation methods. These are 30% for the jet to tau fake rate, 25% for the electron to tau fake rate, 30% for Z and WW and 11.7% for the signal. These add in quadrature between different backgrounds to get the total systematic error of the background estimate. However, in order to get the combined electron plus muon channels estimate, these systematic errors have to be symply added (not in quadrature because these are fully correlated) for each background type.

From the expected numbers of background and signal events, we can evaluate the sensitivity of this analysis to the observation of the signal. The probability p_{exp} for the background alone (2.75 ± 0.75) to fluctuate to a value equal to or greater than the first integer greater than the expected sum of signal and background events (2.75+2.24=5.0) is **16%**. This expected p-value is obtained by the convolution of a Poisson distribution of mean $B = 2.75$ and a gaussian function of width $\sigma = 0.75$:

$$1 - p_{exp}(B, S+B = 5, \sigma) = \frac{\int_0^\infty dx \times e^{-((x-B)/\sigma)^2/2} \times e^{-x} \sum_{n=0}^{4} x^n/n!}{\int_0^\infty e^{-((y-B)/\sigma)^2/2} dy}$$

With an integrated luminosity of 350 pb^{-1}, this analysis is thus not sensitive to a first 3 σ observation of the signal.

Figures 4.1 and 4.2 show the behaviour of the probability $1 - p_{exp}(B, 5, \sigma)$ for the background B to fluctuate to 5 events (or more) with variations of the quoted numbers of background events B and variations of the systematic error on the background σ.

	Electron + tau 359 pb^{-1}	Muon + tau 344 pb^{-1}
jet$\rightarrow \tau$ fakes	$0.91 \pm 0.27_{stat.} \pm 0.29_{sys.}$	$0.92 \pm 0.28_{stat} \pm 0.29_{sys.}$
e$\rightarrow \tau$ fakes	$0.10 \pm 0.035_{stat.} \pm 0.025_{sys.}$	$0.05 \pm 0.025_{stat.} \pm 0.012_{sys.}$
$Z \rightarrow \tau_{\rightarrow l}\tau_{\rightarrow had}$+jets	$0.39 \pm 0.04_{stat.} \pm 0.13_{sys.}$	$0.32 \pm 0.03_{stat.} \pm 0.10_{sys.}$
$WW \rightarrow \tau\nu_\tau l\nu_l + jets$	$0.034 \pm 0.004_{stat.} \pm 0.011_{sys.}$	$0.027 \pm 0.003_{stat.} \pm 0.008_{sys.}$
Total Background	**1.43**$\pm 0.28_{stat.} \pm 0.32_{sys.}$	**1.32**$\pm 0.28_{stat.} \pm 0.31_{sys.}$
	2.75$\pm$0.75	
SIGNAL ($t\bar{t} \rightarrow l + \tau$) (assuming $\sigma(t\bar{t}) = 7.3$ pb and $m_{top} = 175$ GeV)	$1.32 \pm 0.05_{stat.} \pm 0.15_{sys.}$	$0.92 \pm 0.05_{stat.} \pm 0.11_{sys.}$
	2.24 ± 0.26	

Table 4.2: *Summary of predicted signal and backgrounds in 350 pb^{-1}*

These show a remarkable robustness of a result expressed in terms of a probability for the observed events to be due to background fluctuations only.

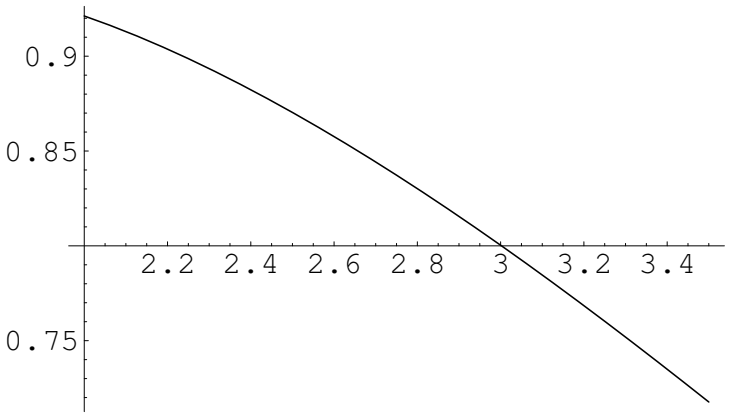


Figure 4.1: Dependence with the expected number of background events of the probability for the background alone to fluctuate to 5 events or more, with a fixed systematic error of 0.75.

4.2.2 Observation in data: 350 pb^{-1} result

Opening the blind box in the 350 pb^{-1} data, **5** events compatible with tau dilepton events are observed, summarized in the table 4.3. The first two events, corresponding to runs of the first 195pb^{-1} of data, are the exact same two events that were observed by the older 195pb^{-1} analysis [17].

This is in total agreement with the expectations and so well inside the sensitivity of this analysis.

These five events were scrutinized. The figures 4.3 and 4.4 present the distributions for six kinematical variables, H_T, $\not\!\!E_T$, τ-jet E_T, lepton (electron or muon) E_T and the

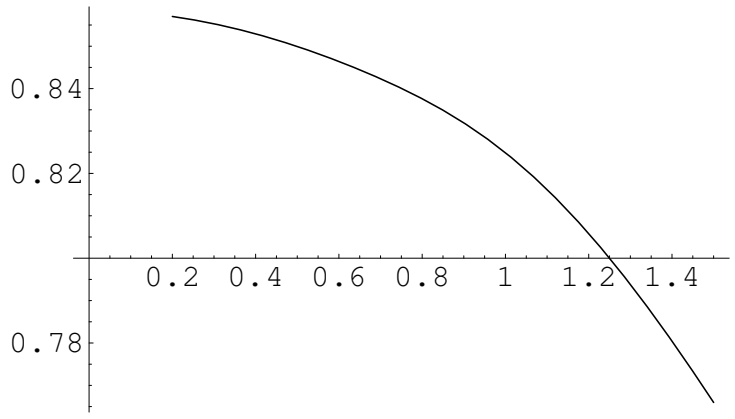


Figure 4.2: Dependence with the systematic error on the background of the probability for the background alone to fluctuate to 5 events or more, with a fixed expected mean number of 2.75 background events.

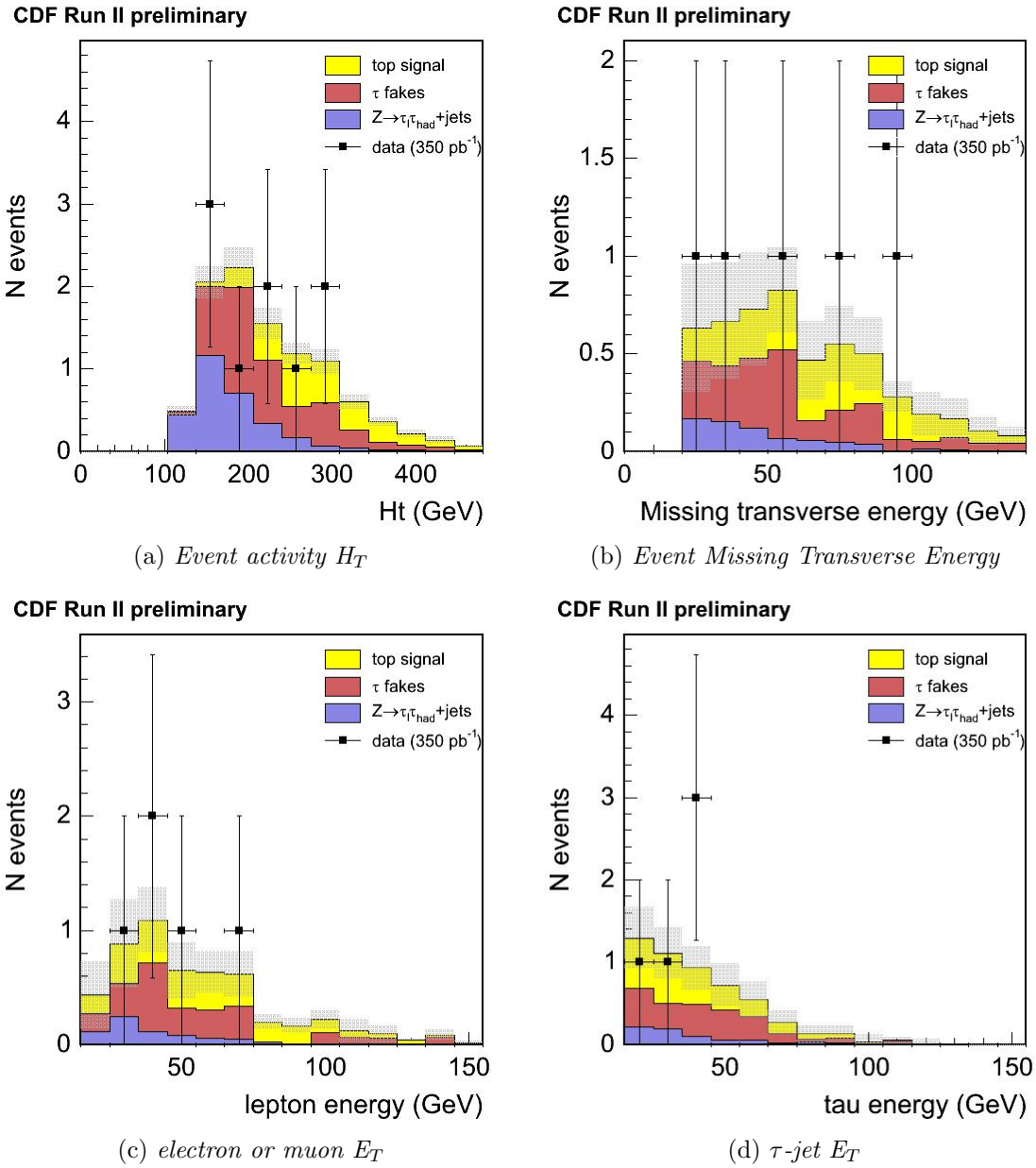
Run Number	Event Number	Channel
151434	158200	electron+τ
167299	2376337	electron+τ
183557	543561	CMX+τ
185379	300012	CMUP+τ
185971	1331561	CMUP+τ

Table 4.3: *List of the 5 observed events in the 350 pb^{-1} analysis.*

two leading jet E_T. 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. With this small statistics, the distributions don't show any obvious disagreement.

Also the detailed kinematical characteristics of the 5 observed events are gathered in the Table 4.4.

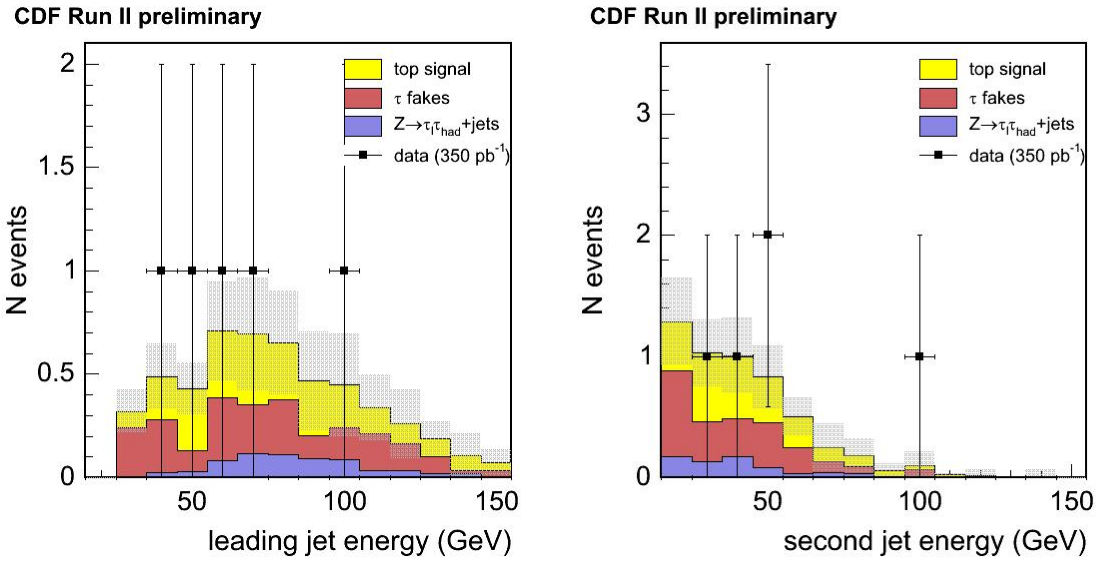
The result is given in terms of probability for the 5 observed events to be due to background only, probability estimated by folding a Poisson distribution of mean 2.75 with a gaussian distribution of standard deviation 0.75, to be : **16%**. In other terms, this constitutes a 67% ($\sim$ 1 sigma) evidence of the existence ot the tau dilepton channel in $t\bar{t}$ production.



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

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

Figure 4.3: Comparison of the 5 observed events distributions with the standard model expectations



(a) *leading jet* E_T (b) *Second jet* E_T

Figure 4.4: Comparison of the 5 observed events distributions with the standard model expectations

CDF RunII preliminary	Event 1	Event 2	Event 3	Event 4	Event 5
Type	electron	electron	μ (CMX)	μ (CMUP)	μ (CMUP)
Missing E_T (GeV)	71.9	53.5	99.8	29.2	35.9
Missing E_T Φ (rad)	-1.52	-1.42	1.46	0.89	0.76
Tau E_T (GeV)	19.7	39.6	43.7	44.4	30.9
Tau pseudorapidity	-0.43	-0.95	0.02	0.62	-0.35
Tau Φ (rad)	3.80	3.75	2.05	3.10	5.60
Tau z_0 (cm)	-51.1	3.7	-22.8	7.9	41.3
Tau N tracks	3	1	3	3	3
Lepton E_T (GeV)	67.7	41.9	41.6	29.4	53.1
Lepton pseudorapidity	-0.79	-0.27	-0.80	-0.23	0.07
Lepton Φ (rad)	1.71	1.06	5.99	1.20	2.93
Lepton z0 (cm)	-51.3	-3.35	-23.0	9.4	41.3
Number of jets	2	3	2	4	2
1st jet E_T (GeV)	35.3	69.3	61.9	101.8	48.6
1st jet pseudorapidity	-0.74	-0.25	-0.61	1.93	0.38
1st jet Φ (rad)	1.25	2.89	0.82	5.64	1.06
2nd jet E_T (GeV)	33.2	39.8	55.0	96.2	47.7
2nd jet pseudorapidity	0.52	-1.49	-1.20	0.41	0.95
2nd jet Φ (rad)	5.00	0.22	2.65	2.29	4.44
Ht (GeV)	228	244	302	301	216

Table 4.4: *Kinematical characteristics of the 5 observed events in 350 pb^{-1}*

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

4.3.1 1 fb^{-1} analysis expectations and sensitivity

4.3.1.1 The loose likelihood selection

In the table 4.6, 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 4.6), 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.

4.3.1.2 The tight likelihood selection

In the table 4.7, 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 4.5, we gather the different p-values that can be got, depending on the number of events really observed in data, with the corresponding probability for it

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 4.5: *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.*

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

	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 4.6: *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 4.7: *Summary of predicted signal and backgrounds in 1 fb$-$1 with the tight likelihood cut Log($L0_{10}$) > 0.*

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

4.3.2.1 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 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:

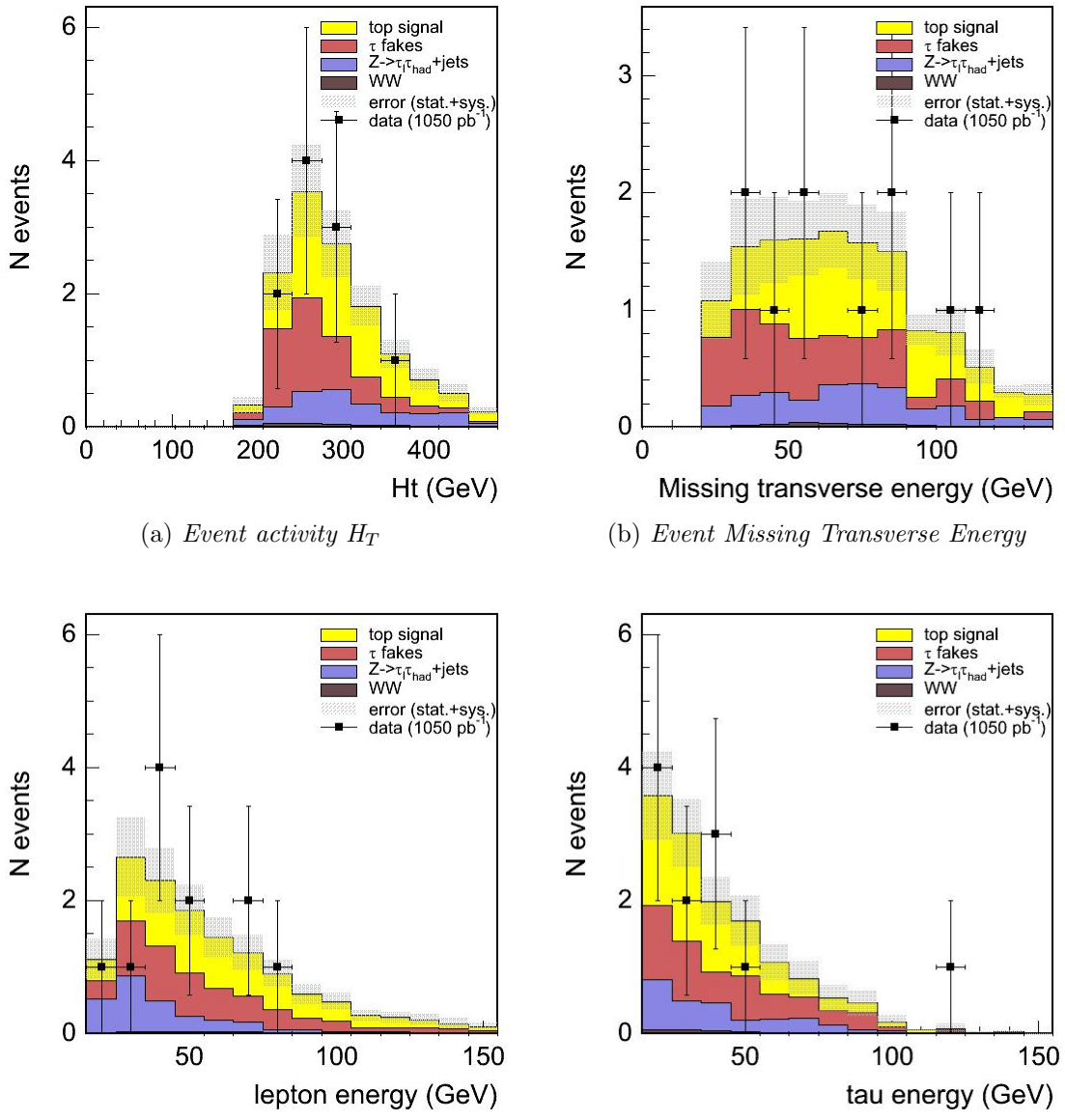
$$\mathbf{r_\tau < 1.5} \text{ at } 95\% \text{ CL.}$$

The eleven events were scrutinized. The figures 4.5 and 4.6 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.

4.3.2.2 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 4.5) that the measured p-value is **7.6%**, and that we got a **1.17 sigma evidence** for the signal.



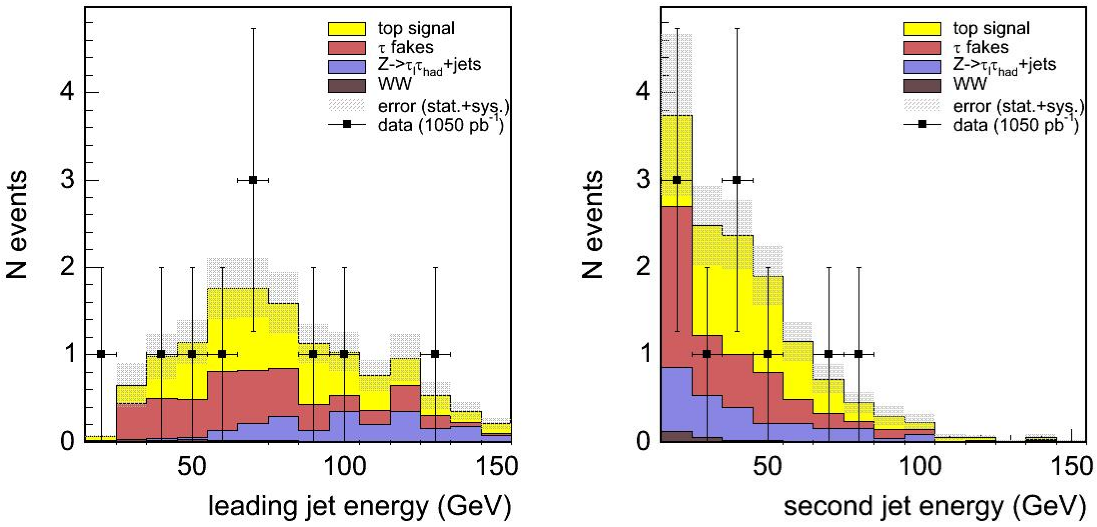
(a) *Event activity* H_T

(b) *Event Missing Transverse Energy*

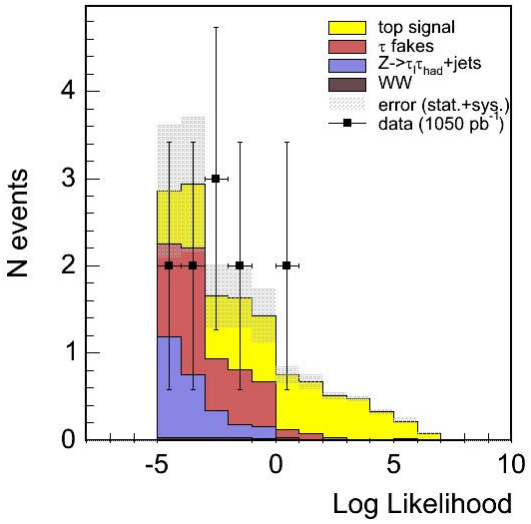
(c) *electron or muon* E_T

(d) τ*-jet* E_T

Figure 4.5: Comparison of the 5 observed events distributions with the standard model expectations.



(a) *leading jet* E_T (b) *Second jet* E_T



(c) *L0 likelihood*

Figure 4.6: Comparison of the 11 observed events distributions with the standard model expectations.

Chapter 5

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.

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.

Bibliography

[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] Fermilab Beam Division, Operations Rookie Books, http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html

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

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

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

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

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

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

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

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

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

[13] 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)

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

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

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

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

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

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

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

[21] 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

[22] CDF note 8148, Feb.23 2006

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

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

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

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

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

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

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

[30] 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.

[31] 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).

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

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

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

[35] 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)

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

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

[38] 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).

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

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

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

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

[43] 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

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

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

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

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

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

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

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

[51] 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)

[52] 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)

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

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

List of Figures