Measurement of the $t\bar{t}$ production cross section using heavy flavor tags in $W + \geq 3$ jet events in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV

A thesis presented

by

Fotios K. Ptochos

to

The Department of Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Physics

Harvard University

Cambridge, Massachussetts

June, 1998

To my mother and father

στη Μάρω

στη Ρίκα

and to the memory of

an exceptional person, friend and colleague

Γιώργου Μιχαήλ

Acknowledgments

All these years in graduate school would not be a joy if I was not surrounded by so many great people and friends. Their constant support and enthusiasm all these years was invaluable. I am happy, I finally have the opportunity to acknowledge them here.

First and foremost I would like to thank Melissa Franklin and Paolo Giromini.

I was fortunate enough to have Melissa as my advisor. It is hard to find the right words to thank her for the support, guidance, freedom, friendship and her unbelievable endurance to my stubbornness throughout all these years. A simple thank you Melissa is not and will never be enough. Paolo was the person who taught a tremendous amount of physics, how to be patient and try to understand a plot, attack a problem to its final solution and how not to confuse San Antonio with the pig. His stories and sense of humor during the long walks in the prairie were always at the right time and the great escape from the stinky trailers. His ability to understand problems and guide through them are equally amazing to his difficult character.

Sanjib Mishra for his friendship and ingenuity. His endless enthusiasm for particle physics was the light in some of the darkest periods of graduate school. George Brandenburg has helped me out in many occasions both as director of HEPL and as advisor in the years I was climbing and hanging from the CMX. Many thanks to Phil Schlabach for taking responsibility in the troublesome years of CMX and for the unforgettable search for "pig tails" in all the auto-stores around Fermilab. Many thanks go also to Weiming for numerous discussions and suggestions and his efforts to explain me the secrets of SECVTX and the top analysis. I would like to thank Michael Schmitt who was kind and patient enough to go through the hundred of pages without falling asleep (although Paolo will never believe this) and for trying to

make sense out of my incomprehensible sentences.

My life and sanity would not be the same without the presence of three great great persons. George Archontis, Tom Baumann and Maria Spiropulu were always there in the bad and the good moments and they could always find a way to turn things around at the drop of a hat. Their friendship is invaluable and I hope it will remain in the future. Andrew Gordon "IT IS YOU !!!", simply a genius, and Peter Hurst because he is always the great Peter. Maria Tzamarioudaki was willing to listen to my complaints, offer a happy voice from the other side of the phone line, a hefty loan and the honey. Vaia Papadimitriou for her interest on my work and the endless and amusing cat stories. Young-Kee Kim offered great hospitality and great company down at the pit. Alex Samuel and Robin $\epsilon\lambda\iota\acute{\alpha}$ Coxe were great for being around HEPL, asking questions and inspiring with their determination.

Many thanks go to the Italian clan, Giorgio Apollinari, Stefano Miscetti, Maura Barone, the "grepping" Andrea Parri, Emilio Meschi, Fabio Happacher, Stefano Belforte, Marco Cordelli, Irene Fiori, that made a Greek feel at home, happy and comfortable among them. Many discussions with them help me clarify my confusions on many things and they were always happy to spend their time to teach me something new or talk to me and encourage me. Many thanks go also to Dee and Steve Hahn for helping out in numerous occasions and organizing and inviting me to various parties. I would also like to thank JJ for being patient with my CPU tricks and for being always very helpful with my stupid problems.

Ashutosh Kotwal was a great roommate and friend when I was trying to discover the Midwest and the buffaloes. The rest of my fellow Harvardianos from the early to the current generation, Colin Jessop, Jaco Konigsberg, Colin Gay, David Kestenbaum, Rowan Hamilton, Thomas Dignan, Ted Liu, Robyn and Jun, Michael Schmitt, that

made the life at Fermilab and in Cambridge pleasurable and taught me a great deal of things. Carter for being a smiling person and Werner for keeping me company during these last days in Cambridge. Ed Frank, the fellow colleague across my office on the fourth floor of HEPL, was always there to tell me stories from his graduate experiment and share a word of wisdom. Vera Malkjovic for teaching me cognitive psychology and being with me during the hard first years.

HEPL was always a pleasant place to be and work, thank to the great and fun staff. Guy Sciacca was always more than helpful to assist on any material he could and he was always ready to share a humorous point of view on any subject. John Oliver, Jack O'kane and Sarah Harder were always very helpful for anything I had problems with and the same with Mary Platt in the early years. Down in the basement, Jim McIllaney, Frank Dalyrimple, Eddie Lucia, Rony Gale, Steve Sansoni, were awesome during the period of constructing the CMX and were always happy to help and share a six-pack of Sam Adams. Special thanks go to Rick Haggerty for being such a great person (despite of being Irish) and making me happy with just a "good morning". Definitely I'll miss them all.

My first years of postgraduate career I spent working on the ICARUS project at CERN. This was my first real experience with the world of particle physics and detectors. I would like to thank Max Ferro-Luzzi for introducing me to this wonderful world and giving the opportunity and freedom for creative work. While being at CERN, Gianni Carugno, Francesco and Valeria Pietropaolo and Claude Detraz offered many ideas and introduce me more to the mystic world of physics. Their hospitality, friendship and warmth were important ingredients to survive my first long time away from Greece. Following the year at CERN, I spent my first two years of graduate school working on warm liquid calorimeters at Harvard. I would like to thank Steve

Geer for giving me the opportunity to come to Harvard and everything he taught me during that period.

I would like to acknowledge and thank the efforts of the CDF collaboration and Fermilab staff for making this work possible. Special thanks go to the CDF people who develop the tools and machinery that went into this thesis. Special thanks to the CDF top group for their continuous and enthusiastic support throughout all this analysis.

On the other side of the Atlantic my long lasting friends *Αλεξάνδρα*, *Ακις*, *Γιώργος*, Sergei, *Χρύσα* were always a constant source of encouragement and support. During all these years of me being away for so long, they could always find a way to get together with me and make me feel I was never away. They would always invent something new to fill the time lapses and renew the memories and friendship. Many thanks for keeping company to my parents and make them forget their loneliness. They are simply great. Many many many thanks to *Αλεξάνδρα* for all her phone calls up to the very last moment. They were so important.

Many thanks would have been addressed to my dear friend George Michail. His tragic death changed many many things in my life. His gentle and great personality, his warmth, his laughter, his passion and talent for physics and perfection, his happy face will never be forgotten. The minimum I can do is to dedicate this thesis to his honor and memory and to his mother *Ρίκα*. Her courage and strength in this most difficult time are phenomenal. She is a wonderful person and she could always find the strength and humor to encourage me through my little problems. Many many many thanks to *Ρίκα Μιχαήλ*.

Finally and most importantly, I am thankful to my wonderful parents, *Μαρίκα* and *Κώστα*, and my brother *Τάσο*, for their unconditional support, encouragement

and love all these years. Without them this would not be possible.

Measurement of the $t\bar{t}$ production cross section using heavy flavor tags in $W + \geq 3$ jet events in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV

Fotios K. Ptochos

Advisor: Melissa E. B. Franklin

Abstract

This thesis presents the measurement of the $t\bar{t}$ production cross section using 110 $pb^{-}1$ of $p\overline{p}$ collisions at $\sqrt{s}$=1.8 TeV collected using the Collider Detector at Fermilab (CDF). Assuming Standard Model couplings, events consistent with containing a W boson produced in association with at least three jets are used for the search of events originating from $t\bar{t} \rightarrow W^{+}bW^{-}\overline{b}$ decays. The presence of high momentum electrons and muons associated with large energy imbalance transverse to the beam direction are the characteristic signatures used to identify events with $W \rightarrow \ell + \nu$ decays.

In order to further reduce the QCD background contribution from W production in association with jets, three algorithms are used to determine the presence of a heavy flavor b-quark jet in the event. Two of the algorithms use the very fine position resolution of the silicon vertex detector in order to identify either displaced vertices or displaced tracks contained inside a jet. The presence of b-quark in the event is also inferred by the identification of a soft lepton from its semileptonic decay ($b \rightarrow \ell\nu X$ or $b \rightarrow c \rightarrow \ell\nu X$). This is the basic ingredient of the third algorithm used in the analysis.

The background to $t\bar{t}$ signal, consists of $Wb\bar{b}$, $Wc\bar{c}$, Wc, single top, misidentified Z's produced in association with heavy flavor jets, $Z \rightarrow \tau^+\tau^-$ and diboson (WW, WZ, ZZ) production. The contribution of this background is calculated with a combination of data and Monte Carlo simulated events. Non-heavy flavor jets misidentified as b-quarks consist a major source of background and its contribution is determined directly from the data.

The $W+ \geq 3$ jet sample consists of 252 events before b-quark identification. The algorithm based on the presence of a displaced secondary vertex in a jet, identifies 29 events containing a b-quark jet with a background expectation of 8.12 ± 0.99 events yielding a a $t\bar{t}$ cross of $\sigma_{t\bar{t}} = 4.83 \pm 1.54$ pb using acceptances for a top quark mass of 175 GeV/c^2. The algorithm based on the presence of displaced tracks in a jet, identifies 41 candidate events with a background contribution of 11.33 ± 1.36 events, yielding a $t\bar{t}$ cross section of $\sigma_{t\bar{t}} = 7.33 \pm 2.10$ pb. Finally, 25 events are found consistent with containing jets from b-quark semileptonic decays with expected background of 13.22 ± 1.22 events, resulting to a $t\bar{t}$ cross section of $\sigma_{t\bar{t}} = 8.37 \pm 3.98$ pb. Based on a kinematic fit of events containing b-quark jets, the top mass is measured to be M_{top} = 175.9 GeV/c^2. For the measured mass the $t\bar{t}$ cross sections for all three b-quark identification algorithms are in good agreement with the theoretical calculations which are in the range of 4.75 pb to 5.5 pb for a top quark mass of M_{top}=175 GeV/c^2.

Contents

List of Figures

List of Tables

Chapter 1

The Standard Model and the top quark

To date, the best understanding of the fundamental constituents of matter and the interactions with each other at small distance scales are expressed in a theory called the *Standard Model* of particle physics. Predictions made with this theoretical model have been confirmed by many experimental measurements and the agreement between theory and experiment has been remarkable.

The top quark was one of the last missing particles which existence has been predicted by the theory. In 1995, the CDF and D0 collaborations reported the discovery of the top quark [2, 3] using data collected in $p\bar{p}$ collisions at center of mass energy of 1.8 TeV at the Fermilab Tevatron. The top quark discovery completes the set of twelve particles believed to be the fundamental constituents of matter while it offers one more triumph for the Standard Model. Within the context of the Standard Model, the top quark decays almost exclusively into a bottom quark and a W boson. The W boson can decay into a charged lepton and a neutrino, or it can decay into

two quarks. At the Tevatron, top quarks are produced mainly in pairs and therefore $t\bar{t}$ events contain always a pair of W bosons and a pair of b-quarks. The top quark was discovered in events where one of the W bosons produced in the decay of a $t\bar{t}$ pair decays into a charged lepton and a neutrino.

This chapter begins with a brief overview of the Standard Model and continues with the theoretical and experimental arguments provided by the Standard Model for the necessity of the existence of the top quark before its discovery. Expectations on top quark based on precision electroweak measurements are also presented. After the existence of the top quark is motivated, a discussion of the top production mechanisms at the Tevatron and its decay modes is presented. An overview of the current results establishing the existence of the top quark are presented in the last sections of the chapter.

1.1 The Standard Model

1.1.1 Forces and matter

The Standard Model of physics uses quantum field theories to describe three of the four forces by which the fundamental particles interact. The most familiar of these forces, the gravity, is not included. A consistent quantum theory of gravity does not yet exist but it is by far the weakest of the four and has negligible effect at the subatomic level. The other forces are the electro-magnetic, strong and weak forces.

Maxwell was the first to formulate the electro-magnetic force. All electrically charged particles act as sources of electro-magnetic fields, which hold electrons and nuclei together to form atoms and atoms together in larger structures. The application of relativistic quantum mechanics to the electro-magnetic field gave rise to the

Quantum Electrodynamics, or QED. This describes the interactions of electrically charged particles, mediated by a massless boson, the photon. It was the original quantum field theory.

The weak force is responsible for nuclear beta decay. It affects all particles and acts on short-range, being mediated by the massive vector bosons $W^{\pm}$ and Z^0. In the 1960's Glashow, Salam and Weinberg showed that the electro-magnetic and weak forces were in fact two aspects of a single electroweak theory [7]. This symmetry appears at high energies where the masses of the $W^{\pm}$ and Z^0 are negligible. The Higgs mechanism gives the bosons their masses and therefore breaks the symmetry at lower energies. It also predicts the existence of a further neutral scalar particle, the Higgs boson, whose mass is not predicted and which has not yet been observed.

The strong force binds the constituent particles of the nucleons and holds the nucleus together. Particles experiencing the strong force are said to have a color charge. Such particles interact via a massless boson, the gluon. However unlike the photon, which is electrically neutral, the gluons themselves carry color charge and can therefore interact with one another. The field theory which describes the strong interaction is Quantum Chromodynamics, or QCD [6, 7].

The fundamental particles within the Standard Model are divided into two classes: The spin-$\frac{1}{2}$ fermions which make up matter and the spin-1 bosons which mediate the fundamental forces. In addition, as said before, the mechanism of the electroweak symmetry breaking predicts a Higgs boson with spin 0. The Standard Model fermions are shown in Table 1.1. They are classified in two categories, the quarks, which have color charge and fractional electric charge (1/3 or 2/3) and interact via the strong force, and the leptons which do not. Both groups undergo electroweak interactions. Most familiar of the leptons is the electron (e). The up (u) and down (d) quarks

Leptons		Quarks	
Particle	Electric charge	Particle	Electric charge
e	-1	u	$+2/3$
ν_e	0	d	-1/3
μ	-1	c	$+2/3$
ν_μ	0	s	-1/3
τ	-1	t	$+2/3$
ν_τ	0	b	-1/3

Table 1.1: The fundamental particles within the Standard Model. The charge is given in units of e, the magnitude of the charge carried by the electron.

consist the building blocks of the protons and neutrons, which combine with the electrons to make an atom. Quarks, in contrast to leptons, do not appear as free particles. This is a consequence of the structure of the strong force which binds them together. The phenomenon is called confinement. Quarks appear in nature either in pairs called *mesons*, or triplets called *baryons*. The electron is accompanied by another lepton, the electron neutrino (ν_e). The neutrinos are believed to be either massless, or very light and carry no electric or color charge and interact with matter only via the weak force.

Together the electron, the electron neutrino, the u and the d quarks form a generation of particles. There are two more generations of particles, each made up of a charged lepton and its corresponding neutrino and two quarks. Every particle has identical quantum numbers to its partners in the other generations, apart from flavor which characterizes the generation the particle belongs, and mass which increases moving from one generation to the next. The second generation of particles consists

of the muon (μ) and its partner, the muon neutrino ν_μ, the charm (c) and strange (s) quarks. Finally, the third generation consists of the bottom (b) and top (t) quarks, along with the tau (τ) and tau neutrino (ν_τ). Apparently the existence of the top quark was implied by the discovery of its partner, the bottom quark.

The Standard Model does not explain why there are three and only three generations of particles. However, if any further generation existed containing a massless or very light neutrino, it would be observed as part of the width of the Z^0 boson which is attributed to particles which can not be observed (invisible width). Nevertheless, measurements of the Z^0 invisible width at SLC and LEP [81] suggest that there are 2.991 ± 0.016 generations with light neutrinos.

1.1.2 Perturbation theory

In QED, the strength of the interaction between two electrically charged objects depends on the product of their charges (in units of e which is the electron charge) and on the intrinsic strength of the force. This last quantity is

$$\alpha = \frac{e^2}{4\pi}$$

the electro-magnetic coupling constant.

The particles in a quantum field theory interact through the exchange of field quanta which in the case of QED is the photon field. The simplest such interaction is shown in Figure 1.1 in terms of a Feynmann diagram. The Feynmann diagrams represent particular terms in the perturbation series expansion of the transition amplitude in terms of the interaction coupling constant. Using appropriate rules, *the Feynmann rules*, the amplitude, $\mathcal{M}$, of any such diagram can be calculated which

leads to the calculation of the cross section for the specific interaction. A factor of $\sqrt{\alpha}$ is introduced into $\mathcal{M}$ for every vertex where a charged particle interacts with a photon. In other quantum field theories, similar diagrams represent the fundamental interactions, with coupling constants determined by the intrinsic strength of each force. Figure 1.1(b) shows an example of a correction to the simplest case, where a virtual photon disintegrates to an electron-positron pair which then annihilates to form a photon once more. This is an example of a loop diagram. An example of vertex corrections is shown in Figure 1.1(c), where a photon is exchanged between the incoming and outgoing electron.

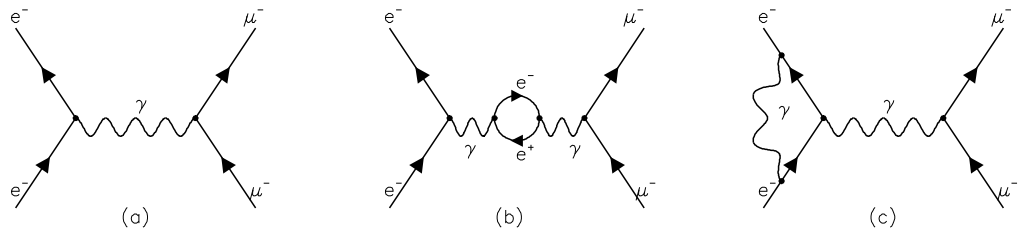

Figure 1.1: Electro-magnetic interaction between an electron and a muon (a) by exchange of a single photon, (b) and (c) with 1 loop corrections. A factor of $\sqrt{\alpha}$ corresponds to every vertex a charged particle interacts with a photon.

The two extra vertices in the diagram of Figure 1.1(b) introduce a factor of α into $\mathcal{M}$ compared to the simplest diagram. In fact, all of the diagrams whose contribution

must be summed to give the total amplitude for the interaction can be arranged to form a power series in α. When the coupling constant is less than 1, as for QED and the electroweak theory and in some circumstances with QCD, each power of α is smaller than the preceding one. Hence, it appears that each term can be considered as a small correction to the series up to that point, which is the basis of perturbation theory.

Unfortunately, the amplitudes calculated for some of the loop diagrams appear to be infinite. To deal with this, one must apply the renormalization theory techniques. In this case, the loop corrections are split into infinite and finite parts. It turns out that for field theories with local gauge invariance, such as those within the Standard Model, the infinite parts of the corrections cancel to all orders in α.

Renormalization involves redefining the parameters of the theory, such as the charge and mass of an electron in QED. For example, to measure the charge of an electron, one would observe its behavior in the presence of an electro-magnetic field. But, since this behavior is affected by loop corrections, it is impossible to observe the "bare" charge. After renormalization, the electric charge entering in the expression of $\mathcal{M}$ is the one that is physically measured at low energies [6].

The finite parts of the corrections still remain. Some of them can be summed to all orders in α to a give a finite, momentum dependent correction to the coupling constant. In QED, this can be viewed as the result of a cloud of virtual electron-positron pairs created by processes such as in Figure 1.1(b). The positrons in this cloud are attracted towards the electron, screening its charge. The measured charge then depends on the scale at which the electron and its cloud are probed. As the energy increases and the distance scale decreases, the observed charge increases. This variation with energy scale leads to variation of the coupling constant constanst α

and this behavior is referred to as the *running coupling constant.*

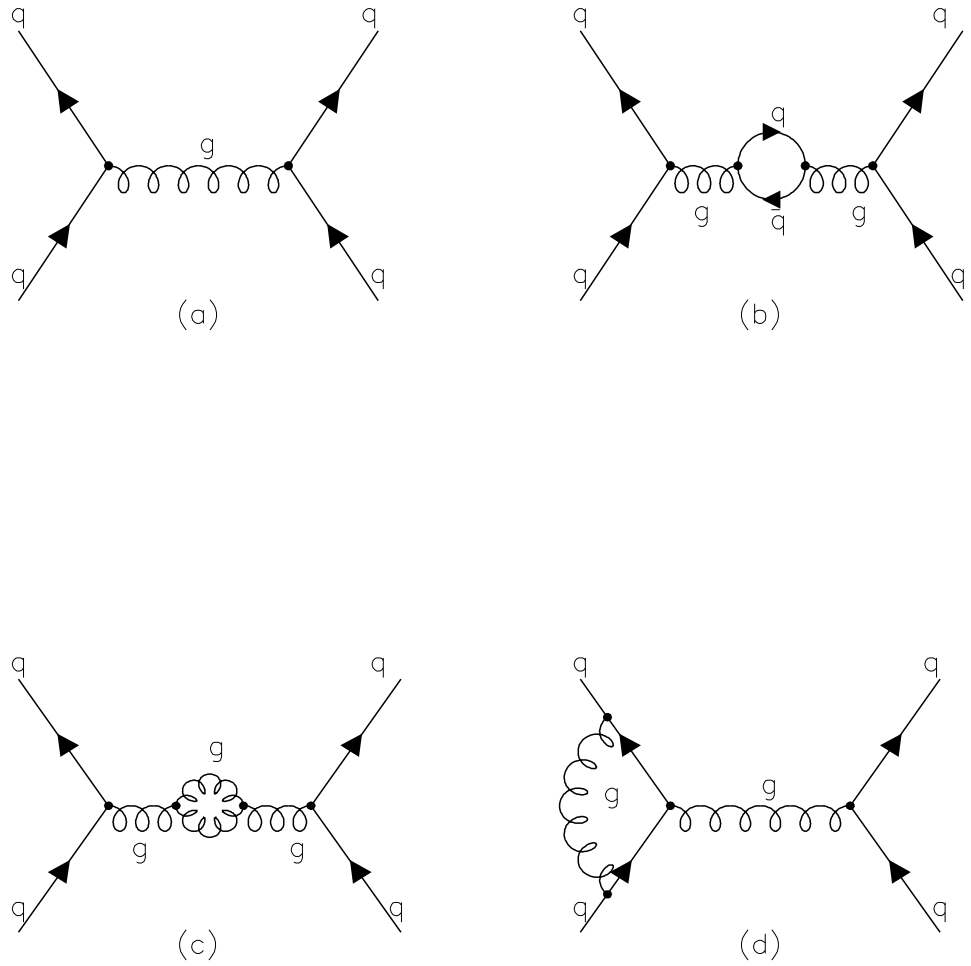

Figure 1.2: Strong interaction scattering between two quarks (a) with the exchange of a single gluon. Diagrams (b), (c) and (d) represent higher order 1 loop corrections to (a). Since QCD is a non-Abelian theory, the higher order diagrams include contributions from both quark and gluon loops.

After renormalization, calculations are based on an experimentally measured value of α, taken at some reference energy scale μ. It is clear that physical quantities, as $\mathcal{M}$, should not depend on the value of μ. Thus, the dependence of $\mathcal{M}$ on μ must cancel

the dependence of α on μ. This leads to the Renormalization Group Equation [6].

The theory of strong interactions, QCD, can be partly treated using the perturbation theory. As with QED, the observed value of the coupling constant, α_s, depends on the energy scale. However, gluon self-coupling causes α_s to *decrease* for small distances and high energies. Figure 1.2 show examples of strong scattering between two quarks. The non-Abelian nature of the theory is revealed in diagram 1.2(c) where gluon loops contribute to the strong scattering. This leads to the phenomena of *asymptotic freedom* and *quark confinement* at small and large distances respectively. For high-energy interactions, such as when quarks are probed inside a hadron, they can be treated almost as free particles and calculations can be made using perturbation theory. As the energy scale falls, α_s, approaches and exceeds unity and perturbation theory does not apply. Exact calculations are not possible in this non-perturbative regime. Instead one relies on phenomenological models.

1.1.3 The Electroweak Theory

The electroweak model of Glashow, Weinberg and Salam is based on the group $SU(2)_L \times U(1)$. The underlying gauge fields are W_i^μ(i=1,2,3) for the SU(2) factor and B^μ for the U(1) factor, with fundamental coupling constants g and g' respectively. The physical weak and electro-magnetic vector bosons are expressed as a linear combination of these fields:

$$W_i^\mu = \frac{1}{\sqrt{2}}(W_1^\mu \mp iW_2^\mu) \tag{1.1}$$

$$Z^\mu = W_3^\mu \cos\theta_W - B^\mu \sin\theta_W \tag{1.2}$$

$$A^\mu = W_3^\mu \sin\theta_W + B^\mu \cos\theta_W \tag{1.3}$$

respectively, where

$$\sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}} \tag{1.4}$$

$$\cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}} \tag{1.5}$$

The Fermi constant for weak decay can be deduced:

$$G_F = \frac{\sqrt{2}g^2}{8M_W^2} \tag{1.6}$$

The $W^{\pm}$ and Z^0 are given masses (M_W and M_Z respectively) via the Higgs mechanism [6]. It is these masses which cause the apparent weakness of the weak force. At low energies, interactions are highly suppressed by the need to produce a weak vector boson.

The $SU(2)_L$ field couples only to left-handed fermions. These transform as weak isospin doublets, with a weak isospin quantum number T = 1/2. The up-type member of each doublet is assigned a third component of isospin $T_3 = +1/2$ while the lower type member has $T_3 = -1/2$:

$$T = 1/2, \quad T_3 = \left\{ \begin{array}{ll} +1/2 & \left(\begin{array}{c} u \\ d' \end{array} \right)_L \left(\begin{array}{c} c \\ s' \end{array} \right)_L \left(\begin{array}{c} t \\ b' \end{array} \right)_L \\ -1/2 & \end{array} \right.$$

$$T = 1/2, \quad T_3 = \left\{ \begin{array}{ll} +1/2 & \left(\begin{array}{c} \nu_e \\ e^- \end{array} \right)_L \left(\begin{array}{c} \nu_\mu \\ \mu^- \end{array} \right)_L \left(\begin{array}{c} \nu_\tau \\ \tau^- \end{array} \right)_L \\ -1/2 & \end{array} \right.$$

The right-handed components have $T_3 = 0$ and are weak isospin singlets.

In the case of quarks, the eigenstates with which the weak force interacts are not

quite the same as the mass eigenstates. Instead, the down-type quarks, namely d, s and b, undergo mixing under the operation of the *CKM matrix* (Cabbibo-Kobayashi-Maskawa matrix [4]), into d', s' and b':

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

As a result of the mixing, b quarks are able to decay to c quarks via the charged weak interaction:

$$b \rightarrow c + W^-$$

The W subsequently decays semileptonically, to a charged lepton and its neutrino partner, or hadronically to a pair of quarks.

The $b \rightarrow c$ transition is possible because the s' weak eigenstate, which appears in the same doublet as the c quark, contains a fraction V_{cb} of the b mass eigenstate. The transition takes place between the elements of the $(c, s')_L$ doublet. It is also possible for b to decay via the $(u, d')_L$ doublet. Because the decay of the b quark can proceed only through the slow weak force and because it is further inhibited by the factor V_{cb}=0.036-0.046 (or the even smaller V_{ub}=0.002-0.005), it has a long lifetime compared to other quarks. The lifetime is $\sim$1.5 ps [81]. The relatively long lifetime of the b quark sets the basis for the techniques developed for its identification. In contrast to quarks, there is no flavor mixing process among leptons and it appears that the lepton number is conserved separately for each flavor.

1.1.4 QCD - The theory of strong interactions

The strong interactions are described by a local non-Abelian gauge theory of quarks and gluons in which SU(3) is the gauge group and the gluons are the gauge bosons. In each flavor, a quark has three color states (Red, Green and Blue) to form a triplet in the fundamental representation of $SU(3)_C$ while eight gluons form an octet. The covariant derivative acting on a quark field is

$$\mathcal{D}_{\mu,jk} = \delta_{jk}\partial_\mu - ig_3\lambda^a_{jk}G^a_\mu \tag{1.7}$$

resulting to the Lagrangian

$$\mathcal{L} = \bar{\psi}_j(i\gamma^\mu\mathcal{D}_{\mu,jk} - M_j\delta_{jk})\psi_k - \frac{1}{4}F^{\mu\nu}_a F_{a\mu\nu} \tag{1.8}$$

where the indices a, j and k refer to color and assume the values $a = 1, \cdots, 8$ and j, $k = 1, 2, 3$. ψ_j is the 4-component Dirac spinors representing the quark fields, G^μ_a are the gluon fields, λ_a are the generators of the $SU(3)_C$ symmetry group, g_3 is the strong coupling coefficient and M_j is the quark mass. The gluon field tensor is defined as

$$F^{\mu\nu}_a \equiv \partial^\mu G^\nu_a - \partial^\nu G^\mu_a - g_3 f_{abc} G^\mu_b G^\nu_c$$

where f_{abc} are the structure constants of the SU(3).

The strong interaction potential between two quarks has the form

$$V(r) \propto \frac{\alpha_s(r)}{r}$$

and the dependence of α_s on r is such that $V(r) = \lambda r$ when $r \to \infty$. Due to the

self-coupling of the gluons, the lines of color force attain a constant flux at large separations. This means that as more energy is injected into a $q\bar{q}$ bound state new $q\bar{q}$ states are generated from vacuum. This explains the reason single quarks are not found isolated in nature but they can only be found in colorless composite states. The phenomenon is called *confinement*. At very small separations the potential is very small and the quarks behave as free particles as far as strong interactions are concerned. This phenomenon is called *asymptotic freedom*.

1.2 Why top must exist

As mentioned above, the Standard Model does not predict the number of generation but it requires that the quarks come in pairs. Since the discovery of the b quark, extensive studies of its decays over many years and in different experiments led to accurate measurements of its electric charge and weak isospin. These measurements suggest that the bottom quark is part of a new generation of quarks and according to theoretical arguments, it is indeed part of a doublet. This implies the b-quark should have an isospin partner, the top quark. Furthermore, precision electroweak measurements over the years offered not only indirect evidence for the existence of the top quark but also information about its mass. Even in the absence of any direct evidence for top, there would be good reason to believe that it exists, as otherwise the Standard Model would not be consistent. Finally, the long awaited discovery of the top quark was announced in 1994 and its existence was confirmed in the following year. The accumulated evidence over the years prior to its discovery were so strong that the discovery did not come as a surprise as in the cases of the charm and bottom quarks. On the contrary the surprise and at the same mystery surrounding the top

quark are associated to its large mass. According to recent studies by the CDF and D0 collaborations, the top mass is measured to be 175.9±4.8(stat)±4.9(syst) GeV/c^2 [24] and 172.2±5.2(stat)±4.9(syst) GeV/c^2 [26], respectively. So the top quark is much heavier than the rest of the quarks while its mass is close to the scale where the electroweak symmetry breaking occurs.

1.2.1 Cancellation of Anomalies

In the Standard Electroweak model, the existence of triangle diagrams of the form illustrated in Figure 1.3 suggest that there are divergences that cannot be treated by the usual renormalization procedures. These divergences cannot be removed unless there are other fields in the theory that cancel these anomalies. However, with the requirement of the existence of three generations of leptons and quarks, such anomalies cancel out making the theory renormalizable. Each flavor of fermion which can go around the loop make a contribution to the process. The divergences cancel out by adding the contributions to the process of each fermion in a generation, provided that the sum of the electric charge for these fermions is zero. The contribution of

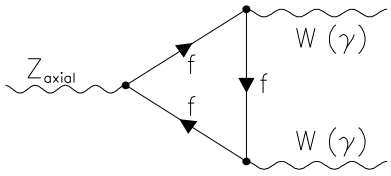

Figure 1.3: Examples of triangular graphs that give rise to the chiral anomaly.

each fermion is proportional to $N_c g_A^f Q_f^2$ where N_c is the possible number of colors for each fermion (so each quark contribution is counted three times), g_A^f is the axial coupling to the Z of each fermion and Q_f is the fermion charge. The contribution of each quark doublet exactly cancels the contribution of its corresponding lepton doublet. Therefore, the presence of the top quark is required to ensure the cancellation of these anomalies. However, this method of eliminating the problem with this type of anomaly diagrams is not the only possibility, and therefore just by itself the argument is not sufficient to conclude the presence of the third generation quark doublet.

1.2.2 $B^0 - \overline{B}^0$ mixing

The B^0 and $\overline{B}^0$ mesons can mix with each other through box diagrams like the ones shown in Figure 1.4, involving internal quark lines. The top contribution domi-

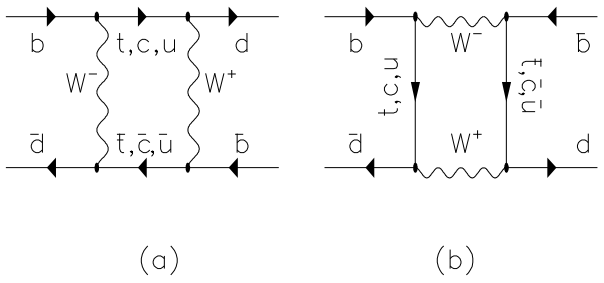

Figure 1.4: Box diagrams for the $B^0 - \overline{B}^0$ mixing.

nates in these diagrams because the matrix element for the process is proportional to $M_q^2 |V_{qb} V_{qd}^\star|^2$ and therefore depends on the internal quark mass. Results on the $B^0 - \overline{B^0}$ mixing suggest that in order to explain the observed mixing within the Stan-

dard Model, the top quark mass has to be greater than 45 GeV/c^2 [12]. However, as in the previous case, the above argument is not conclusive, since one could build models in which other states can give the needed contribution.

1.2.3 $B^0/B_s \to \mu^+\mu^-$

The decay of $B^0/B_s \to \mu^+\mu^-$ which proceed through the diagram shown in Figure 1.5 is also suppressed because of the GIM mechanism. The relevant quantities involved in these processes are V_{td} and V_{ts} depending on the flavor of the second quark forming the b meson. The Standard Model predicts that the branching ratios for the B_d and

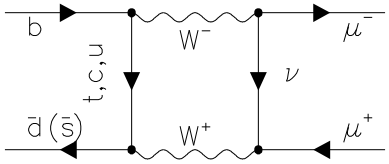

Figure 1.5: Box diagram contributing to the $B^0/B_s \to \ell^+\ell^-$ rare decays.

B_s decays to $\mu^+\mu^-$ are $(1.5 \pm 1)10^{-10}$ and $(3.5 \pm 1)10^{-9}$ respectively, assuming the b-quark is part of an isospin doublet. If instead it were a singlet then there would be no GIM suppression and the branching fractions would be significantly higher. Recent results from the CDF experiment [13] set the upper limits for these tansitions to $Br(B_d^0 \to \mu^+\mu^-) < 8.7 \times 10^{-7}$ and $Br(B_s^0 \to \mu^+\mu^-) < 2.6 \times 10^{-6}$ at 95% confidence level.

1.2.4 Suppression of FCNC b quark decays

In the case that top did not exist, both helicity states of the b-quark would be SU(2) singlets. In this case, the b would not interact with W's and could not decay through the usual weak processes. However, the b-quark is observed to decay. Therefore, in order for the Standard Model to accomodate for the b-decays and at the same time to allow the b-quark to be part of an SU(2) singlet, the assumption that the b mixes with a lighter quark of an SU(2) doublet had to be made. This lighter quark could then decay normally via a virtual W, as shown in Figure 1.6(a). If this were the

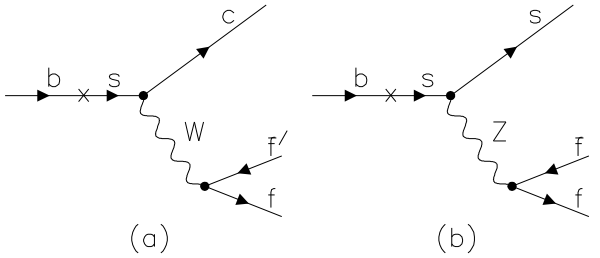

Figure 1.6: Hypothetical b-quark decay via quark mixing (a). Additional decay that would be allowed if the decay (a) were allowed.

case, then the corresponding process involving a Z, shown in Figure 1.6(b), would also be present, with a cross section of at least 12% of the former process. However, the experimental upper limits for this ratio is several order of magnitude less than the predicted value. Therefore the suppression of FCNC processes for the b-quark decays is a strong indication that the bottom quark is not an SU(2) singlet and it has to be in an SU(2) doublet with the top quark as a partner.

1.2.5 Forward-Backward asymmetry in $e^+e^- \rightarrow b\bar{b}$

As discussed before, in the context of the electroweak standard model, particles are grouped into SU(2) spin multiplets. The helicity state of each particle is assigned a quantum number called the *weak isospin,* T_3. The value of the particle's isospin is integral or half-integral depending on whether the particle belongs to an SU(2) singlet or doublet. The strength of the weak interactions depends in part on the values of the isospin and therefore the isospin value of a particle can be measured.

This was done for the reaction $e^+e^- \rightarrow (\gamma, Z) \rightarrow b\bar{b}$, which proceeds through a photon or Z exchange diagrams. It is expected that the angular distribution of the b's in the photon exchange processes to be symmetric, but that from the Z exchange should not be because the axial and vector couplings of the Z to fermions depend on the isospin of the coupled fermions. The interference between these two processes give rise to an asymmetric angular distribution for the b production. The forward-backward asymmetry is defined as

$$A_{FB}^{b\bar{b}} \equiv \frac{\sigma_F^b - \sigma_B^b}{\sigma_F^b + \sigma_B^b}$$

where σ_F^b and σ_B^b are the cross-sections of b-jets in the forward and backward direction relative to the e^- beam direction.

The amount of this asymmetry depends on the weak isospin of the left-handed b-quark T_{3L}^b. The left (L_b) and right-handed (R_b) coupling of a b-quark singlet can be expressed as:

$$L_b = T_{3L}^b - e_b \sin^2 \theta_W$$

$$R_b = T_{3R}^b - e_b \sin^2 \theta_W$$

where, e_b is the b-quark electric charge and θ_W the Weinberg angle. Below the Z^0 pole, the asymmetry $A_{FB}^{b\bar{b}}$ accepts contributions from both γ and Z exchange processes and it varies as $(L_b - R_b)$. Under the Z^0 pole the asymmetry is given by

$$A_{FB}^{b\bar{b}} = \frac{L_b^2 - R_b^2}{L_b^2 + R_b^2}$$

If the b-quark were an SU(2) singlet, then $T_{3L}^b = T_{3R}^b$ and consequently $A_{FB}^{b\bar{b}} = 0$ under the Z^0 pole. This scenario is contradicted by current measurements of $A_{FB}^{b\bar{b}}$ at LEP [14] which give $A_{FB}^{b\bar{b}} = 0.0983 \pm 0.0024$. On the other hand, the measured value agrees, within errors, with the Standard Model expectation of $A_{FB}^{b\bar{b}} = 0.0995 \pm 0.002$ based on $T_3^b \equiv -1/2$. Therefore, the b-quark is part of an SU(2) doublet.

In fact the values of T_{3L}^b and T_{3R}^b can be calculated via another measurement, which is a function of both components:

$$\Gamma(Z \to b\bar{b}) \propto (T_{3L}^b + \frac{1}{3}\sin^2\theta_W)^2 + (T_{3R}^b + \frac{1}{3}\sin^2\theta_W)^2$$

The constant of proportionality has been calculated to good accuracy. Substitution of the measured values of $A_{FB}^{b\bar{b}}$ and $\Gamma(Z \to b\bar{b})$ to the above expression yields $T_{3L}^b = -0.504^{+0.018}_{-0.011}$, and $T_{3R}^b = -0.008^{+0.056}_{-0.056}$.

1.2.6 Precision electroweak measurements

Because of the comparatively weak strength of the electroweak interaction, it is possible to use perturbation theory to predict experimental observable with high precision. The Standard Model has 18 input parameters which are: the three coupling strengths, α, α_s and G_F of the electro-magnetic, strong and weak forces respectively, the masses

M_Z and M_H of the Z^0 and Higgs bosons, the 9 masses of the quarks and charged leptons and the three amplitudes and one phase needed to determine the CKM matrix. Of all these parameters, only α, G_F, M_Z and the masses of the three charged leptons are known with very good accuracy. In fact, the Z mass along the W mass can be predicted by the theory, given the value of $\sin\theta_W$.

$$M_W = (\frac{\pi\alpha}{\sqrt{2}G_F})^{\frac{1}{2}} \sin^{-1}\theta_W \quad (1.9)$$

$$M_Z = (\frac{\pi\alpha}{\sqrt{2}G_F})^{\frac{1}{2}} (\cos\theta_W sin\theta_W)^{-1} \quad (1.10)$$

In this case $\sin\theta_W$ becomes a free parameter of the model. However, the current measurements of M_Z at LEP give better accuracy and therefore is instructive to be used to extract other parameters of the model. The top and bottom quark masses and the CKM matrix elements that link the first two generations are less well known. The masses of the lightest quarks, α_s and the remaining CKM matrix elements are poorly known. Finally, the mass of the Higgs boson is still unknown. However, limits have been placed on the possible value of the Higgs mass. For example, the lower limit on the Higgs mass from direct searches is approximately 77 GeV/c^2 [14].

At tree level, predictions for the observables included in the electroweak fit are made solely in terms of the three best-known parameters: α, G_F and M_Z. Such observables include the total and partial widths of the Z and the polarization of its decay products. They also include the forward-backward and left-right asymmetries for fermions produced in the Z decay.

Beyond tree level, it is necessary to take into account loop corrections which depend on α_s and the fermion and Higgs masses. Figure 1.7 shows examples of diagrams involving the top quark and the Higgs which contribute to the radiative

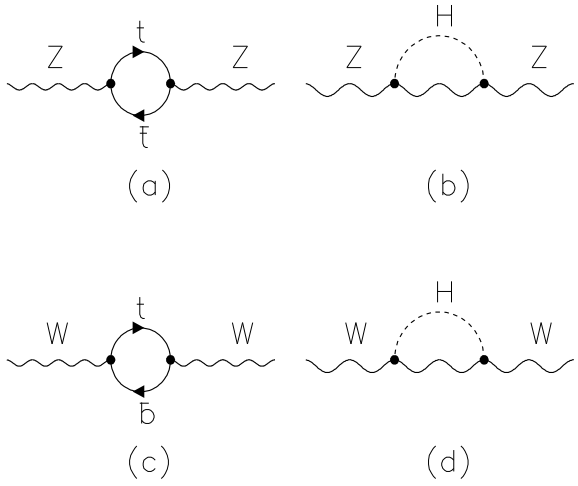

Figure 1.7: Diagrams for radiative corrections to the Z and W boson masses involving the top quark and the Higgs boson.

corrections of the W and Z boson masses.

The top quark mass plays a very important role in these radiative corrections because of the large mass difference between the top quark and the next heaviest quark, the bottom quark. Most of the radiative corrections have quadratic dependence on the top mass in the form M_{top}^2/M_Z^2, whereas they have logarithmic dependence on the Higgs mass with terms of the form $\ln(M_{Higgs}/M_Z)$.

When these corrections are made, the theoretical predictions can be compared to the measured values. By making a fit to the experimental results, call *the electroweak fit*, constraints on the remaining input parameters of the predictions are obtained. Important information is obtained with this technique on the Higgs, the top quark masses and α_s. In addition, the mutual consistency of the observed results, as measured by the χ^2 of the fit, gives an indication of the validity of the Standard Model

predictions. Any parameter of the fit whose experimental value cannot be accomodated in the fit may be an indication of physics beyond te Standard Model. With the results of LEP and SLD being precise enough, the loop corrections need to be included in the fit. This means that it is possible to observe physics that enters the loops but cannot be otherwise observed. Such an example it would be a particle which were quite heavy to be directly produced.

Of the parameters used in the electroweak fit, one of the important ones is the ratio $R_b = \Gamma_{Z^0 \to b\bar{b}}/\Gamma_{Z^0 \to hadrons}$. In the $\Gamma_{Z^0 \to b\bar{b}}$, the coupling between the top and bottom quarks have particularly large impact when vertex corrections are considered. All the other corrections which depend on the Higgs mass and α_s, are common for

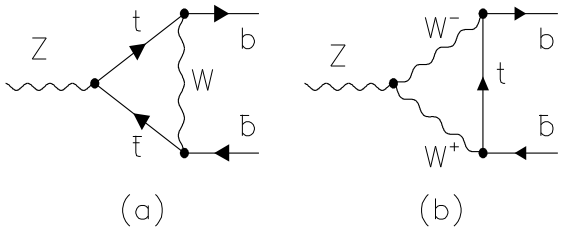

Figure 1.8: Diagrams involving the top quark and contribute to the $Zb\bar{b}$ vertex corrections.

all other quark flavors and in the ratio R_b, as defined above, these corrections can be factored out. The value of R_b is the only indirect measurement of the top mass independent of the Higgs mass, because all the terms depending on the Higgs mass cancel out in the ratio. Figure 1.8 show the one-loop corrections to the $Zb\bar{b}$ vertex involving the top quark.

The W mass also depends on the top quark and the Higgs masses through loop diagrams as the one shown in Figure 1.7(c) and (d). A precise measurement of the W mass for a fixed Higgs mass constrains the top mass. Turning the argument around, precise measurement of W and top masses can be used to provide information on the Higgs mass or signal the existence of new physics. The constraints on the Higgs mass are relatively weak because of the radiative corrections have a logarithmic dependence on the Higgs mass.

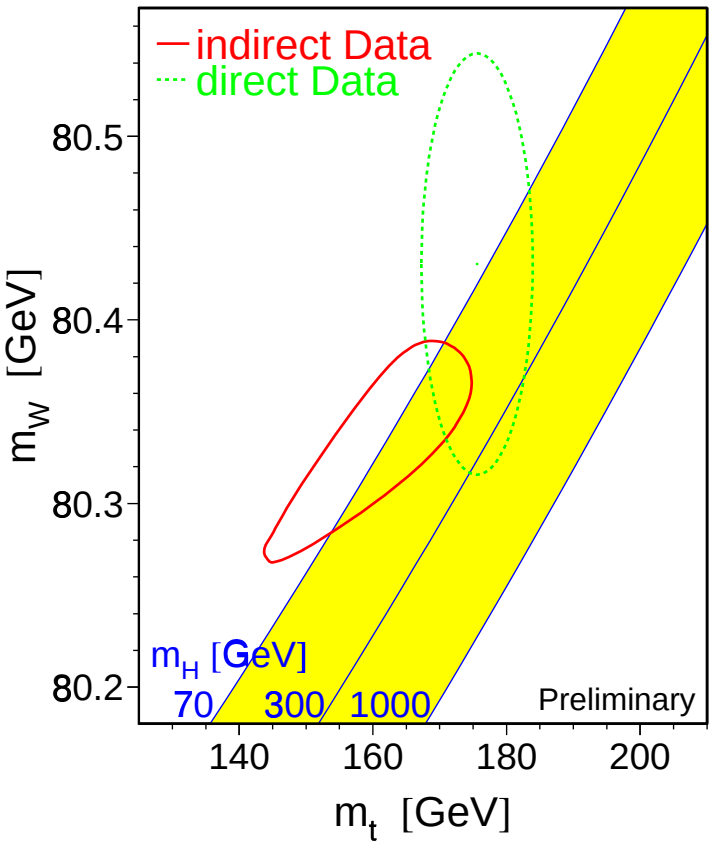

Figure 1.9: Comprarison of the indirect measurements of m_w and m_t using LEP, SLD and νN data (solid contour) to the direct measurements from the Tevatron and LEP-II data (dashed contour). In both cases the 68% contours are plotted. Also shown is the Standard Model relationship for the masses as a function of the Higgs mass [14].

Combining the indirect information from the LEP and SLD measurements and the W mass measurements a global fit for m_t is performed [14]. The fit returns a value for $m_t = 173.1 \pm 5.4$ treating the Higgs mass as a free parameter.

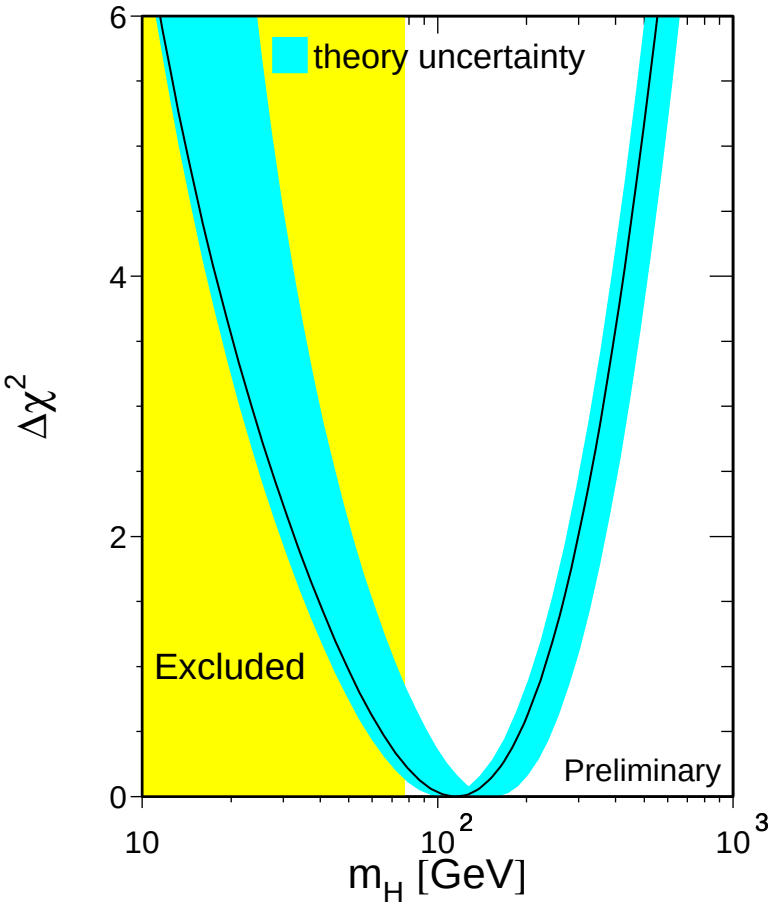

Figure 1.10: $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ versus m_H curve. The line is the result of the fit using all the data values of all electroweak parameters and leaving the Higgs mass as a free parameter. The band represents an estimate of the theoretical error due to missing higher order corrections while the vertical band represents the 95% confidence level exclusion limit on m_H based on the direct search. The plot is from Reference [14].

Compilation of all Standard Model parameters measured at LEP along with the current measurements of the W and top masses at the Tevatron yield a prediction

for the Higgs mass of 115^{+116}_{-66} GeV/c^2. Figure 1.10 shows the observed value of $\Delta\chi^2 \equiv \chi^2 - x^2_{min}$ obtained from the fit to the electroweak data versus the Higgs mass. An upper limit of $m_H = 420$ GeV/c^2 at 95% confidence level on the Higgs mass can be inferred taking also into account the error on the theoretical predictions [14].

In conclusion all the neutral current data as well as the W and top mass measurements, are in agreement with each other. Despite the large number of very precise measurements, there is still little information on the Higgs mass. The data seem to prefer a small value of the Higgs mass as shown in Figure 1.10. It is also interesting to note in Figure 1.9 that the neutral current data seem to prefer a lower mass also for the top quark if the direct measurements of the top mass from the Tevatron are excluded from the fit. By doing the fit without the direct top mass measurement one can check whether the predictions for the top mass from the radiative corrections are in agreement with the direct measurement. Fitting the data in terms of m_t, m_H and $\alpha_s(m_Z)$ one finds a top mass of 161^{+10}_{-8} GeV/c^2, while if the direct measurement of the W mass is ignored then the fit prefers a value of $m_t = 157^{+10}_{-9}$ GeV/c^2.

1.3 Top production at the Tevatron

The top quark is very heavy and can only be produced in colliders with large center of mass energy. It can be produced in e^-e^+ colliders through the process $e^-e^+ \rightarrow Z(\gamma) \rightarrow t\bar{t}$ where the Z or γ are off-shell. Since the highest available center of mass energy e^-e^+ collider is $\sqrt{s} \sim 170$ GeV, top cannot be produced at e^-e^+ colliders. It could be produced at an ep collider like the one at HERA which operates at 310 GeV center of mass energy but the top production cross section is too small for observation. At Fermilab, the Tevatron which is a $p\bar{p}$ collider, operates at center of mass energy

of 1.8 TeV, the highest available energy ever. Therefore, the top quark can only be produced at the Tevatron which offer sufficiently large energies. Top quarks in $p\bar{p}$ collisions can be produced either in pairs, $p\bar{p} \rightarrow t\bar{t} + X$, or singly, $p\bar{p} \rightarrow tb + X$.

1.3.1 $t\bar{t}$ pair production

At the Tevatron the $t\bar{t}$ pair production is the dominant mechanism for top production for a wide range of top quark masses. The $t\bar{t}$ pair production can be described at leading order by the diagrams shown in Figure 1.11. Two processes contribute to the diagrams. The $q\bar{q}$ annihilation and the gluon fusion. The gluon fusion process is ex-

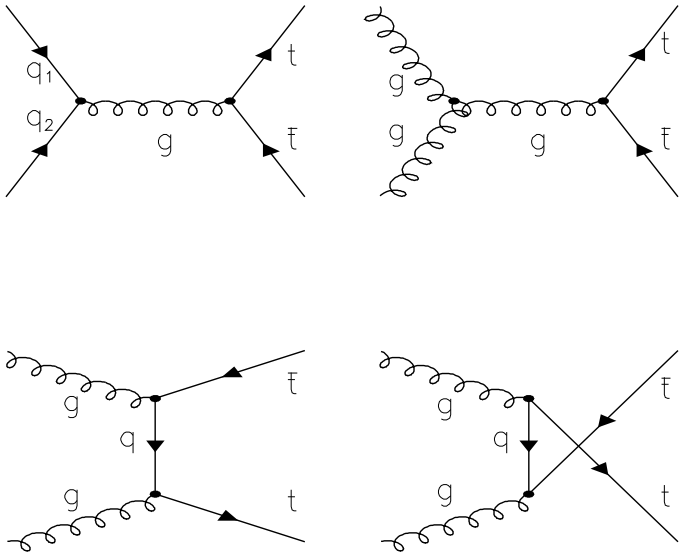

Figure 1.11: Leading order diagrams for $t\bar{t}$ production in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV.

pected to dominate for lower top mass while for higher top masses the $q\overline{q}$ annihilation processes is the dominant one. This happens because as the mass of the top quark increases, the colliding partons need to carry larger fractions of the proton or anti-proton energy in order to increase the center of mass energy of the colliding system. Gluons typical carry smaller fraction of the total proton momentum and therefore the $gg \rightarrow t\bar{t}$ process is suppressed. For a top mass of 150 GeV/c^2 the gg contribution to the $t\bar{t}$ cross section is $\sim$ 20% and decreases to $\sim$ 10% for top mass of the order of 200 GeV/c^2. This is shown in Figure 1.12 where the fractional contribution to the $t\bar{t}$ cross-section of the two processes is plotted as a function of the top mass.

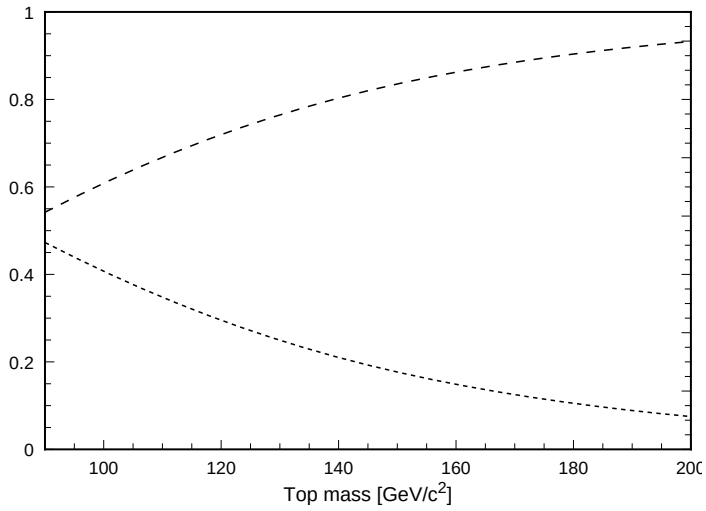

Figure 1.12: Fractional contributions of the $q\overline{q}$ and gg processes at NLO to the $t\bar{t}$ production cross section at the Tevatron as estimated in Reference [80]. The top curve corresponds to the $q\overline{q}$ process and the bottom one to the gg process.

The cross section for $t\bar{t}$ pair production can be calculated using perturbative QCD. It can be written as the product of the parton distribution functions inside the protons

and the parton-parton point cross section

$$\sigma(p\bar{p} \rightarrow t\bar{t}) = \sum_{a,b} \int dx_a F_a(x_a, \mu^2) \int dx_b F_b(x_b, \mu^2) \hat{\sigma}_{ab}(\hat{s}, \mu^2, M_{top})$$

The indices a and b refer to the incoming partons, components of the proton and the anti-proton respectively. The functions F_a and F_b are the number densities or parton distribution functions, of the partons inside the proton and anti-proton evaluated at a momentum scale μ. The function $F_a(x_a, \mu^2)dx_a$ represents the probability that a parton of type a (quark, anti-quark, or gluon) carries a fraction of the proton or anti-proton momentum between x_a and $x_a + dx_a$. $\hat{\sigma}_{ab}$ is the point cross section for the process $a + b \rightarrow t\bar{t}$ and $\hat{s} = x_a x_b s = 4x_a x_b P^2$ is the square of the center of mass energy of the parton-parton interaction, where s refers to the total available energy for the collider ($\sqrt{s}$ = 1.8 TeV) and P to the magnitude of the momentum of the colliding beam.

The parameter μ called the *renormalization scale*, is an arbitrary parameter with dimensions of energy which is introduced in the renormalization procedure. In principle, the cross section should be independent of the value chosen for this parameter but because the cross section calculations are performed to finite order in perturbation theory the dependence on the renormalization scale remains. The cross section usually is evaluated for $\mu^2 = m_t^2$. The parton distribution functions F_a and F_b are extracted from parametrizations of fits to experimental results mostly from deep inelastic scattering experiments.

There are two main sources of uncertainty in the calculation of the $t\bar{t}$ cross section. The first is due to the uncertainty in the renormalization scale, μ, used in the perturbative calculation. The size of this uncertainty is usually estimated by varying

the value of μ around the top quark mass. The second source of uncertainty is due to the knowledge of the parton distribution functions, F_a and F_b, and the assumed value of the QCD parameter Λ_{QCD}. The value of Λ_{QCD} which sets the energy scale at which the perturbative calculation breaks down, affects the μ^2 evolution of both α_s and the parton distribution functions. In particular the extraction of the gluon distribution functions from deep inelastic data also depends on Λ_{QCD}. The uncertainty due to the parton distribution functions is estimated by studying the variations in the calculated cross section for different parametrizations of distribution functions and different values of Λ_{QCD}. The NLO calculations were found to be stable at the $\sim$20% level when the factorization scale varied between $2m_t$ and $0.5m_t$ and the Λ_{QCD} from 90 to 250 MeV [9]. This uncertainty accepts approximately equal contributions from both the choice of the factorization scale and the value of Λ_{QCD}. This study was performed using the MRSA$'$ [10] set of parton densities fitted with different values of the strong coupling constant, α_s. The fitted values of $\alpha_s(M_Z)$ used in the MRSA$'$ sets, are 0.105, 0.110, 0.115, 0.120 for Λ_5 = 100, 140, 190 and 253 MeV and two-loop formula for the α_s calculation. The calculated $t\bar{t}$ cross section is $4.75^{+0.73}_{-0.62}$ pb. The central value is calculated for $\mu_R = \mu_F = m_t$ and Λ_5=152 MeV which corresponds to $\alpha_s(M_Z) = 0.1113$.

The theoretical cross section for $t\bar{t}$ production was known at NLO order in perturbation theory since the late eighties [16]. These calculations were subsequently convoluted with parton distribution functions by Altarelli *et al.* [17] and Ellis [18] who found that the order α_s^3 corrections raised the cross section by $\sim$30%. These calculations were subsequently refined by Laenen *et al.* [80] by including leading logarithm corrections corresponding to initial state gluon bremsstrahlung. The corrections can be resummed to all order in perturbative QCD. The corrections are large

near the $t\bar{t}$ threshold. Their calculation introduces an infrared cutoff $\mu_0 >> \Lambda_{QCD}$ where the resummation is terminated. The resummation diverges as $\mu_0 \to 0$ because of dominant non-perturbative effects. These corrections are positive at all orders of the perturbative calculations. Therefore, the lower limit of the $t\bar{t}$ cross section is estimated by the sum of the full $O(\alpha_s^3)$ and $O(\alpha_s^4)$ correction at $\Lambda_{QCD} = 105$ MeV. The net effect of these additional terms to the magnitude of the $t\bar{t}$ cross section is an

Calculation	Order in perturbation theory	$\sigma_{t\bar{t}}$
Altarelli et al.	NLO	$4.37^{+0.26}_{-1.1}$
Laenen et al.	NLO plus gluon resummation	$4.95^{+0.70}_{-0.42}$
Berger et al.	NLO plus gluon resummation	$5.52^{+0.70}_{-0.42}$
Catani et al.	NLO plus gluon resummation	$4.75^{+0.63}_{-0.68}$

Table 1.2: Summary of $p\bar{p} \to t\bar{t}$ cross sections at the $\sqrt{s} = 1.8$ TeV for a top quark mass of 175 GeV/c^2. The various calculations use different structure functions for the central value. The uncertainties in the Laenen *et al.* and Berger *et al.* calculations do not include variations in the choice of the structure functions.

increase of $\sim 15\%$. However, the additional terms reduce the uncertainty due to the choice of the factorization scale μ^2 and the Λ_{QCD} to $\sim 10\%$. Similar techniques have been performed using a different resummation technique by Berger and Contopanagos [79] yielding cross sections $\sim 10\%$ higher than those obtained by Laenen [80]. These calculations are based on the Principle Value Resummation techniques [79] and are independent of the arbitrary infrared scale μ_0. The uncertainties in the calculation are estimated by varying the renormalization and factorization scale μ by a factor of two around the top mass. Recent evaluations of the soft gluon resummation by Catani *et*

al. [78] suggest that its contribution is much smaller than originally calculated. Table 1.2 gives a summary of the various theoretical calculations of $t\bar{t}$ production cross section at the Tevatron for top mass of 175 GeV/c^2. Figure 1.13 shows the different $t\bar{t}$ production cross section calculations as of a function of the top quark mass.

Two things are important to note in Figure 1.13. First, there is a rapid decrease in the cross section with increasing top mass. This implies that the mass of the top quark, along with its cross section, provides a sensitive test of the QCD calculations. Even with a relatively small sample of $t\bar{t}$ events it would be possible to detect deviations from the prediction, which could lead to new physics. For example a cross section larger than the predicted value may indicate that there is a new process contributing to the same final state. On the other hand a cross section smaller than the predicted one may indicate that there are additional decay channels beyond those in Standard Model, which would indicate the presence of new particles. Secondly, one notices that the cross section is of the order of $4-6$ picobarns, while the total inelastic $p\bar{p}$ cross section at $\sqrt{s} = 1.8$TeV is $\sim 50mb$ which is 10 order of magnitude higher than the $t\bar{t}$ cross section. Therefore, one top event can be found in approximately 10 billion events.

As mentioned before the $q\overline{q}$ process is the dominant one for $t\bar{t}$ production. At LO the $t\bar{t}$ production cross section for this process is given by:

$$\hat{\sigma} = \frac{8\pi\alpha_s^2}{27\hat{s}}\sqrt{1-\frac{4m_t^2}{\hat{s}}}(1+\frac{2m_t^2}{\hat{s}})$$

The parton-parton cross section rises as a function of $\hat{s}$ from zero at threshold ($\hat{s} = 4m_t^2$) and reaches a maximum at $\hat{s} = 5.6m_t^2$ and then falls asymptotically as $1/\hat{s}$. When convoluted with the parton densities the maximum of $q\overline{q} \rightarrow t\bar{t}$ cross section is

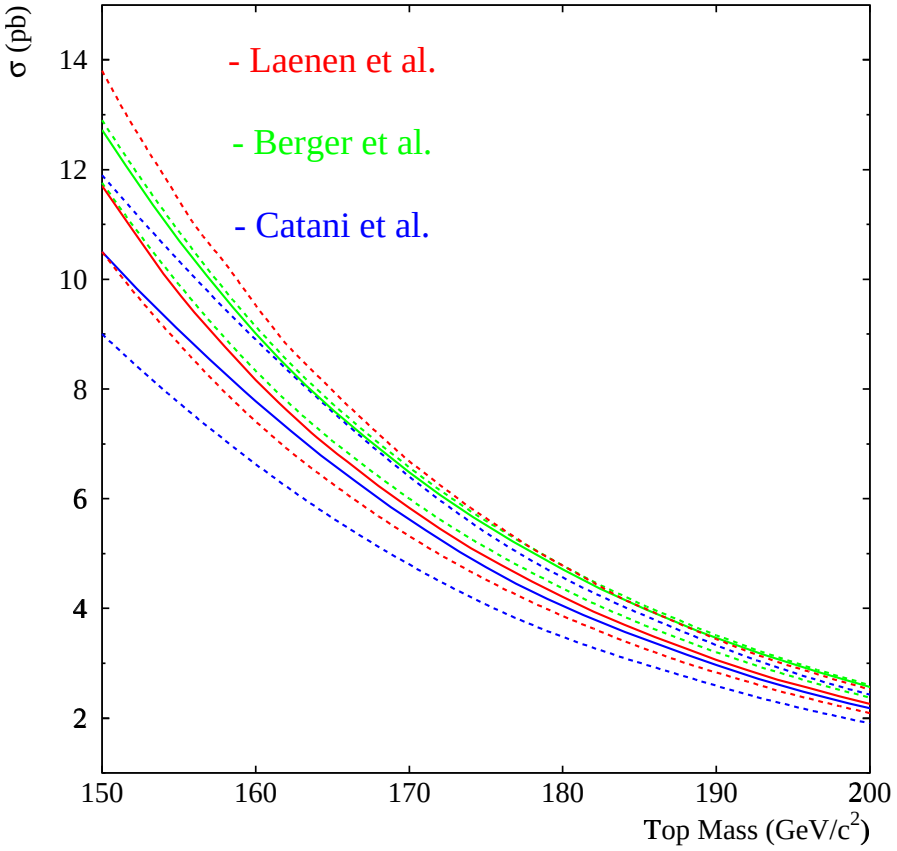

Figure 1.13: The theoretical cross sections for $t\bar{t}$ production expected for $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV as a function of the top quark mass. Shown are the calculations of Laenen *et al.* [80](solid), Berger *et al.* [79](dashed) and Catani *et al.* [78](dotted).

shifted down to $\hat{s} \approx 4.5m_t^2$. Therefore, the most probable energy for a top quark is $E \approx 1.1m_t$ and the most probable momentum is $P \approx 0.5m_t$.

The leading order diagrams (see Figure 1.11) lead to a back-to-back topology for the t and $\overline{t}$ in the transverse plane. This configuration is slightly modified by higher order corrections. Because the top momentum is not large compared to its mass, the decay products are not significantly boosted along the original top quark flight path, leading to nearly spherical events.

1.3.2 Single top production

There are two processes which produce a single top quark, rather than a $t\bar{t}$ pair. These are, the W-gluon fusion process as shown in Figure 1.14(a) and $q\bar{q}$ annihilation, $q\overline{q} \rightarrow t\overline{b}$ shown in Figure 1.14(b). Diagrams (c) and (d) correspond to leading and next-to-leading order contributions to $W - g$ fusion. In the case of collinear gluon emission, the large logarithms can be summed into a b-distribution function and the leading order diagram becomes the one shown in (e) [83].

The interesting feature of both production mechanisms is that they proceed via the weak interaction. Both processes probe the charged-current weak interaction of the top quark. The single top quark production cross section is proportional to the square of the CKM matrix element V_{tb}. V_{tb} however, cannot be measured directly in top quark decays since top is very short-lived. On the contrary, measurement of the single top production cross section can directly be used to measure V_{tb}. The cross section for both processes has been calculated to next to leading order in Reference [83]. The resulting cross sections are $\sigma_{W\ g} = 1.8 \pm 0.5$ pb and $\sigma_{W^{\star} \rightarrow t\bar{b}} = 0.74 \pm 0.045$ pb, for $W - g$ fusion and $q\bar{q}$ annihilation respectively.

The process $q\overline{q} \rightarrow t\overline{b}$ is especially powerful to study properties of the top quark because the initial state partons are light quarks with relatively large momentum fraction, x, and therefore the parton densities are well understood. Also because of the similarities of the process to the Drell-Yan process, $q\overline{q} \rightarrow \ell\overline{\nu}$, the total cross section has been studied in great detail. On the other hand this process suffers from a smaller cross section than the $W - g$ fusion mode and a larger depedence on the mass of the top quark.

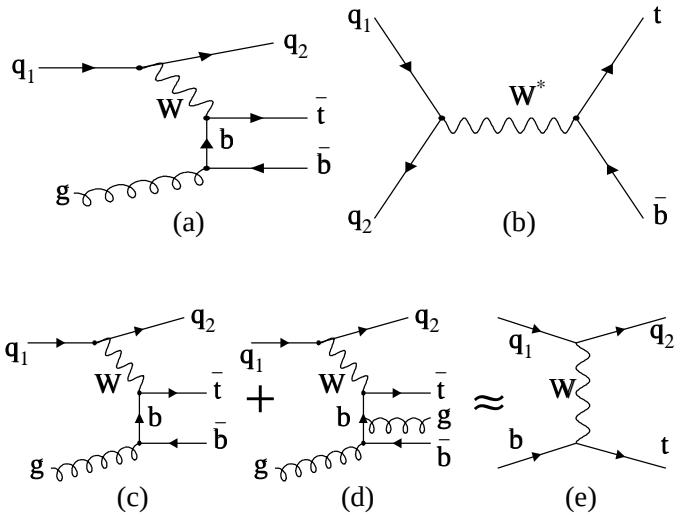

Figure 1.14: Diagrams contributing to the single top production at the Tevatron. Leading order diagrams to $W-g$ fusion (a) and to $q\bar{q}$ annihilation (b). Diagrams (c) and (d) correspond to leading and next-to-leading order contributions to $W-g$ fusion. In the case of collinear gluon emission, the large logarithms can be summed into a b-distribution function and the leading order diagram becomes the one shown in (e) [83].

1.4 Top quark decay

According to the CKM matrix, the top quark decays predominantly into a real W and a b quark, $t \to Wb$. Decays into Ws and Wd final states can also occur, but are suppressed because the corresponding CKM matrix elements are fairly small. The suppression factors are of the order $|V_{ts}|^2/|V_{tb}|^2 \approx 10^{-3}$ and $|V_{ts}|^2/|V_{tb}|^2 \approx 5 \times 10^{-4}$.

The parton hadronization is a non-perturbative process and it is estimated that takes place at a time-scale of $\Lambda_{QCD}^{-1} \approx (200MeV)^{-1} \approx 10^{-23}$ seconds. The top quark, because of its large mass, is very short-lived. The partial width for the decay $t \to Wb$

is given by [15]

$$\Gamma(t \to Wb) \approx 175 MeV (\frac{m_t}{m_W})^3$$

A top quark of mass of $m_t = 175$ GeV/c^2 should have a width of 2 GeV and a lifetime of $\sim 4 \times 10^{-25}$ seconds. Therefore it decays before it can form any hadronic state. According to the model of top quark hadronization [19] the t and $\bar{t}$ produced in the hard scatter are linked by color strings to the remnants of the proton and the anti-proton. When the separation of the quarks and the proton remnants exceeds about 1 fm, the color strings are expected to break and produce fragmentation particles. Since the top quark is so short-lived, it decays before it travels that distance (the decay distance a top quark travel in one lifetime is ~ 0.04 fm). Even in the case that the top quark undergoes hadronization, its effects would be extremely hard to observe because the expected fractional energy loss of the top quark during hadronization is small. However, the fragmentation of the b quark produced in the decay might be affected by the top hadronization, because the color string would link the b-quark with a light quark produced in the top hadronization rather than a light quark from the proton remnants.

The lack of hadronization effects have an interesting feature which can be used as a counter-argument to prove that top decays without hadronization. If the top quark decays before hadronization, it preserves its spin orientation. The effect could be measured because the top decays via the weak interaction which is sensitive to its spin orientation. Therefore the angular distribution of the top decay products can be used to measured its spin. The problem however is that top pairs are produced via the strong interaction which is parity conserved and therefore the $t\bar{t}$ pair is unpolarized. However there is correlation of spins between the top and anti-top. Observation of

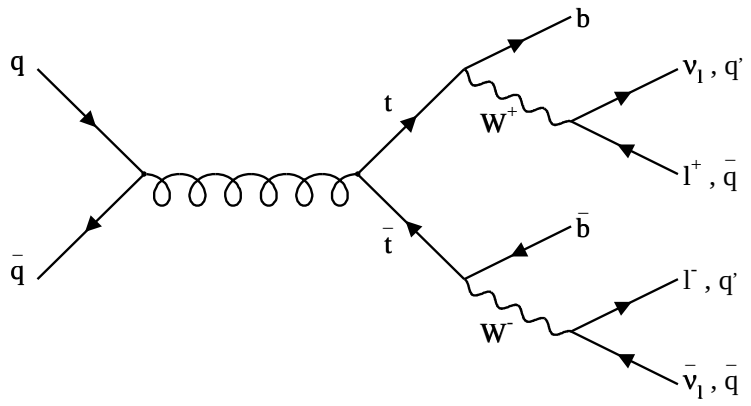

Figure 1.15: A Schematic view of the Standard Model decay modes of a $t\bar{t}$ pair.

such a spin correlation will prove that the top quark decays before hadronization.

1.4.1 Top decay modes

As discussed above nearly all top quarks will decay to a W and a b quark. The b quark undergoes fragmentation and hadronization forming a jet of final state particles. It is the decay of the W that adds variety to the top quark final state. Figure 1.15 shows a schematic view of the Standard Model decay modes of a $t\bar{t}$ pair. The W may decay into any pair of particles forming one doublet (except of course the kinematically forbidden tb doublet). As the masses of the particles in the other doublets are far less than the W mass, the phase space available for decay into any doublet is nearly equal. Therefore, the rate of the decay into each allowed doublet is identical, and when one takes into account that there are three sets of quark doublets since the quarks come with three types of color charge, one finds that the branching ratios of a W into the various doublets are: $BR(W \to \ell\nu) = 1/9$ and $BR(W \to q\overline{q}') = 1/3$.

Table 1.3 summarizes the possible Standard Model decay modes of a $t\bar{t}$ pair with their approximate branching ratios.

Top Decay Channels

W^+ \ W^-	$e^-\nu_e$	$\mu^-\nu_\mu$	$\tau^-\nu_\tau$	jets (ud, cs)
$e^+\nu_e$	1/81	1/81	1/81	6/81
$\mu^+\nu_\mu$	1/81	1/81	1/81	6/81
$\tau^+\nu_\tau$	1/81	1/81	1/81	6/81
jets (ud, cs)	6/81	6/81	6/81	36/81

2 modes x 3 colours = 6

3 Classes of Signals

Dilepton : **BR = 9/81** $\left\langle \begin{array}{l} ee,\ \mu\mu,\ e\mu \rightarrow \ \mathbf{4/81 = 5\%} \\ e\tau,\ \mu\tau,\ \tau\tau \rightarrow \ \mathbf{5/81 = 6\%} \end{array} \right.$

Lepton + Jets : **BR = 36/81** $\left\langle \begin{array}{l} e,\ \mu + \text{jets} \rightarrow \ \mathbf{24/81 = 30\%} \\ \tau + \text{jets} \rightarrow \ \mathbf{12/81 = 15\%} \end{array} \right.$

All Jets : **BR = 36/81 = 44%**

Table 1.3: Branching ratios for $t\bar{t}$ decay modes assuming Standard Model couplings.

The $t\bar{t}$ events are classified according to the number of W bosons that decay leptonically. In what follows, the τ's are not considered in the lepton counting because the τ hadronic decays are almost indistinguishable from the standard quark and gluon jets and in general result to very small acceptance. However, the τ leptonic decays to

an electron or muon are counted as resulting from W decays. In approximately 5% of $t\bar{t}$ decays, both W's decay leptonically. This mode is called the *dilepton channel* and experimentally is characterized by the presence of two high P_T leptons accompanied by large transverse energy imbalance due to the escaping neutrinos and two b-quark jets. The final state is very clean but the branching ratio is fairly small. About 30% of the time, one of the W's will decay to a lepton and a neutrino while the other W decays hadronically. This mode is called the *lepton + jets channel.* The experimentally observed final state for this channel consists of a high P_T lepton, large transverse energy imbalance, and (usually) four jets, two from the hadronic W decay and two b-quark jets. This channel resembles the standard QCD $W+$ jets production. However, the QCD background can be reduced significantly by a combination of kinematic requirements and identification of the b-jets in the event. Finally, the largest branching fraction (44%) corresponds to the case that both W's decay hadronically. This mode consists the *all hadronic channel* and is identified by the presence of (usually) six jets in the final state, four jets coming from the W's and two b-quark jets and not significant transverse energy imbalance. Despite the large branching ratio for this channel, the QCD backgrounds make the identification of the $t\bar{t}$ signal quite difficult. These backgrounds can be reduced to a manageable size requiring at least one of the jets in the event to be identified as a b-quark candidate. About $\sim$ 21% of the times, there will be a τ lepton in the final state of a $t\bar{t}$ event.

1.5 The discovery of the top quark

In 1995 the CDF and D0 collaborations announced [2, 3] the observation of the top quark. Each experiment observed a 5σ excess of $t\bar{t}$ candidate events over the expected

background. Both experiments also reconstructed peaks in the mass distribution corresponding to the top quark mass.

A brief summary of all current results on top quark from the CDF and D0 collaborations is presented in the two following sections.

1.5.1 Results on the top quark from CDF

The CDF results from Run 1A and Run 1B are based on $\sim 110 pb^{-1}$ of data. The top decays at CDF are identified in all the decays channels (dilepton, lepton+jets and full hadronic) and the results are summarized below.

In the dilepton channel 9 events are identified (one ee, one $\mu\mu$ and 7 $e\mu$ events) with an expected background of 2.4 ± 0.5 events. Background that contributes to the dilepton channel arise from lepton pair Drell-Yan production, diboson production (WW, WZ, ZZ), $Z \rightarrow \tau\tau$ decays $b\bar{b}$ production and background from QCD jets that fake leptons. The $t\bar{t}$ acceptance in the dilepton channel for top mass of $m_t = 175$ GeV/c^2 is estimated to be $0.74 \pm 0.08\%$. The resulting top cross section from the dilepton channel alone is $\sigma_{t\bar{t}} = 8.2^{+4.4}_{-3.4}$pb [20].

The CDF has also searched for $t\bar{t}$ events from dilepton events in the $e\tau$ and $\mu\tau$ decay modes plus at least two jets [21]. As mentioned before, the identification of hadronic τ decays is very difficult in $p\bar{p}$ collisions because the resulting τ-jets are very similar to jets from quarks or gluons. Two methods used to identify τ-jet candidates. The first method is based on the identification of "one-prong" τ decays. It requires an isolated track with large transverse momentum, $P_T \geq 15$ GeV/c, and consistency between the energy deposited in the calorimeter and the track momentum ($E/P > 0.5$). Tracks consistent with electrons or minimum ionizing particles are rejected. The second method is based on calorimetry properties of τ-jets

and accepts contribution from a larger variety of τ decay modes. The two methods identify the same 4 candidate events (two $e\tau$ and two $\mu\tau$). The total background is 1.28 ± 0.29 events for the track-based method and 2.50 ± 0.43 events for the calorimetry based method. The estimated cross sections with the calorimetry and track-based methods are $\sigma_{t\bar{t}} = 10.2^{+16.3}_{-10.2}(\text{stat})\pm1.6(\text{syst})$ pb and $\sigma_{t\bar{t}} = 29.1^{+26.3}_{-18.4}(\text{stat})\pm4.7(\text{syst})$ pb, respectively.

The search for $t\bar{t}$ events in the lepton+jets channel is the subject of this thesis. The search for $t\bar{t}$ events is pursued with three different methods. The goal of each analysis is to reduce the large background from QCD W+jet production. In order to enhance the signal to background ratio, each method requires the presence of a b-quark jet in the event. The difference between the three methods is based on the technique used to identify, *tag*, b-quark jets. In one analysis, b-jets are identified by the presence of low-P_T leptons emerging from the b semileptonic decays, $b \to c\ell\nu$ or $b \to cX \to \ell\nu X$. It is known as the *Soft Lepton Tagging (SLT)* method. The other two techniques identify B-hadrons by exploiting their long lifetime. In one method a b-jet is identified by searching for a secondary vertex significantly displaced with respect to the $p\bar{p}$ interaction vertex. It is known as the *Secondary Vertex Tagging or (SECVTX)*. The other method uses the combination of impact parameter of tracks in a jet to assign a probability the jet being consistent with the zero lifetime hypothesis. This method is known as the *Jetprobability Tagging*. Every analysis uses a common $W+$ jets sample selected by requiring a high P_T lepton (electron or muon), large transverse energy imbalance and at least three high P_T jets (E_T $\geq$15 GeV) in the rapidity region of $|\eta| \leq 2$. The sample consists of 252 $W+$ $\geq$3 jets events. The SECVTX tagging algorithm identifies 29 events with at least one b-tagged jet and the expected background is 7.99 ± 0.98 events. The derived $t\bar{t}$ cross section is $\sigma_{t\bar{t}} = 4.9 \pm 1.6$ pb.

The Jetprobability method identifies 41 $t\bar{t}$ candidate events with at least one tagged b-jet with expected background of 11.11$\pm$1.34 events. The corresponding $t\bar{t}$ cross section is $\sigma_{t\bar{t}} = 7.8 \pm 2.2$ pb. Finally, the SLT method identifies 25 $t\bar{t}$ candidate events with an expected background of 13.22$\pm$1.34 events. The resulting $t\bar{t}$ cross section is $\sigma_{t\bar{t}} = 8.9 \pm 4.4$ pb. Each cross section is calculated assuming a top quark mass of $m_t = 170$ GeV/c^2.

The results of this thesis do not represent the results of the published CDF analysis. However, this analysis uses a different treatment of the background calculations and different b-tagging efficiency in the case of the SECVTX algorithm. The Jetprobability algorithm consists a new approach in the search for $t\bar{t}$ events. For completeness, the "official" CDF results in the lepton+jets channel are presented below. The SECVTX method identifies 34 $t\bar{t}$ candidate events with expected background of 9.2$\pm$1.5 events. The corresponding $t\bar{t}$ cross section is $\sigma_{t\bar{t}} = 6.2^{+2.1}_{-1.7}$ pb. The SLT method identifies 40 $t\bar{t}$ events with an expected background contribution of 22.6$\pm$2.8 events. The corresponding $t\bar{t}$ cross section is $\sigma_{t\bar{t}} = 9.2^{+4.3}_{-3.6}$ pb.

In the full hadronic channel two method were used and both require the presence of at least 5 jets in the event. The first analysis requires the presence of at least one b-quark jet in the event and imposes strict kinematic requirements to reduce the background from QCD events. The second method requires the presence of at two least b-quark jets in the event. The first method identifies 187 $t\bar{t}$ candidate events with a background of 142$\pm$12 events while the second method identifies 157 events with expected background of 120$\pm$18 events. The resulting cross sections in the two channels are $\sigma_{t\bar{t}} = 9.6^{+4.4}_{-3.6}$ pb for the first method and $\sigma_{t\bar{t}} = 11.5^{+7.5}_{-7.0}$ pb for the second method. The combined cross section is $\sigma_{t\bar{t}} = 10.1 \pm 1.9(\text{stat})^{+4.1}_{-3.1}(\text{syst})$ pb [22].

Combining the results from all decays channels examined (excluding the hadronic

τ which are dominated by large statistical error) the $t\bar{t}$ cross section reported by CDF is $\sigma_{t\bar{t}} = 7.6\pm^{+1.8}_{-1.5}$ pb for top mass of $m_t = 175$ GeV/c^2 [23]. The quoted uncertainty includes both statistical (± 1.2 pb) and systematic effects.

The CDF has also measured the mass of the top quark in the lepton+jets and dilepton decay channels. For the lepton+jets channel a constrained fit is performed on lepton+4 jet events assuming that result from $t\bar{t} \to WbW\bar{b} \to \ell\nu qq'b\bar{b}$. The sample is divided in four subsamples ordered according to the decreasing signal to background ratio: events with 2 SVX tags, events with a single SVX tag, events with SLT tags but not SVX tags and events with no tags. The mass resolutions of each subsample is determined and then each set is fit to a combination of background and $t\bar{t}$ for different top masses. The top quark mass that maximizes the likelihood of the fit in each subsample is taken as the measurement of the top mass on this subsample. The results of the mass measurement from all the subsamples are combined by using a global likelihood function. The final measurement is $m_t = 175.9 \pm 4.8(\text{stat}) \pm 4.9(\text{syst})$ GeV/c^2 [24]. The largest contribution to the systematic uncertainty result from uncertainties in the jet energy scale and uncertainty in the knowledge of initial and final state gluon radiation. The top quark mass as measured in the dilepton channel is $m_t = 161 \pm 17(\text{stat}) \pm 10(\text{syst})$ GeV/c^2 [20].

1.5.2 Results on the top quark from D0

The results on top quark from the D0 collaboration are based on $\sim$125pb^{-1} of data collected during the Run 1 Tevatron collider run. A brief summary of the D0 results are presented in this section. Details of the different analysis can be found in several recent publications [25, 26].

The D0 collaboration observed $t\bar{t}$ events in the dilepton and lepton + jets channels.

In the dilepton channel 5 events are observed (one ee, one $\mu\mu$ and three $e\mu$ events) with expected background of 1.4±0.4 events. In the same channel, a different approach orthogonal to the standard dilepton search is also pursued. This method requires one high P_T electron, very large transverse energy imbalance ($\geq$ 50 GeV) and two or more jets in the event. The method, called the *$e\nu$ channel* contains top signal mainly from dileptons and lepton+jets top decays which fail the standard kinematic requirements. Four $t\bar{t}$ candidate events are found with expected background of 1.2±0.4 events. The resulting $t\bar{t}$ cross section based on the above two methods is $\sigma_{t\bar{t}} = 6.3 \pm 3.3$ pb for top mass of m_t = 173.3 GeV/c^2. This value corresponds to the central top mass value measured at the experiment (discussed below).

In the lepton+jets channel two different methods are pursued in order to reduce background contribution and enhance the signal to background ratio. In the first method, topological and kinematic requirements are imposed which eliminate large fraction of QCD background. This method identifies 19 $t\bar{t}$ candidate events with expected background of 8.7±1.7 events. In the second method, large background rejection is achieved by requiring the presence of a low momentum muon inside a jet signaling the presence of semileptonically decaying b-hadron. The method is analogous to the Soft Lepton tagging technique used in CDF. There are 11 candidate events with this method with an expected background of 2.4±0.5 events. The $t\bar{t}$ cross section obtained with the two methods is $\sigma_{t\bar{t}} = 4.1 \pm 2.0$ pb for the lepton+jets method and $\sigma_{t\bar{t}} = 8.2 \pm 3.5$ pb for the lepton+jets+μ tag.

The $t\bar{t}$ cross section measured by the D0 collaboration from a combination of all the decay channels is $\sigma_{t\bar{t}} = 5.5 \pm 1.4(\text{stat}) \pm 0.9(\text{syst}) \pm 0.6(\text{gen})$ pb [25].

The top mass is measured separately in the lepton+jets and dilepton channel. A constrained fit method is used similar to the reconstruction technique followed at

CDF. A top likelihood selection is applied on the data to reduce the background without introducing significant bias in the background mass distribution. The largest systematic results from the uncertainty in the jet energy and the Monte Carlo modeling of the background and signal. The measured top mass is $m_t = 173.3 \pm 5.6(\text{stat}) \pm 6.2(\text{syst})$ GeV/c^2 [26]. In the dilepton channel, the measured top mass is $m_t = 168.4 \pm 12.3(\text{stat}) \pm 3.7(\text{sys})$ GeV/c^2. Combination of the mass measured in the lepton+jets and in the dilepton channels gives $m_t = 172.1 \pm 5.2(\text{stat}) \pm 4.9(\text{sys})$ GeV/c^2.

Chapter 2

The Tools: The Tevatron and the CDF detector

The Fermilab Tevatron collider provides the highest energy collisions ever observed at an accelerator. It is a large superconducting magnet collider which collides equal energy beams of protons and antiprotons at a center of mass energy of $\sqrt{s} = 1.8$ TeV. The Tevatron layout is diagrammed in figure 2.1. There are two detectors built to study $p\bar{p}$ interactions at the tevatron. The Collider Detector at Fermilab (CDF) [30] and D0 [31] at a different location on the accelerator ring, are designed in a general purpose way, offering the potential to study thoroughly the physics processes describing the particle interactions and dynamics and the discovery of new particles and physics.

2.1 The Tevatron

The accelerator employs counter rotating beams of protons and antiprotons of energy 900 GeV each, figure 2.1). The protons and antiprotons travel in bunches inside an

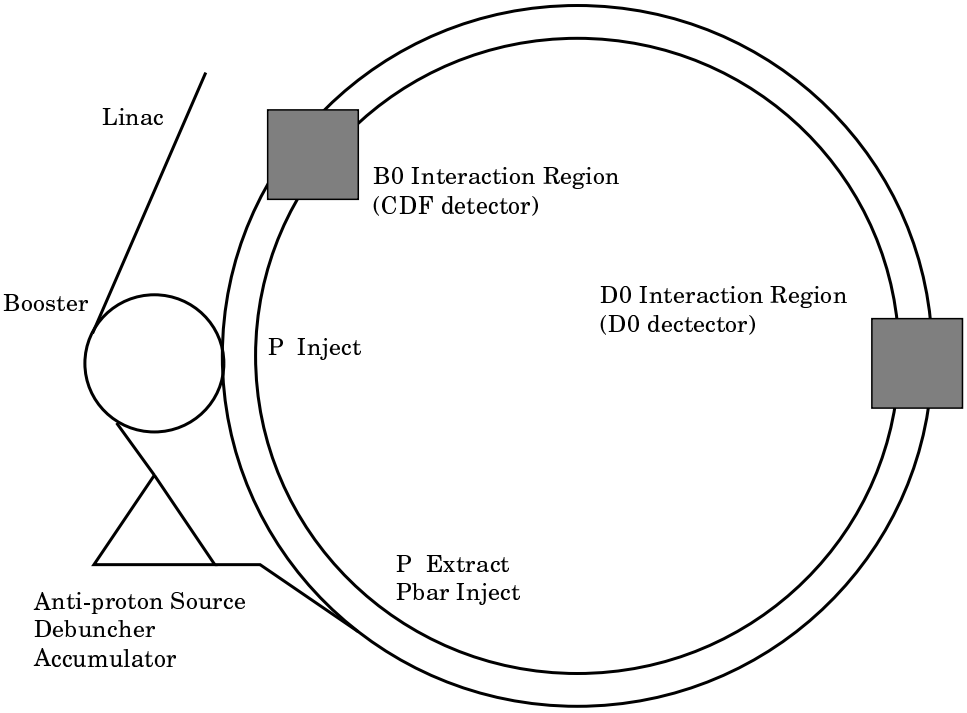

Figure 2.1: The Tevatron accelerator complex at Fermilab. The two interaction regions (CDF - B0) and D0 are also indicated.

approximately circular beampipe held at pressure of $\sim 10^{-10}$ Torr. The acceleration process is rather complex and takes place in several stages before reaching the maximum energy.

The protons used in the Tevatron come from hydrogen gas which is ionized to form H^- ions. The resulting ions are accelerated to 750 KeV in a Cockroft-Walton electrostatic generator. The H^- atoms pass through a carbon foil which strips off the electrons leaving only the protons. The remaining protons are accelerated to 400 MeV in a linear accelerator (LINAC), 150 meters in length. After the LINAC the protons are transferred to a Booster ring, a synchrotron accelerator of 475 meters in circumference, where they are accelerated to 8 GeV and formed into bunches. The proton bunches are then collected and injected into the Main Ring, which is a

synchrotron accelerator of 6300 meters in circumference. It is worth mentioning that the Main Ring was removed in 1996-1997 under the Fermilab's main injector upgrade program.

The Main Ring consists of alternating dipole and quadrupole magnets for bending and focusing of the proton beam respectively, and RF cavities that boost the proton energy to 150 GeV. When the proton bunches reach their peak energy, they are coalesced into one and are either injected in the Tevatron which lays underneath the Main Ring or sent to the fixed target experimental facilities. The whole process take place in about one minute yielding typically 10^{12} protons divided in 6 equally spaced bunches.

Protons of the Main Ring are also used for the production of antiprotons. The antiprotons are produced as secondary particles when protons from the Main Ring collided with an external Tungsten target. For this process, the protons are accelerated up to 120 GeV before extraction. The antiprotons are selected, focused through a Lithium lens and sent to the Debuncher/Accumulator. This part of the accelerator is used for *stochastic cooling* [29] of the antiprotons, a process which reduces their momentum spread, and accumulation, called *"stacking"*. The stacking process continues until there are about 6×10^{11} antiprotons accumulated. Then antiprotons are injected into the Main Ring where they are accelerated to 150 GeV, and afterwards sent to the Tevatron.

In the Tevatron, protons and antiprotons are accelerated to 900 GeV and kept in the ring by the 5.7 Tesla magnetic field of the superconducting dipole magnets. The Tevatron operates with 6 proton and 6 antiproton bunches each containing approxomately 7×10^{10} antiprotons. The revolution frequency is f = 47.7 KHz and so the period between bunch crossings at the interaction regions is 3.5 μsec. To increase the

luminosity at each interaction region, two low β quadrupole magnets at either end of the collision halls are activated. The phase of the two RF systems, which act on the protons and antiprotons separately, are adjusted, advancing or retarding the beams so their center of collision is at the center of each detector.

2.1.1 The 1992-1995 Collider Run

The last collider run, called *Run 1*, lasted from August of 1992 till July of 1993 (*Run 1A*) and from January of 1994 till July of 1995, (*Run 1B*). The total amount of data collected during the course of the run corresponds to an integrated luminosity of $\int \mathcal{L}$ of $109.4 \pm 7.2 pb^{-1}$.

The integrated luminosity, $\int \mathcal{L}$, depends on the instantaneous luminosity, $\mathcal{L}$, of the collider. The instantaneous luminosity for a $p\bar{p}$ collider is given by [8] and [27]:

$$\mathcal{L} = \frac{f B N_p N_{\bar{p}}}{4\pi\sigma^2}$$

where the parameters are:

- f: Revolution frequency
- B: Number of proton and antiproton bunches (B = 6)
- N_p: Number of protons per bunch (typically 2×10^{11})
- $N_{\bar{p}}$: Number of antiprotons per bunch (typically 7×10^{10})
- σ: The transverse cross sectional area of each bunch (typically $5 \times 10^{-5} cm^2$)

The highest luminosity achieved was $2.8 \times 10^{31} / cm^2 sec$ during Run 1B data taking. Typical starting instantaneous luminosities achieved were $1.6 \times 10^{31} / cm^2 sec$ and $0.54 \times$

$10^{31}/cm^2sec$ for Run 1A and Run 1B respectively. Due to transverse spreading of the beam and losses from collisions the luminosity falls exponentially. Taking into account the total inelastic cross section of $\sim$ 50 mb and a luminosity of $1.6 \times 10^{31}/cm^2sec$, the average interaction rate at the Tevatron is $\sim$ 1 MHz.

2.2 The CDF Detector

The CDF detector is a general purpose detector, built to analyze the $p\bar{p}$ collisions produced at the tevatron. The detector is capable of measuring the momentum and the energy deposition of particles emerging from the interaction region with good tracking and calorimeter resolution. An isometric view of the CDF detector is shown in Figure 2.2.

The proton antiproton beams collide roughly at the center of the CDF detector. At the interaction region the beam is approximately circular in cross section, with a radial spread of $\sim 40\mu$m and has a Gaussian longitudinal profile with a standard deviation of $\sim$ 30 cm. A penetrating particle produced at the interaction point encounters in succession a beryllium beam pipe, three tracking chambers, a superconducting solenoid, calorimeter cells and muon chambers. The Silicon Vertex detector (SVX) surrounds the interaction region and provides information about the track impact parameter and displaced vertices. Following the SVX, there is the Vertex Time Projection detector (VTX) which reconstructs the z position of the $p\bar{p}$ interaction. The Central Tracking Chamber (CTC) surrounds the previous two detectors and provides particle tracking information. The three tracking detectors are inside an axial magnetic field of 1.412 Tesla which provides the appropriate curvature for charged particle momentum measurement. The solenoid is surrounded by the electro-

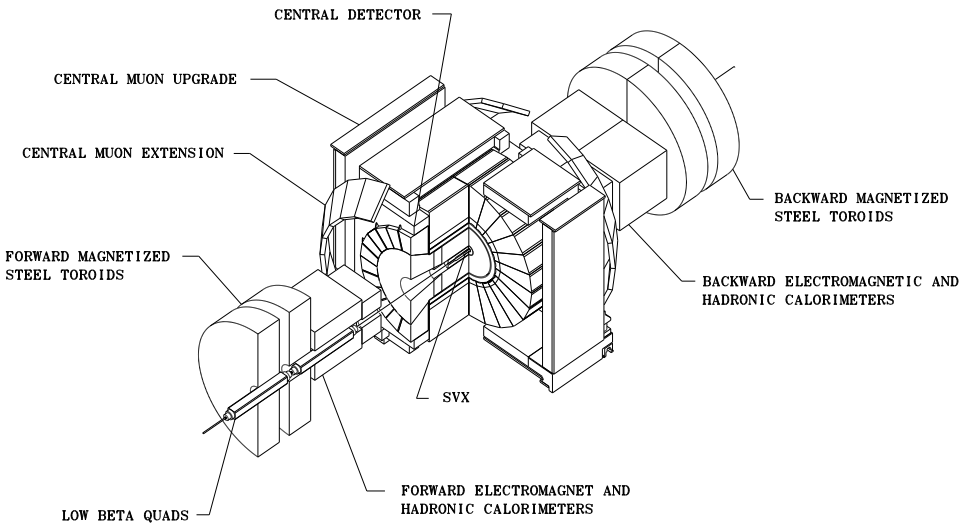

Figure 2.2: Isometric View of the 1992 CDF detector.

magnetic and hadronic calorimeters which measure the energy of interacting particles or jets. Furthest from the beam line and shielded by thick steel plates are the muon chambers which are used for the detection of muons. A cross sectional view of the detector is shown in figure 2.3. All the detector components are described in details elsewhere [30]. Here we provide a brief description of the components relevant in this analysis.

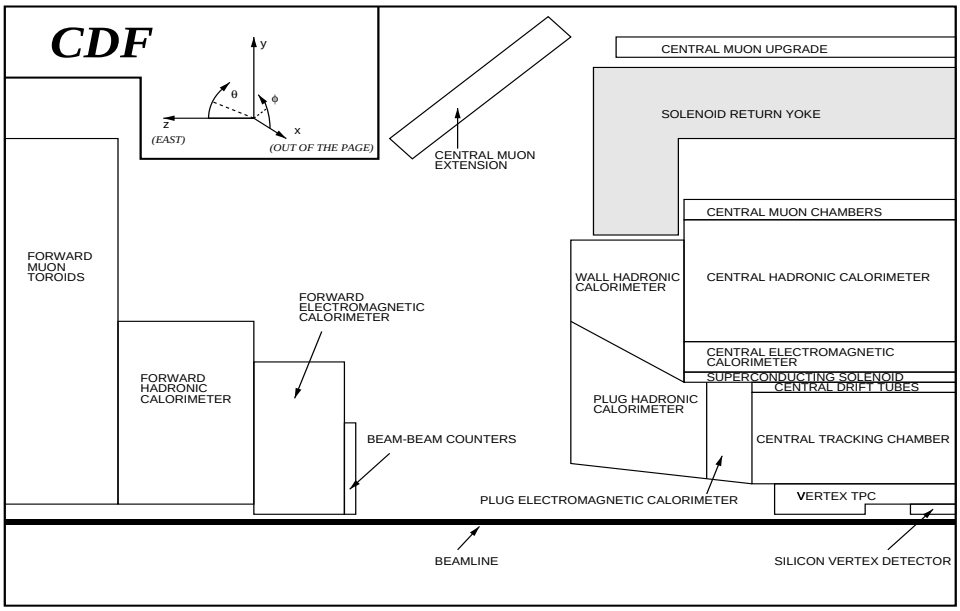

Figure 2.3: A cross sectional view of a quadrant of the CDF detector. The detector is at forward-backwards symmetric about the interaction region (lower right corner of the figure). The CDF co-ordinate system is described in the upper left corner of the figure.

2.2.1 CDF co-ordinate system

The origin of CDF co-ordinate system is at the geometrical center of the detector. The positive z-axis is pointing along the beamline in the direction of the protons. The positive y-axis is normal to the accelerator plane pointing upwards while the positive x-axis is normal to the other two axes with positive direction as to define a right handed co-ordinate system. In order to describe the particle kinematics or detector location, most often spherical co-ordinates are used. The angular co-ordinate ϕ is the azimuthal angle about the z-axis and $\phi = 0$ is the positive x-axis. The second angular co-ordinator is the polar angle, θ, from the z-axis and goes to zero at the positive z-axis direction. Instead of using the angle θ, a new variable is used, *pseudorapidity*

η, which is connected to the polar angle with the relation:

$$\eta \equiv -\ln \tan \frac{\theta}{2}$$

The pseudorapidity, η, is equivalent to the *rapidity, y*, of a particle in the limit of p $\gg$ m, where p is the momentum of the particle and m its mass. The rapidity, y, is defined by:

$$y \equiv \frac{1}{2} \ln \frac{E + p_z}{E - p_z}$$

Under a boost in the z-direction to a frame with a velocity β, rapidity transforms as $y \rightarrow y + \tanh^{-1} \beta$. Therefore the rapidity distribution, dN/dy, is invariant under Lorentz transformations along the z direction.

The CDF calorimetry is divided up into *projective towers*, angular segments in η and ϕ which point back to the origin of the co-ordinate system.

As already mentioned, not all interactions take place at z=0. The spread of the position of the interaction vertex implies that the η and ϕ position of the towers with respect to the interaction vertex is not fixed. Two terms are used to specify the calorimetry location. One called *detector* η, which gives the location in the calorimeter with respect to the origin, and the other the *event* η, which gives the location in the calorimeter with respect to interaction vertex.

In CDF the transverse momentum and energy of particles, called P_T and E_T, are directly measured. While the transverse momentum, P_T, is extracted from the curvature of the particle's track in the magnetic field, the transverse energy E_T is directly related to the knowledge of the polar angle θ of the calorimeter tower with respect to the interaction vertex. Due to the high luminosity conditions during the period of Run 1, more than one $p\bar{p}$ interaction can take place in the same beam

crossing. The average number of interactions per crossing is ~ 2.8 for typical instantaneous luminosity value of $\mathcal{L} = 1.6 \times 10^{31}/cm^2 sec$. A careful determination of the interaction vertex is essential for the correct measurement of the energy and location of calorimeter towers contributing to the formation of jets and measurement of the energy attributable to non-interactive particles which escape detection.

2.2.2 The Tracking System

CDF employs three systems for tracking of charged particles: the SVX, the VTX and the CTC.

The Silicon Vertex Detector (SVX)

The Silicon Vertex Detector (SVX) is the detector closest to the beam pipe and provides excellent position resolution which is useful for heavy flavor identification. The SVX is a silicon microstrip detector installed during Run 1A. Due to radiation damage, the SVX was replaced by a similar detector, SVX', for Run 1B. A brief description of both devices is given below. The name convention is used for both detectors and where necessary the indication of SVX or SVX' will be used to indicate differences between the two. A detailed description for both detectors can be found in References [32, 33].

The SVX consists of two identical barrels, aligned along the beam line. The barrels are facing each other at z=0, with a gap of 2.15 cm between them. An isometric view of the SVX detector is shown in figure 2.4. The total active length of the detector is 51 cm. Because of the longitudinal spread of the $p\bar{p}$ interactions along the beam line, roughly 60% of all interactions take place in the SVX fiducial volume. The pseudorapidity coverage of the SVX is $|\eta| \leq 1.9$.

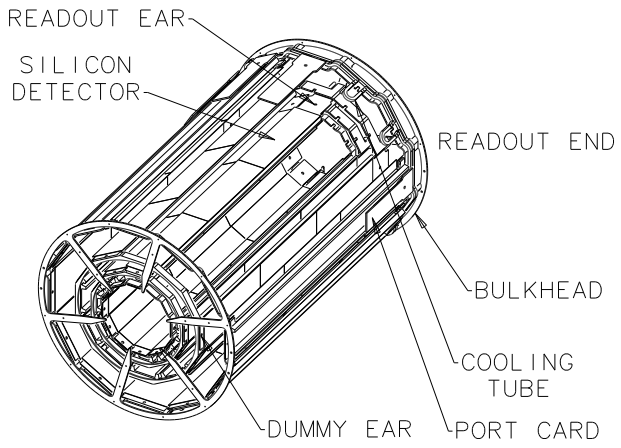

Figure 2.4: An isometric view of an SVX barrel. The detector consists of two such barrels laid end to end, at z=0.

Each SVX barrel is divided into 12 azimuthal wedges of 30° each, and into 4 concentric radial layers of silicon strips. Every layer in turn contains 12 *ladders*, each 25.5 cm in length. Figure 2.5 shows an SVX ladder. Each ladder composed of three DC-coupled (for SVX) and AC-coupled (for SVX') single sided silicon microstrip sensors *wafers*, each 8.5 cm long. The wafers are wired-bounded end-to-end to form the primary mechanical and electrical units. Each ladder is rotated 3° about its longitudinal axis in order to provide overlap between adjacent ladders. The readout strips of the silicon are aligned parallel to the barrel axis. The pitch of the readout strips is 60 μm in the inner three layers and 55μm for the outermost layer. The readout end of the ladder is microbounded to an off-board custom thick-film hybrid referred as an "ear". The ear carries 2,3,4 or 6 SVXD (for SVX) or SVXH3 chips (for SVX') for layers 0 to 3 respectively. The SVXH3 chip provides 20% less noise and 30% more gain than the SVXD chip. A copper-kapton film at the end of the ear,

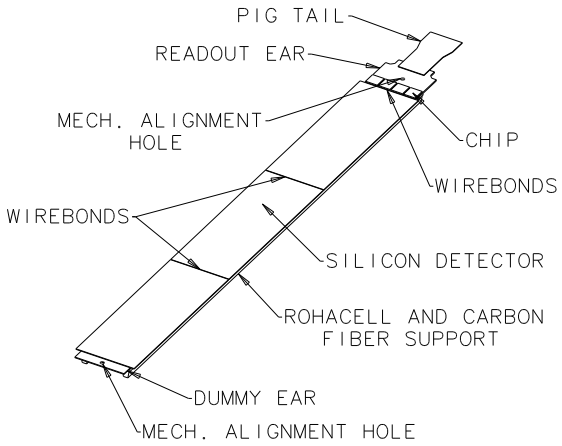

Figure 2.5: Schematic view of an SVX ladder.

called *pig tail*, provides the connection of the chips to the data acquisition system.

The radius of each layer is the same for both detector except the innermost one. The radius of layer 0 is 3.005cm for SVX and 2.861cm for the SVX'. The radii of the rest of the layers are 4.256, 5.687 and 7.866cm for layers 1 to 3 respectively. The change in the design specification for layer 0 in addition to a 1° rotation between the adjacent ladders of layer 0 for the SVX', was implemented to improve the ϕ coverage by eliminating 1.26° gap present in the initial detector. The first layer is positioned as close as possible to the beam in order to provide the best possible measurement of the track impact parameter. The middle layers (1 and 2) provide redundancy near the high track density and high radiation environment. Layer 3 provides good matching between the outer tracking systems and the vertex detector.

Because of the strip geometry, the SVX detector provides only 2-D information in the $r - \phi$ plane. The readout is done in *sparse mode*, i.e only those strips that are significantly above threshold are read out. There are 12 channels attached on every chip bringing the total number of readout channels to 46080 for the whole detector. The DC-coupled SVX necessitates that the device be operated in a quadruple sample

and hold mode. In this mode, the charge is integrated twice (once "on beam" and once "off beam") and the difference of these integrations give the signal used for further processing. This is done in order to eliminate the baseline shift due to the varying strip-to-strip leakage current (typically less than 2 nA) which would also be integrated during the sampling time. However since SVX' is AC-coupled the leakage current is not seen by the readout electronics, hence the SVX' operates in a double sample and hold mode. In this case there is need of one only integration at "on beam" time. This results in the reduction of the SVX' noise by $1/\sqrt{2}$ compared to the SVX detector. The signal-to-noise ratio for the SVX detector was 9.5 and 16.0 for the SVX', during the time of installation. Typical readout time is $\sim$ 2 ms, the longest among the components of the CDF detector.

The conversion of the raw charge from individual strips to track hits is performed after a channel by channel pedestal subtraction followed by a clustering algorithm. In the clustering, data from contiguous strips are grouped together forming a cluster. The threshold for cluster formation depends on the noise of each strip and the number of non-dead strips in the group. If a strip fails the threshold requirement the algorithm splits the cluster and the procedure is repeated in each remaining group. The cluster position is calculated as a charge weighted centroid using the charge and center position of the individual strips in the cluster. The spatial resolution of the cluster is assigned based on the number of strips and the total charge. For clusters with more than 3 strips or total charge greater than 11.7 fc, the position error is calculated according to (strip pitch$\times$number of strips)$/\sqrt{12}$. For clusters with 3 or fewer strips, the resolution is obtained from data by studying the residual distributions of the final track fits. The position resolution of an individual SVX hit has been measured from data and determined to be approximately $\sim 15\mu m$ in the transverse

plane and the hit finding efficiency is $\sim 98\%$ per layer.

Combining SVX and CTC information yields impact parameter resolution asymptotically approaching $15\mu m$. Figure 2.6 shows the distribution of the impact parameter resolution of the SVX' detector.

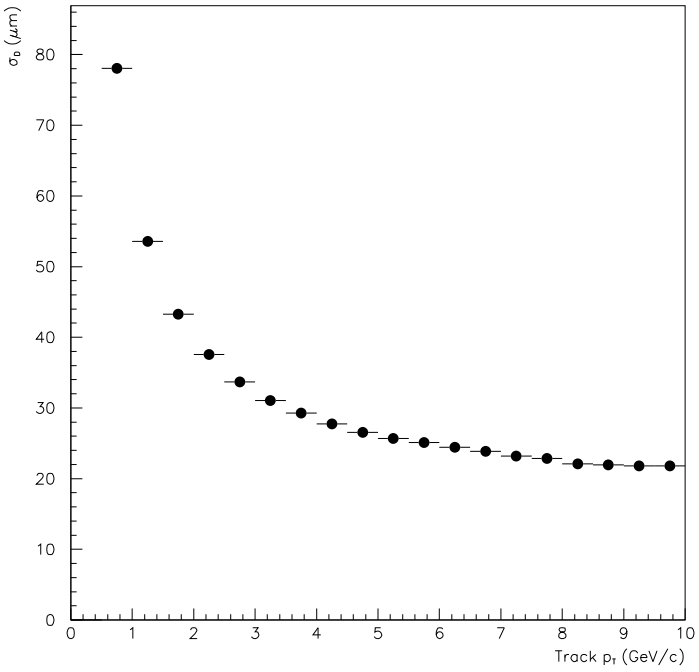

Figure 2.6: The SVX impact parameter resolution as a function of the P_T of the tracks. For small transverse momentum tracks the resolution is dominated by multiple scattering which scales like $1/P_T$. For higher momentum tracks the effect of multiple scattering decreases and the resolution approaches asymptotically the detector's intrinsic resolution. The plotted distribution includes also the primary vertex resolution.

The Vertex Time Projection Chambers (VTX)

The VTX is an Argon-Ethane time projection chamber composed of 8 modules and surrounds the SVX detector. Each module is octagonal, segmented into 8 wedges. The endcaps of the wedges are segmented into two sets of wires, one running perpendicular to the beam and the other one being perpendicular to the radial centerline of the wedges. Hits are produced on the sense wires from electrons produced by the ionization of the Argon-Ethane when a charged particle passes through the detector. By measuring the drift times, the position of a track can be reconstructed in the r-z plane. Combining all the tracks reconstructed in the VTX, the z position of a vertex can be determined. The resolution of the detector along the z-direction is 1-2 mm depending on the number of tracks in the event. The VTX provides z vertex information for pseudorapidities up to $|\eta| < 3.25$. The vertex information provided by the VTX is used by the CTC for full 3-D reconstruction of tracks.

The Central Tracking Chamber (CTC)

The CTC detector is the last tracking system used in CDF and conventionally the most important since it provides the full 3-D tracking information. It is a 3.2 m long cylindrical drift chamber contained inside the CDF's superconducting magnetic solenoid. Radially the chamber extends from 0.3 m (inner radius) to 1.3 m (outer radius). The chamber consists of 84 layers of sense wires grouped into 9 *superlayers*, 5 axial and 4 stereo. Figure 2.7 shows the CTC wire arrangement as seen in the r-ϕ plane. Each axial superlayer consists of 12 sense wires which run in the direction parallel to the beam line. The stereo superlayers consist of 6 sense wires each, tilted with respect to the axial wires by $\pm 3^{\circ}$. While the axial superlayers provide r-ϕ information the addition of the stereo superlayers allow full 3-D track reconstruction.

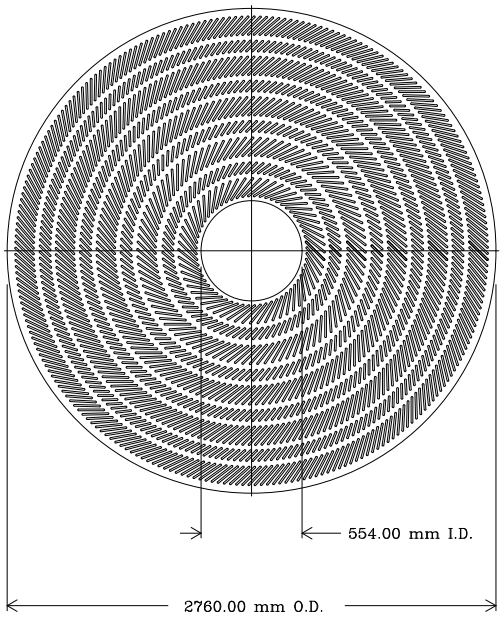

Figure 2.7: The superlayer structure of the central tracking detector.

Both axial and stereo superlayers are divided into cells in which the field wires form a drift field of 1350 V/cm, so that the maximum drift distance is less than 40mm corresponding to a maximum drift time of 800ns. Each cell is tilted by 45° with respect to the radial direction to compensate for the Lorentz angle β from the crossed E and B fields and make the drift direction azimuthal. $\tan\beta = \frac{u(E,B=0)B}{KE}$, where $u(E, B = 0)$ is the drift velocity (in m/s) in the absence of the magnetic field, E is the electric field strength (in V/m) and B is the magnetic field strength (in T). The parameter k depends on the gas mixture and is approximately 0.7 for the gas mixture used in the operation of the CTC.

Each sense wire is read out through a pre-amplifier whose analog output is fed into an amplifier shape discriminator (ASD). The ASD differential ECL signals are fed into multiple hit Fastbus Lecroy 1879 TDCs (time to digital converters).

The r-ϕ, r-z hit coordinates can be determined within $200\mu m$ and 6mm respectively. The track momentum is given by measuring the curvature of the track helix. The track P_T is determined with a resolution of $\frac{\delta P_T}{P_T^2} < 0.002$ for $|\eta| \leq 1.0$. The double track resolution is of the order of 3.5mm determined by the TDC pulse width. Tracks with $P_T \leq 400$ GeV/c are too soft and curl inside the magnetic field while tracks of $P_T \geq 1$ GeV/c can be reconstructed with an efficiency of 98%.

Combining CTC and SVX information the momentum resolution for tracks with $|\eta| \leq 1.0$ is determined to be:

$$\frac{\delta P_T}{P_T} = \sqrt{(0.0009 P_T)^2 + (0.0066)^2}$$

where P_T has units of GeV/c. The track reconstruction efficiency drops fast for $|\eta| \geq 1$, while there is no tracking information for $|\eta| \geq 1.7$.

2.2.3 The calorimeter

The solenoid and the tracking are surrounded by the CDF calorimeter. The calorimeter at CDF has 2π coverage in azimuth and extends up to $|\eta|$ <4.2 in pseudorapidity. The CDF calorimetry is divided up into *projective towers*, angular segments in η and ϕ which point back to the origin of the co-ordinate system. Figure 2.8 shows the calorimetry $\eta - \phi$ segmentation map for a quadrant of the detector. The calorimeters are divided into three major regions according to their pseudorapidity coverage. The *Central* rapidity region, $|\eta| < 1.1$ is based on scintillator calorimetry and the more forward regions on gas. The forward region consists of the *Plug* which covers the region $1.1 < |\eta| < 2.4$, and the *Forward* which covers the region from $2.4 < |\eta| < 4.2$. The absorber for all hadronic calorimeters is iron while the absorber for all electromag-

netic calorimeters is lead. Some of the most important parameters of the calorimeters are summarized in Table 2.1.

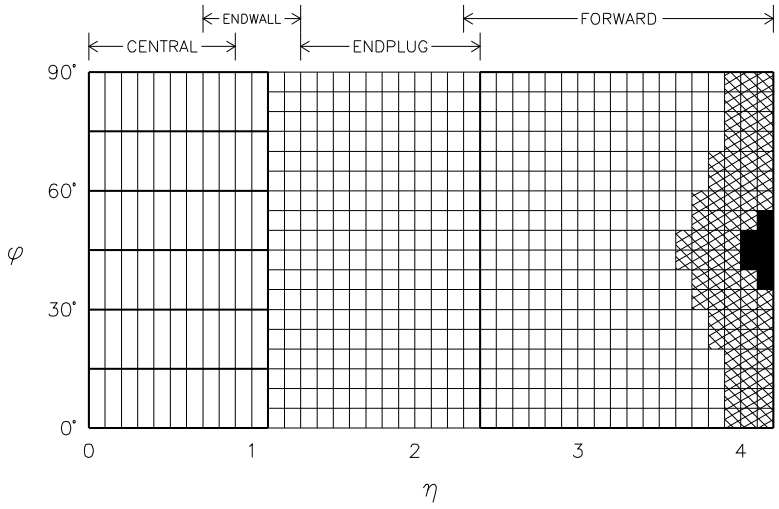

Figure 2.8: $\eta - \phi$ calorimeter segmentation map. The thick lines indicate module or chamber boundaries. The cross-hatched area has only partial hadron depth coverage because of the space required for the low beta quadrupoles. The shaded area has no coverage.

Each of the three regions has an electro-magnetic calorimeter, namely CEM, PEM, FEM, followed by a hadronic calorimeter, namely CHA/WHA, PHA, FHA. This design is dictated by the fact that electrons shower earlier in the calorimeter than the hadrons. This also helps for particle identification since electrons and photons deposit practically all their energy in the electro-magnetic calorimeter with small leakage in the hadronic compartment. Comparing the energy deposited in the electro-magnetic and hadronic towers, electrons/photons can be distinguished from hadrons.

System	η Range	Energy Resolution	Thickness
CEM	$\lvert\eta\rvert < 1.1$	$\frac{13.7\%}{\sqrt{E_T}} \oplus 2\%$	18 X_0
PEM	$1.1 < \lvert\eta\rvert < 2.4$	$\frac{22\%}{\sqrt{E}} \oplus 2\%$	18-21 X_0
FEM	$2.2 < \lvert\eta\rvert < 4.2$	$\frac{26\%}{\sqrt{E}} \oplus 2\%$	25 X_0
CHA	$\lvert\eta\rvert < 0.9$	$\frac{50\%}{\sqrt{E_T}} \oplus 3\%$	4.5 λ_0
WHA	$0.7 < \lvert\eta\rvert < 1.3$	$\frac{75\%}{\sqrt{E}} \oplus 4\%$	4.5 λ_0
PHA	$1.3 < \lvert\eta\rvert < 2.4$	$\frac{106\%}{\sqrt{E}} \oplus 6\%$	5.7 λ_0
FHA	$2.4 < \lvert\eta\rvert < 4.2$	$\frac{137\%}{\sqrt{E}} \oplus 3\%$	7.7 λ_0

Table 2.1: Summary of CDF calorimeter properties. The symbol $\oplus$ signifies that the constant term is added in quadrature in the resolution. Energy resolutions for the electro-magnetic calorimeters are for incident electrons and photons, and for the hadronic calorimeters are for incident isolated pions. Energy is given in GeV. Thicknesses are given in radiation lengths (X_0) and interaction lengths (λ_0) for the electromagnetic and hadronic calorimeters, respectively.

Central Calorimeters

The central and wall calorimetry, including the central electro-magnetic (CEM), the central hadronic (CHA) and the endwall hadronic (WHA) calorimeters cover the pseudorapidity range from about -1.1 to 1.1. The central calorimeters are segmented in $\eta - \phi$ towers of dimensions $\delta\phi = 15^\circ$ in ϕ and approximately $\delta\eta = 0.1$ in η. The CEM covers the η region from -1.1 to 1.1, the CHA from -0.7 to 0.7 and the WHA covers the region $0.7 < |\eta| < 1.2$. The CEM and CHA are packaged together and form a barrel of 1.3 meter inner and 2.2 meters outer radius respectively, around the beampipe. The central calorimetry is divided into four large units, called *arches*, east and west, north and south, that make-up the barrel. Each arch is divided in 12 azimuthal segments called *wedges*, making 48 in all. Each wedge contains 10 η-towers

for a total of 20 towers.

All three calorimeters use scintillators as their active material with good energy resolution (see Table 2.1). This is important in the central rapidity regions since the total energy E is practically the same as the $E_T = E \sin\theta$ ($\sin\theta$ close to 1.0) and their respective resolutions are comparable. In the plug and forward regions, the same E_T corresponds to a much higher E and thus a better E_T resolution.

The central electro-magnetic calorimeter consists of 31 layers of 5 mm thick scintillator made of polysterene, sandwiched by lead sheets of 0.32 cm thickness which act as absorber. The CEM extends from 173 cm at inner radius to 208 cm at outer radius. The total amount of the CEM material is approximately 18 radiation lengths. The energy resolution for the electro-magnetic showers is

$$(\frac{\sigma}{E})^2 = (\frac{13.7\%}{\sqrt{E_T}})^2 + (2\%)^2$$

where the energy, E, is measured in GeV. A schematic drawing of a single electromagnetic wedge is shown in figure 2.9.

Between the solenoid and CEM there is a set of proportional chambers employed in run 1 to aid in distinguishing photons which can interact in the solenoid coil from hadrons which are less likely to interact. The chambers which serve as a preradiator, *CPR*, consist of sense wires separated by 2.2 cm and operates in Argone-Ethane gas. There are 2 chambers in η for each 15° wedge of CEM.

Proportional chambers with strip and wire readout (CES) are located at a depth of six radiation lengths in the CEM calorimeter (184 cm from the beam pipe), corresponding to approximately the shower maximum of the electromagnetic showers. The CES provides both z and r-ϕ position of the electromagnetic shower and it provides a

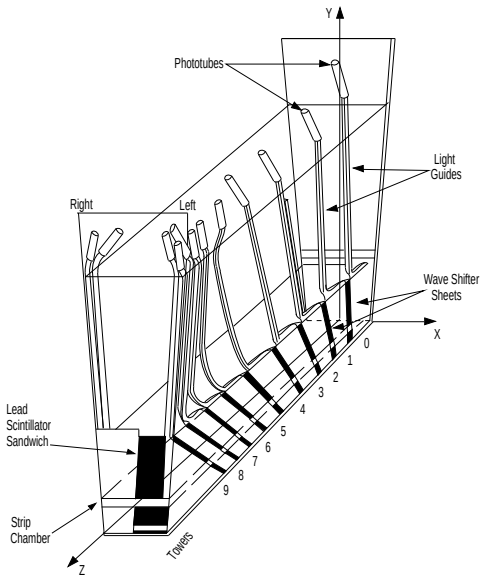

Figure 2.9: A schematic drawing of central electromagnetic calorimeter wedge. Each wedge contains 10 towers in η each covering 0.1 units of pseudorapidity. The wedge covers 15° in azimuth. Strip and wire proportional chambers are inserted at the expected shower maximum for electrons.

handle for distinguishing moderate energy single photons from photons produced in π^0 decays ($\pi^0 \rightarrow \gamma\gamma$) and can result to two significantly overlapping electromagnetic showers.

The central hadronic calorimeter modules are in the same wedge as the CEM, following the CEM. Their active area consists of 30 layers of iron-scintillator sandwich (corresponding to a total of 6 interaction lengths) The energy resolution of the CHA was measured with single isolated pions and is

$$(\frac{\sigma}{E})^2 = (\frac{50.0\%}{\sqrt{E_T}})^2 + (3\%)^2$$

where as before the energy is measured in GeV. The Wall Hadron calorimeter (WHA) resembles the CHA but its modules are located into the yoke of the magnet and are

part of the return flux path of the solenoid. It consists of steel-scintillator sandwich. The energy resolution of the WHA as measured with testbeam pions is found to be:

$$(\frac{\sigma}{E})^2 = (\frac{75.0\%}{\sqrt{E_T}})^2 + (4\%)^2$$

Gas Calorimeters: The Plug and the Forward

The gas calorimeters cover the $|\eta|$ range from 1.1 to 4.2 corresponding to polar angles of 30° to 2° respectively. Besides some small differences, plug and forward calorimeters are essentially the same. Both detectors use gas sampling calorimetry of 50/50 Argone-Ethane gas mixture with a small percentage of isopropyl alcohol mixture, with proportional tubes and cathode pad readout. The alcohol admixture is used to prevent the development of a continuous glow discharge when a large dose of radiation illuminates a small region. The calorimeters are subdivided into towers each covering 5° in ϕ and 0.1 in rapidity.

The Plug electro-magnetic calorimeter (PEM) consists of 34 layers of proportional tube planes separated by 2.7 mm of lead. The proportional tubes are constructed from resistive plastic and are epoxied to sheets of copper-clad G10 which forms the cathode. The cathode is etched into pads which form the basis of the towers. Each plane is divided into 4 quadrants with tubes running vertically. Projective tower layers are ganged into three groups corresponding to three different depth segments. As a charged particle passes through, the gas mixture is ionized producing a cloud of slowly moving positive ions and high mobility electrons. The electrons are collected fast on the wire while the positive ions induce a charge on the cathode pad which is connected to an integration circuit. The charge collected on the cathode pad gives a measure of the energy of the particle. The energy resolution of the plug

electromagnetic calorimeter was measured with testbeam electrons to be

$$(\frac{\sigma}{E})^2 = (\frac{22.0\%}{\sqrt{E}})^2 + (2\%)^2$$

The plug hadronic calorimeter has 20 gas sampling layers separated by 5 cm of steel, corresponding to about 6.5 interaction lengths of material. The tubes and corresponding wires lie along the ϕ direction in twelve 30° wedges. The pad signals of all 20 layers in the same tower are ganged together. The energy resolution of PHA was determined with testbeam pions to be

$$(\frac{\sigma}{E})^2 = (\frac{90.0\%}{\sqrt{E}})^2 + (4\%)^2$$

The forward electromagnetic (FEM) and hadronic (FHA) calorimeters cover the polar angle region from 2° to 10° at both ends of the CDF detector. Both sides are divided azimuthally into four quadrants. The FEM consists of 30 layers of gas proportional tube planes interleaved with 4.5 mm thick lead sheets. The tower size is also $\delta\eta = 0.1$ and $\delta\phi = 5°$. The energy resolution of FEM was determined with testbeam electrons to be

$$(\frac{\sigma}{E})^2 = (\frac{26.0\%}{\sqrt{E}})^2 + (2\%)^2$$

The Forward hadronic calorimeter covers the same angular region as its electromagnetic partner and consists of 26 layers of chambers separated by 5 cm thick steel plates. The FHA energy resolution was measured with testbeam pions and it is

$$(\frac{\sigma}{E})^2 = (\frac{137.0\%}{\sqrt{E}})^2 + (4\%)^2$$

2.2.4 The Central Muon System

The CDF muon system consists of four separate detectors. The central muon chambers (CMU) and its complementary system, the central upgrade chambers (CMP), cover the angular region corresponding to $|\eta| \leq 0.6$. The central muon extension chambers (CMX) cover the region $0.6 \leq |\eta| \leq 1.0$. The last two detectors were installed before run 1A with main purpose to increase the acceptance for muons and reduce the background of the CMU muons due to penetrating pions. This is a major background for the identification of muons resulting from semileptonic b-decays where the lepton is inside the b-jet. The fourth muon system is the forward muon spectrometers (FMU) which cover the region of $2.0 < |\eta| < 3.6$. Each spectrometer consists of a magnetized steel toroid with three layers of drift chambers sandwiched by two layers of scintillators.

The CMU and CMP

The CMU consists of four layers of drift chambers located outside the CHA. The CHA acts as a hadron absorber for the CMU. Muons with $P_T \geq 1.5$ GeV/c and $|\eta| < 0.6$ are detected by the CMU. Figure 2.10 shows the location of the CMU chambers in the central calorimeter wedge.

There are 3 CMU towers for every central calorimeter wedge. Each tower contains four layers of tubes. Each tube contains a sense wire running parallel to the beam line and with length equal to the central calorimeter wedge. Figure 2.11 shows the arrangement and definition of single CMU tower. Sense wires from alternating tubes are ganged together and are offset from the other pair by 2 mm in order to resolve the left right ambiguity. The z co-ordinate of the stub is determined by charge division along the wire. The ϕ co-ordinate is determined by the drift time. Hits on the

muon chambers are reconstructed to form a *stub* and are matched to CTC tracks to determine the momentum of the muon. The CMU system has a spatial resolution of 1.2 mm in z and 0.25 mm in ϕ.

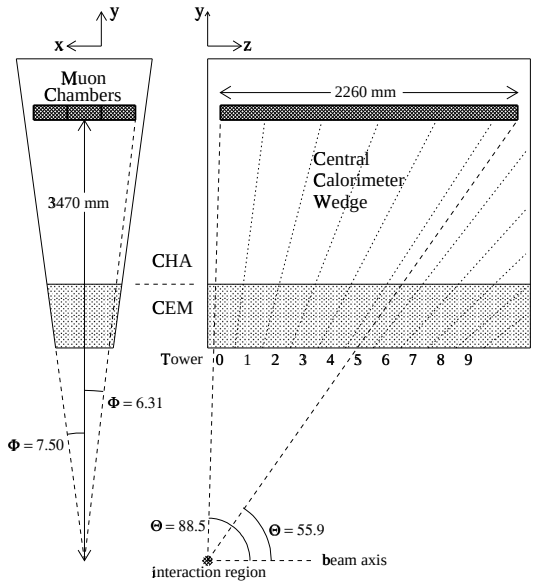

Figure 2.10: Location of the CMU detector in the central calorimeter wedge seen in an azimuthal and polar slice of the wedge.

Because of the small interaction length of the material between the interaction vertex and the muon chambers ($\sim$ 5.4 pion interaction length), a significant amount of charged pions ($\sim$ 0.5% [36]) can penetrate the hadron calorimeter without interacting and produce stubs on the CMU chambers, called *punch through*. To reduce the background from hadronic punch through, three pion interaction lengths of steel (60 cm of steel) were added at both sides of the detector while the return yoke's steel which is also of three pion interaction lengths, was used as an absorber.

Four layers of drift tubes were mounted behind these iron walls for muon identifi-

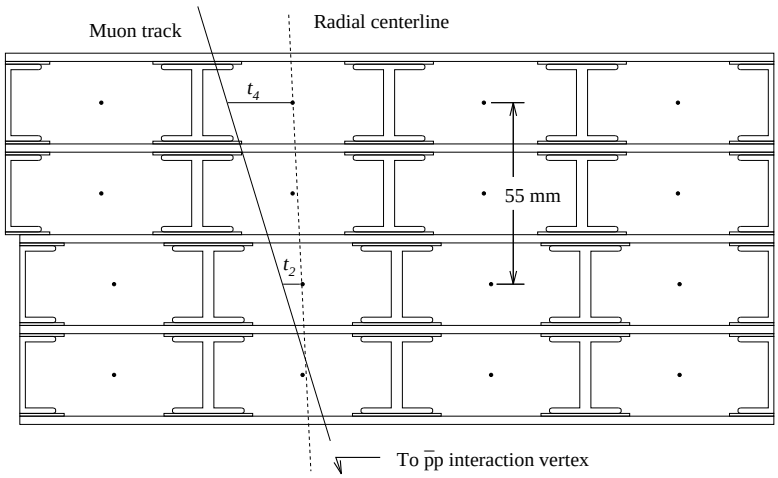

Figure 2.11: Single CMU tower. A tower consists of radial layers of 4 drift tubes. Wires on even and odd layers are radially aligned forming 2 pairs per tower. The pairs are offset to each other by 2 mm to resolve left-right ambiguities in the reconstruction of the muon stub.

cation (CMP). The additional steel reduces the amount of punch through by a factor 20 while it reduces the amount of background due to decays in flight of π's and K's when they decay to low P_T muons.

The CMP chambers are single wire rectangular cells 640 cm long and 2.5 cm wide $\times$ 15 cm height made of aluminum extrusion. The chambers operate with a 50/50 mixture of Argone-Ethane. Four layers of chambers are glued together forming a four layer "stack". Each stack is mounted behind the additional steel absorber, with its length parallel to the beam line. The tubes in alternate layers are half cell staggered to resolve left right ambiguities and diminish the azimuthal gaps. CMP chambers do not offer 2π azimuthal coverage. The break-down of the azimuthal coverage of the $|\eta| < 0.6$ region by the CMU and CMP chambers is approximately 53% for CMP and CMU, 84% by CMU and 63% by the CMP. Figure 2.12 shows the overlap between the CMU and CMP coverage.

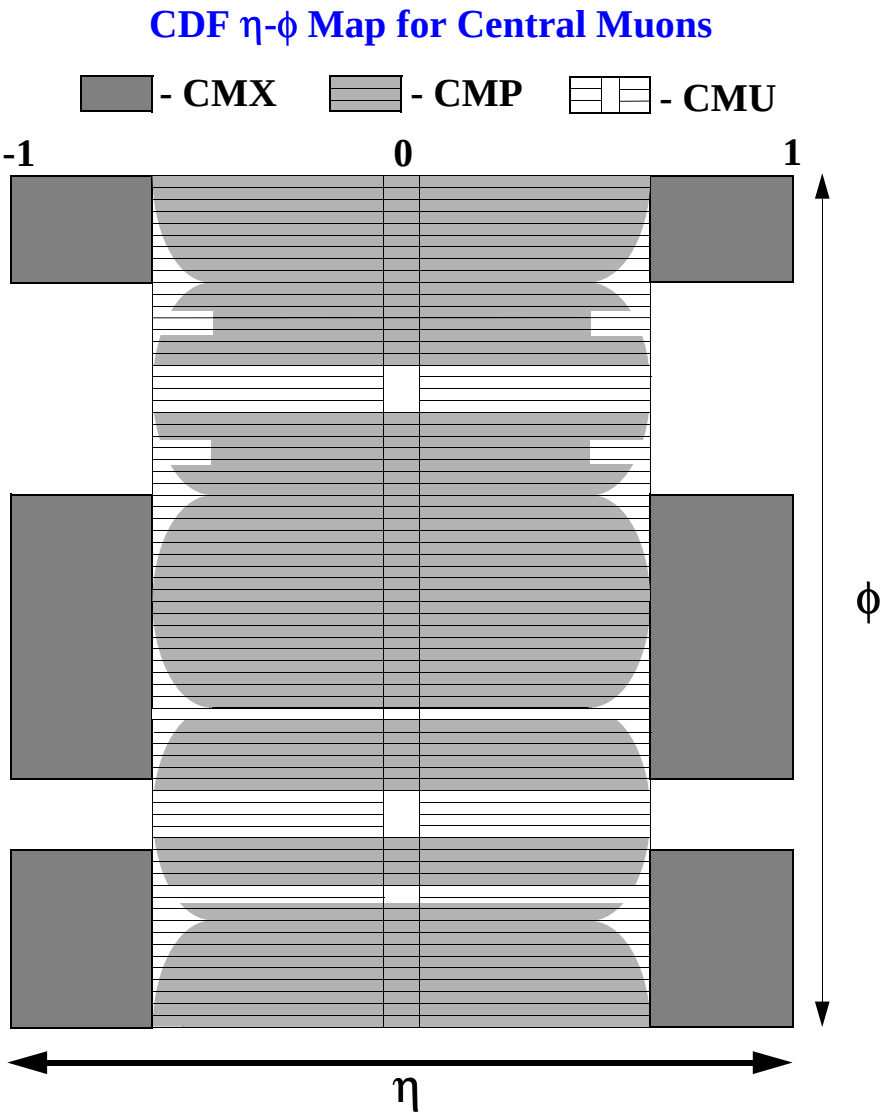

Figure 2.12: Schematic view of the $\eta - \phi$ coverage of the CMU and CMP and CMX muon detectors.

The Central Muon Extension (CMX)

The Central Muon Extension (CMX) extends the central muon coverage in the region of $17^\circ < \theta < 55^\circ$ corresponding to $0.65 < |\eta| < 1.0$ and spans 240° in azimuth.

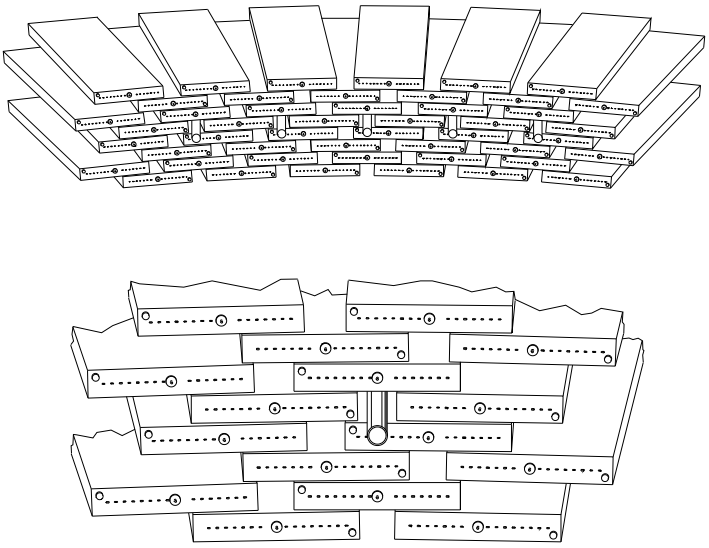

Figure 2.13: Schematic drawing of a CMX wedge. Each wedge consists of 48 drift tubes spanning 15° in azimuth and is mounted on a free standing arch which cover the region of $0.6 < |\eta| < 1$ and 240° in azimuth. The CMX detector consists of 4 such arches.

The CMX chambers are built similar to the ones used in the CMP detector. The chambers operate in the limited streamer mode in a mixture of 50/50 Argone-Ethane gas. The chambers are epoxied together in groups of 8 layers each containing 6 cells and spanning 15° in azimuth. Successive layers are half cell staggered while the upper and lower layers have the same layout. This constitutes a *CMX wedge*. The 8 layers are divided into 4 groups of radially aligned pairs while adjacent pairs are offset by half cell. The positioning of the cells on a wedge is shown in figure 2.13. Each wedge

consists of 48 tubes and is mounted onto a free standing arch to a total of four arches. Each wedge is sandwitched between scintillator counters (CSX) which are used for triggering. Each CMX arch has a conical structure with an opening angle of 41.4° and center of radius 10 m away of the interaction region along the z-axis.

2.3 The Trigger System

A three level trigger system is employed at CDF to accomodate the 1 MHz of overlapping triggers in order to keep all the events from small cross section processes handling efficiently at the same time the high rate of QCD dijet events. The trigger rate is successively reduced from the 280 KHz at the input of Level 1 to $\sim$ 10 Hz at the output of Level 3, with the output Level 3 rate being limited by the recording speed of 8 mm tape drives. The trigger complexity increases according to its level in the cost of decision time.

The Level 1 trigger

Level 1 is the simplest and fastest of the three trigger systems. At this level the decision to accept or reject an event is based on identification of energy clusters in the calorimeter or stubs in the muon chambers.

The energy of the calorimeter is summed in a form of trigger towers with dimensions of 15° in $\phi \times 0.2$ in η for both the electromagnetic and hadronic calorimeters. To accomodate the high rate, a single trigger tower is required to have transverse energy exceeding a preset threshold, different for the three calorimetry regions. The single trigger tower thresholds used in Run 1A and Run 1B are summarized in Table 2.2. The high E_T threshold for the forward calorimeters (51 GeV is the highest threshold) during Run 1B, eliminated any Level 1 triggers from these detectors but it was

necessary to avoid the saturation of Level 1 during very high luminosity conditions.

Detector	Run 1A	Run 1B
CEM	6	8
CHA	8	12
PEM	8	11
PHA	25	51
FEM	8	51
FHA	25	51

Table 2.2: Single trigger tower E_T thresholds used at Level 1 during run 1A and run 1B. E_T is measured in GeV.

The Level 1 muon decision was done requiring a minimum P_T for the candidate muon stub. Since full tracking information is available only at the highest trigger level, the muon P_T at this trigger level is inferred by measuring the angle of incidence of the incoming track, relative to an infinite momentum track emerging from the nominal interaction point ($z = 0$). The angle of incidence is measured by the difference in the arrival time (δt) of hits on a pair of radially alligned wires consisting the Level 1 muon segment or *Level 1 muon stub.*

Comparing the time difference δt to a hardwired preset value a track of minimum P_T is selected. The Level 1 trigger requirements were $P_T^{min} > 6$ GeV/c for CMU muons and $P_T^{min} > 10$ GeV/c for CMX muons.

To reduce background from accidental or uncorrelated hits, a minimum energy of 300 MeV was required in the hadron calorimeter tower associated with the muon stub.

Because the CMX detector is completely unshielded from the tevatron beampipe

which runs through the detector, its corresponding Level 1 trigger rate is very high. This is due to a large background from particles from the collision which scatter in the beampipe or the forward calorimeter and back into the CMX. This background is partially reduced by the hadronic energy requirement in trigger mentionted before. In order to further reduce this background, the Level 1 trigger calculates an arrival time for a CMX stub using the accompanying scintillator array (CSX) and requires it to be consistent with a particle from the interaction vertex.

The Level 1 offers a reduction of $\sim$ 1000 bringing the rate input to Level 2 to $\sim$ 2 kHz, while the decision time was less than the bunch crossing time of 3.5 μs.

The Level 2 trigger

Level 2 of the trigger is the stage where the decision for accepting or rejecting the largest portion of the events is made. It is a Fastbus based hardware trigger system with a much slower decision time ($\sim 20\mu$s) than Level 1. The analog trigger signals from the detector components are brought to the trigger electronics. The trigger towers, defined as 15° in $\phi \times 0.2$ units of rapidity at the first two levels of the trigger system, are summed into energy clusters by the "hardware cluster" finder, forming electromagnetic and jet like energy clusters. For each cluster the E_T, average ϕ, average η, and the ϕ and η widths are determined. All tower transverse energies are calculated with respect to the origin of the detector (z = 0). Also at this level, tracks were reconstructed using a fast track processor (CFT). The CFT processor searches for tracks in the r - ϕ of the CTC and matches them to EM clusters in the central region or muon stubs. The CFT momentum resolution is $\delta P_T/P_T = 3.5\%(GeV^{-1})P_T$. This procedure defines the central electron and muon Level 2 triggers. Specifically for the electron triggers, the algorithm is the following. The EM cluster is formed starting with EM trigger towers above a certain E_T threshold (cluster seeds). Any

adjacent EM towers are added to the cluster if their E_T was above a threshold value (*shoulder towers*). There were two EM clustering procedures based on different seed and shoulder tower E_T thresholds and were used for the construction of different electron E_T triggers. A cut on the electromagnetic fraction of the cluster was imposed as a final determination of an EM candidate cluster. It was required (EM + Had)/EM < 1.125. After the cluster was formed, a CFT track above some P_T threshold specific for the designed trigger was required to point at the cluster.

The muon Level 2 trigger requires a CFT track matched to a Level 1 muon stub. Because of the high trigger rate observed for the CMX and CMU with no CMP overlap, muon triggers from these detectors were prescaled. The prescaling factor was changed according to the instantaneous luminosity of the run.

Besides the lepton Level 2 triggers, triggers designed to collect data based on jet clusters are also implemented. The Level 2 inclusive jet triggers are made by demanding a single trigger tower above threshold at Level 1 and a cluster of electromagnetic and hadronic transverse energy in the calorimeter at Level 2. The Level 2 jet clusters are formed by starting with a seed tower of $E_T > 3$ GeV and summing the E_T of all the contiguous to the seed towers in η and ϕ. The additional towers are required to have $E_T > 1$ GeV. Events are accepted requiring jet threshold energies above $E_T >$ 20, 50, 70, and 100GeV. Because of the large trigger rate, the first three jet triggers were prescaled with the Jet_20 having the largest prescaling factor ($\sim$ 400). Data collected with the jet triggers are used in this analysis for calculating the fake rate of the b-tagging procedure.

A crude estimate of the *missing transverse* energy which reflects the transverse energy imbalance in the calorimeter, called $\not\!\!E_T$, attributed to non-interacting particles, is also possible at this trigger level. $\not\!\!E_T$ is calculated summing vectorially the

transverse energy of all towers in the calorimeter, $\vec{\not\!\!E}_T = -\sum_{i=1}^{n} \vec{E}_T^i$, where E_T^i is the transverse energy of the i-th tower.

The Level 2 electron triggers used in this analysis required an electromagnetic cluster of transverse energy greater than 16 GeV and a CFT track matched to it with transverse momentum $P_T > 12$ GeV/c. Another trigger designed for W electrons requiring an electromagnetic cluster of $E_T > 16$ GeV and $\not\!\!E_T > 20$ GeV was also used in order to cover losses due to the CFT requirement of the first trigger. For the muon Level 2 triggers the requirement was to have a CFT track with $P_T > 12$ GeV/c and pointing to within 5° of a Level 1 muon segment. To ensure good efficiency for top signal, some additional triggers requiring Level 2 jets or $\not\!\!E_T > 35$ GeV at Level 2 were also used in conjunction with the basic muon triggers.

Based on the decision of Level 2 the event rate is reduced from the 2 kHz input from Level 1 down to approximately $\sim$ 20 - 35Hz. Accepted Level 2 events are fed to the next stage of trigger, Level 3 for final processing and decision.

The Level 3 trigger

The Level 3 trigger decision is made after full event reconstruction, including track pattern recognition and cleanup of electronic noise and gas spikes. Since the process is very CPU consuming, accepted Level 2 events are processed on a "farm" of silicon graphics processors running the full CDF reconstruction package. Each CPU has the ability of processing an event while a second one is being read in or written out to disk. The input rate is reduced by 60% with final output rate of $\sim$ 8 Hz. Some loose quality criteria are applied on the leptons which are selected as output of this trigger level. The efficiency of the Level 3 trigger is very high ($\sim$ 98%) limited only by some tracking pattern recognition failures.

2.4 Luminosity measurement

Two planes of scintillator counters at either end of the detector, in front of the forward calorimeters were used for luminosity measurements. The counters known as beam-beam counters (BBC) are giving the hit count from particles produced during beam crossing. They cover the pseudorapidity range $3.24 < |\eta| < 5.90$. The rate (number) of coincidences in these counters, divided by the effective cross section of the counters, gives the instantaneous (integrated) luminosity. The BBC cross section, σ_{BBC} is given by [34]

$$\sigma_{BBC} = \sigma_{tot} \frac{N_{BBC}^{vis}}{N_{tot}}$$

where σ_{tot} is total $p\bar{p}$ cross section as measured in CDF [35], N_{BBC}^{vis} is the BBC triggered events and N_{tot} is the total number of inelastic and elastic events. The calculated $\sigma_{BBC} = 51.15 \pm 1.6$ mb. Using this value of σ_{BBC} the luminosity can be written as:

$$\mathcal{L}(t) = -\frac{f_o}{\sigma_{BBC}} \times \ln(1 - \frac{R_{BBC}}{f_o})$$

where f_o is the frequency of the Tevatron ($f_o = 286.272$ kHz) and R_{BBC} is the BBC coincidence rate. The formula is derived by assuming Poisson distribution for the number of $p\bar{p}$ interactions per beam crossing:

$$P(n) = \exp(- < n >) \times \frac{< n >^n}{n!}$$

and measuring the probability to have at least one interaction per crossing over the

total number of beam crossings

$$P = (1 - P(0)) = \frac{R_{BBC}}{f_o}$$

where $< n >$ is the average number of interactions per crossing and is given by

$$< n >= \frac{\sigma_{BBC}\mathcal{L}}{f_o}$$

Integrating over time both parts of the luminosity equation yields the integrated luminosity, $\int \mathcal{L} dt$. Due to the high luminosity conditions during the run, the observed BBC coincidence rate does not correspond to the true collision rate. Beam gas and beam halo events contribute an accidental rate on top of the true collision rate. This accidental rate depends on the luminosity of the beam. The effect is parametrized as a function of the instantaneous luminosity and is subtracted from the final integrated luminosity [34]. The total integrated luminosity for run 1A is estimated to be 19.3 pb^{-1} with an uncertainty of $\pm$ 3% and the corresponding one for run 1B is estimated to be 90.1 pb^{-1} with an uncertainty of $\pm$ 8%.

Chapter 3

Data Reconstruction and Selection

This chapter describes the selection of the W + *jets* data sample used in this analysis for the search of $t\bar{t}$ production. Leptons emerging from W decays are expected to have large transverse P_T and to be isolated most of the time. The analysis begins by selecting a high purity sample of leptons from the high-P_T lepton (e and μ) samples which are selected at trigger level. The quality criteria applied to this dataset to select W events and reduce the background from other physics processes are presented in the following sections.

Besides the high-P_T lepton sample, other datasets were used in order to obtain the efficiency of the various selection criteria and to study the characteristics of some background processes. The description of the Monte Carlo samples used for the calculation of acceptances and background estimates is also given in this chapter.

Before discussing the selection of the data samples it is useful to give a description of the reconstruction of various physics objects at CDF and the terminology which is used throughout this analysis.

3.1 Offline data reconstruction

3.1.1 Energy reconstruction

The energy in the calorimeter is measured in terms of the ADC counts of each calorimeter electronic channel. The ADC counts are converted to energy with use of detector dependent scale factors determined either from specific data samples collected during the run, or from testbeam data. After suppression of noisy channels and removal of spurious sources of energy, an $\eta - \phi$ array of tower energies is obtained which is further used to construct an array of transverse energies, E_T, using the polar angle, θ, of each tower center with respect to the event vertex. The transverse energy of each tower is given by: $E_T = E \sin\theta$

3.1.2 Track reconstruction

Track reconstruction in CDF is performed using information from all the tracking detectors. Tracks are reconstructed by fitting CTC hits to the three dimensional track-helix equation. The reconstruction starts by fitting hits from the CTC axial superlayers (r-ϕ plane), to a circle. The fitted circle is projected into the CTC stereo superlayers and a three dimensional five parameter fit to a track-helix is performed. For this part of the reconstruction, information from the VTX detector is also used. The five track parameters used in CDF for all track analysis are:

- *curvature*: the 2-dimensional curvature which is proportional to the transverse momentum of the track.

- *cotθ*: the cotangent of the polar angle θ.

- *impact parameter*: the distance of closest approach of the track to the interaction vertex in the transverse plane.

- z_0: the z-coordinate of the point of closest approach.

- ϕ_0: the ϕ direction of the track at the point of closest approach.

Each CTC track is then extrapolated in the SVX fiducial volume and a search for associated SVX hits (defined in Section 2.2.2) is performed. Multiple scattering and ionization loss of energy in the material are taken into account during the extrapolation. The size of the search region, *road search*, is defined based on the track parameters and their uncertainties. Every time a cluster is found, the track parameters, the error matrix and the road size are recalculated and a new search is performed with the new parameters. Since more points are added to the track fit, the χ^2 is recalculated and the cluster assignment to the track candidate is determined according to the goodness of the fit. The procedure is iterated for each SVX layer starting from the outermost layer and continuing to the inner ones until all four layers are examined. An SVX track segment with at least two hits is chosen as the one associated with the CTC track if the value of the total χ^2 is the lowest among the values obtained using the other SVX segments. It is further required that the SVX tracks share no more than two hits with other SVX tracks. SVX track segments linked to CTC tracks are declared *SVX tracks*, otherwise ignored. The track finding efficiency is 98.0% for the SVX detector and better than 98.7% for the SVX$'$. Information from the SVX tracks is used as input to the algorithms used to identify heavy flavor jets from $t\bar{t}$ decays. They are also used for the reconstruction of the $p\bar{p}$ interaction point.

3.1.3 Vertex reconstruction

The event vertex

The vertex algorithm uses tracks reconstructed in the VTX detector, to define a common point along the beam line from which the tracks originate. This common point defines a vertex. Since there can be more than one vertex present in an event, the vertex with the highest VTX hit occupancy associated with it, is declared the *event vertex*. The event vertex, as defined above, is used to determine the polar angle, θ, of each calorimeter tower. The transverse energy of each calorimeter tower is calculated according to the event vertex and consequently all algorithms making use of calorimeter towers for cluster formation or calculation of the energy imbalance in the detector, refer to this vertex.

The lepton vertex

Due to high luminosity conditions during the run, 50% of the events contain multiple interactions which result to additional vertices in the event. The large number of additional vertices result to an uncertainty on the choice of the event vertex. Despite the fact that wrong choice of the event vertex does not affect on average the kinematical characteristics of the events, it can lead to large fluctuations on an event-by-event basis. Furthermore, since the tracks matched to a jet are used by the tagging algorithms (see chapter 4) to determine whether a jet is tagged or not, wrong choice of event vertex can lead to reconstruction of jets with the wrong direction and therefore to track-jet mismatches.

Events used in this analysis contain always a high-P_T lepton and therefore there is always a high-P_T track associated with it. In order to resolve the ambiguity on the choice of the event vertex, the vertex associated with the lepton track is declared as the event vertex and consequently jet clustering, the missing E_T calculation and

b-tagging is performed with respect to this vertex.

3.1.4 Electron reconstruction

Electron reconstruction begins with a clustering algorithm to identify electron showers. An electron cluster consists of a *seed* tower and *shoulder* towers (towers adjacent to the seed). Towers with transverse electromagnetic energy of $E_T \geq 3$ GeV are considered as seed towers. Once a seed tower is found, all adjacent towers in η with $E_T \geq 0.1$ GeV are incorporated in the cluster. The algorithm continues until no more adjacent towers are added to the cluster or the cluster reaches its maximum size. The maximum of the cluster size is restricted to 3×1 in $\eta - \phi$, three towers in η ($\delta\eta$=0.1), and one tower in ϕ ($\delta\phi = 15°$) for central electron candidates. For the plug and forward electromagnetic calorimeters where the tower size is smaller, the maximum electromagnetic cluster size is restricted to 5×5 for the plug and 7×7 towers in $\eta - \phi$ for the forward. An electromagnetic cluster is accepted if the total transverse electromagnetic energy is $E_T \geq 5$ GeV and the ratio of the energy in the hadronic component of the towers contributing to the cluster to the electromagnetic energy is less than 0.125 ($E_{had}/E_{em} \leq 0.125$).

3.1.5 Muon reconstruction

The muon reconstruction starts with the reconstruction of a muon track segment, *stub*, by fitting the hits observed in one of the muon detectors. The stubs are categorized according to the detector they are reconstructed at. At the second step, CTC tracks are extrapolated to the muon chambers and matched to muons stubs in $r - \phi$ is performed. A muon candidate is declared when the muon stub is linked to the nearest CTC track.

3.1.6 Jet reconstruction

A jet is a collimated spray of hadrons formed during the fragmentation of a parton (quark or gluon) emerging from the hard scattering. During the fragmentation process, the emerging particles acquire momentum perpendicular to the parton direction causing the jet to spread over many calorimeter towers. Jets appear approximately circular in $\eta - \phi$ space. Jets of the same E_T appear equal in size independent of their pseudorapidity. These properties are used by the clustering algorithm, described in the next section. Jets are reconstructed by summing up the energy deposited in the calorimeter cells within a fixed cone size in $\eta - \phi$ space of radius defined as $\Delta R = \sqrt{\delta\eta^2 + \delta\phi^2}$. The cone radius is defined around the jet centroid direction.

The size of a jet exhibits a weak dependence on its transverse energy. In the absence of gluon emission the angular spread of the jet of hadrons would decrease roughly linearly with increasing energy of the parton. Because of gluon emission, the size of a jet transverse to a central axis will not decrease linearly, but more slowly (e.g. logarithmically) with energy [6]. Nevertheless, there is no theoretically determined value of how wide a jet is, and different cone sizes can be chosen according to the purpose of a specific analysis. Smaller cone sizes give better resolution on the angles of the jets, while larger cone sizes include more fragmentation products from the original parton providing better energy resolution. In CDF three cone sizes are used for jet clustering $\Delta R = 0.4$, 0.7 and 1. Jet studies at CDF has shown that approximately 70% of the jet energy is contained within a cone of 0.4 radius [42]. Throughout this analysis, the 0.4 cone size used for jet clustering. This cone size is chosen because it gives better discriminating power for detecting $t\bar{t}$ events [59] in which case large jet multiplicity is expected.

3.1.7 Jet clustering algorithm

The jet algorithm begins by creating a list of calorimeter towers above a fixed E_T threshold, set at 1 GeV, to be used as seed towers for the jet finding algorithm.

Seed towers are merged with other contiguous seed towers to form preclusters. Associated seed towers are required to fall monotonically in energy. If an energy valley is observed between adjacent towers, the algorithm separates the preclusters. The list of the preclusters is used as input to the clustering algorithm.

For the growth of the preclusters into clusters, the algorithm first calculates the E_T weighted centroid of the precluster and a cone of radius R (0.4 for this analysis) is formed around this centroid. Then, all calorimeter towers with $E_T \geq 100$ MeV which centroid contained in this cone are incorporated into the precluster to form a cluster. The E_T weighted centroid of the newly formed cluster is calculated taking into account all the towers in the cluster. A new cone is drawn around the new cluster centroid and the procedure is repeated. The process of recomputing a new centroid and finding new or deleting old towers from the cluster list is iterated until the tower list remains unchanged. During the iteration, the towers of the precluster are kept intact regardless of whether they lay inside the circle or not. This procedure is followed to avoid cluster formation far from its initial center.

The final steps of the algorithm take care of overlapping clusters and sharing of calorimeter towers by more clusters. The cluster overlap and tower sharing is a product of the independent formation of clusters. The E_T of the common towers is computed and compared to the E_T of each individual cluster. If the E_T of the common towers accounts for more than 75% of the E_T of the smaller cluster, the two clusters are merged into a single one. Otherwise the clusters are separated and the centroid is recomputed without the common towers. The common towers are split

between the two clusters based on the proximity of their centroid to the two cluster centroids. After the reassignment of common towers, the centroid of each cluster is recomputed and the procedure is iterated until the list of separated towers does not change.

At the end of the clustering algorithm the four-vector (E,p) of every jet is computed summing the energy of all towers (i) contributing to the formation of the jet:

$$\begin{aligned} E &= \sum_i (E_i^{em} + E_i^{had}) \\ p_x &= \sum_i (E_i^{em} \sin\theta_i^{em} + E_i^{had} \sin\theta_i^{had}) \cos\phi_i \\ p_y &= \sum_i (E_i^{em} \sin\theta_i^{em} + E_i^{had} \sin\theta_i^{had}) \sin\phi_i \\ p_z &= \sum_i (E_i^{em} \cos\theta_i^{em} + E_i^{had} \cos\theta_i^{had}) \end{aligned}$$

Based on the above quantities the jet $P_T \equiv \sqrt{p_x^2 + p_y^2}$, $E_T \equiv P_T \frac{E}{P}$ and $\eta \equiv 0.5 \ln \frac{E+p_z}{E-p_z}$ are calculated. Often in CDF the quantity η_D, the pseudorapidity of the jet with respect to the origin of the coordinate system (x=y=z=0), is used to define the location of the jet with respect to calorimeter regions. The jet four-vector calculated above is based on the raw calorimeter energies and is subject to numerous corrections. Jets which are used with the raw energy and momentum quantities are termed *uncorrected* jets.

3.1.8 Jet Energy Corrections

The jet energy is measured by summing up the raw energies of each individual tower included in the calorimeter cluster. This energy assigned to the jet does not correspond to the energy of the parton originating the jet. The resolution in this

measurement is quite poor and typically of the order of $\sigma(E_T)/E_T \approx 1/\sqrt{E_T}$, where E_T is measured in GeV. This poor jet resolution arises due to (a) non-linear response of the calorimeter to low energy charged pions, (b) lower response at the boundaries of the different calorimeter regions and tower thresholds, (c) loss of low momentum fragmentation particles inside the magnetic field, (d) energy deposition in towers outside the jet cone, (e) contribution from underlying-event or additional interactions and (f) energy loss due to minimum ionizing particles or neutrinos present in jets formed by semileptonic decays of b and c-quarks.

A correction function which takes into account these effects is generated and can be applied to jets in the data sample. The correction function is a map of the detector response for different energies and pseudorapidity regions. The procedure for obtaining the response map can be divided into four parts. In the first part the response of the central calorimeter to jets is determined using tracking information from the CTC detector. The CTC information is used in order to study the fragmentation properties of jets and to study the response of the calorimeter to low momentum, $P_T \leq 10$ GeV/c, pions. This information is then used to tune the fragmentation parameters in the Monte Carlo simulations and to include the non-linear response of the detector to low energy pions. At the second step, the response in the central calorimeter is extended into the other regions of the detector where tracking information is unavailable, in order to obtain a uniform calorimeter response across the entire detector. The last two parts of the response function correct for energy lost outside the jet cone and for the contribution to the jet energy from the underlying event. These corrections have been derived with the purpose of correcting *on average* the jet energies. The tuning of the various correction factors for the two collider Runs (1A and 1B) is incorporated into a special CDF offline routine [44]. A brief description of

each part of the response function is given below.

Central jet response - Absolute energy scale

These type of corrections try to adjust the observed jet energy as close as possible to the true parton energy. The corrections are based on Monte Carlo studies by comparing the observed jet E_T to the momentum sum of all particles produced in the fragmentation process and lying in a cone centered about the measured jet axis and originating from the primary partons. A quadratic spline fit is used to parametrize the average jet response as a function of the jet E_T. In order to use these corrections on jets observed in the data, the Monte Carlo jet fragmentation properties and detector response function need to be tuned to the corresponding ones observed in the data.

Jets are mainly composed of pions. The detector response to jets consequently depends on the calorimeter response to the momentum spectra of the charged pions produced during the fragmentation process. The calorimeter response to high P_T pions has been measured in testbeam data using fixed momentum pion beams while the response to pions of $P_T \leq 10$ GeV/c has been measured in minimum bias events using isolated tracks. A non-linear response function is obtained. The measured response as a function of pion momentum is used in the Monte Carlo simulation to reproduce the same detector response [43]. The jet fragmentation properties are then studied using a combination of calorimeter and tracking information associated with the jets in the event. The observed kinematical distributions from the data are compared to the same fragmentation distributions obtained using the Monte Carlo fragmentation modeling. For these studies the ISAJET [71] Monte Carlo generator was used with the Feynmann-Field [73] fragmentation model. The fragmentation parameters in the Monte Carlo are then tuned in order to obtain the same kinematical distributions as the ones observed in the data.

Relative energy corrections

These corrections compensate for the non-linear response of the calorimeter as a function of detector pseudorapidity η_D. Different calorimeter components have different response on particle showers. Also the energy measurement at the boundaries of calorimeter subsystems is reduced since fragmentation particles can go inside cracks and their energy is unmeasured. These type of corrections are based on the calibration of the jet energy with respect to jets measured in the central calorimetry region where the detector response is well understood. These correction factors are obtained using dijet events with one jet required always in a good central region ($0.2 \leq |\eta| \leq 0.7$) away of detector boundaries. Requiring the E_T of the two jets to be balanced, as expected for dijet events, leads to an η energy scale correction map. Figure 3.1(a) shows the relative correction factor as a function of the jet detector pseudorapidity, η_D. The peaks correspond to detector boundaries where the detector response is quite reduced and thus the correction factor large.

Out-of-Cone corrections

These corrections compensate for the loss of energy outside the jet cone. They depend on the size of the jet clustering cone and are derived from Monte Carlo simulation.

Underlying Event and Multiple interactions

These corrections compensate for the contribution to the jet energy from particles originating from the remnants of the proton or antiproton, participating in the primary interaction (called UE corrections) and from the presence of additional interactions in the event (called MUE). The UE correction is applied after both relative and absolute energy corrections while the MUE correction is applied after the relative correction. The energy subtracted from a jet of 0.4 cone size due to underline event is 0.297 GeV. This amount of undeline event energy in the jet corresponds to the case

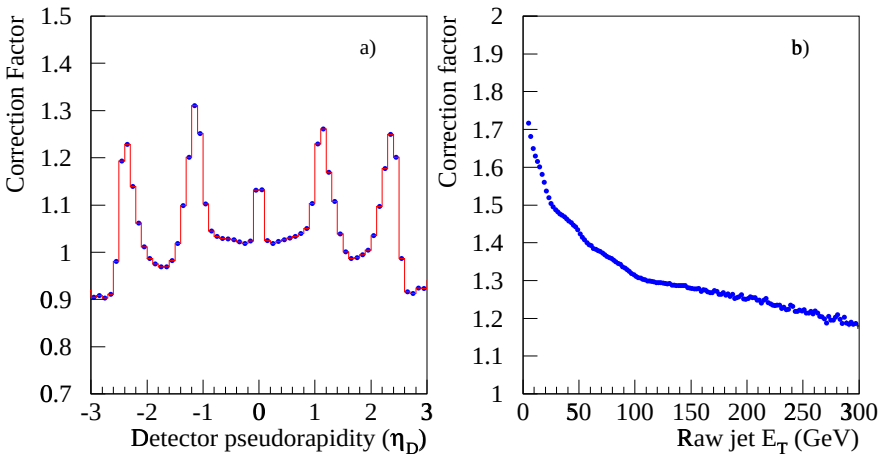

Figure 3.1: a): Distribution of the relative jet energy correction scale as a function of the jet detector pseudorapidity (η_D), b): Average jet energy correction factor as a function of the jet raw E_T. Jets are clustered with cone of 0.4 and in the $|\eta| \leq 3.5$. The peaks shown in plot (a) correspond to detector boundary regions where the detector response is significantly lower and therefore larger corrections apply on the jet energy.

there is one interaction or primary vertex in the event. An additional amount of 0.65 GeV is subtracted from the jet for every additional vertex of class 12 in the event.

The fully corrected jet transverse energy can be expressed in terms of the above corrections in the following form:

$$P_T^{cor} = (P_T^{raw} \times F_{rel} - MUE) \times F_{abs} - UE + F_{o-c}$$

On average the jet corrections increase the raw jet energy by $\sim$ 30% The corrections reproduce on average the true jet E_T but do not reduce energy fluctuations around this average value. The average jet energy correction factor as a function of the jet E_T is shown in Figure 3.1(b), for jets clustered with cone size $R = 0.4$ in the

pseudorapidity range of $|\eta_D| \leq 3.5$

The main sources of uncertainty in the jet E_T are a) detector resolution and time stability of calorimeter energy scales, b) jet fragmentation and c) underlying event. The fractional systematic uncertainty in the jet energy is shown in Figure 3.2 as a function of the jet true E_T [46].

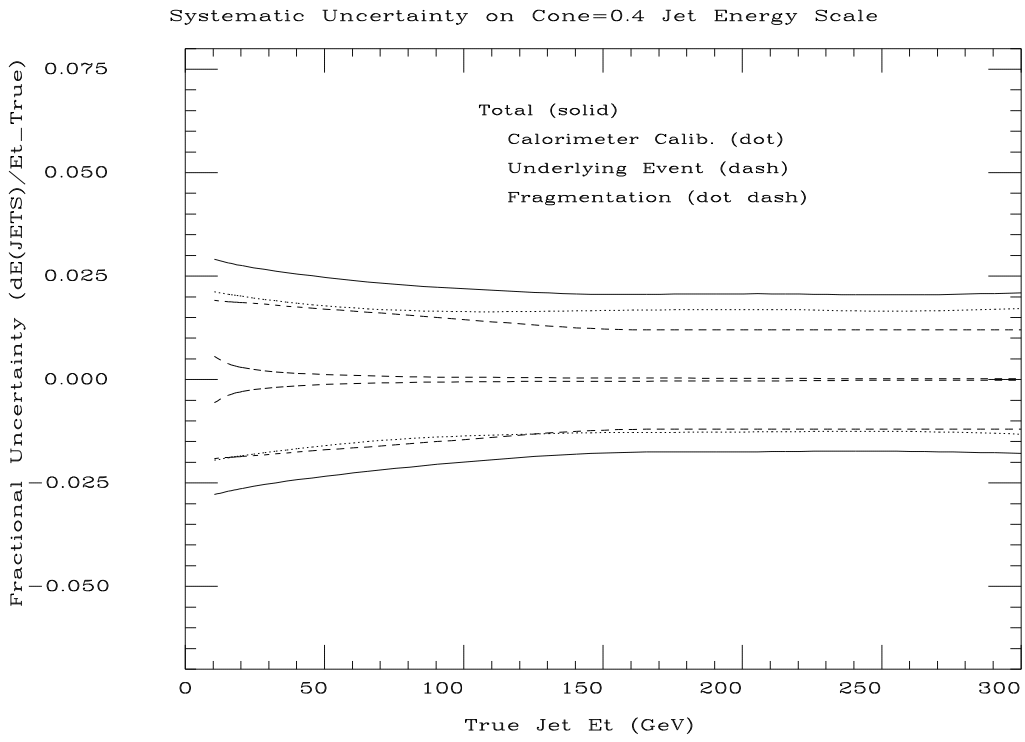

Figure 3.2: The fractional systematic uncertainty in the jet energy scale for 0.4 cone jets as a function of the true jet E_T.

Checks of the obtained jet energy correction function are performed with $\gamma+$ jet and $Z+$jet balancing. In each case the reconstructed boson transverse momentum balances the corrected E_T of the recoiling jet. The fractional energy imbalance is measured to be within 3% of the Z or the photon energy. While the systematic

effects on the reconstructed jet E_T is within 2.5%, the limited knowledge of large-angle gluon emission and how well it is modeled in the Monte Carlo simulations results in a combined 5% systematic uncertainty on the jet energy scale [46].

The correction procedure described above applies to gluon and light-quark jets. The corrections discussed above are based on QCD dijet events and therefore can vary according to the underlying hard scattering process. Additional jet corrections are derived for the measurement of the top mass. These corrections depend on the type of parton they are assigned to when fitting for the top mass. Therefore, there are corrections for the light quarks from a W decay, for b-quarks decaying hadronically, for b-quarks decaying semileptonically in electrons and finally there are separate corrections for b-quarks decaying semileptonically to muons. These corrections are derived from a study of $t\bar{t}$ events generated with the HERWIG Monte Carlo. A detailed description of these corrections can be found in [59]. Since they are not used in this analysis there will not be further discussed.

3.1.9 Missing Transverse Energy ($\not\!\!E_T$) and Neutrino reconstruction

Neutrinos do not interact in the detector since there is not sufficient amount of material. Since the longitudinal component of the colliding partons momentum is unknown, only the transverse component of the neutrino momentum can be measured. From transverse momentum conservation, the presence of undetected neutrinos results in transverse energy imbalance in the detector which is proportional to the neutrino momentum and it is called *missing* E_T or $\not\!\!E_T$.

The $\not\!\!E_T$, is defined as the negative vector sum of the transverse energy in the

calorimeter:

$$\vec{\not\!\!E}_T = -\sum_i E_T^i \hat{\mathbf{n}}_\mathbf{i}, \tag{3.1}$$

where the $\sum$ is performed over all the calorimeter towers $\mathbf{i}$ in the pseudorapidity of $|\eta_i| < 3.5$. E_T^i is the magnitude of the transverse energy of tower $\mathbf{i}$, $\hat{n_i}$ is the unit vector perpendicular to the beam axis pointing at the tower face.

The magnitude of the $\not\!\!E_T$ can be expressed as:

$$\not\!\!E_T = \sqrt{(\sum_i E_T^i \cdot \cos\phi_i)^2 + (\sum_i E_T^i \cdot \sin\phi_i)^2}$$

The resolution of the $\not\!\!E_T$ depends on the response of the calorimeter to the total energy deposited in the event and therefore on the event topology. The $\not\!\!E_T$ resolution is parametrized in terms of the total scalar E_T, termed $\sum E_T$. The resolution is measured with minimum bias events. Minimum bias events are collected requiring a coincidence of hits in both the forward and backward BBC counters. No requirements on the calorimeter energy is made. These minimum bias events are dominated by inelastic $p\bar{p}$ collisions. No significant $\not\!\!E_T$ is expected in these events. A fit to these data, yields [72]

$$\sigma(\not\!\!E_{T_x}) = -0.582 + 0.7418\sqrt{\sum E_T(GeV)}$$

The non-zero term in the expression is due to run-by-run offsets and out-of time accidental energy which are not accounted for.

3.1.10 $\not\!\!E_T$ Corrections

The magnitude and direction of $\not\!\!E_T$ are subject to two types of corrections. The first type corrects the $\not\!\!E_T$ for the presence of muons in the event. The second correction is applied only in cases the energies of the jets in the event are corrected according to the procedure described in Section 3.1.8.

The $\not\!\!E_T$ needs to be corrected for the presence of muons in the event because muons deposit minimum ionizing energy in the calorimeter. Equation 3.1 does not include the full muon energy but only the energy deposited in the muon calorimeter tower. In order to account for the presence of muons in the event, the $\not\!\!E_T$ is recalculated according to the following equation:

$$\vec{\not\!\!E}_T^{\,\mu\ cor} = \vec{\not\!\!E}_T^{\,raw} + \vec{E}_T(\mu\ tower) - \vec{P}_T^{\mu} \tag{3.2}$$

The term $\vec{E}_T(\mu$ tower) removes from the raw $\not\!\!E_T$ the energy of the calorimeter tower traversed by the muon and the term $\vec{P}_T^{\mu}$ adds back to the raw $\not\!\!E_T$ the muon P_T.

The procedure is repeated for all identified muons in the event with $P_T \geq 5$ GeV/c passing the tight selection requirements (see Table 3.4). When Z and dilepton candidate events are selected, the $\not\!\!E_T$ is also corrected for the presence of muons even when the muon is not in the tight lepton list but satisfies the selection criteria of the second leg (see sections 3.6.1 and 3.6.2).

For the rest of the analysis, reference to the event $\not\!\!E_T$, implies that the magnitude and direction of the $\not\!\!E_T$ vector is corrected for the presence of any primary muons in the event.

In order to calculate the $\not\!\!E_T$ in the event after jet energy corrections, the energy in the calorimeter is divided in three distinct categories and corrections applied to each

category separately. The first category, termed *lepton cluster energy*, includes the towers contained in the primary electron clusters. The second category includes the energy deposited in the calorimeter towers contained in jet clusters with E_T above a threshold and is termed *clustered energy*. Finally the third category contains any energy detected in the calorimeter but not collected in a jet or electron cluster and is called *unclustered energy*. Based on this division of tower energy, equation 3.1 can be written in the following form:

$$\vec{\not\!\!E}_T = -\{\vec{E}_T^{ele} + \vec{E}_T^{clust} + \vec{E}_T^{uncl}\} \tag{3.3}$$

$$\vec{E}_T^{ele} = \sum_{prim.\ ele} \vec{E}_T(ele) \tag{3.4}$$

$$\vec{E}_T^{clust} = \sum_{jet} \vec{E}_T(jet) \tag{3.5}$$

Equation 3.4 refers to the contribution to the $\not\!\!E_T$ from the towers in all primary electron clusters in the event. Equation 3.5, represents the sum over all jet clusters with $E_T \geq 10$ GeV and $|\eta_D| \leq 3.5$. From equations 3.1 and 3.3 the event uncluster energy can be inferred. Once each term is corrected, the corrected $\not\!\!E_T$ is reconstructed replacing the corrected terms in equation 3.3:

$$\vec{\not\!\!E}_T{}^{cor} = -\{\vec{E}_T^{ele,cor} + \vec{E}_T^{cl,cor} + \vec{E}_T^{uncl,cor}\} \tag{3.6}$$

The term $\vec{E}_T^{cl,cor}$ represents the clustered energy, after jet energy corrections. In order to avoid double counting of energy, corrections for the energy loss outside the jet cone, and the underlying event contribution are not applied.

$$\vec{E}_T^{cl,cor} = \sum_{jet} \vec{E}_T^{cor}(jet) \tag{3.7}$$

Corrections applied to the electron clusters are quite small and are not taken into account. So the $\vec{E}_T^{ele,cor}$ is assumed to be equal to $\vec{E}_T^{ele}$.

The correction to the unclustered energy, $\vec{E}_T^{uncl}$, is calculated similar to the jet energy corrections [47]. An correction factor, $a_{uncl} = 1.6$, is used to adjust on average the unclustered calorimeter energy for detector nonlinearities, reduced calorimeter response at the boundaries between modules, and detector dependent thresholds. This correction factor can be thought of as the correction factor which could be applied to a jet with E_T equal to the unclustered energy.

3.1.11 Bad Run Removal

Runs associated with detector or readout problems were removed, to ensure good quality of the datasets. This requirement removes 4% of the events in the high-P_T inclusive lepton samples [41].

3.2 The inclusive high-P_T lepton sample

The high-P_T lepton sample consists of electrons and muons, produced in W and Z leptonic decays either directly through $W \to \ell\nu$ and $Z \to \ell^+\ell^-$ or in the decay chain $W \to \tau\nu$, and $Z \to \tau^+\tau^-$ with subsequent τ leptonic decay, $\tau \to \ell\nu\nu_\tau$. QCD multijet processes are additional sources of high-P_T leptons. In this case the leptons can be either real or fake. Real leptons produced in semileptonic decays of heavy quarks, $b \to c\ell\nu$ and $c \to s\ell\nu$, photon conversions in the case of electrons and decays in flight of charged mesons like $\pi/K \to \mu\nu$ in the case of muons. Fake leptons originate from hadron showers and exhibit most of the characteristics of real leptons. There are different QCD processes contributing to fake leptons depending on the

lepton in question. Electron-like objects can be produced via "charge exchange", $\pi^- + p \rightarrow \pi^0 + n$, or overlaps of a $\pi^\pm$ with a π^0. Fake muons can be produced by energetic pions which penetrate the calorimeter and create stubs in the muon chambers.

A series of quality criteria were developed for the separation of the two and selection of real leptons. Also the contribution of QCD events is expected to fall exponentially with increasing lepton P_T while leptons from W and Z decays are very energetic. Furthermore W events are associated with large $\not\!\!E_T$ due to the produced neutrino, while in QCD events, $\not\!\!E_T$ is a result of jet mismeasurement and detector resolution. It is clear also that leptons produced in QCD multijet events are usually accompanied by other hadronic activity while leptons from W and Z decays are often isolated. Because tracking information is fully available for the central part of the detector and in order to ensure large rejection of QCD events, events from the central lepton samples only are used in the analysis.

3.2.1 High purity central electron sample

The identification of central electrons begins at the Level 1 trigger stage, requiring a CEM tower with $E_T \geq$8 GeV. At Level 2, the electromagnetic clustering is performed requiring a CEM cluster of $E_T \geq$ 16 GeV. This requirement is ANDed at trigger level with the requirement of a CFT track with $P_T \geq$ 12GeV/c associated with the electromagnetic cluster. This is one of the two level 2 triggers used to accept events with high-P_T electrons. Because the CFT requirement is only $\sim$ 92% efficient, a second trigger with the requirement $\not\!\!E_T \geq$ 20GeV in place of the CFT requirement is also used in this analysis. The last trigger is designed mainly to accept W events. The logical OR of the two triggers is found to be 100% efficient for W events and

is used in this analysis to accept the electron events. The final sample selection at trigger level is performed at Level 3 where some loose quality criteria are imposed on the variables described below.

Transverse Energy $E_T \geq 20$ GeV
Good fiducial region
$E_{had}/E_{em} \leq 0.05$
$E/P \leq 1.8$
Lateral sharing (Lshr) ≤ 0.2
$\chi^2_{strip} \leq 10$
Track - Shower matching:
$\lvert\Delta(x)\rvert \leq 1.5$ cm, $\lvert\Delta(z)\rvert \leq 3$ cm
$\lvert\Delta(Z^{ele}_{vtx} - Z^{trk}_0)\rvert \leq 5$ cm
$\lvert Z^{ele}_{vtx}\rvert \leq 60$ cm

Table 3.1: Variables used for electron identification and the corresponding required values for the selection of *primary* electrons.

In order to reduce the contribution of fake electrons in the sample a number of electron identification criteria are imposed on the sample. The variables considered along with the corresponding requirements used to select the *the primary* or *tight* electrons are the following:

1. **Electron cluster $\mathbf{E_T} \geq$ 20 GeV.**

2. **Electron in good fiducial volume.** The central electron fiducial volume is defined by the CES strip chambers and the exclusion of uninstrumented regions.

Overall the fiducial volume covers the pseudorapidity region of $|\eta| \leq 1.05$. The fiducial volume requirements are:

- The extrapolated track position in the CES plane is required to lie within 21 cm of the wedge center in the r-ϕ view, to eliminate the possibility of the electron falling in a ϕ crack between wedges or some of the energy being lost in the crack.
- The region $|\eta| < 0.05$ is excluded. This is approximately 9 cm from the edges of the η -tower at $\phi = 90^0$. This is the region where the two halves of the detector meet.
- The region $0.77 < \eta < 1.0$, $75^\circ < \phi < 90^\circ$ is not instrumented and thus excluded from the fiducial volume. This region contains the cryogenic connections to the solenoidal magnet.

3. $\mathbf{E/P < 1.8}$. The quantity compares the energy (E) measured by the calorimeter to the momentum (P) of the track as measured by the CTC and it is required to be close to 1. Due to bremsstrahlung radiation the momentum of the electron is measured lower than the energy in the electromagnetic cluster which usually contains the radiated photons. The tail of E/P is larger for events like $\pi^0 \rightarrow \gamma\gamma \rightarrow \gamma e^+ e^-$, where the the electromagnetic cluster energy is the one of the π^0 but the momentum the one of the two electrons.

4. $\mathbf{E_{Had}/E_{Em} < 0.05}$. This requirements ensures that most of the energy associated with the electron shower is contained in the electromagnetic calorimeter. It differentiates against the hadronic showers which deposit energy in both the electromagnetic and hadronic compartment of the calorimeter.

5. **Lateral sharing (Lshr) $<$ 0.2.** This quantity allows a comparison of the lateral sharing of energy in the towers of the electron cluster to the predicted sharing based on the z position and angle of the electron shower as measured by the CES detector and a parametrization based on test beam electron data. The quantity as derived from test beam has the functional form:

$$Lshr = 0.14 \sum_i \frac{E_i^{adj} - E_i^{prob}}{\sqrt{(0.14\sqrt{E})^2 + (\Delta E_i^{prob})^2}},$$

where E_i^{adj} is the energy (in GeV) measured in a tower adjacent to the seed tower of the electron cluster, E_i^{prob} is the energy (in GeV) expected in this tower based on strip chamber information, $0.14\sqrt{E}$ is the error on the energy estimate from calorimeter resolution and ΔE_i^{prob} is the uncertainty on the expected energy associated with a 1 cm uncertainty in the strip chamber position measurement.

6. $\chi^2_{\mathbf{strip}} < \mathbf{10.}$ This quantity is derived by performing a chi-square fit of the lateral shape of the measured electron shower to the shower shape measured for electrons from a test beam.

7. **Track-shower matching variables:**

   $|\mathbf{\Delta(x)}|$ **(r-ϕ view)** < 1.5 **cm** and $|\mathbf{\Delta(z)}|$ **(r-z view)** $<$ **3.0 cm.**

   These quantities are derived comparing the shower position of the largest pulse height cluster as measured in the CES to the extrapolated position in the CES plane, of the CTC track associated with the electron cluster. A large number of fake electrons which result from overlaps of charged and neutral hadrons can be removed by requiring a tight match between the track and shower position.

8. $|\delta(\mathbf{Z_{ele}^{vertex}} - \mathbf{Z_{track}})| <$ **5.0 cm.** The electron track is required to lie within 5

cm of a well reconstructed vertex.

9. $|\mathbf{Z_{vtx}^{ele}}| < \mathbf{60\ cm.}$ The electron vertex is required to lie within 2 σ of the nominal interaction vertex (z=0). The longitudinal spread of the event vertex about the nominal interaction point is a Gaussian with a sigma of $\sigma = 26$cm.

Table 3.1 summarizes the variables used to identify electrons and the values used to select the primary electron sample. The distributions of the above identification variables are shown in figure 3.3 for electrons in a $Z \rightarrow e^+e^-$ sample.

Since electrons leave clusters of energy in the calorimeter, they are also reconstructed as jets according to the jet clustering algorithm. In this case the cluster includes more towers as described in section 3.1.7. To avoid double counting of electron and jet clusters, when an electron candidate is found passing the strict selection requirements, the energy of the electromagnetic towers contributing to the electron cluster is zeroed and the jet clustering is re-run on the updated tower list.

3.2.2 Conversion removal

Photons arise either directly from the hard scattering (*direct* photons) or from π^0 decays. They can interact with the material in the detector and convert to an electron-positron pair. A conversion removal algorithm was developed to remove this type of electrons. The algorithm [37] is based on the search of an opposite sign track near the electron candidate. Two tracks are considered originating from a photon conversion if their $R\phi$ separation at the point of conversion (the point at which the two helices are parallel in ϕ), and the difference in their polar angle, $\delta \cot\theta$, are close to zero. The quantity δ_{sep} is a signed quantity and it is positive if the two helices do not overlap and negative if they overlap. The quantity is symmetric

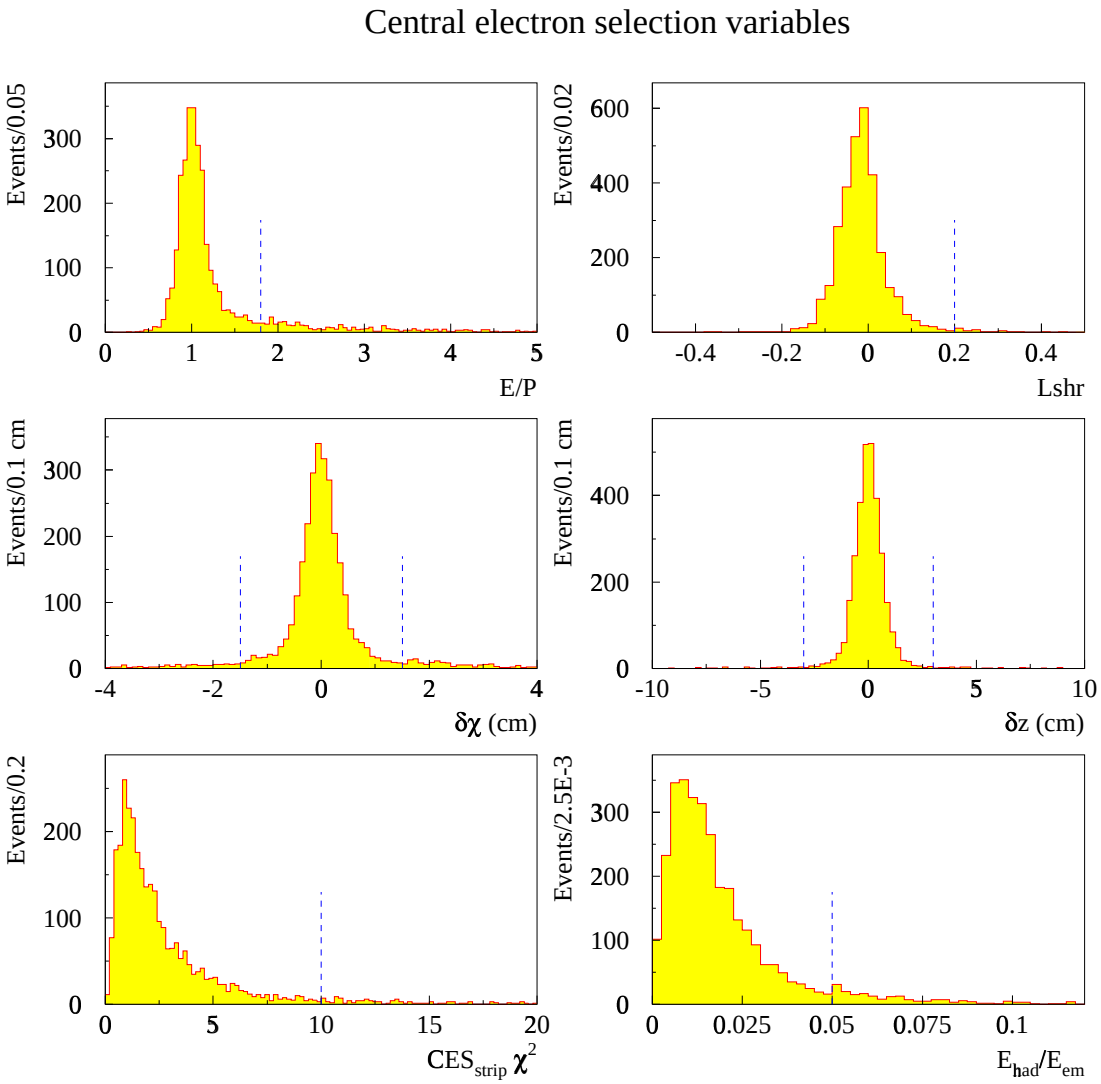

Figure 3.3: Distributions of the variables used for central electron identification as measured in a sample of $Z \rightarrow e^+e^-$ events. The plotted variables correspond to one leg of the Z decay, while the other leg is required to pass all the selection criteria listed in table 3.1. The dotted lines indicate the value used to select primary electrons.

$\lvert\delta(r-\phi)_{sep}\rvert < 0.3$
$\lvert\delta\cot(\theta)\rvert < 0.06$
$-20.0 < R_{conv} < 50.0$ cm
OR
$VTX_{occupancy} < 0.2$

Table 3.2: Criteria used to identify electrons from photon conversions.

around zero for real conversion pairs. For fake conversions, negative separation is the most likely configuration because fakes consist of two prompt tracks arising from the interaction region. The conversion radius, defined as the distance from $x = y = 0$ to the conversion point, is required to be in the range $-20 < R_{conv} < 50$ cm. In many cases the track partner is too soft to be reconstructed and the algorithm fails. In such a case the number of VTX hits associated with the electron candidate is compared to the expected number. If the number of found hits is fewer than 20% of the expected one, the electron candidate is removed as photon conversion. Table 3.2 summarizes the criteria used to identify and remove electrons from photon conversions. Figure 3.4 shows the distribution of the variables used to identify photon conversions in the inclusive high-P_T electron sample. As shown in figure 3.4(e), conversion electrons comprise $\sim 40\%$ of the inclusive high-P_T electron sample. The sharp falling spectrum is indicating the presence of W electrons for $\not\!\!E_T \geq 20$ GeV.

The *efficiency* of the conversion algorithm was measured with a sample of photon conversions selected with the Central Preradiator Detector (CPR) which are a set of proportional chambers placed between the CEM and the magnetic coil and act as a pre-shower detector. The efficiency of the conversion removal algorithm is

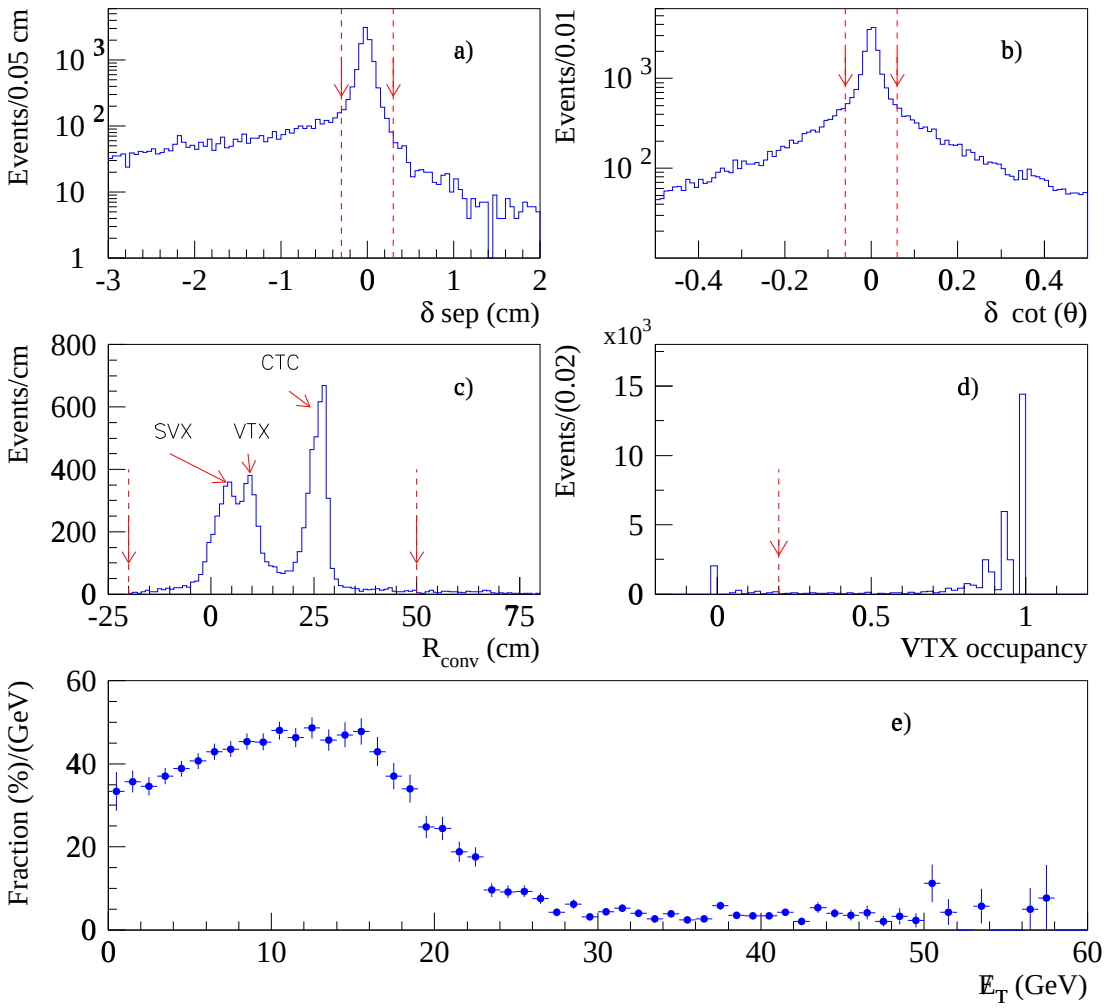

Figure 3.4: Conversion variables. a): separation distribution of the two tracks at tangent point, b): distribution of the $\Delta \cot\theta$, c): the radial distribution of conversion electrons after the cuts indicated in (a) and (b) are applied. d): VTX occupancy of inclusive electrons and e): fraction of photon conversions as a function of the $\not\!\!E_T$. The sharp drop of the distribution at $\sim$ 20 GeV is indicating the presence of W-electrons.

found to be 90.7 $\pm$ 3.8% [38]. The *over-efficiency* of the algorithm, meaning the fraction of non-conversion electrons removed erroneously, is estimated using a sample of $Z \rightarrow e^+e^-$ events and examining the number of events removed under the Z mass peak ($70 \leq M_{e^+e^-} \leq 110$ GeV/c^2). The over-efficiency of the conversion removal algorithm is 2.2 $\pm$ 0.6% [38].

3.3 Inclusive high-P_T central muon sample

Muons at CDF are categorized by detector region and named accordingly. There are the CMU-only, CMP-only, and CMX muons. Muons reconstructed in both CMU and CMP are called CMUP muons (CMU$\oplus$CMP).

Muons for the high-P_T data sample are collected based on a series of Level 2 triggers. At Level 2, muon track segments reconstructed at Level 1, are required to match, within 5°, a CFT track with $P_T \geq 7.5$ GeV/c. Prescaling was applied to some of the Level 2 muon triggers to accommodate their large rate observed in high luminosity conditions. The increased rate was caused due to increased tracking multiplicity which result in higher accidental rate. The prescaling factors applied to the trigger was adjusted dynamically according to the instantaneous luminosity. In order to retain good efficiency for $t\bar{t}$ events, additional Level 2 triggers were implemented. These triggers were based on a combination of muon stub-CFT matching, with a requirement of a calorimeter cluster with $E_T \geq 15$ GeV. Since there was no explicit trigger path for CMP-only muons, an alternative Level 2 trigger based on large $\not\!\!E_T$ in the event, ($\not\!\!E_T \geq 35$ GeV) accompanied by two calorimeter jets was used. The Level 2 trigger path used in this analysis to accept muons is the logical OR of 16 different Level 2 triggers. The main features of these triggers are summarized in

Triggers	Prescaled
CMU-only muons	
1) $\not\!\!E_T$ > 35 GeV, Two Jets with E_T >3 GeV	NO
2) CFT track P_T >12 GeV/c matched to CMU stub	YES
3) CFT track P_T >12 GeV/c matched to CMU stub *and*	
One Jet E_T >15 GeV	NO
CMUP muons	
1) $\not\!\!E_T$ > 35 GeV, Two Jets with E_T >3 GeV	NO
2) CFT track P_T >12 GeV/c matched to CMU and CMP stubs	NO
3) CFT track P_T >12 GeV/c matched to CMU and CMP stubs *and*	
One Jet with E_T > 15 GeV	NO
CMP-only muons	
1) $\not\!\!E_T$ > 35 GeV, Two Jets with E_T >3 GeV	NO
CMX muons	
1) $\not\!\!E_T$> 35 GeV, Two Jets with E_T >3 GeV	NO
2) CFT track P_T >12 GeV/c matched to CMX stub	YES
3) CFT track P_T >12 GeV/c matched to CMX stub *and*	
One Jet with E_T >15 GeV	YES

Table 3.3: The Level 2 trigger path used for the primary muons. Muons are identified Initially at Level 1 with the reconstruction of muon stubs in the CMU, CMUP or CMX detectors. CMP-only muons are accepted through the $\not\!\!E_T + jets$ trigger as indicated in the table.

Table 3.3. The acceptance of many of these triggers is practically negligible for $W + 0$ jet events. Muons accepted at Level 2 are passed to the Level 3 stage of trigger, where full offline reconstruction of muon stubs is run along with a 2-dimensional version of the tracking reconstruction algorithm. Muons are accepted for the inclusive sample if a CTC track with $P_T \geq 18$ GeV/c extrapolates within a specific cutoff distance from a muon stub. The maximum track-stub separation distance, $|\delta x|$, is required to be 10 cm for the CMUP and CMU-only muons, 25 cm for CMX muons and 40 cm for the CMP-only muons. At Level 3, the deposited energy in the hadronic towers traversed by the muon candidate is required to be less than 6 GeV, consistent with the energy deposition expected by an isolated minimum ionizing particle.

Due to the limited muon coverage, muons that miss all of the detectors can not reconstructed. Nevertheless, their presence can be inferred from minimum ionizing tracks in the detector. They are called *CMIO* (central minimum ionizing objects) particles, if the energy deposited in the corresponding calorimeter tower is consistent with the one from a single minimum ionizing particle. CMIO muons are used in this analysis for the identification of Z and dilepton candidate events as it will be discussed in sections 3.6.1 and 3.6.2.

3.3.1 High purity muon sample

Starting from the inclusive sample, a series of additional requirements are imposed on the muon candidates in order to reduce the background from cosmic rays, hadronic punch-through and decays in flight of charged mesons. The requirements are listed below:

1. **Muon $P_T \geq$ 20 GeV.** In order to increase the resolution of the muon momentum, the track associated with the muon stub is "*beam constrained*", meaning that it is constrained to originate from the vertex closest to the muon.

2. **$E_{em} \leq$ 2.0 GeV in the muon tower.** From testbeam studies and cosmic ray data, the average energy deposition in the electromagnetic calorimeter from the passage of a muon is $\sim$ 0.4 GeV and exhibit the characteristic Landau distribution.

3. **$E_{had} \leq$ 6.0 GeV in the muon tower.** This variable similar to the E_{em} refers to the energy deposited in hadronic compartment of the calorimeter by a muon. From cosmic ray data and testbeam studies the average amount of energy deposition is $\sim$ 3 GeV.

Beam constrained momentum $P_T \geq 20.0$ GeV
Impact parameter $\lvert d_0 \rvert \leq 3$ mm
Energy in muon tower:
Electromagnetic: $E_{em} \leq 2.0$ GeV
Hadronic: $E_{had} \leq 6.0$ GeV
Total: $E_{em} + E_{had} \geq 0.1$ GeV
CTC track - muon stub matching:
$\lvert \Delta(x) \rvert \leq 2$ cm for CMU
$\lvert \Delta(x) \rvert \leq 5$ cm for CMP
Logical OR of the above two for CMUP
$\lvert \Delta(x) \rvert \leq 5$ cm for CMX
$\lvert Z_{vtx}^{muo} \rvert \leq 60$ cm
$\lvert \Delta(Z_{vtx}^{muo} - Z_0^{trk}) \rvert \leq 5$ cm

Table 3.4: Variables used for muon identification and the corresponding requirements used for the selection of *primary* muons.

4. $\mathbf{E_{had} + E_{em} \geq 0.1}$ **GeV in the muon tower.** This requirement was used to eliminate "fake" muon created mainly due to tracking reconstruction failures. In such cases a track reconstructed at the wrong ϕ can be matched to an accidental muon stub, while there is no energy deposition in the corresponding calorimeter towers.

5. **Impact parameter** $\mathbf{|d_0| \leq 3}$ **mm.** The impact parameter is defined as the distance of closest approach of the muon track to the beam position in the x-y plane. The requirement is applied to remove events in which the muon does not originate from the interaction region as it happens for cosmic rays and muons

from decays in flight.

6. $|\mathbf{Z_{vrtx} - Z_{track}}| \leq$ **5 cm.** The intercept of the muon track with the beam line is required to be within 5cm of a well defined event vertex. This requirement helps to reject cosmic rays and decays in flight.

7. $|\mathbf{Z_{vrtx}}| \leq$ **60 cm.** As for the electron case, the vertex which the muon track is associated with, is required to be within $\sim 2\sigma$ from the nominal interaction region.

8. **Track-muon stub matching ($|\Delta(x = R\phi)|$).** This variable describes the difference in ϕ between the position of the muon stub and the extrapolated CTC track. Because CMX and CMP muons traverse more material than the CMU ones, they experience more multiple scattering and therefore greater deflections. The matching requirements for the selection of good muons are listed in Table 3.4. For CMUP muons the logical OR of the CMU and CMP requirements is used.

All variables used for muon identification and the corresponding values used for the primary muon selection are summarized in Table 3.4. The distributions of the above identification variables for each detector type, are shown in figures 3.5 to 3.6 for the second leg of in a $Z \rightarrow \mu^+\mu^-$ sample. The first leg is required to satisfy all the identification criteria of a primary muon.

The distribution of the variables used for the selection of central minimum ionizing tracks (CMIO) are shown in figure 3.7.

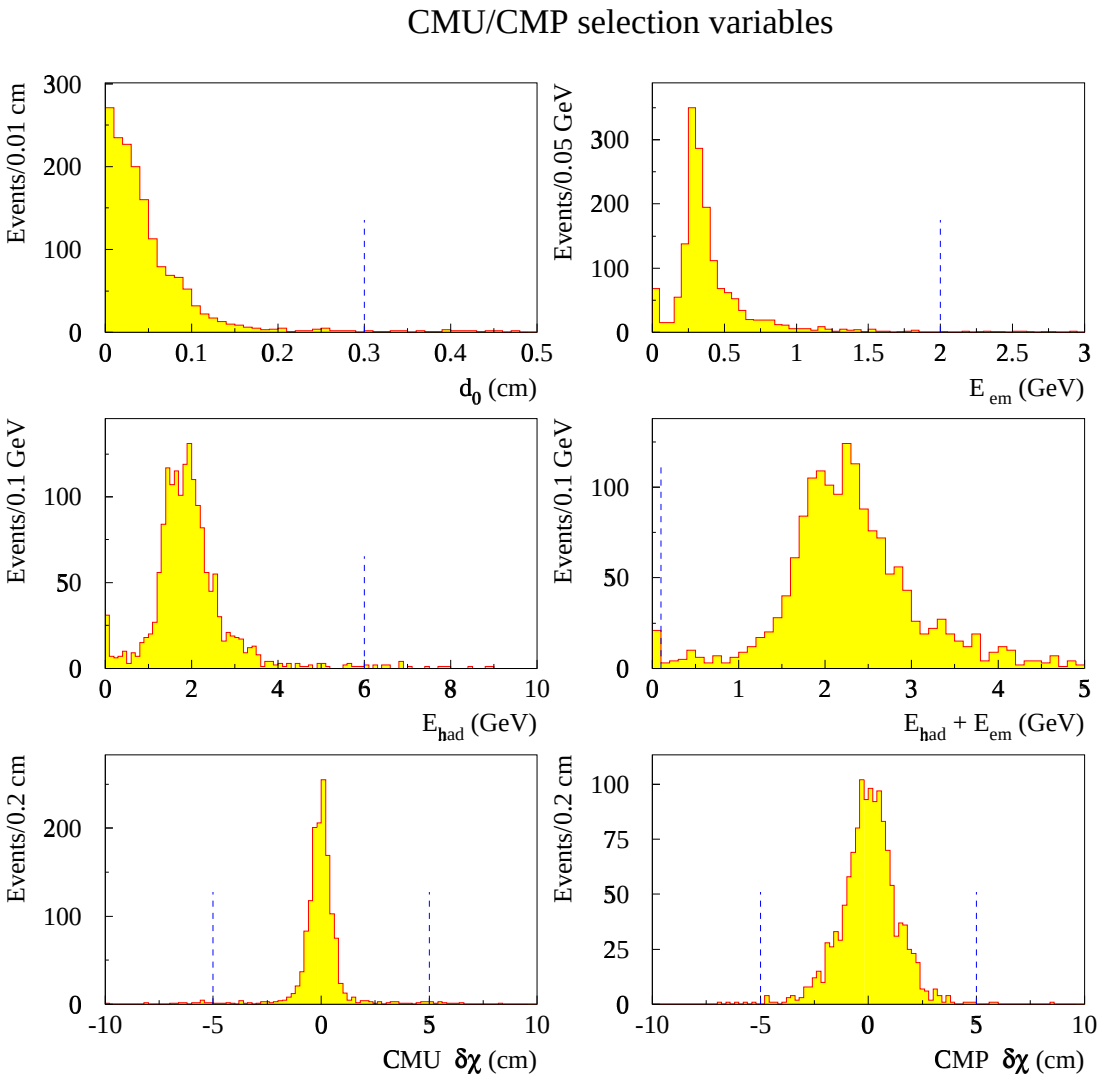

Figure 3.5: Distributions of CMU and CMUP identification variables as measured from one leg of $Z \rightarrow \mu^+\mu^-$ decays. The other muon leg is required to pass the selection criteria of a primary muon listed in Table 3.4. Dotted lines indicate the values used for the selection of primary muons.

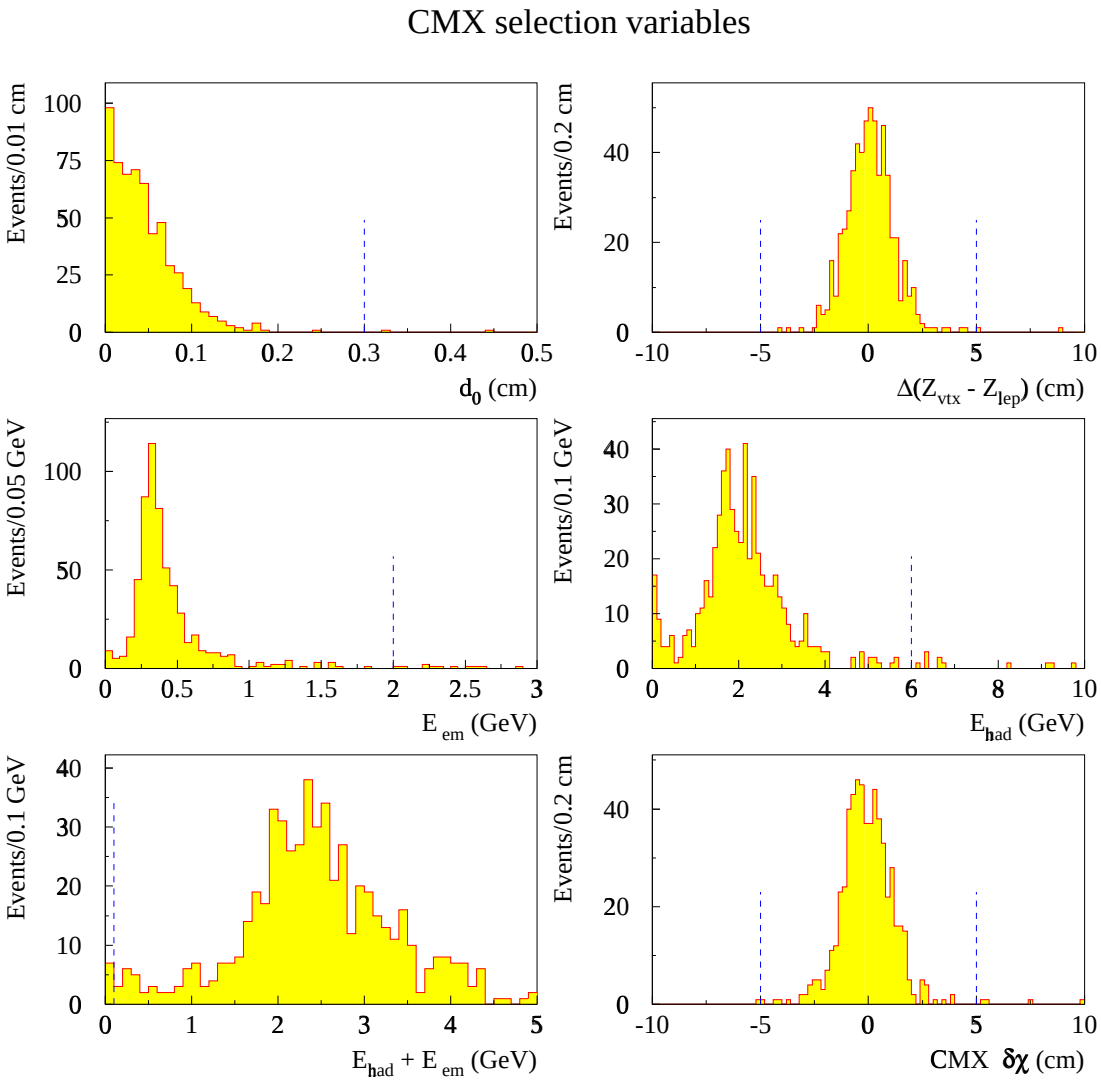

Figure 3.6: Distributions of CMX identification variables as measured from one leg of $Z \rightarrow \mu^+\mu^-$ decays. The other muon leg is required to pass the selection criteria of a primary muon listed in table 3.4. Dotted lines indicate the values used for the selection of primary muons.

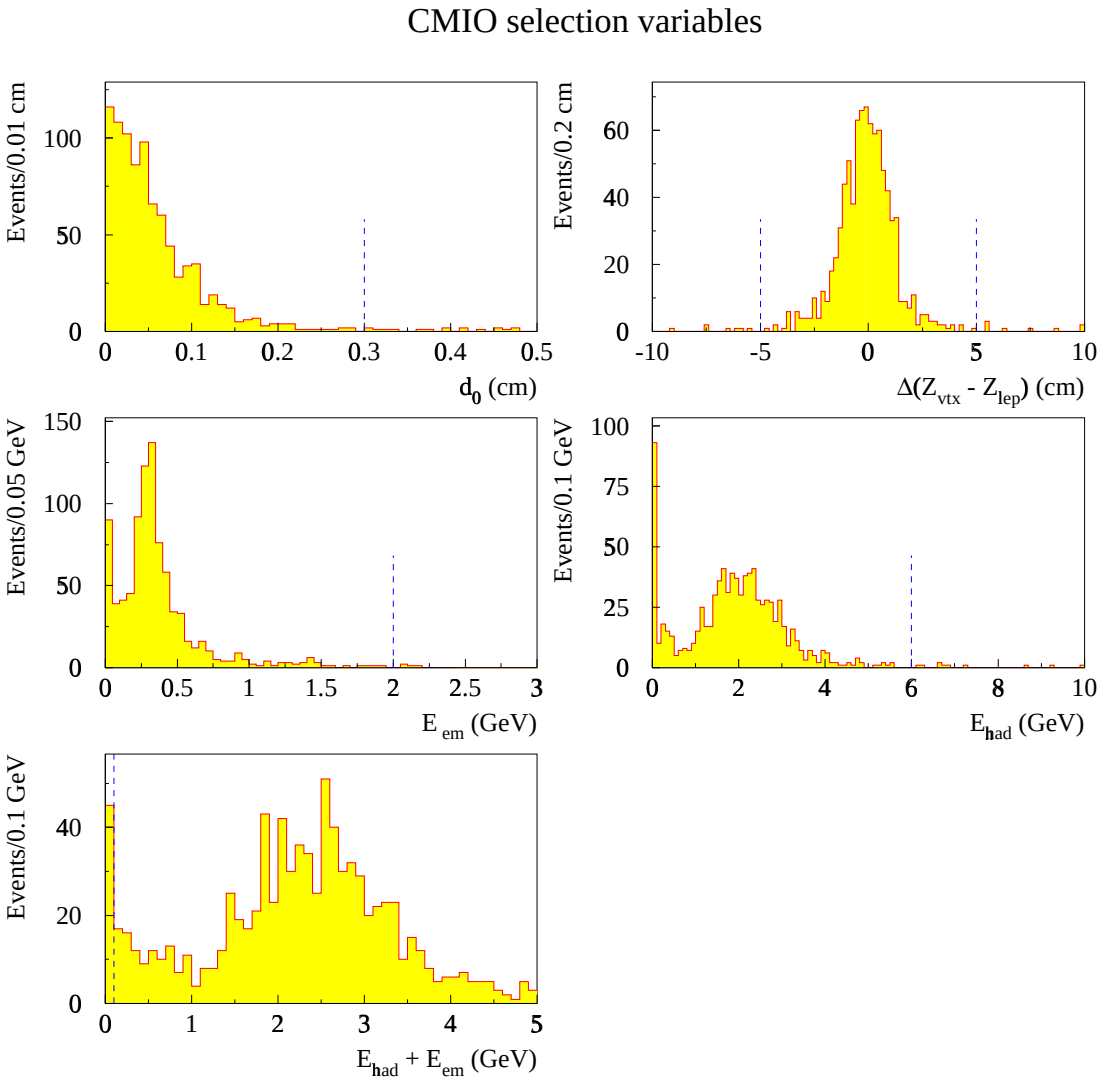

Figure 3.7: Distributions of CMIO identification variables as measured from one leg of $Z \rightarrow \mu^+\mu^-$ decays. The other muon leg is required to pass the selection criteria of a primary muon listed in table 3.4. Dotted lines indicate the values used for the selection of CMIO muons.

3.4 Lepton Isolation

Leptons from W and Z decays are expected to be isolated from other hadronic activity in the event while leptons produced in QCD events, are associated with jets of other particles produced nearby in $\eta - \phi$ space. The isolation variable is used to reduce the background from QCD processes.

The isolatation is defined for electrons and muons in a similar manner. In the electron case, the energy surrounding the electron cluster in a cone of radius $\Delta R = 0.4$ in $\eta - \phi$ space is compared to the energy of the electromagnetic cluster:

$$Isolation(Iso)^{ele} = \frac{E_T^{cone} - E_T^{ele}}{E_T^{ele}}, \tag{3.8}$$

where E_T^{cone} is the total energy (hadronic+electromagnetic) in a cone of radius $R = 0.4$ in $\eta - \phi$ space centered about the electron direction.

For muons the total transverse energy in a cone of $\Delta R = 0.4$ excluding the energy deposited in the tower traversed by the muon, is compared to the muon transverse momentum P_T

$$Isolation(Iso)^{\mu} = \frac{E_T^{cone} - E_T^{tower}}{P_T^{\mu}} \tag{3.9}$$

The primary leptons in W and Z events are required to have isolation $\mathrm{Iso}^{\ell} \leq 0.1$.

The lepton isolation variable is process dependent. Events with a lot of jet activity tend to have less isolated leptons and in such cases, the isolation efficiency is measured with Monte Carlo simulations.

3.5 Lepton Identification efficiencies

The efficiency of the lepton requirements is estimated from a sample of $Z \rightarrow \ell^+\ell^-$. The Z sample used for the efficiency study is selected from the inclusive lepton samples with the following criteria:

1. Two central fiducial leptons (electrons or muons) with $E_T \geq 20$ GeV. At least one lepton is required to pass all the primary lepton requirements listed in Tables 3.1 and 3.4.

2. Both leptons are required to be isolated with $\mathrm{Iso}^{\ell}_{cal} \leq 0.1$

3. The primary lepton vertex is required to be within $|Z^{lep}_{vtx}| \leq 60$ cm

4. The dilepton invariant mass to fall in the mass range $70 \leq M_{\ell\ell} \leq 110$ GeV/c^2.

The selection efficiencies are determined by applying each of the selection requirements on the second leg of the Z. The efficiency is determined according to the following arguments. Assuming there are N_{tot} events of $Z \rightarrow \ell^+\ell^-$ that could enter in the specified mass window before any selection, the events can be divided according to the probability, ϵ, of the leptons to pass any given selection requirement in four categories:

1. Both leptons pass the selection requirements: $N_{pp} = N_{tot}\epsilon^2$

2. First lepton passes the selection and second fails: $N_{pf} = N_{tot}\epsilon(1-\epsilon)$

3. First lepton fails the selection and second passes: $N_{fp} = N_{tot}(1-\epsilon)\epsilon$

4. Both leptons fail the selection requirements: $N_{ff} = N_{tot}(1-\epsilon)^2$

Since the sample selection requires at least one lepton to satisfy all the primary lepton criteria, events belonging to category 4 do not appear in the sample. The events, N_{obs}, in the other three categories are the ones that enter the selected sample and the efficiency ϵ, can be measured by counting the number of events in which both leptons pass a specific cut (category 1) divided by the total number of events in the sample (categories 1, 2 and 3). Due to the sample construction, events in categories 2 and 3 are examined twice and therefore the number of events, N_1, in categories 2 and 3 has to be divided by two to properly account for the number of events considered.

$$\begin{aligned}
\epsilon &= \frac{N_{pp}}{N_{pp} + \frac{1}{2}(N_{pf} + N_{fp})} \\
\epsilon &= \frac{\epsilon^2 N_{tot}}{\epsilon^2 N_{tot} + \frac{1}{2}(2\epsilon(1-\epsilon)N_{tot})} \\
\epsilon &= \frac{2N_{pp}}{2N_{pp} + N_{obs} - N_{pp}} \\
\epsilon &= \frac{2N_{pp}}{N_{obs} + N_{pp}} \\
\epsilon &= \frac{2R}{1+R}
\end{aligned}$$

where R is defined as $R = \frac{N_{pp}}{N_{obs}}$, the ratio of events with both leptons passing the selection requirement over the total number of events in the sample.

The above arguments hold for the case of two indistinguishable leptons which is the case for central-central electrons and CMUP-CMUP Z events. Due to small statistics of the CMX-CMX Z sample, the efficiency of the CMX muon requirements was determined from a sample of CMUP-CMX Zs, requiring the CMUP muon to satisfy all the primary muon requirement listed in Table 3.4. In this case the CMX

Lepton type	Efficiency
Electrons	0.81 ± 0.01
CMUP	0.94 ± 0.01
CMU-only	0.92 ± 0.03
CMP-only	0.92 ± 0.03
CMX	0.92 ± 0.02

Table 3.5: The lepton identification efficiencies including the isolation requirement, as measured in a sample of $Z \rightarrow \ell\ell$ events. The errors are statistical only.

selection efficiencies are determined by:

$$\epsilon = \frac{N_p}{N_{obs}}$$

where N_p is the number of events passing the specific CMX selection and N_{obs} is the total number of events in the sample.

The efficiencies of the criteria applied to the selection of primary electrons and muons are listed in Table 3.5 for each lepton type.

3.6 Specific background vetoes

The sample of selected primary leptons contains still a number of real leptons from processes other than W decays. Some of these processes can be identified on an event-by-event basis directly from data and the corresponding candidate events are removed before selecting the final W sample.

3.6.1 Z-removal

Leptonic decays of the Z result in large P_T electrons or muons which have the same characteristics as leptons from W-decays. Z events are not expected to be associated with large $\not\!\!E_T$. Nevertheless, when one of the leptons escapes detection or is poorly reconstructed, results in large $\not\!\!E_T$ and the event looks like a W candidate.

To identify Z candidates one lepton is required to be in the central detector region and satisfy the primary lepton selection criteria (see Tables 3.1 and 3.4). The second lepton is allowed in any detector region and is required to satisfy "looser" selection criteria as shown in Table 3.6. For every pair, the invariant mass of the two lepton system is formed as follows:

$$M_{ee} = \sqrt{(E^1_{tot} + E^2_{tot})^2 - (E^1_x + E^2_x)^2 - (E^1_y + E^2_y)^2 - (E^1_z + E^2_z)^2}$$
$$M_{\mu\mu} = \sqrt{(|\vec{P}^1| + |\vec{P}^2|)^2 - (P^1_x + P^2_x)^2 - (P^1_y + P^2_y)^2 - (P^1_z + P^2_z)^2}$$

where E_i and $|\vec{P}_i|$ are the energy and momentum of the two electron, muon legs respectively. Events are flagged as Z candidates if the invariant mass of the lepton pair falls in the range $70 \leq M_{\ell\ell} \leq 110$ GeV/c^2. In the case of PEM/FEM electrons where there is inefficient or non-existent tracking coverage, the E_T of the electromagnetic cluster is calculated with respect to the primary electron vertex instead of the highest hit occupancy vertex.

Special consideration is paid to the central loose electrons. Figure 3.8(a) shows the second electron E/P distribution as a function of the electron pair invariant mass. It is clear that a lot of events appear in the Z mass window of (70-110 GeV/c^2) for large values of E/P. Recall that for primary electrons it is required $E/P \leq 1.8$. However, hard photon bremsstrahlung off the electron results in lower electron momentum

Electrons
1) Region CEM or PEM or FEM
2) $E_T \geq 10.0$ GeV
3) $E_{had}/E_{em} \leq 0.12$
4) Isolation ($\Delta R = 0.4$) ≤ 0.15
6) For non fiducial candidates in CEM/PEM:
i) If $0 \leq E/P \leq 10.0$ use track P_T instead of E_T
ii)Otherwise add the component of $\not\!\!E_T$ in lepton direction to the E_T
Muons with reconstructed stub
1) $P_T \geq 10.0$ GeV
2) $Isolation(\Delta R = 0.4) \leq 0.15$
3) Impact parameter $\lvert d_0 \rvert \leq 5$ mm
4) Common vertex of primary and secondary muon $\lvert\Delta(Z_{vtx}^{prim.\ \ell} - Z_{trk}^{\ell})\rvert \leq 10$ cm
5) Opposite charge to primary muon
6) Stub - CTC track matching:
$\lvert\Delta(x)\rvert \leq 10$ cm for all muon types (CMU/CMP/CMX)
7) Energy in μ-tower $E_{had} \leq 10.0$ GeV and $E_{em} \leq 5$ GeV
Minimum Ionizing Tracks (CMIO)
8) Energy in tower traversed by the track:
$E_{had} + E_{em} \leq 10$ GeV and either $E_{em} \leq 2.0$ GeV or $E_{had} \leq 6.0$ GeV
$70 \leq M_{\mu\mu} \leq 110$ GeV/c^2

Table 3.6: Variables and the corresponding selection criteria used to define loose leptons. Loose leptons are paired with same flavor primary leptons in order to identify and remove Z candidate events, as indicated in the table.

and therefore large values of E/P (most often the emitted photon lands at the same calorimeter tower as the electron). Some secondary effects contributing to large values of E/P are due to tracking algorithm failures. Clearly, allowing values of $E/P \leq 10$ for the second leg more Z candidates can be identified.

For non-fiducial electrons in the CEM and PEM regions, the E/P measurement is worse since the electron lands close to a low response detector region and its energy is mismeasured. In this case, the momentum measurement is closer to the true electron energy while E/P measures less than 1. For this type of electrons, the invariant mass of the electron pair is formed using the momentum of the second leg instead of the cluster energy. In Figure 3.8(d), the electron pair invariant mass formed using the energy of the 2nd leg, is compared to the mass obtained using the momentum of the 2nd leg for the cases of non-fiducial electrons with E/P$<$ 1.0. It is evident that more events appear in the Z mass window using the momentum information on the 2nd leg.

The lepton pair invariant mass distribution after all selection criteria listed in Table 3.6 applied, for electron and muon pairs for all jet multiplicities and for events with at least one 0.4 cone jet with $E_T \geq 15$ GeV and $|\eta_D| \leq 2$ is shown in figure 3.9. The described Z-removal selection criteria reduce the acceptance for $t\bar{t}$ events by $\sim 2\%$ for a top mass of $M_{top} = 170\text{GeV/c}^2$.

3.6.2 Dilepton Removal

In addition to the Z removal, another class of events with two leptons in the final state were also removed. These events arise from Drell-Yan lepton pair production, $Z \rightarrow \tau^-\tau^+$ followed by leptonic decay of one or both τ leptons and from $t\bar{t}$ production in which both Ws decay to leptons. There are also some remaining $Z \rightarrow e^+e^-$ events

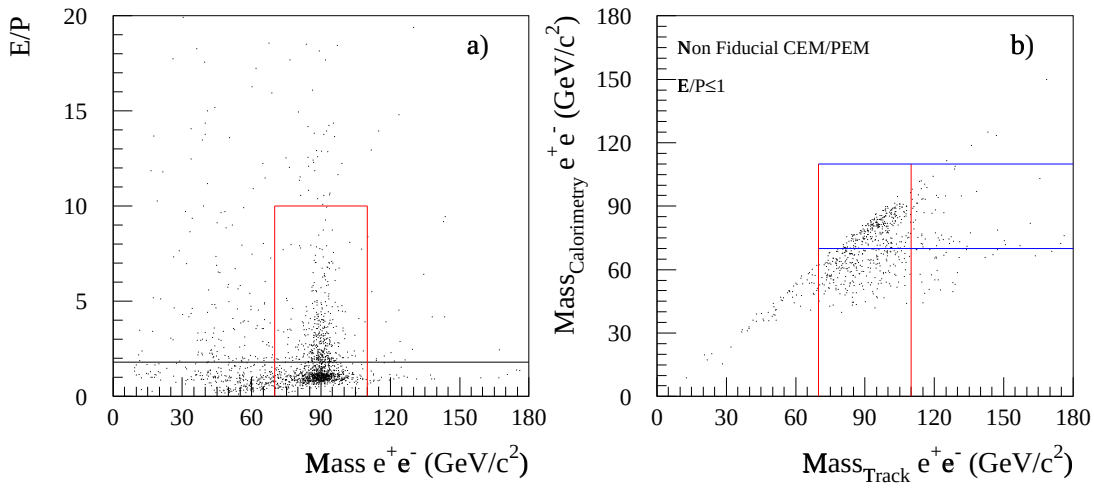

Figure 3.8: a): E/P distribution of loose CEM electrons as a function of the electron pair invariant mass. A large number of events appear in the Z mass region with poorly measured E/P for the 2nd leg. The horizontal line indicates the default value of E/P used to select primary electrons. b): The electron pair invariant mass reconstructed with the momentum of the track associated with the 2nd leg versus the pair mass obtained using the 2nd electron energy, for non-fiducial electrons in CEM and PEM regions with E/P$<$ 1. Vertical lines in all plots indicate the mass window ($70 \leq M_{\ell\ell} \leq 110 \mathrm{GeV/c^2}$) used for the identification of the Z candidate events and horizontal lines the selection applied to second loose lepton.

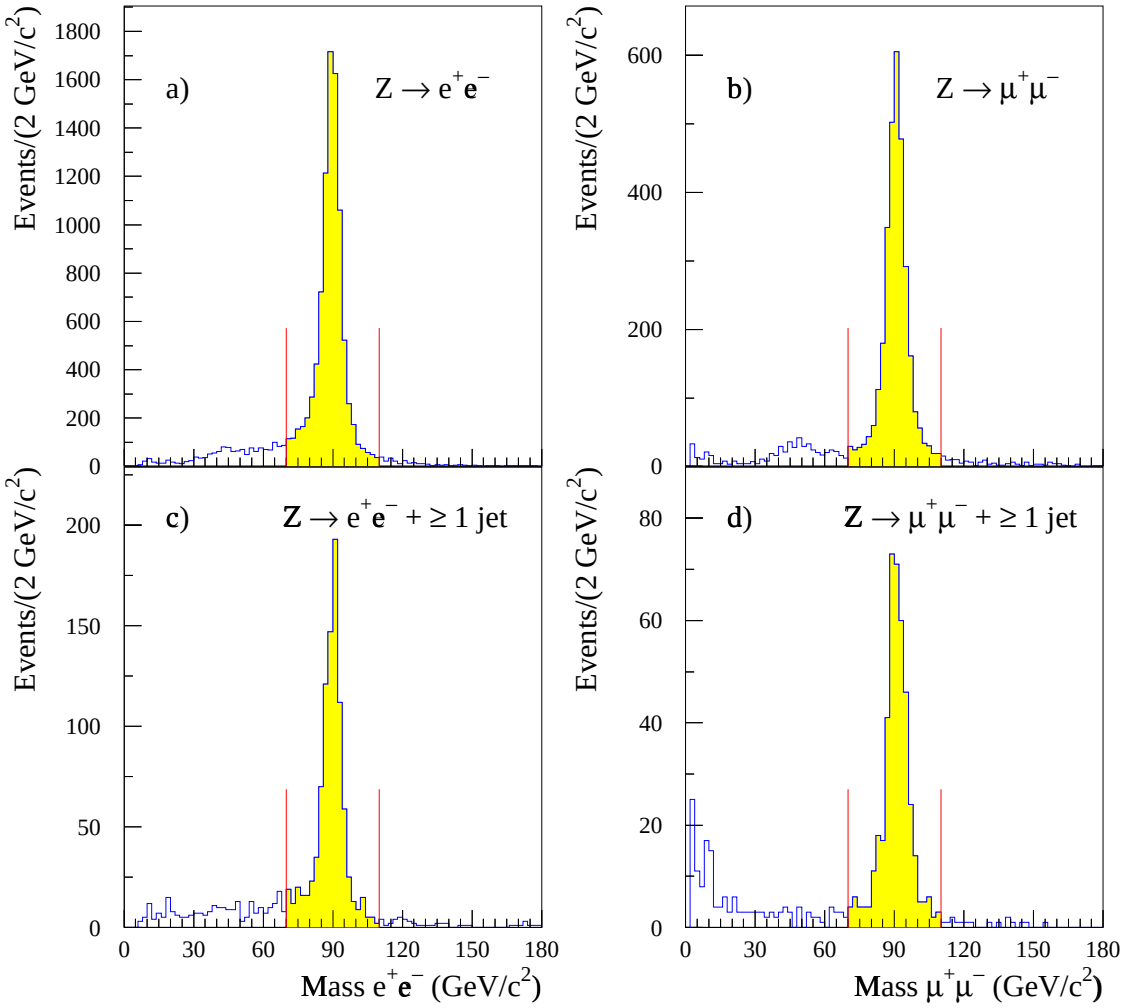

Figure 3.9: Invariant mass of electron and muon pairs in the data for all jet multiplicities and for events with at least one jet clustered with 0.4 cone radius, $E_T \geq 15$ GeV and $|\eta_D| \leq 2$. Shaded areas indicate the mass range used to select Z candidate events. Events at the low tail of the $M_{\mu^+\mu^-}$ distribution (plots (b) and (d)) are due to Υ and J/Ψ production.

in which one electron is not reconstructed because either the track is missing due to limited tracking coverage or tracking reconstruction problems and therefore the electron is reconstructed as an electromagnetic jet, or it falls inside a detector crack and its energy is completely lost. In the last case the remaining electron track in the detector regions that can be reconstructed, can be used to help identify potential Z candidates.

The dilepton removal proceeds as follows:

1) Removal of events with two leptons of same type and opposite charge and with invariant mass $M_{\ell\ell} \geq 2$ GeV/c^2 (see figures 3.9(a) and (b)). One lepton is required to satisfy the primary lepton criteria while the second lepton is required to pass the loose lepton requirements listed in Table 3.6.

2) Removal of events containing two leptons of different type (e,μ) regardless of charge. As before, the requirement is for a tight and a loose lepton.

3) Removal of events containing a primary lepton and an isolated track of $P_T \geq$ 10 GeV/c with charge opposite to the primary lepton and with invariant mass of the lepton-track system $M_{\ell\ trk} \geq 2$ GeV/c^2. Events of this type arise from Z and $t\bar{t}$ decays in which either one of the leptons is lost in a detector crack or the lepton is a τ which decays hadronically.

The criteria used to identify an isolated track leg, (call it a τ-candidate) are listed in Table 3.7. The track isolation requirement, $\sum_{j=1}^{ntrack}$, is defined as the sum of the momenta of all tracks with $P_T \geq 400$ MeV/c in a cone of 0.4 in $\eta - \phi$ space around the candidate track. These tracks are required to have impact parameter $|d_0| \leq$ 1cm and $|Z_{trk} - Z^{\ell}_{vtx}| \leq 5$ cm. Figures 3.10(b) and (c) show the P_T and η distributions of the isolated track candidates in the ℓ+jets sample after removing events with any $\ell\ell$ pair. In a large fraction of the events the track appears in major detector cracks

A 3-dimensional track of $P_T \geq 10$ GeV/c
Track impact parameter $d_0 \leq 0.5$ cm
$\lvert Z_{trk} - Z_\ell \rvert \leq 5$ cm
Isolation, $\sum_{j=1}^{ntracks} P_T^j \leq 2$ GeV/c, in a cone of 0.4 around the candidate track

Table 3.7: The criteria applied for the definition of an isolated track or a τ-candidate.

($|\eta| \leq 0.05$ and $1.0 \leq |\eta| \leq 1.1$). The distributions of $\not\!\!E_T$ and azimuthal separation, $\Delta(\phi_{\not\!\!E_T,trk})$, of the candidate track and $\not\!\!E_T$ are shown in figures 3.10(a) and (d).

There are 1038 events in which the track candidate and the primary lepton are *opposite sign* - *OS* and 554 of these events appear in the region of $\not\!\!E_T \geq 20$GeV. For comparison, the rate of events with *same sign* - *SS* is 235 of which 109 have $\not\!\!E_T \geq 20$ GeV.

The invariant mass distribution of the opposite and same sign events are shown in figure 3.11(a). Figure 3.11(b) shows the invariant mass distribution of the opposite sign $\ell - track$ events after same sign event subtraction. For comparison, the corresponding distribution of a $Z+ \geq 1$ jet PYTHIA simulation is shown in the same figure. The Monte Carlo sample is normalized to the same number of $Z \rightarrow \ell\ell$ events observed in the data. Very good agreement is found between the two samples, justifying the removal of events of this kind. The invariant mass distributions of opposite and same sign events for different jet multiplicities in the $W+ \geq 1$ jet events are shown in figure 3.12.

4) Removal of events with a primary lepton and an electromagnetic jet, *em-jet*. An em-jet is defined as a 0.4 cone jet with $E_T \geq 15$ GeV and $|\eta_D| \leq 2.1$, with large electromagnetic fraction, Emfr= $E_{em}/E_{total} \geq 0.95$ and containing less than three tracks. As already mentioned, the purpose of removing this kind of event

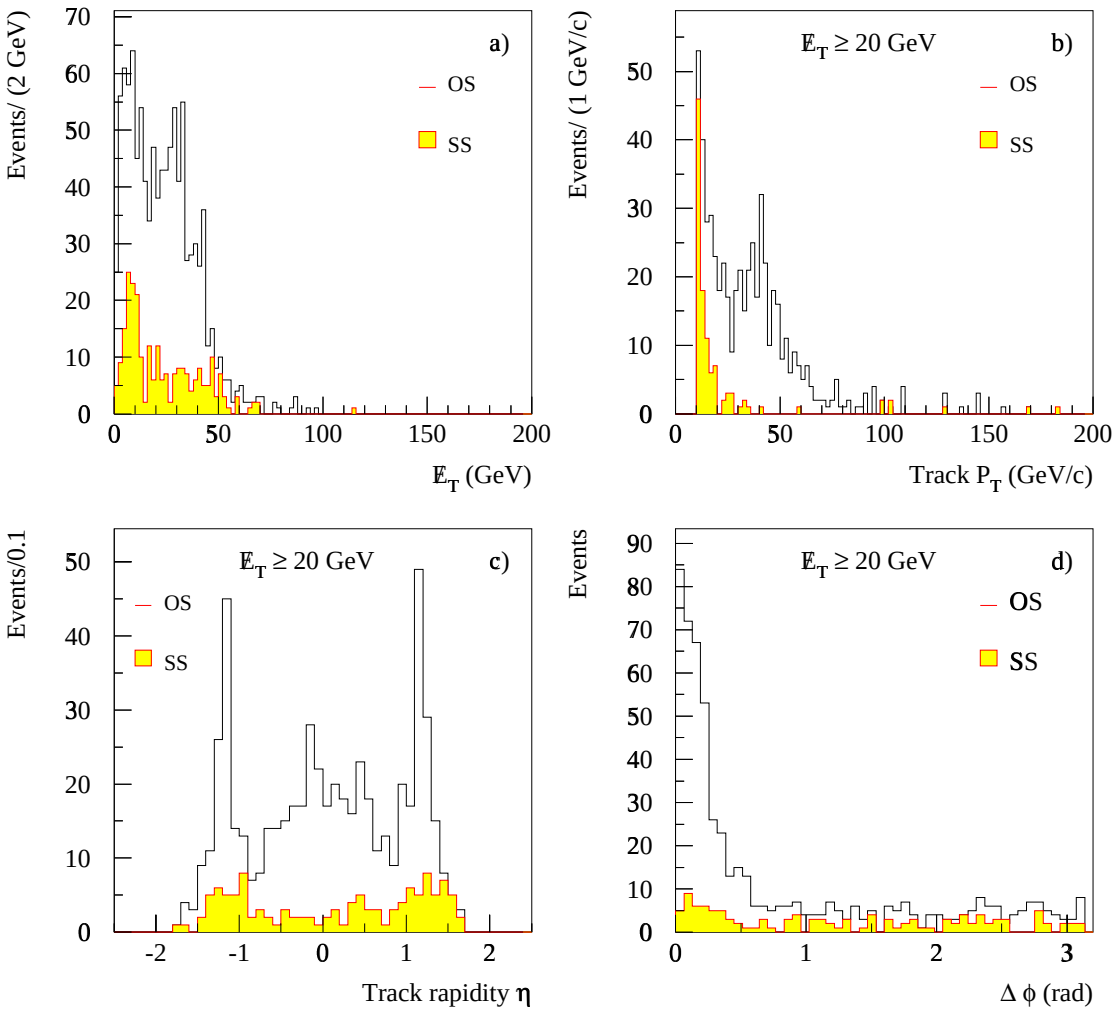

Figure 3.10: Kinematic distributions for events with a primary lepton and an isolated track, after removing events with any lepton pair. a): $\not\!\!E_T$ in the event, b): Track P_T, c): Track η and d): $\Delta\phi$ between track and $\not\!\!E_T$. Shaded histograms show the corresponding distributions for lepton-track pairs of the same charge. The last three distributions correspond to events with $\not\!\!E_T \geq 20$ GeV.

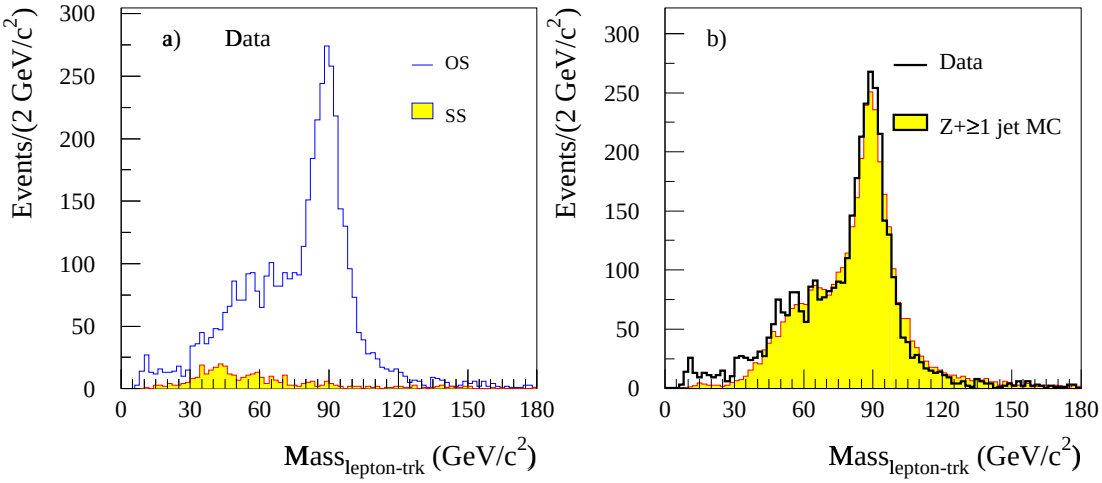

Figure 3.11: Invariant mass distribution of primary ℓ isolated track pairs, in the $\ell+ \geq$ 1 jet sample. a): Data with lepton-track pairs of opposite charge, *OS*, (solid) and pairs of same charge, *SS*, (shaded). b): OS-SS lepton-track invariant mass distribution compared to a PYTHIA $Z+ \geq 1$ jet simulation. The Monte Carlo sample is normalized to the number of $Z \rightarrow \ell\ell$ events observed in the data.

is to eliminate Z candidates in which an electron track is not reconstructed either because the electron undergoes a hard bremsstrahlung emission and the track changes direction pointing to a different calorimeter tower than the electromagnetic cluster, or the tracking algorithm fails completely to reconstruct the track and instead appears as an electromagnetic jet. Contributions to this background arise from decays like $Z \rightarrow \tau^+\tau^-$ in which one of the τ's decays leptonically while the other decays hadronically i.e. $\tau \rightarrow \pi^{\pm}\pi^0$. Because in the leptonic decays, part of the tau energy is carried away by the neutrinos, the invariant mass of the lepton and electromagnetic jet peaks at approximately 60 GeV/c^2. The invariant mass of the primary lepton and em-jet pairs is shown in figure 3.13(a) for events in the lepton+ $\geq$ 1 jet sample before

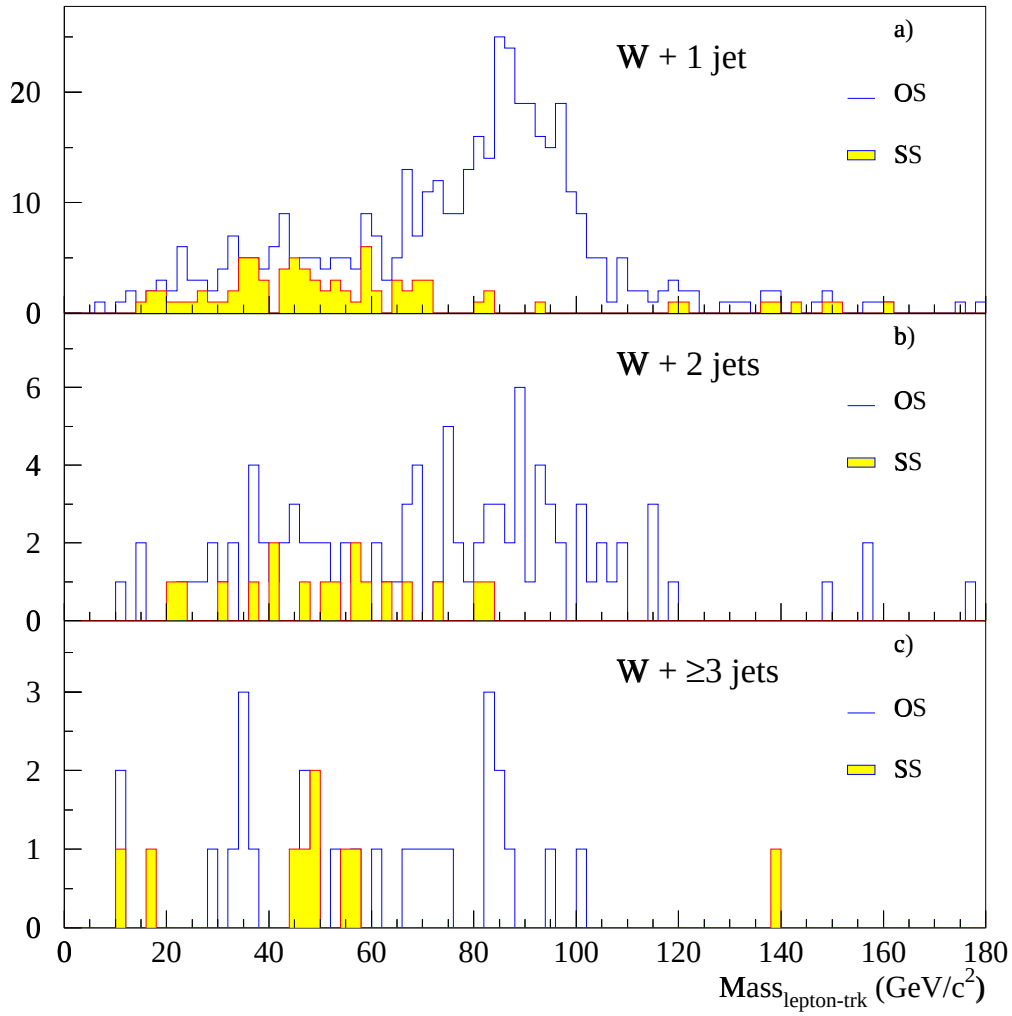

Figure 3.12: Invariant mass distribution of opposite sign (open) and same sign (shaded) ℓ-isolated track pairs and after removing events with any $\ell\ell$ pair in $W + 1$ jet events a), $W + 2$ jet events b) and $W + \geq 3$ jet events c).

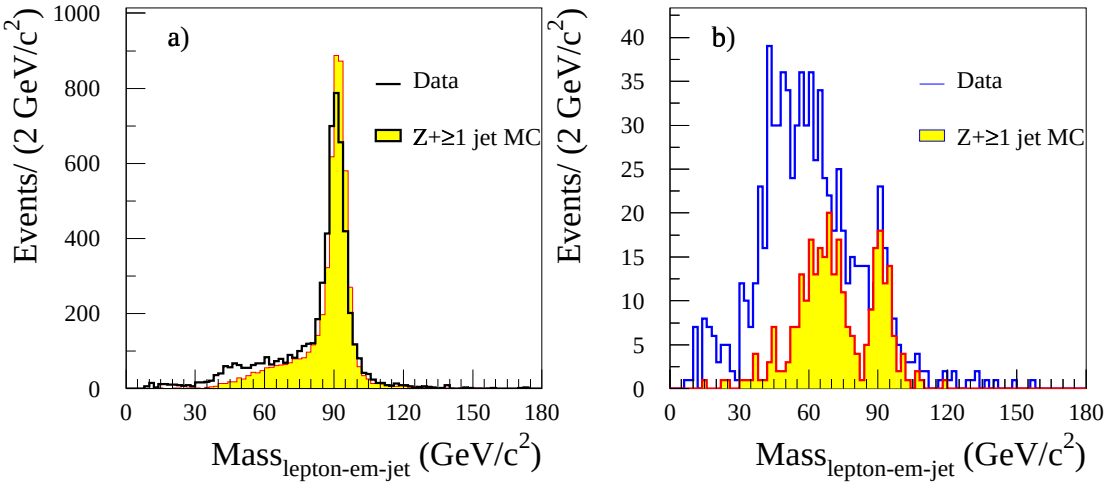

Figure 3.13: Invariant mass distribution of primary ℓ em-jet pairs, in the $\ell+ \geq 1$ jet sample. a) Data, all events b): After removing $\ell\ell$ pairs in data (solid) and PYTHIA $Z+ \geq 1$ jet simulation (shaded). Data and simulation are normalized to the same number of $Z \rightarrow \ell\ell$ events.

removing $\ell\ell$ pairs. After removing $\ell\ell$ pairs, the distribution of the invariant mass of the lepton-em jet pair in the data and in a $Z+ \geq 1$ jet PYTHIA simulation, is shown in figure 3.13(b). The Monte Carlo sample is normalized to the number of $Z \rightarrow \ell\ell$ events seen in the data. Two mass peaks are observed in the Monte Carlo distribution. The peak at $\sim$ 60 GeV/c^2 corresponds to $Z \rightarrow \tau^+\tau^-$ events while the second peak at $90 GeV/c^2$ is due to $Z \rightarrow e^+e^-$ events where one electron is badly reconstructed or undergoes hard bremsstrahlung. In general, there is fair agreement between data and Monte Carlo above $60 GeV/c^2$ but the data show an excess of 500 events for lower masses arising from photon jets and fluctuations of hadronic showers.

Examining a large sample of jet data collected with a trigger requiring a jet of $E_T \geq 50$ GeV (*JET_50 sample*), it is found that 0.4% of the events have a jet satisfying

the em-jet selection criteria. The primary $\ell+ \geq 1$ jet sample consists of 33376 events and therefore 150 events are expected to have an em-jet. The remaining 350 events comprise 2% of the total number of events with a $\ell\ell$ pair and the disagreement between data and Monte Carlo is reasonable. On the other hand, there are 12525 $W+ \geq 1$ jet events after $Z \rightarrow \ell\ell$ removal and the expected number of em-jet pairs removed by mistake is approximately 50 events.

The dilepton removal procedure described above reduces the acceptance for $t\bar{t}$ events in the $W+ \geq 3$ jet signal region by $\sim$ 17.2% for top mass of $M_{top} = 170$ GeV/c^2 while it removes most of the $t\bar{t}$ dilepton candidate events in the $W + 1, 2$ jet region.

3.7 Selection of the W Sample

As mentioned earlier, leptons from W/Z boson decays are expected to be isolated from the rest of the event activity. Also in leptonic W decays, the presence of a very energetic neutrino results in a significant energy imbalance in the event (large $\not\!\!E_T$). These two characteristic features of leptonic W decays are used to select the final W sample which is used in this analysis.

For W selection, the lepton is required to have isolation of $\mathrm{Iso}^{\ell}_{cal} \leq 0.1$. This requirement removes most of the events in which the lepton is produced via the semileptonic decays of heavy quarks while it is 87% and 81% efficient for electrons and muons respectively, in $t\bar{t}$ events with ≥ 3 jets. Comparison of the primary lepton isolation distribution in $\ell+ \geq 1$ jets events before and after requiring $\not\!\!E_T \geq$ 20 GeV is shown in figure 3.14(a). After the $\not\!\!E_T$ requirement is applied, selecting mainly W events, the isolation distribution of the lepton resembles the one of primary

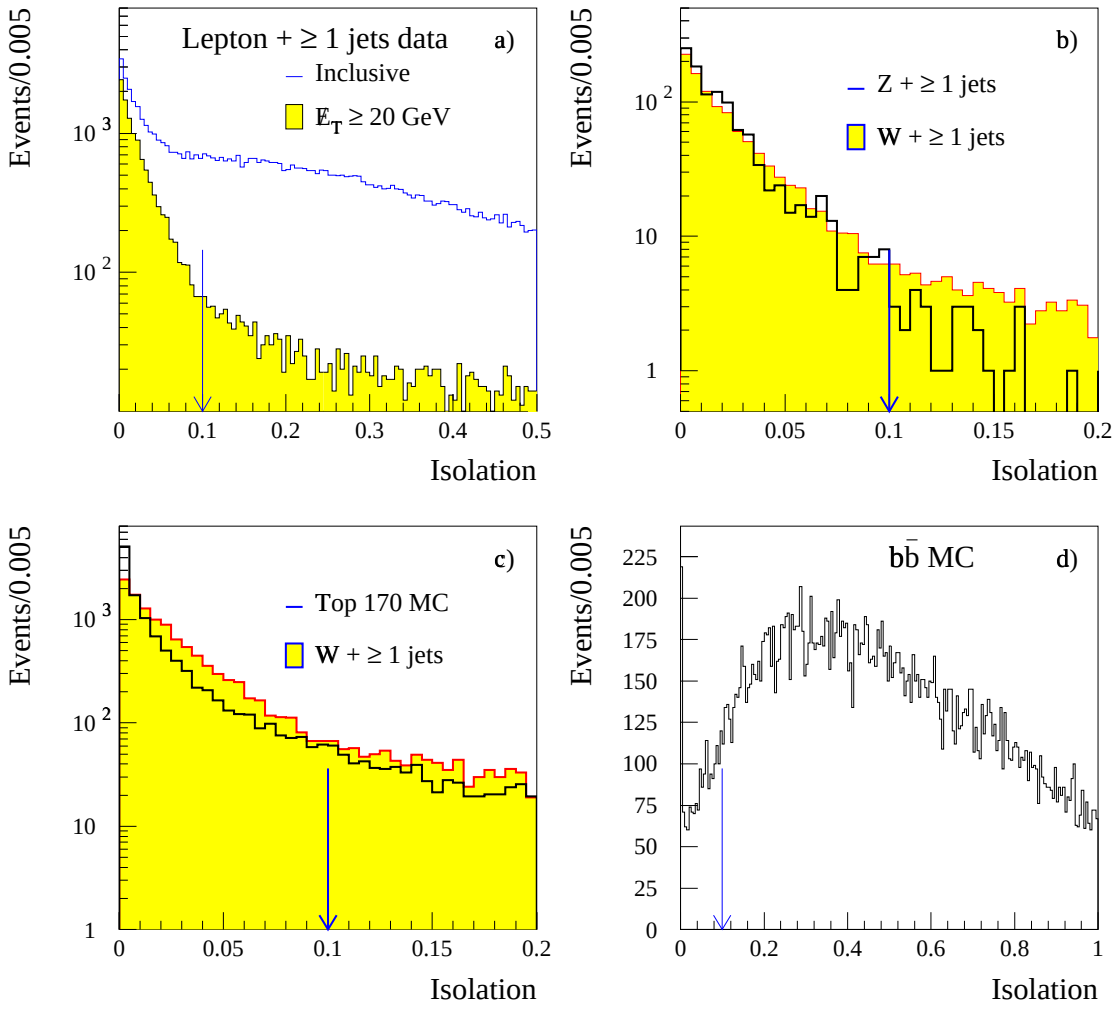

Figure 3.14: Primary lepton isolation distributions in a): inclusive $\ell+ \geq 1$ jet events (solid) and $W+ \geq 1$ jet events (shaded), b): $Z \rightarrow \ell^+\ell^-$ events (solid) and $W+ \geq 1$ jet events (shaded), c): PYTHIA $t\bar{t}$ Monte Carlo simulation and d): PYTHIA $b\bar{b}$ Monte Carlo simulations.

leptons in Z events as shown in figure 3.14(b). Figure 3.14(c) and (d) show the isolation distribution of primary leptons in PYTHIA [70] $t\bar{t}$ and $b\bar{b}$ Monte Carlo events respectively. The final selection for W events relies on the requirement of large

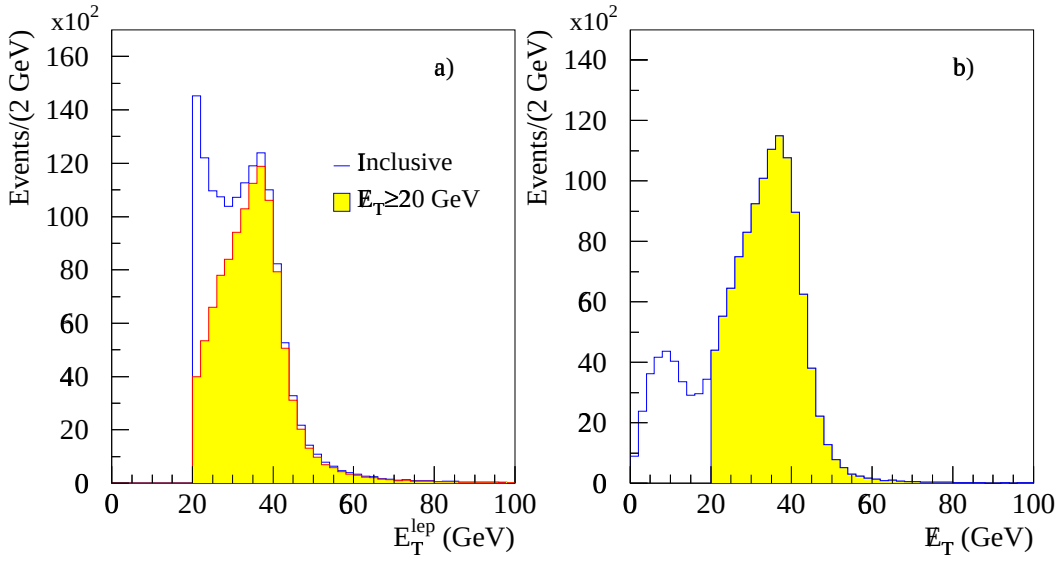

Figure 3.15: a): Distribution of the isolated lepton E_T or P_T before (open) and after (shaded) the application of the $\not\!\!E_T \geq 20$ GeV requirement. b): The distribution of the $\not\!\!E_T$ in the sample of isolated primary leptons. The shaded region corresponds to the $\not\!\!E_T$ distribution of the final W sample. The presence of W leptons is evident by the characteristic Jacobian peaks in both distributions.

transverse energy imbalance in the event. W events are selected requiring $\not\!\!E_T \geq 20$ GeV. The $\not\!\!E_T$ is calculated as described in section 3.1.9 and it is corrected only for the presence of primary muons in the event.

The distribution of the lepton E_T before and after the $\not\!\!E_T$ requirement is shown in figure 3.15(a). Figure 3.15(b) shows the distribution of $\not\!\!E_T$ in the selected W

sample. Two peaks are observed. The peak at low values of $\not\!\!E_T$ is dominated mainly of QCD events and events from Z decays. The $\not\!\!E_T$ for the QCD background events is expected to arise from detector resolution and cracks. The second peak at ~ 40 GeV/c is consistent with the neutrino coming from a leptonic W decay. The presence of W events is evident by the characteristic Jacobian peaks in both E_T and $\not\!\!E_T$ spectra. It is also evident from figure 3.15(a) that requiring high $\not\!\!E_T$, a large fraction of the remaining QCD background contribution is eliminated.

Table 3.8 shows the number of events surviving the selection criteria in the electron and muon samples.

Selection	Electron sample	Muon sample
Initial sample	638949	747345
Lepton quality	156694	110996
Bad run removal	148410	105578
Trigger	143714	93010
Isolation, $I \leq 0.1$	101264	60613
Z removal	92166	57480
Dilepton removal	88239	54347
$\not\!\!E_T \geq 20$ GeV	68306	42100

Table 3.8: Number of events surviving the various W selection requirements in the inclusive electron and muon samples.

3.7.1 Features of W events

The longitudinal momentum of the neutrino cannot be reconstructed since the momenta of the colliding partons are unknown. This prevents the full reconstruction of

the $\ell\nu$ mass system. Instead the transverse mass, $M_T^{\ell\nu}$, of the lepton-neutrino system can be reconstructed and is expected to exhibit the characteristic smeared Jacobian with a peak at $M_T^{\ell\nu} = 80$ GeV/c^2. The transverse mass is defined according to the following equation:

$$M_T^{\ell\nu} = \sqrt{(E_T^\ell + \not\!\!E_T)^2 - (\vec{P}_T^\ell + \vec{\not\!\!E}_T)^2} = \sqrt{2E_T^\ell \not\!\!E_T \cos(1 - \delta\phi_{\ell\vec{\not\!\!E}_T})} \qquad (3.10)$$

where E_T^ℓ refers to the magnitude of the transverse energy or momentum of the electron or muon respectively and $\Delta\phi_{\ell\vec{\not\!\!E}_T}$ to the azimuthal angle between the direction of the lepton and neutrino. Figure 3.16 shows the transverse mass distribution for electron and muon events in the final W sample.

3.8 Jet multiplicity yield in the W sample

The number of W events observed as a function of the jet multiplicity are listed in Table 3.9 for the electron and muon channel separately. Jets are clustered with a cone of 0.4 radius in $\eta - \phi$ space. Each jet is required to have $E_T \geq 15$ GeV and its centroid to lie in the pseudorapidity range of $|\eta_D^{jet}| \leq 2$. The pseudorapidity and E_T selection requirements are imposed to increase the signal to background ratio for large top mass $t\bar{t}$ events. The lowest order matrix elements for the production of $W+$ jet production have been computed for up to 4 partons [50]. The W+N jet cross section is proportional to α_s^N and therefore, for each additional jet in a W event, the cross section falls by a factor of α_s. The exact knlowledge of the $W + N$ jet rate is important not only because it is the dominant background to the $t\bar{t}$ events, but also on other exotic processes. However, there are large uncertainties in the tree-level predictions of the W+N jet cross section mainly due to the dependence of α_s on the

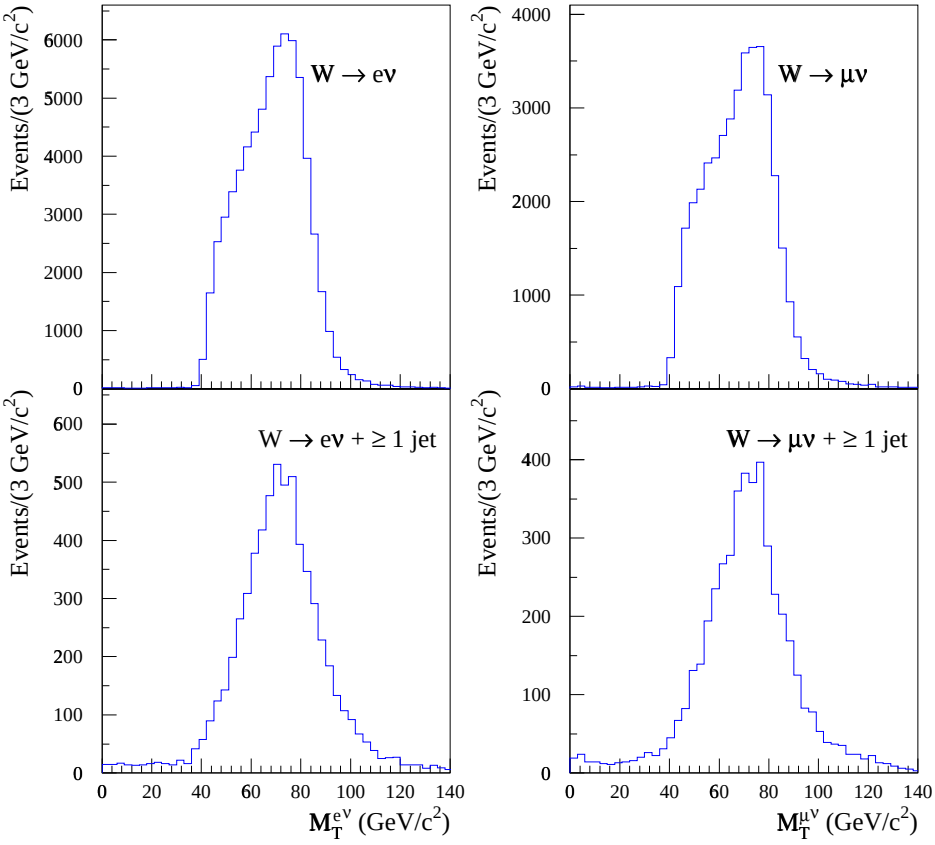

Figure 3.16: Transverse mass distribution of the $\ell - \nu$ system after the final selection of W events in the electron and the muon sample. Also shown the transverse mass distribution of $W+ \geq 1$ jet for $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$.

unphysical factorization and renormalization scale μ. These uncertainties are greatly reduced with the addition of next-to-leading order calculation. Unfortunately, next-to-leading corrections exist only for the zero and 1 jet bins [51]. Nevertheless, some of these uncertainties cancel out when considering ratios of $W + N$ jet events rates. Theoretical expection [49] suggest that the ratio of the number of events between

Multiplicity	No. Events ($W \rightarrow e\nu$)	No. Events ($W \rightarrow \mu\nu$)	Total
$W + 1$ jet	5472	3988	9460
$W + 2$ jets	744	629	1373
$W + 3$ jets	111	87	198
$W + \geq$ 4 jets	26	28	54

Table 3.9: Number of W events as a function of the jet multiplicity in the $W \rightarrow e(\mu) + \nu$ samples, after Z and dilepton removal. Jets are clustered with cone radius of $\Delta R = 0.4$ and are required to have raw $E_T \geq 15$ GeV and $|\eta_D| \leq 2.0$.

adjucent jet multiplicity bins should remains constant:

$$\frac{W + (n+1)\ jets}{W + n\ jets} = \frac{W + (n+2) jets}{W + (n+1) jets} \equiv \alpha$$

This constant factor α should consequently be proportional to the strong coupling constant, α_s. Figure 3.17 shows a simple exponential fit to the electron and muon $W+$ jets data. In both cases the fit is performed in the raw data before any background subtraction and correction or trigger efficiency. The fit returns $\alpha = 0.144 \pm 0.004$ for the electrons and $\alpha = 0.160 \pm 0.005$ for the muon data. Similar fits to the $Z+$ jet data where the production processes are similar to the $W+$ jets, are shown in Figure 3.19 and yield comparable results. Namely, $\alpha = 0.128 \pm 0.01$ and $\alpha = 0.141 \pm 0.015$ for the electron and muon channel respectively. Similar fits were performed to VECBOS [69] simulated $W + N$ jet events generated using two different re-normalization scales. The fit for events generated with normalization scale $\mu^2 = M_W^2 + P_T^{W^2}$ gives $\alpha = 0.147 \pm 0.005$ while for a choice of re-normalization scale equal to the average jet P_T, $\mu = < P_T >^2$ returns a value of $\alpha = 0.165 \pm 0.005$.

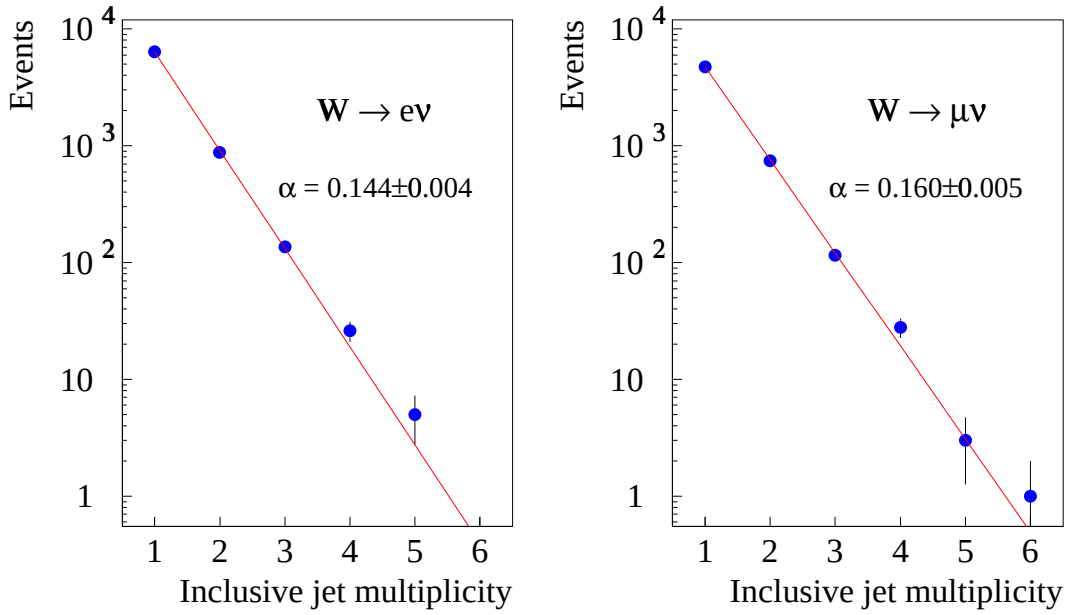

Figure 3.17: Jet multiplicity distributions in the the $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ samples before any background corrections. Events shown are after Z and dilepton removal. Each jet is required to have $E_T \geq$15 GeV and $|\eta_D| \leq 2$ and it is clustered with cone of radius 0.4. The lines represent a simple exponential fit to the data.

The validity of this simple scaling law depends on the minimum jet E_T and minimum jet-jet separation, dr, which is related to the clustering cone size. For large cone sizes, as the number of jets increases going to higher multiplicities, the available phase space to accommodate the extra jet decreases and eventually becomes impossible to add an extra jet. Therefore, the scaling law should break down at higher multiplicities. On the other hand, when the jet cone size becomes very small, this translates to small separation between outgoing partons. As the separation becomes smaller, the collinear singularities, present at leading order calculations, become important and the cross section diverges logarithmically. Of course, this never happens

in the experimental level, because finite size of the calorimeter towers prevents these divergences. The minimum jet E_T similarly controls the infrared diverencies present at leading order calculations when the emitted partons are arbitrary soft or collinear with the incoming partons. However, for reasonable E_T and cone size selection the fixed order perturbative description of a specific inclusive process is appropriate.

The large uncertainties associated with the LO calculations of the $W + N$ jet cross section, especially at high jet multiplicity bins, make difficult to establish any excess of events over the $W + N$ expectation. Nevertheless, since the scaling low seem to hold for reasonable selection criteria, it can be used to make predictions on the amount of $W + N$ jet rates. This method was used by the D0 collaboration [3] to estimate the background contribution of $W + N$ jets in the $t\bar{t}$ signal region.

This jet scaling low is not used anywhere in this analysis to estimate the W+N background contribution to the $t\bar{t}$ signal. Instead, other selection criteria are used to discriminate between $t\bar{t}$ and QCD $W + N$ events.

3.9 Control Samples

For the purpose of understanding the efficiency and background of the heavy flavor tagging algorithms (see chapter 4) used in this analysis, a number of independent datasets are also examined. A brief description of these datasets is given in the following sections.

3.9.1 The Inclusive Jet Data Samples

The study of the tagging rate in events with little heavy flavor (b and c) is essential for determining the probability of obtaining a mistag. Mistags is a major background

to the identification of heavy flavor jets. The inclusive jet samples are dominated by gluon or light-quark jets and therefore are suitable for studying the mistag rate of the b-tagging algorithms. Jets in these samples are called *generic* jets.

The inclusive jet samples are collected based on specific Level 2 triggers. These triggers require a localized energy cluster in the calorimeter with transverse energy above a predefined E_T threshold, termed as *trigger threshold*. Triggers with E_T thresholds set at 20, 50, 70 and 100 GeV are employed. Due to high cross sections, the first three of the above triggers are prescaled. Offline, data samples of each jet trigger threshold are made requiring the events to satisfy the corresponding Level 2 jet trigger. In order to reduce the rate of events arising from detector malfunctions or main ring splashes, the presence of at least one SVX track is also required. The E_T distribution of all the jets in each of the above trigger samples is shown in Figure 3.18. The trigger threshold is apparent in each distribution. Because offline the jet E_T is calculated with respect to the event vertex while at trigger level, the jet E_T is calculated with respect to the detector origin (z=0), an E_T smearing around the trigger threshold is observed.

Besides the above samples which are based on specific jet E_T thresholds, two additional jet samples are also examined. These samples are collected requiring that the scalar total transverse energy, $\sum E_T$, of the event be greater than 175, 300 GeV. One last jet trigger, designed for the search of the full hadronic $t\bar{t}$ decays, requires $\sum E_T \geq 125$ GeV along with four calorimeter clusters of $E_T \geq 15$ GeV is also used for additional cross checks of the results obtained using the other jet triggers.

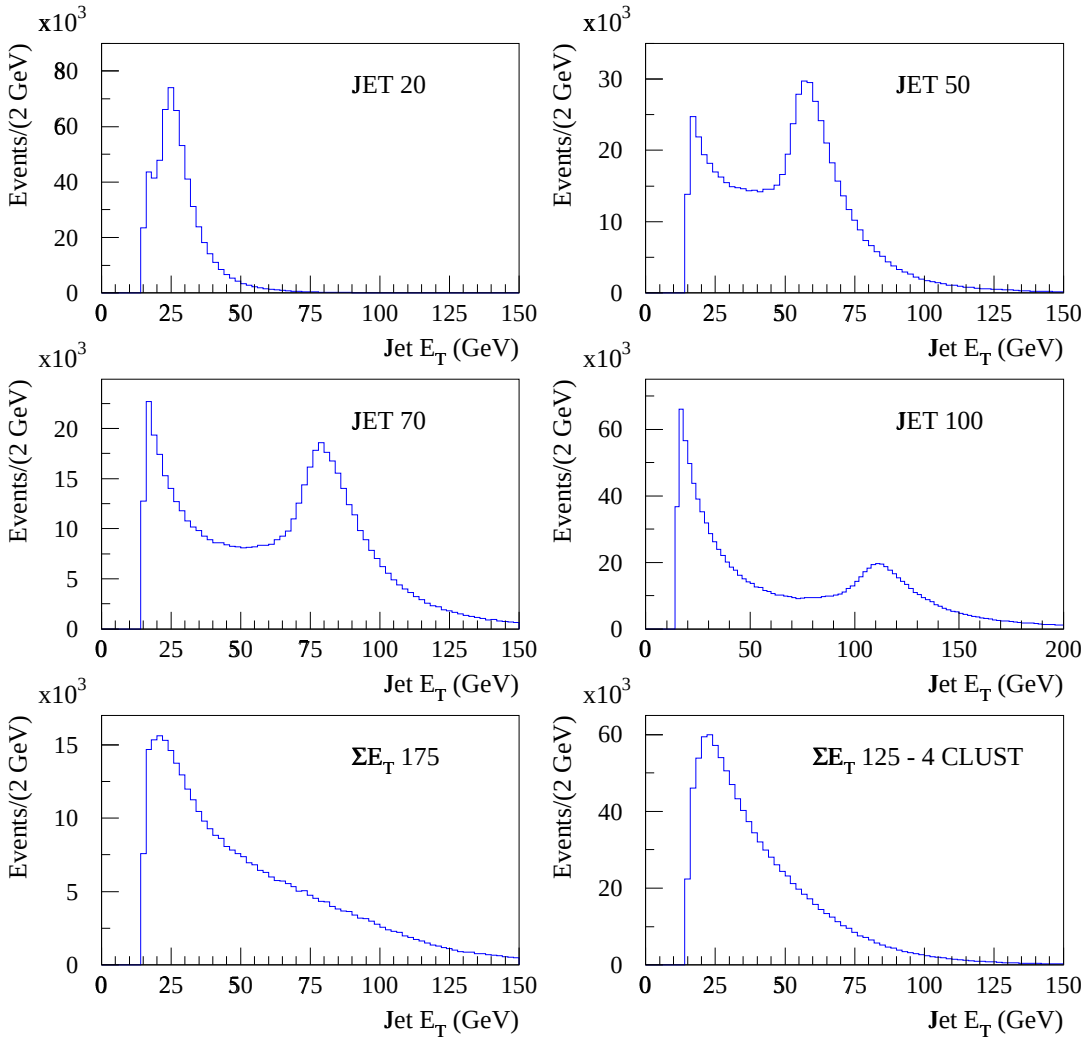

Figure 3.18: E_T distributions of all jets in the various jet samples used in the analysis.

3.9.2 The low-P_T Inclusive Electron Sample

In order to measure the efficiency of the b-tagging algorithms a sample of low-P_T inclusive electrons was used. A fraction of electrons in this sample result from semileptonic decays of b-quarks making the sample suitable for studies of the b-tagging algorithms. Events in this sample are collected requiring a CEM cluster of $E_T \geq 8$ GeV at Level

Electron candidate
Candidate electron satisfies Level 2 trigger
$E_T \geq 9$ GeV at Level 3
Satisfies all primary electron selection requirements in Table 3.1
Cluster isolation, $Iso^{ele}(cone0.4) \geq 0.1$
Electron is associated with a 0.4 cone jet with $E_T \geq 15$ GeV
Kinematic requirements
One 0.4 cone jet other than electron jet with $E_T \geq 15$ GeV and $\|\eta_D\| \leq 2$
Electron jet and away jet back-to-back, $\Delta\phi \geq 2.4$

Table 3.10: Criteria applied in the selection of the $b\bar{b}$ enriched sample using the low-P_T inclusive electron sample.

1 and a CEM cluster with $E_T \geq 9$ GeV matched to a CFT track of $P_T \geq 9.2$ GeV/c at Level 2. In order to enrich the heavy flavor content of the sample, the events are required to contain an additional jet of $E_T \geq 15$ GeV, separated from the electron candidate by $\Delta\phi \geq 2.4$. The b-purity of the electron jets is $\sim 43\%$ in b-jets. A summary of the criteria used to select the b-enriched low-P_T electron sample is shown in Table 3.10.

3.9.3 The Photon Sample

A sample of isolated photons in association with at least one jet provides an alternative sample in which the performance of the tagging algorithms can be tested. The heavy flavor content of the sample is expected to be enriched in c-jets due to $\gamma\ c$ production making the sample useful for testing the tagging algorithms on c-jets. Production of $\gamma\ c$ proceeds either via the Compton process, $g\ c \rightarrow \gamma\ c$ which is the dominant one, or via $q\ \overline{q}$ annihilation followed by gluon splitting, namely $q\ \overline{q} \rightarrow \gamma g \rightarrow \gamma c \overline{c}$. Events from run 1B were selected based on a Level 2 photon trigger

Photon candidates
Transverse energy of γ candidate, $E_T(\gamma)\ \geq\ 23$ GeV
$\lvert\eta_{CEM}\rvert\ \leq\ 1.0$ No 3-D track pointing to the CEM cluster
$E_{had}/E_{em}\ \leq\ 0.055\ +\ 0.00045 \cdot E_\gamma$
E_T in a cone 0.7 around the cluster, $E_T\ \leq\ 2$ GeV
CES cluster profile consistent with that of a γ, CES $\chi^2\ \leq 20$
Event selection
Primary vertex within $\lvert Z_{vrtx}^{prim}\rvert\ \leq\ 60$ cm
One only good photon candidate in the event
At least one SVX track in the event

Table 3.11: Selection criteria used in the making of the photon sample.

which required a CEM cluster with $E_T\ \geq\ 23$ GeV. After full offline reconstruction, photons were selected requiring no 3-dimensional track extrapolating to the CEM cluster. Additional criteria were applied to ensure background rejection from π^0 and η meson decays. Table 3.11 summarizes the criteria applied for the selection of the

photon sample. After requiring the presence of an additional jet with $E_T \geq 15$ GeV and $|\eta_D| \leq 2$, the final sample consists of 3000 $\gamma+ \geq 1$ jet events. The expected background contamination of the sample due to π^0 and η decays is estimated to be $\sim 45 \pm 4.5\%$ [45].

3.9.4 The Z + N jet Sample

The production mechanisms of W and Z in association with jets are very similar and so is the heavy flavor content in the samples. Furthermore, the Z + *jets* sample has very small contamination from $t\bar{t}$ production which dominates the high jet multiplicity bins in the W+jets sample. As described in section 6.10, the Z+jets sample is used as a control sample of the method used to estimate the heavy flavor background to $t\bar{t}$ events.

Multiplicity	No. Events ($Z \rightarrow e^+e^-$)	No. Events ($Z \rightarrow \mu^+\mu^-$)	Total
Z + 1 jet	791	357	1148
Z + 2 jets	107	52	159
Z + 3 jets	9	7	16
Z + $\geq$ 4 jets	3	1	4

Table 3.12: Number of Z events as a function of the jet multiplicity in the $Z \rightarrow e^-(\mu^-)+e^+\mu^+$ samples. Jets are clustered with cone radius of $\Delta R = 0.4$ and required to have raw $E_T \geq 15$ GeV and $|\eta_D| \leq 2.0$.

The Z sample consists of the events removed from the high-P_T primary lepton sample when the Z removal algorithm was applied (see section 3.6.1). The number

of Z events in the mass range $70 \leq M_{\ell\ell} \leq 110$ GeV/c^2 as a function of the jet multiplicity is shown in Table 3.12 and plotted in Figure 3.19 along with a fit to the multiplicity distributions. The fit describes better the $Z + N$ jet data compared to the $W + N$ jet distributions (see Figure 3.17) where the large contribution of $t\bar{t}$ events is expected in the high jet multiplicity bins.

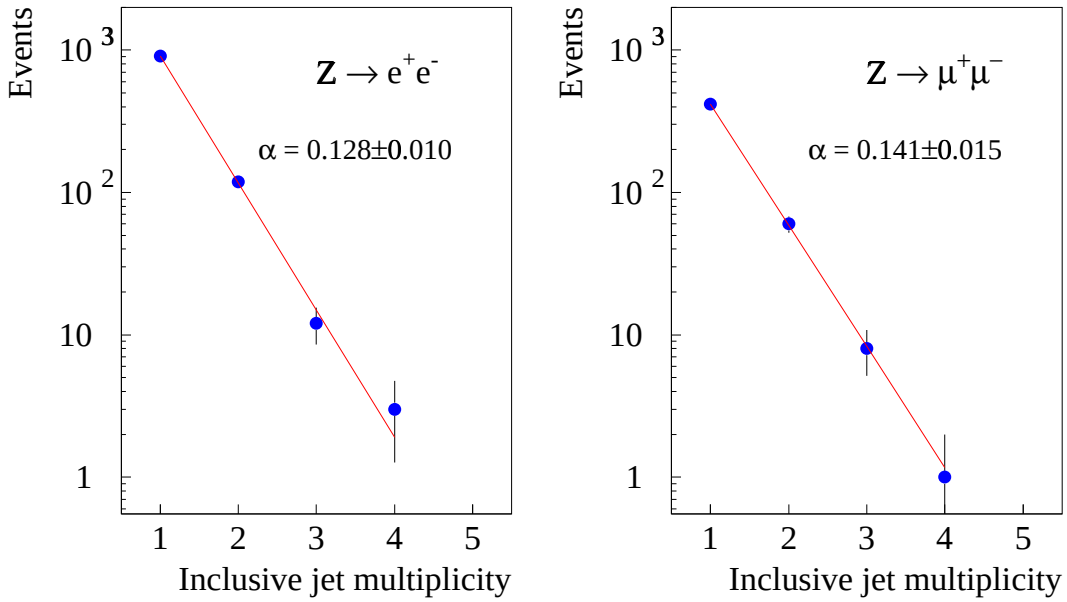

Figure 3.19: Jet multiplicity distributions in the the $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ samples before any background corrections. Each jet is required to have $E_T \geq$15 GeV and $|\eta_D| \leq 2$ and it is clustered with cone of radius 0.4. The lines represent a simple exponential fit of the form $N_j = N_{Z+1\ jet} \times \alpha^{j-1}$ as explained in the text.

3.10 Monte Carlo Samples

Several Monte Carlo generators were used to estimate the acceptance, tagging efficiencies, and background contributions. The generators used in this analysis are HERWIG [68], PYTHIA [70] and VECBOS [69]. The acceptance for $t\bar{t}$, and b-tagging efficiency in $t\bar{t}$ events and some of the backgrounds were estimated using the PYTHIA Monte Carlo versions 5.6 and 5.7. The HERWIG version 5.6 Monte Carlo generator was used for the calculation of some of the backgrounds, for estimating systematic uncertainties in the acceptances calculated using PYTHIA. The HERWIG generator is also used to compare the tagging rates of heavy flavor jets in simulation and inclusive jet data and the obtained normalization factors are then used to derive the heavy flavor content in $W+ \geq 1$ jet events.

Both HERWIG and PYTHIA generators use tree-level matrix element calculations for the hard scattering, convoluted with parametrizations of the parton distribution functions. The outgoing initial and final state partons are then converted into a cascade of gluons and $q\overline{q}$ pairs with energy and angular distributions according to Altarelli-Parisi splitting equations [6]. The strength of these generators is in the way of modelling the parton shower by taking into account the color correlations, *color coherence*, between the initial and final state partons. The parton shower terminates when the invariant mass of the parton falls below the threshold that perturbative QCD breaks down. At this Level the remaining partons are turned into colorless hadrons according to phenomenological models and the process is called *hadronization* or *fragmentation*. HERWIG employs the *cluster* model while PYTHIA employs the *string* model for the fragmentation of light quarks. For heavy quarks (t, b and c-quarks) the fragmentation is modeled by the Peterson fragmentation functions [74]

with $\epsilon_p = 0.006$ for b-quarks and $\epsilon_p = 0.05$ for c-quarks. The generators include a model for the *underlying event* which describes the hadronization products of the beam remnants. The underlying event modelling is tuned to low Q^2 $p\bar{p}$ collision data (soft collisions). The output of the generators includes the four-vectors of all the particles which participate in the parton shower and also the appropriate indices for connecting each particle to its ancestor. The information is kept in a specific data bank called *GENP*. The detector simulation, *QFL* [77], uses the GENP bank and parametrization of detector responses based on testbeam measurements or tuned to the data, to creat the fully simulated data banks for each event. This allows the processing and selection of simulated events to be performed in exactly the same way as in the data.

In order to model correctly the decay of heavy flavor hadrons the CLEO Monte Carlo [76] simulation package, *QQ*, was used. The particle list from the main generators was fed to the QQ simulation package which removes the decay products of heavy flavor mesons from the list and then it re-decays the initiator heavy flavor meson.

The VECBOS Monte Carlo was used to study the kinematics of the W + jets background and the tagging efficiency in $Wb\bar{b}$ and $Wc\bar{c}$ events. VECBOS is a parton-level Monte Carlo program based on the exact computation of the leading order matrix elements of W + n jets (n= 0, 1, 2, 3, 4 jets) production. The output created by VECBOS is an event list containing the weight of the event and the four-momentum vectors of the W decay products and the associated partons. Events from VECBOS are interfaced via a package called *HERPRT* with the HERWIG Monte Carlo. Then, HERWIG provides the evolution of the VECBOS partons, adds initial and final state gluon radiation and provides an underlying event. VECBOS with the HERPRT interface is shown to reproduce the kinematic distributions observed in W+ jets

events in the data [48].

The MRSD0 parton parametrization function was used to generate all the Monte Carlo event samples used in this analysis, unless otherwise stated. This structure function is shown to reproduce the results of W asymmetry at CDF [75].

Chapter 4

Methods of identifying heavy flavor decays

The presence of jets originating from b quark hadron decays is one of the characteristic signature of $t\bar{t}$ events. As described in Chapter 1, two b-quark jets are expected in the final state of a $t\bar{t}$ decay. However, in practice, one or sometimes both of the b-quark jets may fall outside the detector's acceptance or may have too low E_T to be counted. The b-quark content of the final state can also be enhanced due to b's from gluon splitting. In this case a gluon produced from initial or final state radiation can split to a heavy flavor pair with small probability ($\sim$1%).

The average fraction of b-quark jets as a function of the jet multiplicity in PYTHIA simulated $t\bar{t}$ decays are shown in Figure 4.1. In the $W+\geq$3 jet region, the final state of a $t\bar{t}$ decay is fully reconstructed and the average number of b-quark jets (1.8) reaches the expected value.

In contrast, jets from standard QCD $W+n$ jet production originate from light

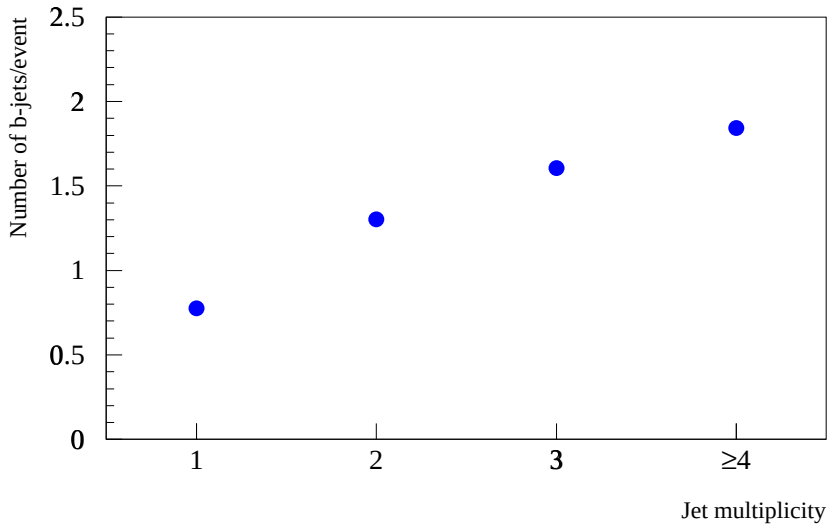

Figure 4.1: The number of b-jets expected in $t\bar{t}$ PYTHIA Monte Carlo events (M_{top}=170 GeV/c^2) as a function of the jet multiplicity. The events are required to pass the W selection criteria as described in the previous chapter.

quark or gluons jets. Therefore any heavy flavor content in this type of events arise from gluon splitting. Using the VECBOS Monte Carlo, it is estimated that the fraction of W $\geq$3 jet events with two b-jets at the final state is $\sim$2%. Table 4.1 shows the fraction of QCD $W+n$ jet with two b-jets at generation level as a function of the jet multiplicity as determined using the VECBOS Monte Carlo. Therefore, a large

Multiplicity	Fraction (%)
$W+1$ jet	1.06$\pm$0.02
$W+2$ jets	1.9$\pm$0.01
$W+3$ jets	2.4$\pm$0.3

Table 4.1: Fraction of $W+$ jet events with two b-jets, as a function of jet multiplicity at parton level as determined using the VECBOS Monte Carlo. Errors are statistical only.

number of QCD $W+\geq$3 jets events contributing background to the $t\bar{t}$ signal can be

eliminated requiring at least one of the jets in $W+ \geq 3$ jet events to be identified as a b-quark candidate.

The method of identifying a b-quark jet is known as *heavy flavor tagging*. The techniques developed to accomplish such a task are the subjects of this chapter.

Heavy flavor identification can be achieved by exploiting some characteristic properties of heavy flavor hadron decays in combination with the excellent position resolution of the silicon vertex detector ($\sigma_d = 16 \cdot (1 + 0.8/P_T^2)\mu$m). The b and c-hadrons have relatively long lifetime, $\tau \sim 1.5$ and $0.4 \sim 1$ ps respectively. Moreover, in $t\bar{t}$ events, b and c-hadrons receive large transverse boosts ($< P_T > \sim 44$ GeV/c) and travel a significant distance in the transverse plane before they decay. The mean distance travelled by b-hadrons in $t\bar{t}$ events is of the order of 3.5 mm. Figures 4.2(a) and (b) show the P_T and η distribution at generation level of b-hadrons in PYTHIA $t\bar{t}$ Monte Carlo events with M_{top}=170 GeV/c^2. The distance travelled by the B-hadrons in the transverse plane, referred to as *transverse decay length or* L_{xy}, is shown in Figure 4.2(c). The b-hadron decay vertex, known as *secondary vertex*, is significantly displaced with respect to the $p\bar{p}$ interaction vertex and consequently the tracks emerging from this secondary vertex, appear with large impact parameter with respect to the primary vertex.

The ability to identify such displaced tracks and vertices depends on the resolution for determining both the trajectory of each track and the position of the primary interaction vertex, from which most tracks emanate. The SVX detector offers a very fine position resolution and is used to resolve these secondary vertices and identify displaced tracks.

Two tagging techniques based on tracking information from the SVX detector were developed for heavy flavor hadron identification. One of the techniques uses

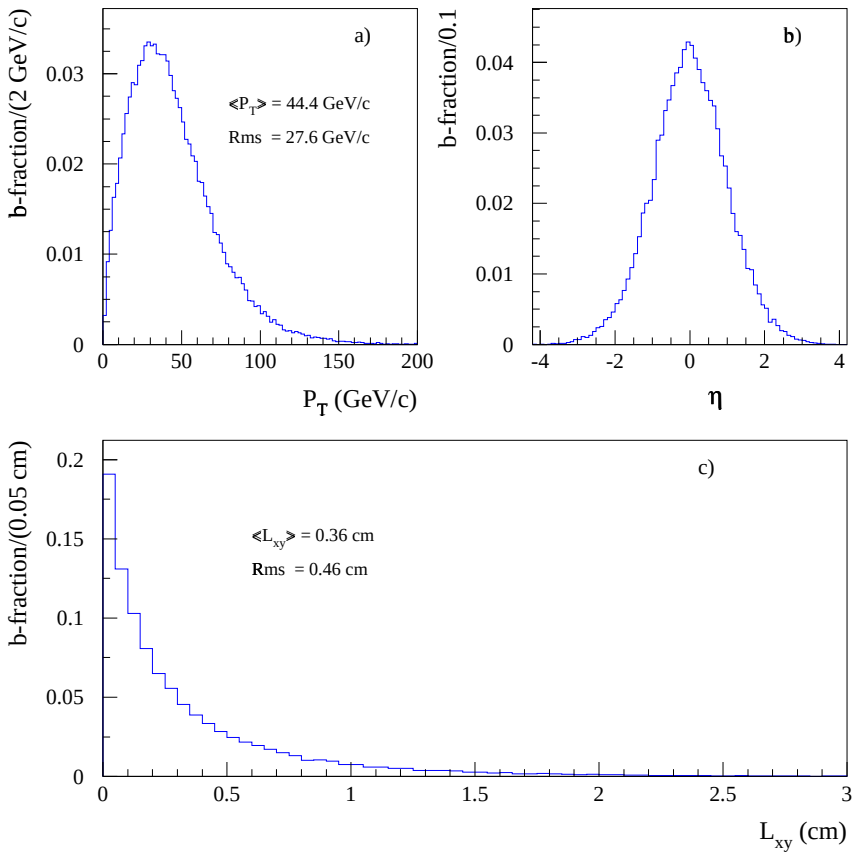

Figure 4.2: The P_T (a), pseudo-rapidity (b) and transverse decay length (c) distributions of b-hadrons before any detector resolution effects as expected in PYTHIA $t\bar{t}$ simulated events (M_{top}=170 GeV/c^2).

at least two well displaced tracks contained inside a jet in order to reconstruct a secondary vertex. This tagging algorithm is known as *secondary vertex tagging or SECVTX*. If a secondary vertex is reconstructed inside a jet and it is significantly displaced with respect to the $p\bar{p}$ interaction vertex, the jet containing this vertex is declared a *heavy flavor tag*. However, secondary vertices can also be reconstructed inside non-heavy flavor jets. These secondary vertices are either physical resulting from K or Λ decays or unphysical resulting from detector resolution effects, tracking inefficiencies or track pattern recognition failures. Tagged jets resulting from detector effects are known as *mistags*. The mistag rate determines the identification power of a tagging algorithm. In the development of a tagging algorithm special attention is paid to achieve high efficiency for identification of b-jets while maintaining the level of mistags to a few percent level.

A second b-tagging technique known as *jetprobability tagging* was also developed. This tagging algorithm uses the impact parameter of individual tracks contained in a jet to determine the probability the track originates from the primary interaction vertex. Comparing the track displacement to the SVX detector resolution, the probability that the track displacement is consistent with resolution effects alone can be determined. Combination of the probabilities of all tracks in a jet leads to an overall probability for the jet. For heavy flavor jets, this probability is expected to be small. A jet is considered tagged if the resulting jet probability is smaller than a predefined value. For this analysis, the maximum allowed jetprobability was chosen to be $jet\text{-}probability \leq 0.05$. This value is relatively loose comparing to the SECVTX algorithm allowing for more background due to mistags and more events with c-jets. As discussed in the following Sections and in Chapter 6, the efficiency to identify c-jets with jetprobability is twice the efficiency of c-tagging efficiency of SECVTX. The looser re-

quirement applied on jetprobability was used to enhance the sensitive to background events allowing for better studies of the various background sources. These studies can further used to recalibrate and re-examine features of the SECVTX algorithm which may have been neglected or hidden due to the small background contribution.

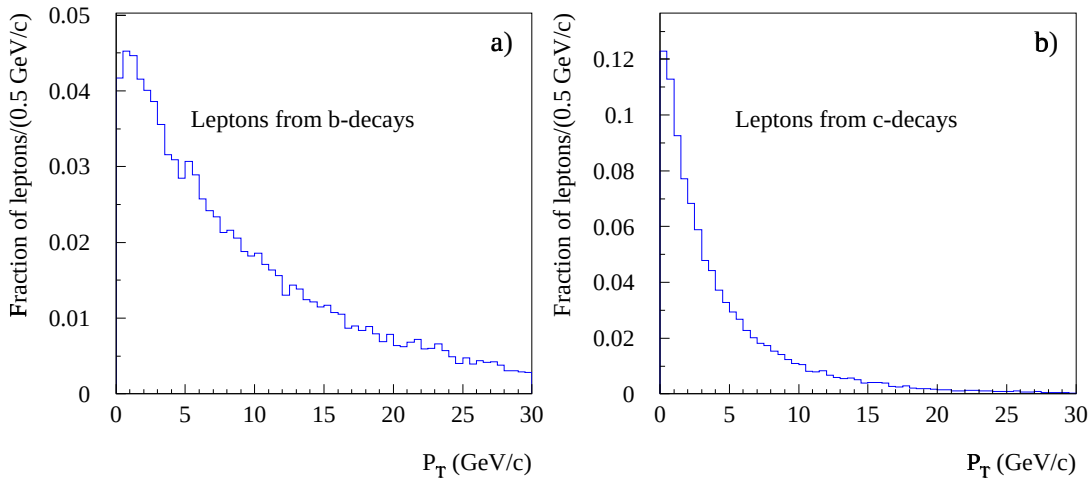

Figure 4.3: The generator level P_T distributions of leptons from b (a) and c-hadron decays (b) in $t\bar{t}$ Monte Carlo events (M_{top}=170GeV/c^2) generated with the PYTHIA Monte Carlo generator.

Some fraction of heavy flavor hadrons decay semileptonically. This property of heavy flavor quark decays is used by a third tagging technique called *soft lepton tagging* or SLT. The produced leptons (electrons or muons) have relatively low momentum compared to leptons produced in W or Z leptonic decays. Figure 4.3 shows the track momentum spectra of leptons from b and c-hadron semileptonic decays in $t\bar{t}$ Monte Carlo events with M_{top}=170 GeV/c^2. These *soft leptons* most often are in the proximity of the jet formed by the remaining hadron decay products and as a

consequence cannot appear in isolation.

The soft P_T spectrum and the poor isolation make the identification of soft leptons rather challenging. The task is even more difficult considering that hadronic showers of generic QCD jets can mimic the electron or muon signature. The soft lepton algorithm developed in CDF, maintains high efficiency for soft leptons from b and c decays while the rate of fake leptons is kept to less than 1% per track.

The description, features and performance of the tagging techniques are discussed in more details in the following sections.

4.1 B-tagging with the SVX

The SVX based B-tagging algorithms use the impact parameter of tracks to determine the presence of a heavy flavor hadron decay in the event. The identification of displaced tracks depends on the resolution for determining the trajectory of each track and the location of the primary interaction vertex with respect to which the track displacement is measured.

The method used to determine the primary vertex position and the concept of the track impact parameter and its relation to the primary and heavy flavor decay vertices are the subject of the next two sections.

4.1.1 Primary vertex finder

Precision measurement of the primary interaction point results in a correct frame of reference for the event topology. As discussed in Section 3.1.3, the vertex reconstruction is performed using tracks with segments in the VTX and CTC detectors. However, the SVX detector provides better position resolution in the r-ϕ plane and

therefore a more accurate location of the interaction vertex can be obtained using only SVX tracks. As a reminder, SVX tracks are the tracks reconstructed in the CTC and VTX detectors with track segments in the SVX detector (see Section 3.1.2).

Recall that the $p\bar{p}$ interaction point has a bivariate Gaussian distribution parallel and perpendicular to the beam axis, with $\sigma_{||} = 30$cm and $\sigma_{\perp} = 35\mu$m. The beam axis and detector axis are not exactly parallel, but they have relative slopes of $\sim$ 4.5 μm/cm in the horizontal plane and $\sim$-3 μm/cm in the vertical plane. At the nominal interaction point (z=0), the displacement of the detector axis with respect to the beam axis varies between 200 and 1200 μm in the horizontal plane and 400 and -1000 μm in the vertical plane. Due to the change of the conditions of the tevatron during data taking, the slopes and displacements drift. However, both relative slope and displacement are measured on an run-by-run basis with an accuracy of 0.4 μm/cm and $\sim$ 10 μm for the offset and displacement respectively.

The primary vertex finding algorithm performs on an event-by-event basis a weighted three-dimensional fit of all the SVX tracks after the application of all appropriate corrections for the detector offset and slope. As a starting point, the position of the best VTX vertex is used. An iterative procedure which minimizes the residuals of the P_T weighted tracks with respect to the fit point is performed, assuming that most tracks in the event come from the primary vertex. Tracks with large impact parameters are removed in each iteration and the fitting procedure is repeated. Clearly, the accuracy of the algorithm can be affected by the number of SVX tracks used in the fit and the global event topology. The uncertainty in the fitted primary interaction vertex coordinates perpendicular to the beam direction is estimated in the range of 6$\sim$26 μm depending on the topology of the event and the number of SVX tracks.

Due to the high luminosity conditions during the run, a large number of events

($\sim 70\%$) are accompanied by additional interactions resulting in vertices occurring at different locations along the beamline. In order to diminish the ambiguity as to which VTX vertex should be used as seed vertex at the beginning of the iteration procedure, the vertex with the highest total transverse momentum of associated tracks is chosen. Tracks are assumed to be associated with a vertex if they extrapolate within 5 cm of this vertex along the z-axis.

4.1.2 Track impact parameter

Tracks displaced with respect to the primary interaction vertex are used as input to the SVX based tagging algorithms. The track displacement is described in terms of the impact parameter D_0, which is the distance of closest approach of the track to the primary vertex in the r-ϕ plane, as shown in Figure 4.4. The sign of the

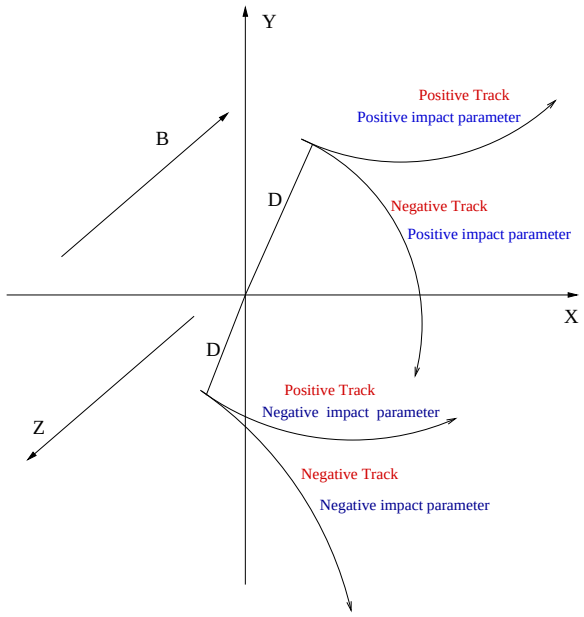

Figure 4.4: A schematic view in the r-ϕ plane showing tracks of charged particles with positive/negative charge and the convention used at CDF for the sign of the impact parameter. The positive z-axis is defined as the direction of the protons. The magnetic field, B, is pointing to the negative z-direction.

impact parameter is determined by the location of the beam in the transverse plane, relative to the circle, projection of the track helix in the transverse plane. If the beam lies inside the circle, the impact parameter is signed negative (positive) for positively (negatively) charged tracks. If the beam lies outside, the sign of the impact parameter is taken as positive (negative) for positively (negatively) charged tracks.

Parameters that contribute to the uncertainty of the impact parameter are the intrinsic SVX resolution ($\sim 13\mu m$), the error in the position of the primary vertex ($\sim 10\mu m$) and the multiple scattering, function of the track P_T and the amount of the transversed material ($60/P_T$ in μm). The impact parameter resolution as a function of the track P_T is shown in Figure 2.6.

For heavy flavor tagging, it is important to quantify the magnitude of the measured track displacement with respect to its associated uncertainty. A variable called *impact parameter significance* is defined as $S_{D_0} = \frac{D_0}{\sigma_{D_0}}$ with the purpose of normalizing the measured track displacement to the associated uncertainty. Figure 4.5 shows the distribution of the impact parameter significance for tracks in generic jets from a sample of events collected with the JET_50 trigger. Also shown, the same distribution for tracks associated with heavy flavor jets taken from a JET_50 HERWIG [68] Monte Carlo sample after full detector simulation. The core of the data distribution is well represented by a Gaussian with $\sigma = 0.98$. The tails of the distribution are due to a combination of non-Gaussian tracking effects and true long-lived particles. The corresponding Monte Carlo distribution is a combination of zero-lifetime tracks originating from the fragmentation of the heavy flavor-quark and of tracks from the b-hadron decay. As a result of the long b-lifetime the Monte Carlo distribution is much broader.

For heavy flavor tagging it is of interest to assign the sign of the impact parameter

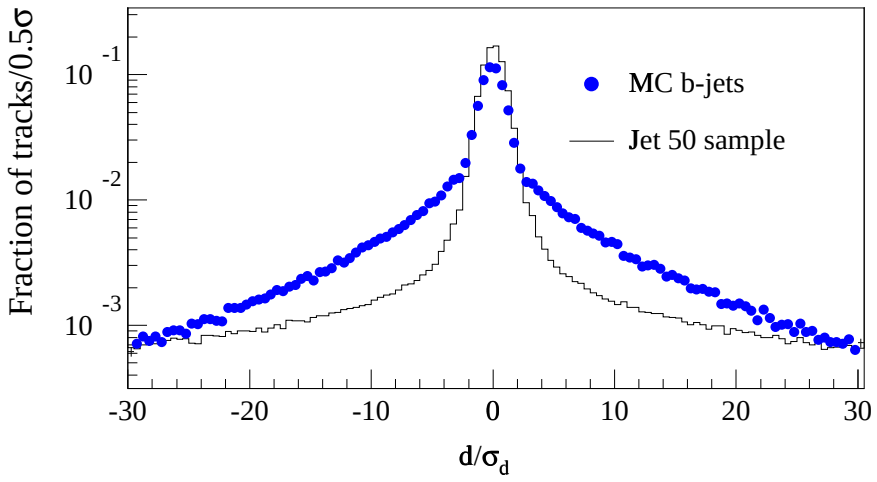

Figure 4.5: Distribution of impact parameter significance, D_0/σ_{D_0}, for tracks in a sample of events collected with the JET_50 trigger (solid) and in b-jets from a JET_50 HERWIG [68] Monte Carlo simulation (points).

relative to the direction of the jet originated from the heavy flavor hadron decay. This convention, termed as *signed impact parameter*, is used to indicate whether the track originates from a displaced vertex in the direction of motion of the examined jet or from an unphysical location. Accordingly, the impact parameter normalized with respect to its uncertainty is termed as *signed impact parameter significance*, S_{D_0}.

The sign of the impact parameter relative to the jet direction is determined according to the following convention:

$$\text{sign of impact parameter} = \begin{cases} \text{positive} & \text{if } \phi_{D_0} < 90^\circ \text{ (i.e. } \phi_1 \text{ in Figure 4.6)} \\ \text{negative} & \text{if } \phi_{D_0} > 90^\circ \text{ (i.e. } \phi_2 \text{ in Figure 4.6)} \\ \text{zero} & \text{otherwise} \end{cases}$$

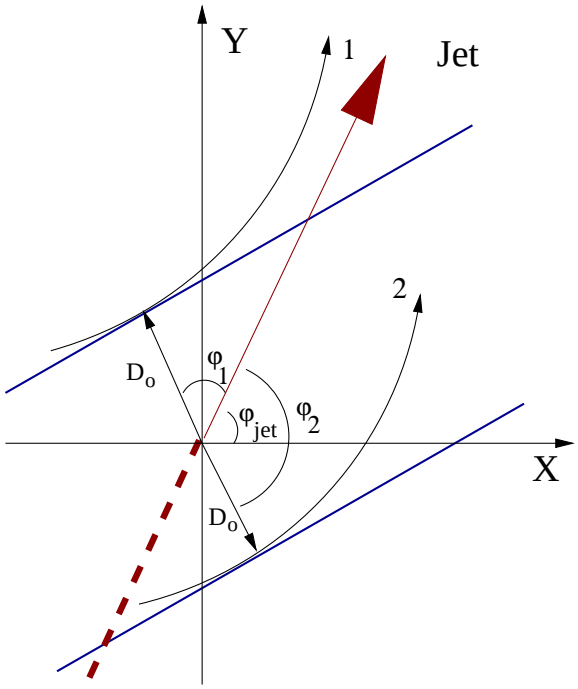

Figure 4.6: The definition of the sign of the impact parameter with respect to the jet axis direction in the r-ϕ plane. The impact parameter is positive or negative if the tangent of the track at the point of closest approach intercepts the jet axis above or below the primary vertex.

Practically it means that if the tangent of a track at the point of closest approach intercepts the jet axis above the primary vertex (in the jet direction) the impact parameter relative to the jet is positive, while if the interception point is below the primary vertex the sign of the impact parameter is negative.

Tracks from long-lived particles are expected to populate mostly the positive side of the S_{D_0} distribution. The negative side of the distribution results from resolution effects and mismeasurement of the jet direction and therefore is populated mostly by prompt tracks. These tracks are expected to be randomly distributed with positive and negative impact parameter significance and in the absence of any heavy decays the impact parameter significance is expected to be symmetric around zero. However, the negative side of the distribution contains also tracks from heavy flavor decays. This behaviour is demonstrated in Figure 4.7 where the distributions of the S_{D_0} for

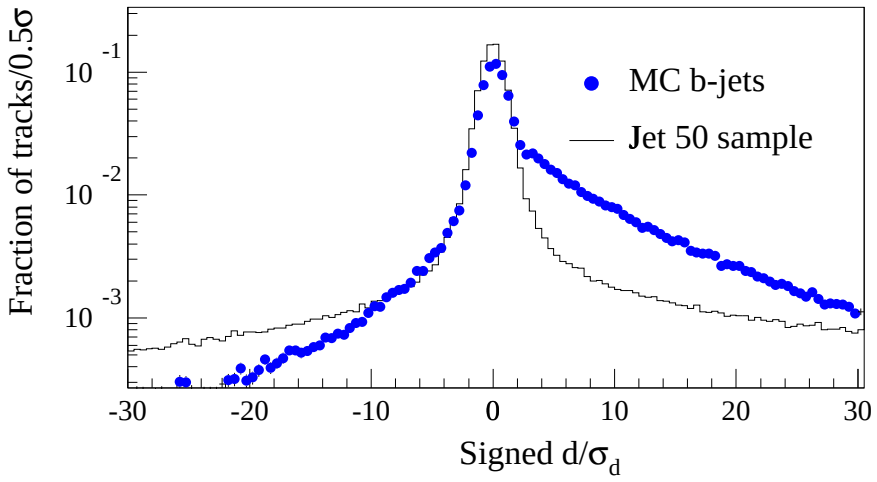

Figure 4.7: Distribution of the signed impact parameter significance for tracks in a sample of events collected with the JET_50 trigger (solid) and in b-hadron jets points in JET_50 HERWIG Monte Carlo simulations (points).

tracks in generic jets from the JET_50 sample and for tracks from b-jets in a JET_50 HERWIG simulation are plotted. The assymetric shape of the data distribution in the positive side indicates the presence of tracks from heavy flavor decays or decays in flight of π's and K's.

An estimate of the amount of heavy flavor tracks in the JET_50 data can be obtained from the distributions shown in Figure 4.7. The positive side of the data distribution can be fitted to a combination of the impact parameter significance distributions of heavy flavor tracks and prompt tracks. The shape of the impact parameter significance in the Monte Carlo is used to model the heavy flavor tracks distributions. The symmetric around zero of the negative side of the data distribution is used to model the distribution of prompt tracks. Performing the fit, it is found that $\sim$10% of the tracks with positive S_{D_0} in the generic jet sample originates from heavy flavor decays. This fraction increases to $\sim$25% when requiring the signed $\frac{D_0}{\sigma_{D_0}} \geq 2.5$

4.2 SECVTX - A Secondary Vertex tagging algorithm

The *SECVTX* algorithm was developed for the identification of heavy flavor hadron decay vertices inside *taggable* jets.

A jet is defined to be SECVTX taggable if it has:

- $E_T \geq 15$ GeV in a cone of radius 0.4.
- Pseudo-rapidity, $\mid \eta_{Det} \mid \leq 2.0$ calculated with respect to z=0.
- at least two SVX tracks passing the "Pass 1" quality requirements shown in Table 4.2, except the impact parameter significance requirement.

4.2.1 Track selection

For the vertex reconstruction, the algorithm uses SVX tracks inside a cone of radius 0.4 around the jet axis. The jet-track association requirement retains 90% of the tracks from b-hadron decays in $t\bar{t}$ events, while it rejects 65% of the tracks arising from the fragmentation of a b-quark [53]. The SVX tracks in a jet are classified into two categories according to the criteria shown in Table 4.2 and explained below:

- Loose or "Pass 1" selection criteria for displaced tracks
  - ▷ In order to reduce the uncertainties in the reconstruction of the secondary vertices, all tracks are required to be 3-D and satisfy some minimum CTC quality criteria. For each track, the CTC track segment is required to contain hits in at least 2 axial superlayers with at least 4 hits per superlayer

Pass 1 selection criteria for tracks
3-D passing the following CTC quality criteria
$\geq$ 2 axial superlayers with $\geq$ 4 hits/layer
$\geq$ 2 stereo superlayers with $\geq$ 2 hits/layer
χ^2_r $<$ 6 (Chi-square per degree of freedom)
$\lvert Z_{primary\ vertex} - Z_{track} \rvert < 5$cm
$\lvert D_0 \rvert < 0.15$cm
Hit pattern requirements
• Tracks with $\geq$3 SVX hits
$P_T \geq 0.5$ GeV/c
At least one good SVX hit defined as follows
Associated to only one track
There are no bad strips in the cluster
Length $\leq$ 3 strips
• Tracks with 2 SVX hits
Two good SVX hits
Both hits in the two innermost or outermost SVX layers
$P_T \geq 1.5$ GeV/c
Not part of K_s or Λ decay vertex
$\frac{D_0}{\sigma_{D_0}}$ $>$ 2.5
Pass 2 selection criteria for tracks
Tracks satisfy the Pass 1 selection criteria
At least one good hit for tracks with 4 SVX hits
At least two good hits for tracks with 3 SVX hits
Tracks with 2 SVX hits are rejected
$P_T \geq 1.0$ GeV/c
$\frac{D_0}{\sigma_{D_0}} > 3.0$

Table 4.2: The selection criteria applied to displaced SVX tracks in order to be considered in the SECVTX b-tagging algorithm.

and hits in at least 2 stereo superlayers with at least 2 hits per superlayer. These requirements ensures that all the track parameters are well measured.

- ▷ Track χ^2 per degree of freedom is required to be ≤ 6
- ▷ To ensure that all tracks come from the same primary vertex they are required to extrapolate within 5 cm in z from the primary vertex location, $|Z_{primary\ vertex} - Z_{track}| \leq 5$ cm.
- ▷ Tracks consistent with photon conversion or decays of K_s mesons ($K_s \rightarrow \pi^+\pi^-$) or Λ baryons ($\Lambda \rightarrow p\pi^-$), originating at the primary vertex, are rejected by requiring a maximum impact parameter value of $|D_0| \leq 0.15$ cm. This requirement rejects 3.1% of the tracks before tagging while it is 99% efficient for tracks from b-decays [59]. In addition, track pairs with opposite charge which form a vertex with mass consistent with the K_s or Λ mass are also removed. These requirements are imposed because K_s and Λ decays produce also secondary vertices with decay lengths of $c\tau(K_s) = 2.7$cm and $c\tau(\Lambda) = 7.9$cm [81].
- ▷ For tracks with hits in 3 or more layers of the SVX detector, one hit is required to be *good*. An SVX hit is defined to be good if it does not include any noisy or dead strips, it is ≤ 3 strips long and there is no contribution from neighboring tracks. Requiring the presence of a good hit, it reduces track reconstruction ambiguities arising when several possible tracks can pass through overlapping hits. In addition, the tracks are required to have $P_T \geq 0.5$ GeV/c.

- ▷ For tracks with hits in only 2 layers of the SVX, stringer requirements are applied. The hits are required to be either in the two innermost or the two outermost SVX layers and both of them are required to be good. For this category of tracks the minimum P_T requirement is raised to $P_T \geq 1.5$ GeV/c.
- ▷ All tracks passing the previous selection criteria are required to have impact parameter significance, $\frac{D_0}{\sigma_{D_0}} \geq 2.5$ with respect to the primary vertex.

- Tight or "Pass 2" selection criteria for displaced tracks
  - ▷ All tracks are required to satisfy the loose quality criteria
  - ▷ Only tracks with hits in 3 or more layers are accepted. Tracks with 4 hits are required to contain at least one good hit while tracks with 3 hits are required to have at least 2 good hits.
  - ▷ Minimum track $P_T \geq 1.0$ GeV/c.
  - ▷ Track impact parameter significance, $D_0/\sigma_{D_0} \geq 3.0$.

4.2.2 Vertex reconstruction

The SECVTX algorithm employs a two-step sequence. The first step, called *"PASS 1"* requires at least three tracks passing the loose selection criteria described above to form the secondary vertex. The second step, called *"PASS 2"*, uses at least 2 tracks passing the tight selection criteria, in order to form the secondary vertex. The "Pass 2" process is performed only if there is no secondary vertex candidate found by the "Pass 1" process.

In detail, during the PASS 1 stage, the algorithm performs the following procedure:

- A list of all jets in the event with $E_T \geq 15$ GeV along with their associated SVX tracks is formed. Tracks are counted in a jet if they lay within $\Delta R \leq 0.4$ from the jet axis. They are required to pass the loose selection criteria presented above.

- The tracks are ranked according to their P_T, impact parameter significance, and number of good SVX hits associated with the track.

- The best two tracks are constrained to come from a single vertex, called *seed vertex*. One of the tracks forming the seed vertex is required to have $P_T \geq 2$ GeV/c.

- The remaining tracks in the list are checked for association to the seed vertex. The search stops and a secondary vertex candidate is declared if at least one track is found to have impact parameter significance, $\frac{D}{\sigma_D} \leq 3.0$, with respect to the seed vertex.

- If no secondary vertex is found then the next best pair of tracks is used to define a seed vertex and the procedure is repeated with the remaining tracks.

- If the list of seed vertices is exhausted and there is no secondary vertex found the algorithm proceeds with Pass 2 stage.

The Pass 2 stage of the algorithm proceeds as follows:

- At least two tracks are required to satisfy the tight selection criteria for the algorithm to perform the process.

- A common vertex is made using all tracks passing the quality requirements.

- Tracks which contribute $\chi^2 > 50$ to the reconstruction of the vertex are dropped and a new fit is performed with the remaining tracks. The procedure is repeated until the list of tracks contributing to the fit does not change.

- If the reconstructed vertex has at least two displaced tracks one of which has $P_T \geq 2$ GeV/c^2, it is declared a candidate secondary vertex.

A flow chart of the SECVTX tag algorithm is shown in Figure 4.8. The major sources of the tagging algorithm inefficiencies result from the kinematic requirements on the tracks (P_T and impact parameter significance) and the track quality requirements. For the latter, the requirement for an unshared SVX hit contributes the most to the algorithm inefficiency. A detailed description of the optimization of the selection requirements used in the algorithm can be found in References [53], [54], and [55].

The displacement of a secondary vertex relative to the primary can be measured only in the transverse to the beam plane since the SVX detector provides information only in the $r - \phi$ plane. The displacement is measured in terms of the transverse decay length, L_{xy}, defined as

$$L_{xy} = \sqrt{(x_p - x_s)^2 + (y_p - y_s)^2}$$

where (x_p, y_p) and (x_s, y_s) are the coordinates of the primary and secondary vertex with respect to the nominal interaction point (x=y=z=0). The sign of the measured displacement is defined relative to the jet direction. It is determined by the dot product of the vector, $\vec{L}$, pointing along the L_{xy} from the primary to the secondary vertex and the unit vector, $\vec{u}$, which points along the jet axis.

A jet is *tagged* if it contains a secondary vertex with transverse decay length less than 2.5 cm and significance, defined as $\frac{L_{xy}}{\sigma_{L_{xy}}}$, greater than 3.0. Typical values of

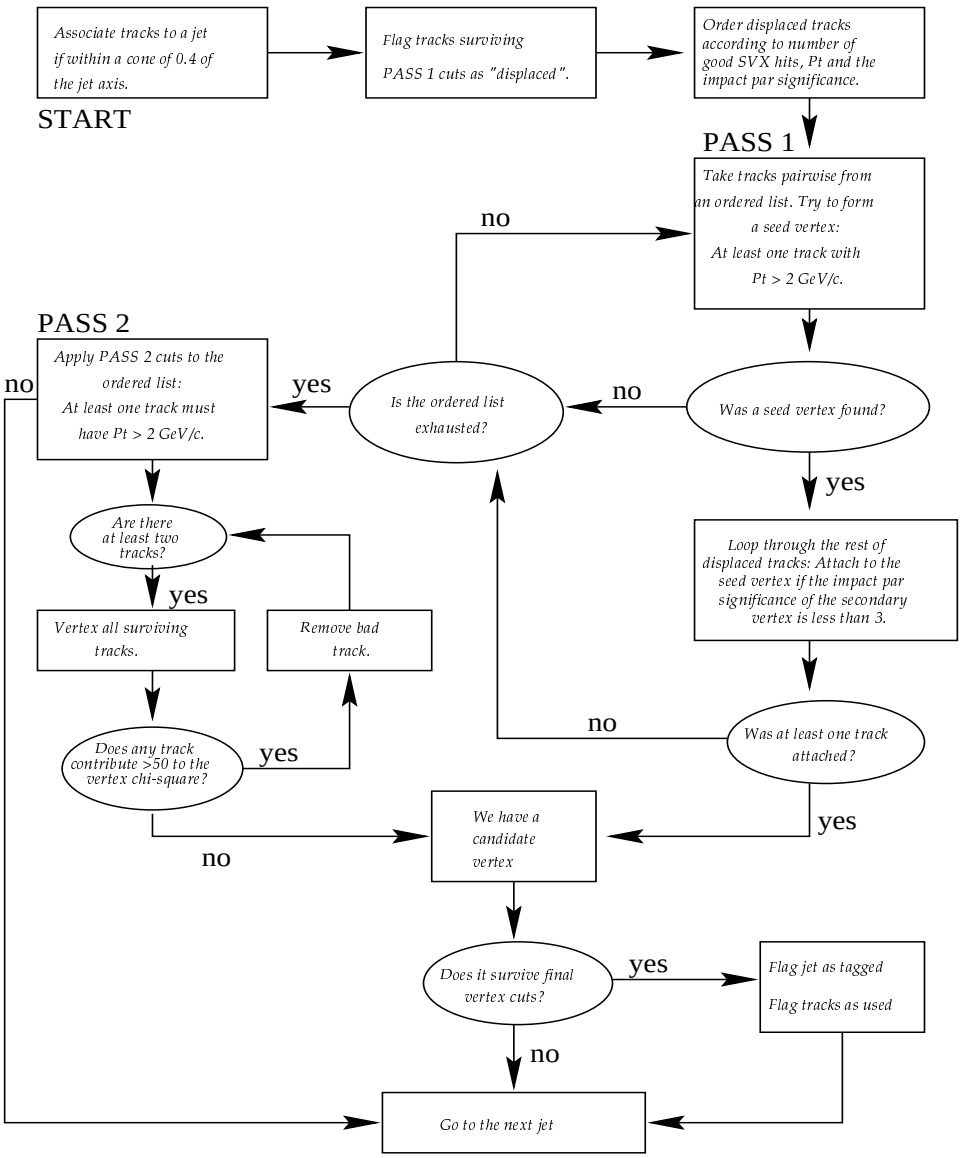

Figure 4.8: Flow chart of the SECVTX *b*-tagging algorithm.

$\sigma_{L_{xy}}$ is of the order of $\sim 130\mu$m, which is much smaller than the distance traveled by b-quark hadrons in $t\bar{t}$ events. The requirement on the maximum value of the L_{xy} is applied to ensure that the secondary vertex is inside the region defined by the SVX innermost layer radius.

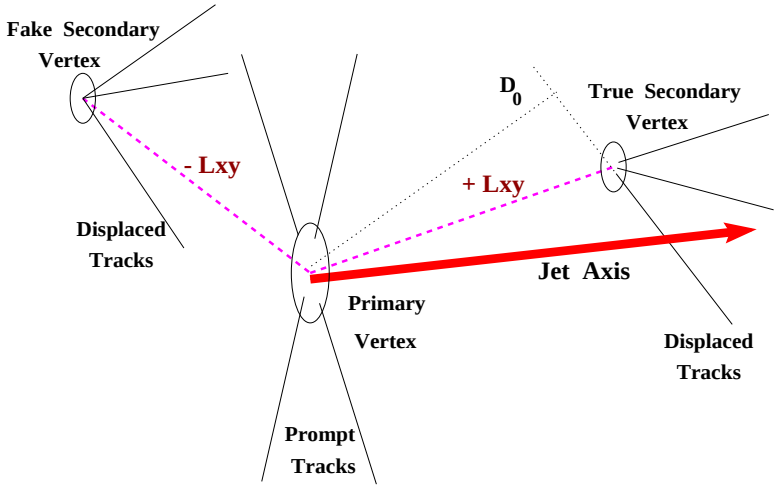

Figure 4.9: Schematic view of a real and a fake SVX b-tag in the r-ϕ plane in which the transverse displacement from the primary vertex, L_{xy}, is measured.

Figure 4.9 shows a view of two SECVTX secondary vertices with positive and negative transverse decay length respectively. If the tagged jet results from a secondary vertex with positive (negative) L_{xy} the tag is termed accordingly positive (negative) tag.

Heavy flavor decays result in secondary vertices with positive L_{xy} displacement while negative tags result from inefficiencies of the tracking reconstruction algorithm and detector resolution effects. Therefore secondary vertices with negative L_{xy} are not physical. However, tracks in light quark or gluon jets can be displaced and result in secondary vertices. Secondary vertices of this origin are called mistags and are expected to have L_{xy} symmetrically distributed about zero [59]. For heavy flavor identification, only jets with positive L_{xy} are considered as "b-tags". The negative

tags are representative of the level of the mistag contribution to the positive tags. Nevertheless, heavy flavor decays can also result in negative tags. The heavy flavor contribution to the negative tagging rate is discussed in Chapter 5.

4.2.3 Performance of the SECVTX algorithm

To confirm that the SECVTX algorithm identifies heavy flavor jets its performance is examined in the inclusive low P_T electron sample, described in Section 3.9.2. This sample is rich in heavy flavor decays because the electron comes from the semileptonic decay ($b \to e\nu c$, or $c \to e\nu s$) of a b or a c-quark in the event. As it is discussed in details in Section 4.5, $43 \pm 5\%$ of the electron jets in this sample before tagging are due to heavy flavor decays.

To verify that the tagged jets associated with the electrons include a high fraction of heavy flavor jets, the decay length of the secondary vertex associated to the jet is converted into an estimate of the decay length, *pseudo-*$c\tau$, defined as:

$$c\tau = L_{xy}\frac{M_V}{P_T^V}$$

where M_V is the mass of the reconstructed vertex formed by the combination of the tracks participating to the vertex and P_T^V is the transverse momentum of the vertex based on the P_T of the attached tracks. The prefix "pseudo" is used to indicate that the tagging algorithm does not fully reconstruct the decay vertex, since some of the decay products remain unattached to the vertex and there is also a neutrino involved in the semileptonic decays which escapes detection.

Figure 4.10 shows a comparison of the pseudo-$c\tau$ distributions of tagged electron jets in the $b\bar{b}$ data sample and the corresponding HERWIG simulation which uses

the world average B-hadron lifetime (1.5 psec). The Monte Carlo is normalized to the data using the number of electron jets in the data associated with heavy flavor decays (43±5%) and Data/Monte Carlo b-tagging efficiency scale factor of 1.25±0.13 as discussed in details in Section 4.5. The region close to zero pseudo-cτ is depleted due to the SECVTX selection criteria which are biased towards higher L_{xy}'s and consequently higher cτ's. The good agreement between the two distributions indicate

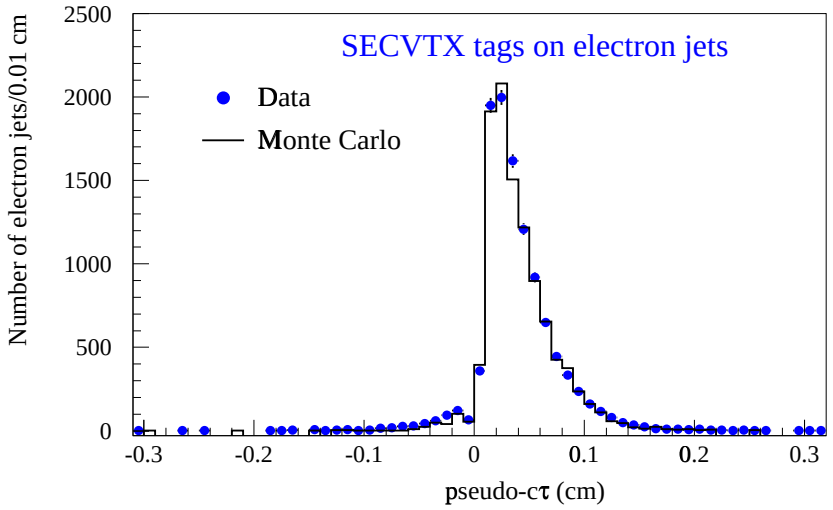

Figure 4.10: Comparison of the pseudo-cτ distributions of electron jets with SECVTX tags in the inclusive low P_T electron sample (points) and in the corresponding HERWIG simulation (histogram). The simulation includes electrons from all QCD heavy flavor production mechanisms. The heavy flavor hadrons produced by HERWIG were decayed according to the CLEO Monte Carlo [76] decay tables and the resulting events were simulated with the CDF detector simulation.

that the algorithm really identifies heavy flavor jets while the efficiency to tag a non-heavy flavor jet (mistag) is very small. In fact, the b, c and non-heavy flavor tagging efficiency in the Monte Carlo are in the ratio of approximately 40:10:1 implying that most of the tags are due to b-jets.

4.3 Jetprobability tagging algorithm

As described earlier, the SECVTX algorithm discriminates heavy flavor jets from generic jets based on the identification of a secondary vertex. However, some selection criteria are imposed both on the tracks participitating in the reconstruction of the vertex and on the displacement of the vertex. Recall that all tracks considered in the secondary vertexing are required to have $D_0/\sigma_{D_0} \geq 2.5$ and the L_{xy} significance of the secondary vertex is required to be $L_{xy}/\sigma_{L_{xy}} \geq 3.0$. The requirement on the track impact parameter significance affects the number of tracks considered in the vertexing and this in turn affects the accuracy on the location of the secondary vertex. Also as the B-hadron decay vertex approaches the primary vertex, the vertex finding efficiency is decreasing since the imposed selection criteria bias the L_{xy} towards larger values.

However, information about a particle's lifetime can be obtained using only the impact parameter of tracks in a jet without imposing any vertexing constraints. A second b-tagging algorithm, called *jetprobability*, was developed based only on the impact parameter of tracks in a jet and the probability of the given track ensemble in the jet to be consistent with non-lifetime properties.

The description of the basic concepts upon which the algorithm is built is the subject of the following sections. The efficiency of the algorithm is examined in Sections 4.5. The algorithm was introduced first by the ALEPH collaboration [57] in studies of $Z \rightarrow b\bar{b}$ decays but its features make it suitable for b-tagging at CDF.

4.3.1 The basic idea of the jetprobability algorithm

As discussed in Section 4.1.2, decay products from long-lived particles have large displacement relative to the primary vertex and are almost always emitted in the

direction of flight of the decaying particle. In contrast, particles with no lifetime, originate at the primary vertex and the corresponding impact parameter should in principle be zero. However, resolution effects or tracking algorithm mistakes can cause prompt tracks to appear displaced. These tracks are expected to be randomly displaced with respect to the jet direction of flight. This led to the definition of the signed impact parameter, and signed impact parameter significance. These variables are used to indicate whether a track originates from a location in front of the primary vertex in the direction of the jet or from a location behind the primary vertex. The signed impact parameter significance distributions, S_{D_0}, of tracks in generic jets and in b-jets from Monte Carlo simulations is shown in Figure 4.7.

Since the negative side of the signed impact parameter significance consists mainly of prompt tracks it can be used to determine the SVX resolution function. Knowing the SVX resolution function, the $|S_{D_0}|$ of any track in a jet can be compared to this resolution function and a probability the track displacement is consistent with resolution effects can be derived. This probability, called *track probability*, can take any value between zero and one. A probability value close to one indicates that the examined track originates from the primary vertex while a value close to zero indicates that the track is rather displaced and is very unlikely to emerge from the primary vertex. Tracks from zero lifetime objects are expected to be uniformly distributed with probability between zero and one. In contrast, the track probabilities of the heavy flavor decay products are expected to concentrate around zero since their displacement cannot be described by resolution effects alone. Therefore, comparing the displacement of tracks in a jet to the SVX resolution function, a test of the hypothesis of whether the observed track displacement is consistent with resolution effects alone or from a long lived particle can be made.

Taking the combination of probabilities of all tracks in the jet, a *jet probability* can be defined. It represents the probability that the given ensemble of tracks in the jet is consistent with resolution effects alone and therefore the jet is consistent with the zero lifetime hypothesis. The jet probability is also a continuous function and can take any value between zero and one. Jets with probability close to zero are likely to result from b or c-hadron decays, while generic jets are expected to be equally distributed with jet probability values between zero and one.

In summary, using the signed impact parameter of tracks in a jet a continuous probability function can be defined and it can be used to discriminate between heavy flavor and generic jets. For every jet containing one or more SVX tracks passing some quality criteria, a probability value can be assigned at all times and this value characterizes the lifetime properties of the specific jet. In contrast to the SECVTX algorithm which specifies whether a jet is a heavy flavor candidate based on the presence or not of a secondary vertex, the jetprobability algorithm assigns a probability value to all jets regardless of vertexing constraints. The probability assignement can be made at all times regardless of the impact parameter significance of the tracks forming the track ensemble. The fact that the jet probability corresponds to a continuous distribution offers two alternative approaches in studying the heavy flavor content of jets. First, the full shape of the jet probability distribution can be used in combination with the probability shapes from heavy flavor and background control samples in order to study the heavy flavor composition of a given sample. Second, selecting jets with probability less than a maximum value, samples with different heavy flavor composition can be selected. Furthermore, the maximum jet probability selection value underlines in principle the amount of non-heavy flavor allowed in the selected sample.

4.3.2 Determination of the SVX impact parameter resolution function

In order to measure the SVX resolution function, tracks originating from the primary vertex are needed. These tracks can be displaced positively or negatively with respect to the jet direction. A sample of prompt tracks can be selected using only tracks with negative signed impact parameter significance. Because heavy flavor decays can contribute tracks with negative impact parameter significance (see Section 5.1.6) it is instructive to use samples with small heavy flavor content in order to minimize the bias to the negative impact parameter significance distribution from the presence of real heavy flavor products. For this reason, tracks in a sample of generic jets collected with the JET_50 trigger are used. It is also important to reject badly measured tracks and tracks from K^0 and Λ decays. The kinematic and quality criteria used to select the SVX tracks for extracting the SVX impact parameter resolution function are listed in Table 4.3.

The SVX resolution function, $\mathcal{R}(|S_{D_0}|)$, is determined by fitting the negative impact parameter significance distribution of the selected prompt tracks. If all errors were strictly Gaussian the resolution function, $\mathcal{R}(|S_D|)$, would take the form $|S_{D_0}|e^{-|S_{D_0}|^2/2\sigma^2}$. However, in practice there are substantial non-Gaussian tails in the error due to the non-Gaussian behaviour of the tracking. The resolution function is well approximated with an exponentially decreasing tail of the form $e^{-|S_{D_0}|/\lambda}$ added to a combination of two independent Gaussian components describing the core of the distribution.

In order to account for differences in the tracking resolution between the data and the Monte Carlo simulations, resolution functions are derived independently for

the Monte Carlo sample. For this purpose, a JET_50 QCD Monte Carlo sample is used. The simulated JET_50 sample is created with the HERWIG generator and

Requirements for SVX resolution function
$P_T \geq 1.5$ GeV/c
Within a cone of radius 0.4 around a jet with $E_T \geq 15$ GeV
CTC quality criteria as in Section 4.2
3-D track
≥ 2 axial superlayers with ≥ 4 hits/superlayer
≥ 2 stereo superlayers with ≥ 2 hits/superlayer
$\|D_0\| \leq 0.1$ cm
$\|Z_{trk} - Z_{primary\ vertex}\| \leq 5$cm
$\frac{\chi^2}{Nhits} \leq 6$
Not part of K_s or Λ decay vertex
Hit Pattern
At least two SVX hits
Tracks with 2 SVX hits one of which at layer 0 are excluded

Table 4.3: Kinematic and quality criteria used to select SVX tracks in order to measure the SVX impact parameter resolution.

generating a standard $2 \rightarrow 2$ parton process (process 1500). The generated events are then passed through the full detector simulation, QFL$'$ [77]. In order to simulate the trigger requirement, events with at at least one jet after detector simulation with observed $E_T \geq$50 GeV are selected. In the Monte Carlo sample, the tracks used to extract the SVX resolution function are selected with the same criteria as in the data. No requirement on the track's origin is imposed in order to include the effect of the heavy flavor contributions on the resolution function which is present in the data.

The tracks are classified in twelve independent categories according to the num-

ber of associated SVX hits (Nhits=2:4) and according to the number of those hits shared by other tracks (Nshared=0:4). This classification is necessary because the error of the impact parameter strongly depends on these parameters. For each of the categories, a probability distribution that characterize the resolution function is obtained. Each such normalized distribution is fitted to the functional form described above (two Gaussian plus an exponential tail). In practice, both positive and negative sides of the distribution are fitted simultaneously by adding a second exponential tail to parametrize the positive side of the distribution. However, only the negative parametrization is used to define the resolution function. Samples of the fits which used to define the resolution functions for six classes of tracks are shown in Figures 4.11. In all cases, the distributions are quite asymmetric with the excess on the positive side attributable to tracks from heavy flavor decays.

4.3.3 The track probability

Having define the SVX impact parameter resolution function, $\mathcal{R}(S_{D_0}, N_{hits}, N_{shared})$, function of the number of SVX hits (N_{hits}) and shared hits (N_{shared}), it can be used to define the track probability:

$$P(S_{D_0}) = \frac{\int_{-\infty}^{-|S_{D_0}|} \mathcal{R}(S_{D_0}) dS_{D_0}}{\int_{-\infty}^{0} \mathcal{R}(S_{D_0}) dS_{D_0}} \tag{4.1}$$

This definition of track probability gives essential the probability to observe the given track with impact parameter significance $-|S_{D_0}|$ or smaller and represents the probability that the track is prompt. The track probability as given in the above expression is properly normalized so that the integral of the probability is one. Tracks with negative S_{D_0} are expected to have flat track probability distribution since they are used

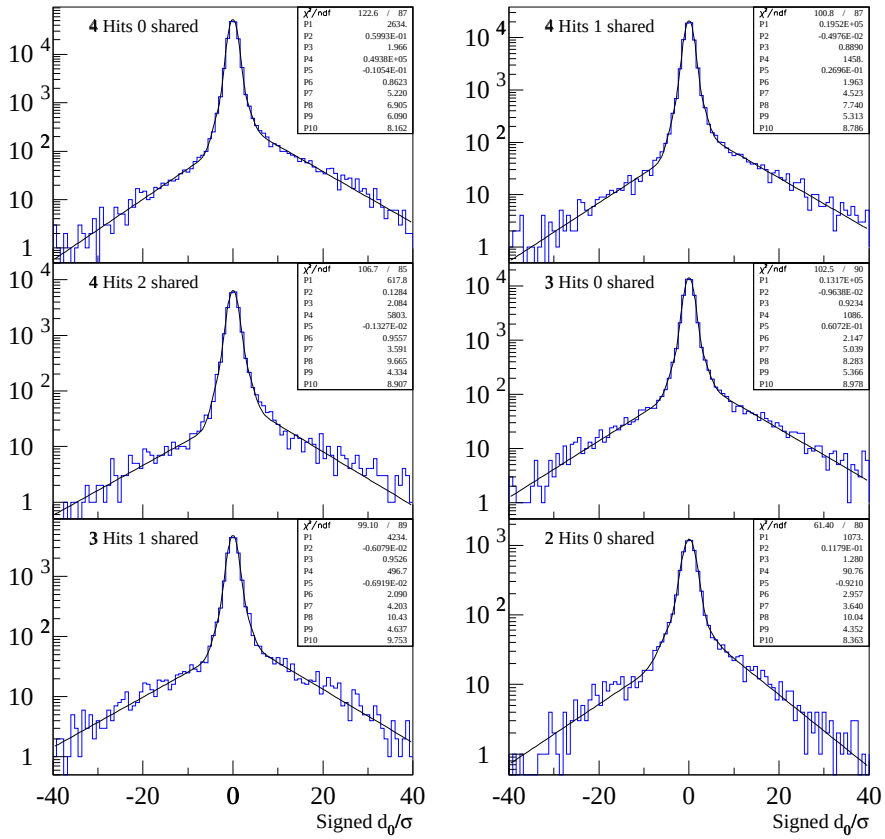

Figure 4.11: Distributions of signed impact parameter significance of tracks used to derive the SVX resolution function. Tracks associated with jets collected with the JET_50 trigger are used for the parametrization of the resolution function. The fit function consists of two Gaussian and two exponential distributions one for the negative side and one for the positive side of the distribution. The resolution function for each case is determined by the fit on the negative side of the distribution. The excess of the positive side is due to heavy flavor decay products.

in the construction of the resolution functions. The "flatness" of the negative track probability distribution reflects the goodness of the parametrization used for the resolution function.

Figure 4.12 shows the distributions of track probability for tracks with positive (open) and negative (shaded) signed impact parameter significance. The probability distributions of negative impact parameter significance tracks are indeed flat over most of the range. However, a small excess for probability values close to zero is observed in some cases. Two effects contribute to this excess. One is due to mismeasured tracks appearing in the tails of the negative S_{D_0} distribution. It implies that the S_{D_0} distribution of these tracks is not described adequately by the exponential term of the resolution functions. The second effect is due to tracks from heavy flavor decay products which are mis-signed and appear with negative S_{D_0}. The probability distribution of tracks with positive signed impact parameter significance resembles in most of the region the corresponding distribution of tracks with negative signed impact parameter significance. However, a large excess of tracks at small values of track probability is observed. This excess is attributable to tracks from decays of long lived particle.

Figure 4.13(a) shows the track probability distributions for tracks with positive and negative signed impact parameter significance from the simulated JET_50 Monte Carlo sample. The probability distribution of all negative impact parameter significance tracks regardless of origin is flat over the entire probability range as expected because these tracks are used in the construction of the resolution function. Similarly to the data, tracks with positive S_{D_0} have a peak around zero probability values due to tracks from heavy flavor decays. Plots (b), (c) and (d) in Figure 4.13 show the track probability distributions for non-heavy flavor jets, b and c-jets respectively. The

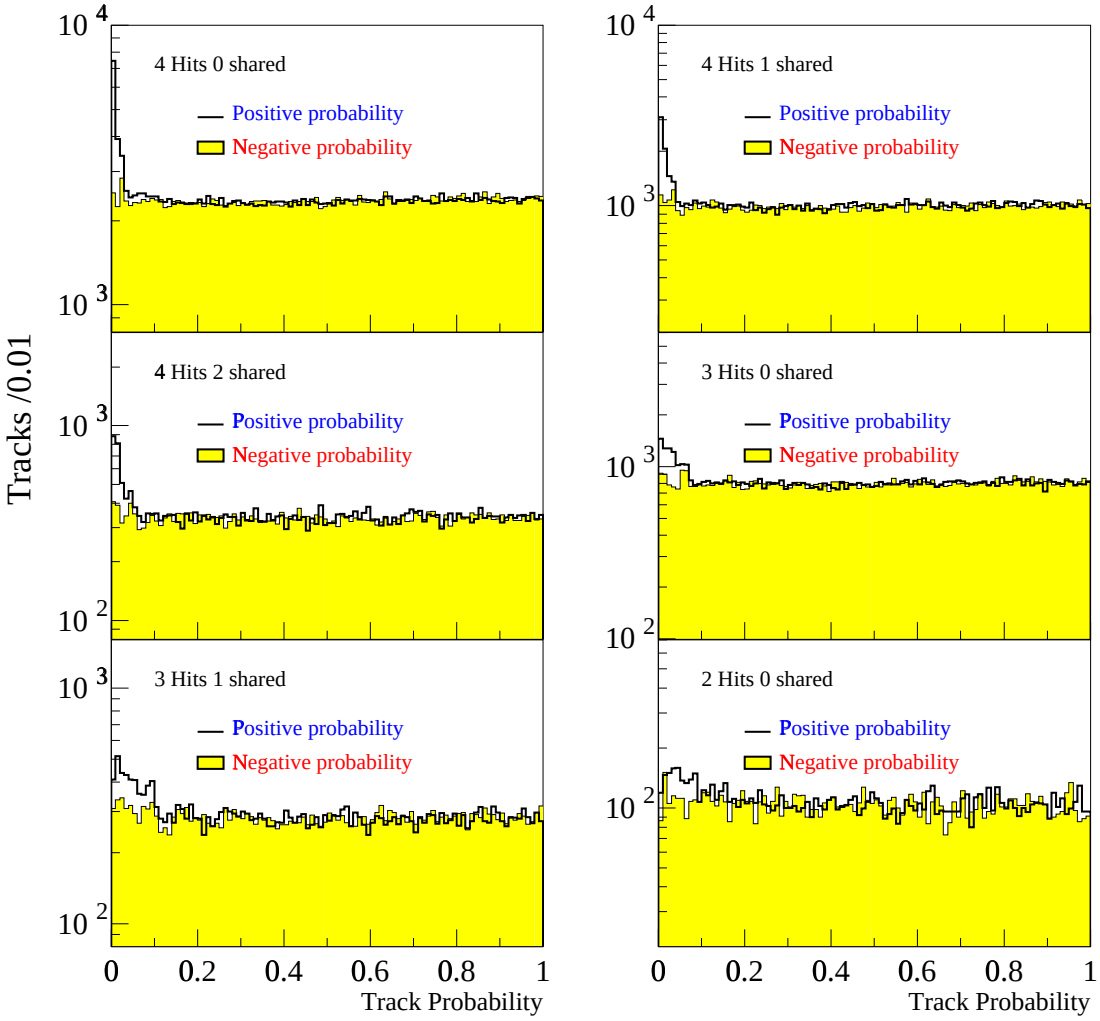

Figure 4.12: Track probability distributions for six classes of tracks. A sub-sample of these tracks was used to derived the corresponding resolution functions shown in Figure 4.11. The probability distributions corresponding to tracks with positive S_{D_0} (open) and to tracks with negative S_{D_0} (shaded) are shown. There is a peak at zero for positive track probabilities attributable to tracks from heavy flavor decays. The negative track probabilities are uniformely distributed over the entire range as expected.

distributions corresponding to tracks in heavy flavor jets are complementary to the ones corresponding to generic jets since the combination of these tracks was used to derive the resolution function. Therefore, the distribution corresponding to generic jets shows a "deep" at small values of track probability. This region corresponds to tracks with large $|S_{D_0}|$ which are expected to originate from heavy flavor decays. In fact, both the positive and negative probability distributions from tracks in b and c-jets show a large excess around zero. The distribution corresponding to tracks from c-jets is wider than the ones corresponding to tracks from b-jets reflecting the shorter lifetime of the c-hadrons. Comparison between the number of tracks under the peak of positive and negative track probability distributions show that $\sim 15\%$ of tracks from heavy flavor decays are mis-signed.

There is a very appealing feature in the use of the track probability. According to the definition, the track impact parameter significance is converted to a normalized probability through the resolution functions. This procedure is performed independently on data and Monte Carlo. The resulting track probability does not depend on the details in the modeling of the tracking algorithm, especially on the resolution which is difficult to model. The tagging algorithm is based on the ensemble of track probabilities in a given jet and therefore is independent of the tracking details between data and Monte Carlo. Any differences in the performance of the algorithm between data and Monte Carlo are left in the difference on the number of tracks considered in the ensemble and the uncertainties in the modeling of the parton fragmentation and the heavy flavor lifetime which are part of the Monte Carlo generator. In Section 4.5 the efficiency of the algorithm in the data and Monte Carlo is compared and in fact proves the above expectation.

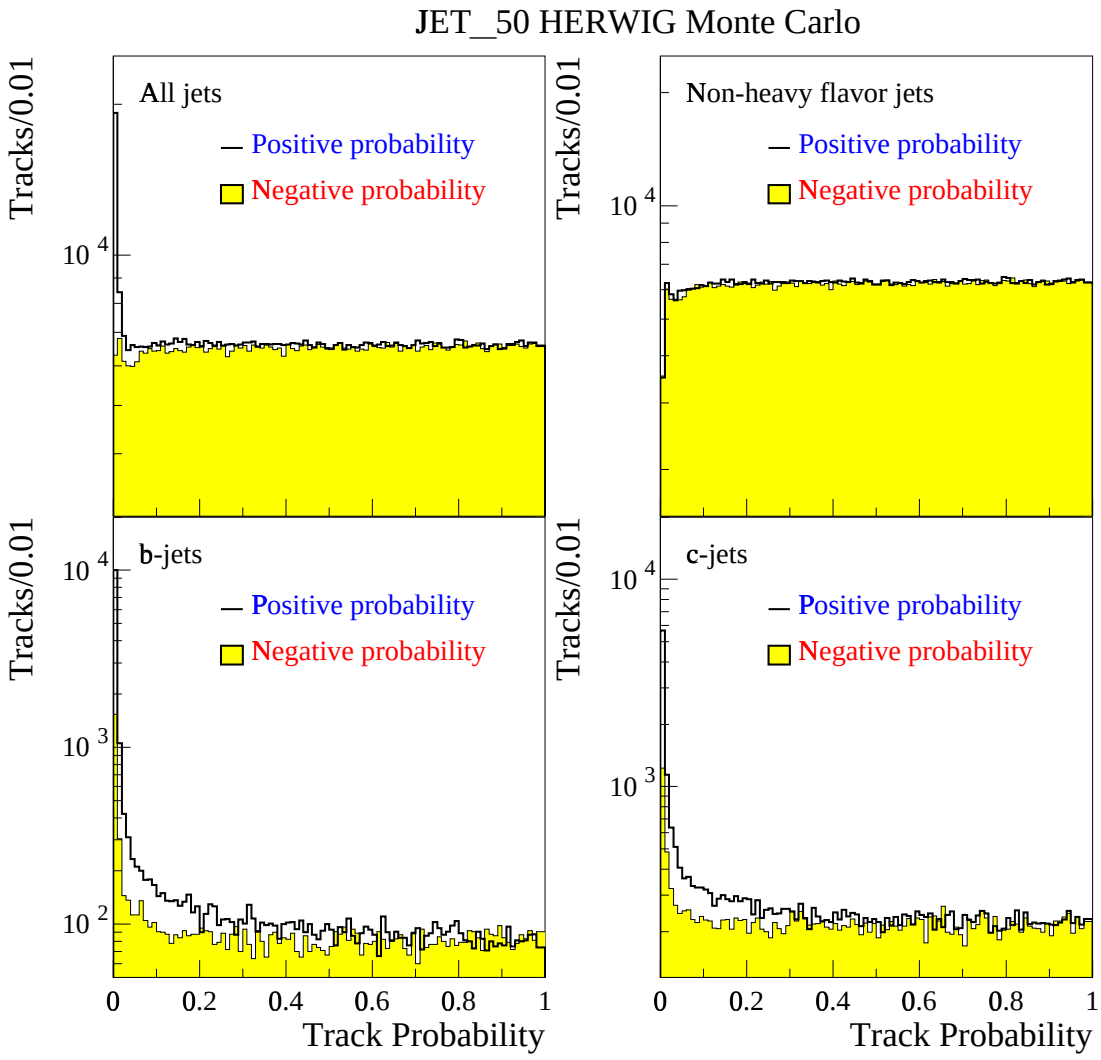

Figure 4.13: Positive(open) and negative(shaded) track probability distributions for tracks in Monte Carlo simulated jets with different heavy flavor component. The upper left plot includes tracks from all jets regardless of their origin. These tracks were used to derive the resolution function in the Monte Carlo simulation. The rest of the plots show the track probability distributions for tracks in generic, b and c-jets respectively. The distributions corresponding to heavy flavor jets are complementary to the ones for generic jets. Tracks from b-jets are sharply peaked at zero track probability, whilst the corresponding distributions for c-jets are broader due to the shorter c-lifetime. Tracks in generic jets have a uniform track probability distribution over the whole range. The deep at zero reflects the way the resolution functions were derived.

4.3.4 Converting the track probabilities to jet probability

As discussed in the previous section, the track probability variable offers a handle for discriminating tracks of long lived particles from tracks originating at the primary vertex. However, decays of heavy flavor hadrons result in jets with large track multiplicity. Thus, studies based on individual track probabilities lack the full information on the correlations among the tracks in the decay. It is instructive to take the combination of track probabilities and "grow" it into an overall probability which characterizes the jet.

The *jet probability* is defined as the probability that the jet would have tracks with the observed set of track probabilities and consequently S_{D_0}'s or any combination less probable. As in the case of track probabilities, the jet probability corresponds to a continuous variable with values in the range of zero and one. For the definition of the jet probability only same sign impact parameter significance tracks are considered. Therefore two jet probability values can be defined for the same jet. The *positive jet probability* is defined using all tracks in the jet with positive signed impact parameter significance, S_{D_0}. Accordingly, the *negative jet probability* is defined using all tracks in the jet with negative S_{D_0}. The negative jet probability takes also values between 0 and 1 and the prefix negative is used to indicate the use of tracks with negative signed impact parameter significance.

The jet probability definition can be easily visualized for the case of a jet containing only two tracks with probabilities P_1 and P_2, as shown in Figure 4.14. The combination of the two track probabilities defines the curve of constant probability $\prod = P_1 \cdot P_2$. The probability, $\mathcal{P}$, that the two track probabilities result in a combination with probability $\prod' < \prod$ is given by the area of the shaded region in Figure 4.14. The area of this region is equivalent to the jet probability for the given jet. In this

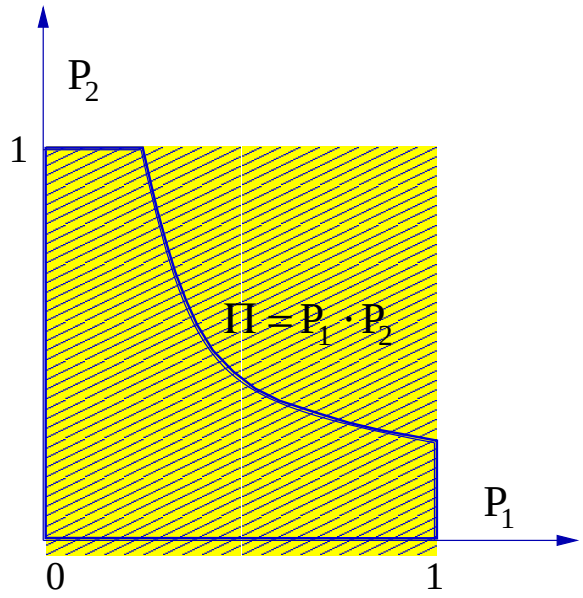

Figure 4.14: Definition of the jet probability for a jet containing two tracks with probability P_1 and P_2. The shaded area enclosed by the curve of constant probability, $\prod = P_1 \cdot P_2$, defines the jet probability. It represents the set of two-track combinations with probability less or equal to $\prod$.

simple case, the integration of the area gives:

$$\mathcal{P}_{jet} = \prod(1 - \ln \prod) \tag{4.2}$$

By induction, the jet probability for a general case of a jet containing N tracks is given by

$$\mathcal{P}_{jet} = \prod \sum_{k=0}^{N-1} \frac{(-\ln \prod)^k}{k!} \tag{4.3}$$

where, $\prod = P_1 P_2 \cdots P_N$, is the product of the individual track probabilities of the selected tracks.

All SVX tracks in a jet which satisfy the kinematic and quality criteria listed in Table 4.3 are considered in the construction of the jet probability. In addition, it is required that the SVX tracks have at least 2 good SVX hits as defined in Table 4.2. A minimum number of 2 such SVX tracks with positive (negative) S_{D_0}'s are required

Tracks satisfy the criteria in Table 4.2
At least 2 good SVX hits as defined in Table 4.2
For positive jetprobability: jet has at least two tracks with positive S_{D_0}
For negative jetprobability: jet has at least two tracks with negative S_{D_0}

Table 4.4: Requirements imposed on SVX tracks and jets for the definition of the positive/negative jet probability.

in a jet in order to form the positive (negative) jet probability. Jets which satisfy the above criteria are called *jetprobability taggable*. The list of the criteria applied for the definition of positive and negative jet probability are shown in Table 4.4.

Figure 4.15 shows the positive and negative jet probability distributions for jets in a sample of events collected with the JET_50 and JET_140 triggers. The positive jet probability, shown in Figure 4.15(a), includes tracks from heavy flavor decays and consequently an excess of jets close to zero probability is observed. Figure 4.15(b) shows the distribution of the negative jet probability. The distribution is quite flat over most of the probability range. This is expected, because the negative jet probability is formed of tracks with negative S_{D_0} which were used in the construction of the resolution function and their track probability is by construction flat. A small excess at low value of negative jet probability is observed and is attributtable to heavy flavor jets. In fact this excess disappears, as shown in figure 4.15(c), when plotting the negative jet probability of those jets which have positive jet probability in the range of $0.1 \sim 1$. Selecting jets away from the peak around zero positive jet probability ensures that the jets are free of lifetime effects and consequently the negative jet probability is flat. The contribution of heavy flavor decays to the negative jet probability is estimated in Chapter 5 and the above mentioned effect is discussed in greater detail.

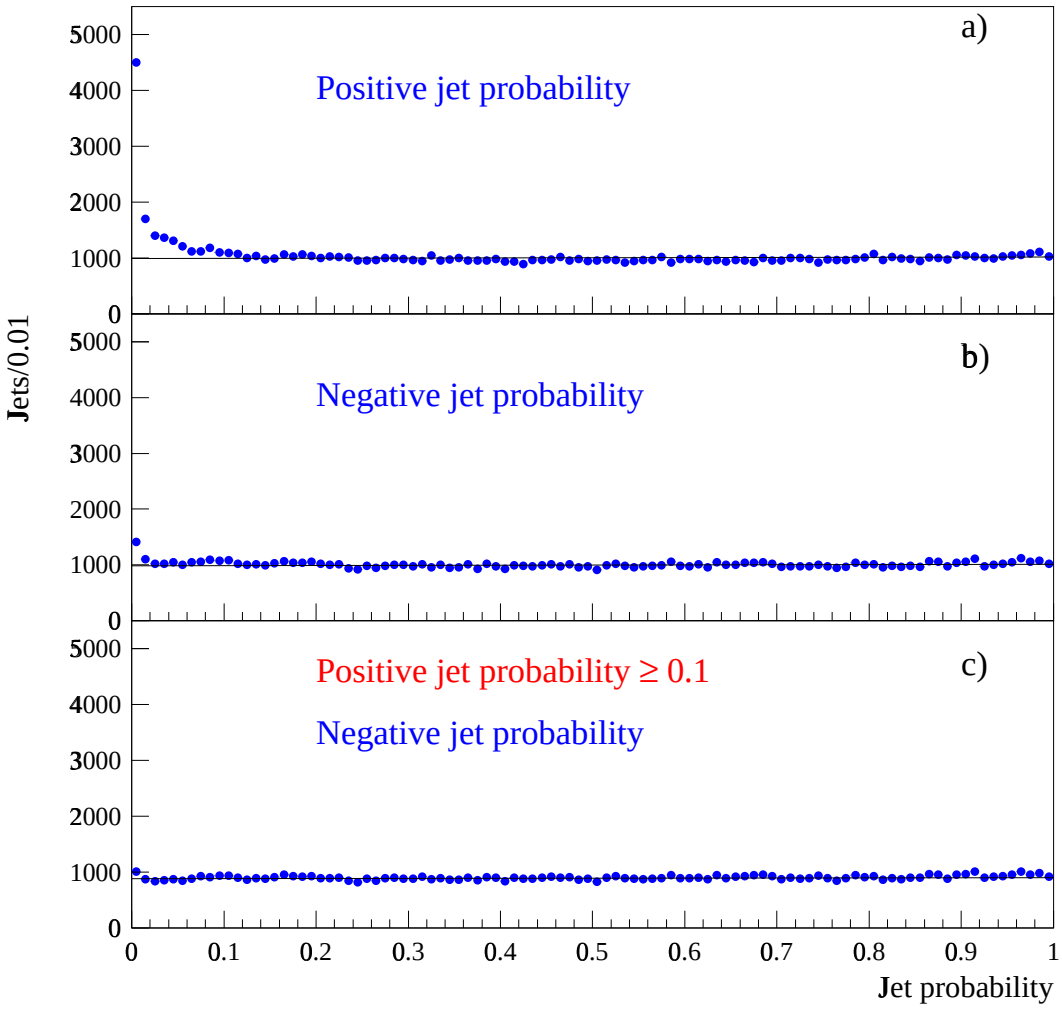

Figure 4.15: Distributions of positive (a) and negative (b) and (c) jet probability for jets in a sample formed by a mixture of events from the JET_50 and JET_140 samples. The negative jet probability distribution shown in (c) corresponds to jets with positive jet probability greater than 0.1. This selection requirement removes most of the jets from heavy flavor decays. The lines in all plots represent a simple likelihood fit to the negative jetprobability distribution shown in (b), with a first degree polynomial of the form $P_0 + P_1 x$. The slope in the fit translates to a 1.6% change over the entire region of jetprobability. In figure (c) the fit result is scaled by the number of jets (90442/101050).

In conclusion, the positive jet probability distribution offers a separation power to discriminate heavy flavor jets against primary or non-heavy flavor jets. On the other hand, the negative jet probability can be used to study mistakes of the algorithm and mistags, the same way the negative L_{xy} tags of the SECVTX algorithm are used.

4.3.5 Definition of jetprobability tags

The jetprobability tags are derived using SVX tracks contained in a jet of cone 0.4 around the jet-axis. As described before, the jets are required to have $E_T \geq$15 GeV and $|\eta_{det}| \leq$2.0.

The jetprobability values are calculated with the derived resolution parametrizations for each class of SVX tracks in the jet and the resulting track probabilities are subsequently converted to jet probabilities. The positive and negative jetprobability values are derived using tracks with positive and negative signed impact parameter significance, respectively.

A jet is considered to be tagged when it contains 2 or more good SVX tracks (see Table 4.4) with *positive* signed impact parameter resulting to a *jetprobability value of* $\leq$ *0.05*. This is a positive jetprobability tag. Similarly, a jet is considered a *negative jetprobability tag* if it contains 2 or more good SVX tracks with *negative* signed impact parameter and with resulting *jetprobability* $\leq$ *0.05*.

In a perfect world, requiring jetprobability values less than 0.05 should correspond to a 5% mistag rate. Fitting the jet probability distribution shown in Figure 4.15(b) to a first degree polynomial function, $P_0 + P_1 \cdot x$, in the interval 0.1$\sim$1.0 the fit returns as results $P_0 = 100340 \pm 826$ and $P_1 = -1608 \pm 1508$ with χ^2/ndf=1.3 and $\rho = -0.9$. Extrapolating the fit results in the $0.0 \sim 0.05$ jetprobability region, 5015$\pm$36 negative jetprobability tags are expected while 5607 tags are observed. The

excess of 592 observed negative jetprobability tags over the expected ones from the fit is attributed to heavy flavor jets and it is of the order of 10.5±0.2% of the number of negative jetprobability tagged jets (jetprobability≤0.05). The expected mistag rate after subtracting the heavy flavor contribution is 5% of the total number of jets (101050 jets) in the sample.

Figure 4.15(c) shows the negative jetprobability distribution of jets whose heavy flavor content is depleted by requiring their corresponding positive jetprobability to be greater than 0.1. The extrapolation of the fit in the region of jetprobability 0.0∼0.05 gives 4441±34 negative jetprobability tags while 4455 are observed which corresponds to 4.94% of the total number of jets.

Figure 4.16 show a comparison of the positive jetprobability distribution of electron jets in the low P_T inclusive electron sample and the corresponding HERWIG simulation. The presence of an additional jet tagged by SECVTX is required in order to reduce the contribution of non-heavy flavor e-jets. Data and simulation are normalized to each other according to the method described in Section 4.5.

4.4 Tracking degradation

The tagging efficiencies for both $t\bar{t}$ events and all other processes contributing tags to the $W+ \geq 1$ jet sample are estimated based on Monte Carlo simulation. Because of this dependence on the CDF detector simulation and especially on the simulation of the CTC and SVX tracking efficiencies, it is necessary to cross-check the modeling of the b-quark jets in the simulation with data.

Effects that can contribute to an erroneous determination of the tagging efficiency in the simulation include:

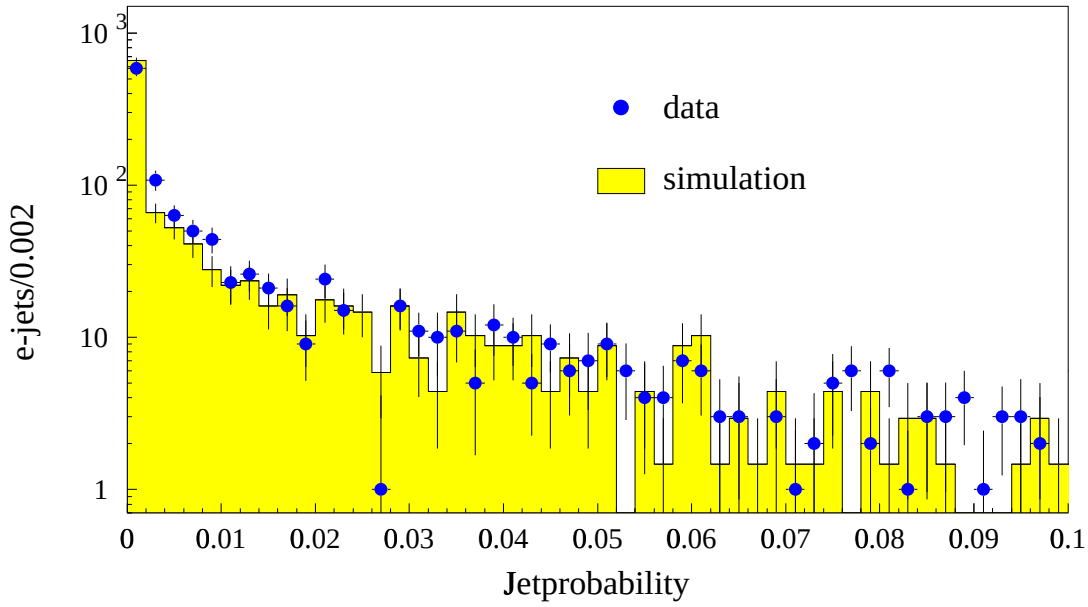

Figure 4.16: Comparison of the positive jetprobability distributions of electron jets in the low P_T inclusive electron data sample and the corresponding HERWIG simulation. The presence of an away jet tagged by SECVTX is required.

- The tracking efficiency in the Monte Carlo is different than in the data. This implies that the number of tracks present in a jet is different in data and simulation. Such a difference affects the tagging efficiency in two ways. Firstly, fewer jets can be flagged as taggable (it is always required a jet to have at least two goof SVX tracks to be considered by the SVX based tagging algorithms). Secondly, fewer tracks in a jet can be available for being considered by the tagging algorithms.

- The modeling of the impact parameter resolution is wrong in the Monte Carlo.

- The modeling of the b/c decays, fragmentation and lifetime is not correct in the Monte Carlo.

The biggest effect that contribute to the discrepancies between the tagging effi-

ciency in the Monte Carlo and the data is due to the modeling of the CTC and SVX tracking efficiency.

By studying individual tracks, it is determined that the CDF tracking simulations are very optimistic. The reconstruction efficiency of tracks in the Monte Carlo simulations is higher than the efficiency observed in the data. The tracking reconstruction efficiency in the data was determined embedding random Monte Carlo tracks in jet data and rerunning the track reconstruction algorithm. Events from the JET_50 sample were used for this study. Counting the number of times the embedded track is found after full tracking reconstruction, the track reconstruction efficiency is determined. In measuring the track finding efficiency only tracks coming the primary vertex are used. This method is used because as described in Section 3.1.2 tracks are reconstructed from individual hits in the CTC and SVX detectors. The difference between the track finding efficiency in the data and in the simulation defines the degradation factor to be applied.

Table 4.5 show the measured track reconstruction efficiency in the data using the track embedding technique and the track reconstruction efficiency measured in the Monte Carlo [62]. The error on the above procedure is derived by comparing the

Sample	CTC-only	SVX-only	CTC+SVX	Luminosity	Total
Track embedding	0.94	0.93	0.87	0.95 ± 0.02	0.83 ± 0.04
Simulation	0.994	0.992	0.983		0.983

Table 4.5: Track reconstruction efficiency in the data and in the simulation.

number of hits used per superlayer by the embedded track and tracks in the JET_50 data. Good agreement is found for the outer superlayers whilst a difference of up to 20% is found for the inner superlayer. Reducing the CTC wire efficiency for the

embedded tracks, the tracking efficiency reduces by 4%. However, while all embedded tracks are prompt, tracks in the data can result from K or Λ decays and therefore the above difference is overestimated. Based on the above difference a conservative systematic uncertainty of 4% is assigned to the above described procedure [62].

In practice, the reduced efficiency is parametrized as a function of the number of hits around the object track. When the number of hits is high, the probability of reconstructing a given track decreases due to the ambiguities introduced by the presence of extra hits around the examined track. The tracking reconstruction in the simulation includes the individual chamber efficiencies but does not incorporate the effect of reduced track finding efficiency due to overlapping hits and therefore given the amount of hits in the CTC or SVX detectors, a track can always be reconstructed. In order to compensate for this effect, the loss of track reconstruction efficiency is parametrized as a function of the number of hits around the embedded track. The obtained parametrization is then used in the Monte Carlo simulation to degrade the track finding efficiency.

Analyzing Monte Carlo data, the track degradation is performed in two steps. First for each track in the simulation, the number of hits around the object track is found and a quality factor is assigned according to the hit density. This quality factor corresponds to specific value of track reconstruction efficiency as determined by the embedding track method. The examined track is kept or discarded drawing a random number between 0 and 1 and comparing to the corresponding track reconstruction efficiency.

Since the b-tagging depends on the number of tracks in the CTC+SVX tracks in a jet, the track degradation results in a lower b-tagging efficiency. Furthermore, the procedure of introducing the track degradation in the Monte Carlo takes also into

account automatically E_T dependent effects on the tagging efficiency since as the jet E_T increases the number of overlapping tracks in a jet it is also increasing and consequently the density of hits. As a consequence the tracking reconstruction inefficiency becomes increasingly higher and therefore the tagging efficiency is reduced.

Studying the effect of track degradation on the tagging efficiency in $t\bar{t}$ simulated events show that the b-tagging efficiency after degradation reduces to 85.5$\pm$0.9% of the efficiency before degradation. Comparison between the b-tagging efficiency in the degraded and non-degraded $t\bar{t}$ Monte Carlo are discussed in Section 7.5.

The overall b-tagging performance of the degraded Monte Carlo is examined in the following section, comparing the b-tagging efficiency in the low P_T inclusive electron data sample which is enriched in semileptonic b decays and in the corresponding HERWIG simulation.

4.5 The b-jet tagging efficiency and the measurement of the data to Monte Carlo b-tagging efficiency scale factor

In Section 4.4, the track degradation was introduced in the simulation in order to model the reduced track finding efficiency observed in the data when tracks are contained in a jet. In this section, the tagging efficiencies of the SECVTX and Jetprobability are evaluated in the data and in the simulation which includes the tracking degradation and a *scale factor* is derived in order to adjust the two tagging efficiencies.

In the case the tracking algorithm, efficiency, impact parameter resolution and the heavy flavor decays are modeled correctly, the measured tagging efficiency in the data should be the same to the one measured in the simulation. Any differences should be taken into account in the so called data to Monte Carlo *tagging efficiency scale factor* defined as:

$$SF = \frac{\epsilon_{b-tag}^{data}}{\epsilon_{b-tag}^{MC}}$$

The tagging efficiency in the data is measured using the low-P_T central inclusive electron sample of run 1B. The sample is selected requiring

- an electron with $E_T \geq 10$ GeV, passing all the electron quality criteria used to identify electrons from W decays (see Table 3.1).
- To reduce contribution from W or Z electrons, it is required that the electron isolation is Iso^{ele} $\geq$0.1.
- The electron is also required to be within a cone of 0.4 around the axis of a taggable jet (a jet with at least two good SVX tracks and $|\eta| \leq 2.0$) with $E_T \geq$15 GeV. This jet is called *e-jet*
- It is also required the presence of an additional taggable jet with $E_T \geq$15 GeV and $|\eta| \leq 2.0$. This jet is called *a-jet*.
- The primary vertex is required to be inside the SVX fiducial volume $|Z_{svx}| \leq 35$ cm.
- Runs flagged as bad are removed from the sample.

The sample selected with the above criteria consists of 55248 events.

	$b\bar{b}$	b f.e.	$g \rightarrow b\bar{b}$	$c\bar{c}$	c f.e.	$g \rightarrow c\bar{c}$
e-jet with SECVTX tag	3200	125	986	151	8	79
e-jet with JPB tag	3970	170	1195	446	25	193
a-jet with SECVTX tag	1448	118	155	73	6	25
a-jet with JPB tag	1835	145	200	169	22	55

Table 4.6: Heavy flavor content of tagged electron and away jet in the simulated low P_T inclusive electron sample.

To measure the tagging efficiency in the Monte Carlo, a similar sample to the data was generated. The simulated sample was created using the HERWIG generator. A generic $2 \rightarrow 2$ process (HERWIG process 1500) was generated with minimum parton P_T at hard scattering of $P_T \geq 13$ GeV. The generic QCD process was selected in order to include electrons from all possible heavy flavor production mechanisms (b and c direct production, flavor excitation and guon splitting). The MRS(G) parton parametrization was used and the masses of the b and c quarks were set to $m_b = 4.75$ GeV/c^2 and $m_c = 1.5$ GeV/c^2, respectively. The generated events were interfaced with the CLEO Monte Carlo for modeling the decay and kinematics of the heavy flavor hadron decays. After CLEO, events were selected at generator level requiring an electron with $P_T \geq$8 GeV and $|\eta| \leq 1.2$ in order to simulate the low-P_T inclusive electron trigger. The selected events were simulated with full detector simulation including track degradation. Approximately 2 billion $2 \rightarrow 2$ events were generated resulting to a sample of $\sim$300K events after detector simulation. The resulting sample was then analyzed with the same selection criteria described for the data. The resulting sample consists of 16547 events.

Table 4.6 shows the heavy flavor composition of the tagged electron and away jets in the simulated sample. Table 4.7 shows the heavy flavor composition of the simulated events with double SECVTX or Jetprobability tags.

Double SECVTX tags						
	$b\bar{b}$	b f.e.	$g \to b\bar{b}$	$c\bar{c}$	c f.e.	$g \to c\bar{c}$
e-jet	448	41	54	4	1	1
a-jet	455	33	51	4	0	6
Double Jetprobability tags						
	$b\bar{b}$	b f.e.	$g \to b\bar{b}$	$c\bar{c}$	c f.e.	$g \to c\bar{c}$
e-jet	771	73	116	37	2	12
a-jet	742	55	84	38	6	17

Table 4.7: Heavy flavor content of the simulated inclusive low-P_T electron sample when both electron and away jet are tagged.

The tagging efficiency is measured examining the rate of tagged e-jets in events with a tagged away jet. This is the so called *"double tag"* method. The tagging efficiency is given by:

$$\epsilon_{b \to eX} = \frac{N_{doubles}}{N^{hf}_{a-jet}}$$

This technique is used because it offers a higher purity b-sample. The same method is applied in the simulated events.

Table 4.8 show the tagging rates in the data and in the data and in the simulation. Mistags in the data are calculated with the mistag probability matrices discussed in Chapter 5, while in the simulation refer to tagged jets not matched to a heavy flavor hadron at generator within a cone of 0.4 around the jet axis. The mistags for double tagged events are calculated adding the probabilities of mistagging the e-jet and the a-jet and subtracting the product of the two. The column QCD tags refer to the amount of tags in the away jet due to generic QCD jets and it is calculated using the positive tagging probability matrices. The number of estimated QCD tags on the away jet shown in table is after subtracting the expected mistag contribution.

The data sample before tagging is approximately 40% pure in electrons from heavy

flavor decays [2, 52, 53]. Therefore, a fraction of the events with tags on the away jet results from generic QCD jets and it is not associated with a real heavy flavor e-jet. These events need to be removed from the sample with a tagged away jet in order to estimate the e-jet tagging efficiency on real heavy flavor jets. To estimate this QCD fraction on the away jet, the knowledge of the b-purity of the sample is necessary. The events with generic QCD tags in the away jet are estimated using the positive tagging probability matrices (see Chapter 5 and scaling the result according to the b-purity of the sample. Traditionally, the b-purity of the sample was measured by looking at the rates of tagged e-jets which contained also a soft μ identified by the SLT algorithm. Using this method, the b-purity of the sample was found to be $43\pm8\%$ [2, 52, 53] assuming no electron contribution from c-production.

Since the current simulation includes e-jets from all possible heavy flavor production mechanisms and it is about to be compared to the data the knowledge of the b-purity of the e-jets in the data is not necessary from the beginning. Instead it is derived parallel to the b-tagging efficiency scale factor with the following iterative procedure:

1. Data and simulation are normalized to the same number of tagged e-jets after mistag removal:

$$\alpha = \frac{N^{e-tag} - N^{e-tag}_{mstg}}{MC^{e-tag} - MC^{e-tag}_{mstg}}$$

   N^{e-tag} and N^{e-tag}_{mstg} represent the number of events with tagged and mistagged e-jets in the data and MC^{e-tag} and MC^{e-tag}_{mstg} refer to the corresponding number of events in the simulation.

   Using the above expression the number of events before tagging with the e-jet

Data - Total of 55248 events						
	SECVTX	mistags	QCD tags	JPB tags	mistags	QCD tags
e-jet	8158	84.3		9123	335.3	
a-jet	3640	112.8	921.8	4584	517.6	1455.2
SECVTX tag on the a-jet						
e-jet	1126	23.8		1225	35.3	
JPB tag on the a-jet						
e-jet	1292	80.4		1402	101.1	
Simulation - Total of 16547 events						
e-jet	4549	0		5999	0	
a-jet	1832	7		2609	190	
SECVTX tag on the a-jet						
e-jet	550	1		743	1	
JPB tag on the a-jet						
e-jet	767	43		1011	69	

Table 4.8: Rates of events with a tagged electron or away jet observed in the low P_T electron data and in the simulation. QCD tags in the data refer to the number of positive tags expected from generic QCD jets and are estimated using the positive tagging matrices. The mistags are calculated using the mistag matrices described in Chapter 5. The mistags for the double tags are estimated adding the probabilities for mistagging the e-jet and the a-jet and subtracting the square of the two.

from a heavy flavor decay, N^e_{hf}, is given by:

$$N^e_{hf} = \frac{\alpha \cdot MC^e}{SF}$$

where MC^e is the total number of events in the simulation and SF is the tagging efficiency scale factor to be measured. Initially, it is assumed that $SF = 1$.

2. For the difference between the events with a tagged away jet the data and in the normalized simulation, the amount of tags in the a-jet due to non-heavy flavor e-jets is computed using the rate of positive QCD tags. The rate of positive QCD tags should therefore applied to the number of events with non-heavy flavor e-jets, N^e_{fake}, which using the previous step is given by:

$$N^e_{fake} = N^e - N^e_{hf}$$

where N^e is the total number of events in the data. However, in practice the amount of expected QCD tags is estimated by applying the positive tagging rate on all the events, N^{a-tag}_{QCD} and therefore the resulting amount of QCD tags shown in Table 4.8 need to be scaled to the number of events with non-heavy flavor e-jets:

$$N^{a-tag-QCD}_{e-fake} = N^{a-tag}_{QCD} \cdot \frac{N^e_{fake}}{N^e}$$

This amount of expected QCD tags is subtracted from the number of observed tags in the away jet to give the total number of events, N^{a-tag}_{hf}, with a tagged away jet associated with an e-jet from heavy flavor decay.

$$N^{a-tag}_{hf} = N^{a-tag} - N^{a-tag-QCD}_{e-fake}$$

3. The ratio of double tags N_{double} to N_{hf}^{a-tag} gives the tagging efficiencies in the data ϵ^d in the data and in the simulation ϵ^{MC}. The ratio of the two tagging efficiencies defines the scale factor:

$$SF = \frac{\epsilon^d}{\epsilon^{MC}}$$

Moreover the fraction, F_{hf}, of events with e-jet from heavy flavor decay is given by:

$$F_{hf} = \frac{N_{hf}^e}{N^e}$$

4. The derived scale factor is used to normalize data and simulation to the same number of tagged e-jets meaning that the iteration cycle is concluded and it starts again from step 1. A new scale factor is derived.

5. The iteration procedure is repeated until the scale factor is stable within 1%.

Sample	SF	F_{hf}
SECVTX e-jet, SECVTX a-jet	1.23±0.08	43.5±2.9 %
JPB e-jet, SECVTX a-jet	0.96±0.05	45.3±2.4 %
JPB e-jet, JPB a-jet	1.04±0.09	42.6±3.8 %
SECVTX e-jet, JPB a-jet	1.26±0.11	42.6±3.9 %

Table 4.9: Data to Monte Carlo tagging efficiency scale factor derived comparing the ratio of double tags to the number of tagged a-jets in the data and in the simulation. F_{hf} is the resulting fraction of e-jets due to heavy flavor in the data.

Using the rates of electron and away jets shown in Table the scale factor, SF, is 1.23±0.08 for SECVTX and 0.96±0.05 for Jetprobability when the away jet is also tagged by SECVTX. The error is due to sample statistics (with largest contribution

from the Monte Carlo) and added in quadrature with a 10% systematic error due to the mistag rate calculation. The results of this method are summarized in Table 4.9 for different combinations of tagged e-jet and a-jets.

Using the above procedure, the SECVTX and jetprobability tagging efficiencies per b-jet are also derived. The average SECVTX b-jet tagging efficiency for electron jets with $E_T \geq$15 GeV is found to be $\epsilon^{data}_{SECVTX} = 36.9 \pm 1.9\%$ in the data and $\epsilon^{data}_{SECVTX} = 30.1 \pm 1.1\%$ in the simulation. The corresponding data and Monet Carlo b-jet tagging efficiencies for e-jets tagged by jetprobability when the away-jet is tagged by SECVTX are $\epsilon^{data}_{JPB} = 39.2 \pm 2.1\%$ in the data and $\epsilon^{data}_{JPB} = 40.7 \pm 1.1\%$ in the simulation.

Figures 4.17 and 4.18 compare E_T distributions of the electrons, e-jets and SECVTX tagged e-jets and a-jets in single double tagged events in data and Monte Carlo. Data and simulation are normalized to the same number of tagged e-jets using the scale factor and heavy flavor fraction F_{hf} shown in Table 4.9. Kinematical distributions related to the tags in the data and simulation are shown in Figures 4.19 and 4.20. The comparison includes the distributions of pseudo-cτ, mass and transverse momentum of the secondary vertex as determined by the summed P_T of the tracks participating to the vertex, for SECVTX tagged e-jets and tagged e-jets when the a-way jet is also tagged by SECVTX. The distributions associated with the e-jet is chosen because of the small contribution of mistags whose shape can not be determined.

Comparison with other measurements of the scale factor

The b-purity of the e-jets before tagging is found to be 43.5$\pm$2.9 % in very good agreement with the results of the direct method described in Reference [53]. The scale factor derived in the above reference was 1.04$\pm$0.07 but the simulation did not

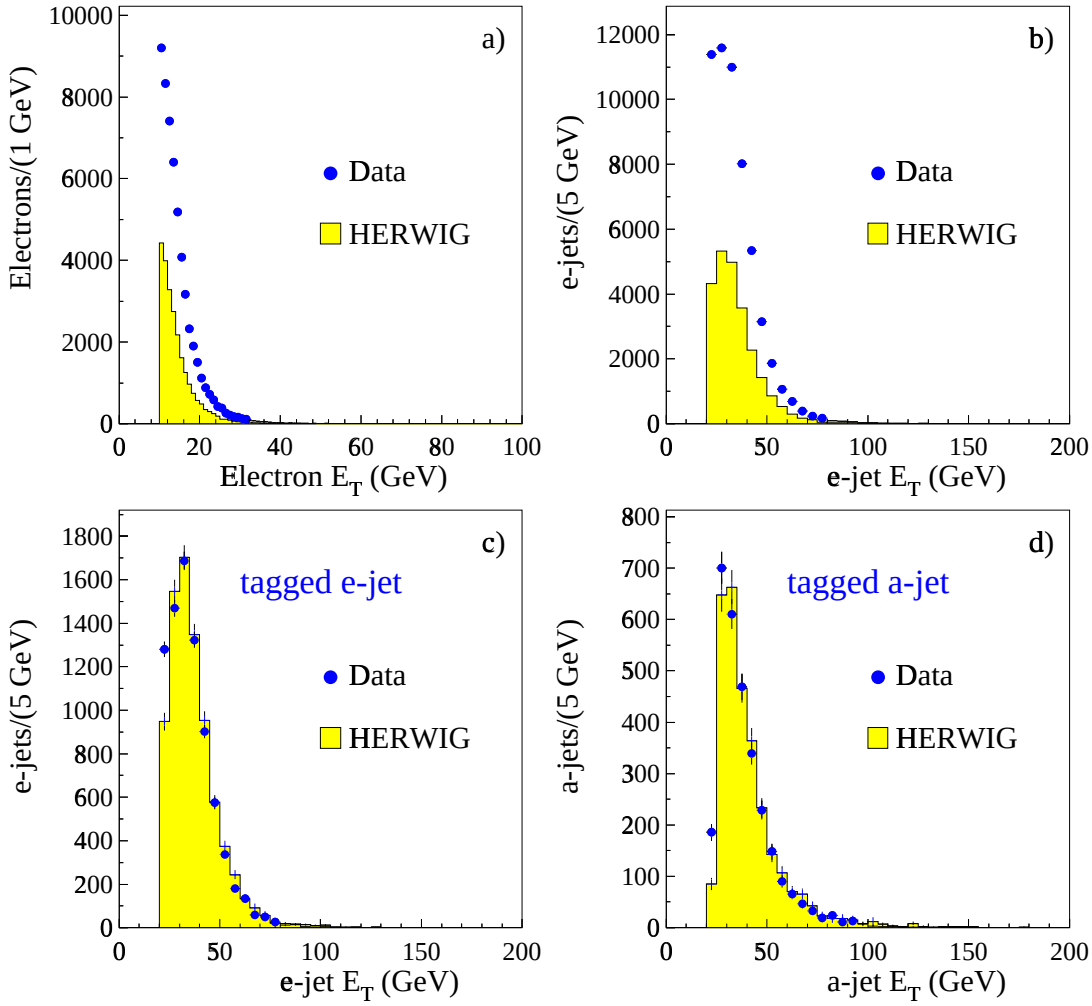

Figure 4.17: The distributions of transverse energy for the electron (a), e-jet (b), SECVTX tagged e-jet (c) and SECVTX tagged a-jet (d) in the low P_T inclusive electron data (•) and in the corresponding HERWIG simulation (shaded) used to derive the tagging efficiency scale factor. The simulation is normalized to same number of tagged e-jets in the data using the scale factor and F_{hf} in Table 4.9. The distributions shown in (c) and (d) are after subtracting mistags and the expected shape for the QCD tagged a-jets.

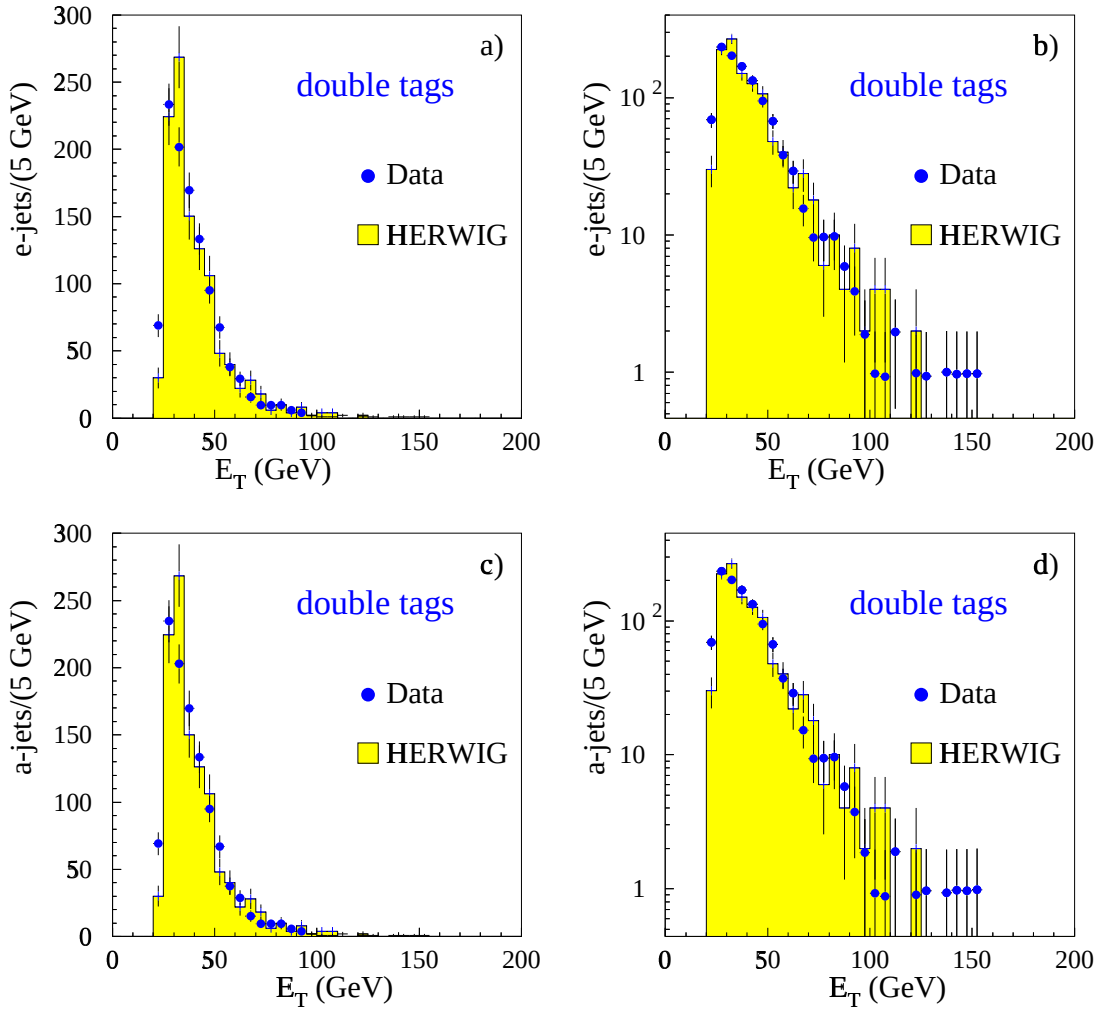

Figure 4.18: The distributions of transverse energy for tagged e-jets (a,b) and tagged a-jets (c,d) in data (•) and simulation (shaded) in events where both jets are tagged by SECVTX.

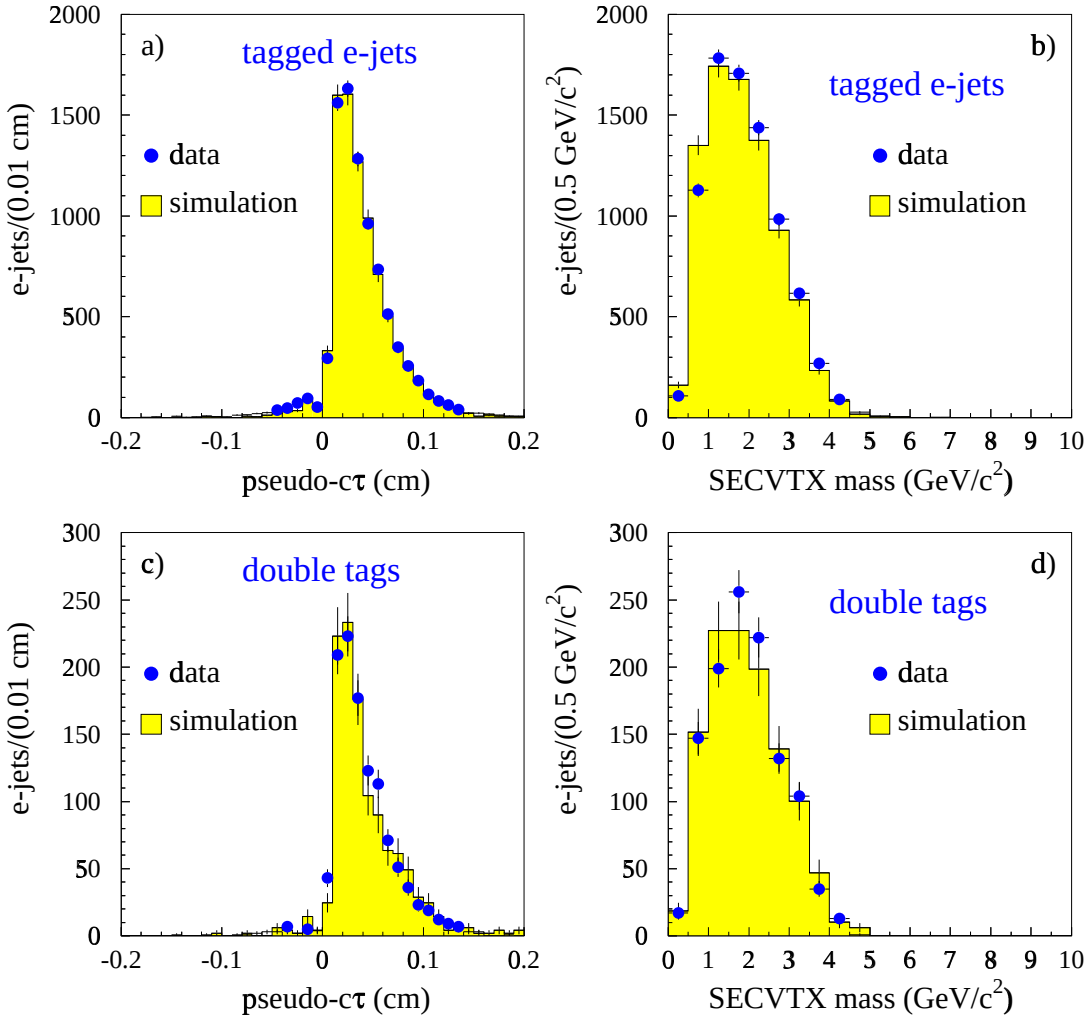

Figure 4.19: The distributions of pseudo-$c\tau$ and secondary vertex mass in the data ($\bullet$) and in the normalized HERWIG simulation (shaded) for SECVTX tagged e-jets (a,b) and tagged e-jets when the a-jet is also tagged by SECVTX (c,d).

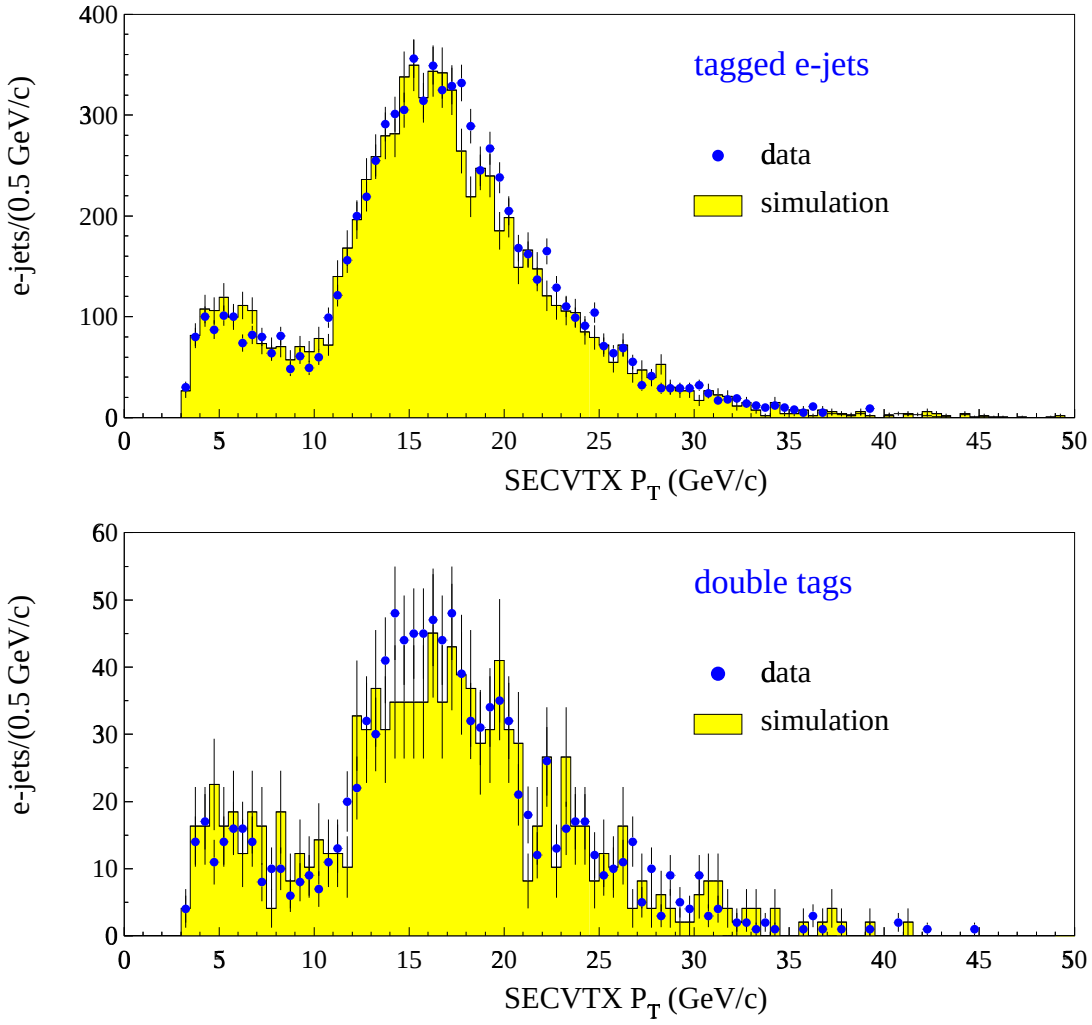

Figure 4.20: The distributions of the secondary vertex P_T in the data (•) and in the normalized HERWIG simulation (shaded) for SECVTX tagged e-jets (top) and tagged e-jets when the a-jet is also tagged by SECVTX (bottom).

include tracking degradation and it contained b's only from direct production and flavor excitation. The data sample used in that study was also smaller than the one used in the study presented above, covering $\sim$25 pb^{-1} of the run 1B data sample. The data sample has the tag composition shown in Table 4.10

Data - Total of 9658 events						
	SECVTX	mistags	QCD tags	JPB tags	mistags	QCD tags
e-jet	1459	13.7		1673	58.3	
a-jet	659	19.1	161.9	827	87.7	257.3
SECVTX tag on the a-jet						
e-jet	210	4.17		222	6.2	
JPB tag on the a-jet						
e-jet	237	13.9		249	18.6	

Table 4.10: Rates of events with a tagged electron or away jet observed in the low P_T inclusive electron data sample used in the measurement of the tagging efficiency scale factor in References [2, 52, 53].

To compare the obtained results to the ones of reference [53], the same data sample was used and the Monte Carlo was run without tracking degradation. Also, events with b's from direct production and flavor excitation are selected. Assuming a b-purity of 43%, a scale factor of 1.06$\pm$0.06 is obtained to be compared with the value of 1.04$\pm$0.07. Including all possible heavy flavor production mechanisms but still ignoring track degradation, the results become, $SF = 1.10 \pm 0.12$ and $F_{hf} = 44 \pm 5\%$. When including track degradation on the full simulation a scale factor of 1.25$\pm$0.13 and $F_{hf} = 43 \pm 5\%$ is obtained.

Since the published CDF $t\bar{t}$ cross section measurement is based on this smaller sample and since other b studies have been performed on this sample, it is adopted

as the default sample also for this analysis and the scale factor and b-purity is recalculated. The results are shown in Table 4.11

Sample	SF	F_{hf}
SECVTX e-jet, SECVTX a-jet	1.25±0.13	43±5 %
SECVTX e-jet, JPB a-jet	1.32±0.14	41±5 %
JPB e-jet, SECVTX a-jet ($\star$)	0.95±0.10	43±5 %
JPB e-jet, SECVTX a-jet	0.94±0.10	48±6 %
JPB e-jet, JPB a-jet	0.97±0.11	46±6 %

Table 4.11: Data to Monte Carlo tagging efficiency scale factor derived comparing the ratio of double tags to the number of tagged a-jets in the data and in the simulation. F_{hf} is the resulting fraction of e-jets due to heavy flavor in the data. ($\star$): This scale factor is derived using the F_{hf} value in the first line.

After these tests, the following conclusions are drawn:

- Using track degradation, the scale factor for SECVTX tags in b-jets is 1.25±0.13. The scale factor is referred to b-jets only since, as shown in Table 4.7, the full simulation includes a small fraction of events with c's.

- The heavy flavor content of the sample is 43±5% consistent with the value of 43±8% derived in References [52].

- The scale factor for jetprobability is 0.95±0.1 quite consistent with a value of 1. This value for the jetprobability scale factor is expected, since the simulation is self calibrated. As discussed in Section 4.3.3, separate SVX resolution functions are derived for data and simulation.

Test of the E_T dependence of the scale factor

In order to examine any E_T dependence of the scale factor, the iteration procedure is repeated as a function of the minimum e-jet E_T requirement. In this way, the purity of the sample, F_{hf}, is recalculated on each resulting sample and the scale factor is derived. Furthermore, any differences in the E_T spectrum of the jets between data and Monte Carlo are taken into account since data and simulation are normalized via the iteration procedure. This is important in order to make a real comparison between the tagging efficiencies in data and simulation without biasing the comparison on any differences in the jet E_T spectrum in the data and in the Monte Carlo. Actually, a small difference in the jet E_T spectrum in data and simulation is expected since HERWIG is a parton shower Monte Carlo starting from $2 \rightarrow 2$ process and tends to produce a slightly harder jet E_T spectrum than the one observed in the data.

The procedure is performed in steps of 5 GeV for the e-jet E_T, starting from 15 GeV up to 50 GeV. Due to limited statistics the procedure can be performed for higher values of e-jet E_T minimum.

Figure 4.21 shows the derived scale factors and sample b-purity as a function of the minimum e-jet E_T requirement. The procedure is performed separately for SECVTX and jetprobability tagged e-jets while the away jet is required to be tagged by SECVTX. The values of the first points in each plot correspond to the values shown in Table 4.9. Clearly, there is no E_T dependence observed on the scale factor. The decrease in the amount of real heavy flavor e-jets at higher energies is due to the increased background due to $\pi^0\pi^{\pm}$ overlap.

Fitting the minimum E_T distributions of the scale factors for SECVTX and jetprobability with first degree polynomials of the form $A_{SEC} + B \cdot E_T^{min}$ and $A_{JPB} + B \cdot E_T^{min}$ the following results are obtained:

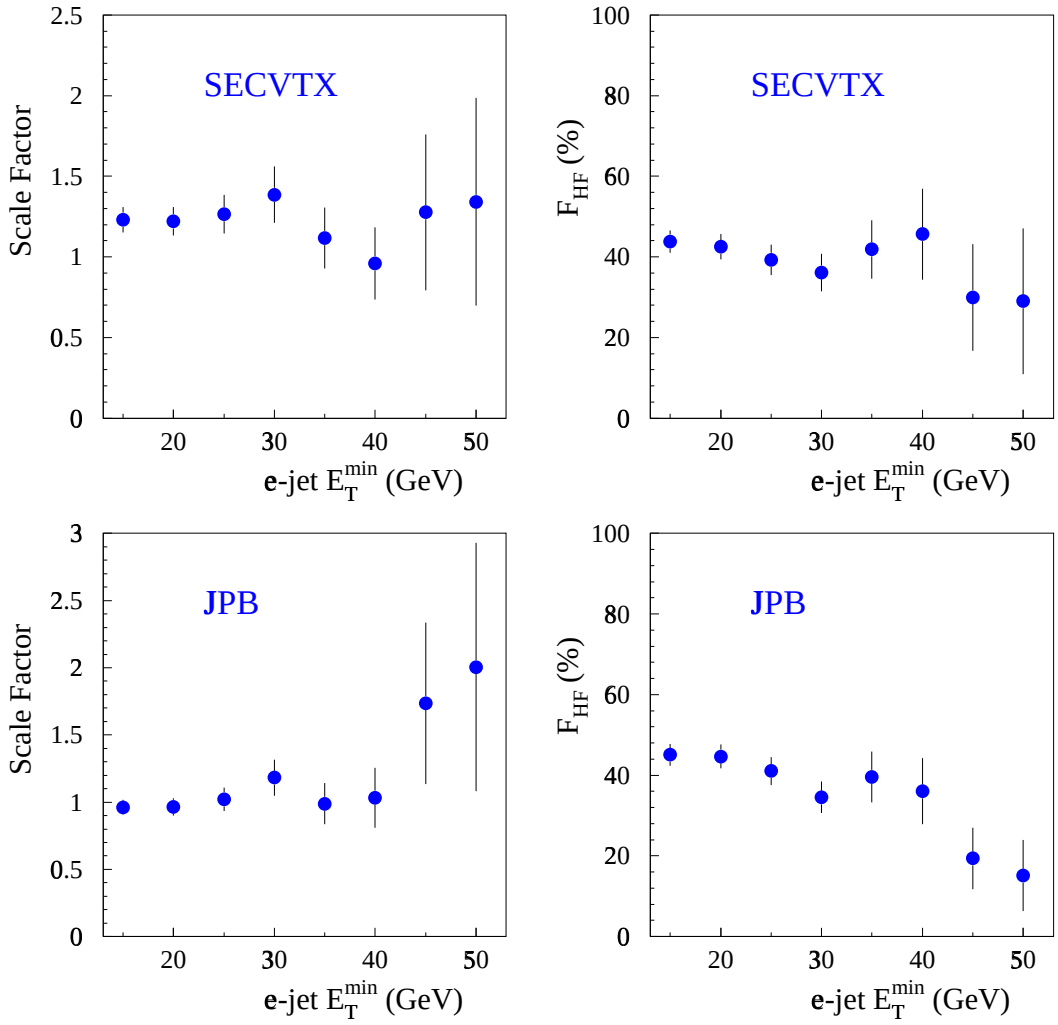

Figure 4.21: The SECVTX and Jetprobability tagging efficiency scale factors and F_{hf} in the inclusive electron sample as a function of the minimum transverse energy of the e-jet, in events where the a-jet is tagged by SECVTX.

$A_{SEC} = 1.15 \pm 0.10$, $A_{JPB} = 0.91 \pm 0.09$ and $B = 0.003 \pm 0.004$.

The χ^2 of the fit is 8 for 13 degrees of freedom. Once again it is clear that the E_T dependence of the scale factor is consistent with none.

4.6 Cross check of the tagging efficiency scale factor using jet data

Still one might worry that the low P_T electron sample can not be used to determine the tagging efficiency and data to Monte Carlo scale factor to be applied to b-jets from $t\bar{t}$ decays. The tagging efficiency measured in the low P_T electron sample refer to the semileptonically decaying jets and therefore there is always a track missing in the jet due to the escaping neutrino. Also the requirement for the presence of an electron, forces the presence of at least one reconstructed track in the jet. Furthermore, the statistics of the inclusive low P_T electron sample is not too high to examine the b-jet tagging efficiency in the region of $E_T^{b-jet} \geq 40$ GeV where most b's from $t\bar{t}$ decay are produced.

To address this concern, the inclusive jet samples collected with different jet E_T hardware trigger thresholds are used. The E_T spectrum of the jets in these samples span the region from 15 GeV up to 300 GeV and therefore offer a good sample to examine the tagging rates for a large spectrum of jet E_T.

For this study the jet samples collected with the JET_50 and JET_100 triggers are used. The events are selected in the following manner:

- one taggable jet with $E_T \geq$ trigger threshold
- em one and only one additional taggable jet with $E_T \geq 15$ GeV

- polar angle between the two jets $\delta\phi \geq 2.4$
- $|z_{svx}| \leq 60$ cm

The same selection criteria are also applied to the corresponding HERWIG generated jet samples (a description of the simulated QCD samples is given in Section 6.6.3). For each sample, the rates of events with double jetprobability, $N2_{JPB}$, and SECVTX tags, $N2_{SEC}$, are counted. The tagging rates in the data are shown in Table 4.12. The corresponding double tagging rates in the simulation along with the heavy flavor composition of the sample is shown in Table 4.13.

	JET_50		JET_100	
	tags	fakes	tags	fakes
SECVTX	129	21.8±2.2	123	40.7±4.1
JPB	228	114.4±11.5	257	165.3±16.5

Table 4.12: Rates of double jetprobability and SECVTX tags in the JET_50 and JET_100 data samples. Fakes account for fake · fake and real · fake and are calculated according to the mistag matrices discusssed in Chapter 5.

	JET50	JET100
	tags	tags
SECVTX	131 (116 *bb*, 11 *cc*, 4 *bc*)	76 (65 *bb*, 8 *cc*, 3 *bc*)
JPB	212 (170 *bb*, 34 *cc*, 8 *bc*)	130 (97 *bb*, 27 *cc*, 6 *bc*)

Table 4.13: Rates of double jetprobability and SECVTX tags in the JET_50 and JET_100 HERWIG simulated samples. The heavy flavor composition of the tags is also listed.

The ratio of double jetprobability tags to double SECVTX is formed in the data

and in the corresponding HERWIG simulations:

$$R_{data} = \frac{N2_{JPB}}{N2_{SEC}} = 1.08 \pm 0.21 \quad \text{and} \quad R_{sim} = \frac{N2^{sim}_{JPB}}{N2^{sim}_{SEC}} = 1.65 \pm 0.14$$

If the scale factors are uncorrelated for double tagged events, the ratio of the above two ratios can be written as:

$$R^2 = \frac{R_{data}}{R_{sim}} = (\frac{SF_{JPB}}{SF_{SEC}})^2$$

Using the tagging rates observed in the data and in the simulation as listed in Tables 4.12 and 4.13 a value of R=0.80$\pm$0.08 is obtained. This value of R can be compared to the value obtained with the low P_T inclusive electron sample, (see Table 4.11) of R=0.76$\pm$ 0.11. The agreement is pretty good implying that the scale factor obtain with the low P_T inclusive electron sample is valid also for hadronic b-decays and for jets of wider E_T spectrum. The method also implies that the scale factor is not affected by any correlation between the jets in the event.

Investigation for the presence of a scale factor for c-jets

The inclusive jet sample offer also a way to examine the presence of a tagging efficiency scale factor for c-quark jets. For this test the single tagging rates of Jetprobability and SECVTX are examined. The rates of single jetprobability and SECVTX tags in the data and in the corresponding simulation are given in Table 4.14.

The ratio of Jetprobability to SECVTX tag rates is used again and a value of R_{data}=1.32$\pm$0.13 is obtained using the rates shown in Table 4.14. The corresponding ratio in the simulated sample yields, R_{sim}=1.49$\pm$0.03.

In order to extract the SECVTX scale factor for c-jet, SF^c_{SEC}, the following ex-

Data				
	JET_50		JET_100	
	tags	fakes	tags	fakes
SECVTX	3334	730±73	3782	1239±130
JPB	6016	2472±247	6804	3528±353
Simulation				
	JET_50		JET_100	
	b	c	b	c
SECVTX	1771	812	805	409
JPB	2137	1753	984	787

Table 4.14: Rates of single jetprobability and SECVTX tags in the JET_50 and JET_100 data and HERWIG simulations.

pression can be used:

$$R_{data} = \frac{5661 \cdot SF_{JPB}}{2576 \cdot SF_{SEC} + 1221 \cdot SF^{c}_{SEC}}$$

where, R_{data}=1.32±0.13 and SF_{SEC}=1.25±0.13 as measured before. The assumption made in the above expression is that the jetprobability scale factor for c-jets is $SF^{c-jet}_{JPB} = 1 \pm 0.1$. This assumption is reasonable considering that the b-tagging efficiency scale factor for jetprobability is 0.95±0.1. Furthermore, in order to solve the above equation it is assumed that data and simulation have the same b and c composition.

With the above provisions and assumptions, the resulting SECVTX tagging efficiency scale factor for c-jets is $SF^{c}_{SEC} = 0.88 \pm 0.56$.

4.7 Soft Lepton (SLT) tag algorithm

The soft lepton tagging algorithm is intended to identify leptons from b or c semileptonic decays, $b \to c\ell\nu$ or $c \to s\ell\nu$. The fraction of $t\bar{t}$ events with a lepton (e or μ) from b or c semileptonic decays is quite large. The branching fractions for semileptonic decays of bottom and charm quark is approximately 10% per lepton species. Since there are two b quarks in each $t\bar{t}$ event, there is $\sim$60% probability to find a soft lepton candidate. Taking also into account the soft lepton contribution from events with a c-quark from $W \to c\bar{s}$ decays (this accounts practically for half of the hadronic W decays) one finds that on average there is one soft lepton candidate in every $t\bar{t}$ event.

However, leptons from heavy quark semileptonic decays are harder to identify compared to leptons from W/Z decays. The P_T spectrum of the leptons from heavy flavor quark semileptonic decays, shown in Figure 4.3, is much softer than the P_T spectrum of lepton from W/Z decays. In addition, these leptons are part of the hadronic jet formed by the fragmentation and decay products of the b-quark and hence they are expected to be found in the b-jet proximity. Consequently, the soft leptons are less isolated as opposed to leptons from W/Z decays. Figure 4.22(b) shows the $DR_{\ell j}$ distribution of soft leptons from b-semileptonic decays in $t\bar{t}$ Monte Carlo events. $DR_{\ell j}$ is defined as the distance in $\eta - \phi$ space between the soft lepton and the rest of the b-hadronic jet. In 95% of the cases the soft lepton is within $DR_{\ell j} \leq 0.4$ from the b-jet. The corresponding soft lepton isolation distribution in the same simulation is shown in Figure 4.22(b).

Similar to the SVX taggers, SLT is a track based algorithm. Recall that muons in

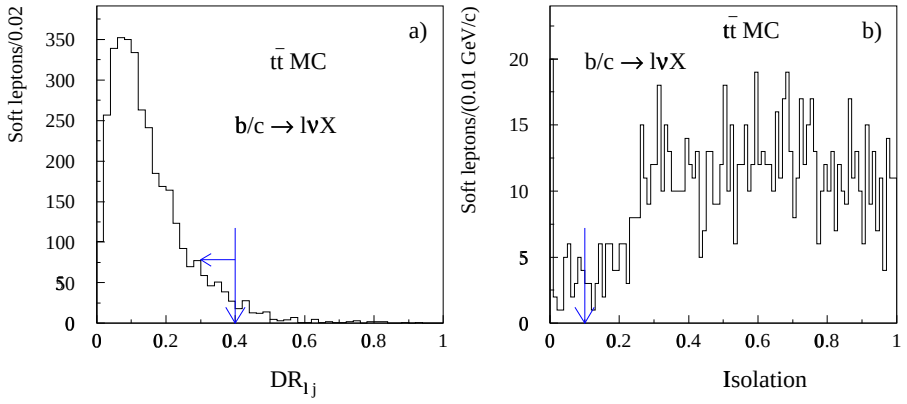

Figure 4.22: a): The distance, $DR_{\ell j}$, of a soft lepton from the rest of the b-hadronic jet. The soft lepton is within $DR_{\ell j} \leq 0.4$ from the b-jet 95% of the times. Also shown, the sof lepton isolation distribution. The arrow indicates the maximum value of the isolation requirement applied to the primary leptons. Both distributions are measured in $t\bar{t}$ PYTHIA Monte Carlo with top mass $m_t = 170$ GeV/c^2.

CDF are defined in track manner but electrons are defined by a combination of track and energy cluster in the calorimeter. The electron clustering algorithm has a lot of implicit isolation requirements making the efficiency of resolving soft electrons from semileptonic decays quite smaller than the corresponding for soft muons.

The SLT algorithm is described in great details in References [58, 59]. The identification criteria used and the corresponding efficiencies are presented briefly in the following sections.

4.7.1 Description of the algorithm

The algorithm considers only tracks with transverse momentum greater than $P_T \geq 2$ GeV/c, to ensure large rejection of tracks from generic jets which have a much softer P_T spectrum. This minimum P_T requirement is also imposed because muons tracks

of $P_T \leq 1.4(2.0)$ GeV/c do not reach the CMU(CMP) detector [36].

All tracks are required to have CTC track segments in at least 2 axial and 2 stereo superlayers to ensure good 3-D reconstruction. The track impact parameter, D_0, is also required to be $D_0 \leq 3$mm in order to reject a large fraction of cosmic rays and also muons tracks from $\pi^{\pm}$'s and K's decays in flight.

All surviving tracks are required to extrapolate to the fiducial regions of the central electromagnetic shower detector (CES) and preradiator detector (CPR) to be considered as soft electron candidates or to fiducial regions of the muon chambers to be considered as soft muon candidates. The search for soft lepton tracks is restricted in the η region of $\eta \leq 1$ because of the limited detector muon coverage, and tracking inefficiencies for $1 \leq |\eta| \leq 1.6$ which makes difficult the identification of electrons in the plug region.

Once a fiducial track is found it is declared as a soft lepton candidate and it is further tested for consistency with the definition requirements of a soft electron or muon.

The soft lepton requirements were studied with low P_T electrons from photon conversions and muons from J/ψ decays. The goal in designing the selection requirements was to achieve high efficiency for the identification of b electrons while maintaining the rate of "fake" leptons from tracks in generic QCD jets at less than 1% per track.

4.7.2 Soft electron identification

The track is required to extapolate to the central electromagnetic shower detector region (CES) which is covered by the CTC and the central preradiator detector (CPR). The CES and CPR fiducial regions are shown in Table 4.15. Also shown, is a list

of all the selection requirements used in the SLT algorithm to identify soft electrons along with the efficiency of each requirement alone.

Once an electron fiducial track is found, the algorithm tries to associate the track to a CES energy cluster. If a match is found, the matched cluster is required to be consistent in size and shape with expectations for electron showers. The quality of the CES shower profile is determined by the value of the χ^2 (χ^2_{strip} and χ^2_{wire}) obtained comparing the observed shower shape to the one measured for electrons from testbeam data. The energy weighted mean of three wires and three strips of the CES, gives the position of the CES cluster in the $r-\phi$ and $r-z$ plane. It is further required, the distance of the cluster position from the track extrapolated position (δx, δz) to be consistent with expectations for electrons from photon conversions. The electron shower is fully contained within 5 strips and 5 wires of the CES. The energy of the CES cluster is measured by summing up the energy of five strips (wires) around the extrapolated position of the track candidate. Figure 4.23(a) and (b) shows the distributions of E^5_{wire}/P and E^5_{strip}/P, defined as the ratio of the energy measured in the wires (strips) of the CES to the momentum, P, of the track matched to the CES cluster, for electrons from photon conversions and tracks in generic jets. The electron distributions exhibit the expected peak at 1 while the distributions are quite broad reflecting the limited energy resolution of the CES detector.

In order to further discriminate real electrons from fakes, the algorithm uses the charge information, Q_{CPR}, from the central preradiator detector (CPR) and the ionization rate as measured from the charge associated with the CTC hits, Q_{CTC}, used in the track in order to place some additional quality requirements on the candidate track. Figure 4.23(c) and (d) show the distributions of these two quantities as measured for electrons from photon conversions and track candidates in QCD jets.

Soft electron selection requirements		
Fiduciality		
Local CES radial position $\lvert X_{CES}\rvert \leq 22cm$		
Local CES Z position $6.22cm \leq \lvert Z_{CES}\rvert \leq 237.45cm$		
Exclude region of solenoid cryogenic connections		
Local CPR radial position $\lvert X_{CPR}\rvert \leq 17.78cm$		
Local CPR Z position $9.0cm \leq \lvert Z_{CPR}\rvert \leq 118.0cm$		
Local CPR Z position $125.0cm \leq \lvert Z_{CPR}\rvert \leq 235.26cm$		
Isolation independent requirements	$\epsilon_{\mathbf{ele}}(\%)$	$\epsilon_{\mathbf{jet}}^{\mathbf{track}}(\%)$
CES selection requirements		
$E_s^5/P \geq \begin{cases} 0.24+0.03\cdot P & P<12 \text{ GeV/c} \\ 0.6-(P-12)\cdot 0.0125 & 12<P<20 \text{ GeV/c} \\ 0.5 & P>20 \text{ GeV/c} \end{cases}$	89.7±0.7	24
$E_w^5/P \geq \begin{cases} 0.24+0.03\cdot P & P<12 \text{ GeV/c} \\ 0.6-(P-12)\cdot 0.0125 & 12<P<20 \text{ GeV/c} \\ 0.5 & P>20 \text{ GeV/c} \end{cases}$	87.5±0.7	23
$\lvert\Delta x\rvert \leq Max(0.7cm, 1.82cm - 0.1867P)$	92.7±0.6	59
$\lvert\Delta z\rvert \leq 2$ cm	94.8±0.5	47
$\chi^2_{wire}/6 \leq 16$	98.2±0.3	90
$\chi^2_{strip}/6 \leq 16$	98.9±0.2	92
Other requirements		
$Q_{CPR} \geq 4744 - 11592(P/P_T) + 7923(P/P_T)^2$	84.0±0.8	53
$Q_{CTC} \geq 29.15 + e^{1.671-0.08P}$ for $P \leq 15$ GeV/c	85.4±0.8	51
Isolation dependent requirements		
$0.7 \leq E/P \leq 1.5$	62±1	22
$E_{had}/E_{em} < 0.1$	46±1	23

Table 4.15: Soft electron selection requirements and the corresponding efficiencies. P is the track momentum in GeV/c. The CPR and CES coordinates have $z = 0$ at η=0 and $x = 0$ at the center of the calorimeter wedge. ϵ_{ele} corresponds to the identification efficiency as determined from conversion electron data requiring $P_T \leq 5GeV$. ϵ_{jet}^{track} is the selection efficiency for fiducial tracks in QCD jets. Errors in ϵ_{ele} are statistical only while the statistical errors in ϵ_{jet}^{track} are less than 1%. The efficiencies for the E/P and E_{had}/E_{em} requirements are determined from b-jet electrons in $t\bar{t}$ HERWIG Monte Carlo simulations with $M_{top} = 175$ GeV/c^2. The efficiency is for each selection requirement alone [58], while many requirements are correlated.

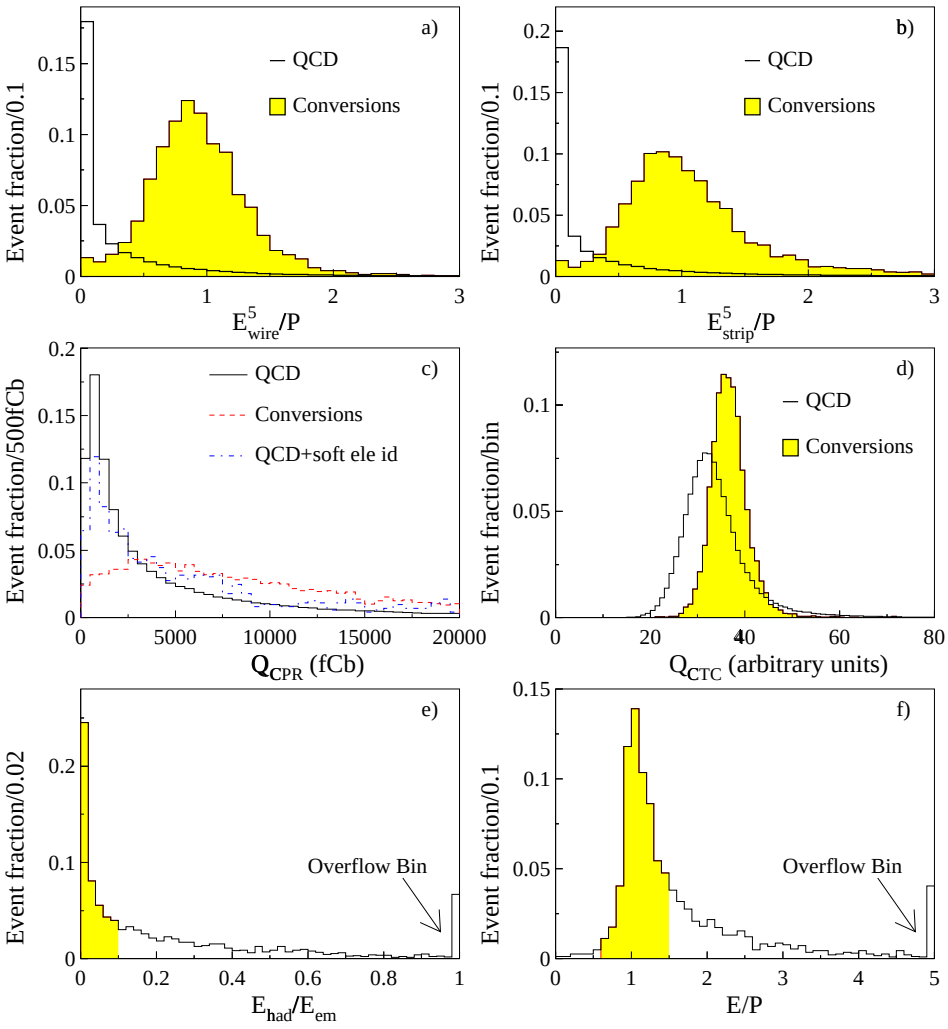

Figure 4.23: Distributions of the variables used to identify soft electrons. The distributions (a) to (d) correspond to electrons from photon conversion data sample and compared to the corresponding distributions of tracks in jet data. The distributions (e) and (f) correspond to electrons from heavy flavor jets in HERWIG $t\bar{t}$ simulations with $M_{top} = 175$ GeV/c^2.

After the CES, CPR and DE/dx requirements the electron fake rate is of the order of 2.5% per track. To further suppress the contribution of fake electrons the information on the E_{had}/E_{em} and E/P is used. These two variables are similar to the ones used in the selection of the electrons from W/Z decays (see Section 3.2.1). However, in order to optimize the detection of non-isolated electrons the electromagnetic clustering algorithm is modified relative to the one used in the reconstruction of the primary electrons. Recall, the electron clustering uses by default three towers in η and one tower in ϕ to reconstruct the electron cluster. Instead, the SLT algorithm reconstructs an electron cluster using only the energy of the tower the track candidate extrapolates to. In the case the track is within 2 cm of two tower boundaries in η, both towers are used in the clustering. The E_{had}/E_{em} quantity is defined as the ratio of the energy in the hadronic compartment of the tower or towers in the cluster to the energy in the electromagnetic compartment of the towers. Similarly, E/P is defined as the ratio of the electromagnetic energy of the electron cluster (one or two towers) to the momentum of the track. The efficiency of the selection on the two variables is determined using Monte Carlo samples since the nearby activity from the remaining of the heavy flavor hadron jet affects directly these variables. Figure 4.23(e) and (f) shows the distribution of E_{had}/E_{em} and E/P as measured for soft electrons from b's in HERWIG $t\bar{t}$ Monte Carlo events with $M_{top} = 175$ GeV/c^2. The shaded areas indicate the selection requirements applied to these quantities.

4.7.3 Soft electron identification efficiency

The efficiency of each selection requirement applied independently of the other requirements is shown in Table 4.15. For the isolation independent variables, the efficiency is determined using electrons from photon conversion data while the isola-

tion dependent variables (E/P and E_{had}/E_{em}) are determined from Monte Carlo. Figures 4.24(a) to (d) show the efficiency of the isolation independent variables as a function of the soft electron track P_T. In order to verify that the efficiency of these variables is indeed independent of the isolation, the combined efficiency is plotted as a function of the $\sum_P^{0.2}$ defined as the scalar sum of the momenta of all tracks, in a cone of radius $\Delta R = 0.2$ around the soft electron candidate. Figure 4.24(e) shows the combined efficiency of the CES, CPR and dE/dx requirements as a function of the $\sum_P^{0.2}$ after correcting for the P_T dependence shown in Figures 4.24(a) to (d). The efficiency distribution is flat within statistical uncertainties. As described, the efficiency of the E/P and E_{had}/E_{em} requirements are measured in Monte Carlo events. It is verified [58] using electrons from photon conversions data that the Monte Carlo models these variables quite satisfactory. Figure 4.24(f) shows a comparison between the efficiency of the E_{had}/E_{em} requirement as measured for low P_T electrons from photon conversion data and for electrons from b's in $t\bar{t}$ Monte Carlo events. Within the systematic uncertainties the agreement between the two distributions is fairly good.

The efficiency of E_{had}/E_{em} is determined to be $56\pm2\%$ for electrons from b's and $35\pm2\%$ for electron from c's [58]. The corresponding E/P selection efficiencies are $72\pm1\%$ and $48\pm2\%$ for b and c jets respectively. Despite the significantly lower efficiency of these two requirements compared to the other identification requirements, a factor of 8 reduction in the fake rate is achieved justifying the use of these requirements.

After all soft electron identification requirements applied, the rate of fake electrons from QCD jets is reduced to $\sim$0.3% per track [58].

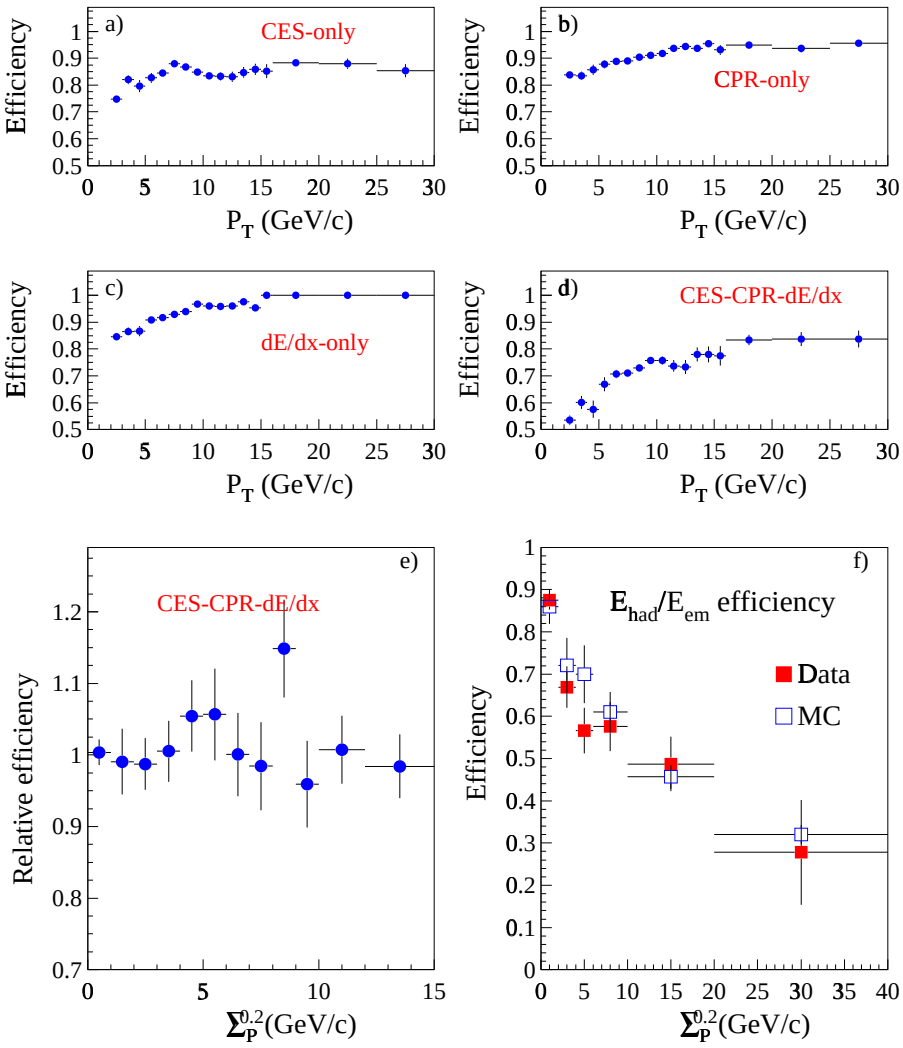

Figure 4.24: Efficiency distributions of the soft electron selection requirements as a function of the P_T of the candidate track. The efficiencies shown in (a) to (c) are derived by examining the shown requirement alone. The efficiency shown in (d) corresponds to the efficiency of the CES, CPR and dE/dx requirements combined. The combined efficiency of the CES, CPR and dE/dx requirements as a function the $\sum_P^{0.2}$. is shown in (e). Plot (f) shows a comparison of the efficiency of the E_{had}/E_{em} requirement as measured for electrons from photon conversion data (boxes) and for b electrons (open boxes) in Monte Carlo events.

4.7.4 Soft muon identification

To identify muons from b and c-hadron decays, track segments reconstructed in the muon chambers are matched to CTC tracks which are required to extrapolate within the boundaries of the corresponding muons detectors. The track extrapolation procedure calculates the distance of the track from the closest edge of each detector and the expected deflection due to multiple scattering. As for primary muons, the soft muons are also classified in four categories according to the detector the muon stub is found. The CMUP type includes muons with reconstructed stubs in both the CMU and CMP detectors which are well matched to a CTC track which extrapolates within good fiducial volume of the detector. The track is required to have $P_T \geq 3$ GeV/c because muons of lower P_T cannot reach the CMP detector. Selecting tracks with $P_T \geq 3$ GeV/c avoids also the CMUP turn-on region which occurs at 2.8 GeV/c. The CMP-only type includes muons that have a reconstructed stub in the CMP detector and matched to a CTC track of $P_T \geq 3$ GeV/c. This category includes muons which cannot be classified as CMUP because either they miss completely the CMU detector or because despite the presence of the CMU stub, the CTC track does not extrapolate within the good fiducial volume of the CMU detector. Similar considerations hold for the CMU-only type of soft muons. This category includes muons with reconstructed stubs in the CMU detector which are matched to CTC tracks of $P_T \geq 2$ GeV/c. The last category includes muons with reconstructed stubs in the CMX detector and matched to a CTC track which extrapolates within the boundaries of the CMX detector.

Different track-stub matching criteria are applied according to the muon type. The different criteria are necessary because the amount of material transversed by muon tracks is different for the different muon types. Since multiple scattering depends on

CMUP muons	
Number of CMU TDC hits ≥ 3	
Number of CMU ADC hits ≥ 3	
$\lvert\Delta z\rvert_{CMU} < MAX(3\sigma$, 8 cm)	
$\lvert\Delta x\rvert_{CMU} < MAX(3\sigma$, 2 cm)	
$E_{had} < \sum_P^{0.2}$ +6 GeV	if $P_T > 6$ GeV/c
CMU-only muons	
Number of CMU TDC hits ≥ 3	
Number of CMU ADC hits ≥ 3	
$\lvert\Delta z\rvert_{CMU} < MAX(3\sigma$,8 cm)	
CMCLUS ≤ 5	
$\chi^2_{\delta x \delta \phi\ CMU} < 10$	if $P_T < 20$ GeV/c
$\lvert\Delta x\rvert_{CMU} < MAX(3\sigma, 2cm)$	if $P_T \geq 20$ GeV/c
$E_{had} < \sum_P^{0.2}$ +6 GeV	if $P_T > 6$ GeV/c
CMP-only muons	
$\chi^2_{\delta x \delta \phi\ CMP} < 10$	if $P_T < 10$ GeV/c
$\lvert\Delta x\rvert_{CMP} < MAX(3\sigma, 5cm)$	if $P_T \geq 10$ GeV/c
$\lvert\Delta \phi\rvert_{CMP} < 0.1$	if $P_T \geq 10$ GeV/c
$E_{had} < \sum_P^{0.2}$ +6 GeV	if $P_T > 6$ GeV/c
CMX muons	
$\chi^2{}_{\Delta z\ CMX} < 9$	
$\chi^2{}_{\Delta x\ CMX} < 9$	
$\chi^2_{\delta x \delta \phi\ CMX} < 9$	if $P_T \leq 5$ GeV/c
$\lvert\Delta \phi\rvert_{CMX} < 0.1$	if $P_T > 5$ GeV/c

Table 4.16: Soft muon selection requirements. σ represents the expected CTC track-muon stub mismatch as calculated by adding in quadrature the resolution of the corresponding muon chamber and the expected deflection due to multiple scattering.

the amount of matterial, different track-stub mismatch is expected for muons of the same P_T but different detector type. The mismatch is larger for CMX muons because they travel long distance exiting the calorimeter and before hitting the muon chambers. The next larger mismatch corresponds to CMP muons because they transverse more matterial while the smaller mismatch is expected for CMU muons which have the less shielding. However, the angular deflection a muon experiences due to multiple scattering falls as $1/P_T$ and the expected mismatch reduces for high P_T muons.

Muon stubs reconstructed in the CMU provide information for both x and ϕ position of the muon with a resolution of 250 μm and for the z position with a resolution of 1.2 mm. Muon stubs in the CMP are two dimensional and can only provide $x - \phi$ information. Stubs in the CMX detector are three dimensional yielding information in x, z and ϕ directions with poor resolution ($\sim$2 cm) in the z-position measurement.

Therefore two effects contribute to track-stub mismatch. One is due to detector resolution and the other is due to the multiple scattering. The calculated mismatch includes the contribution of each effect added in quadrature. Based on the expected mismatch, χ^2 quantities are constructed for each matching variable and also for the correlations between them. The matching requirements applied to each muon type are listed in Table 4.16.

To maintain high efficiency for non-isolated muons, the minimum ionizing requirements used to identify muons from W or Z decays (see Section 3.3.1) are relaxed. Instead a modified minimum ionizing requirement is imposed on muon candidates with $P_T \geq 6$ GeV/c. The requirement is intended to reduce the background from hadronic punch-through events. It is required that the energy, E_{had}, measured in the hadronic compartment of the tower transversed by the muon to satisfy the following

relation: $E_{had} < \sum_P^{0.2} + 6GeV$. $\sum_P^{0.2}$ is defined as the scalar sum of the momenta of all tracks in a cone $\Delta R = 0.2$ around the muon candidate. For muons with P_T below 6 GeV/c, the requirement does not apply because the calorimeter information is not as effective at rejecting background events from hadron punch-through. This minimum ionizing requirement is measured in $t\bar{t}$ Monte Carlo events to be more than 98% efficient for muons from b and c decays.

Some additional quality requirements are also applied. Minimum numbers of ADC and TDC hits associated with a CMU stub are required to ensure good quality stub. The requirement on the maximum number of muon chambers, CMCLUS, associated with at least one TDC hit and are adjucent to the ones used in the stub reconstruction is imposed in order to reduce background from punch-through events in the case of CMU-only muons.

As shown in Figure 5.21, after the application of the soft muon selection requirements, the muon fake rate reduces to approximately 1% per track candidate. In contrast the efficiency of the soft muon selection requirements, discussed below, is ~90% efficient for real muon tracks.

4.7.5 Soft muon selection efficiency

The efficiency of the soft muon requirements is measured with muons from J/ψ and $Z \rightarrow \mu\mu$ data. The first sample is used to determine the efficiency for low P_T muons while the second sample is used for high P_T soft muon candidates. The efficiency of the matching requirements as a function of the track candidate P_T is shown in Figure 4.25 Another efficiency under consideration is the muon stub reconstruction efficiency. The stub reconstruction efficiency for each muon type is measured using $Z \rightarrow \mu\mu$ and $J/\psi \rightarrow \mu\mu$ data. It is found that the stub reconstruction efficiency is

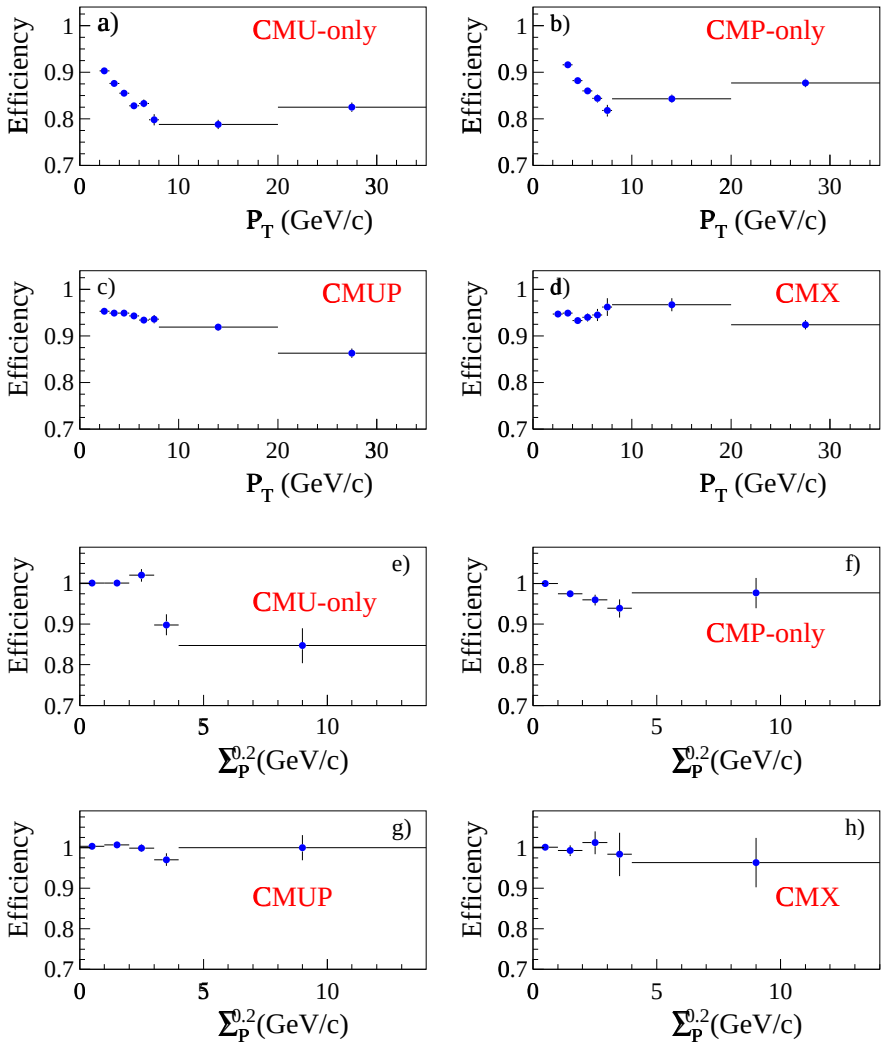

Figure 4.25: Efficiency distributions of the soft muon matching requirements as a function of the P_T of the candidate track for CMUP (a) CMU-only (b), CMP-only (c) and CMX (d) muons. The same efficiencies are plotted in (e) to (h) as a function of the isolation quantity $\sum_P^{0.2}$. For these efficiencies the P_T dependence shown in plots (a) to (d) has been removed. The efficiencies are obtained from J/ψ data after background subtraction using the mass sidebands and from $Z \rightarrow \mu\mu$ data.

98.3±0.3% for CMU, 99.8±0.3% for CMP and 98.9±1% for CMX stubs respectively.

4.7.6 Implementation of the SLT algorithm in the Monte Carlo

The meaured soft lepton identification efficiencies which are parametrized as a function of the soft lepton P_T (see Figures 4.24 and 4.25) are implemented in the Monte Carlo simulations. In contrast to the SVX based taggers, soft lepton candidates in the Monte Carlo simulation are required to be associated with leptons from b or c semileptonic decays, or leptons from W, Z and τ decays. Therefore, soft leptons in the Monte Carlo are always real leptons. Once a track associated with a lepton at generator level is found, it is propagated in the detector and extrapolated in the relevant detector regions and it is checked for fiduciality following the same procedure as in the data. According to the soft lepton type, the relevant selection efficiency parametrization is applied on the candidate track in order to select the final soft lepton candidates. For the case of soft electron the efficiency of the E/P and E_{had}/E_{em} requirements are determined from Monte Carlo and therefore the candidate electron track in the Monte Carlo is required to satisfy these two requirements explicitly.

In order to account for the reduced CTC tracking reconstruction efficiency for tracks inside jets, the track degradation procedure desribed in Section 4.4 is also implemented in the Monte Carlo modeling of the algorithm.

4.7.7 Definition of SLT tags

Since leptons from heavy flavor semileptonic decays are contained within a cone of 0.4 around the heavy flavor jet direction, the definition of soft lepton tags is restricted

only to soft lepton candidates contained within a cone of 0.4 around the axis of a jet with $E_T \geq 15$ GeV. A jet is assumed to be tagged by the soft lepton algorithm if it contains at least one track with $P_T \geq 2$ GeV/c identified as a soft lepton tag by the SLT algorithm. The adopted definition of soft lepton tags is different with respect to the CDF analyses presented in References [59, 58, 23] where any track identified as a soft lepton is considered as an SLT tag.

Figure 4.26 show the invariant mass distribution of primary and soft leptons for all $W+ \geq 1$ jet events with SLT tags for different lepton types and charges. Since the $W+ \geq 1$ jet sample is selected after Z and removal of all dilepton pairs, it is expected that the contribution of primary - soft lepton pairs of the same flavor due to Z or Drell-Yan production to be minimum. In addition, since soft leptons are searched for inside jets, the contribution from Z and Drell-Yan production is even smaller, since leptons in these processes tend to be isolated. As shown in Figure 4.26 there is a little hint of SLT tags in the ee pair from Drell-Yan or Z production

However, there is a handful of events where the soft muon is consistent with being the second leg of a Z embedded in a jet. As such, these events are removed from the sample. It is further required for the definition of a soft muon tag that the soft muon does not form an invariant mass in the mass window of $70 \leq M_{\mu\mu} \leq 110$ GeV/c^2 with a primary muon of opposite charge. On the other hand the muon pairs from Drell-Yan and Υ production is only a very small fraction of muons from Z production and therefore are not removed.

According to the SLT analysis presented in Reference [58], soft lepton tags are looked for anywhere in the event. In order to reduce the contribution of pairs due to Drell-Yan production, events with isolated soft leptons of the same flavor and opposite charge to the primary lepton in th event are removed. However, in the present analysis

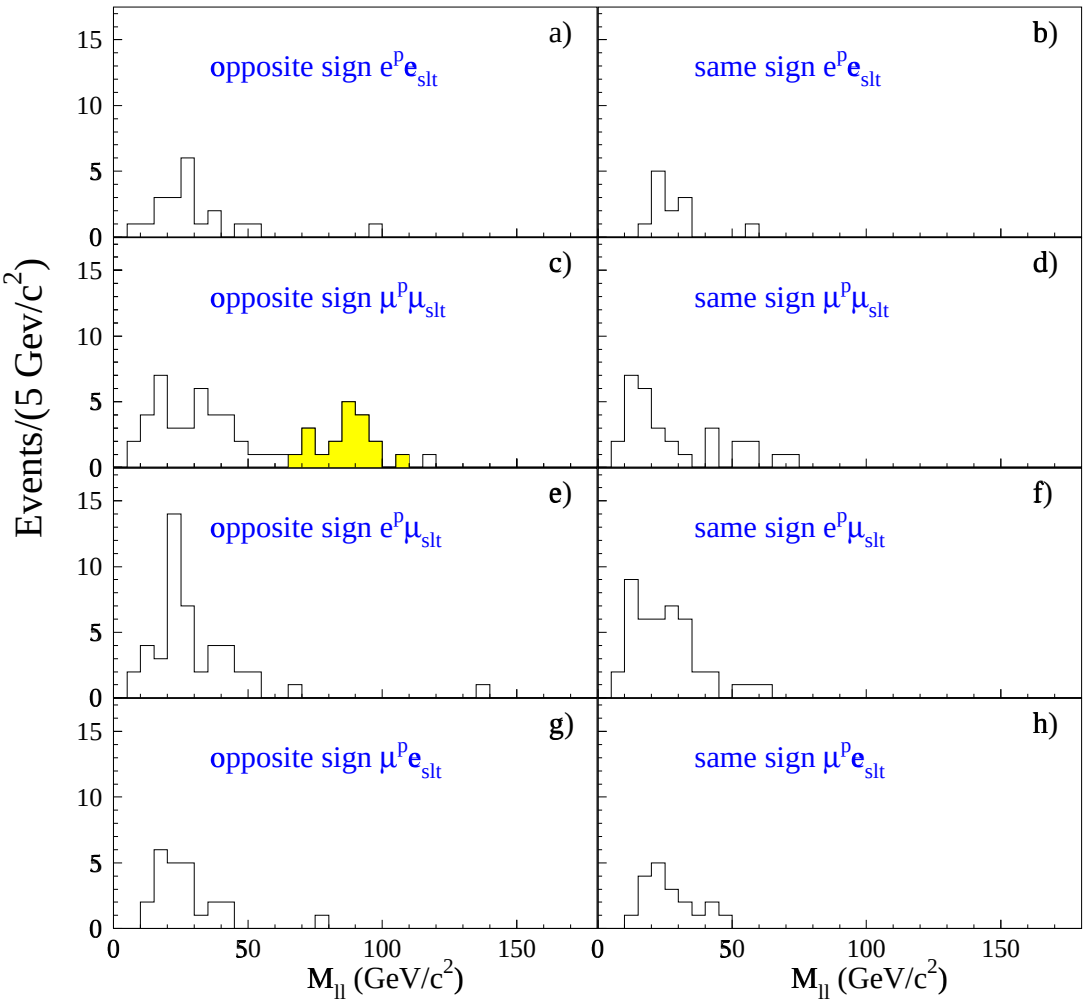

Figure 4.26: Invariant mass distributions of the primary and soft leptons in $W+\geq 1$ events. An excess of events in the Z mass window of $70 \leq M_{\mu\mu} \leq 110$ for opposite sign primary muon and soft muon is seen. These events are removed from the sample.

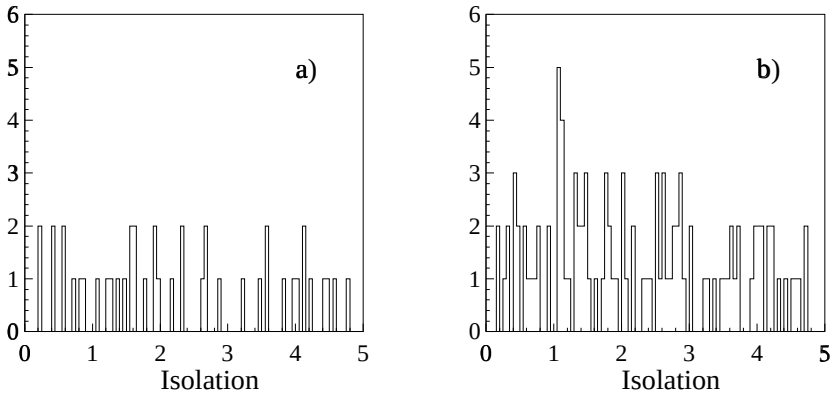

Figure 4.27: Calorimetric isolation distribution of soft leptons in the $W+\geq 1$ jet sample. (a): events where the primary and soft leptons are of the same flavor and opposite charge. (b): the remaining events.

this can never happen since the soft leptons are always inside jets and therefore experience poor isolation. Figure 4.27 show the calorimetric isolation distribution of soft leptons in $W+\geq 1$ jet events for events where the primary and soft lepton are of the same flavor and opposite charge (a) and for all other combinations. Both distributions are very flat over the entire range of the distribution and no apparent cluster of events is observed in low isolation.

4.8 Application of the tagging algorithms on the $W+\geq$1 jet sample

This section summarizes the tagging yields obtained applying the three tagging algorithms on the selected $W+\geq$1 jet sample.

As described in the previous sections, tags are associated with jets with raw transverse momentum $E_T \geq$15 GeV and $|\eta| \leq$2.0. A jet is considered tagged by SECVTX if the reconstructed secondary vertex has positive transverse decay length, $L_{xy} \leq$2.5 cm and significance $L_{xy}/\sigma_{L_{xy}} \geq$3.0. A jet is tagged by jetprobability if it contains at least two good jet probability tracks with positive signed impact parameter significance and jetprobability $\leq$0.05. A SLT tag is required to have $P_T \geq$2 GeV/c and to lie within a cone of $DR \leq$0.4 around the jet centroid. For events with a primary muon and a SLT muon tag of opposite sign, the invariant mass of the muon pair is required to be outside the mass window $70 \leq M_{\mu^-\mu^+} \leq 110$ GeV/c^2.

The number of events with one (single tag) or two tagged jets (double tags) are counted independently for each tagging algorithm and are summarized in Table 4.17 as a function of jet multiplicity.

	$W+n$ jet multiplicity						
	$W+1$ jet	$W+2$ jets		$W+3$ jets		$W+\geq 4$ jets	
All events	9460	1373		198		54	
Tagger	single	single	double	single	double	single	double
SECVTX	66	35	5	10	6	11	2
Jetprobability	125	62	6	21	5	12	3
SLT	146	56	–	17	–	8	–

Table 4.17: Yields of $W+\geq$1 jet events tagged by SECVTX, Jetprobability and SLT as a function of jet multiplicity.

Events tagged by one algorithm can also be tagged by one of the other two algorithms or both. In such case either the same jet is tagged, *multitag*, or the tag is found on a different jet in the event. The breakdown of the multitagged events is shown in Tables 4.18 to 4.20 for each jet multiplicity and for each tagger separately. There are no SLT double tagged events.

	Events with SECVTX tags			
	$W+1$ jets			
	Single tags		Double tags	
	Common	Different	1 common	2 common
Total	66		–	
JPB	51	–	–	–
SLT	1	–	–	–
JPB⊕SLT	1	–	–	–
	$W+2$ jets			
Total	35		5	
JPB	26	–	1	4
SLT	6	–	–	–
JPB⊕SLT	4	–	2	–
	$W+3$ jets			
Total	10		6	
JPB	5	1	1	5
SLT	2	–	3	–
JPB⊕SLT	1	–	3	–
	$W+\geq 4$ jets			
Total	11		2	
JPB	7	2	1	1
SLT	2	1	–	–
JPB⊕SLT	2	–	–	–

Table 4.18: Breakdown of the $W+n$ jet SECVTX tagged events, according to the number of jets tagged by the other two taggers. The column marked *common*, contains events in which the same jet is tagged by more than one taggers, while the column marked *different* contains events tagged by the other taggers but the tags are on different jets. The symbol $\oplus$ indicates that the event contains a jet tagged by both algorithms.

	Events with Jetprobability tags			
	$W+1$ jets			
	Single tags		Double tags	
	Common	Different	1 common	2 common
Total	125		–	
SECVTX	51	–	–	–
SLT	3	–	–	–
SECVTX⊕SLT	1	–	–	–
	$W+2$ jets			
Total	62		6	
SECVTX	26	–	1	4
SLT	7	1	2	–
SECVTX⊕SLT	5	–	1	–
	$W+3$ jets			
Total	21		5	
SECVTX	6	1	–	5
SLT	3	1	3	–
SECVTX⊕SLT	2	–	3	–
	$W+\geq 4$ jets			
Total	12		3	
SECVTX	7	2	1	1
SLT	2	1	–	–
SECVTX⊕SLT	2	–	–	–

Table 4.19: Breakdown of the $W+n$ jet events with Jetprobability tags, according to the number of jets tagged by the other two taggers. The column marked *common*, contains events in which the same jet is tagged by more than one taggers, while the column marked *different* contains events tagged by the other taggers but the tags are on different jets. The symbol $\oplus$ indicates that the event contains a jet tagged by both algorithms.

	Events with SLT tags			
	$W+1$ jets		$W+2$ jets	
	Single tags		Single tags	
	Common	Different	Common	Different
Total	146		56	
SECVTX	1	–	8	1
JPB	3	–	9	1
SECVTX⊕JPB	1	–	6	–
	$W+3$ jets		$W+\geq 4$ jets	
Total	21		12	
SECVTX	5	–	2	1
JPB	5	1	2	2
SECVTX⊕JPB	4	–	2	–

Table 4.20: Breakdown of the $W+n$ jet events with SLT tags, according to the number of jets tagged by the other two taggers. The column marked *common*, contains events in which the same jet is tagged by more than one taggers, while the column marked *different* contains events tagged by the other taggers but the tags are on different jets. The symbol $\oplus$ indicates that the event contains a jet tagged by both algorithms.

Chapter 5

Mistags and fakes

Tagged jets do not always originate from heavy flavor quark decays. Light quark or gluon jets can also be tagged. These tags are termed as *mistags*. Mistags are the result of detector resolution, tracking inefficiencies and mistakes of the tagging algorithms and constitute the instrumental background of the SVX based tagging algorithms. In the case of the SLT algorithm, fake tags are either due to jet fluctuations or due to real leptons which do not originate from any leptonic decays of W, Z, τ or heavy flavor hadrons. Conversion electrons or muons from decays in flight of charged π and K mesons, or very penetrating pions or π^0's are typical sources of SLT fake tags. For the SECVTX and Jetprobability algorithms, mistags constitute an important source of background to the $t\bar{t}$ signal while for the SLT algorithm, fakes is the dominant background.

Since mistags and fakes are due to non-heavy flavor jets, the tagging rate in a pure non-heavy flavor jet sample represents the true mistag rate for any tagging algorithm. The jet samples collected with the inclusive Level 2 QCD triggers contain small fraction of heavy flavor decays ($10 \sim 15\%$) and therefore are suitable for measuring

the tagging rate of non-heavy flavor jets.

The methods used to calculate the mistag rate for the three tagging algorithms are presented in the following sections. In Section 5.1.2 the fake rates for SECVTX and Jetprobability are calculated based on the rate of negative tags observed in the generic jet data. Section 5.2 describes the parametrization of the SLT fake rate based on the fraction of tracks which satisfy the soft lepton requirements. Because, the generic jet samples contain some heavy flavor decays, their contribution to the mistag rate is accounted for in Sections 5.1.4 and 5.2.1, using a combination of Monte Carlo simulations and data. The reliability of the derived mistag parametrizations is tested in different data samples and presented in sections 5.1.7, 5.1.8 and 5.2.2.

5.1 The SECVTX and Jetprobability mistags

5.1.1 Outline of the method

The SECVTX and Jetprobability negative tag yields are almost exclusively due to track reconstruction algorithm confusions and failures. Since these mismeasurements are equally likely to produce a positive or a negative tag, the negative tagging probability is appropriate to study the tagging rate of non-heavy flavor jets.

The method used to derive the mistag probabilities consists of three steps. First, the rates of positive and negative SECVTX and Jetprobability tags are measured in QCD inclusive jet data and are parametrized as a matrix, function of the jet transverse energy and jet track multiplicity. This first step determines the *tagging probability matrices* to be applied to any jet in any data sample.

At the second step, the contribution of heavy flavor decays to the negative tags is estimated and the negative tagging probability matrices are adjusted accordingly.

The heavy flavor contribution is assumed to be proportional to the amount of heavy flavor present in each QCD sample examined and the constant of proportionality is derived from heavy flavor jets in Monte Carlo simulations. These adjusted negative tagging matrices correspond to the true mistag rate due to non-heavy flavor jets.

At the last step, the effect of multiple interactions in the tagging rate is examined and a parametrization is derived, function of the $\sum P_T$ of the tracks associated with vertices other than the primary interaction vertex. This parameterization is then used to adjust the estimates of the mistag matrices, derived from the first two steps.

5.1.2 SECVTX and Jetprobability tagging rate parametrizations

The ability to reconstruct a secondary vertex in a jet depends on the jet transverse energy and the number of SVX tracks in the jet. Similarly for the Jetprobability, the efficiencies to identify a heavy flavor jet depends on the E_T of the jet since the heavy flavor hadron acquires larger boosts with increasing E_T and therefore its decay products are further displaced. It also depends on the number of SVX tracks in the jet, since heavy flavor decays result in larger track multiplicity jets. Therefore the fraction of heavy flavor jets is expected to raise for jets with higher track multiplicity. This larger track multiplicity results in better resolution in the vertex reconstruction. However as the jet E_T or track multiplicity increases so does the track hit density in the jet and hence the probability for tracking mistakes and confusions raises, resulting in higher mistag rates.

Due to those dependencies, both the positive and negative jet tagging rate are

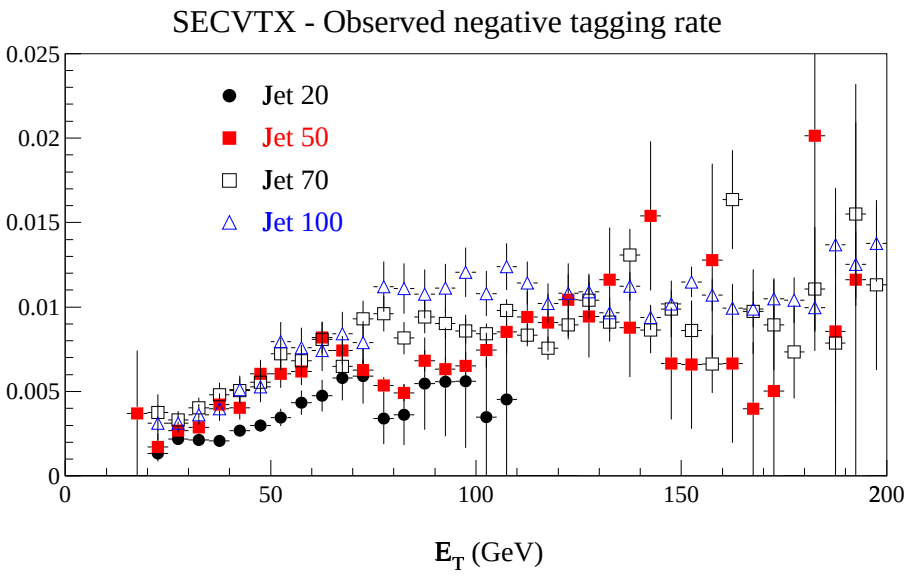

Figure 5.1: Distributions of SECVTX negative tagging rate of all jets in the inclusive JET_20, JET_50, JET_70 and JET_100 trigger samples as a function of the jet corrected E_T. The tagging rate changes by a factor of ~ 2 when moving across samples of lower E_T trigger threshold to samples with higher trigger threshold.

parametrized as a function of the jet E_T and number of good SVX tracks (defined in Section 4.2) in the jet. Samples of generic jets from the inclusive JET_20, JET_50, JET_70, JET_100 QCD Level 2 triggers and also jets from the $\sum E_T$_300 Level 3 trigger are used for the parametrization. The large number of data samples used is appropriate to map the jet tagging rates over a wide spectrum of jet E_T. The samples used in the parametrization cover adequately jets with E_T in the range $15 \leq E_T \leq 200$ GeV.

As discussed in Sections 3.1.7 and 3.1.8 the calorimeter includes unistrumented regions and regions of lower response. Consequently, the energy of jets in those regions is subject to large fluctuations. The tagging efficiency depends on the jet E_T and jets with E_T much lower than the trigger threshold are probably mismeasured. If these

jets were used to map the negative tagging probability in the QCD samples, it would lead to wrong estimate of the mistag rate. Figure 5.1 shows the observed negative tagging rate of all jets in the inclusive JET_20, JET_50, JET_70 and JET_100 trigger samples as a function of the jet corrected E_T. As seen, the tagging rate changes when moving from a sample of lower E_T trigger threshold to another one with higher E_T trigger threshold. Clearly, if one were to use a tagging rate parametrization derived from one QCD sample to predict the tagging rate in another sample of different E_T trigger threshold, he would predict a negative tag yield very different than the actual tag yield in the sample. Therefore, jets with E_T much smaller than the sample's harware E_T trigger threshold are assumed to be mismeasured and there should not be used in the tagging rate parametrization. However, the differences in the tagging rates are not only due to mismeasured jets. The heavy flavor content of the sample is different for different samples and also the nature of jets contributing tags changes (quark or gluon jets). The corrections to the negative tagging rate parametrizations that take into account these effects are discussed in Section 5.1.4.

5.1.3 Construction of tagging probability matrices

The positive and negative tagging probability matrices are derived using well measured jets selected according to the following procedure:

- The E_T dependence of the tagging rate is mapped using the inclusive Run 1B QCD samples: JET_20, JET_50, JET_70, JET_100 and $\sum E_T$_300. These inclusive jet samples will be referred to as *the QCD samples.*

- Well measured jets are selected requiring, the *jet corrected* E_T to be greater than the hardware trigger threshold of each sample. Specifically, jets with corrected

$E_T \geq$30, 60, 80, 110 and 150 GeV from the JET_20, JET_50, JET_70, JET_100 and $\sum E_T$_300 samples are used for the tagging rate parametrizations. These jets are termed as the *leading jets*.

- The E_T region below 30 GeV is mapped with jets from a sub-sample of the JET_20 sample. This sub-sample consists of events which contain at least two leading jets with corrected $E_T \geq 30$ GeV. The jets below the trigger threshold are referred to as *the non-leading jets*.

- Maximum $\not\!\!E_T$ selection requirements are imposed on the events to ensure that jets do not arise due to detector electronic malfunctions, main ring sprays or cosmic rays. The maximum value of allowed $\not\!\!E_T$ in the event varies according to the sample examined. Specifically, the $\not\!\!E_T$ in the event is required to be $\not\!\!E_T \leq$ 40, 50, 60, 70 and 120 GeV for the JET_20, JET_50, JET_70, JET_100, $\sum E_T$_300 samples, respectively.

- All the selected leading jets are required to be away from major detector cracks ($|\eta_D| \sim 0.0$ and $|\eta_D| \sim 1.0$). Only jets which centroids are in the region $0.2 \leq |\eta_D| \leq 0.8$ are selected.

- Jets pointing to the region which contains the cryogenic connections to the solenoidal magnet ($0.4 \leq \eta_D \leq 1.0$ and $1.4 \leq \phi \leq 1.5$) are also rejected.

- Event quality criteria are also applied. More explicitly, events from runs with known detector or trigger hardware problems are rejected. The primary vertex is required to be within $|Z_{vrtx}| \leq 60$ cm.

Figure 5.2 shows the corrected E_T distribution of all jets selected with the above procedure before and after applying SECVTX tagging. The corresponding distribu-

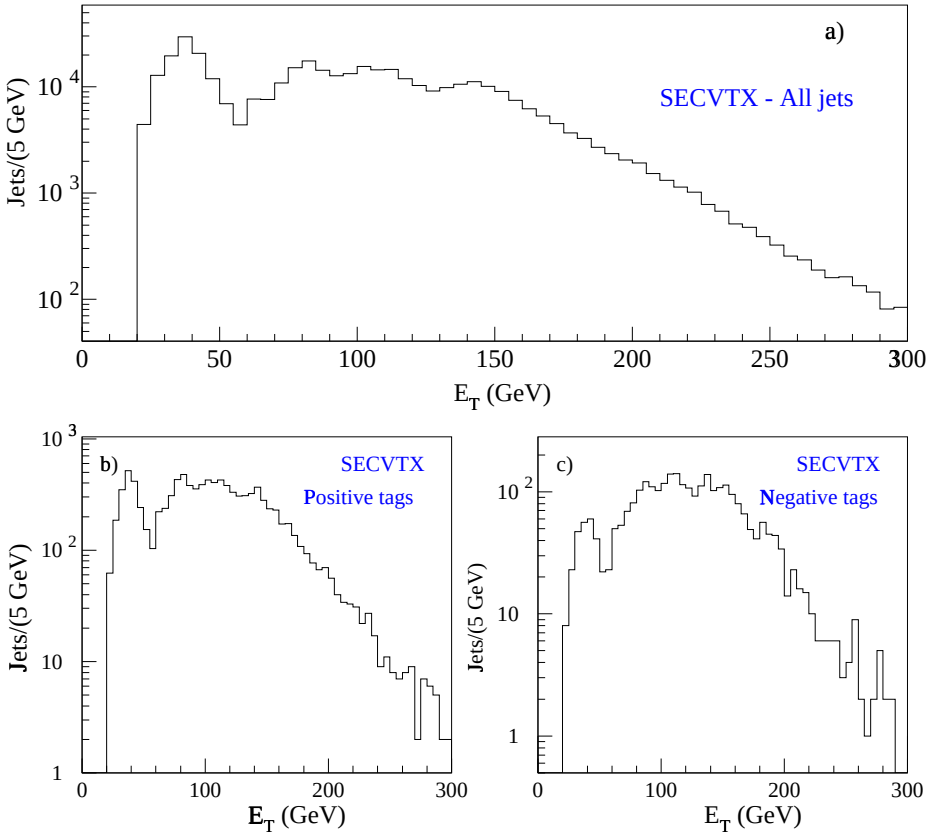

Figure 5.2: The distribution of corrected E_T of all jets used in the parametrization of the positive and negative SECVTX tagging probabilities before and after SECVTX tagging.

tions for Jetprobability are shown in Figure 5.3. For the case of the Jetprobability algorithm, jets are considered in the construction of the positive (negative) tagging probability matrices if they contain at least two good jetprobability tracks (see Table 4.4) with positive (negative) signed impact parameter significance. A jet is considered positively (negatively) tagged by the jetprobability algorithm if its positive (negative) jet probability is less than 0.05.

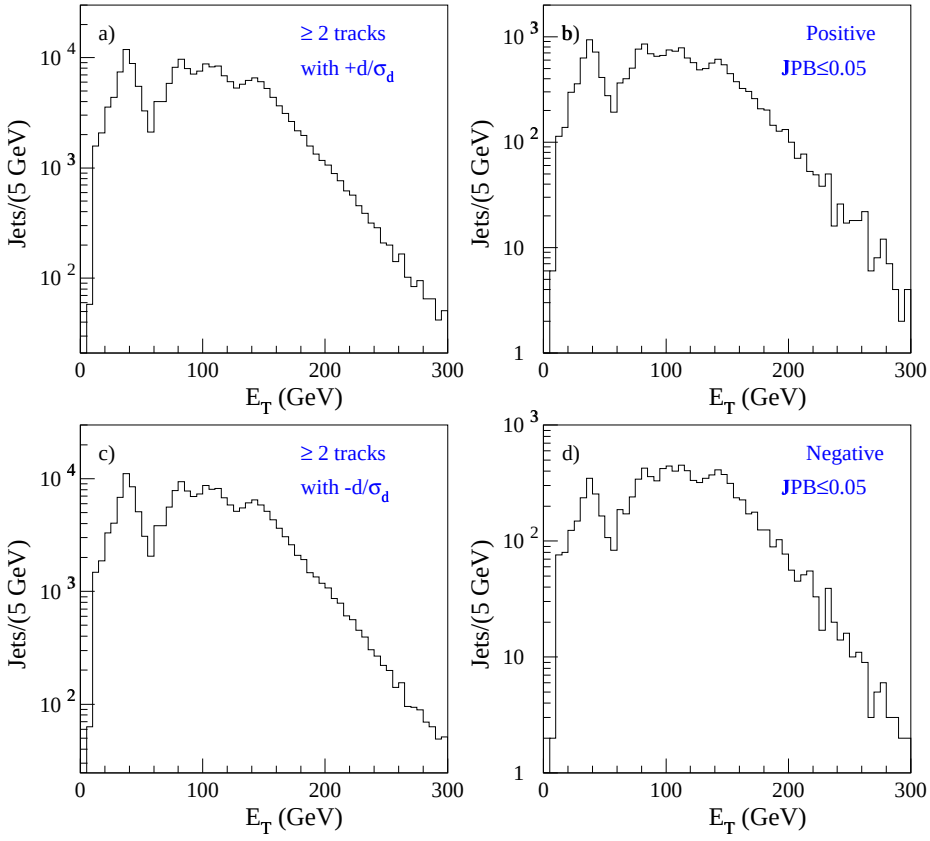

Figure 5.3: The distribution of corrected E_T of all jets with at least two tracks with positive (a) and (b) or negative (c) and (d) impact parameter significance before and after requiring the corresponding jet probability to be $\leq$0.05.

The dependence of the tagging rate on the jet corrected E_T is partitioned as fine as possible into 10 bins of E_T. The bin size is chosen as to reflect the uncertainty in the measurement of the jet transverse energy ($\sim$ 10%). To reduce statistical uncertainties, jets with 120 $\leq$ E_T $\leq$180 GeV are binned in bin of 30 GeV and only one bin is used for jets with corrected E_T $\geq$ 200 GeV. Table 5.1 shows the E_T range of each bin used in the parametrization of the tagging probabilities.

bin #	1	2	3	4	5	6	7	8	9	10
lower bound (GeV)	0	20	35	50	65	80	100	120	150	180
upper bound (GeV)	20	35	50	65	80	100	120	150	180	∞

Table 5.1: Lower and upper bound of the E_T bins used in the construction of the tagging rate matrices.

bin #	1	2	3	4	5	6	7	8
lower bound	2	3	4	5	6	8	10	14
upper bound	2	3	4	5	7	9	13	∞

Table 5.2: Lower and upper bound of the N_{SVX}^{trk} bins used in the construction of the tagging probability matrices.

The number of SVX tracks in a jet, N_{SVX}^{trk}, ranges from 2 to 15. Approximately 90% of the jets have N_{SVX}^{trk} less than 10. The jets are therefore binned into 8 bins with size as shown in Table 5.2.

Tagging probability matrices for SECVTX and Jetprobability are constructed separately for positively and negatively tagged jets. The tagging probability $\mathcal{R}_{ij}$ in the i jet E_T bin and j track multiplicity bin, is obtained as the ratio of the number of positive (negative) SECVTX or Jetprobability tags in this bin, N_{ij}^{tag}, to the total number of jets in the bin, N_{ij}^{tot}:

$$\mathcal{R}_{ij} = \frac{N_{ij}^{tag}}{N_{ij}^{tot}}$$

The statistical uncertainty in $\mathcal{R}_{ij}$, $\Delta\mathcal{R}_{ij}$, is obtained assuming binomial statistics. Figure 5.4 show the positive and negative SECVTX tagging probabilities as a function of the jet corrected E_T and jet track multiplicity, respectively. Both positive and negative tagging probability matrices show very strong dependence on both the jet

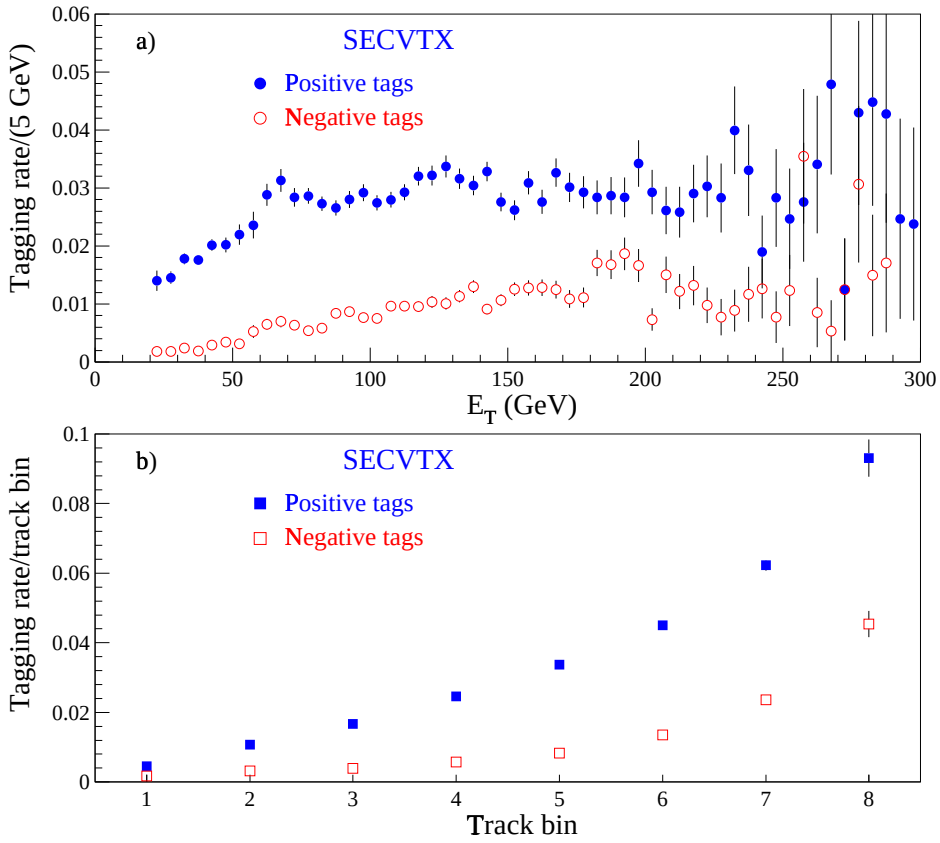

Figure 5.4: The positive and negative SECVTX tagging rates as a function of the jet corrected E_T (a) and as a function of the number of SVX tracks in a jet (b), respectively. The tagging rate as a function of number of tracks is plotted in terms of the track bins used in the construction of the matrices 5.2.

E_T and number of SVX tracks in the jet.

The positive and negative Jetprobability tagging rates as a function of the jet corrected E_T and number of SVX tracks in the jet are shown in Figure 5.5. The positive jet probability shows strong dependence on the number of tracks in the jet and a mild dependence on the jet E_T. For negative jet probability, the tagging rate shows a mild dependence on the jet E_T while it has almost no dependence on the

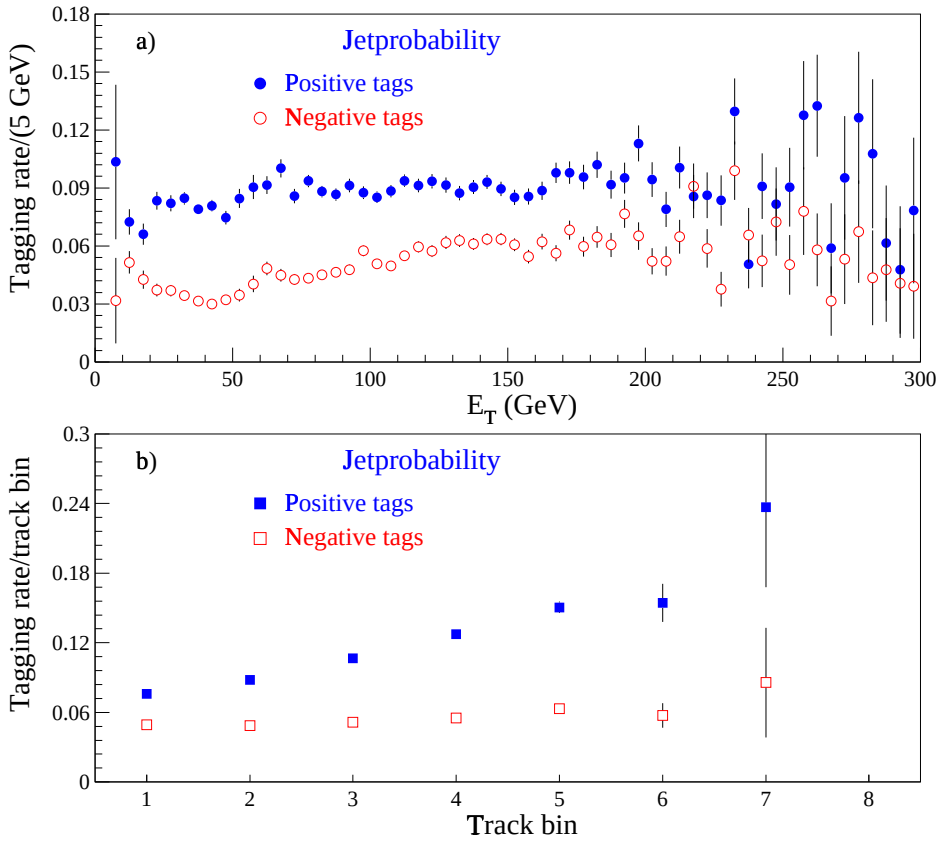

Figure 5.5: The positive and negative Jetprobability tagging rates as a function of the jet corrected E_T (a) and as a function of the number of SVX tracks in a jet (b), respectively. The tagging rate as a function of number of tracks is plotted in terms of the track bins used in the construction of the matrices 5.2.

number of tracks in the jet and it is $\sim$ 5%. By construction, the Jetprobability algorithm normalizes the overall probability to the number of tracks used. If the heavy flavor fraction were independent of the track multiplicity, the Jetprobability tagging rate should also be independent. Since the fraction of heavy flavor expected in the negative Jetprobability jets is small then the Jetprobability dependence on the number of tracks is also small. In contrast, the positive tagging rate increases

with track multiplicity. This is an indication that the heavy flavor fraction raises as the track multiplicity increases. The Jetprobability improves the results of tracking confusion by parametrizing the resolution function as a function of the number of SVX hits shared by other tracks but the probability for observing the specific ensemble of track impact parameters correspond to a single number, the jet probability. The behavior of the negative Jetprobability tagging rate is different than the corresponding one for the SECVTX algorithm. The SECVTX algorithm combines tracks in order to reconstruct secondary vertices and it is more sensitive to the combinatorics of the various tracks participating to the vertexing. These combinatorics grow rapidly as the track multiplicity increases. Therefore both positive and negative tagging probabilities strongly depend on the jet track multiplicity.

Before concluding this section it is important to check whether the selection criteria applied to the jets participating in the construction of the tagging probability matrices bias either the vertex position of the events with taggable jets or the jet η distribution. It is worrisome that selecting well measured jets away from cracks may bias the number of available taggable jets which then it would affect the jet tagging rate. Figure 5.6 shows a comparison of the primary vertex z-position distribution for events contributing jets in the construction of the tagging probability matrices to the vertex position distribution of events with at least one taggable jet with $|\eta_D| \leq 2.0$. It is clear that the requirements applied to the selection of well measured leading jets does not bias the vertex distribution.

In order to check whether the leading jet selection requirements biases the η distribution of the taggable rate, the pseudorapidity distributions of all positively and negatively SECVTX tagged jets are compared to the distributions obtained by weighting the η distribution of all taggable jets in the event with the tagging probability

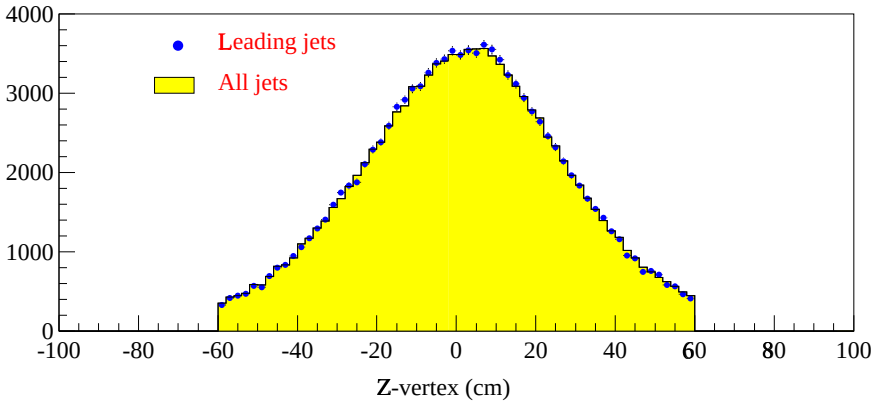

Figure 5.6: Comparison between the primary vertex z-position distribution for events contributing jets in the construction of the tagging rate matrices (points) and the distribution of all events with a taggable jet with $|\eta_D| \leq 2.0$ (shaded histogram). The two distribution are normalized to the same area.

matrix. In this case the requirement on a jet to be away from a major detector crack is relaxed. If there were a bias, the estimated rates would not reproduce the dip in the observed tag yields at $\eta_D \sim 0.0$ where the two SVX barrels met and therefore there is no tracking information. Also, the tagging rate predictions would not reproduce the observed tagging yields in regions excluded by the selection requirements. The results of this comparison are shown in Figure 5.7. The predictions reproduce very well the observed distributions.

5.1.4 Heavy flavor contribution to the negative tagging rate

It was mentioned earlier, that the heavy flavor content is not the same across the QCD samples used to derive the tagging probability matrices. The contributions

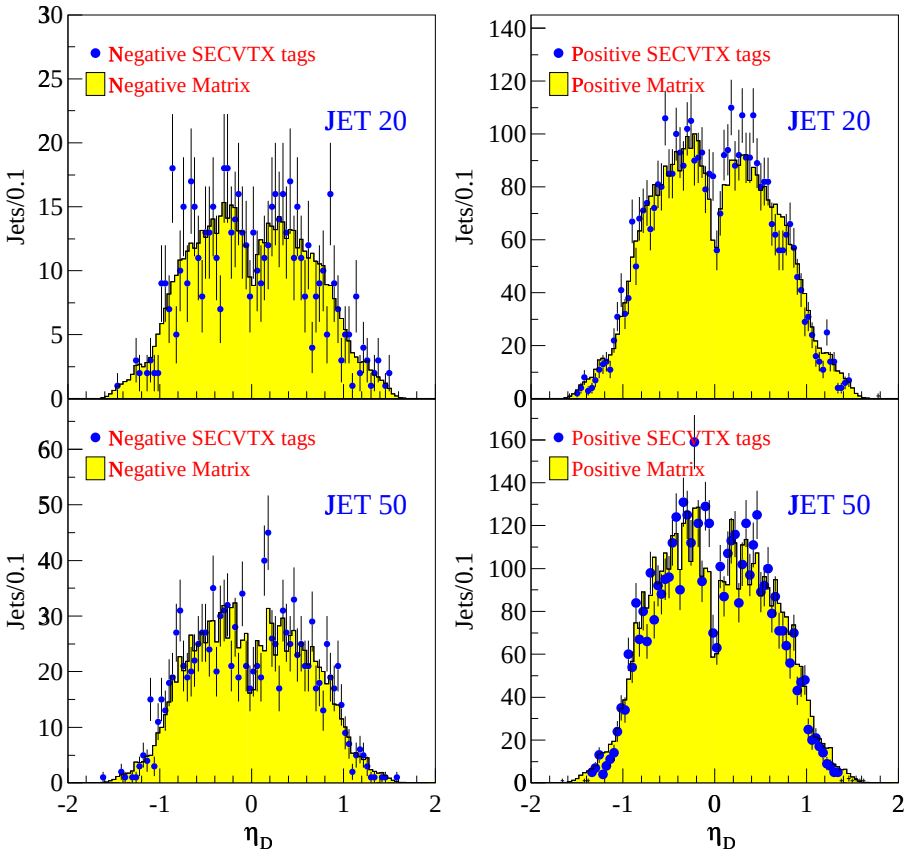

Figure 5.7: The pseudorapidity, η_D, distribution of the observed positively (negatively) SECVTX tagged jets in the JET_20 and JET_50 samples (points) compared to the predicted distribution obtained by weighting the η distribution of all taggable jets in the event with the positive (negative) tagging probability matrix (shaded).

from the three heavy flavor production mechanisms, namely direct production, flavor excitation and gluon splitting, change as the sample's jet E_T threshold increases and this directly affect the heavy flavor content of each sample.

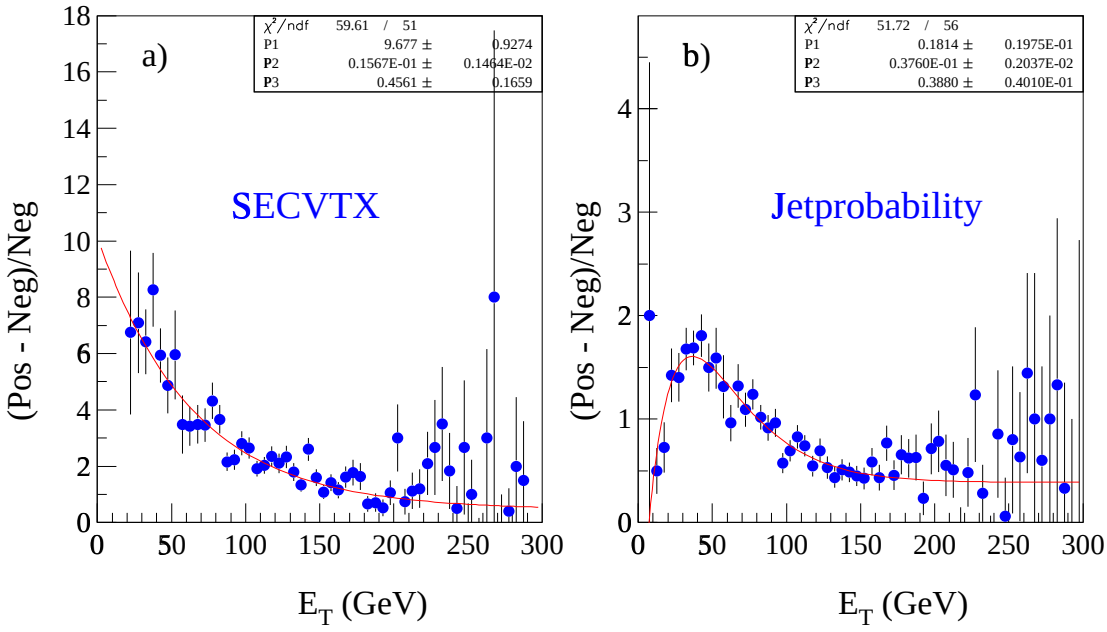

Figure 5.8: Ratio of the SECVTX (Jetprobability) positive tag excess, defined as the difference between the observed positive and negative tags, $P - N$, to the negative tags, N, as a function of the tagged jet corrected E_T for all leading jets in the QCD samples. (a): SECVTX and (b):Jetprobability tags. The lines represent the fits described in text.

The change in the heavy flavor content can be described in terms of the difference between the observed positive and negative tagging yields in the different samples. Recall, the negative tags are unphysical and are representative of the instrumental background which is expected to contribute equally on the positive tags. Therefore

the excess of positive over the negative tags can be explained in terms of heavy flavor decays plus some contribution of residual K_s and Λ decays which is of the order of less than 10% [59]. Figure 5.8 show the E_T dependence of the ratio, $\frac{(P-N)}{N}$ where P and N are the yields of positively and negatively tagged leading jets in all the QCD samples.

Fitting the data distributions shown in Figure 5.8 the E_T dependence of the heavy flavor content can be parametrized with the following expressions:

$$\frac{(P-N)}{N} = \begin{cases} 0.456 + 9.7 \cdot e^{-0.0156 \cdot E_T} & \text{for SECVTX} \\ (E_T - 10GeV) \cdot 0.181 \cdot e^{-0.0376 \cdot E_T} + 0.388 & \text{for Jetprobability} \end{cases} \tag{5.1}$$

It is reasonable to assume that the contribution of heavy flavor decays to the negative tags grows with the amount of heavy flavor present in the sample. Since the excess of positive tags, $(P-N)$, represents the amount of heavy flavor present in the sample, it is assumed that the heavy flavor contribution to the negative tags is proportional to $(P-N)$. The constant of proportionality is derived using heavy flavor jets from Monte Carlo simulations according to the following procedure.

- The heavy flavor jets are classified into two major categories. One category consists of those jets which contain only one heavy flavor hadron. These jets are called *"quark jets"*. The second class consists of those jets which contain two or more heavy flavor hadrons as shown in Figure 5.9. These jets are termed as *"gluon jets"* because of the production mechanism responsible for the presence of two heavy flavor objects (gluon splitting, $g \rightarrow b\bar{b}(c\bar{c})$).

- The ratios, $R = \frac{N}{P}$, of negative to positive tag yields for both classes of b and c-jets are measured in Monte Carlo simulations. Separate parametrizations,

$R_1(E_T)$ and $R_2(E_T)$, of the ratio R, are obtained for both "quark" and "gluon" jets by fitting the corresponding Monte Carlo distributions.

- The heavy flavor contribution on each E_T bin of the negative tagging probability matrix is obtained by multiplying the positive tag excess corresponding to the specific E_T bin with the R_1 parametrization which corresponds to "quark" jets. For jets entering in the E_T bins below 30 GeV the R parametrization for "gluon" jets is used. At this step, it is assumed that *leading jets are "quark" like while non-leading jets are "gluon" like.*

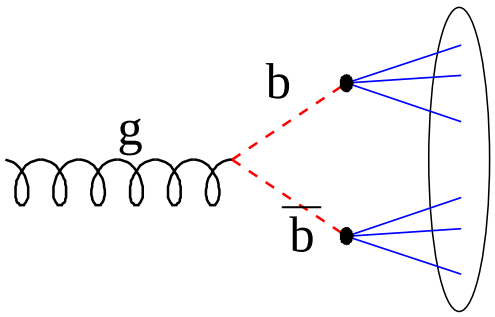

Figure 5.9: A schematic view of a jet containing two heavy flavor hadrons. Jets of this type arise due to gluon splitting and are termed gluon jets.

"Quark" jets are studied with PYTHIA Monte Carlo simulations of processes like $t\bar{t}$, WW, WZ, ZZ, WH and also VECBOS $W + b\bar{b}(c\bar{c})$. These simulations produce heavy flavor jets that contain most of the times one heavy flavor hadron. "Gluon" jets are studied with HERWIG Monte Carlo simulations of the $W+ \geq 1$ jet process. Heavy flavor jets in these simulations arise mainly from gluon splitting to $b\bar{b}$ or $c\bar{c}$

pairs. The selection of jets containing one or two heavy flavor hadrons is based on information at the generator level. Table 5.3 summarizes the tagging rates observed in the Monte Carlo simulations for "quark" and "gluon" jets. The jetprobability

SECVTX					
Type	no. of jets	Pos. tags	Neg. tags	Mistags	ϵ_{tag}
b-quark	50431	14249	288	0	0.283±0.002
$g \to b\bar{b}$	5087	1002	85	0	0.197±0.005
c-quark	12376	846	35	0	0.068±0.002
$g \to c\bar{c}$	8781	389	20	0	0.044±0.002

Jetprobability						
Type	no. of jets	Pos. tags	Neg. tags	Mistags	ϵ_{tag}	$\frac{\epsilon_{tag}^{JPB}}{\epsilon_{tag}^{SVX}}$
b-quark	50079	16671	2368	270	0.327±0.002	1.16±0.01
$g \to b\bar{b}$	5087	1265	253	15	0.246±0.008	1.25±0.05
c-quark	11805	1624	330	105	0.128±0.004	1.88±0.08
$g \to c\bar{c}$	8781	838	204	56	0.089±0.003	2.02±0.11

Table 5.3: Tag rates of quark and gluon jets observed in Monte Carlo events after full detector simulations. Quark jets and gluon jets have different E_T spectra. In the SECVTX case there are no mistags from heavy flavor jets as indicated in 4th column. The mistags for Jetprobability are derived by fitting the jet probability shape of heavy flavor jets in the region 0.1∼1 and extrapolating the fit in the region between $0 \sim 0.05$. ϵ_{tag} is the positive tagging efficiency for SECVTX or Jetprobability respectively. Also shown in the last column, the ratio of the Jetprobability positive tagging efficiency to the SECVTX efficiency. This ratio can be used to estimate the relative heavy flavor b and c-jet content in any sample.

mistags on b and c-quark or gluon jets shown in the 4th column of the Table, are derived by fitting the shape of the jet probability distribution of these jets in the range 0.1∼1 and extrapolating the fit results in the region $0 \sim 0.05$.

In the simulation, jets which are tagged by SECVTX and are not correlated to any heavy flavor hadron at generator level, are equally distributed with positive and negative tags. The negative tagging rate of these jets correspond to 10% of the rate predicted when applying the negative probability matrix. This implies that the tracking simulation is very efficient and it does not produce non-heavy flavor secondary vertices.

The distributions of $R = \frac{N}{P}$ as a function of the jet E_T for all jet classes, as obtained from the different simulations, are shown in Figure 5.11. The b and c "quark"-jets of the same class exhibit similar behavior, although the R distribution for jets containing c-hadrons appear slightly higher that then corresponding b-"quark" distributions. However, the tagged c jet distributions are subject to limited statistics. On the contrary, for jets of the same E_T, the R values are larger for "gluon" jets than for "quark" jets. This different behavior can be explained according to Figure 5.10. When a jet contains two heavy flavor hadrons there are two possible secondary vertices that can be reconstructed. However, since there is some spatial separation between these vertices, tracks from one b-decay vertex can be combined with tracks from the other b-vertex to form a secondary vertex with negative L_{xy} as shown in Figure 5.10. Parametrizations of the ratio R as a function of the jet E_T for all types of jets are obtained by fitting the distributions shown in Figure 5.11 with first degree polynomials of the form $P_0 + P_1 \cdot R$. The same polynomial is used to fit the R distributions of b and c-hadron jets belonging to the same class of jets. The fit is dominated by the b-hadron jet distributions because of higher statistics. The "gluon" jets which populate lower jet E_T regions, are fitted using a polynomial of the same slope to the one obtained when fitting the corresponding "quark" jets distributions.

The functional form that describes best the ratio, R, of negative to positive tags

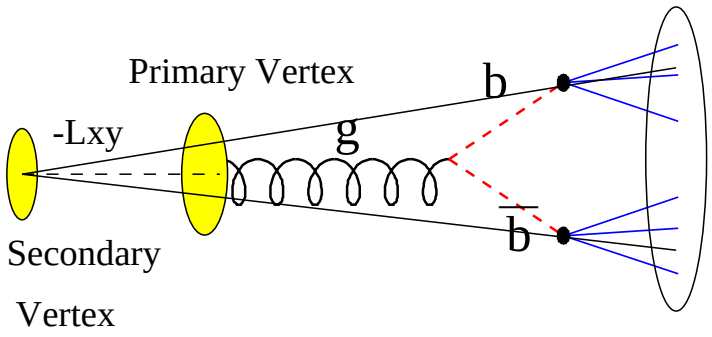

Figure 5.10: A schematic view of mistags produced by jets containing two heavy flavor hadrons from gluon splitting. Jets containing heavy flavor hadrons from gluon splitting result to higher negative tagging rate because tracks from the two b-vertices can be combined to form a secondary vertex with negative L_{xy} or negative jetprobability.

in the Monte Carlo simulations for the "quark" type of jets is given by the following expression:

$$R_1(E_T) = \begin{cases} 0.0088 + 0.000158 \cdot E_T & \text{for SECVTX} \\ 0.039 + 0.00117 \cdot E_T & \text{for Jetprobability} \end{cases} \tag{5.2}$$

The corresponding functional form for "gluon" jets is given by:

$$R_2(E_T) = \begin{cases} 0.075 + 0.000158 \cdot E_T & \text{for SECVTX} \\ 0.14 + 0.00117 \cdot E_T & \text{for Jetprobability} \end{cases} \tag{5.3}$$

To account for the heavy flavor contribution to the negative tags, each bin of the

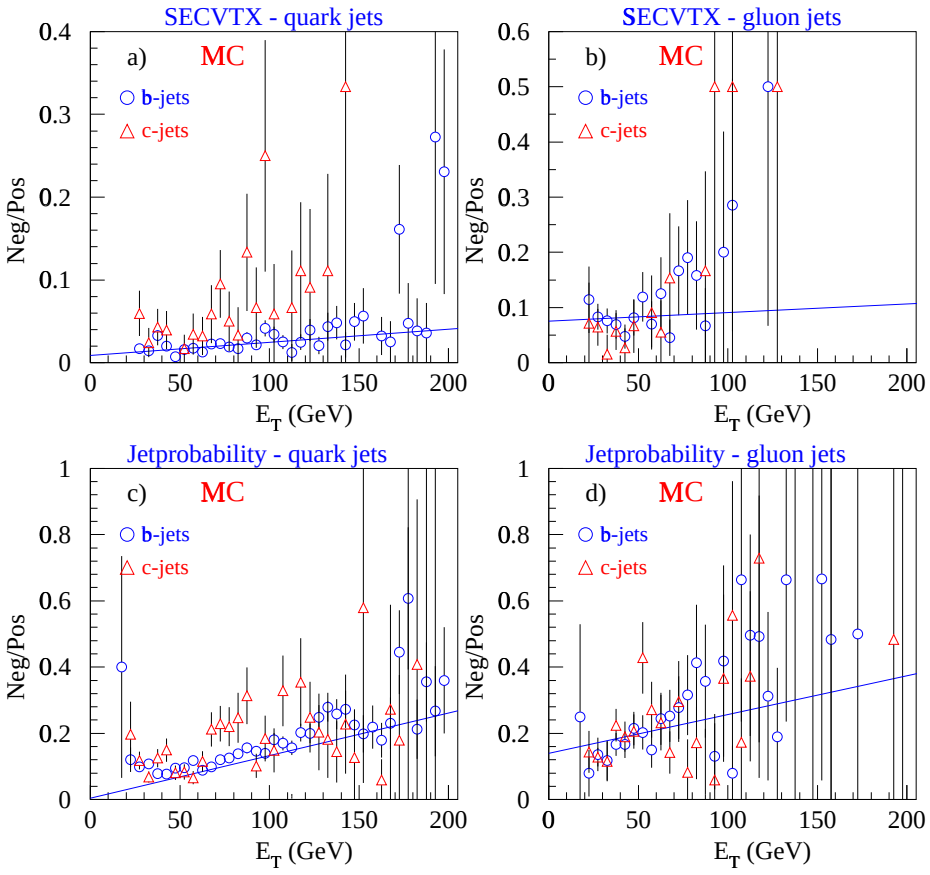

Figure 5.11: Ratio of negative to positive tags as a function of the jet E_T for quark and gluon jets containing b-hadrons (open circles) and c-hadrons (triangles), in Monte Carlo simulations. (a): SECVTX, quark jets (b): SECVTX, gluon jets (c):Jetprobability, quark jets and (d): Jetprobability, gluon jets. The lines represent the fits to the distributions as described in text.

negative tagging matrix is adjusted according to the following expression:

$$Mistags(E_T) = \begin{cases} N - (P - N) \cdot R_1(E_T) & \text{for jet } E_T \geq 30 \text{ GeV} \\ N - (P - N) \cdot R_2(E_T) & \text{for jet } E_T \leq 30 \text{ GeV} \end{cases} \quad (5.4)$$

where P and N are the positive and negative tagging yields of the well measured leading jets. This adjusted negative tagging probability matrix represents the *true mistag* probabilities due to detector resolution effects. The rescaled mistag matrices for both algorithms are given in Appendix A.

Table 5.4 shows the estimated fraction of heavy flavor decays in each E_T bin of the negative tagging probability matrix.

E_T range	1	2	3	4	5	6	7	8	9	10
SECVTX (%)	-	12	10	7	8	6	6	6	5	5
JPB (%)	10	19	15	13	15	12	12	10	12	15

Table 5.4: Fraction (%) of negatively SECVTX and Jetprobability tagged jets due to heavy flavor decays as a function of the jet corrected E_T. The E_T range is given in terms of the bins used in the construction of the tagging probability matrices 5.1.

In equation 5.4, the observed positive excess of leading jets is multiplied by the R_1 parametrization which corresponds to "quark" jets while the gluon parametrization, R_2, is used for the part of matrix covered by non-leading jets ($E_T \leq$30 GeV). This assumption is tested in QCD HERWIG Monte Carlo simulations of the generic $2 \rightarrow 2$ parton process. Three large Monte Carlo samples were generated in order to simulate the JET_20, JET_50 and JET_100 data samples. These samples reveal that 95.8%

of the leading tagged jets in the simulation are due to quarks while almost 100% of the non-leading jets are due to gluon jets. However, the validity of the assumption is tested also in all available Run 1B QCD samples as it is will be discussed in Sections 5.1.7 and 5.1.8. It will be shown that the sum of the true mistag yields as calculated using the mistag matrix plus the heavy flavor contribution to the negative tags predict very well the observed negative tag yields.

5.1.5 The effect of multiple interactions to the negative tagging rates

Because of the high luminosity conditions during the run, there are more than one interactions in the event. These interactions contribute additional tracks and energy in the event that can result to larger tracking algorithm confusions, or mismeasurements of the jet centroid. These effects directly influence the tagging rate. In the QCD samples which are used to derive the tagging probability matrices $\sim$70% of the events contain additional interactions. The effect of the multiple interactions is different when other samples are examined. This change is due to the different criteria applied in the sample selection which can bias towards less energetic multiple interactions. For example the isolation requirement present in the selection of the $W+\geq 1$ jet sample or the photon+jets sample selects against very energetic additional interactions. Indeed, the effect of multiple interactions to the tagging rate was noticed in the isolated photon sample when the predicted negative tagging yields overestimated the observed negative tagging yields. In this sample, the selection criteria (see Section 3.9.3 require also less than 2 GeV calorimeter energy in a cone of radius 0.7 around the photon candidate. Essentially, this requirement selects events with small

contribution from multiple interactions.

In order to determine the effect of additional interactions, the negative tagging rate is studied in QCD events as a function of $\sum P_T^V$, the sum of the transverse momentum of all tracks associated with additional vertices displaced by more than 5 cm from the primary vertex.

The distribution of $\mathcal{P}_N / < \mathcal{P}_N >$, defined as the negative tagging probability relative to the average negative tagging probability in the QCD sample, is shown in Figure 5.12 as a function of $\sum P_T^V$. Unity corresponds to the values of average negative tag probability of $< \mathcal{P}_N^{SVX} >= 0.0073$ and $< \mathcal{P}_N^{JPB} >= 0.0495$ for SECVTX and Jetprobability respectively. Using the distributions shown in Figure 5.12(a) and

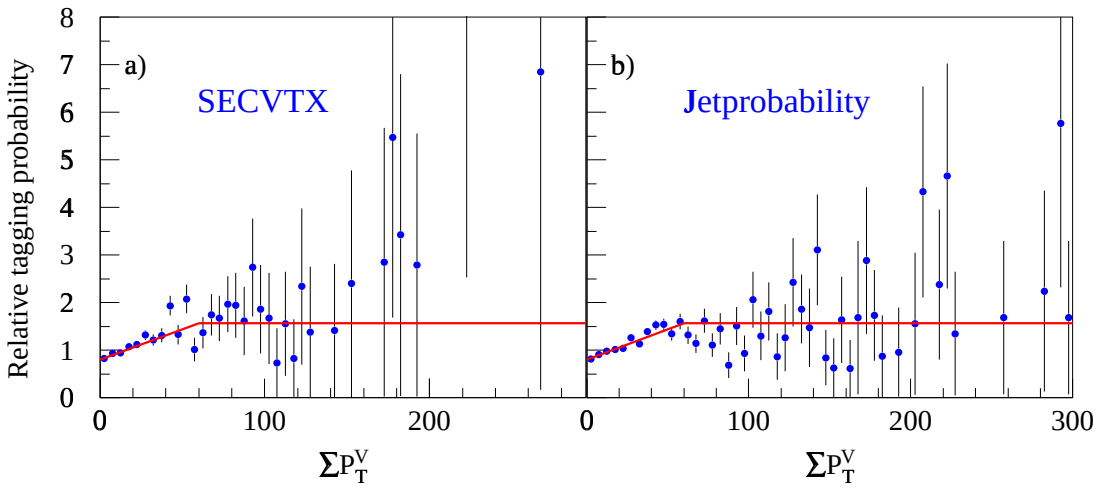

Figure 5.12: The distribution of $\mathcal{P}_N / < \mathcal{P}_N >$, defined in text, for SECVTX (a) and Jetprobability (b) as a function of $\sum P_T^V$, in the QCD samples. The lines indicate the parametrization used to derive the effect of multiple interactions to the negative tagging rates.

(b), the average negative tagging rate is corrected for the contribution of additional interactions using the following expression:

$$\mathcal{F}(\sum P_T^V) = \begin{cases} 0.8 + 0.0128 \cdot \sum P_T^V & \text{for } \sum P_T^V < 60 \text{ GeV/c} \\ 1.57 & \text{for } \sum P_T^V \geq 60 \text{ GeV/c} \end{cases}$$

This parametrization reproduces the measured negative tag rate within 3% in all QCD samples.

The amount of mistags calculated using the true mistag matrix is multiplied by the above $F(\sum P_T^V)$ parametrization in order to obtain the final estimate of mistags in any sample. For the samples used in the construction of the matrix the effect is already included in the matrix.

5.1.6 Test of the negative to positive tag ratio in the inclusive low-P_T electron sample

The calculation of the heavy flavor contribution to the negative tagging rate was derived based on the Monte Carlo predictions for the ratio, R, of negative to positive tagged jets. In order to test whether the Monte Carlo simulates correctly this ratio, a sample enriched in heavy flavor decays is needed. As discussed in Section 4.5, the b-purity of the low P_T inclusive electron sample is $\sim 43 \pm 4\%$ and therefore it is a good sample to test the R_1 and R_2 ratios since most of the negative tags are due to heavy flavor. For the comparisons, the HERWIG Monte Carlo sample of low P_T electrons used in Section 4.5 is used. As a reminder, the sample contains electrons from all possible heavy flavor production mechanisms (direct production, flavor excitation, gluon splitting) of b and c-jets.

For the purpose of this test, the measured negative tagging yields of the electron and recoiling jet (away jet) in both data and simulation samples are compared to the predictions based on the sum of mistags, N_{mtg}, and heavy flavor contribution to the negative tags, N_{hf}^{neg}. N_{mtg} is calculated using the mistag matrices derived in Section 5.1.4. The heavy flavor contribution to the negative tags of each sample is calculated according to the following expression:

$$N_{hf}^{neg} = (N_{pos}^{tag} - N_{mtg}) \cdot (F_g \cdot R_2 + (1 - F_g) \cdot R_1)$$

where N_{pos}^{tag} is the positive tag yield, R_1 and R_2 are the parametrizations for the ratios of negative to positive tags for quark and gluon jets and F_g is the fraction of gluon jets in the simulation. As discussed in Section 6.6.3, the cross section for gluon splitting in the HERWIG simulation is 35% lower than the cross section estimated in the data. Therefore, when calculating in the data the contribution of heavy flavor jets to the negative tags, the gluon splitting contribution is increased by 35%.

	$b\bar{b}$ direct and F.E.	$g \to b\bar{b}$	$c\bar{c}$ direct and F.E	$g \to c\bar{c}$
e-jet pretagged	60.8%	19.2%	13.8%	6.2%
SVX tag on e-jet	72.3%	22.5%	3.2%	2.0%
JPB tag on e-jet	68.2%	20.9%	7.4%	3.5%
a-jet SVX tag on e-jet	87.7%	2.0%	6.1%	4.2%

Table 5.5: Heavy flavor content (%) of the positively tagged electron and away jet as measured in the low P_T inclusive electrons HERWIG Monte Carlo sample.

The percentage fractions of the various processes contributing positive SECVTX and Jetprobability tags in the electron and away-jet in the simulated sample are

listed in Table 5.5. In this table, heavy flavor jets from direct production and flavor excitation are counted as "quark" jets. Because the b-purity of the pretagged electron jets is $43 \pm 4\%$, the away jet is examined only when the electron jet is SECVTX tagged.

A comparison of the measured and predicted negative tagging rates for both electron and away jets as determined in the data and simulated samples is shown in Table 5.6

SECVTX								
	Data				HERWIG			
	N_{pos}^{tag}	N_{mtg}	N_{neg}^{tag}	$N_{mtg} + N_{hf}^{neg}$	N_{pos}^{tag}	N_{mtg}	N_{neg}^{tag}	$N_{mtg} + N_{hf}^{neg}$
lepton jet	9047	82.9	452	420.2	5643	0	184	180.6
away jet	1129	19.0	59	43.3	596	0	8	8.8
Jetprobability								
	Data				HERWIG			
	N_{pos}^{tag}	N_{mtg}	N_{neg}^{tag}	$N_{mtg} + N_{hf}^{neg}$	N_{pos}^{tag}	N_{mtg}	N_{neg}^{tag}	$N_{mtg} + N_{hf}^{neg}$
lepton jet	10285	385.5	1527	1533.4	7511	75.8	1000	905.1
away jet	1285	75.6	173	187.4	827	56.0	125	121.0

Table 5.6: Tag rates in the low P_T inclusive electron data sample and in the corresponding HERWIG simulation. Mistag rates are calculated with the mistag matrices. N_{hf}^{neg} is the predicted rate of negative tags due to heavy flavor decays.

The comparisons between measured and predicted negative SECVTX and Jetprobability tags in the data and Monte Carlo simulation are shown in Figures 5.13 and 5.14.

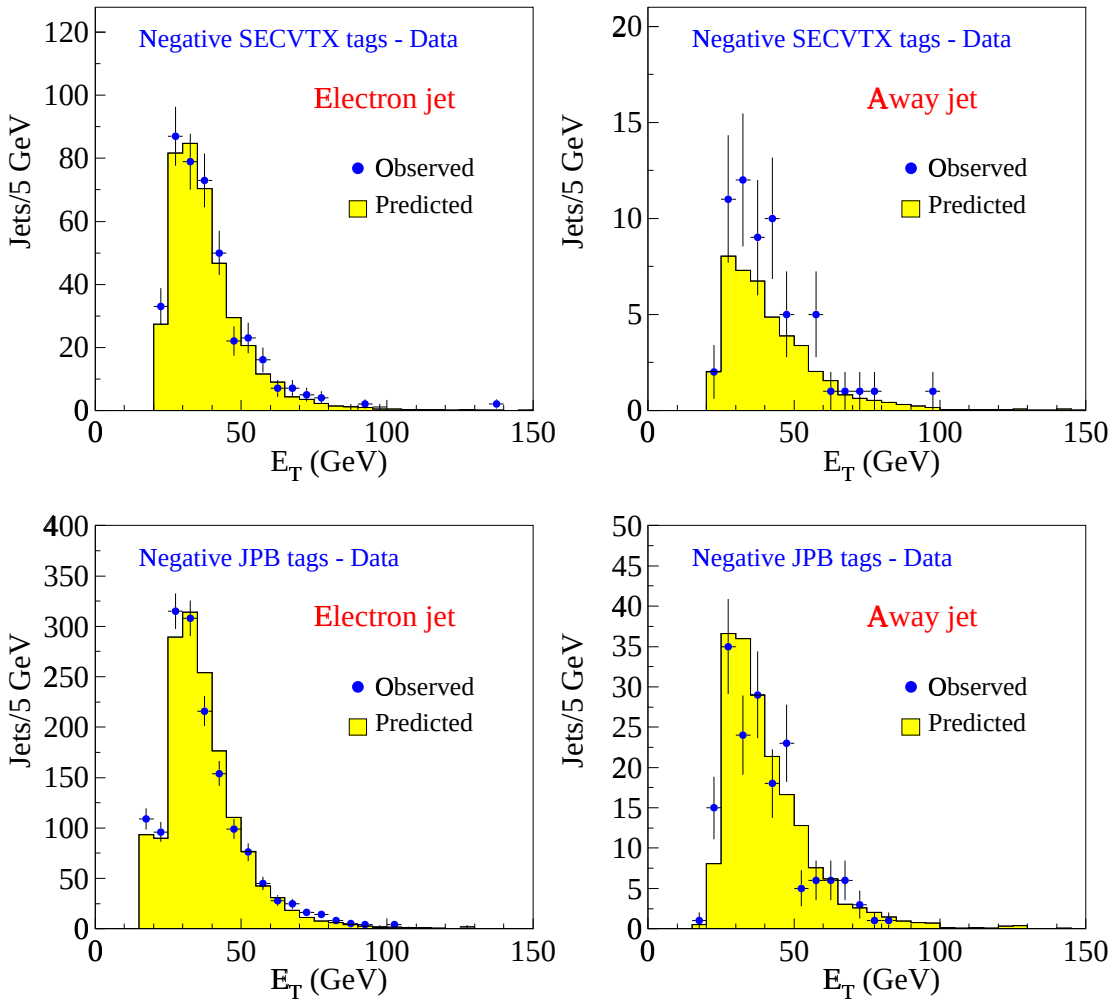

Figure 5.13: Corrected E_T distributions for observed (points) and predicted (shaded) negative tags in the low P_T inclusive electron data sample.

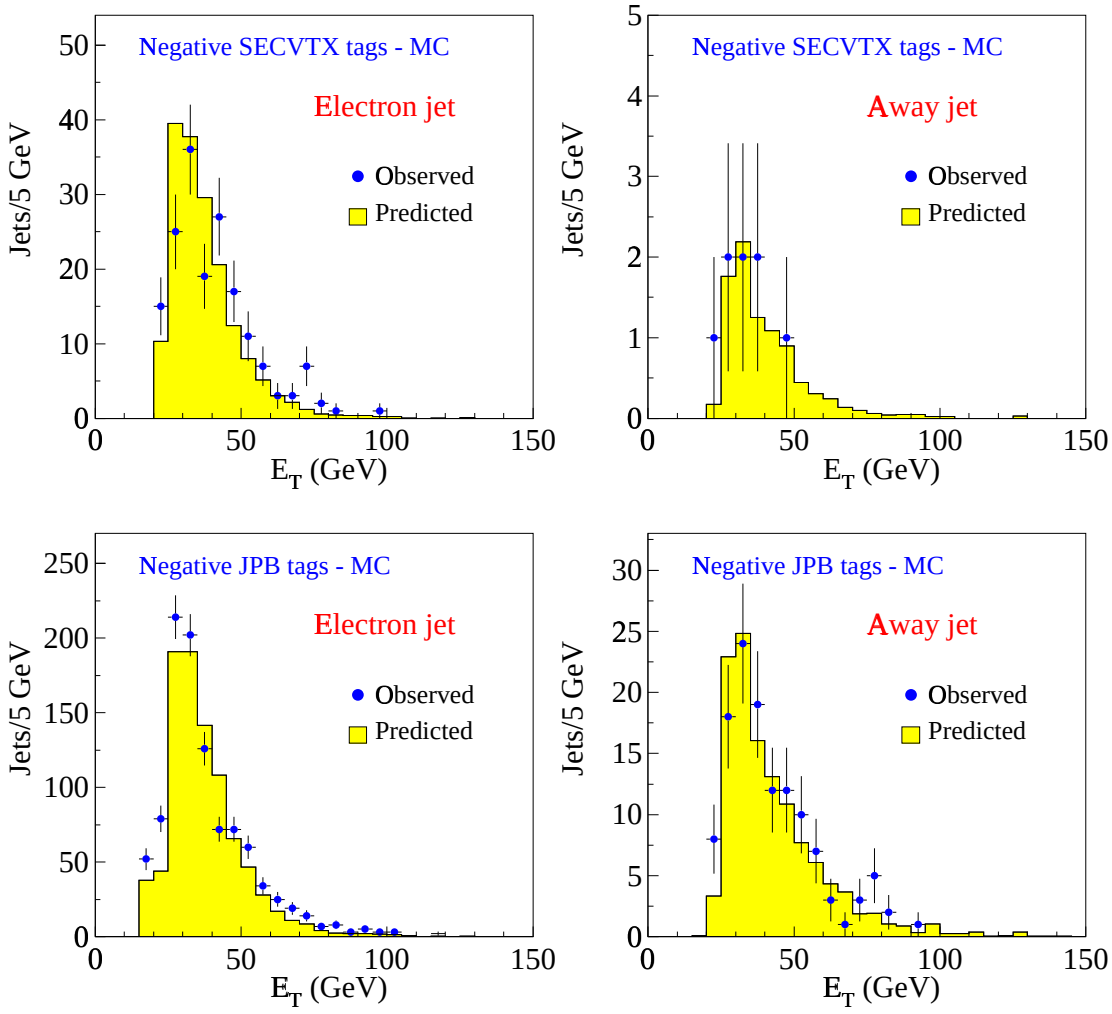

Figure 5.14: Corrected E_T distributions for observed (points) and predicted (shaded) negative tags in the low P_T inclusive electron HERWIG Monte Carlo sample.

5.1.7 Application of the SECVTX mistag parametrization on all the QCD samples

The parametrization procedure described in the previous sections is tested here in all available QCD samples. The procedure to be valid, it has to be able to predict the negative tagging rate observed in all QCD samples, using the mistag matrix parametrization and adding the heavy flavor contribution to the negative tags. As described, the heavy flavor contribution is calculated from the difference between the positive and negative tags multiplied with the parametrizations $R_1(E_T)$ and $R_2(E_T)$ for jets above and below the trigger threshold, respectively.

For the purpose of the checks, the $\not\!\!E_T$ requirements and the requirements imposed on the selection of leading jets for the construction of the tagging rate matrices, are relaxed and the parametrization is applied on all jets in the events, including the ones in the calorimetry cracks. The tests are carried out in all samples used in the parametrization and also on all other available QCD samples. The $\sum E_T$_175, $\sum E_T$_125_4Clusters and the isolated photon sample, described in Section 3.9, are used as independent samples.

Figure 5.15 serves as an illustration of the procedure followed in order to compare the observed negative tagging rate in the QCD samples to the one expected from pure mistags and heavy flavor. Figure 5.15(a) shows a comparison between the corrected jet E_T distribution of the observed positive tags in the JET_100 sample to the expected mistag distribution obtained convoluting the corrected jet E_T spectrum with the mistag matrix probabilities. In Figure 5.15(b) the mistag rate predictions are compared to the negative tagging yields. As shown, the mistag rate which does

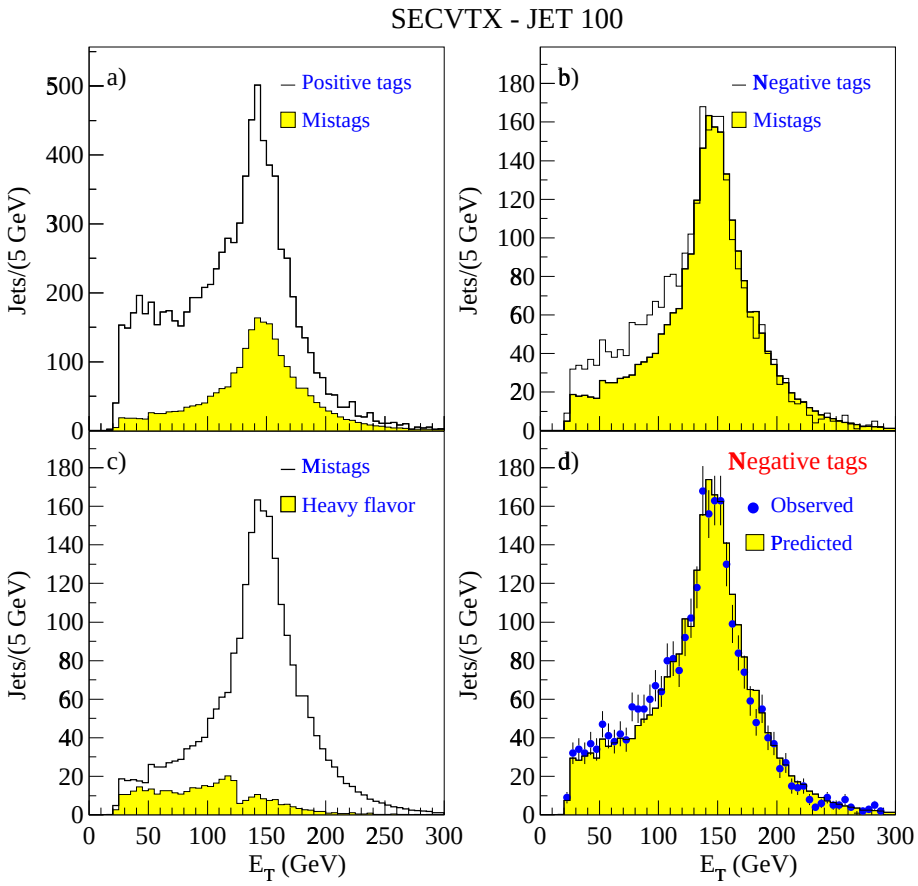

Figure 5.15: Corrected E_T distributions of jets with SECVTX tags in the JET_100 sample. The plots illustrate the procedure followed to compare the observed negative tagging rate to the one expected using the mistag rate predictions and adding the heavy flavor contribution. (a): Observed positive tags (open) compared to the calculated mistags (shaded). (b): Observed negative tags (open) compared to the calculated mistags (shaded), which does not include heavy flavor contributions. (c): mistags (open) compared to the predicted heavy flavor contribution to the negative tags (shaded). (d): Observed negative tags (points) compared to the mistags plus the heavy flavor contribution to the negative tags (shaded). Very good agreement is observed between the two distributions.

not include any heavy flavor contribution is lower than the observed negative tagging yield. Figure 5.15(c) shows the E_T distribution of mistags and the distribution of the heavy flavor contribution to the negative tags as obtained by multiplying the difference of positive tags minus mistags (shown in Figure 5.15(a)) by R_1 if a jet has E_T $\geq$120 GeV (leading jet) and by R_2 if a jet has E_T $<$120 GeV (non-leading jet). The estimated heavy flavor contribution to the negative tags is $\sim$13.6% of the observed negative tag yield. Figure 5.15(d) shows a comparison between the observed and predicted yield of negative tags. The predicted yield is derived by adding the two distributions shown in Figure 5.15(c). Very good agreement is observed between the predicted and observed yields of negative tags.

Following the same procedure, comparisons between the corrected jet E_T distributions for the observed negative SECVTX tags and the distributions obtained convoluting the jet E_T spectra with the mistag rate predictions can be performed for all samples as shown in Figures 5.16 and 5.17. For the case of the all hadronic sample, ($\sum E_T$_125_4Clusters) there is no clear distinction as of which jets to be counted as leading or non-leading ones. From Monte Carlo simulations it was determined that leading jets in this sample correspond to 2/12 of all jets. This ratio is then used for R_1 and R_2 when calculating the heavy flavor contribution. For the $\sum E_T$_175 sample, it is assumed that the sample is formed mainly by dijet events with uncorrected E_T $\geq$80 GeV. Therefore jets with corrected E_T $\geq$110 GeV are assumed to be leading jets and are scaled by R_1 while jets below 110 GeV are taken as gluon jets and multiplied by R_2. Finally for the photon sample, it is assumed that the tagged jets are resulting from quark jets ($\gamma + c$ production) and the jets are scaled according to the R_1 parametrization.

The mistag rate calculation plus the heavy flavor contributions reproduces the

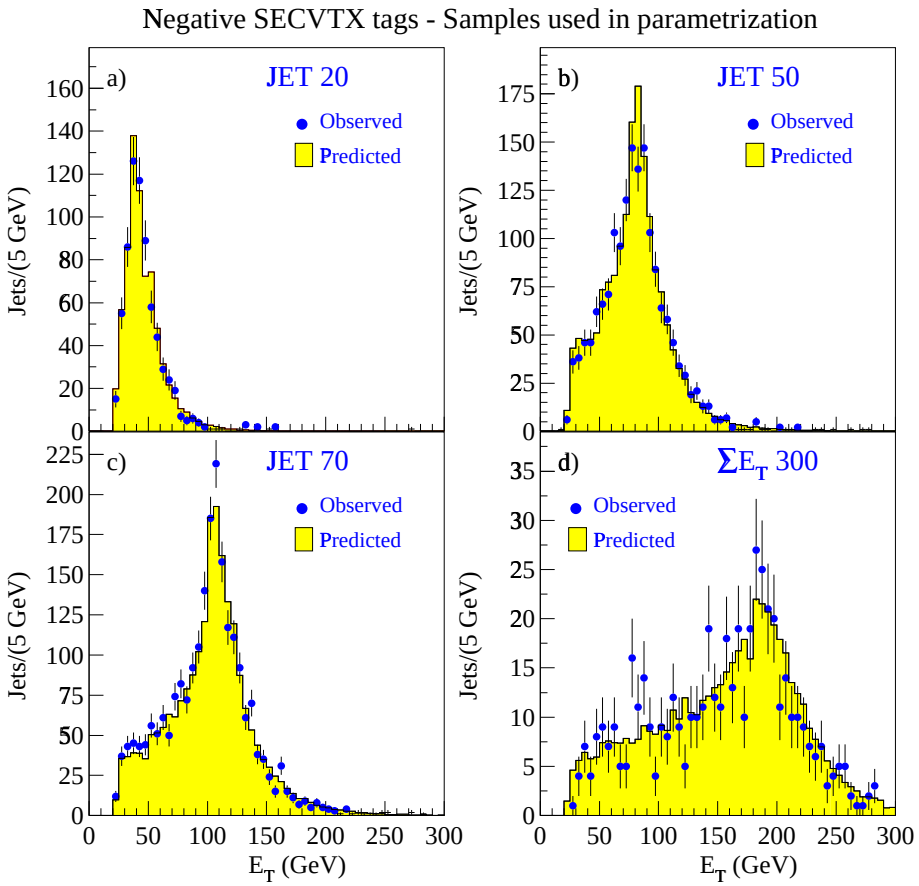

Figure 5.16: Comparison between the corrected E_T distributions of negative SECVTX tags and the ones obtained from the mistag calculations adding the heavy flavor contribution to the negative tags. The shown distributions correspond to the samples used in the construction of the mistag probability matrices.

observed negative tagging yields remarkably well for all the tested samples. It should be emphasized that the parametrization is tested even for the samples used in the construction of the mistag matrices. As expected, very good agreement is found for jets above the trigger threshold because these jets are used in the construction of the mistag matrix. However, the agreement in the E_T region below trigger threshold it would be true if the difference $R_1 - R_2$ is correct.

Table 5.7 summarizes the rates of observed and predicted negative SECVTX tags in all the available QCD samples. Also shown for comparison, the rate of positive SECVTX tags and the calculated heavy flavor contribution to the negative tag yields plus mistags. The predictions reproduce the observed negative tag yield within 10% for all samples tested. Based on the level of the observed agreement, a 10% systematic error is assigned on the SECVTX mistag rate predictions.

5.1.8 Application of the Jetprobability mistag parametrization on all QCD samples

The reliability of the Jetprobability mistag rate predictions is tested along the same guidelines set for the SECVTX calculations. The same samples are used and all jets in the events including the ones going into the calorimeter cracks are taken into account. The $\not\!\!E_T$ requirements are also relaxed.

Figure 5.18(a) shows the corrected jet E_T distribution of positive Jetprobability tags in the JET_100 sample, compared to the one obtained weighting the jet E_T spectrum with the mistag probabilities. Figure 5.18(b) shows a similar comparison between the observed negative Jetprobability tags and the calculated mistags while

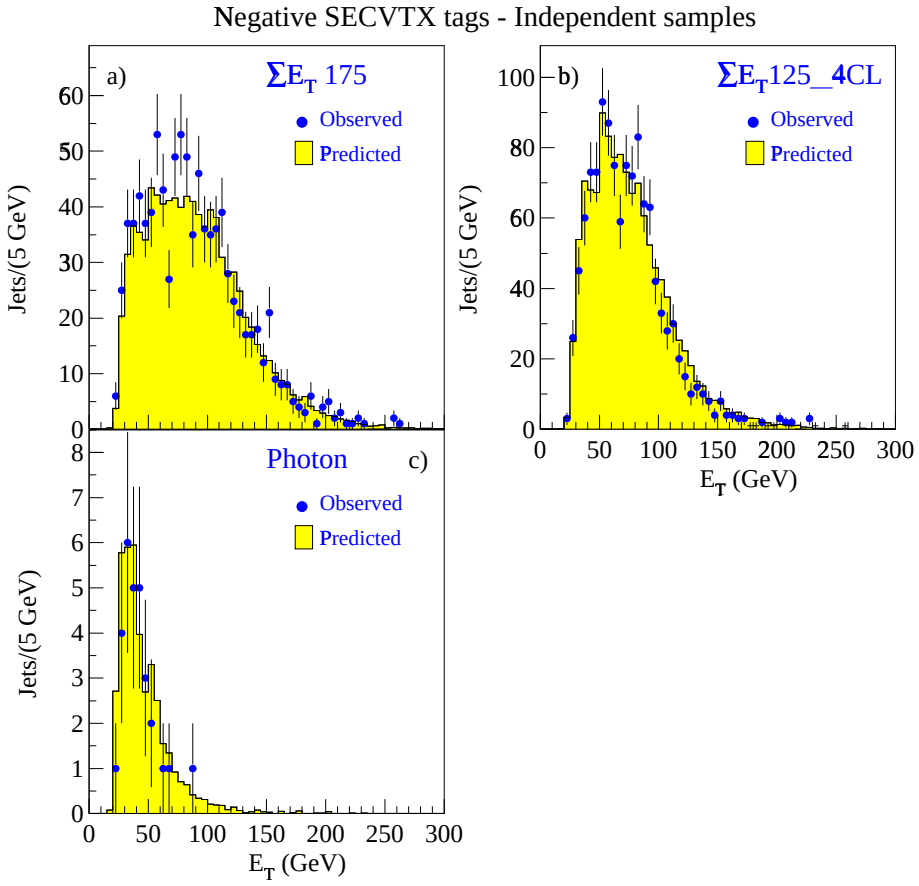

Figure 5.17: Comparisons between the jet corrected E_T distributions of negative SECVTX tags and the ones obtained from the mistag calculations adding the heavy flavor contribution to the negative tags, as observed in QCD samples not used in the mistag rate parametrization.

	Pos. tags	Neg. tags	Mistags	Mistags + heavy flavor
Samples used in the parametrization				
JET_20	4731	699	652	722
JET_50	6874	1648	1426	1695
JET_70	7758	2248	1858	2192
JET_100	8335	2723	2385	2756
$\sum E_T$_300	1507	501	438	521
Independent samples				
$\sum E_T$_175	3790	947	675	908
$\sum E_T$_125_4CL	5637	1203	897	1249
Isolated γ	284	29	35	40

Table 5.7: Summary of the rates of observed and predicted positive and negative SECVTX tags in all available Run 1B QCD samples. The comparison between the second and fourth column shows that the method used to estimate the mistags reproduces the negative SECVTX tag rate within less than 10% in all samples.

Figure 5.18(c) shows a comparison of the expected shape of mistags to the expected shape of jets from heavy flavor decays contributing to the negative tags. Adding the expected mistags and heavy flavor contributions, the calculated negative tag yield agrees very well with the observed yield as shown in Figure 5.18(d).

Figures 5.19 and 5.20 show comparisons of the corrected E_T distributions of observed negative Jetprobability tags in all QCD samples to the ones obtained from the mistag matrix predictions plus the heavy flavor contributions to the negative tags. The calculations reproduce the E_T spectra of the observed negative tags extremely well for all samples.

Table 5.8 summarizes the rates of observed and predicted negative Jetprobability

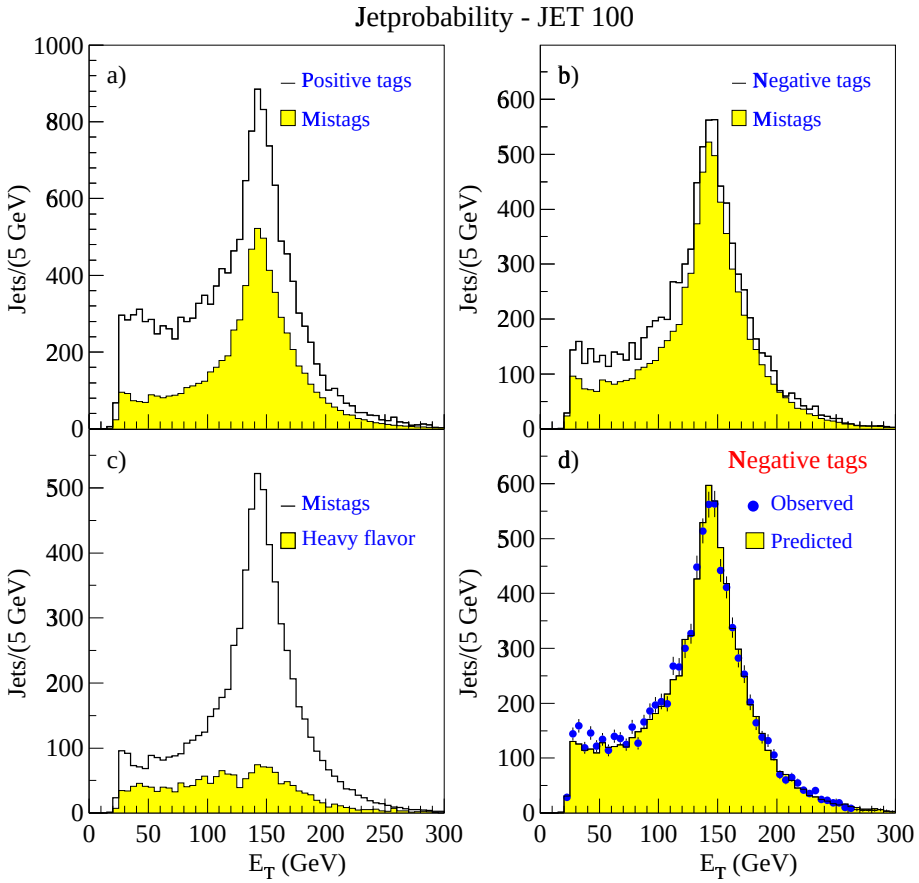

Figure 5.18: Corrected E_T distributions of jets with Jetprobability tags in the JET_100 sample. The plots illustrate the procedure followed to compare the observed negative tagging rate to the one expected using the mistag matrix and adding the heavy flavor contribution. (a): Observed positive tags (open) compared to the calculated mistag rate (shaded). (b): Observed negative tags (open) compared to the calculated mistags (shaded), which does not include heavy flavor contributions. (c): mistags (open) compared to the predicted heavy flavor contribution to the negative tags (shaded). (d): Observed negative tags (points) compared to the mistags plus the heavy flavor contribution to the negative tags (shaded). Very good agreement is observed between the two distributions, compared to the ones shown in (b).

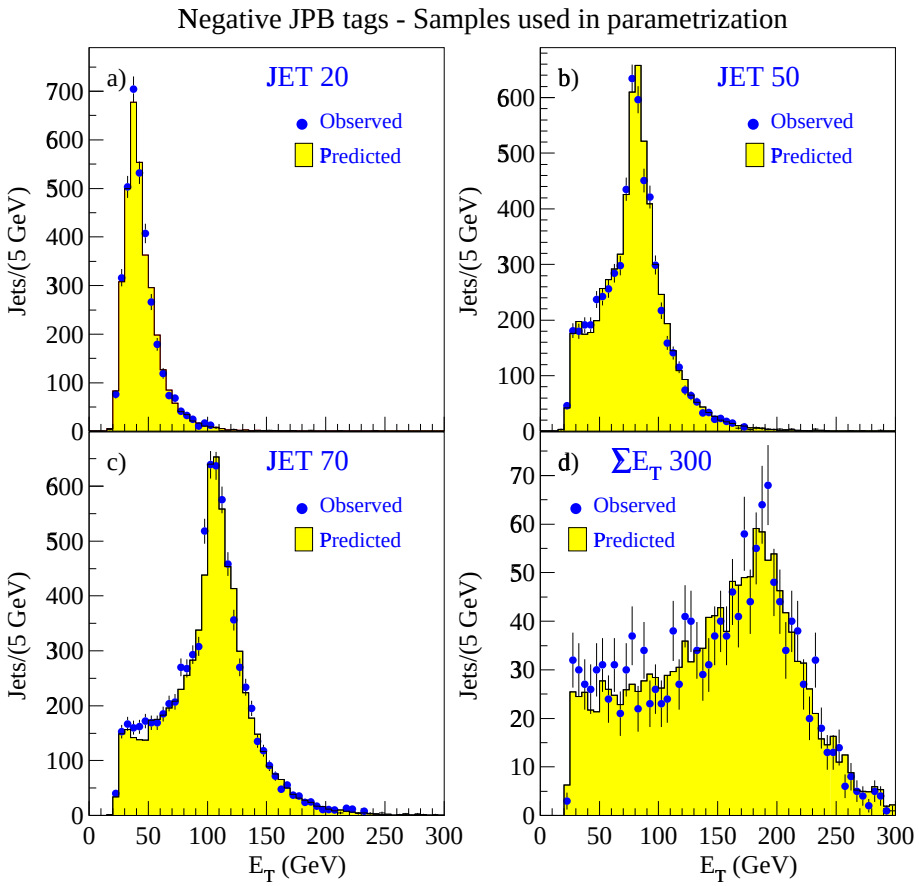

Figure 5.19: The corrected E_T distributions of negative Jetprobability tags compared to the ones obtained from the mistag parametrization plus heavy flavor contribution to the negative tags. The shown distributions correspond to samples used in the derivation of the mistag matrices.

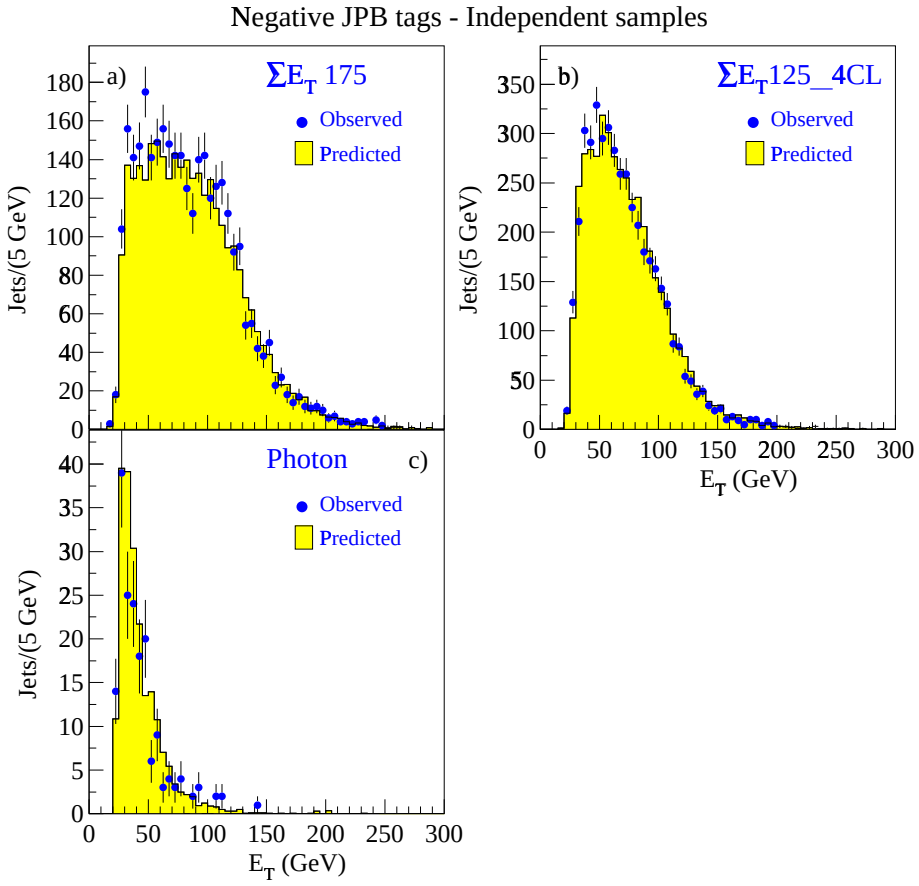

Figure 5.20: The corrected E_T distributions of negative Jetprobability tags compared to the ones obtained from the mistag parametrization plus heavy flavor contribution to the negative tags, for the QCD samples not used in the mistag parametrizations.

tags in all available QCD samples. As for SECVTX, the mistag matrix predictions plus the contribution of heavy flavor decays to the negative tags reproduce the observed negative tag yield within 10% in all the samples tested. Based on the level of the observed agreement, a 10% systematic error is assigned on the Jetprobability mistag rate predictions.

	Pos. tags	Neg. tags	Mistags	Mistags+heavy flavor
Samples used in the parametrization				
JET_20	8418	3414	2919	3421
JET_50	12124	5970	4948	6156
JET_70	13254	7567	6020	7437
JET_100	14528	8827	7010	8721
$\sum E_T$_300	2712	1581	1162	1566
Independent samples				
$\sum E_T$_175	6217	3235	2227	3069
$\sum E_T$_125_4CL	9283	4407	3166	4481
Isolated γ	537	179	176	209

Table 5.8: Summary of the rates of observed and predicted positive and negative Jetprobability tags in all available Run 1B QCD samples. The comparison between the second and fourth column shows that the method used to estimate the mistags reproduces the negative Jetprobability tag rate within less than 10% in all samples.

5.2 The SLT fake rate

The rejection of *fake SLT tags* due to hadrons satisfying the soft electron and muon selection requirements is studied with tracks in generic jet events from a combination of JET_20, JET_50 and JET_70 inclusive QCD triggers. The SLT fake rate is defined [58] as the number of tracks passing the soft lepton selection criteria described in section 4.7 to the total number of tracks which satisfy the soft lepton fiducial requirements. It is extracted separately for electrons and muons. According to the above definition, the SLT tagging rate includes contributions from both fake and real leptons. It can be considered as similar to the positive tagging rate derived for the SECVTX and jetprobability tagging algorithms. The contribution of real SLT tags due to heavy flavor jets is derived in Section 5.2.1.

Since muons are categorized according to the detector the stub is found, the muon fake rate is derived separately for each muon type. Because there is large overlap among the CMU-only, CMP-only and CMUP muon fiducial regions, a combined fake rate is used for these muon types. To account for possible P_T dependence of the background and also for the P_T dependence of the SLT selection requirements, the fake rate is parametrized as a function of the fiducial track P_T.

Muon fake rate

The muon fake rate is defined as the number of soft muon tags (CMU/P or CMX) divided by the total number of tracks extrapolated to the CMU/P or CMX fiducial volume. There is a trigger bias when deriving the soft muon fake rate. The trigger bias appears when a muon is part of a jet. Muons are minimum ionizing particles and hence they can carry a large fraction of the jet momentum without depositing any significant energy in the calorimeter. This means that the energy of the jet containing

the muon is measured systematically low. Because of the exponential distribution of the jet E_T this effect has further implications. Jets that fire the trigger are more likely to be in good fiducial regions of the detector, away from cracks, and therefore their energy is better measured, at least with respect to another jet in the event that it does not fire the trigger. These trigger jets are less likely to contain a muon, since

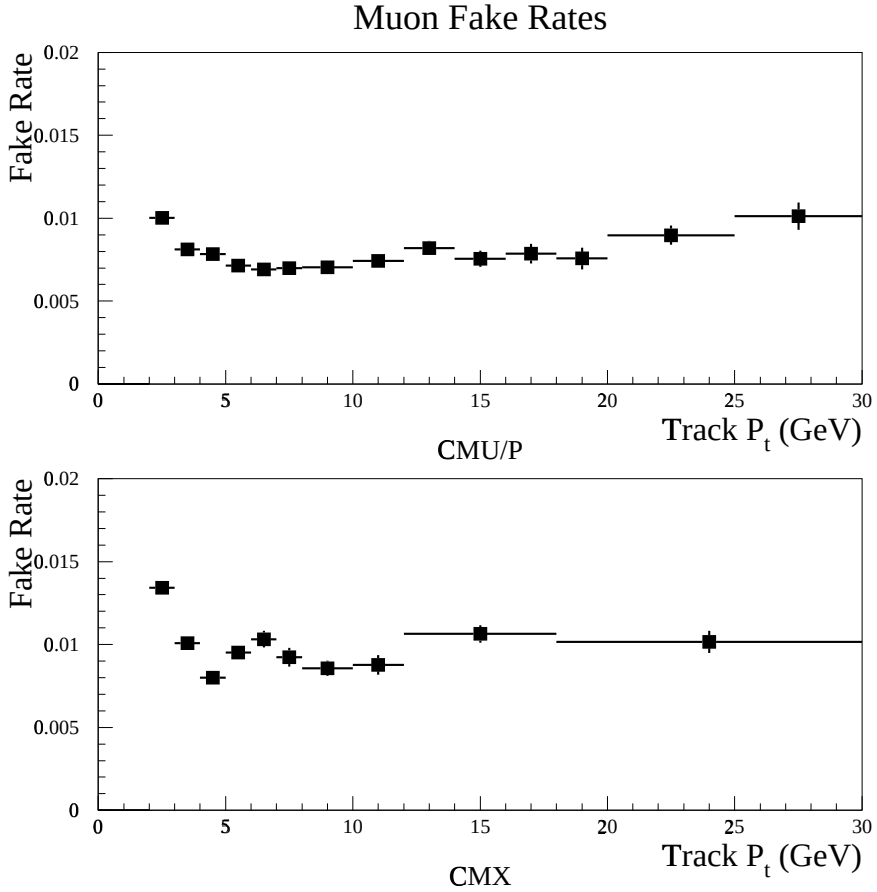

Figure 5.21: Fake rate for CMU/P and CMX soft muon tags as derived from tracks extrapolating to the CMU or CMP (top) and to the CMX (bottom) fiducial regions. The drop in the fake rate with increasing track P_T is due to the tighter criteria applied to the track-muon stub matching variables. The matching is expected to be better for higher P_T muons since they undergo smaller multiple scattering. The kink observed in the CMX fake rate is due to the change of the CMX selection criteria at P_T =5 GeV/c. The plot is from Reference [58].

their energy is well measured. If there were a muon in a trigger jet, then it would have carried away a fraction of the jet's energy and hence the jet would have failed the trigger threshold. According to the same argument, jets below the trigger threshold are more likely to contain a muon. However, the trigger bias is entirely removed, when the fake rate is calculated using tracks well separated ($\Delta R \geq 0.7$) from the jet that triggered the event.

Estimating the muon fake rate without the tracks of the leading jet in the event, it is found to be $\sim 15\%$ higher than the fake rate obtained using all the tracks in the event. The muon fake rate as a function of the track P_T is shown in Figure 5.21. The fake rate is $\sim 1\%$ for tracks with $P_T \sim 2$ GeV/c and decreases for tracks of higher P_T. As the track P_T increases, the effect of multiple scattering for real muons decreases resulting in better muon stub-track matching resolution. This effectively results in tighter stub-track matching selection and consequently smaller fake contribution.

Electron fake rate

Similarly, the electron fake rate is determined by the ratio of the number of soft electron tags found to the number of tracks extrapolated in the CEM fiducial region. Because some the soft electron selection requirements (E_{had}/E_{em}, E/P) depend on the hadronic environment surrounding the candidate track, the electron fake rate is parametrized as a function of the track P_T and relative tracking isolation, $\sum_P^{0.2}/P$. Recall that $\sum_P^{0.2}$ is the scalar sum of the momentum of all other tracks within a cone of 0.2 in $\eta - \phi$ space, and P is the momentum of the track candidate. The electron fake rate is determined for three bins of $\sum_P^{0.2}/P$. Namely, it is determined for $\sum_P^{0.2}/P < 0.2$, $0.2 < \sum_P^{0.2}/P < 5$ and $\sum_P^{0.2}/P > 5$. The measured electron fake rate for the three bins of the $\sum_P^{0.2}/P$ variable is shown in Figure 5.22.

SLT jet fake rate

The SLT fake rate as determined above is based on the *track-tag* rate observed in the QCD data. This fake rate is adequate when the SLT tags are considered independently of whether the soft lepton is found inside a jet or not. This approach was followed by the top quark analysis presented in references [58] and [59]. In this analysis, SLT tags

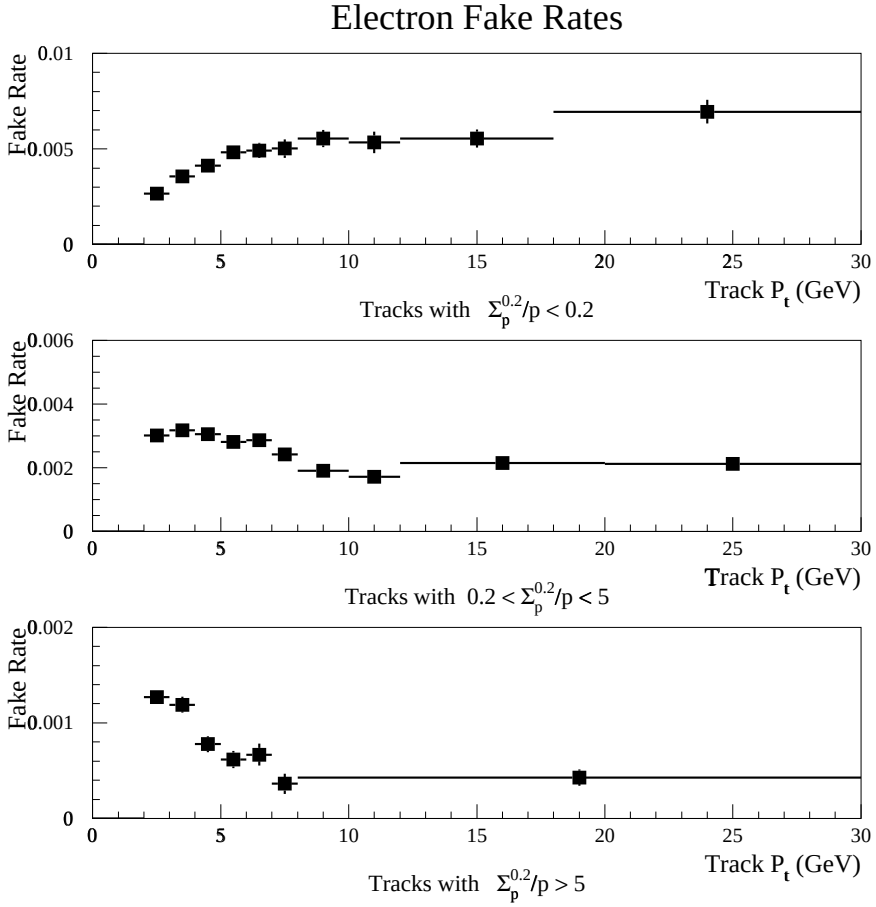

Figure 5.22: Fake rate for soft electrons as a function of the track P_T as derived from tracks extrapolating to the CEM fiducial region. The fake rate is divided in three bins of the $\sum_p^{0.2}/P$ because of the dependence of the E/P and E_{had}/E_{em} selection requirements on the track momentum. The fake rate is decreasing moving to higher values of $\sum_p^{0.2}/P$ because less tracks are likely to satisfy the E/P and E_{had}/E_{em} requirements. The plot is from Reference [58].

are required to lie in a cone of 0.4 about the centroid direction of a jet with $E_T \geq 15$

GeV and therefore the SLT track-tag rate needs to be converted to a *jet-tag* rate. Jets are considered tagged by the SLT algorithm if they contain at least one track satisfying the soft lepton requirements. The sum of the tagging probabilities, P_i, of all the SLT fiducial tracks contained in a cone of 0.4 about the jet axis, is defined to be the SLT jet fake rate,

$$P_{fake}^{jet} \sum_{i=1}^{Ntracks} P_i \tag{5.5}$$

5.2.1 The "true" SLT fake rate

The samples used to extract the SLT fake rate contains events with heavy flavor decays and therefore, the fake rate is inflated by this contribution. Since the method which is used to calculate the $t\bar{t}$ background relies on the exact calculation of the contribution of all Standard Model processes with heavy flavor decays, the heavy flavor content included in the mistag rate should be subtracted to avoid double counting.

The method used to estimate the amount of heavy flavor in the QCD jet samples with SLT tags, uses the signed impact parameter significance distribution (see Section 4.1.2) of tagged SLT tracks in jets from the JET_50 data sample. This distribution is fitted with the shapes of the corresponding distributions of b and c-jets from Monte Carlo simulations and the shape of fake SLT tags. Figure 5.23(a) and (b) shows the signed impact parameter significance distribution of SLT tagged tracks in b and c-jets. These distributions are obtained from simulations of processes which may contain heavy flavor decays like $t\bar{t}$, WW, WZ, ZZ, $Wb\bar{b}$ and $Wc\bar{c}$, and cover a large P_T spectrum of heavy quark decays and different heavy flavor production mechanisms. The distribution for fake SLT tags is derived using tracks in jets from the JET_50 sample requiring that the event does not contain any SECVTX or SLT tags. The anti-tag requirement is imposed to further deplete the heavy flavor content

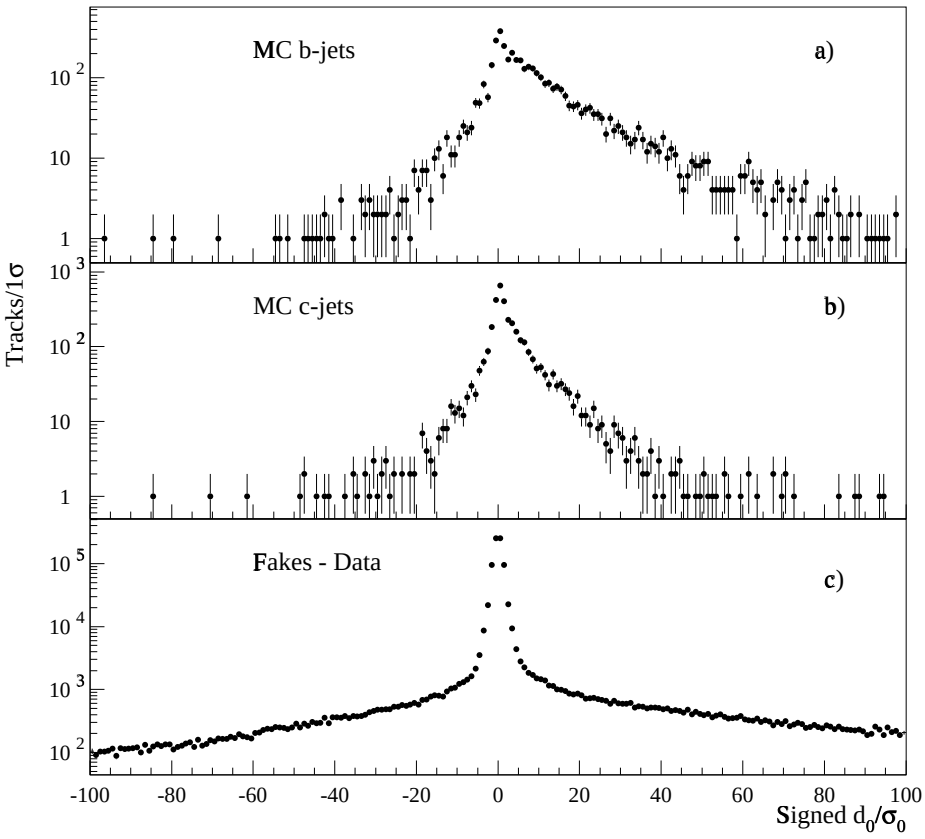

Figure 5.23: Distributions of impact parameter significance of tracks corresponding to SLT tags. The distributions for b-jets (a), and c-jets (b) are from Monte Carlo simulations of any process which may contain heavy flavor decays, like $t\bar{t}$, ZZ, WW and WZ. The distribution of pure fakes (c), is obtained from SLT fiducial tracks contained in jets with no SECVTX or SLT tags. The three distributions are used as templates on the fit to determine the amount of SLT tags due to heavy flavor decays in the generic jet samples.

of the inclusive jet sample. The impact parameter significance distribution of fake SLT tags is shown in Figure 5.23(c). The distribution is expected to be symmetric around zero if there is no heavy flavor contribution because it is then dominated only by resolution effects. However, a small asymmetry is observed towards the positive side of the distribution. This asymmetry is caused by particles from decays in flight, K_s and Λ decays and heavy flavor decays. Monte Carlo studies show that ~4.5% of the jets contributing tracks in the fake SLT distribution originate from heavy flavor decays. Because of the SECVTX anti-selection requirement, tracks from heavy flavor decays are expected mostly in the core of the distribution.

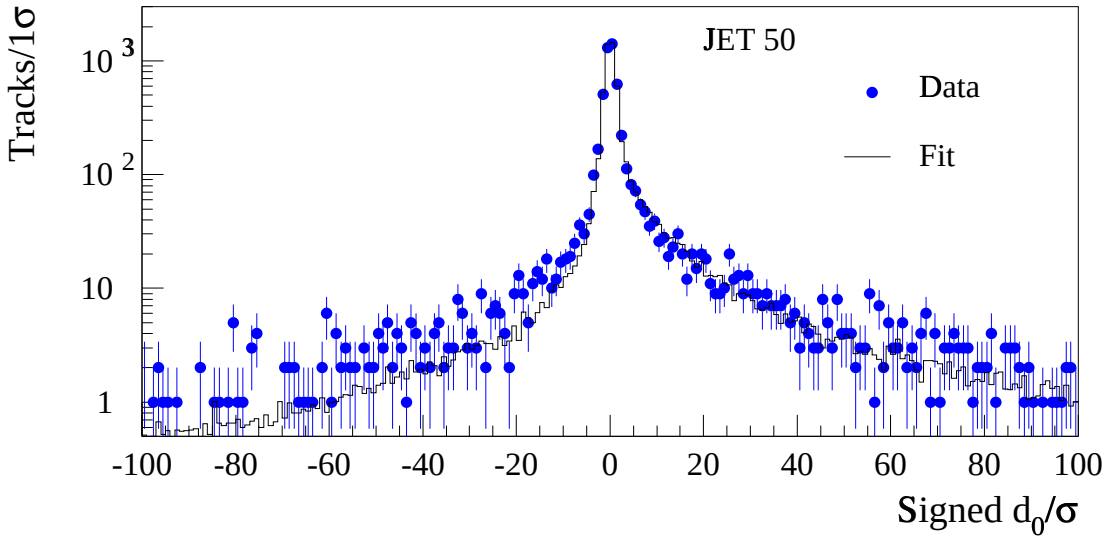

Figure 5.24: Distribution of impact parameter significance of SLT tagged tracks contained in jets with $E_T \geq$15 GeV, in the JET_50 sample. (points). The solid histogram represents the fit-results obtained by fitting the above distribution with the three template distributions shown in Figure 5.23. The fit determines the rate of non-heavy flavor jets tagged by SLT to be 74±3.2% of the total SLT jet tagging rate.

Figure 5.24 shows the signed impact significance distribution of SLT tags in jets of the JET_50 data sample along with the results of the fit. The composition of the SLT tags determined from this fitting procedure is:

- Fakes: $74.0 \pm 3.2\%$
- b's: $10.5 \pm 2.3\%$
- c's: $14.5 \pm 4.3\%$

In order to check for any bias in the fitting procedure due to the slightly asymmetric shape of the fake SLT distribution, the fit was repeated using as shape for the positive impact parameter significance the shape of the negative distribution. In other words, the negative side of the distribution is "flipped" around zero. This is done because the negative side of the distribution is dominated by zero lifetime tracks displaced due to SVX resolution effects. Using this exactly symmetric distribution as a template for fake SLT tags and repeating the fit a value of $72.8 \pm 3.4\%$ for the fake contribution is obtained which is within the errors of the first fit results.

Based on the above procedure the SLT jet mistag rate is rescaled by $74 \pm 3.2\%$ to account for the contribution of heavy flavor decays. This rescaled fake rate corresponds to the *true* SLT mistag rate.

5.2.2 Reliability of the SLT fake rate and systematic uncertainty

The reliability of the SLT fake rate parametrization was tested in Reference [58]. For these tests the data sets of 20, 50 and 70 GeV jet triggers were used. For these checks, the SLT fake rate parametrization obtained using a single jet sample was

applied to the combination of the other two samples. In this manner the examined samples are always independent of the sample used to obtain the parametrization. A comparison of the observed and predicted SLT tag rate based on this method is shown in Table 5.9. Electrons fake rates compare least well across the samples, while muons rates are in agreement to a few percent level. Based on these studies a 10% systematic uncertainty is assigned in the SLT fake rate parametrization according to reference [58].

Additional checks of the SLT fake rate parametrization are performed on a variety of independent data samples. For these tests the SLT jet fake rate is compared to the observed yields. Table 5.10 summarizes the results of the comparison between the predicted and observed SLT yields. Since there is a trigger bias present in the muon yields for the inclusive-jet triggers, the comparison is performed with and without the trigger jet. However, when more than one jets are above the trigger threshold, both jets are examined. When the trigger jet is not included there is good agreement between the observed and predicted jet tag rates. A disagreement appears for the $\sum E_T$ 300 samples. However, this disagreement is also due to the trigger bias because there is no dedicated Level 2 trigger for this sample but only a Level 3 trigger. Therefore, events in the samples are mostly coming through the Level 2 JET_100 trigger. The last three samples shown in Table 5.10 serve as completely independent samples free of any trigger bias. The agreement between predicted and observed rate is fairly good.

Figure 5.25 show the E_T distribution of jets tagged by SLT compared to the one expected from fakes, in the $\sum E_T$ 175 and 300, photon and $\sum E_T$_125_4Cluster samples. The distribution of the SLT fakes is obtained by weighting the E_T of each jet with the expected fake probability for this jet. The jet fake probability is derived

Electron fake rate				
Parametrization	Test Sample	Predicted(P)	Observed(O)	(O-P)/O
JET 20	JET50+JET70	8814 ± 201	6939	−27 %
JET 50	JET20+JET70	6596 ± 114	6790	+3 %
JET 70	JET20+JET50	5912 ± 101	6459	+8 %
Muon fake rate				
CMU/P fake rate				
Parametrization	Test Sample	Predicted(P)	Observed(O)	(O-P)/O
JET_20	JET50+JET70	12019± 219	11782	−2 %
JET_50	JET20+JET70	10128± 138	10015	−1 %
JET_70	JET20+JET50	8886± 112	8939	+0.6 %
CMX				
Parametrization	Test Sample	Predicted(P)	Observed(O)	(O-P)/O
JET_20	JET50+JET70	6091± 147	6283	+3 %
JET_50	JET20+JET70	5459± 101	5428	−1 %
JET_70	JET20+JET50	5004± 87	4879	−3 %

Table 5.9: Comparison of the observed and predicted SLT track-tag rates in the JET_20, JET_50 and JET_70 data samples. For each case, the fake parametrization obtained from one sample is applied to the combination of the other two samples in order to obtain the SLT tag rate predictions in independent samples. The muon fake rate predictions agree with the observed tag rate to a few percent level, while larger variations are observed in the electron tag rates. The table is from reference [58].

Test Sample	Predicted(P)	Observed(O)	(O-P)/O
Samples used in the parametrization			
JET_20	7937.3	7617	−4.2%
JET_50	10227.8	9344	−9.5%
JET_70	11739.7	10653	−10.2%
Without leading jet			
JET_20	4996.6	5087	1.8%
JET_50	7155.5	7245	1.2%
JET_70	8317.8	8385	0.8%
Independent samples			
JET_100	12429.3	10817	−14.9%
JET_100 (no leading jet)	8832.2	8561	−3.2%
$\sum E_T$ 300	2361.1	2026	−16.5%
$\sum E_T$ 175	4910.4	4930	0.1%
$\sum E_T$_125_4CL	15779.5	16755	5.8%
Photon	536.6	521	−3.0%

Table 5.10: Comparison of the observed and predicted jets with a SLT tags The predicted tag rate is based on a track-tag rate parametrization obtained using a mixture of the 20, 50 and 70 GeV jet trigger samples. A trigger bias as explained in the text is present in the 20, 50, 70 and 100 jet samples. Better agreement is observed when comparing the observed and predicted rates without counting the leading jet. Good agreement is observed in the last three samples which are free of any trigger bias. Note the agreement in the photon sample where the heavy flavor content is different than in the inclusive jet samples.

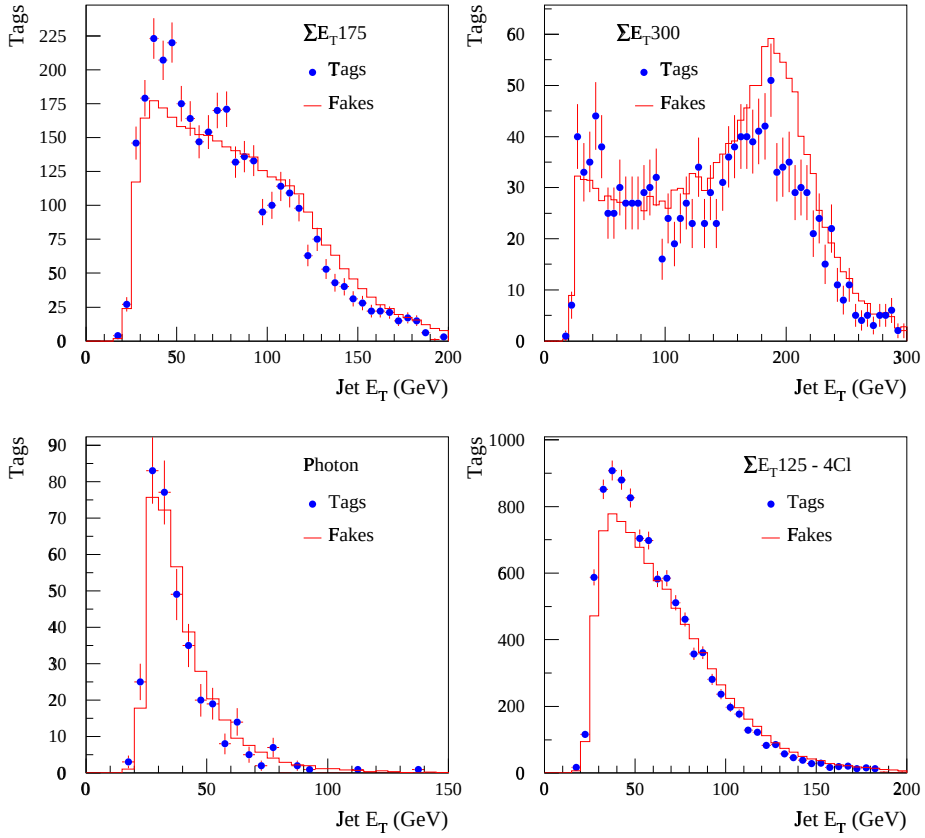

Figure 5.25: Comparison of the E_T distributions of jets tagged by SLT (points) and the corresponding one as predicted by the fake rate (histogram), for the $\sum E_T$ 175 and 300 GeV, the isolated photon and $\sum E_T$_125_4Clusters samples respectively.

from the fake probability of each track in the jet (equation 5.5). Recall that the SLT fake rate is a track-based fake rate and therefore there is no dependence on the jet E_T. However, it is useful to compare the observed and predicted SLT tagged jet E_T spectra. This comparison shows whether the derived track-based parametrization reproduces the jet shape, especially for high-E_T jets, in a sensible matter and that there are no additional biases. Fair agreement is observed between the predicted and observed jet E_T distributions.

5.3 A last check - Comparison of the excess of positive tags in data and in HERWIG simulations

As a final check, the difference between the positive tags and mistags observed in the inclusive jet data is compared to the tag yields from heavy flavor hadrons in inclusive jet HERWIG Monte Carlo simulations. Three HERWIG samples were generated with minimum parton P_T of 10, 40 and 80 GeV/c in order to simulate the corresponding Level 2 jet trigger for the JET_20, JET_50 and JET_100 samples. Events in the data and simulation samples are selected requiring at least one jet above trigger threshold and the presence of at least one taggable jet (a jet with $E_T \geq 15$ GeV, $|\eta_D| \leq 2$ containing at least two good SVX tracks). The HERWIG predictions for heavy flavor production via direct production, flavor excitation and gluon splitting are rescaled according to the normalization factors derived in Section 6.6.3.

Table 5.11 lists the positive tag yields observed in each sample, along with the estimated mistags as calculated with the mistag matrix. The SLT mistags are calculated using the jet fake probabilities scaled by 74% to account for the heavy flavor contribution to the fake rate. The last two columns in Table 5.11 compare the excess of positive tags, defined as the difference between the observed positive tags and mistags, to the HERWIG positive tag yields. Tags in the simulation samples are counted if the corresponding jets are associated with a heavy flavor hadron at generation level within a cone of 0.4 in $\eta - \phi$ space.

Excellent agreement is observed between the excess of positive tags of each tagging algorithm in the data and in the HERWIG simulations.

JET 20 (194009 events)				
	DATA			HERWIG
Tagger	Positive tags(P)	Mistags(M)	P - M	Positive tags
SECVTX	4674	616	4058±92	4052±108
JPB	8343	2801	5542±295	5573±116
SLT	4994	3962	1032±402	826±45
JET 50 (151270 events)				
	DATA			HERWIG
Tagger	Positive tags(P)	Mistags(M)	P - M	Positive tags
SECVTX	6536	1360	5176±158	5314±72
JPB	11533	4700	6833±482	6740±73
SLT	6408	5241	1167±530	1116±31
JET 100 (129434 events)				
	DATA			HERWIG
Tagger	Positive tags(P)	Mistags(M)	P - M	Positive tags
SECVTX	7682	2227	5455±239	5889±112
JPB	13365	6494	6871±659	7263±113
SLT	7483	6367	1116±642	1160±46

Table 5.11: Number of tagged jets in inclusive QCD jet data and simulations. The amount of SECVTX and Jetprobability mistags and SLT fakes are calculated with the derived mistag parametrizations which do not include heavy flavor contributions. The errors on the data positive excess are calculated assuming 10% systematic error on the mistag predictions. The HERWIG predictions and errors are according to the normalization factors derived in Section 6.6.3. Tags in the simulation are related to true heavy flavor jets.

Chapter 6

Heavy flavor composition of the $W+ \geq 1$ jet sample

Tagged events in the $W+ \geq 1$ jet sample can be classified into two categories. Events that contain real heavy flavor tags and events that contain fake tags or mistags.

The most relevant to this analysis mechanism contributing heavy flavor jets to the $W+ \geq 1$ jet sample is the $t\bar{t}$ decay which results in a final state containing a pair of W's and a pair of b quarks. As already discussed, $t\bar{t}$ events are expected to contribute the most in the $W+ \geq 3$ jet region and it naturally consists the search region for $t\bar{t}$ decays. Also the production and decay of a single top quark via $W-$gluon fusion or $q\bar{q}$ annihilation contributes heavy flavor jets mainly in the low $W+$ jet multiplicity bins ($W+1$ and $W+2$ jet bins).

Jets containing heavy flavor hadrons can also be produced in association with a W boson either via the standard QCD mechanisms of $W+$ jets production or via the decay of a resonance produced in association with a W as it is the case of WW, WZ production. In the case of QCD $W+$jets production, heavy flavor jets can emerge

either at a Wcs vertex ($p\bar{p} \rightarrow Wc$) or they are produced when a final state gluon splits into a pair of heavy flavor quarks ($g \rightarrow b\bar{b}$ and $g \rightarrow c\bar{c}$). In the case of WW and WZ production, heavy flavor jets can be produced in the hadronic decay of one of the bosons ($W \rightarrow cs$ and $Z \rightarrow b\bar{b}(c\bar{c})$). There are also processes that contribute real tags to the $W+ \geq 1$ jet sample but do not include any real W at the final state. The QCD $Z+$ heavy flavor production, $Z \rightarrow \tau^+\tau^-$ where one of the τ's can be tagged, $b\bar{b}$ production and ZZ production are sources of non-W events with heavy flavor jets at the final state. These events enter in the $W+ \geq 1$ jet sample when the leptonic decay of one of the partons or bosons is combined with a jet mismeasurement due to a hadronic shower fluctuation or detector effects resulting in a final state that can satisfy the $W+$ jets selection criteria.

As discussed in details in Chapter 5, mistags and fakes arise from track mismeasurements or hadrons misidentified as leptons, in light quark or gluon jets in all of the above mentioned processes. The contribution of the mistags to the $W+ \geq 1$ jet tagged sample is estimated with the use of the mistag probability matrices.

The knowledge of the $W+ \geq 1$ jet tagged sample composition is important not only for the search for $t\bar{t}$ decays but also for searches of more exotic processes. Such processes involve either the production of Higgs in association with a W [66] or the production of a new particle via Technicolor interactions [67]. The production cross sections for such interactions ranges between 0.1 and 10 pb for the Tevatron energies. Both processes result in the same final state events consisting of a W boson and a pair of heavy flavor quark jets ($p\bar{p} \rightarrow W^* \rightarrow W + H \rightarrow W + b\bar{b}$ and $p\bar{p} \rightarrow \rho_T^{\pm} \rightarrow W\pi_T^0 \rightarrow Wb\bar{b}$). However, the search for such exotic processes can be performed after the contribution of all previously mentioned Standard Model processes are properly calculated and the $t\bar{t}$ cross section is measured. Any discrepancy between the observed

and predicted tagging yields in the $W+\geq 1$ jet sample can then be used as the initiator for searches of new physics processes or re-evaluation of the procedure used to calculate the contribution of the different Standard Model processes. For the remaining of this analysis, contributions to the $W+\geq 1$ jet tagged sample from Higgs production or other non-Standard Model processes are not taken into account.

The following sections of the chapter describe in details the method used to estimate the contribution of each Standard Model process to the $W+\geq 1$ jet sample. Initially, the contribution of each individual process is estimated under the assumption that there no contribution from $t\bar{t}$ events in the $W+\geq 1$ jet sample. The contribution of each individual process to the $W+\geq 1$ jet tagged sample is estimated solely from Monte Carlo simulations of each individual process. Where possible, the Monte Carlo predictions are compared to data and proper normalization factors are derived to adjust the Monte Carlo predictions. Large Monte Carlo samples of each contributing process were generated and processed with full detector simulations. All generated events before detector simulation were interfaced with the CLEO [76] Monte Carlo simulation in order to properly model the heavy flavor decay branching fractions and decay kinematics. The Monte Carlo simulations include the simulation of the muon trigger path and the relevant Data/Monte Carlo scale factor for the lepton identification efficiency. Also, in order to propagate to the tagging efficiency the effect of the reduced reconstruction efficiency of tracks inside jets, the track degradation procedure is applied to all Monte Carlo samples. Finally, the tagging yields measured in the simulations of the different processes, are adjusted according to the measured Data/Monte Carlo tagging efficiency scale factor for b and c-hadron jets.

This method of explicitly calculating the heavy flavor contribution of each individual process to the $W+\geq 1$ jet sample is known as *Method* II [59] as opposed to

Method I calculation which was used in the $t\bar{t}$ cross section measurement with SLT tags presented in References [59, 23, 58]. Method I estimates the contribution of heavy flavor produced in association with a W or a Z based on the positive tagging probability matrices derived from the QCD generic samples. A discussion of the advantages and disadvantages of the two methods of estimating the heavy composition of the $W+ \geq 1$ jet sample is presented in Section 6.6.2.

The $t\bar{t}$ contribution is calculated in Chapter 8 based on the derived composition of the $W+ \geq 1$ jet sample which includes only contribution from Standard Model processes. Any measured excess of tagged events in the $W+ \geq 3$ jet is attributed to $t\bar{t}$ events and it is used to calculate the $t\bar{t}$ cross section. After including the top contribution, the heavy flavor content of the $W+ \geq 1$ jet sample is recalculated based on the new composition of the pretagged $W+ \geq 1$ jet sample. While the $W+ \geq 3$ jet sample is used to evaluate the $t\bar{t}$ contribution, the $W+1$ and $W+2$ jet bins are used as control sample to check whether the procedure used to estimate the heavy flavor contribution of the all the other processes is correct.

6.1 Contribution from non-W events

This class of events consists of events in which the primary lepton does not originate from a W decay but it is fake. The presence of a fake lepton in addition to mismeasurement or complete loss of another jet in the event produces significant $\not\!\!E_T$ and thus fake W signature. These events are referred to as *non-W* events. The observed fake or real lepton most often is accompanied by large hadronic activity from the remaining parton fragmentation products and therefore is less isolated than leptons from W decays. Non-W events can also contribute tags to the $W+ \geq 1$ jet sample. In

cases where a gluon splits to a $b\bar{b}$ or a $c\bar{c}$ pair, or there is direct $b\bar{b}$ or $c\bar{c}$ production, one of the heavy flavor jets can produce a high P_T lepton, while the other jet can be tagged. Part of the non-W tagging rate arises also from mistagging a light quark or a gluon jet. Therefore, the calculation of the amount of non-W tagged events is considered as the integral of the contributions from all hadronic non-W sources.

The non-W contribution to the tagged $W+\geq 1$ jet sample, is extracted from the data. The calculation is divided in two parts. First, the fraction of non-W events, F_{non-W}, on the pre-tagging $W+\geq 1$ jet sample is estimated and at the second step the tagging rate, ϵ^{tag}_{non-W}, of non-W events is determined for each tagging algorithm. The expected number of tagged events in each jet multiplicity bin, is calculated using the following expression:

$$N^{tag}_{non-W} = N_W \times F_{non-W} \times \epsilon^{tag}_{non-W}$$

where, N_W is the number of observed W candidate events before tagging in each jet multiplicity bin.

The non-W fraction before tagging is estimated by extrapolating the lepton isolation from a region away from the W signal into the W signal region. In practice this is done by dividing the Isolation versus $\not\!\!E_T$ plane in four regions, as shown in Figure 6.1. Region **A** corresponds to events with $\not\!\!E_T \leq$10 GeV and lepton isolation, Iso$^{lep} \leq 0.1$; region **B** corresponds to events with $\not\!\!E_T \leq$10 GeV and Iso$^{lep} \geq$0.1; region **C** corresponds to events with $\not\!\!E_T \geq$20 GeV and I$^{lep} \geq$0.1; finally region **D** corresponds to the W signal region ($\not\!\!E_T \geq$20 GeV and Iso$^{lep} \leq$0.1) and includes contamination from the non-W contribution.

Under the assumption that $\not\!\!E_T$ and Isolation are uncorrelated for the hadronic

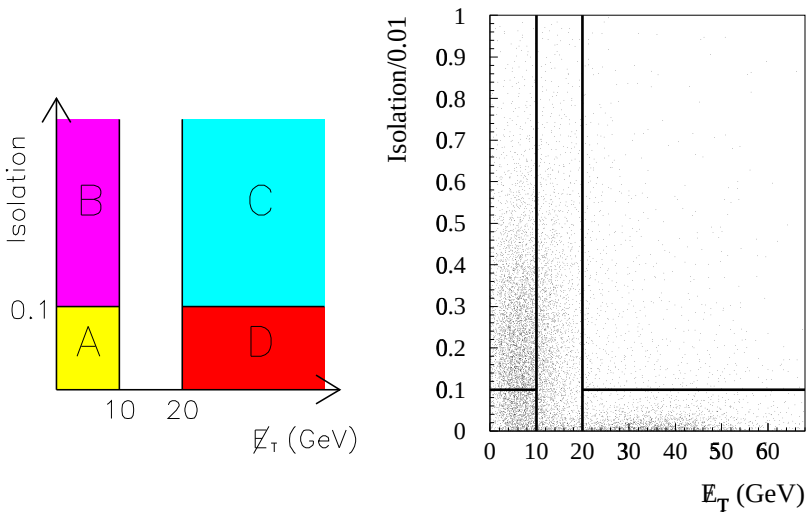

Figure 6.1: The four regions in the Iso^{lep} vs $\not\!\!E_T$ plane used to estimate the non-W contribution to the $W+\geq 1$ jet sample (a). Region A, corresponds to $\not\!\!E_T$ $\leq$10 GeV and Iso^{lep} $\leq$0.1, region B to $\not\!\!E_T$ $\leq$10 GeV and Iso^{lep} $\geq$0.1, region C to $\not\!\!E_T$ $\geq$20 GeV and Iso^{lep} $\geq$0.1 and finally region D corresponds to the W signal region of $\not\!\!E_T$ $\geq$20 GeV and Iso^{lep} $\leq$0.1. Isolation distribution of primary leptons from the inclusive lepton sample versus the $\not\!\!E_T$ in the event is shown in (b) along with the boundaries of the four regions.

events, the amount of non-W events in the signal region can be estimated from

$$N_{non-W} = N_C \times \frac{N_A}{N_B}$$

where N_A, N_B and N_C are the number of events in regions A, B and C. The non-W fraction is then given by $F_{non-W} = \frac{N_{non-W}}{N_D}$, where N_D is the number of events in the W signal region. Since $Z \rightarrow \ell\ell$ and Drell-Yan events are sources of isolated leptons and in general there is small $\not\!\!E_T$ associated with these events, they populate region A. Including these events in region A will overestimate the non-W event fraction, thus

they are removed from the sample according to Z and dilepton removal procedure described in sections 3.6.1 and 3.6.2.

Figure 6.2(a) and (c) show the lepton isolation distributions for events with $\not\!\!E_T$ $\geq$20 GeV (solid) and $\not\!\!E_T$ $\leq$10 GeV (points) normalized to the observed ratio of Region C/Region B for ℓ+1 jet and $\ell+$ $\geq$2 jets respectively. The analogous dis-

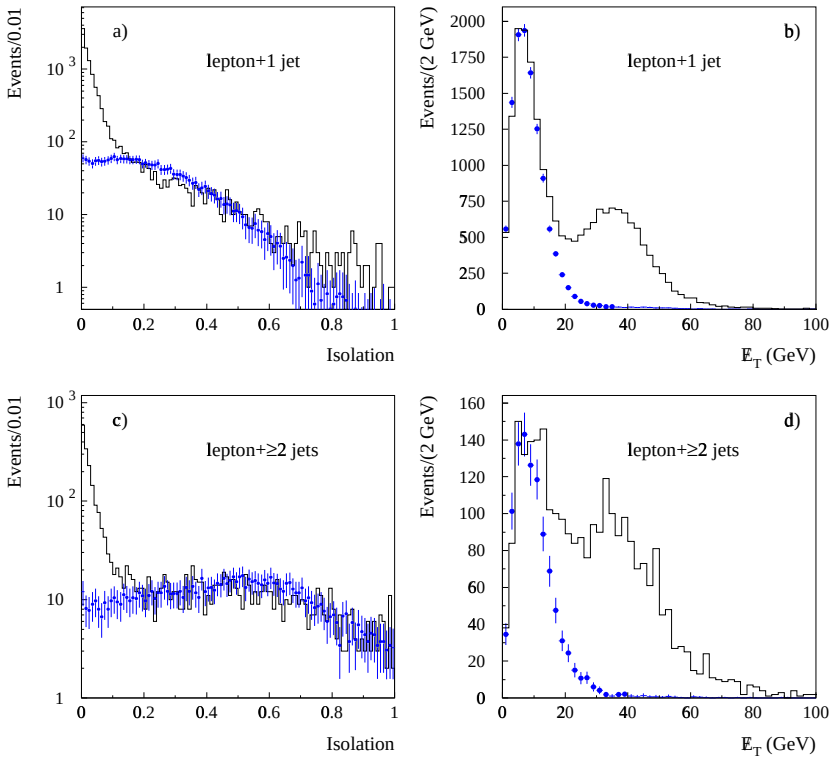

Figure 6.2: The lepton isolation distribution for events with $\not\!\!E_T$ $\geq$20 GeV (solid) compared to the isolation distribution of events with $\not\!\!E_T$ $\leq$10 GeV (points) normalized to the ratio C/B for ℓ+1 jet (a) and $\ell+$ $\geq$2 jet events (c). Also shown, are the $\not\!\!E_T$ distribution of events with Iso^{lep} $\leq$0.1 (solid) compared to the $\not\!\!E_T$ distribution of events with Iso^{lep} $\geq$0.1 (points) normalized to the A/B ratio, for ℓ+1 jet (b) and $\ell+$ $\geq$2 jet events (d) respectively. In all cases, there is good agreement between the compared distributions in the regions enriched in hadronic events ($\not\!\!E_T$ $\leq$10 GeV or Iso^{lep} $\geq$0.1) indicating that $\not\!\!E_T$ and Iso^{lep} are uncorrelated for this class of events.

tributions for $\not\!\!E_T$ are displayed in Figure 6.2(b) and (d), where the $\not\!\!E_T$ distribution for events with Iso^{lep} $\leq$0.1 (solid) is compared to the $\not\!\!E_T$ distribution of events with Iso^{lep} $\geq$0.1 (points) normalized to the ratio of Region B/Region A. The good agreement in the hadronic event rich regions ($\not\!\!E_T$ $\leq$10 GeV or Iso^{lep} $\geq$0.1) indicate that the isolation and $\not\!\!E_T$ are to a good approximation uncorrelated for non-W hadronic events.

Table 6.1 lists the estimated fraction of non-W events as a function of jet multiplicity in the event.

F_{non-W}	$W+1$ jet	$W+2$ jets	$W+3$ jets	$W+\geq$4 jets
Before tagging	5.92$\pm$0.2%	5.20$\pm$0.2%	6.28$\pm$1.0%	9.45$\pm$3.1%
SECVTX	1.46$\pm$0.1%	2.09$\pm$0.7%	6.15$\pm$3.1%	6.15$\pm$3.1%
Jetprobability	2.24$\pm$0.2%	3.55$\pm$0.9%	4.61$\pm$2.7%	4.61$\pm$2.7%
SLT	1.60$\pm$0.2%	2.93$\pm$0.8%	3.08$\pm$2.2%	3.08$\pm$2.2%
Events	$W+1$ jet	$W+2$ jets	$W+3$ jets	$W+\geq$4 jets
Before tagging	560.4$\pm$14.9	71.3$\pm$2.7	12.4$\pm$2.0	5.1$\pm$1.7
SECVTX	8.18$\pm$0.78	1.49$\pm$0.47	0.76$\pm$0.38	0.31$\pm$0.16
Jetprobability	12.55$\pm$0.95	2.53$\pm$0.61	0.57$\pm$0.33	0.24$\pm$0.14
SLT	8.97$\pm$0.84	2.09$\pm$0.56	0.38$\pm$0.27	0.16$\pm$0.11

Table 6.1: The (%) fraction of non-W events in the pre-tagging $W+\geq$1 jet sample and the non-W tagging rate for events with SECVTX, Jetprobability and SLT tags as determined from region A. The product of the number of W candidate events in each jet multiplicity bin, times the fraction of non-W events shown in the first raw times the non-W tagging rate shown in the rows 2$\div$4 predicts the amount of non-W tagged events shown in the last rows of the table.

Since the determination of the non-W contribution before tagging is relying on the

assumption that non-W events are uncorrelated in $\not\!\!E_T$ and isolation space, the non-W contribution after tagging could be estimated using the same technique. However, the heavy flavor content of non-W events depends on the lepton isolation. Indeed such a dependence is observed as shown in Figure 6.3 where the tagging rate is plotted versus the isolation of the primary lepton for ℓ+1 jet events with $\not\!\!E_T$ $\leq$10 GeV and $\not\!\!E_T$ $\geq$10 GeV. This dependence can be understood from the fact that region B is a $b\bar{b}$-enriched sample. When the away jet is tagged, then the lepton is more likely to be the product of the other b's semileptonic decay and therefore is less isolated. For this reason, events from region A which has the same isolation distribution as the W signal region is used for the calculation of the non-W tagging efficiency. Based on the number of tagged events in region A the tagging rate is given by $\epsilon^{tag}_{non-W} = \frac{N_A^{tags}}{N_A}$. Using this tagging rate, the fraction of non-W events, F_{non-W}, in the W region and the number of W candidate events, N_D, the expected non-W tagged events in every jet multiplicity bin, is estimated with the following expression:

$$N^{tag}_{non-W} = N_D \times F_{non-W} \times \epsilon^{tag}_{non-W}$$

The procedure is verified by using the tagging rate measured in region B to predict the number of tagged events observed in region C.

Table 6.1 summarizes the non-W tagging rate for each tagging algorithm as a function of jet multiplicity along with the final predictions for the expected non-W tagged contribution to the $W+$ $\geq$1 jet sample. Because of low statistics in the $\geq$3 jet multiplicity bins, the tagging rate in region A is calculated from the combination of the two bins and the result is assigned to both bins.

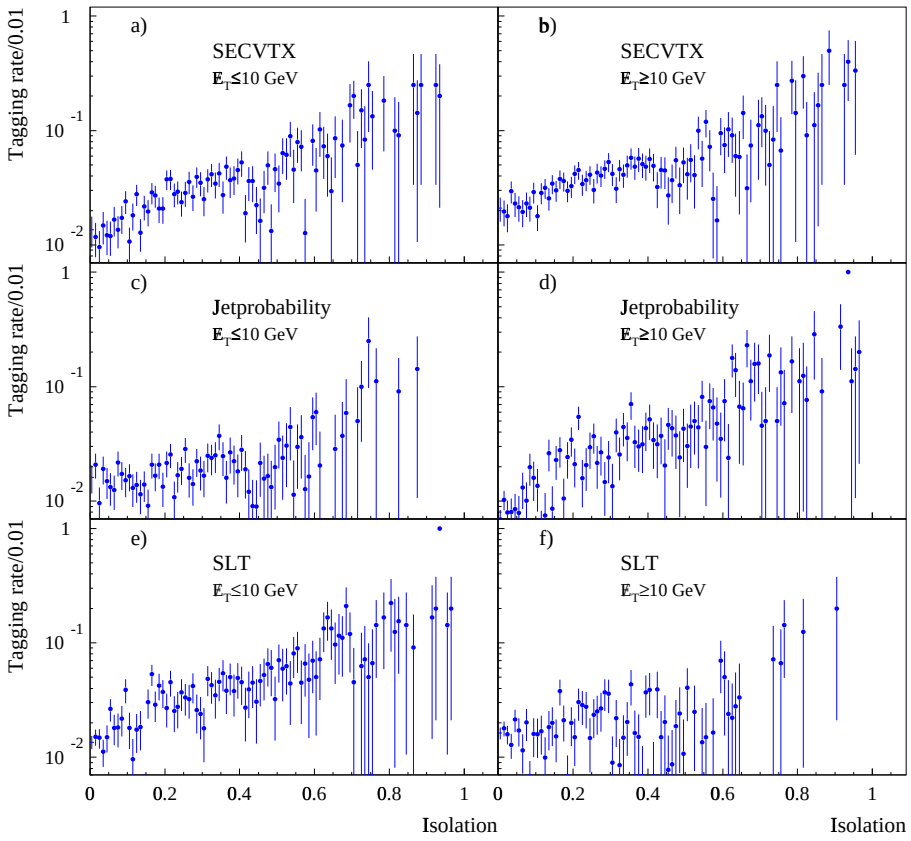

Figure 6.3: Tagging rates (%) as a function of the isolation of the primary lepton in ℓ+1 jet events. The SECVTX tagging efficiencies versus Iso^{lep} for events with $\not\!\!E_T \leq$10 GeV (a) and $\not\!\!E_T \geq$10 GeV (b). The corresponding distributions for Jetprobability and SLT are shown in (c) and (d) and (e) and (f) respectively. A dependence of the tagging rate on the isolation of the primary lepton is observed for all tagging algorithms.

6.2 Contribution from $Z \rightarrow \tau^+\tau^-$

The τ leptonic decays and its relatively long lifetime ($\sim$0.3 psec) are the reasons for examining this process separately from the rest of the $Z+N$ jet contribution. Events of this type, enter in the $W+\geq$1 jet sample when one of the τ's decays leptonically to a high P_T lepton, which satisfies the primary lepton criteria, while the three neutrinos produced in the decay can give rise to significant $\not\!\!E_T$. Depending on the decay pattern of the other τ, its decay product can be tagged by any of the tagging algorithms. In

cases where the second τ decays leptonically it can be identified by the SLT algorithm while when it decays hadronically to three or more $\pi^{\pm}$'s the resulting jet can be tagged by either the SECVTX or the Jetprobability algorithm.

The $Z \rightarrow \tau\tau$ contribution to the $W+ \geq 1$ jet sample is estimated using a PYTHIA Monte Carlo sample of $\sim$300K $Z \rightarrow \tau^{+}\tau^{-}+ \geq 1$ jet events. The normalization of the Monte Carlo sample to the data is derived with the following expression:

$$Norm_{Z\rightarrow\tau\tau} = Br(Z \rightarrow \tau\tau) \cdot \frac{N_{gen}^{Z\rightarrow X}}{N_{gen}^{Z\rightarrow\tau\tau}} \cdot \frac{N_{data}^{Z\rightarrow\ell\ell}}{N_{Z\rightarrow X}^{Z\rightarrow\ell\ell}}$$

where,

- Br($Z \rightarrow \tau\tau$)=0.0336, is the branching fraction of $Z \rightarrow \tau\tau$ [81].
- $N_{gen}^{Z\rightarrow X}$ is the number of generated events of an inclusive $Z+ \geq 1$ jet PYTHIA Monte Carlo sample.
- $N_{gen}^{Z\rightarrow\tau\tau}$ is the number of generated events of the $Z \rightarrow \tau\tau+ \geq 1$ jet PYTHIA Monte Carlo sample.
- $N_{Z\rightarrow X}^{Z\rightarrow\ell\ell}$ is the number of $Z \rightarrow \ell\ell+ \geq 1$ jet events identified in the inclusive Z Monte Carlo sample.
- $N_{data}^{Z\rightarrow\ell\ell}$ is the number of $Z \rightarrow \ell\ell+ \geq 1$ jet events identified in the data.

The contribution to the $W+ \geq 1$ jet tagged sample, is estimated requiring the tag to be associated with the τ lepton in the SLT and SECVTX case. For the Jetprobability case no association requirement is imposed in order to properly account for mistags from this process.

The expected number of events from this process contributing to the $W \geq 1$ jet sample is summarized in Table 6.2

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
Before tagging	35.2±2.2	13.1±1.2	1.6±0.4	0.3±0.2
with SECVTX tag	0.96±0.30	0.70±0.25	0.17±0.12	0.00±0.00
with JPB tag	2.35±0.47	1.13±0.32	0.17±0.12	0.09±0.09
with SLT tag	0.54±0.23	0.09±0.09	0.09±0.09	0.00±0.00

Table 6.2: Summary of the expected number of $Z \rightarrow \tau^+\tau^- + \geq 1$ jet events contributing to the $W+\geq 1$ jet sample before and after tagging, as a function of the jet multiplicity.

6.3 Contribution from single top production

Single top production was discussed in Section 1.3.2. There are two processes which produce a single top quark, rather than a $t\bar{t}$ pair: the $W-$gluon fusion process shown in Figure 6.4(a) and $q\overline{q} \rightarrow t\overline{b}$ via a W^* as shown in Figure 6.4(b).

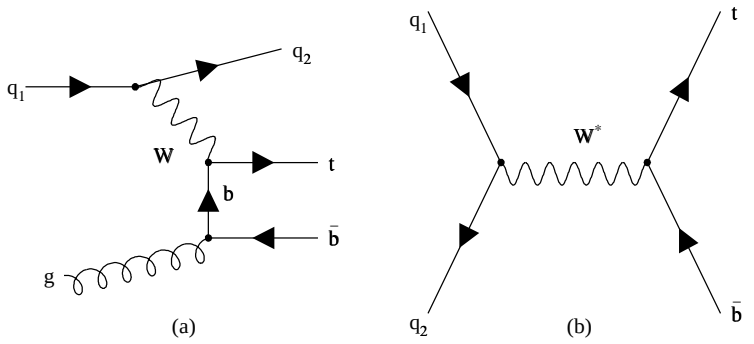

Figure 6.4: Leading order diagrams contributing to single top production at the Tevatron. (a): $W-g$ fusion; (b): quark-antiquark annihilation.

The single top contribution is estimated from Monte Carlo simulations. Two

different generators were used to measure the acceptance and tagging efficiency of single top events produced via the two channels. Single top production via $W-$gluon fusion was simulated using a sample of $\sim$100K events generated with the HERWIG Monte Carlo generator. The luminosity of the Monte Carlo sample is normalized to the data using a cross section for $W-g$ fusion of σ_{W-g}=1.8$\pm$0.5 pb, according to the theoretical calculations presented in Reference [83]. The theoretical cross section includes a 30% systematic uncertainty due to the uncertainty in the gluon distribution function.

The $q\overline{q} \rightarrow W^* \rightarrow t\overline{b}$ annihilation process was simulated with a sample of $\sim$200K events generated with the PYTHIA [70] Monte Carlo. The decay of $W^* \rightarrow t\overline{b}$ was selected among all possible W^* decay channels. Using the cross section value of $\sigma_{W^* \rightarrow t\bar{b}}$=0.74$\pm$0.045 pb as calculated in [83] and the Br($W^* \rightarrow t\overline{b}$) branching faction measured in the simulation, the Monte Carlo sample is normalized to the luminosity of the data.

Sample	Acceptance(%)		Tagging efficiency (%)					
			SECVTX		Jetprobability		SLT	
	W^*	W_g	W^*	W_g	W^*	W_g	W^*	W_g
$W+1$ jet	1.07	2.39	4.1	24.9	4.5	22.8	1.2	6.9
$W+2$ jet	0.33	3.02	30.5	30.0	26.1	28.8	9.3	7.9
$W+3$ jet	0.07	0.67	39.4	33.8	35.4	33.4	9.4	10.8
$W+\geq 4$ jet	0.02	0.08	46.6	42.7	33.1	39.5	17.1	13.6

Table 6.3: Acceptance (in %) and tagging efficiency (in %) for single top events produced via $W-g$ fusion and $q\overline{q}$ annihilation.

Table 6.3 summarizes the acceptance for single top events for both $W-g$ fusion

and $q\bar{q}$ annihilation, as a function of the jet multiplicity. The shown acceptances include the Data/Monte Carlo scale factor for the lepton id efficiency and the trigger efficiency (see Chapter 7). The event tagging efficiency in single top events is determined from the fraction of events containing at least one tagged event over the total number of events passing the analysis cuts in each jet multiplicity bin. The calculated tagging efficiency is then scaled according to the measured Data/Monte Carlo tagging efficiency scale factor. The resulting event tagging efficiencies for each production process and each tagging algorithm are summarized in Table 6.3.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
W^*				
Before tagging	10.40±1.73	3.17±0.53	0.69±0.13	0.18±0.04
with SECVTX tag	0.42±0.10	0.97±0.20	0.27±0.07	0.08±0.03
with JPB tag	0.47±0.10	0.83±0.17	0.24±0.06	0.06±0.02
with SLT tag	0.10±0.03	0.30±0.07	0.07±0.02	0.03±0.01
$W-g$ fusion				
Before tagging	4.70±1.49	5.92±1.88	1.31±0.42	0.16±0.05
with SECVTX tag	1.17±0.39	1.77±0.60	0.44±0.15	0.07±0.03
with JPB tag	1.07±0.36	1.70±0.57	0.44±0.15	0.06±0.02
with SLT tag	0.34±0.12	0.46±0.15	0.14±0.05	0.02±0.01

Table 6.4: Expected number of single top events in the W sample before and after tagging. The expectation are shown for the W^* and $W-g$ fusion production channels. The overall contribution of single top events in the tagged sample is estimated by adding the contribution of each process. The single top contribution in the $W+\geq 3$ jet signal region is less than one event for all taggers.

With the acceptance $\mathcal{A}$ and tagging efficiency ϵ_{tag}, calculated for each process in every jet multiplicity bin, the expected number of single top events in the data is

calculated with the following expression:

$$N_{exp} = \sigma \cdot \mathcal{A} \cdot \epsilon_{tag} \cdot \mathcal{L}$$

where, σ, refers to the cross section of the specific process and $\mathcal{L}$ refers to the luminosity of the data. The total number of tagged single top events expected in the W sample as a function of jet multiplicity is calculated by adding the contribution of each process. The results are summarized in Table 6.4.

6.4 Contribution from WW, WZ, ZZ production

The production of heavy boson (W and Z) pairs followed by the leptonic decay of one of the produced bosons, constitutes an additional source of a high P_T lepton,

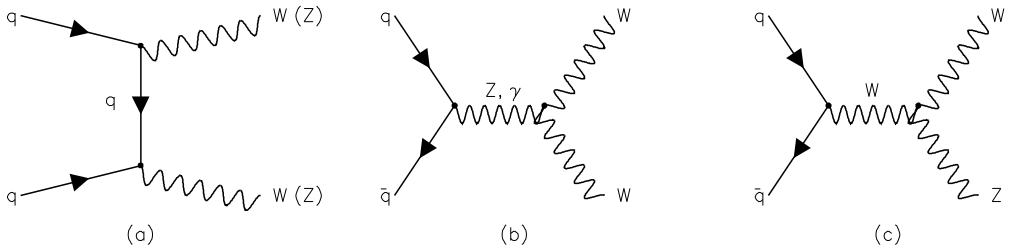

Figure 6.5: Leading order diagrams for diboson production at the Tevatron. Diagram (a) contributes for all processes, while diagrams (b) and (c) contribute to the WW and WZ production respectively.

large $\not\!\!E_T$ and at least two jets produced by the fully hadronic decay of the second boson. Therefore events from diboson production look similar to the QCD $W + 2$

jets production. However, loss of jets outside the detector fiducial region or emission of hard gluons at the initial or final state alter the jet multiplicity of the final state. In addition, the hadronic decays of $Z \to b\bar{b}, c\bar{c}$ and $W \to c\bar{s}$, produce high-E_T heavy flavor jets that can be tagged. Figure 6.5 shows the leading order diagrams for the diboson production. The contribution of diboson production to the the $W+ \geq 1$ jet sample, is estimated with PYTHIA Monte Carlo samples Individual samples of $\sim$23K, $\sim$30K and $\sim$40K events, were generated for the WW, WZ and ZZ processes, respectively. The expected number of events for each process is calculated, normalizing the number of Monte Carlo events passing the $W+ \geq 1$ jet selection criteria to the luminosity of the data sample (108.8 ± 7.86 pb^{-1}). For the normalization, the luminosity of the generated sample is calculated using the number of generated events and the theoretical cross section of each diboson production process as determined in [82]. The cross sections used for the normalization of each sample are $\sigma(WW)=9.5 \pm 0.7$ pb, $\sigma(WZ)=2.6 \pm 0.3$ pb and $\sigma(ZZ)=1.0 \pm 0.2$ pb.

In the cases of WZ and ZZ production, the tagged jets originate from b-quark decays. In order to properly calculate the expected event tagged yield due to WZ and ZZ production, the estimated tagged yield is scaled by the Data/Monte Carlo b-tagging efficiency scale factor. In contrast, there is no need to rescale the event tag yield for the WW production since most of the tags are due to c-quark jets from the hadronic decay of one of the W's ($W \to cs$).

The expected diboson contribution to the $W+ \geq 1$ jet sample before and after tagging as a function of the jet multiplicity is summarized in Table 6.5.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
		WW		
Before tagging	31.2±5.4	31.1±5.4	5.2±1.0	0.8±0.2
with SECVTX tag	0.33±0.13	0.81±0.24	0.28±0.12	0.00±0.00
with JPB tag	0.89±0.25	1.75±0.40	0.57±0.18	0.04±0.04
with SLT tag	0.21±0.10	0.71±0.22	0.04±0.04	0.00±0.00
		WZ		
Before tagging	4.4±0.9	4.8±1.0	0.9±0.2	0.1±0.0
with SECVTX tag	0.18±0.06	0.53±0.14	0.10±0.04	0.00±0.00
with JPB tag	0.24±0.07	0.60±0.15	0.15±0.05	0.01±0.01
with SLT tag	0.04±0.02	0.20±0.06	0.05±0.02	0.00±0.00
		ZZ		
Before tagging	0.3±0.1	0.4±0.1	0.1±0.0	0.0±0.0
with SECVTX tag	0.02±0.01	0.04±0.01	0.02±0.01	0.00±0.00
with JPB tag	0.01±0.01	0.05±0.02	0.03±0.01	0.00±0.00
with SLT tag	0.01±0.01	0.02±0.01	0.01±0.01	0.00±0.00

Table 6.5: Contributions to the $W+\geq 1$ jet sample from diboson (WW, WZ, ZZ) production, before and after tagging.

6.5 Contribution from mistags

The contribution of fake tags is estimated according to the method described in Chapter 5.

The SECVTX and Jetprobability mistags are calculated applying to $W+\geq 1$ jet sample the mistag probability matrices derived from the generic jet samples. As discussed in details in Chapter 5, these matrices represent the true mistag rate due to instrumental effects and are free of any heavy flavor contribution. In addition, corrections for the effect of multiple interactions are also taken into account in the mistag calculation. The SLT mistags are derived applying the track-based jet fake rate parametrization described in Section 5.2. The predicted rate is scaled by 74±3.2% to

account for the contribution of heavy flavor decays to the SLT mistags as derived in section 5.2.1.

Table 6.6 summarizes the observed negative tag yield and the estimated mistags in the $W+ \geq 1$ jet sample. The expected mistag yields for all tagging algorithms refer

Sample	SECVTX		Jetprobability		SLT
	Mistags	Negatives	Mistags	Negatives	Mistags
$W+1$ jet	11.5±1.15	19	48.5±4.8	66	108.40±10.8
$W+2$ jets	4.01±0.40	7	16.7±1.6	25	33.05±3.31
$W+3$ jets	1.06±0.11	2	3.7±0.4	8	7.80±0.78
$W+ \geq 4$ jets	0.39±0.04	0	1.2±0.1	7	3.20±0.32

Table 6.6: The SECVTX, Jetprobability and SLT mistag yields the in $W+ \geq 1$ jet sample as calculated using the fake probability matrices for SECVTX and Jetprobability and the track-based fake parametrization for SLT. The observed negative SECVTX and Jetprobability tag yields are also shown for comparison. The difference between the negative and predicted yields is due to heavy flavor contributions as it is shown in Section 8.6.

to the true mistag rate and do not include any heavy flavor contribution. For the case of SECVTX and Jetprobability, the observed negative tag yields are also shown for comparison. As a reminder, the negative tags are indicative of the level of mistags in the sample but they also accept contributions from heavy flavor decays. Therefore the observed disagreement between the predicted mistags and observed negative tags is due to heavy flavor decays and this will be shown in details in Section 8.6.

The mistag rates as calculated above, include the mistags from all processes contributing events to the $W+ \geq 1$ jet sample. However, in this chapter the contribution of each individual process to the tagged $W+ \geq 1$ jet sample is estimated from Monte Carlo simulation and data and a fraction of the contributed tags can also be due to mistags. Therefore, special attention is needed to avoid double counting of the

mistag rates. As already discussed, the SECVTX and SLT tagging algorithms do not include any mistags in the Monte Carlo simulations and there is no double counting of mistags for these two algorithms when estimating the contribution of the various processes. For the Jetprobability algorithm however, the Monte Carlo simulations include the correct amount of mistags and therefore the contribution of all processes before tagging need to be subtracted from the observed number of W candidate events before applying the mistag probability matrix.

Mistags are double counted for all tagging algorithms in the case of the non-W events. As described in Section 6.1, the contribution of this class of events to the $W+\geq 1$ jet sample is estimated directly from data and therefore it includes the appropriate amount of mistags. To avoid double counting, the mistag rate for all tagging algorithms is calculated on the number of $W+\geq 1$ jet candidate events after subtracting the pre-tagging contribution of non-W events.

Putting everything together, the calculation of the SECVTX and SLT mistag rates in the $W+\geq 1$ jet sample is performed applying the fake rate parametrizations to the pretagged $W+\geq 1$ jet sample after subtraction of non-W events, according to the following expression:

$$N_{mistags}^{W+jets} = N_{mistags}^{total} \times (1 - F_{non-W})$$

$N_{mistags}^{W+jets}$ is the expected number of mistags. $N_{mistags}^{total}$ are the mistag yields shown in Table 6.6 as calculated applying the mistag parametrizations to the total number of W candidate events in each jet multiplicity bin. F_{non-W} is the fraction of non-W events before tagging as calculated in section 6.1.

For Jetprobability, the mistags need to be rescaled to the fraction of W+non-heavy

flavor jets, $F_{W+no\ h.f.}$, in the sample, according to the following procedure:

$$\begin{aligned} N_{mistags}^{W+jets} &= N_{mistags}^{total} \times F_{W+no\ h.f.} \\ F_{W+no\ h.f.} &= \frac{N_{W+no\ h.f.}}{N_{W+jets}} \\ N_{W+no\ h.f.} &= N_{W+jets} - \sum_i B_i \end{aligned}$$

$$\sum_i B_i = N_{diboson} + N_{single\ top} + N_{W+h.f.} + N_{Z+h.f.} + N_{non-W} + N_{Z\to\tau\tau}$$

where N_{W+jets} is the number of pretagged W data events in each jet multiplicity bin and $N_{W+no\ h.f.}$ is the corresponding number of W+non heavy flavor jet events including misidentified $Z+\geq$1 jet events. $\sum_i B_i$ refers to the sum of the events before tagging from all the processes contributing events to the $W+\geq$1 jet sample.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
with SECVTX mistag	10.82±1.08	3.80±0.38	0.99±0.1	0.35±0.04
with JPB mistag	41.87±4.24	12.93±1.35	2.67±0.28	0.84±0.09
with SLT mistag	102.00±10.2	31.1±3.11	7.34±0.73	3.01±0.30

Table 6.7: Predicted SECVTX, Jetprobability and SLT mistags in the $W+\geq 1$ jet sample, as a function of jet multiplicity. A 10% systematic uncertainty is assigned on the mistag rate predictions based on the uncertainty of the mistag rate parametrization.

Table 6.7 summarizes the expected mistag rates in the $W+\geq$1 jet sample as a function of the jet multiplicity. The predictions include a 10% systematic uncertainty due to the uncertainty in the mistag rate parametrization (see Chapter 5).

6.6 Estimate of the $W/Z+$ heavy flavor production

Heavy quark in $W+$ jets events can be produced either singly as in the processes $g\bar{s}(\bar{d}) \rightarrow Wc$ or in pairs, when a final state gluon splits into a $c\bar{c}$ or $b\bar{b}$ pair. Similar production mechanisms are responsible for heavy flavor production in association with a Z boson. Figures 6.6, 6.11 and 6.12 show examples of tree level Feynmann diagrams for the production of W/Z in association with heavy flavor jets.

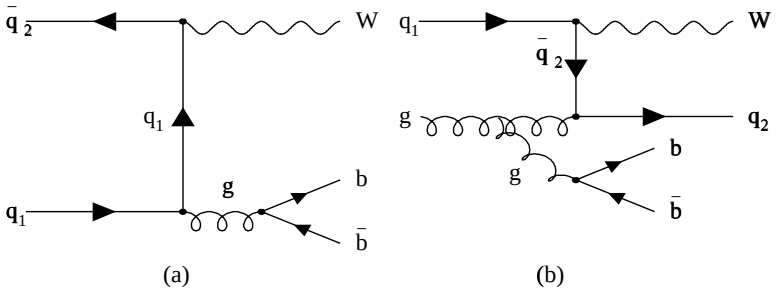

Figure 6.6: Examples of tree level Feynmann diagrams for $W+$ heavy flavor production. (a) Leading-order and (b) higher order diagrams.

However, the theoretical calculation for the production cross section is associated with large uncertainties due to its dependence on the factorization and normalization scale Q^2, the mass of the b and c quark and the choice of the structure function [60].

Two methods were developed to estimate the contributions of the $W/Z+$ heavy flavor production to the $W+ \geq 1$ jet sample. These methods are known as *Method* I and *Method* II. Method I is based solely on the positive tagging probability rates (see Chapter 5) observed in the inclusive jet data and the naive assumption that the

heavy flavor content of the inclusive jet data is the same as in $W+ \geq 1$ jet events. Method II is based on the up to date understanding of heavy flavor production and it relies on results obtained with a mixture of VECBOS [69] $W + N$ jets matrix element Monte Carlo and the HERWIG [68] parton shower Monte Carlo.

Method I was used only in the results of the SLT algorithm presented in References [58, 2, 23] and it is described briefly in the following Section. From the discussion that follows it is clear that Method II is superior and more complete than Method I and it is used for calculating the contribution of $W/Z + b\bar{b}$ in the $W+ \geq 1$ jet events.

6.6.1 Method I

This method assumes that the fraction of heavy flavor jets in the inclusive jet sample is the same as in the $W+$jets sample or larger. Figure 6.7 shows examples of the production mechanisms of heavy flavor jets in the inclusive jet samples. Three mechanisms are responsible for heavy flavor production in the inclusive jet samples:

- *Direct production:* this process produces a pair of back-to-back heavy flavor quarks. Figure 6.7(a) to (d) shows examples of direct production of heavy flavor jets. The initial state can be either gg or $q\bar{q}$.

- *Gluon splitting* in which a gluon at the final state splits to a pair of heavy flavor quarks as shown in Figure 6.7(e) to (h).

- *Flavor excitation* examples of which are shown in Figure 6.7(i) and (j). In this process, a heavy quark pulled from the sea of a proton comes in collision with a parton from the antiproton. The mechanism produces two heavy flavor jets at the final state. However, one of them is the partner of the heavy quark

participating in the hard scattering and is most often does not have significant P_T and it is emitted in large rapidities.

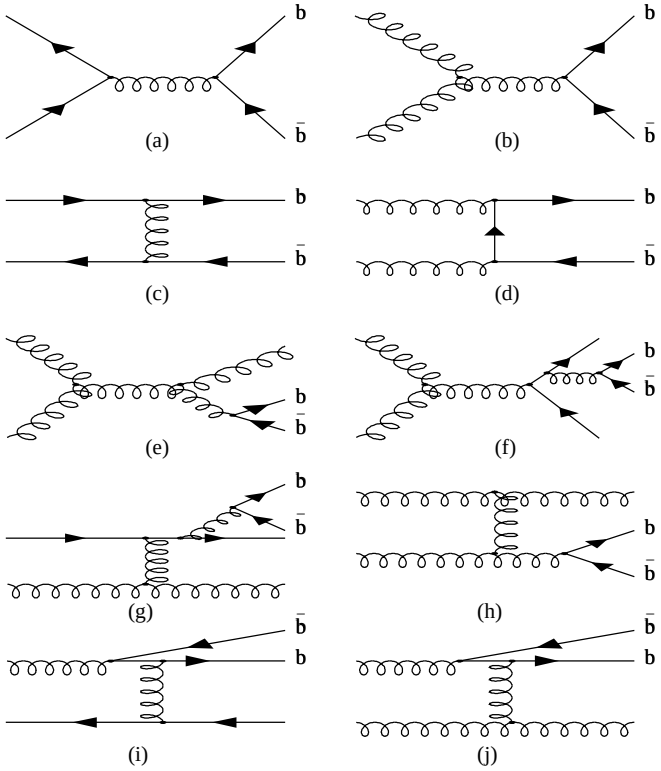

Figure 6.7: Some of the tree level Feynmann diagrams for QCD heavy flavor production. Diagrams (a) to (d) are examples of direct production, (e)-(h) are examples of heavy flavor production via final state gluon splitting and (i)-(j) are examples of flavor excitation. The clear distinction among the mechanisms disappears at next-to-leading order calculations due to interference terms.

The fraction of heavy flavor jets in the inclusive jet sample can be easily derived by measuring the tagging rate of the three tagging algorithms in inclusive jet samples. This procedure was actually performed in Chapter 5 when the positive tagging

probability matrices were derived for the SECVTX and Jetprobability taggers. The SLT fake rate parametrization is also derived with the inclusive jet data. The tagging rate in the inclusive jet data includes contributions from mistags due to track mismeasurements in light quark or gluons and real heavy flavor jets. Since the tagging probabilities are derived from data, there is no need for an a priori knowledge of the efficiency of each tagging algorithm. Therefore, application of these matrices to the W+jets sample it would predict simultaneously the sum of mistags and heavy flavor contributions. In this fashion any theoretical uncertainty on the absolute rate of inclusive $W + N$ jet production would drop out, being replaced by the observed cross section.

A priori there are several reasons to believe this assumption is too naive. First of all, the inclusive jet sample accepts large contributions from direct production of b and c-quark jets. This contribution is absent in the case of W sample, where all heavy quarks come from gluon splitting [60]. Secondly, the average fraction of gluon-initiated final state jets is different in the inclusive jet and W+jets samples. Approximately half of the jets in the $W+$ jet sample are gluon jets whilst gluon jets dominate especially at the low P_T spectrum of the inclusive jet samples. Both these reasons point to a smaller fraction of heavy flavor jets in the W sample than in the inclusive jet sample. This also indicates that the method overestimates the $W+b\bar{b}(c\bar{c})$ contributions and therefore serves as an upper conservative limit. Furthermore, the probability to produce heavy quark pairs in the two samples from gluons of the same E_T can be different because the Q^2 scale of the collision producing a W is much larger than the corresponding one in the inclusive jet case.

6.6.2 Method II

A better method of calculating the $W+$ heavy flavor contributions is based on the calculation of the fraction of $Wb\bar{b}(c\bar{c})$ events, relative to the total number of $W + N$ jet production events. It is expected that the calculation of the relative fractions of events with heavy quark jets to be less affected by theoretical uncertainties than the absolute rates. The absolute predictions can easily obtained by multiplying the calculated event fractions with the number of observed $W + N$ jet events and the corresponding tagging efficiencies for $Wc\bar{c}$ and $Wb\bar{b}$ events. The predictions can be tested on the inclusive sample of $W+$one b-jet, which does not have contributions from $t\bar{t}$ events. For this calculation, it is important to keep into account all mass effects in order to avoid divergences associated with the integration over the second b phase space [60].

The Monte Carlo calculations of $Wb\bar{b}$ production [60] were carried out at lowest order and they can be used only to predict the $W + b\bar{b}$ production in the $W + 2$ jet multiplicity bin. However, comparisons of the results obtained with a HERWIG simulation of $W + 1$ jet show good agreement with the full matrix element calculations [61]. Also several studies performed in hadronic collisions have indicated that jets produced in the scheme incorporated in the HERWIG Monte Carlo agree both in rate and in shape of distributions with what is observed [63, 64]. This gives confidence that HERWIG models correctly the gluon radiation. Another advantage of using the HERWIG Monte Carlo for the $Wb\bar{b}$ calculations is that its predictions on the fraction of b-jets can be measured in inclusive jet samples and be compared to the fractions observed in the data. This comparison, discussed in Section 6.6.3, is used to calibrate the whole simulation package (HERWIG+CLEOMC+detector simulation) using the tagging rates observed in the inclusive jet data.

The main disadvantage in using a parton shower Monte Carlo is that it underestimates the rate of events with large invariant mass between the two b-quarks. Most of the b's produced in the parton shower approximation of the gluon splitting tend to have small opening angle. Therefore, the predictions on the $Wb\bar{b}$ events fractions obtained with HERWIG are valid only when the two b-quarks are very close and practically result in a single b-jet in the event.

In order to estimate the $Wb\bar{b}$ event fractions in the region of large $b\bar{b}$ mass the VECBOS $W + N$ $(N = 0 : 4)$ jet matrix element Monte Carlo [69] is used and only those diagrams involving a $q\bar{q}$ pair are selected. With the VECBOS Monte Carlo, higher order processes like the one shown in Figure 6.6(b) are also included and calculated explicitly. However, the VECBOS matrix element calculations do not include quark mass effects and as a result, the cross section diverges when the

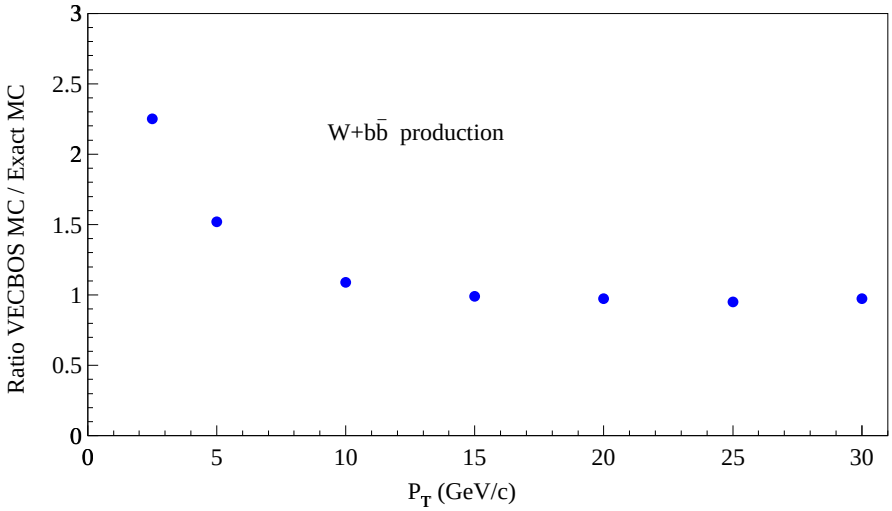

Figure 6.8: Ratio between the VECBOS Monte Carlo $Wb\bar{b}$ production cross section at parton level and the exact matrix element calculations [60] for $W + 2$ jet events as a function of the minimum transverse momentum of the b-quark.

transverse momentum of the quarks is very soft or their spatial separation, DR, goes to zero. Therefore, the VECBOS simulation would yield two hard and well separated b-jets. To gain some confidence that the VECBOS simulation reproduces the results of the exact matrix element calculations of $W + b\bar{b}$ production of Reference [60], the $W+b\bar{b}$ cross section for $W+2$ jet events is compared for the two calculations. The ratio of the two cross sections is shown in Figure 6.8 as a function of the minimum P_T of the two b-quarks. For the comparison, the same Q^2 scale, structure function and minimum DR separation between the two b-quarks are applied in both simulations [61]. As expected, the VECBOS cross section diverges for small values of the b-quark P_T but for $P_T \geq$15 GeV/c good agreement is observed between the results of the two calculations and therefore one can assume the effect of the b-quark mass is negligible for hard b-jets.

From the above discussion, it is clear that there is no single Monte Carlo generator to be used in the calculation of the $W+$ heavy flavor quark pair production. Each Monte Carlo generator discussed above has its advantages and limitations. However, they can be used in regions of phase space that are expected to perform the best. The features of each Monte Carlo calculations lead to the conclusion that in order to calculate the fraction of the $W + N$ jet events with heavy flavor jets , the results from the HERWIG parton shower Monte Carlo need to be combined with the results from the VECBOS $W + N$ jet matrix element generator. Using the mixture of the results from the two generators all possible configurations for the $b\bar{b}$ dynamics can be accounted for. The HERWIG Monte Carlo is used to predict the fraction of events with small mass between the two b-quarks resulting usually in a single b-jet in the event. On the other hand, VECBOS predicts events with hard and well separated b-quarks resulting usually in two b jets in the event. Therefore, the VECBOS Monet

Carlo is used to calculate the relative fraction of $W+$ heavy flavor events with two b-jets.

It is also clear that the probability to tag a $W + b\bar{b}$ event depends strongly on the angular separation between the two b-quarks at the final state since events with well separated b-quark jets have higher probability to be tagged compared to events where the two b quarks are close to each other. For this reason, the probability to tag at least one b-jet in $Wb\bar{b}$ events is estimated separately for events generated with the HERWIG and VECBOS Monte Carlos.

Same arguments as above hold for $Wc\bar{c}$ production and production of heavy flavor jets in association with a Z since mechanisms for producing $Z+$ jets are quite similar to the $W+$ jets production.

The results of the Method II calculations are the subjects of the following sections.

6.6.3 Comparisons of the heavy quark rates in inclusive jet data and HERWIG simulations

As mentioned in Section 6.6.2, the production of $Wb\bar{b}$ or $Wc\bar{c}$ proceeds via splitting of one of the gluons at the final state. It was also mentioned that the calculation of the $Wb\bar{b}$ and $Wc\bar{c}$ contribution relies on HERWIG simulation and therefore it depends on how well HERWIG models the gluon splitting to a heavy flavor pair. Furthermore, it also depends on how well the full simulation package, HERWIG+CLEO+Detector simulation, reproduces the tagged rates in different samples. Therefore the calibration of the simulation package is essential for the calculation of the $Wb\bar{b}$ and $Wc\bar{c}$ production.

The calibration of the simulation package is performed by comparing the HER-

WIG predictions with the tagging rate observed in the inclusive JET_20, JET_50 and JET_100 Run 1B data samples. For this purpose, the inclusive $2 \rightarrow 2$ QCD process was used to generate events with HERWIG. The events were further processed with the CLEO Monte Carlo for correct modeling of the heavy flavor decays and dynamics and finally passed through full detector simulation including the track degradation procedure.

In order to simulate the three different trigger samples, events were generated with the following requirements:

- Minimum parton $P_T \geq 10$, 40 and 80 GeV/c at generation level
- Maximum parton pseudorapidity $|\eta| \leq 4.5$
- At least one jet at detector level with $E_T \geq 20$, 50, 100 GeV in order to simulate the trigger requirement of the data samples.
- Structure function used: MRS(G)
- b and c-quark masses of m_b=4.75 GeV/c^2 and m_c=1.5 GeV/c^2.
- Sample sizes: 10.8, 2.2 and 1 million events were generated for the simulation of the JET_20, JET_50 and JET_100 samples respectively.

For what follows, events both in data and simulation samples are also required to have one jet above the trigger threshold and at least one taggable jet (as a reminder taggable is a jet with $E_T \geq 15$ GeV, $|\eta| \leq 2$ containing at least two good SVX or Jetprobability tracks according to the tagging algorithm considered).

As discussed in Section 6.6.2, the inclusive jet data samples include heavy flavor jets produced via direct production, heavy flavor excitation and gluon splitting.

Examples of Feynmann diagrams for the three production mechanisms are shown in Figure 6.7. In the simulations, the three processes are distinguished based on the information for the partons involved in the hard scattering process $1+2 \rightarrow 3+4$ and the origin of the b and c hadrons after parton shower. The generator information is also used to identify the flavor of the tagged jet. Table 6.8 shows the rates of heavy flavor jets before tagging in a subsample of the simulated JET_50 sample partitioned according to flavor type and production mechanism.

Type	Direct prod.	Flavor excitation	Gluon splitting	Total
b	0.44±0.05%	1.6±0.1%	2.5±0.1%	4.5±0.2%
c	0.45±0.05%	3.1±0.1%	7.7±0.2%	11.2±0.3%

Table 6.8: Heavy flavor event fraction (in %) before tagging in a JET_50 HERWIG inclusive jet simulation split according to flavor type and production mechanism.

In the data the identification of the three processes is highly non-trivial. However, there are some properties associated with each production mechanism which can be used to distinguish among them. Heavy flavor jets produced via direct production are expected to be back-to-back and also are more likely to give rise to double tags in the event. On the other hand, gluon splitting produces a pair of heavy flavor jets quite close in $\eta - \phi$ space. If two jets are close and at the same time are both tagged then it is more likely to come from gluon splitting. Figure 6.9 shows the distributions of the minimum distance between two b-jets in the JET_50 simulation sample for the three production mechanisms before and after tagging with SECVTX.

Taking these properties into account, the data samples are divided into five categories and the tagging rate in each class is examined and compared to the corresponding Monte Carlo rates. Where appropriate the DR distributions are also used

in the comparisons. The 5 classes of events considered are:

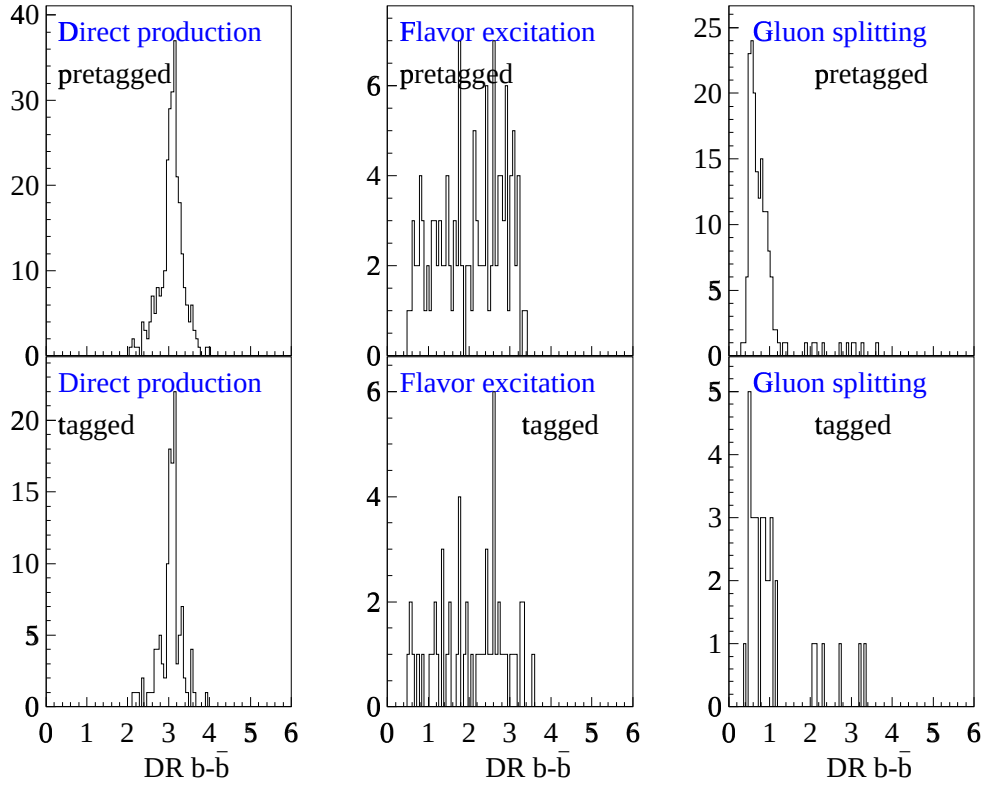

Figure 6.9: Distributions of the minimum distance between two b-jets in the JET_50 simulation before and after tagging with SECVTX for direct production, heavy flavor excitation and gluon splitting.

1. Events with at least one tagged jet without any other selection

2. Events with three or more jets one of which is tagged. This class of events is enriched in gluon splitting.

3. Events with a jet of $E_T \geq 10$ GeV within $DR \leq 1.2$ from a tagged jet. Events of this class are enriched in gluon splitting.

4. Events with double tags. This class contains mostly events from direct production.

5. Events with two tagged jets within $DR \leq 1.2$. This class contains heavy flavor jets from gluon splitting.

Table 6.9 lists the tagging fraction for each class of events in the simulated JET_20, JET_50 and JET_100 samples split according to flavor and production mechanism. The comparison of the HERWIG predictions with the tagging rate in the data taking into account the HERWIG fractions for the contribution of the three production mechanisms (actually 6 considering together b and c jets) is performed with a likelihood fit. In the fit the c/b fraction is required to be the same as in HERWIG within errors. A 10% uncertainty is assigned to the c/b fraction for direct production while for flavor excitation and gluon splitting an uncertainty of $\pm 20\%$ is assigned to the c/b ratio. It should be noted here that in principle one could measure the c/b ratio in the 5 classes of events by fitting the corresponding $c\tau = M_T^V \cdot Lxy/P_T^V$ distributions. However, trying to estimate the c/b fraction with different fitting procedure and different tagging algorithms the obtained results on the c/b ratio were rather unreliable and contradictory independently of the fact the total heavy flavor fraction was stable in all fitting procedure. Also, in the case of b-SECVTX tags, the fraction in the simulation is increased by 25% to account for the b-tagging efficiency scale factor of 1.25 measured in Section 4.4. The c-tagging efficiency scale factor is taken to be 1 after track degradation.

The HERWIG predictions in the three data samples and the five subclasses of events are fitted by adjusting each individual production mechanism cross section with a fit parameter. There are total of 20 degrees of freedom in the fit and the resulting

JET 20						
	b			c		
ev. type	dir.	h.f. exc.	$g \to b\bar{b}$	dir.	h.f. exc.	$g \to c\bar{c}$
all ev. (tags/ev)	0.200 ± 0.013	0.404 ± 0.018	0.135 ± 0.010	0.062 ± 0.007	0.197 ± 0.013	0.155 ± 0.011
$\geq$ 3 jets (tags/ev)	0.178 ± 0.046	0.488 ± 0.076	0.381 ± 0.067	0.036 ± 0.021	0.202 ± 0.049	0.440 ± 0.072
tag with extra jet (tags/ev)	0.002 ± 0.001	0.005 ± 0.002	0.014 ± 0.003	0.001 ± 0.001	0.004 ± 0.002	0.011 ± 0.003
double tags (tags/ev)	0.121 ± 0.032	0.026 ± 0.015	0.009 ± 0.016	0.009 ± 0.009	0.009 ± 0.009	0.000 ± 0.016
$\geq$ 2 tags with $DR \leq$1.2	0.000 ± 0.008	0.000 ± 0.008	0.000 ± 0.008	0.000	0.000	0.000
JET 50						
	b			c		
ev. type	dir.	h.f. exc.	$g \to b\bar{b}$	dir.	h.f. exc.	$g \to c\bar{c}$
all ev. (tags/ev)	0.290 ± 0.011	0.653 ± 0.017	0.737 ± 0.018	0.072 ± 0.002	0.263 ± 0.011	0.417 ± 0.014
$\geq$ 3 jets (tags/ev)	0.231 ± 0.016	0.636 ± 0.026	0.971 ± 0.033	0.053 ± 0.008	0.226 ± 0.016	0.532 ± 0.024
tag with extra jet (tags/ev)	0.023 ± 0.003	0.071 ± 0.006	0.217 ± 0.010	0.004 ± 0.001	0.025 ± 0.003	0.086 ± 0.006
double tags (tags/ev)	0.200 ± 0.018	0.094 ± 0.013	0.065 ± 0.010	0.016 ± 0.004	0.007 ± 0.003	0.009 ± 0.002
$\geq$ 2 tags with $DR \leq$1.2	0.000 ± 0.008	0.008 ± 0.003	0.026 ± 0.005	0.000	0.000	0.000
JET 100						
	b			c		
ev. type	dir.	h.f. exc.	$g \to b\bar{b}$	dir.	h.f. exc.	$g \to c\bar{c}$
all ev. (tags/ev)	0.431 ± 0.022	0.557 ± 0.025	1.185 ± 0.037	0.123 ± 0.012	0.224 ± 0.016	0.688 ± 0.028
$\geq$ 3 jets (tags/ev)	0.351 ± 0.027	0.571 ± 0.035	1.506 ± 0.056	0.086 ± 0.013	0.197 ± 0.020	0.840 ± 0.042
tag with extra jet (tags/ev)	0.064 ± 0.009	0.097 ± 0.010	0.503 ± 0.024	0.009 ± 0.003	0.031 ± 0.006	0.210 ± 0.015
double tags (tags/ev)	0.268 ± 0.029	0.108 ± 0.018	0.142 ± 0.021	0.026 ± 0.009	0.005 ± 0.003	0.010 ± 0.005
$\geq$ 2 tags with $DR \leq$1.2	0.000 ± 0.003	0.006 ± 0.003	0.054 ± 0.009	0.000	0.000	0.000

Table 6.9: Rates (in %) of SECVTX tags in the Herwig QCD simulation.

All QCD samples	
process	cross section weight
b dir	0.48 ± 0.25
b flav. exc.	1.84 ± 0.38
$g \to b\overline{b}$	1.17 ± 0.17
c dir	0.48 ± 0.27
c flav. exc.	2.01 ± 0.67
$g \to c\overline{c}$	1.39 ± 0.43
JET_50 and JET_100 only	
process	cross section weight
b dir	0.43 ± 0.29
b flav. exc.	1.33 ± 0.49
$g \to b\overline{b}$	1.38 ± 0.21
c dir	0.43 ± 0.30
c flav. exc.	1.36 ± 0.63
$g \to c\overline{c}$	1.49 ± 0.45

Table 6.10: Result of the fit of the Herwig simulation to JET_20, JET_50 and JET_100 data when using SECVTX tags. Each Herwig cross section is weighted by a fit free parameter. Ratios $\frac{c}{b}$ are constrained to the Herwig prediction within errors. The b-tagging efficiency in the simulation is increased by 25%.

$\chi^2 = 24$. The weight factors to be applied to the HERWIG predictions are shown in Table 6.10. Table 6.11 shows a comparison between the $b + c$ heavy flavor tagged rates observed in the data samples to the $b + c$ rate as predicted by HERWIG after adjusting the production cross section with the results of the previous fit. Performing the fit using only the JET_50 and JET_100 data the normalization factors for the HERWIG predictions shown in the lower part of Table 6.10 are obtained.

The same procedure is repeated for Jetprobability tags. The results of the fit of the HERWIG predictions to the data are shown in Table 6.12. The fit returns $\chi^2 = 7$ for 6 degrees of freedom.

The results of the fit with the Jetprobability tagging rates compare quite well with

JET 20				
	Data	Herwig		
	$b+c$	$b+c$	b	c
all ev. (tags/ev)	2.090 ± 0.210	1.889±0.054	1.247	0.643
$\geq$ 3 jets (tags/ev)	3.330 ± 0.360	2.8227 ± 0.247	1.788	1.039
double tags (tags/ev)	0.366 ± 0.082	0.217 ± 0.064	0.182	0.035
$\geq$ 2 tags with $DR \leq$1.2	0.020 ± 0.007	0.000 ± 0.028	0.000	0.000
JET 50				
	Data	Herwig		
	$b+c$	$b+c$	b	c
all ev. (tags/ev)	3.420 ± 0.360	3.904 ± 0.056	2.757	1.146
$\geq$ 3 jets (tags/ev)	4.300 ± 0.450	4.250 ± 0.090	3.026	1.224
tag with extra jet (tags/ev)	0.750 ± 0.080	0.668 ± 0.023	0.496	0.172
double tags (tags/ev)	0.560 ± 0.118	0.574 ± 0.044	0.539	0.034
$\geq$ 2 tags with $DR \leq$1.2	0.058 ± 0.014	0.071±0.014	0.071	0.000
JET 100				
	Data	Herwig		
	$b+c$	$b+c$	b	c
all ev. (tags/ev)	4.200 ± 0.460	4.750 ± 0.095	3.279	1.471
$\geq$ 3 jets (tags/ev)	5.150 ± 0.550	5.346 ± 0.136	3.735	1.612
tag with extra jet (tags/ev)	1.550 ± 0.160	1.360 ± 0.049	1.000	0.360
double tags (tags/ev)	0.636 ± 0.132	0.808 ± 0.069	0.772	0.036
$\geq$ 2 tags with $DR \leq$1.2	0.110 ± 0.025	0.116±0.019	0.116	0.000

Table 6.11: Comparison of tagging rates (in %) in jet data and in Herwig simulations tuned according to Table 6.10.

process	cross section weight
b dir	0.51 ± 0.36
b flav. exc.	1.59 ± 0.65
$g \to b\overline{b}$	1.15 ± 0.44
c dir	0.51 ± 0.37
c flav. exc.	1.51 ± 0.63
$g \to c\overline{c}$	1.10 ± 0.45

Table 6.12: Result of the fit of the Herwig simulation to JET_50 and JET_100 data when using Jetprobability tags. Each Herwig cross section is weighted by a fit free parameter. Ratios $\frac{c}{b}$ are constrained to the Herwig prediction within errors.

the results obtained when using SECVTX tags. Therefore the two procedures are combined and the fit is performed only in the JET_50 and JET_100 sample for either Jetprobability or SECVTX. The results of the combined fit are shown in Table 6.13. The fit has $\chi^2 = 22$ for 22 degrees of freedom.

SECVTX or Jetprobability	
process	cross section weight
b dir	0.47 ± 0.22
b flav. exc.	1.37 ± 0.41
$g \to b\overline{b}$	1.39 ± 0.19
c dir	0.47 ± 0.24
c flav. exc.	1.30 ± 0.45
$g \to c\overline{c}$	1.35 ± 0.36

Table 6.13: Result of the fit of the Herwig simulation to JET_50 and JET_100 data when using SECVTX or Jetprobability tags. Each Herwig cross section is weighted by a fit free parameter. Ratios $\frac{c}{b}$ are constrained to the Herwig prediction within errors. The b-tagging efficiency for SECVTX tags is increased by 25%.

The results of this fit are used as correction factors for the simulated gluon splitting rates:

- $g \rightarrow b\bar{b} = 1.39 \pm 0.19$
- $g \rightarrow c\bar{c} = 1.35 \pm 0.36$

These correction factors are used to scale the $Wb\bar{b}$ and $Wc\bar{c}$ event fractions used in the calculation of the $Wb\bar{b}$ and $Wc\bar{c}$ contributions as discussed in the next Section.

6.7 The $Wb\bar{b}$ and $Wc\bar{c}$ contribution

As discussed in Section 6.6.2 the calculation of the $Wb\bar{b}$ contribution is performed in two parts according to the number of b-jets at the final state. At the first part, the HERWIG Monte Carlo used to calculate the fraction and tagging efficiency of $Wb\bar{b}$ and $Wc\bar{c}$ events with exactly one b or one c-jet at the final state. At the second part, the VECBOS $W + N$ jet exact matrix element Monte Carlo is used to calculate the fraction and tagging efficiency of $Wb\bar{b}$ and $Wc\bar{c}$ events with two b or two c-jets at the final state.

6.7.1 $Wb\bar{b}$ and $Wc\bar{c}$ events with 1 b and 1 c-jet

The fraction of W events with *exactly 1 b-jet*, $F1_{Wb\bar{b}}^{[Njet]}$, is calculated as a function of the jet multiplicity, $[Njet]$, in the event using the HERWIG Monte Carlo generator. The samples are created running the $W + 1$ jet matrix elements in HERWIG, followed by parton shower and fragmentation. Before detector simulation, B-hadrons are redecayed using the decay tables and B-hadron decay dynamics incorporated in the CLEO [76] Monte Carlo. Full detector simulation is then applied to the resulting samples. Events are further required to satisfy the $W+ \geq 1$ jet selection criteria and the event yields for $W+$ jets and for $W+$ jets with exactly 1 heavy quark jet at the

final state, are counted as a function of jet multiplicity. A jet, at detector level, with $E_T \geq$15 GeV and $|\eta_D| \leq$2 is defined to be a b-jet, if there is a B-hadron at generation level which direction is within a cone of $DR \leq$0.4 around the jet axis. The HERWIG samples were generated using the MRSD0$'$ structure function from the PDF library and requiring a minimum parton P_T of 10 GeV/c which is sufficiently lower than the jet raw E_T threshold of 15 GeV required at detector level. Also in the generation, the mass of the b and c quarks were set at $m_b = 4.75$ GeV/c^2 and $m_c = 1.5$ GeV/c^2 respectively according to values used in the theoretical calculations (The HERWIG default values are $m_b = 5.25$ GeV/c^2 and $m_c = 1.8$ GeV/c^2 respectively). However, the uncertainty associated with the choice of the b and c-quark masses is included in the the normalization of the HERWIG predictions to inclusive jet data described in Section 6.6.3.

In order to increase the statistics of the $W+$ heavy flavor samples, events with at least one b quark at the end of the parton shower stage were accepted for fragmentation and detector simulation. These samples were normalized to the inclusive $W+$ jets samples using the number of $W+1$ jet matrix element events generated in each case.

The fraction of $W + b\bar{b}$ events with only 1 b-jet at the final state are shown in Table 6.14 as a function of jet multiplicity in the event. The shown fractions include the Data/HERWIG scale factor for the gluon splitting cross section, $SF_{g \rightarrow b\bar{b}} = 1.39 \pm 0.19$, determined in Section 6.6.3.

The same to the above procedure is followed in order to calculate the relative event fraction of $W + c\bar{c}$ events with exactly 1 c-jet at the final state. What is of interest also in this case, is the fraction of events with exactly one c-jet arising from gluon splitting. This is emphasized in the $Wc\bar{c}$ case because events with a

Sample	Wc (%)	$Wb\bar{b}$		$Wc\bar{c}$	
		1 b-jet (%)	2 b-jets (%)	1 c-jet (%)	2 c-jets (%)
$W+1$ jet	4.8±1.4	0.80±0.11		2.01±0.54	
$W+2$ jet	7.2±2.2	1.28±0.18	1.20±0.38	3.73±1.00	1.40±0.52
$W+3$ jet	7.5±2.3	1.88±0.31	1.90±0.62	5.31±1.48	2.30±0.91
$W+\geq 4$ jet	7.5±2.3	3.54±1.06	2.40±0.77	6.08±2.45	3.00±1.13

Table 6.14: Fractions (in %) of $W+\geq 1$ jet events with heavy flavor jets as a function of jet multiplicity. The event fractions are determined from a combination of HERWIG (for events with only 1 b-jet) and VECBOS (for events with 2 b-jets) Monte Carlo simulations of $W+\geq 1$ jet events. The event fractions for $Wc\bar{c}$ and $Wb\bar{b}$ have been multiplied by the Data/HERWIG scale factors of $SF_{g \rightarrow c\bar{c}} = 1.35 \pm 0.36$ and $SF_{g \rightarrow b\bar{b}} = 1.39 \pm 0.19$ as determined by comparing the gluon splitting in generic jet data and Monte Carlo simulations.

$Wc\bar{s}$ vertex result also in events with a single c-jet at final state. However, the contribution of Wc direct production is treated as a separate process as discussed in Section 6.7.5. The fractions of events with exactly 1 c-jet from gluon splitting as a function of jet multiplicity are listed in Table 6.14. The shown fractions have been multiplied with the Data/HERWIG scale factor for the gluon splitting cross section, $SF_{g \rightarrow c\bar{c}} = 1.35 \pm 0.36$, determined in Section 6.6.3.

Figure 6.10(a) shows the fraction of events with one b and one c jet as a function of jet multiplicity as obtained running $W+1$ jet matrix element in the HERWIG Monte Carlo.

The quoted errors on the events fractions calculated with HERWIG, includes both statistical and systematics error added in quadrature. The statistical errors on the $Wb\bar{b}$ fractions as a function of jet multiplicity are ±1.7%, ±3.3%, ±10% and ±26.6% from $W+1$ to $W+\geq 4$ jet bins respectively. The corresponding statistical errors on the $Wc\bar{c}$ fractions are ±1.3%, ±2.5%, ±7.9% and ±30.2% from $W+1$ to $W+\geq 4$ jet bins. Since the HERWIG predictions for the gluon splitting are normalized to the

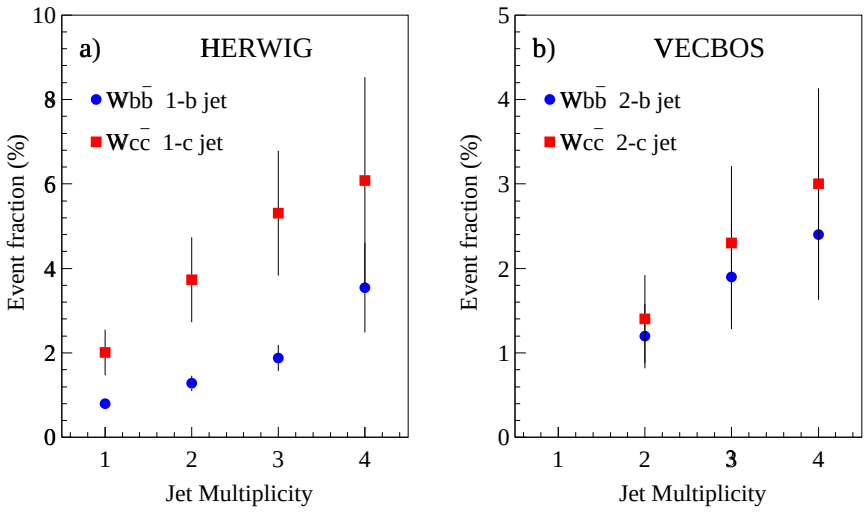

Figure 6.10: Fraction of $Wb\bar{b}$ and $Wc\bar{c}$ events with exactly 1 b or c-jet and exactly 2 b or c-jets.

data, the systematic errors on the calculated event fractions reflect the error of the $SF_{g \to b\bar{b}}$ and $SF_{g \to c\bar{c}}$.

6.7.2 $Wb\bar{b}$ and $Wc\bar{c}$ events with 2 b and 2 c-jets

The fraction of events with 2 b-jets at the final state are calculated with the VECBOS Monte Carlo. VECBOS samples were generated in two ways. In order to obtain the total number of generic $W+$ jet events, the VECBOS Monte Carlo is run accepting events from all contributing diagrams. The number of $Wb\bar{b}$ events are generated selecting only those diagrams that contain a $q\bar{q}$ pair and replacing it with a $b\bar{b}$ pair. This is known as the *forced $b\bar{b}$ or $c\bar{c}$ option* of VECBOS. The exchange of the $q\bar{q}$ with a $b\bar{b}$ pair is done assuming flavor universality (20% of the times it is going to be a $b\bar{b}$ pair) and neglecting the b-quark mass which as discussed in Section 6.6.2 has small effect for b-jets with $P_T \geq$15 GeV/c (see also Figure 6.8. The samples generated with the forced and unforced option of the VECBOS Monte Carlo were interfaced via the

HERPRT interface package to the HERWIG Monte Carlo for initial and final state evolution and fragmentation. The output of the VECBOS+HERPRT package is run through the CLEO Monte Carlo and finally through detector simulation.

The VECBOS Monte Carlo samples were generated using the MRSD0$'$ structure function and using a renormalization scale $Q^2 = M_W^2$ at the matrix elements computations. The same scale was used also in the HERPRT interface package for the initial and final state evolution. To check dependences on structure function and Q^2 scale, separate Monte Carlo samples were also generated with the MRS(G) structure function and two other Q^2 scales: $Q^2 = M_W^2 + P_T^{W^2}$ and $Q^2 =< P_T^{jet} >^2$. In the first scale M_W and P_T^W refer to the mass and transverse momentum of the W. The second scale is defined as the average P_T of the outgoing partons. In all VECBOS samples used in this calculation, the following kinematic requirements were applied at generation level:

- Parton $P_T \geq 8$ GeV/c
- Maximum parton pseudorapidity $|\eta_{parton}| \leq 2.5$
- Minimum separation between final state partons, $DR = \sqrt{\delta\eta^2 + \delta\phi^2} \geq 0.4$
- Minimum W lepton $P_T \geq 12$ GeV/c
- Maximum lepton pseudorapidity $|\eta_\ell| \leq 1.2$

Since VECBOS is an exact matrix element Monte Carlo, individual samples were generated for each parton multiplicity bin for both generic (running VECBOS with all diagrams) $W + N$ partons and $W + N$ partons with a heavy flavor quark pair. In any VECBOS sample after full simulation (including HERPRT, CLEO Monte Carlo for heavy flavor decays and detector simulation) the events are required to satisfy

the $W+ \geq 1$ jet selection criteria and the jet multiplicity in the event to be exactly the same to the number of partons in the matrix element generation. In the case of the forced heavy flavor VECBOS samples, the events are also required to have two b or c-jets at final state. For every sample the "accepted" cross section is calculated, $\sigma = N_W^{[Njet]}/(N_{gen} \cdot \sigma_{VECBOS})$ and the fractions of $Wb\bar{b}$ or $Wc\bar{c}$ events are calculated from the ratio of the $W + N$ jet accepted cross section with a $b\bar{b}$ pair to the generic $W + N$ jet accepted cross section:

$$F2_{W+hf}^{[Njet]} = \frac{\sigma_{Wb\bar{b}}^{[Njet]}}{\sigma_{W\ all}^{[Njet]}}$$

Table 6.14 lists the fraction of $Wb\bar{b}$ and $Wc\bar{c}$ events with 2 b and 2 c-jets at final state as determined from all VECBOS samples.

The event fractions are calculated after full simulation at detector level because there are several sources that can affect the acceptance of $Wb\bar{b}$ events differently than the acceptance of generic $W + N$ jet events. Also if the calculation were to be performed at parton level then the fraction of $W + b\bar{b}$ and $W + c\bar{c}$ events would be exactly the same since the matrix element Monte Carlo treats either case equally. After full simulation, effects as b or c hadron semileptonic decays, heavy quark fragmentation, jet energy scale for b and c-jets relative to light quark or gluon jets and contributions from initial state radiation result in a lower acceptance for $W+$heavy flavor events relative to the generic $W + N$ jet events. Therefore the uncertainty related with the fraction of events with $W+$ heavy flavor is related to the uncertainty introduced by the above mentioned factors. Table 6.15 lists the contribution of each of the above sources to the uncertainty of the $Wb\bar{b}$ and $Wc\bar{c}$ event fractions.

More specific, the effect of the jet energy scale was studied by changing the jet E_T

threshold by 5% and measuring the change in the acceptance for $Wb\bar{b}$ and generic W events. This leads to $\pm 12\%$ change in the acceptance. The knowledge of the branching fractions of b and c hadron decays contribute less than $\pm 5\%$ to the uncertainty while the effect of the uncertainty in the Peterson fragmentation function contributes to less than 10% to the jet acceptance for $Wb\bar{b}$. An overall $\pm$10to the fraction of $Wb\bar{b}$ and $Wc\bar{c}$ fractions due to the heavy quark decay and fragmentation modeling.

Parton level	
m_b, m_c	$\leq 5\%$
Q^2 scale and structure function	5%
Minimum parton P_T requirement	10%
Total	$\pm 15\%$
Simulation and detector level	
Jet energy scale	$\pm 12\%$
B-fragmentation and decay modeling	$\pm 10\%$
Initial and final state radiation	$\pm 15\%$
Total	$\pm 22\%$
Total systematic uncertainty	$\pm 30\%$

Table 6.15: Sources and the corresponding uncertainty contributing to the systematic error of the $Wb\bar{b}$ and $Wc\bar{c}$ event fractions with 2 b and 2 c-jets at the final state.

Initial state radiation affects mostly the generic $W + N$ jet event acceptance because of the different initial state parton participating in the hard scattering. For example, in generic $W + 2$ jet events, $\sim$50% of the initial state is $q\bar{q}$ and $\sim$50% is qg. In contrast, $Wb\bar{b}$ events with 2 heavy flavor quarks at final state include only diagrams with a $q\bar{q}$ pair at the initial state. The P_T spectrum of the initial state radiation is softer for a $q\bar{q}$ initial state than for a gq initial state. Also jets arising from initial state radiation can compensate for any loss in the generic $W + N$ jet acceptance due to soft partons produced at the hard scattering. However, initial state radiation jets

can not contribute to $W+$ heavy flavor events because in the selection requirements ask always for two heavy flavor jets at the final state. The effect of the initial state radiation was studied by changing the Q^2 used in the HERPRT interface package. Three different Q^2 were used: $Q^2 = M_W^2$, $Q^2 =< P_T^{jet} >^2$ and $Q^2 = M_W^2 + P_T^{W^2}$. For small Q^2 scales, the contribution from initial state radiation. is less. By studying the change of the $Wb\bar{b}$ event fraction with respect to the ratio $R = N_{W+\geq 3\ jet}/N_{W+\geq 2\ jet}$, for the three different Q^2 scales, a $\pm 15\%$ systematic uncertainty is assigned to the value of $Wb\bar{b}$ fractions.

An additional $\pm 15\%$ systematic uncertainty is assigned to the values of $Wb\bar{b}$ fractions by studying the effect of the re-normalization scale used in the VECBOS matrix element generation and the effect of higher order corrections to the $Wb\bar{b}$ fraction. The last effect was studied using $W + 3$ jet matrix element Monte Carlo and varying the minimum P_T of the third parton down to very small values. While the cross section diverges for very small values, the ratio of $Wb\bar{b}/W_{all}$ remains finite.

The overall uncertainty assigned to the $Wb\bar{b}$ and $Wc\bar{c}$ event fractions is $\pm 30\%$. The errors are completely correlated among all the fractions.

6.7.3 Tagging efficiency in $Wb\bar{b}$ and $Wc\bar{c}$ events

The tagging efficiency for $W+$ heavy flavor jets is defined as the probability to tag at least one jet in the event:

$$\epsilon^{[Njet]}_{W+hf\ \geq 1\ tag} = \frac{N^{[Njet]}_{W+hf\ \geq 1\ tag}}{N^{[Njet]}_{W+hf}}$$

The tagging efficiency is measured as a function of the jet multiplicity $[Njet]$, separately for events with exactly one heavy flavor jet, using the HERWIG Monte Carlo

sample and for events with two heavy flavor jets using the VECBOS Monte Carlo samples. For the case of two heavy flavor jets, the efficiency to tag at least two jets in the event is also measured according to a similar expression:

$$\epsilon^{[Njet]}_{W+hf\ \geq 2\ tags} = \frac{N^{[Njet]}_{W+hf\ \geq 2\ tags}}{N^{[Njet]}_{W+hf}}$$

Table 6.16 shows the tagging efficiencies for $W+$ heavy flavor jets as a function of the number of heavy flavor jets in the event and jet multiplicity.

	$1\ b-jet$	$2\ b-jets$		$1\ c-jet$	$2\ c-jets$		Wc
	$\epsilon_{\geq 1\ \text{tag}}$	$\epsilon_{\geq 1\ \text{tag}}$	$\epsilon_{\geq 2\ \text{tags}}$	$\epsilon_{\geq 1\ \text{tag}}$	$\epsilon_{\geq 1\ \text{tag}}$	$\epsilon_{\geq 2\ \text{tags}}$	
SECVTX							
$W+1j$	24.7±0.8	–	–	4.56±0.29	–	–	4.1±0.4
$W+2j$	21.7±1.7	46.1±1.8	10.6±1.2	3.69±0.49	11.7±1.1	0.4±0.2	4.5±0.6
$W+\geq 3j$	20.7±4.4	47.0±4.0	10.7±2.8	3.88±1.26	14.4±2.3	0.0±0.0	4.5±0.6
Jetprobability							
$W+1j$	23.9±0.7	–	–	9.9±0.4	–	–	8.7±0.4
$W+2j$	20.4±1.4	40.9±1.5	10.0±0.9	7.8±0.7	24.9±1.4	2.8±0.5	10.8±1.0
$W+\geq 3j$	21.9±3.9	43.5±3.4	9.3±1.9	13.0±2.2	25.8±2.9	1.7±0.8	16.7±2.9
SLT							
$W+1j$	7.6±0.9	–	–	3.8±0.5	–	–	3.2±0.4
$W+2j$	7.8±1.2	13.4±1.7	0.3±0.2	3.3±0.6	6.6±1.0	0.1±0.1	4.4±0.6
$W+\geq 3j$	3.8±1.9	9.4±2.3	0.5±0.5	5.5±1.6	8.5±2.1	0.0±0.0	6.9±0.6

Table 6.16: Tagging efficiencies (in %) in $Wb\bar{b}$, $Wc\bar{c}$ and Wc events. Jets are clustered with a cone of radius 0.4 and required to have $E_T \geq 15$ GeV and $|\eta| \leq 2$. The tagging efficiencies include the appropriate Data/Monte Carlo tagging efficiency scale factors for b and c-quark jets.

The shown efficiencies include the effect of the reduced reconstruction efficiency of tracks contained in a jet and the effect of the high luminosity conditions to the

track reconstruction efficiency. Also there is a factor of 0.97 for the combined tagging efficiency between Run 1A and Run 1B. The SECVTX and Jetprobability b-tagging efficiencies have been rescaled according to the Data/Monte Carlo scale factors determined in Section 4.5 by comparing the b-tagging efficiency after track degradation in low-P_T electron data and Monte Carlo $b\bar{b}$ samples. As a reminder, the b-tagging efficiency scale factors were measured to be $SF_{b-tag}^{SECVTX} = 1.25 \pm 0.13$ for SECVTX and $SF_{b-tag}^{JPB} = 0.95 \pm 0.1$ for Jetprobability. The corresponding scale factors for the c-tagging efficiencies are propagated with the track degradation procedure. The SLT tagging efficiency is measured after track degradation and there is no additional scale factor since the efficiencies measured with the data are incorporated directly in the Monte Carlo simulations. Because of the limited Monte Carlo statistics in the $W+ \geq 3$ jet multiplicity bin, the two multiplicity bins are collapsed into one and the measured tagging efficiency is used for both $W + 3$ and $W+ \geq 4$ jet bin. This is justified by the fact the tagging efficiency is quite flat across the jet multiplicity bins. For the case of the Jetprobability however, the simulation includes a $\sim 5\%$ mistag rate per jet. Therefore it is likely that the tagging efficiency in the $W + 3$ jet bin would be overestimated by 1% while the corresponding efficiency in the $W+ \geq 4$ jets would be underestimated by 4%. This difference is added to the systematic error.

Errors on the tagging efficiency reflect the uncertainty from the Monte Carlo statistics and the uncertainty associated with the value of the b-tagging efficiency scale factors. The statistical error ranges between $\pm 3 \sim 16\%$ for the single tagging efficiency and it is at the level of $\pm 25 \sim 50\%$ for the double tagging efficiency. Comparisons of the tagging efficiency measured in HERWIG Monte Carlo samples requiring two b-jets in the event to the one measured in VECBOS samples are shown in Table 6.17. The measured tagging efficiencies in the two different simulations are

VECBOS						
	SECVTX		Jetprobability		SLT	
Sample	2 b-jets	2 c-jets	2 b-jets	2 c-jets	2 b-jets	2 c-jets
$W+2$ jets	39.0±1.6	12.1±1.1	43.8±1.6	25.6±1.4	13.3±1.1	6.6±0.8
$W+\geq 3$ jets	39.8±3.5	14.8±2.4	46.6±3.6	26.5±3.0	9.4±2.1	8.5±1.9
HERWIG						
	SECVTX		Jetprobability		SLT	
Sample	2 b-jets	2 c-jets	2 b-jets	2 c-jets	2 b-jets	2 c-jets
$W+2$ jets	37.8±3.5	9.7±2.3	43.1±3.6	18.8±2.8	16.0±2.7	6.1±1.9
$W+\geq 3$ jets	40.0±6.3	9.6±4.1	50.0±6.5	17.4±5.2	13.3±4.4	7.7±3.7

Table 6.17: Single tagging efficiencies (in %) in HERWIG and VECBOS events with 2 b and 2 c jets.

within the statistical error of the measurement implying that once the events are divided in events with one or two heavy flavor jet events, the tagging efficiency is independent of the modeling of the kinematics of the event. The shown efficiencies in Table 6.17 include only the effect of track degradation.

6.7.4 Estimate of the $Wb\bar{b}$ and $Wc\bar{c}$ contributions

At this point, all the ingredients needed to calculate the $Wb\bar{b}$ and $Wc\bar{c}$ contributions to the $W+\geq 1$ jet tagged sample are available and the calculation as a function of jet multiplicity, j, proceeds using the following expression:

$$N_{W+hf}^{[Njet]} = N_{Real\ W}^{[Njet]} \cdot (F1_{W+hf}^{[Njet]} \cdot \epsilon 1_{\geq 1\ tag}^{[Njet]} + F2_{W+hf}^{[Njet]} \cdot \epsilon 2_{\geq 1\ tag}^{[Njet]})$$

where:

- $N_{W+hf}^{[Njet]}$: is the number of $Wb\bar{b}$ or $Wc\bar{c}$ event contributing tags in the j jet multiplicity bin.

- $N^{[Njet]}_{Real\ W}$: is the number of W candidate events in the j jet bin after subtracting from the observed number of W candidates the contribution of non-W's, residual $Z+$ jets, single top, and events from diboson production discussed in the previous sections.

- $F1^{[Njet]}_{W+hf}$: is the fraction of $Wb\bar{b}$ or $Wc\bar{c}$ events in the j multiplicity bin with exactly 1 b or 1 c-jet at the final state, as listed in Table 6.14.

- $\epsilon 1^{[Njet]}_{\geq 1\ tag}$: is the probability to tag at least one jet in $Wb\bar{b}$ or $Wc\bar{c}$ events with 1 b or 1 c-jet at final state as given in Table 6.16.

- $F2^{[Njet]}_{W+hf}$: is the fraction of $Wb\bar{b}$ or $Wc\bar{c}$ events in the j multiplicity bin with 2 b or 2 c-jets at the final state, as listed in Table 6.14.

- $\epsilon 2^{[Njet]}_{\geq 1\ tag}$: is the probability to tag at least one jet in $Wb\bar{b}$ or $Wc\bar{c}$ events with 2 b or 2 c-jets at final state as given in Table 6.16.

The combination of single and double tag efficiencies it can be used to measured the yields of $W+$ heavy flavor events with single and double tags in the $W+ \geq 1$ jet sample. It is should be mentioned that double tagged events appear only in events with 2 b or 2 c-jets as given by the VECBOS Monet Carlo. The calculation proceeds according to the following expressions:

$$N^{[Njet]}_{W+hf+1\ tag} = N^{[Njet]}_{Real\ W} \cdot [F1^{[Njet]}_{W+hf} \cdot \epsilon 1^{[Njet]}_{\geq 1\ tag} + F2^{[Njet]}_{W+hf} \cdot (\epsilon 2^{[Njet]}_{\geq 1\ tag} - \epsilon 2^{[Njet]}_{\geq 2\ tag})]$$

$$N^{[Njet]}_{W+hf+2\ tag} = N^{[Njet]}_{Real\ W} \cdot F2^{[Njet]}_{W+hf} \cdot \epsilon 2^{[Njet]}_{\geq 2\ tag}$$

where $\epsilon 2^{[Njet]}_{\geq 2\ tag}$ is the double tagging efficiency for $Wb\bar{b}$ or $Wc\bar{c}$ events with 2 b or 2 c-jets in the event. Table 6.18 lists the number of $Wb\bar{b}$ and $Wc\bar{c}$ tagged events

expected in the $W+\geq 1$ jet sample as a function of jet multiplicity in the event.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
real W	8614±16	1218±8.67	169.3±3.3	47.0±1.85
$Wb\bar{b}$ before tagging	69.0±9.5	29.9±5.2	6.5±1.2	2.8±0.6
$Wc\bar{c}$ before tagging	173.2±46.2	62.4±13.7	12.9±2.9	4.2±1.3
SECVTX				
$Wb\bar{b}$ single tags	17.10±2.43	8.49±1.77	1.85±0.45	0.75±0.19
$Wb\bar{b}$ double tags		1.52±0.52	0.35±0.15	0.12±0.05
$Wc\bar{c}$ single tags	7.90±2.17	3.62±0.90	0.91±0.29	0.31±0.10
$Wc\bar{c}$ double tags		0.06±0.04	0.00± 0.00	0.00±0.00
Jetprobability				
$Wb\bar{b}$ single tags	16.49±2.33	7.62±1.55	1.82±0.42	0.74±0.18
$Wb\bar{b}$ double tags		1.44±0.48	0.31±0.12	0.10±0.04
$Wc\bar{c}$ single tags	17.14±4.63	7.33±1.75	2.10±0.55	0.71±0.21
$Wc\bar{c}$ double tags		0.48±0.20	0.07±0.04	0.02±0.01
SLT				
$Wb\bar{b}$ single tags	5.26±0.95	3.09±0.70	0.41±0.14	0.16±0.06
$Wb\bar{b}$ double tags		0.05 ±0.03	0.02±0.02	0.01±0.01
$Wc\bar{c}$ single tags	6.55±1.93	2.60±0.66	0.83±0.26	0.28±0.1
$Wc\bar{c}$ double tags		0.02±0.02	0.00±0.00	0.00±0.00

Table 6.18: Yields of single and double SECVTX, Jetprobability and SLT tagged $Wb\bar{b}$ and $Wc\bar{c}$ events as a function of jet multiplicity.

6.7.5 The Wc contribution

Events with a W accompanied by a single c-heavy flavor jet arise from events with a $Wc\bar{s}$ vertex, mainly when the s-quark comes from the strange sea in the proton structure function. The leading order diagrams for Wc production are shown in Fig-

ure 6.11(a). The Wc production can also proceed via the $gg \rightarrow Wc\bar{s}$ mechanism shown in Figure 6.11 (b). This process contributes mainly in the $W + 2$ jet multiplicity bin and it is included in the HERWIG Monte Carlo calculations through the parton shower stage. Its contribution is explicitly calculated in the VECBOS Monte Carlo. Comparison between the Wc event fractions predicted by VECBOS and the ones predicted by HERWIG, show very good agreement as a function of jet multiplicity [62] and therefore the HERWIG generator is used for the calculation of the Wc contribution.

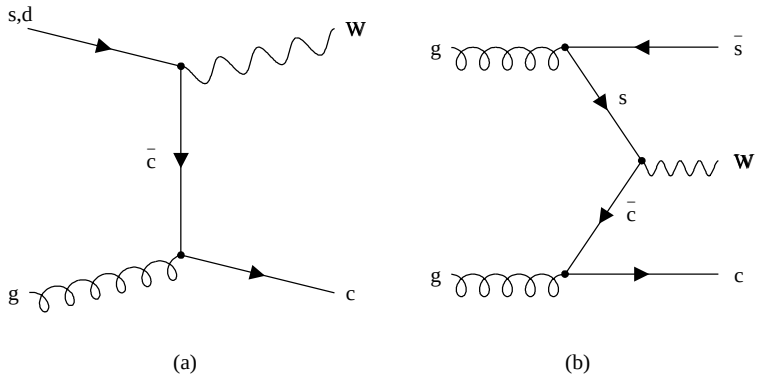

Figure 6.11: Tree level diagrams for Wc production. Diagram (a) shows the dominant production mechanism, $sg \rightarrow Wc$ with a $\sim$10% contribution from $dg \rightarrow Wc$. An uncertainty of $\sim$ 30% is assigned to the contribution of the Wc due to the uncertainty on the strange content of the proton. Diagram (b) corresponds to higher order in α_s, contribution of Wc events to the $W + 2$ jets. This diagram is calculated explicitly in the VECBOS Monte Carlo, while it is included in the HERWIG Monte Carlo through the parton shower approach.

The calculation begins by computing in the Monte Carlo samples, the fraction of the $W + N$ jet events that satisfy the $W+ \geq 1$ jet selection criteria and contain a single c-quark. Calculating the Wc event fractions directly from the Monte Carlo

avoids the problem of having to normalize the overall $W+$ jets theoretical predictions, which contain large uncertainties due to Q^2 dependence.

The Wc production is dominated by the excitation of a s-quark from the proton sea and therefore the Wc event fraction, F_{Wc}, is sensitive to the knowledge of the s-quark content of the proton and consequently it is sensitive to the choice of the structure function used in the calculations. Studies of the F_{Wc} using a variety of structure functions shows that the fraction of Wc events changes by $\sim$30% [62] and [65]. The observed variation is taken as the systematic uncertainty of the Wc calculations. The central value of F_{Wc} used in this analysis is calculated based on the MRSD0′ structure function parametrization. The measured F_{Wc} as a function of jet multiplicity is shown in Table 6.14.

HERWIG Monte Carlo samples of $W+ \geq 1$ jets are used to calculate the Wc event fractions. In each jet multiplicity bin, the Wc event fraction is calculated as the ratio of W events that pass all $W+ \geq 1$ jet selection criteria and contain a Wcs vertex to the total number of W events satisfying the analysis requirements in this jet multiplicity bin. The efficiency, ϵ_{Wc}^{tag}, for tagging the resulting c-quark jet is also estimated with the same Monte Carlo simulations. The tagging efficiency for each tagging algorithm is calculated as a function of the jet multiplicity and it is listed in Table 6.16. The expected number of Wc tagged events in each multiplicity bin is determined according to the following expression:

$$N_{Wc\ tags} = N_W^{real} \times F_{Wc} \times \epsilon_{Wc}^{tag}$$

N_W^{real} is the number of W candidate events after subtraction of events due to fake W's, diboson and single top production, and misidentified $Z+$ jets including events

from $Z \rightarrow \tau\tau$ production as derived in the previous sections.

Table 6.19 lists the number of Wc events expected to be tagged by any of the tagging algorithms as a function of the jet multiplicity.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
real W	8614±16	1218±8	169.8±3.3	47.0±1.85
before tagging	413.5±124.0	87.7±26.3	12.7±3.8	3.5±1.1
with SECVTX tags	16.90±5.39	3.98±1.31	0.58±0.19	0.16±0.05
with JPB tags	35.06±10.65	9.16±2.88	2.06±0.72	0.57±0.20
with SLT tags	13.40±4.35	3.85±1.34	0.88±0.38	0.24±0.10

Table 6.19: The expected number of Wc events before and after tagging as a function of the jet multiplicity in the event. The first row of the table shows the number of real W candidate events derived after subtracting from the observed W yield the event contributions from all other non-W processes.

6.8 $Z+$ heavy flavor production

Production of heavy flavor jets in association with a Z boson proceeds via similar to the $W+$ heavy flavor production mechanisms. The $Z+$ heavy flavor contributions can be divided into four distinct classes according to the flavor type and production mechanism:

1. Events with a $b\bar{b}$ pair at the final state originating from the splitting of one of the produced gluons. Tree level diagrams of this process are shown in Figure 6.12(a) and (b). This process is identical to the $W + b\bar{b}$ production.

2. Events with a $c\bar{c}$ pair at the final state similar to the $b\bar{b}$ production.

3. Events with a $Zb\bar{b}$ vertex from a process like $gg \rightarrow Zb\bar{b}$, resulting in one or two b-quark jets at the final state. Tree level diagrams of this process are shown in Figure 6.12(c) and (d).

4. Events with a $Zc\bar{c}$ vertex resulting in one or two c-quark jets at the final state. The process is analogous to the Wc production mechanism and the Wc diagrams shown in Figure 6.11.

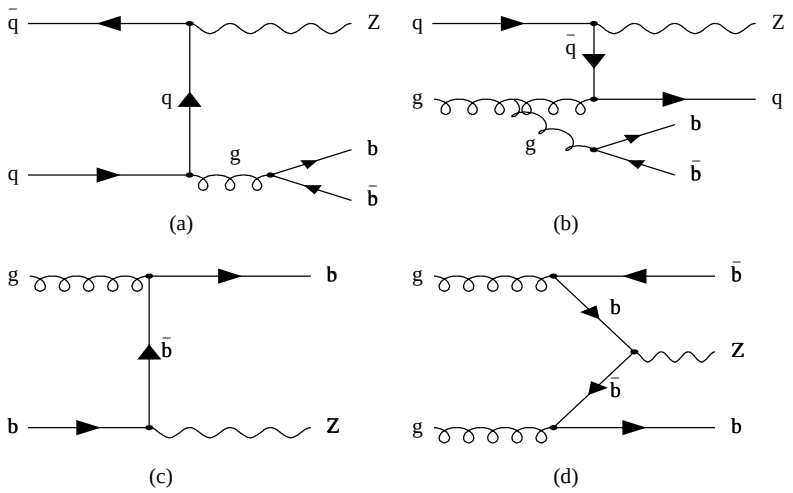

Figure 6.12: Leading order Feynman diagrams for $Z + b\bar{b}, c\bar{c}$ production. Diagrams (a) and (b) are identical to the ones producing $W + b\bar{b}$, $c\bar{c}$ events. Diagrams (c) and (d) contribute to events with a $Zb\bar{b}$ vertex which do not have an analogous $Wb\bar{b}$ production mechanism. Replacing the b quarks with c the analogous to $Wc\bar{c}$ diagrams are obtained. In this case, diagrams (c) and (d) are similar to the Wc production diagrams shown in Figure 6.11.

The fraction of Z events associated with heavy flavor production is calculated in similar manner to the $Wb\bar{b}$ and $Wc\bar{c}$ calculations. The event fractions are mesured

before tagging using VECBOS Monte Carlo samples interfaced with the HERWIG Monte Carlo for the parton shower. To estimate the fraction of events with a $Zb\bar{b}$ and $Zc\bar{c}$ vertex, the PYTHIA generator is used because the available version of the HERWIG Monte Carlo does not include the $Z + 1$ jet process. The procedure is detailed in the following two sections.

The obtained $Z+$ heavy flavor event fractions are then multiplied with the corresponding tagging efficiencies as determined in the $W+$ heavy flavor calculations (see Table 6.16) and the estimated residual number of Z events in the W signal region:

$$N^j_{Z+hf} = N^j_{Z\ res} \cdot F^{tag}_{Z+hf}$$

$$F^{tag}_{Z+hf} = F^{ij}_{Z+hf} \cdot \epsilon^{ij}_{Z+tag\ hf}$$

N^j_{Z+hf} is the number of $Z+$ heavy flavor tagged events contributing to the $W + j$ jets tagged sample, $N^j_{Z\ res}$ is the number of misidentified Z's in the $W + j$ jet bin. F^{tag}_{Z+hf} is the $Zb\bar{b}$, $Zc\bar{c}$ or Zc event fraction after tagging, F^{ij}_{Z+hf} is the fraction of $Z+$ heavy flavor events in the jth jet multiplicity bin. i represents the number of b or c jets at the final state that can be tagged. $\epsilon^{ij}_{Z+tag\ hf}$ is the efficiency to tag a $Z + j$ jets event with one or two heavy flavor jets.

In the N^j_{Z+hf} expression given above, it is assumed that the heavy flavor content of misidentified Z events is the same with the heavy flavor of all $Z+ \geq 1$ jet events. The assumption is valid for $Z \rightarrow \mu^-\mu^+$ events since the Z is misidentified either because one of the muon is either lost due to the limited detector muon coverage or because it fails some of the identification criteria. In any case, the misidentified or lost muon does not alter the final jet multiplicity of the event. In contrast, for $Z \rightarrow e^-e^+$ events, mismeasurement of one of the Z electrons can alter the jet multiplicity because

the electron is counted as one of the jets in the event. Unlike the other jets in the event, the electron jet has no heavy flavor content. In this case, the heavy flavor contribution is calculated by multiplying the fraction of events with a an electron jet with the tagging efficiency of the previous bin before adding it to the estimates of the current jet multiplicity bin. To account for this problem, a $Z+ \geq 1$ jet sample generated with the PYTHIA generator was used to estimate the fraction of $Z+ \geq 1$ jet events satisfying the $W+ \geq 1$ jet selection criteria and containing a jet from one of the Z electron legs. The jet-electron association is performed by requiring matching of the observed jet with one of the Z legs at generator level. It is found that $15.8 \pm 5.7\%$ of the residual $Z + 2$ jets in the W sample are actually $Z + 1$ jet events with an electron counted as a jet. Similarly for the $Z+ \geq 3$ jets, the measured fraction of events with an electron jet is $8.3 \pm 8.7\%$. Accordingly, the above expression for N^j_{Z+hf} is modified as follows:

$$N^j_{Z+hf} = N^j_{Z\ res} \cdot F^{j\ tag}_{Z+hf} + N_{Z\ res\ e-jet} \cdot F^{(j-1)\ tag}_{Z+hf}$$

$$F^{j\ tag}_{Z+hf} = F^{ij}_{Z+hf} \cdot \epsilon^{ij}_{Z+tag\ hf}$$

6.8.1 The $Zb\bar{b}$ and Zb contribution

The fraction of $Z + b\bar{b}$ events are calculated for two classes of events. The first class contains events in which the b quarks originate from final state gluon splitting. As for the $Wb\bar{b}$ case, the gluon splitting can result in one or two b jets in the event. Because HERWIG does not contain the matrix elements for the $Z + 1$ jet process, the fraction of events with one and two b-jets are calculated with the VECBOS Monte Carlo. To estimate the fraction of events with one b-jet at final state, the HERPRT interface is

used and therefore it is left to HERWIG Monte Carlo to simulate the gluon splitting. For the two b-jets case, the event fraction is estimated only with a VECBOS sample.

The second class, contains events with a $Zb\bar{b}$ vertex. The event fraction for this class is estimated with the PYTHIA Monte Carlo generator which includes both processes shown with the Feynmann diagrams 6.12(c) and (d). The fraction of events with a $Zb\bar{b}$ vertex is shown in Table 6.20. The shown event fractions include all possible production mechanisms for b-jets in association with a Z. The $Z+$ heavy flavor fractions before applying the Data/Monte Carlo normalization for the gluon splitting are calculated in Reference [62].

	$Zb\bar{b}$ vertex (%)		$Z + g \to b\bar{b}$ (%)		$Z + g \to c\bar{c}$ (%)	
Sample	1b-jet	2b-jets	1b-jet	2b-jets	1c-jet	2c-jets
$Z + 1$ jet	0.78±0.11		0.80±0.11		2.17±0.60	
$Z + 2$ jet	1.08±0.20	1.01±0.20	1.08±0.33	1.45±0.21	4.55±1.32	1.88±0.51

Table 6.20: Fraction of events with a $Zb\bar{b}$ vertex as calculated from a $Z + 1$ jet PYTHIA Monte Carlo sample and the fraction of Z events with b and c-jets from final state gluon splitting. The above fractions include the 1.39±0.19 and 1.35±0.36 factors for the normalization of the $g \to b\bar{b}$ and $g \to c\bar{c}$ between Data and Monte Carlo.

The $Zb\bar{b}$ event fractions after tagging are calculated using the event fractions from Table 6.20 and multiplying with the relevant tagging efficiencies from Table 6.16. Comparison between the $Zb\bar{b}$ and $Wb\bar{b}$ event fractions after tagging shows that the ratio of event fractions is $Zb\bar{b}/Wb\bar{b} = 2.0 \pm 0.5$ for the 1 and 2 jet multiplicity bins. Therefore it is assumed that the same ratio holds for the $\geq$3 jet multiplicity bins.

Summarizing, in every jet multiplicity bin j, the $Zb\bar{b}$ event fractions after tagging, $F^{tag}_{Zb\bar{b}}$, is calculated from the corresponding $Wb\bar{b}$ event fraction multiplying with 2 ± 0.5,

$$F^{tag}_{Zb\bar{b}} = (2.0 \pm 0.5) \cdot F^{tag}_{Wb\bar{b}}$$

6.8.2 The $Zc\bar{c}$ and Zc contribution

The $Zc\bar{c}$ contribution due to final state gluon splitting is calculated similar to the $Z + b\bar{b}$ case. The corresponding event fractions after multiplying with the 1.35 ± 0.36 scale factor for the $g \to c\bar{c}$ cross section between Data and Monte Carlo are shown in Table 6.20. Calculating the $Wc\bar{c}$ event fractions after tagging, it is found that the ratio, $Zc\bar{c}/Wc\bar{c} = 1 \pm 0.3$. Therefore the $F^{tag}_{Zc\bar{c}}$ is calculated with the expression:

$$F^{tag}_{Zc\bar{c}} = (1.0 \pm 0.3) \cdot F^{tag}_{Wc\bar{c}}$$

The production mechanism of Zc is considered analogous to the Wc production. The Zc event fraction is derived from the Wc event fraction scaling for the relative fraction of the c/s content of the proton as given by the structure function, and the ratio of the Zc/Wc coupling strengths:

$$F_{Zc} = (0.3 \pm 0.15) \cdot F_{Wc}$$

The number of residual Z events in the $W+ \geq 1$ jet sample before tagging and the $Zb\bar{b}$, $Zc\bar{c}$ and Zc event tagging fractions are shown in Table 6.21. The amount of $Z+$ heavy flavor events for each production mechanism can be estimated by multiplying the corresponding event tagging fractions with the number of residual Z's. The heavy

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
Before tagging	234.8±14.5	38.5±5.9	7.9±2.4	0.7±07
with electron jet		7.3±2.4	0.7±0.7	0.0±0.0
SECVTX event tag fraction (%)				
Zc	0.059±0.02	0.098±0.03	0.102±0.03	0.10±0.03
$Zb\bar{b}$ single tags	0.397±0.06	1.394±0.29	2.189±0.53	3.183±0.80
$Zb\bar{b}$ double tags		0.250±0.09	0.418±0.17	0.508±0.21
$Zc\bar{c}$ single tags	0.092±0.03	0.297±0.07	0.536±0.17	0.663±0.21
$Zc\bar{c}$ double tags		0.005±0.00	0.0±0.0	0.0±0.0
Jetprobability event tag fraction (%)				
Zc	0.122±0.04	0.226±0.07	0.365±0.13	0.365±0.13
$Zb\bar{b}$ single tags	0.383±0.05	1.252±0.25	2.151±0.49	3.168±0.77
$Zb\bar{b}$ double tags		0.236±0.08	0.364±0.14	0.441±0.17
$Zc\bar{c}$ single tags	0.199±0.05	0.602±0.14	1.240±0.32	1.503±0.45
$Zc\bar{c}$ double tags		0.039±0.02	0.039±0.02	0.050±0.03
SLT events tag fraction (%)				
Zc	0.047±0.02	0.095±0.03	0.156±0.06	0.156±0.06
$Zb\bar{b}$ single tags	0.122±0.02	0.507±0.11	0.489±0.16	0.691±0.24
$Zb\bar{b}$ double tags		0.008±0.01	0.020±0.02	0.024±0.03
$Zc\bar{c}$ single tags	0.076±0.02	0.214±0.06	0.489±0.15	0.588±0.20
$Zc\bar{c}$ double tags		0.001±0.00	0.0±0.0	0.0±0.0

Table 6.21: The estimated number of residual Z in the $W+\geq 1$ jet sample before tagging and the event tagged fractions for SECVTX, Jetprobability and SLT. The $Z+$ heavy flavor contribution is estimated multiplying the event tagged fractions with the number of $Z+$ jets in the sample. For Z events in which one jet originates from the mismeasured electron, the heavy flavor contribution is added to the previous bin.

flavor contribution of events with a jet originating from a Z electron is added to the previous jet bin.

6.9 Summary of the $W+ \geq 1$ jet sample composition before and after tagging

Tables 6.23 to 6.25 summarize the total number of expected events tagged by SECVTX, Jetprobability and SLT in the $W+ \geq 1$ jet sample as a function of jet multiplicity. The results correspond to 108.8 ± 7.86 pb^{-1} of data collected during the run 1A and 1B of the collider run. A comparison between the expected contributions and the observed number of tagged events for each tagging algorithm is shown in Figures 6.13 to 6.15. Using the expected composition of the tagged $W+ \geq 1$ jet sample and the tagging efficiencies for each contributing process, the composition of the W sample before tagging can be derived. The breakdown of the $W+ \geq 1$ jet composition before tagging is listed in Table 6.22.

As discussed before, the derivation of the $W+ \geq 1$ jet composition is based on the null hypothesis for $t\bar{t}$ production. An excess of tagged events over the expected Standard Model contributions is observed in the $W+ \geq 3$ jet region for all three tagging algorithms. Data and Monte Carlo predictions show very good agreement in the $W+1$ jet region. This region in which small contribution from $t\bar{t}$ production is expected, serves also as a control sample for the method used to calculate the contribution of the different processes. A small excess of events over the predictions is also observed in the $W+2$ jet region.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
non-W	560.4± 14.9	71.3± 2.7	12.4± 2.0	5.1± 1.7
WW	31.2± 5.4	31.1± 5.4	5.2± 1.0	0.8± 0.2
WZ	4.4± 0.9	4.8± 1.0	0.9± 0.2	0.1± 0.0
ZZ	0.3± 0.1	0.4± 0.1	0.1± 0.0	0.0± 0.0
$Z \rightarrow \tau\tau$	35.2± 2.2	13.1± 1.2	1.6± 0.4	0.3± 0.2
$Z+$ jets	234.8± 14.5	38.5± 5.9	7.9± 2.4	0.7± 0.7
single top	15.1± 2.3	9.1± 2.0	2.0± 0.4	0.3± 0.1
Wc	413.5±124.0	87.7± 26.3	12.7± 3.8	3.5± 1.1
$Wb\bar{b}$	69.1± 9.5	30.0± 5.2	6.5± 1.2	2.8± 0.6
$Wc\bar{c}$	173.3± 46.3	62.5± 13.7	12.9± 2.9	4.2± 1.3
$W+$ non$-$h.f.	7958.0±133.7	1037.7± 31.2	137.4± 5.6	36.4± 1.9

Table 6.22: Breakdown of the composition of the $W+\geq 1$ jet sample before tagging as a function of the jet multiplicity. The estimated and observed numbers of events correspond to 108.8±7.86 pb^{-1} of data collected at CDF. The calculations of the different contributions it is based on the assumption that there is no contributions to the $W+\geq 1$ jet sample from $t\bar{t}$ events.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
with single tags	66	35	10	11
with double tags		5	6	2
mistags	10.82± 1.08	3.80± 0.38	0.99± 0.10	0.35± 0.04
non-W	8.18± 0.78	1.49± 0.47	0.76± 0.38	0.31± 0.16
WW, WZ, ZZ	0.52± 0.14	1.38± 0.28	0.40± 0.13	0.00± 0.00
$Z \rightarrow \tau\tau$	0.96± 0.30	0.70± 0.25	0.17± 0.12	0.00± 0.00
single top	1.60± 0.41	2.74± 0.63	0.71± 0.16	0.15± 0.04
Wc	16.90± 5.39	3.98± 1.31	0.58± 0.19	0.16± 0.05
$Wc\bar{c}$ single tags	7.90± 2.17	3.62± 0.90	0.91± 0.29	0.31± 0.10
$Wc\bar{c}$ double tags		0.06± 0.04	0.00± 0.00	0.00± 0.00
$Wb\bar{b}$ single tags	17.10± 2.43	8.49± 1.77	1.85± 0.45	0.75± 0.19
$Wb\bar{b}$ double tags		1.52± 0.52	0.35± 0.15	0.12± 0.05
Zc	0.14± 0.04	0.03± 0.01	0.01± 0.00	0.00± 0.00
$Zc\bar{c}$ single tags	0.22± 0.06	0.10± 0.03	0.04± 0.02	0.00± 0.00
$Zc\bar{c}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$Zb\bar{b}$ single tags	0.93± 0.14	0.46± 0.12	0.17± 0.06	0.02± 0.02
$Zb\bar{b}$ double tags		0.08± 0.03	0.03± 0.02	0.00± 0.00
total single tags	65.26± 6.46	26.87± 2.57	6.64± 0.74	2.07± 0.28
total double tags		1.59± 0.52	0.35± 0.15	0.12± 0.05
excess with single tags	0.74±10.34	8.13± 5.78	3.36± 2.68	8.93± 1.47
excess with double tags		3.41± 1.37	5.65± 0.61	1.88± 0.35
excess of tagged events	0.74±10.34	11.54± 5.94	9.01± 2.75	10.81± 1.51

Table 6.23: Summary of the observed and predicted number of events with SECVTX tags in the $W+\geq 1$ jet sample. Good agreement is observed in the $W+1$ jet bin while an excess of events over the expectation in the $W+\geq 3$ jet bin is observed. Small excess of events is observed also in the $W+2$ jet bin.

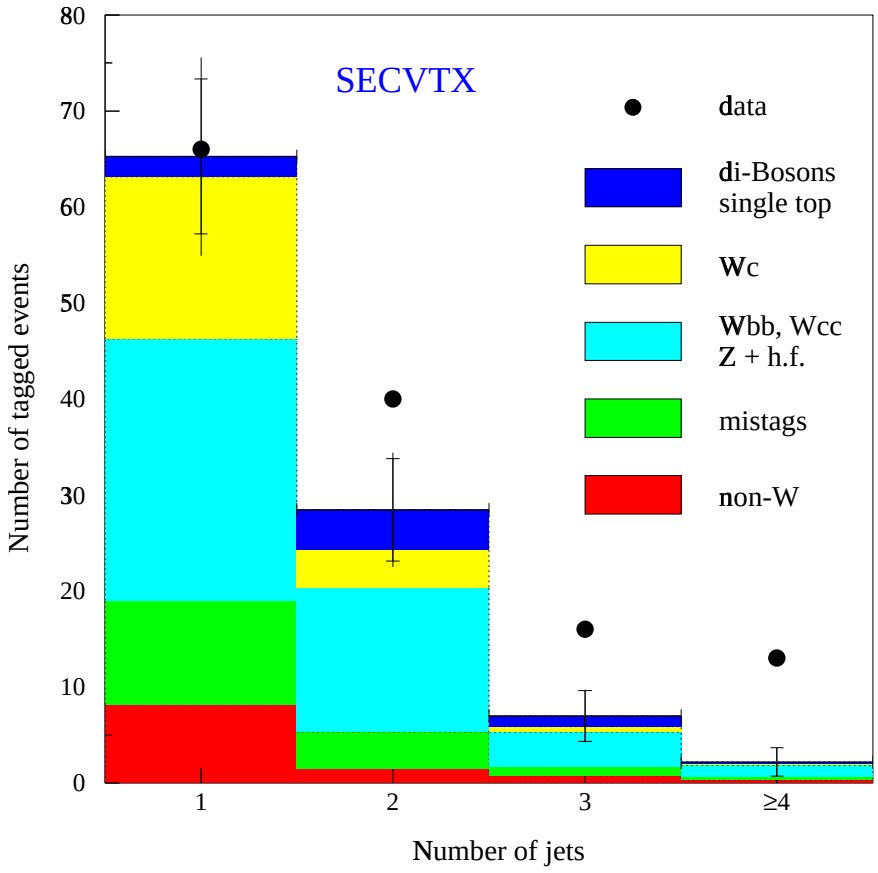

Figure 6.13: Comparison between the observed and predicted number of events with SECVTX tags in the $W+\geq 1$ jet sample. Jets are clustered with cone radius 0.4 and required to have $E_T \geq 15$ GeV and $|\eta| \leq 2$. The vertical bars represent the overall uncertainty on the expected number of events and the horizontal ticks on the bars corresponds to the contribution from the statistical uncertainty alone. There is very good agreement between observed and predicted tag rates in the $W+1$ jet bin while large excess is observed in the $W+\geq 3$ jet region as expected from $t\bar{t}$ production.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
with single tags	125	62	21	12
with double tags		6	5	3
mistags	41.87± 4.24	12.93± 1.35	2.67± 0.28	0.84± 0.09
non-W	12.55± 0.95	2.53± 0.61	0.57± 0.33	0.24± 0.14
WW, WZ, ZZ	1.15± 0.26	2.39± 0.43	0.74± 0.19	0.05± 0.04
$Z \to \tau\tau$	2.35± 0.47	1.13± 0.32	0.17± 0.12	0.09± 0.09
single top	1.54± 0.38	2.53± 0.60	0.68± 0.16	0.12± 0.03
Wc	35.06±10.65	9.16± 2.88	2.06± 0.72	0.57± 0.20
$Wc\bar{c}$ single tags	17.14± 4.63	7.33± 1.75	2.10± 0.55	0.71± 0.21
$Wc\bar{c}$ double tags		0.48± 0.20	0.07± 0.04	0.02± 0.01
$Wb\bar{b}$ single tags	16.49± 2.33	7.62± 1.55	1.82± 0.42	0.74± 0.18
$Wb\bar{b}$ double tags		1.44± 0.48	0.31± 0.12	0.10± 0.04
Zc	0.29± 0.09	0.08± 0.03	0.03± 0.01	0.00± 0.00
$Zc\bar{c}$ single tags	0.47± 0.13	0.20± 0.06	0.09± 0.04	0.01± 0.01
$Zc\bar{c}$ double tags		0.01± 0.01	0.00± 0.00	0.00± 0.00
$Zb\bar{b}$ single tags	0.90± 0.14	0.42± 0.10	0.16± 0.06	0.02± 0.02
$Zb\bar{b}$ double tags		0.07± 0.03	0.03± 0.01	0.00± 0.00
total single tags	129.81±12.64	46.41± 4.07	11.14± 1.12	3.39± 0.39
total double tags		1.92± 0.52	0.37± 0.12	0.13± 0.04
excess with single tags	-4.81±17.02	15.59± 7.94	9.86± 3.52	8.61± 1.88
excess with double tags		4.08± 1.48	4.63± 0.62	2.87± 0.36
excess of tagged events	-4.81±17.02	19.68± 8.07	14.49± 3.58	11.48± 1.92

Table 6.24: Summary of the observed and predicted number of events with Jetprobability tags in the $W+\geq 1$ jet sample. Good agreement is observed in the $W+1$ jet bin while an excess of events in the $W+\geq 3$ jet bin is observed. Small excess of events is observed also in the $W+2$ jet bin.

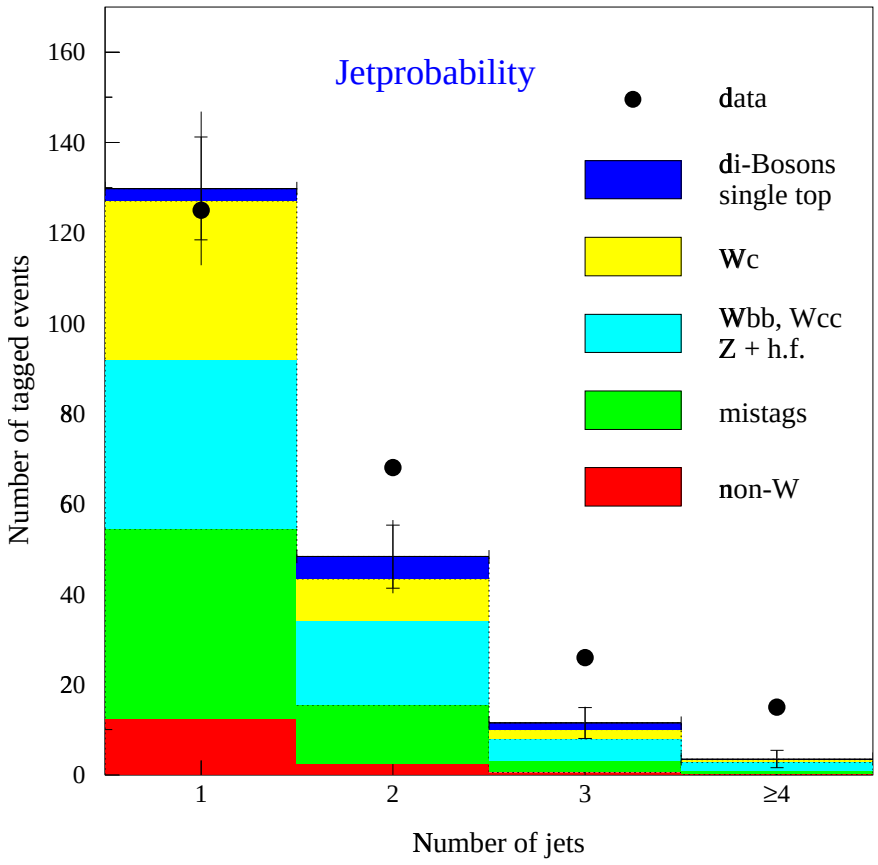

Figure 6.14: Comparison between the observed and predicted number of events with Jetprobability tags in the $W+ \geq 1$ jet sample. Jets are clustered with cone radius 0.4 and required to have $E_T \geq 15$ GeV and $|\eta| \leq 2$. The vertical bars represent the overall uncertainty on the expected number of events and the horizontal ticks on the bars corresponds to the contribution from the statistical uncertainty alone. There is very good agreement between observed and predicted tag rates in the $W+1$ jet bin while an excess of events is observed in the $W+ \geq 3$ jet region.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
with single tags	146	56	17	8
with double tags		0	0	0
mistags	102.00±10.20	31.10± 3.11	7.34± 0.73	3.01± 0.30
non-W	8.97± 0.84	2.09± 0.56	0.38± 0.27	0.16± 0.11
WW, WZ, ZZ	0.26± 0.10	0.94± 0.23	0.10± 0.05	0.00± 0.00
$Z \to \tau\tau$	0.54± 0.23	0.09± 0.09	0.09± 0.09	0.00± 0.00
single top	0.44± 0.12	0.75± 0.17	0.21± 0.05	0.05± 0.02
Wc	13.40± 4.35	3.85± 1.34	0.88± 0.38	0.24± 0.10
$Wc\bar{c}$ single tags	6.55± 1.93	2.60± 0.66	0.83± 0.26	0.28± 0.10
$Wc\bar{c}$ double tags		0.02± 0.02	0.00± 0.00	0.00± 0.00
$Wb\bar{b}$ single tags	5.26± 0.95	3.09± 0.70	0.41± 0.14	0.16± 0.06
$Wb\bar{b}$ double tags		0.05± 0.03	0.02± 0.02	0.01± 0.01
Zc	0.11± 0.04	0.03± 0.01	0.01± 0.01	0.00± 0.00
$Zc\bar{c}$ single tags	0.18± 0.05	0.07± 0.02	0.04± 0.02	0.00± 0.00
$Zc\bar{c}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$Zb\bar{b}$ single tags	0.29± 0.05	0.17± 0.05	0.04± 0.02	0.00± 0.01
$Zb\bar{b}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
total single tags	137.99±11.33	44.78± 3.58	10.33± 0.92	3.91± 0.36
total double tags		0.07± 0.04	0.02± 0.02	0.01± 0.01
excess with single tags	8.01±16.32	11.22± 7.59	6.67± 3.34	4.09± 2.01
excess with double tags		-0.07± 0.26	-0.02± 0.13	-0.01± 0.08
excess of tagged events	8.01±16.32	11.16± 7.59	6.65± 3.35	4.09± 2.01

Table 6.25: Summary of the observed and predicted number of events with SLT tags in the $W+\geq 1$ jet sample. Good agreement is observed in the $W+1$ jet bin while an excess of events in the $W+\geq 3$ jet bin is observed. Small excess of events is observed also in the $W+2$ jet bin.

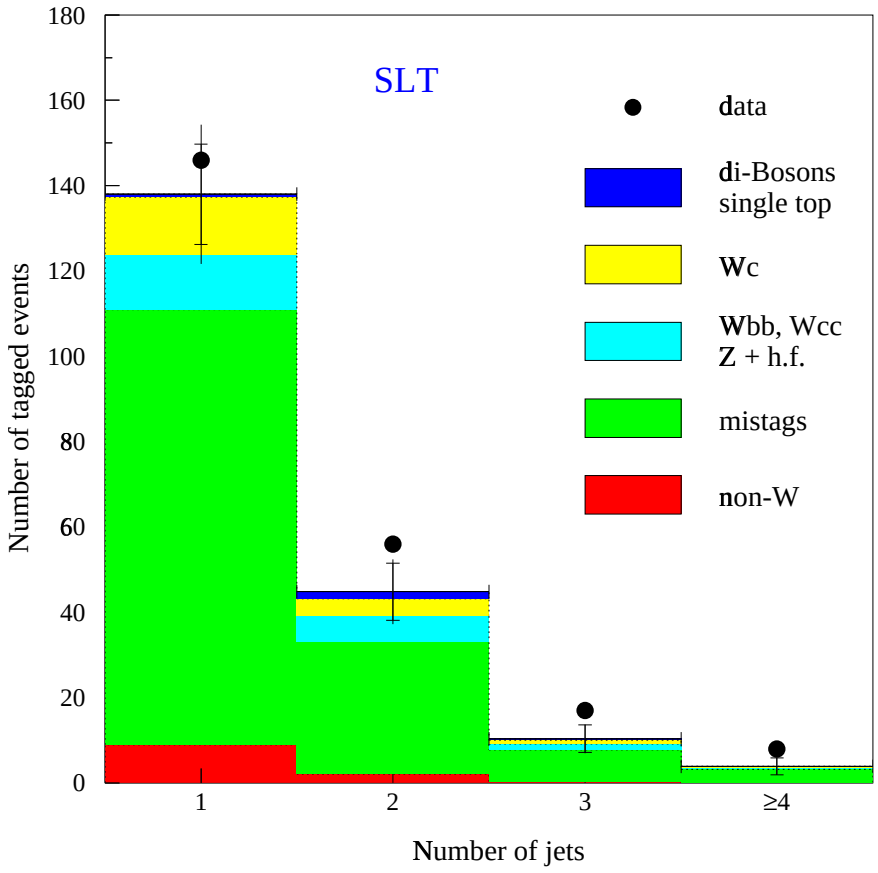

Figure 6.15: Comparison between the observed and predicted number of events with SLT tags in the $W+\geq 1$ jet sample. Jets are clustered with cone radius 0.4 and required to have $E_T \geq 15$ GeV and $|\eta| \leq 2$. The vertical bars represent the overall uncertainty on the expected number of events and the horizontal ticks on the bars corresponds to the contribution from the statistical uncertainty alone. There is very good agreement between observed and predicted tag rates in the $W+1$ and $W+2$ jet bin while an excess of events is observed in the $W+\geq 3$ jet region.

6.10 Cross check of the heavy flavor calculations in the $Z+ \geq 1$ jet sample

The production mechanisms of W and Z in association with jets are very similar. Some differences in the production diagrams of $Z+$ heavy flavor and $W+$ heavy flavor production, have been accounted for as described in section 6.8. In addition, the contribution of $t\bar{t}$ events to the $Z+ \geq 1$ jet sample is very small making the $Z+ \geq 1$ jet sample ideal for testing the $W/Z+$heavy flavor calculation procedure described in the previous sections. Also, if there is a Standard Model process which has not been accounted for or if it has been miscalculated in the calculations of the $W+ \geq 1$ jet sample, it could likely produce the same effect in the $Z+ \geq 1$ jet sample. The underlying assumption for this test procedure is that there is no non-Standard Model process which can contribute in both W and Z processes. Any excess of tagged events in the $Z+ \geq 1$ jet sample will signal either inconsistencies in the heavy flavor calculations or poor understanding of the tagging efficiencies and in the most exciting scenario, contribution from new physics. Unfortunately, the combined cross section times branching ratio for $p\bar{p} \rightarrow Z \rightarrow \ell^+\ell^-$ is an order of magnitude smaller than the cross section for $p\bar{p} \rightarrow W \rightarrow \ell\nu$, resulting in limited statistics for the $Z(\rightarrow \ell^+\ell^-)+ \geq 1$ jet channel.

The $Z+ \geq 1$ jet event sample is selected from the inclusive high-P_T lepton sample as described in section 3.9.4. One lepton is required to pass all selection criteria used for the W sample selection (see Tables 3.1 and 3.4), while looser selection criteria apply on the second lepton (see Table 3.6). Events with the lepton pair invariant mass in the window of $70 \leq M_{\ell\ell} \leq 110$ GeV/c^2 are flagged as Z candidates. The invariant mass distribution and jet multiplicity of the selected Z candidate events are

shown in Figures 3.9 and 3.19 of Section 3.6.1.

The $Z+\geq 1$ jet sample before tagging contains contribution from diboson (WW, WZ, ZZ) production, single top and $t\bar{t}$ production. The contribution of each of the above components is evaluated using the same Monte Carlo samples and normalization procedure as the ones described in the previous sections. For the $t\bar{t}$ contribution, a PYTHIA Monte Carlo sample generated with M_{top}=170 GeV/c^2 is used. The integrated luminosity of the Monte Carlo sample is calculated assuming a value for the $t\bar{t}$ cross section of $\sigma_{t\bar{t}}=5.8^{+0.8}_{-0.5}$ pb based on the calculation of Laenen et al. [80] for a top mass of 170 GeV/c^2.

The tagged event yields obtained after applying each tagging algorithm to the $Z+\geq 1$ jet sample are shown in Table 6.26.

number of events	$Z+1$ jet	$Z+2$ jet	$Z+3$ jet	$Z+\geq 4$ jet
Initial sample	1148	159	16	4
with single SECVTX tags	10	3	0	1
with double SECVTX tags		2	0	0
with single JPB tags	11	5	1	2
with double JPB tags		0	0	0
with single SLT tags	16	3	0	1
with double SLT tags		0	0	0

Table 6.26: Observed events with at least one tag in the $Z+\geq 1$ jet sample as a function of the jet multiplicity.

The observed tags contain contribution from $Z+$ heavy flavor production ($Zb\bar{b}$, $Zc\bar{c}$, Zc, Zb), mistags, and contributions from diboson, single top and $t\bar{t}$ production.

The contribution to the tagging rate from diboson, single top and $t\bar{t}$ production, is calculated from Monte Carlo simulations after scaling the b-quark tagging efficiency by the measured data to Monte Carlo scale factor.

The $Z+$ heavy flavor event rate is estimated by applying the event fractions determined in Section 6.8 to the observed events of the pre-tagged $Z+ \geq 1$ jet data sample after subtracting the pre-tagging contribution of diboson, single top and $t\bar{t}$ production. The resulting event yields are then multiplied with the appropriate b and c jets tagging efficiencies shown in Table 6.16.

Table 6.27 shows the composition of the $Z+ \geq 1$ jet sample before tagging as calculated according to the above procedure.

number of events	$Z+1$ jet	$Z+2$ jet	$Z+3$ jet	$Z+\geq 4$ jet
initial sample	1148	159	16	4
WW	0.8±0.2	0.2±0.1	0.0±0.0	0.0±0.0
WZ	2.2±0.5	1.7±0.4	0.3±0.1	0.1±0.0
ZZ	1.2±0.3	1.6±0.4	0.3±0.1	0.0±0.0
Zc	16.5±4.9	3.3±1.0	0.3±0.1	0.1±0.0
$Zb\bar{b}$	18.3±2.5	7.6±1.3	1.1±0.2	0.4±0.1
$Zc\bar{c}$	23.0±6.1	7.9±1.7	1.1±0.3	0.3±0.1
$Z+$ non−h.f.	1085.3±8.3	135.3±2.5	12.2±0.4	2.9±0.1
single top	0.1±0.0	0.0±0.0	0.0±0.0	0.0±0.0
$t\bar{t}$	0.6±0.1	1.4±0.3	0.5±0.1	0.2±0.0

Table 6.27: The $Z+ \geq 1$ jet sample composition before tagging.

The amount of mistags in each multiplicity bin is calculated with the true mistag probability parametrizations. To avoid double counting the same procedure described in Section 6.5 is also applied in the $Z+ \geq 1$ jet sample. Because the processes contributing background to the $Z+ \geq 1$ jets sample are estimated using Monte Carlo simulations, only the Jetprobability mistags need to be recalculated.

The measured and predicted SECVTX, Jetprobability and SLT tag yields are shown in Tables 6.28 to 6.29.

number of events	$Z+1$ jet	$Z+2$ jet	$Z+3$ jet	$Z+\geq 4$ jet
initial sample	1148	159	16	4
SECVTX tags				
with single tags	10	3	0	1
with double tags		2	0	0
mistags	1.27±0.34	0.34±0.07	0.08±0.02	0.01± 0.01
WW, WZ, ZZ	0.09±0.03	0.18±0.05	0.03±0.01	0.00± 0.00
Zc	0.67±0.21	0.15±0.05	0.02±0.00	0.00± 0.00
$Zc\bar{c}$,$Zb\bar{b}$ single tags	5.56±0.70	2.59±0.46	0.40±0.08	0.14± 0.03
$Zc\bar{c}$,$Zb\bar{b}$ double tags		0.39±0.13	0.06±0.03	0.02±0.01
single top	0.01±0.01	0.00±0.00	0.00±0.00	0.00± 0.00
$t\bar{t}$ single tags	0.22±0.05	0.44±0.09	0.20±0.05	0.03± 0.01
$t\bar{t}$ double tags		0.23±0.06	0.07±0.02	0.03±0.01
Total single tags expected	7.83±0.81	3.70±0.48	0.73±0.10	0.20 ± 0.03
Total double tags expected		0.62±0.14	0.13±0.03	0.04±0.01

Table 6.28: Predicted and observed events with SECVTX tags in the $Z+\geq 1$ jet sample. Jets are clustered with cone radius 0.4 and required to have $E_T \geq 15$ GeV and $|\eta| \leq 2$.

A comparison between the observed and predicted $Z+\geq 1$ jet tag rates is shown in figures 6.16 for SECVTX, Jetprobability and SLT tags respectively.

A χ^2 test between the observed and expected tagging rates in the $Z+\geq 1$ jet sample, using Poisson statistics to describe fluctuations on the expected number of background events, convoluted with a Gaussian smearing to account for the systematic uncertainties on the mean number of background expected events, yields a $\sim$75% probability that the estimated tagging rates are consistent with the observed ones for the $Z+1,2$ jet events. The corresponding probability that the predicted tagging rates are consistent with the observed ones in the $Z+\geq 3$ jet events, varies between $\sim$15% for the SECVTX and SLT and $\sim$5% for the Jetprobability case.

number of events	$Z+1$ jet	$Z+2$ jet	$Z+3$ jet	$Z+\geq 4$ jet
initial sample	1148	159	16	4
Jetprobability tags				
with single tags	11	5	1	2
with double tags		0	0	0
mistags	5.65± 0.57	1.51± 0.15	0.34± 0.04	0.05± 0.01
WW, WZ, ZZ	0.13± 0.03	0.24± 0.06	0.02± 0.01	0.02± 0.01
Zc	1.39± 0.44	0.35± 0.11	0.05± 0.02	0.01± 0.00
$Zc\bar{c}, Zb\bar{b}$ single tags	6.63± 0.87	2.85± 0.45	0.50± 0.09	0.17± 0.03
$Zc\bar{c}, Zb\bar{b}$ double tags		0.42± 0.13	0.06± 0.02	0.02± 0.01
single top	0.01± 0.00	0.00± 0.00	0.00± 0.00	0.00± 0.00
$t\bar{t}$ single tags	0.17± 0.02	0.43± 0.06	0.19± 0.03	0.04± 0.01
$t\bar{t}$ double tag		0.17± 0.02	0.06± 0.01	0.03± 0.00
Total single tags expected	13.98± 1.13	5.37± 0.49	1.11± 0.10	0.30± 0.04
Total double tags expected		0.59± 0.13	0.12± 0.02	0.05± 0.01
SLT tags				
with single tags	16	3	0	1
with double tags		0	0	0
mistags	12.65±1.27	3.66±0.37	0.57±0.06	0.15± 0.02
WW, WZ, ZZ	0.04±0.02	0.09±0.03	0.01±0.01	0.01± 0.01
Zc	0.55±0.17	0.17±0.05	0.02±0.01	0.01± 0.00
$Zc\bar{c}, Zb\bar{b}$ single tags	2.26±0.36	1.10±0.19	0.16±0.03	0.06± 0.01
$Zc\bar{c}, Zb\bar{b}$ double tags		0.02±0.01	0.00±0.00	0.00±0.00
single top	0.00±0.00	0.00±0.00	0.00±0.00	0.00± 0.00
$t\bar{t}$ single tags	0.04±0.00	0.19±0.02	0.08±0.01	0.01± 0.00
$t\bar{t}$ double tags		0.00±0.00	0.00±0.00	0.00±0.00
Total single tags expected	15.54±1.33	5.21±0.42	0.85±0.07	0.24± 0.02
Total double tags expected		0.03±0.01	0.01±0.00	0.00±0.00

Table 6.29: Predicted and observed events with Jetprobability and SLT tags in the $Z+\geq 1$ jet sample. Jets are clustered with cone radius 0.4 and required to have $E_T \geq 15$ GeV and $|\eta| \leq 2$.

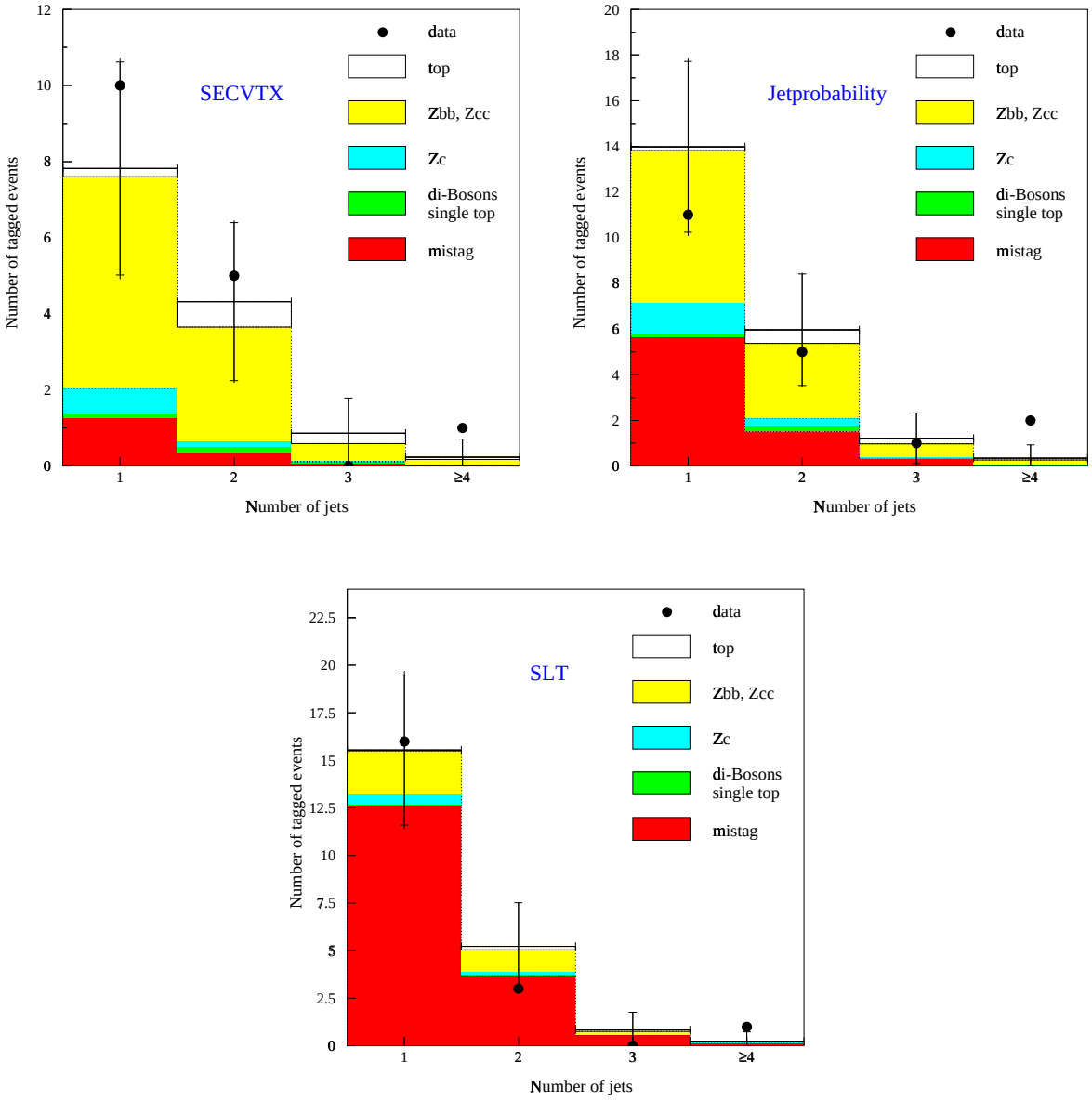

Figure 6.16: Comparison of the observed and predicted number of $Z+ \geq 1$ jet events with SECVTX, Jetprobability and SLT tags, as a function of the jet multiplicity. Error bars represent the overall uncertainty on the expected number of tags and the horizontal ticks on the bars, mark the contribution from the statistical uncertainty.

Chapter 7

Calculation of the acceptance for $t\bar{t}$ events

In this chapter the calculation of the acceptance for $t\bar{t}$ events in the *lepton + jets* channel is presented. The acceptance, $\mathcal{A}_{t\bar{t}}$, is a measure of the efficiency for identifying a $t\bar{t}$ event and is defined as the number of $t\bar{t}$ events, $N_{t\bar{t}}^{obs}$, surviving all the analysis requirements described in the previous chapter, divided by the total number of Monte Carlo generated $t\bar{t}$ events, $N_{t\bar{t}}^{gen}$.

$$\mathcal{A}_{t\bar{t}} = \frac{N_{t\bar{t}}^{obs}}{N_{t\bar{t}}^{gen}}$$

The calculation can be factorized into several terms which are measured either in control samples from data or Monte Carlo simulations. The total acceptance can be written in the following form:

$$\mathcal{A}_{t\bar{t}} = \mathcal{A}_{\ell+jets} \cdot \epsilon_{trigger} \cdot \epsilon_{tag},$$

where

- $\mathcal{A}_{\ell+jets}$ is the fraction of $t\bar{t}$ events within the detector geometric acceptance, passing all the event selection requirements described in sections 3.2.1 and 3.3.1 and have at least 3 jets with $E_T \geq 15$ GeV and $|\eta_D| \leq 2$.

- $\epsilon_{trigger}$ is the fraction of lepton + jet events passing all the geometric and kinematic selections and satisfying one of the Level 2 lepton triggers.

- ϵ_{tag} is the fraction of selected $W+ \geq 3$ jet events satisfying the trigger requirements containing at least one jet flagged as heavy flavor candidate.

The measurement of each of the above components and its associated uncertainties is presented in the following sections. The acceptance calculation is performed assuming top mass of 170 GeV/c^2. The dependence of the acceptance on the top mass is also examined in Monte Carlo samples where the top mass was set at 150 and 190 GeV/c^2. The part of the $t\bar{t}$ acceptance before tagging of heavy flavor jets is common for all three tagging algorithms and also for both parts of run 1 (Run 1A and 1B). The tagging efficiency in $t\bar{t}$ events for the SECVTX, jetprobability and SLT tagging algorithms is presented in separate sections.

7.1 Determination of $\mathcal{A}_{\ell+jets}$

This section describes the computation of the fraction of $t\bar{t}$ events that can be detected in the lepton + jets channel after application of all the lepton selection requirements, including also the Z and dilepton removal.

The calculation is based on simulated $t\bar{t}$ events generated using the PYTHIA [70] Monte Carlo generator and processed with the CDF detector simulation, QFL [77].

The generated sample is processed with the same offline reconstruction code used for the data and the resulting simulated sample is analyzed applying all the selection requirements as for real data.

The lepton+jets kinematic acceptance can be written in the following way, factorizing the contribution of the various selection requirements:

$$\mathcal{A}_{\ell+jets} = \mathcal{A}_{lepton} \cdot \epsilon_{\ell\ id} \cdot \epsilon_{\not\!\!E_T} \cdot \epsilon_{jet} \cdot \epsilon_{topology}$$

where,

- $\mathcal{A}_{lepton} \cdot \epsilon_{\ell\ id}$ is the fraction of $t\bar{t}$ events that have an electron with $E_T \geq 20$ GeV or a muon with $P_T \geq 20$ GeV/c, within the fiducial geometric acceptance of the detector, which also passes the primary lepton selection criteria. By construction this term includes the branching fraction, $Br(t\bar{t} \rightarrow \ell X)$.

- $\epsilon_{\not\!\!E_T}$ is the fraction of the selected lepton events that have $\not\!\!E_T \geq 20$ GeV.

- ϵ_{jet} is the fraction of good lepton events with $\not\!\!E_T \geq 20$ GeV and containing at least 3 jets with $E_T \geq 15$ GeV and $|\eta_D| \leq 2$.

- $\epsilon_{topology}$ is the fraction of $t\bar{t}$ events with a primary lepton surviving the Z and dilepton removal requirements.

The $\not\!\!E_T$ distribution in $t\bar{t}$ events with at least one good lepton of $E_T(P_T) \geq 20$ is shown in figure 7.1(a), for three different top masses. The jet multiplicity distribution for the $t\bar{t}$ events with a good lepton of $P_T \geq 20$ GeV/c and $\not\!\!E_T \geq 20$ is shown in figure 7.1(b) for top masses of 150, 170 and 190 GeV/c^2. Jets are counted for $E_T \geq 15$ GeV and $|\eta_D| \leq 2$.

The results of the acceptance calculation as a function of top mass and jet multiplicity, for each of the contributing terms are summarized in Table 7.1.

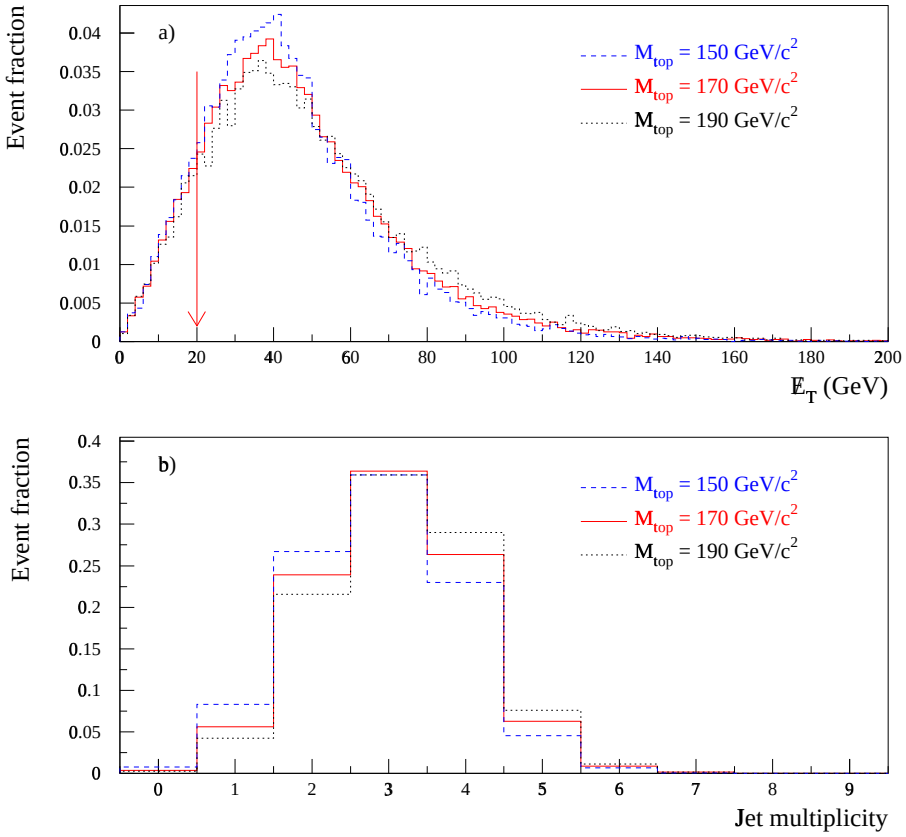

Figure 7.1: $\not\!\!E_T$ distribution for $t\bar{t}$ events which contain a primary lepton of $P_T \geq$ 20 GeV/c a). Jet multiplicity distribution for $t\bar{t}$ events containing a primary lepton of $P_T \geq 20$ GeV/c and $\not\!\!E_T \geq 20$ GeV b). The distributions are shown for top mass of $M_{top} = 150$ GeV/c^2 (dashed), $M_{top} = 170$ GeV/c^2 (solid) and $M_{top} = 190$ GeV/c^2 (dotted).

In determining the raw acceptance there is no requirement imposed on the origin of the lepton. In most of the cases, the selected lepton originates from the decay of one of the two W's in the event while the other W decays into hadron jets. There is also contribution from events in which the primary lepton originates from leptonic tau

decay through the decay chain $W \rightarrow \nu_\tau\tau \rightarrow \ell\nu_\ell\nu_\tau$. Contribution from this process

Selection	$\ell + 1$ jet %	$\ell + 2$ jet %	$\ell + 3$ jet %	$\ell + \geq 4$ jets %	$\mathcal{A}_{t\bar{t}}^{raw}$ %
	$M_{top} = 150$ **GeV/c^2**				
Lepton·ϵ_{id}	1.45±0.04	4.72±0.06	6.40±0.07	5.08±0.07	11.48±0.10
$\not\!\!E_T$	1.30±0.03	4.17±0.06	5.61±0.07	4.42±0.06	10.04± 0.09
Z-removal	1.21±0.03	3.99±0.06	5.52±0.07	4.36±0.06	9.88±0.09
Dilep-removal	0.64±0.02	2.56±0.05	4.45±0.06	3.83±0.06	**8.27±0.08**
	$M_{top} = 170$ **GeV/c^2**				
Lepton·ϵ_{id}	0.99±0.02	4.24±0.05	6.53±0.06	6.12±0.06	12.66±0.08
$\not\!\!E_T$	0.89±0.02	3.79±0.05	5.78±0.06	5.34±0.05	11.12±0.08
Z-removal	0.83±0.02	3.61±0.05	5.63±0.06	5.27±0.05	10.90±0.08
Dilep-removal	0.41±0.02	2.19±0.04	4.41±0.05	4.62±0.05	**9.03±0.07**
	$M_{top} = 190$ **GeV/c^2**				
Lepton·ϵ_{id}	0.69±0.02	3.51±0.06	5.84±0.07	6.17±0.07	12.02±0.10
$\not\!\!E_T$	0.69±0.02	3.51±0.06	5.84±0.07	6.17±0.07	12.02±0.10
Z-removal	0.65±0.02	3.34±0.05	5.71±0.07	6.07±0.07	11.79±0.10
Dilep-removal	0.29±0.02	1.87±0.04	4.44±0.06	5.30±0.07	**9.74± 0.09**

Table 7.1: The fraction (%) of $t\bar{t}$ events passing various event selection criteria, before trigger requirement, as a function of jet multiplicity and top mass. Errors are due to Monte Carlo statistics. The last column shows the $t\bar{t}$ acceptance for events containing a lepton + $\geq$ 3 jets, before any corrections applied to the lepton identification efficiencies.

varies between 6-7% for top masses in the range of 150 to 190 GeV/c^2. Despite the aggressive removal of dilepton events there are still some events contributing to the final $t\bar{t}$ acceptance in the lepton+jets channel. In this case both W's decay leptonically but one of the leptons is either not reconstructed or it is a tau which decays hadronically forming a jet. The remaining dileptons contribute to the ℓ+ jets of $t\bar{t}$ events approximately 4.8% to 6.3% for top masses in the range of 150 and 190 GeV/c^2.

After all selection the final $t\bar{t}$ sample contains approximately equal amount of electrons and muons. The relative contribution of each muon type is 49.2% CMUP, 17.3% CMU-only, 10.9% CMP-only and 22.6% for CMX muons.

7.1.1 Systematic Uncertainties

Systematic uncertainties contributing to the determined acceptance arise mainly from modeling of the Initial and Final State Radiation (*ISR* and *FSR*), the uncertainty in the jet energy scale and the process modeling by different Monte Carlo generators.

The error due to the ISR and FSR modeling is estimated by taking half the difference between the $\mathcal{A}_{\ell+jets}$ calculated with the default Monte Carlo parameters and the one obtained in samples with the ISR and FSR turned off separately. The resulting relative uncertainty is $\pm 5\%$ and $\pm 2\%$ for the modeling of ISR and FSR radiation, respectively [53].

To determine the uncertainty due to jet energy scale, the jet energies in the Monte Carlo were varied within $\pm 10\%$ of the default values. From the changes in the acceptance, a relative uncertainty of $\pm 5\%$ is assigned due to jet energy scale.

The dependence of the acceptance on the modeling of the $t\bar{t}$ process from different Monte Carlo generators was studied in events generated with the HERWIG [68] Monte Carlo. The results were compared to the ones obtained in the default $t\bar{t}$ sample generated with the PYTHIA Monte Carlo generator. A relative systematic uncertainty of $\pm$ 5% is assigned to the calculated acceptance based on the results from the two Monte Carlo samples.

The dependence of the primary vertex fiducial requirement ($Z_{vertex} \leq 60$ cm) and the dependence of the lepton isolation on the instantaneous luminosity also contribute to the systematic uncertainty of the $t\bar{t}$ acceptance calculation. The vertex fiducial

requirement was studied with $W \rightarrow \mu\nu$ events as a function of the luminosity and number of jets. A relative systematic uncertainty of $\sim \pm 2\%$ is assigned on the vertex fiducial requirement, assuming the efficiency of the vertex requirement is flat as a function of the luminosity. Using the $W \rightarrow \mu\nu$ data sample, a relative systematic uncertainty of $\pm 1\%$ is estimated assuming the isolation efficiency is flat a function of the luminosity [53]. Combining the two contributions, a relative $\pm 2\%$ systematic uncertainty is assigned to the acceptance due to luminosity related effects.

Finally, the $t\bar{t}$ acceptance does not depend (¡ 1% uncertainty) on the choice of the parton distribution functions used in the generation of the $t\bar{t}$ samples.

The components contributing to the systematic uncertainty of the kinematic acceptance of $t\bar{t}$ events are summarized in Table 7.2. A total systematic error of $\pm 9\%$ is obtained by the sum in quadrature of all the above mentioned uncertainties.

Source	Uncertainty
Initial state radiation	$\pm$ 2%
Final state radiation	$\pm$ 5%
Monte Carlo dependence	$\pm$ 5%
Jet energy scale	$\pm$ 5%
Luminosity dependence	$\pm$ 2%
Total	$\pm$ 9%

Table 7.2: Components contributing to the systematic uncertainty on the $\mathcal{A}_{t\bar{t}}$.

7.2 Trigger efficiency - $\epsilon_{trigger}$

The trigger requirement used for the selection of the W event sample is a combination of the logical OR of several electron and muon triggers, designed to maintain high

efficiency for $t\bar{t}$ events.

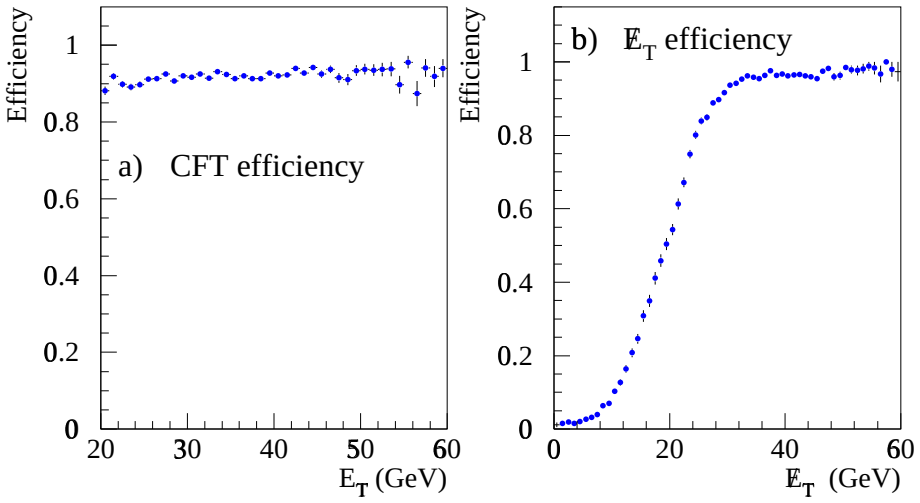

Figure 7.2: Efficiency of the CFT (a) and $\not\!\!E_T$ (b) requirements of the Level 2 electron triggers. The logical OR of the two triggers is measured to be $\sim$ 100% efficient for electrons in the W sample.

As described in section 3.2.1, electrons are accepted based on two Level 2 triggers. One of the triggers requires a CFT track of $P_T \geq 12$ GeV/c pointing to an electromagnetic cluster of $E_T \geq 16$ GeV. Figure 7.2(a) shows the measured efficiency of CFT as a function of the electron E_T. Since the CFT efficiency is $\sim$ 92%, electrons are also accepted with another trigger which does not require the presence of a CFT track but instead is based on the requirement of $\not\!\!E_T \geq 20$ GeV. Figure 7.2(b) shows the efficiency of the $\not\!\!E_T$ trigger as a function of $\not\!\!E_T$. The efficiency of each trigger is measured starting from a control sample satisfying the other trigger and examining the number of events satisfying the trigger in question. This method determines the efficiency of the CFT and $\not\!\!E_T$ requirements. The efficiency of the electromagnetic cluster E_T is measured with CEM triggers of lower E_T. The latter is found to be

100% for electrons with $E_T \geq 20$ GeV. The logical OR of the electron triggers is measured to be 99.8% for electrons in the W sample. Weighting the electron E_T and $\not\!\!E_T$ spectra in $t\bar{t}$ events with the efficiency curves shown in figure 7.2 the OR of the electron triggers is more than 99.9% efficient for top events and for the rest of this analysis is assumed to be 100%.

Muons must satisfy one of the 16 Level 2 triggers used in this analysis (see Table 3.3). The large number of triggers was used to optimize the efficiency because the CMU-only and CMX triggers were prescaled. In order to compensate for the loss in efficiency due to the trigger prescaling, unprescaled triggers requiring the presence of a jet in the event were implemented. These triggers were designed to maintain high efficiency for $t\bar{t}$ events which as shown in figure 7.1 have many jet at final state. Since there was no explicit trigger path for CMP-only muons, two triggers requiring $\not\!\!E_T \geq 35$ GeV are used. In order to estimate the trigger efficiency, a trigger simulation program implementing the efficiency of each trigger was used. The results of the trigger simulation were tested using $W+ \geq 1$ jet HERWIG simulated events and compared to the rate observed in the data. Since the electron trigger efficiency was found to be 100% efficient, the HERWIG sample was normalized to the number of $W(\rightarrow e\nu)+ \geq 1$ jet events in the data after background subtraction. A comparison of the rates of $W(\rightarrow \mu\nu)+ \geq 1$ jet events in the data and simulation for each muon type is shown in Table 7.3.

The efficiency of the muon trigger requirements for $t\bar{t}$ events was measured starting with $t\bar{t}$ events containing a primary muon (see Section 3.3.1) and requiring the muon to satisfy the logical OR of the triggers in the trigger simulation. The muon trigger efficiency was found to be $\epsilon_{\mu\ trig} = 0.854 \pm 0.086$ for a top mass of 170 GeV/c^2. Averaged over primary electrons and muons the trigger efficiency in $t\bar{t}$ events of top

muon-type	Data No. of events	HERWIG No. of events
CMUP	2044±50	1974±43
CMX	1092±36	973±22
CMU-only	521±25	517±12
CMP	112±12	134±3
TOTAL	3769±68	3598±80

Table 7.3: Rates of $W(\rightarrow \mu\nu)+ \geq 1$ jet events in the data and in the HERWIG simulation by muon-type. The data and the simulation are normalized to the same number of $W(\rightarrow e\nu)+ \geq 1$ jet. Errors on data include the muon ID efficiency. Errors on the simulation includes a 2.2% error due to electron statistics and the electron ID efficiency.

mass 170 GeV/c^2, is $\epsilon_{trigger} = 0.92\pm0.02$, where the error is statistical only. A systematic uncertainty of $\pm$ 10% is assigned to the method of the muon trigger efficiency calculation.

7.3 Corrections to some Monte Carlo efficiency

Because the detector simulation does not reproduce exactly all the variables used to select the primary leptons, small correction factor, ϵ_ℓ^{cor}, should be applied to the lepton identification efficiencies, in the case of Monte Carlo events. The final acceptance is obtained by multiplying the previously estimated acceptance with this correction factor. The correction factors are estimated separately for Run 1A and Run 1B.

Corrections for run 1B

The correction factor is determined by comparing the lepton identification efficiencies measured in data using $Z \rightarrow \ell^-\ell^+$ events (see Section 3.5) to the ones measured in $W \rightarrow e(\mu) + \nu + \geq 1$ jets HERWIG Monte Carlo events. In both cases, the efficiencies are calculated after the isolation requirement, $I_\ell^{cal} \leq 0.1$, is applied to the examined

lepton.

The measured lepton identification efficiencies in the Monte Carlo simulation are summarized in Table 7.4 and are also compared to the ones determined in the data. The last column of Table 7.4 indicates the degradation factors to be applied to the Monte Carlo efficiencies for each lepton type. A common correction factor,

Lepton type	ϵ_{id}^{Data}	ϵ_{id}^{MC}	$\epsilon_{\ell\ id}^{cor}$
Electron	0.812±0.01	0.875±0.004	0.927±0.012
CMU-CMP	0.942±0.01	0.974±0.003	0.967±0.011
CMU-only	0.924±0.03	0.971±0.005	0.952±0.031
CMX	0.915±0.02	0.968±0.004	0.945±0.021

Table 7.4: Comparison of lepton identification efficiencies in data and in $W+ \geq 1$ jet HERWIG Monte Carlo simulation. The identification efficiencies in the simulation need to be degraded with the correction factors indicated in the last column.

$\epsilon_{\mu\ id}^{cor} = 0.96 \pm 0.02$, can then be obtained for all muon types by weighting each individual correction factor by the relative acceptance of each muon type. As shown in Table 7.4 the corresponding correction factor for electrons is $\epsilon_{e\ id}^{cor} = 0.927 \pm 0.012$. Averaging over the expected number of electrons and muons, after trigger simulation, the correction factor for any lepton type is determined to be $\epsilon_{\ell\ id-1B} = 0.942 \pm 0.126$.

The error includes $\pm$ 1% uncertainty due to the limited statistics of the Z sample used to determine the lepton identification efficiencies in the data, summed in quadrature with the $\pm$ 9% systematic uncertainty derived in Section 7.1.1 and $\pm$ 10% uncertainty assigned in the calculation of the trigger efficiency in the simulation.

Corrections for run 1A

The correction for run 1A is derived by comparing the rates of $W+ \geq 1$ jet events in run 1A and 1B after normalizing for the corresponding luminosities.

Comparing the observed rate of $W \rightarrow \mu+ \geq 1$ jet events in run 1A to the expected one calculated by scaling the run 1B rate with the ratio of luminosities of the two runs, $\frac{\int \mathcal{L}_{1B}}{\int \mathcal{L}_{1A}} = 4.5 \pm 0.4$, it is found that the muon trigger efficiency in run 1A is lower than the one in run 1B, by 7.8%. The observed inefficiency is attributed mainly to the CMX trigger hardware problems. On the other hand, the rate of $W \rightarrow e+ \geq 1$ jet events is found to scale pretty well with the luminosity (1165 found compared to 1156 expected).

Comparing the total rate of $W+ \geq 1$ jet events of the two runs to the expected one, the run 1B correction factor $\epsilon^{cor}_{\ell\ id}$ obtained before, needs to be degraded by 0.994 to account for the full luminosity of both runs. A conservative uncertainty of 100% is assigned to the calculation of the run 1A muon trigger efficiency.

The final correction factor for scaling Monte Carlo samples to the data is determined to be $\epsilon^{cor}_{\ell\ id\ 1A+1B} = 0.936 \pm 0.126$

7.4 Total $\mathcal{A}_{t\bar{t}}$ before tagging

The $t\bar{t}$ acceptance for $W+ \geq 3$ jet events before heavy flavor tagging and for different top masses is summarized in Table 7.5. The acceptance increases with the top mass, mainly due to the increased fraction of events passing the kinematic requirement for ≥ 3 jets of the selection criteria.

The $t\bar{t}$ acceptance used in this thesis, corresponds to the one obtained for top mass of 170 GeV/c^2 and is equal to $\mathcal{A}_{t\bar{t}\ before\ tagging} = 7.789 \pm 1.105\%$

	$M_{top} = 150\text{GeV/c}^2$	$M_{top} = 170\text{GeV/c}^2$	$M_{top} = 190\text{GeV/c}^2$
$\mathcal{A}_{\ell+jets}$	8.273±0.083%	9.032±0.069%	9.735±0.089%
$\epsilon_{trigger}$	90.95±0.300%	92.14±0.218%	93.21±0.242%
$\epsilon^{cor}_{\ell\ id}$	93.60±12.6%	93.60±12.6%	93.60±12.6%
$\mathcal{A}_{t\bar{t}\ before\ tagging}$	7.043±0.951%	**7.789±1.105%**	8.493±1.146%

Table 7.5: The total $t\bar{t}$ acceptance as a function of top mass before tagging of heavy flavor jets. Errors in the first two raws are due to statistical uncertainty while all the systematic errors are included in the uncertainty of $\epsilon^{cor}_{\ell\ id}$. The errors in the final acceptance include both systematic and statistical uncertainties summed in quadrature.

7.5 Tagging efficiency in $t\bar{t}$ events

The efficiency of each tagging algorithm on $t\bar{t}$ events is measured applying each tagger on $t\bar{t}$ events generated with the PYTHIA Monte Carlo generator and after full detector simulation. Events with $W+ \geq 3$ jets passing all analysis criteria are used as the $t\bar{t}$ signal region. The tagging efficiency is defined as the fraction of $t\bar{t}$ events in the signal region with at least one tagged jet:

$$\epsilon^{event}_{\geq\ 1\ tag} = \frac{N_{\geq\ 1\ tag}}{N_{W+\geq\ 3\ jet}} \tag{7.1}$$

where $N_{W+\geq\ 3\ jet}$ and $N_{\geq\ 1\ tag}$ are the number of $t\bar{t}$ events in the $W+ \geq 3$ jet region before and after tagging respectively. Counting the events with at least one tag, no requirements on the origin of the tagged jet are imposed. Thus a tagged jet can originate either from a b or a c-jet in the event. It could also originate from a lighter quark or gluon in which case it is a mistag. The knowledge of the mistag rate in the Monte Carlo simulation is essential in order to properly account for the total mistag contribution to the $t\bar{t}$ background calculations.

As already discussed in Chapter 4, the detector simulation is $\sim$ 100% efficient

in finding and reconstructing tracks in an event and this strongly affects the tagging efficiency. In order to account for the real tracking efficiency, the tracking degradation algorithm see Section 4.4) is applied to all simulated samples. Tracks are discarded according to the reconstruction efficiency which is parametrized as a function of the density of hits around each track. The higher the density of hits around a track the higher the probability for tracking reconstruction failure. This way, the effect of the tracking degradation is automatically included to the resulting jet tagging efficiency.

Taking into account the probability to find one or two b-tags in a $t\bar{t}$, the event tagging efficiency in equation 7.1 can be expressed in the form:

$$\epsilon_{tag}^{event} = \epsilon_{tag}^{jet\,2} + 2 \cdot \epsilon_{tag}^{jet} \cdot (1 - \epsilon_{tag}^{jet}) \tag{7.2}$$

$$\epsilon_{tag}^{event} = 1 - (1 - \epsilon_{tag}^{jet})^2 \tag{7.3}$$

where ϵ_{tag}^{jet} is the efficiency to tag one of the b-jets in the event. This b-jet tagging efficiency includes also the efficiency of finding a taggable b-jet in the event. Recall that taggable jet is the jet in the SVX fiducial volume containing at least two good SVX tracks. In equation 7.2, events with three or more tags, which is a very small fraction of the tagged events, are counted as double tags. The first term in the right side of equation 7.2 represents the probability to tag two b-jets while the second term is the probability to tag one only b-jet.

As discussed in section 4.5, for the case of the SECVTX and Jetprobability algorithms the b-jet tagging efficiency in the simulation, even after track degradation, is different than the one measured in the data using events containing semileptonic decays of b-quarks. A scale factor, $SF = \frac{\epsilon_{jet}^{data}}{\epsilon_{jet}^{MC}}$, was introduced to account for the difference in the tagging efficiencies measured in the data and in the simulation. In

order to propagate the effect of the scale factor to the event tagging efficiency, the b-jet tagging efficiency determined from equation 7.3, needs to be multiplied by the scale factor. The final expression used to determine the event tagging efficiency for the case of SECVTX and Jetprobability becomes:

$$\epsilon_{tag}^{event} = 1 - (1 - SF \cdot \epsilon_{tag}^{jet})^2 \tag{7.4}$$

7.5.1 SECVTX tagging efficiency

Based on equation 7.1, the efficiency of detecting a $t\bar{t}$ event in the Monte Carlo with at least one jet tagged by SECVTX, before track degradation, is found to be $\epsilon_{\geq 1\ tag}^{event}$ = 0.485±0.004 where the error is statistical only. The track degradation reduces the jet tagging efficiency by 85.5±0.9% while the average event tagging efficiency becomes $\epsilon_{\geq 1\ tag}^{event}$ = 0.429±0.004. The jet tagging efficiency before and after track degradation as a function of the jet E_T is shown in figure 7.3(a). Dividing the two distributions,

Jet type	$W+1$ jet		$W+2$ jets		$W+\geq 3$ jets	
b-jets/event	0.775		1.303		1.753	
c-jets/event	0.051		0.150		0.392	
QCD-jets/event	0.174		0.547		1.523	
Tag type	SECVTX	JPB	SECVTX	JPB	SECVTX	JPB
b-tag/event	0.241	0.283	0.377	0.439	0.518	0.599
c-tag/event	0.001	0.005	0.016	0.028	0.030	0.058
QCD-tag/event	0.001	0.011	0.004	0.011	0.005	0.024

Table 7.6: Average number of b, c and non-heavy flavor jets expected in $t\bar{t}$ events before and after tagging as a function of jet multiplicity.

a scale factor for the effect of the track degradation on the jet tagging efficiency, is

obtained. The scale factor as a function of the jet E_T is shown in figure 7.3(b).

As mentioned before there are no requirements on the origin of the tagged jet. As a result the tagging efficiency includes some contribution from mistags. It is found though that for the SECVTX case the amount of mistags in the simulation is of the order of 0.1% and for the rest of the analysis it is assumed that the Monte Carlo simulation does not include any SECVTX mistags. SECVTX tagged jets in the simulation are treated as resulting only from heavy flavor jets in the event. Table 7.6 shows the jet flavor composition in $t\bar{t}$ events before and after tagging, as a function of the observed jet multiplicity, for top mass of $M_{top} = 170$ GeV/c^2. In section 4.5,

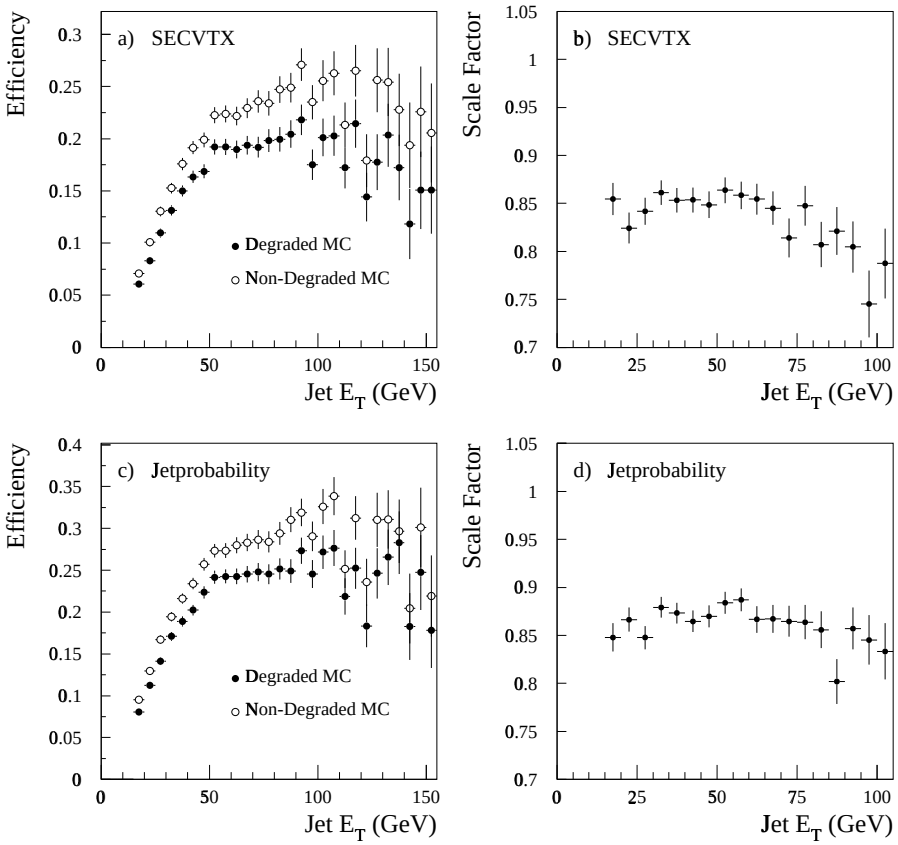

Figure 7.3: Jet tagging efficiency for SECVTX (a) and Jetprobability tags (c) in PYTHIA simulated $t\bar{t}$ events with $W+ \geq 3$ jets, before (open) and after track degradation (solid) and the resulting jet tagging efficiency scale factor for SECVTX (b) and Jetprobability (d) as a function of the tagged jet E_T.

the SECVTX tagging efficiency for b-jets was found to be higher in the data than the corresponding one in the Monte Carlo simulation. It was determined that a scale factor, SF=1.25$\pm$0.13 was needed to adjust the two tagging efficiencies. Since the average tagging efficiency of Run 1A plus Run 1B is a factor of 0.971$\pm$0.009 lower than the run 1B-only efficiency, the data to Monte Carlo scale factor becomes SF = 1.231$\pm$0.13. Making use of equations 7.3 and 7.4 the efficiency for detecting a $t\bar{t}$ event with at least one SECVTX tagged jet is found to be:

$\epsilon^{\mathbf{event}}_{\geq \mathbf{1\ tag}} = \mathbf{0.503 \pm 0.051}$ for a top mass of $M_{top} = 170$ GeV/c^2.

A total systematic uncertainty of 10% is assigned to the tagging efficiency. This uncertainty is mainly due to the 7% systematic uncertainty assigned to the track degradation procedure [56], and smaller contribution due to the dependence of the tagging efficiency on the instantaneous luminosity and radiation damage of the SVX detector. The event tagging efficiency as a function of the top mass is shown in Table 7.7, where a small variation of $\sim 4\%$ is observed for top masses between 150 and 190 GeV/c^2.

7.5.2 Jetprobability tagging efficiency

Similar to the SECVTX case, the $t\bar{t}$ event tagging efficiency using the Jetprobability tagging algorithm is measured to be $\epsilon^{event}_{\geq 1\ tag} = 0.536 \pm 0.004$ when no track degradation is applied to the simulation. After track degradation, the jet tagging efficiency is reduced by $85.85 \pm 0.9\%$ and the average event tagging efficiency is found to be $\epsilon^{event}_{\geq 1\ tag} = 0.488 \pm 0.004$. The jet tagging efficiency before and after degradation as a function of the jet E_T is shown in figure 7.3(c). The resulting scale factor for the jet tagging efficiency as a function of the jet E_T is shown in figure 7.3(d). Figure 7.4 show the positive and negative jetprobability distribution of jets in $W+ \geq 3$ jet events in the data (a) and in PYTHIA $t\bar{t}$ Monte Carlo simulation (b). Fitting the negative jetprobability distribution to a first degree polynomial, $P_0 + P_1 \cdot x$, in the region between $0.1 \div 1.0$ and extrapolating in the region between $0.0 \div 0.05$ an estimate of the mistag rate is obtained. It is found that $5.1 \pm 0.8\%$ of the jets in the $W+ \geq 3$ jet data have a negative jetprobability tag, while the corresponding number in the $t\bar{t}$ simulation is found to be $5.4 \pm 0.3\%$. The simulation then reproduces the mistag rate observed in the data and a 10% conservative systematic uncertainty is assigned to

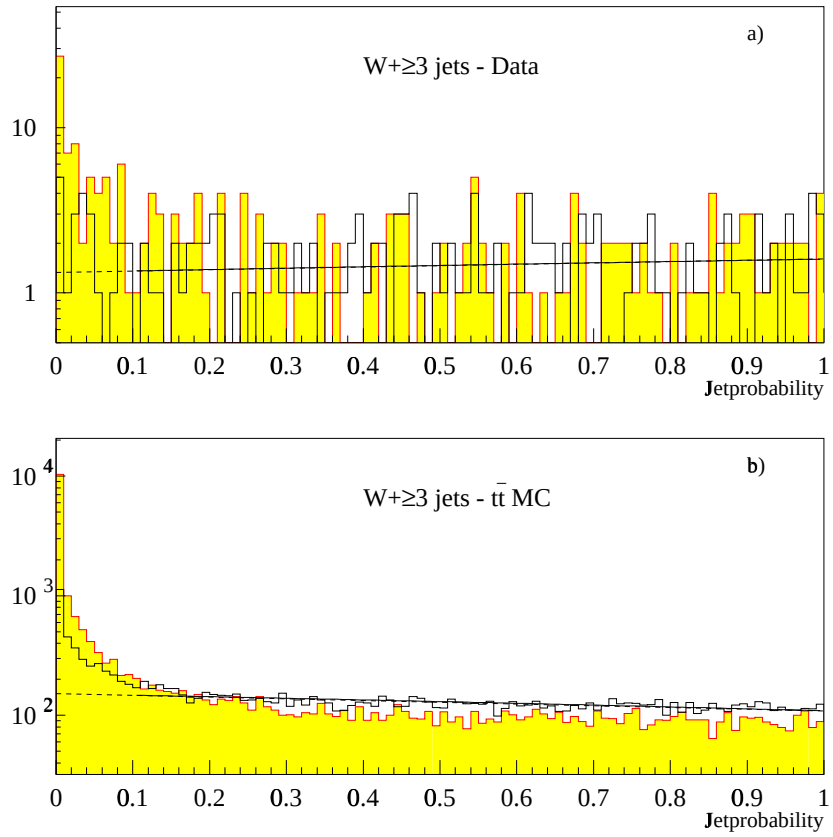

Figure 7.4: Positive (shaded histogram) and negative (open histogram) Jetprobability distributions in $W+ \geq 3$ jet events in data (a) and in PYTHIA $t\bar{t}$ Monte Carlo simulation (b). The lines represent fits to the negative Jetprobability distributions in the range 0.1∼1.0. Extrapolating the fit results in the region 0.0∼0.05 the mistag rate in data and simulations can be obtained.

the predicted mistag rate in the simulation. Repeating the same fitting technique in the $W+1$ and $W+2$ jets sample in data and HERWIG Monte Carlo it is found that the $4.2\pm0.2\%$ of the jets in the HERWIG simulation of $W+1$ and $W+2$ jets have a negative jetprobability tag whilst the same method determines that $3.9\pm0.2\%$ of the $W+1,2$ jets in the data have a negative jetprobability tag. Figure 7.5(a) and (b) show the positive and negative jetprobability distribution of jets in the $W+1,2$ jet events in the data. Figure 7.5(c) show the corresponding distributions in the $W+\geq1$ jet HERWIG simulation. The slope in the jetprobability distribution observed in the

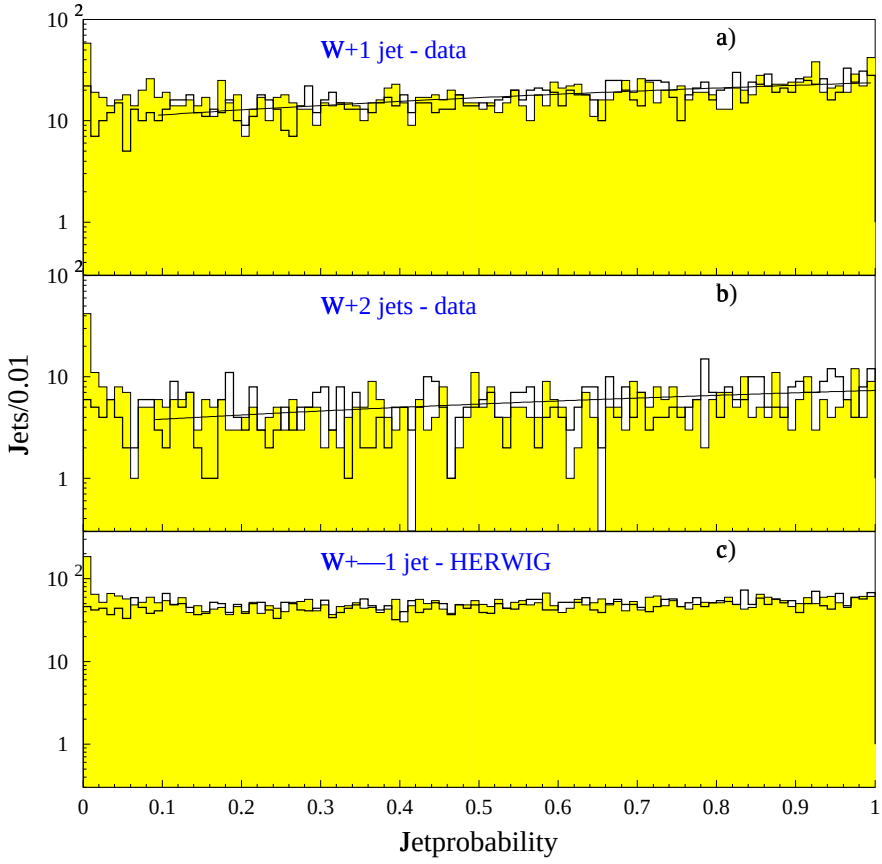

Figure 7.5: Positive (shaded histogram) and negative (open histogram) Jetprobability distributions in $W + 1, 2$ jet events in data (a) and (b) and in HERWIG $W + \geq 1$ jet Monte Carlo simulation (c). The lines represent fits to the negative Jetprobability distributions in the range 0.1∼1.0. Extrapolating the fit results in the region 0.0∼0.05 the mistag rate in data and simulations can be obtained.

$W + 1$ jet case is due to the primary vertex finding algorithm. In events with low track and jet multiplicity, more tracks of a jet are being used in the determination of the primary vertex. This result to a pull of the vertex closer to the jet, affecting the sign and the impact parameter of the tracks in the jet. The effect disappears for events with higher track and jet multiplicity as show in Figure 7.4.

The data to Monte Carlo scale factor for Jetprobability was determined in section 4.5 and was found that the b-tagging efficiency in the simulation needs to be

degraded by SF = 0.95±0.1 to match the one observed in the run 1B low-P_T electron data sample. Taking also into account that the run 1A plus run 1B averaged tagging efficiency is 0.971 of the run 1B-only tagging efficiency, a scale factor of $SF = 0.922 \pm 0.1$ is needed to adjust the differences in the b-tagging efficiency in the data and simulations. Using equations 7.3 and 7.4, the efficiency of observing a $t\bar{t}$ event with at least one jet tagged by Jetprobability is:

$\epsilon^{\mathbf{event}}_{\geq \mathbf{1\ tag}} = \mathbf{0.455 \pm 0.046}$, for top mass of $M_{top} = 170\mathrm{GeV/c^2}$.

The event tagging efficiency as a function of the top mass is shown in Table 7.7, where a small variation of $\sim 4\%$ is observed for top masses between 150 and 190 $\mathrm{GeV/c^2}$.

The sources contributing to the systematic uncertainty of the tagging efficiency are identical to the ones mentioned in the SECVTX case, and a 10% total systematic uncertainty is assigned to the Jetprobability tagging efficiency.

7.5.3 Soft Lepton Tagging efficiency

The soft lepton tagging efficiency in $t\bar{t}$ events in the signal region ($W+ \geq 3$ jets), is measured, as for the other algorithms, from the fraction of events with at least one jet containing a soft lepton tag divided by the total number of pretagged $t\bar{t}$ events. In contrast to the other two taggers, the tagged soft lepton is required to be *real*, meaning that the tagged lepton is required to come from the semileptonic decay of b or a c-quark, or the decay of a τ or W. These tags are considered as the source of true tags in each $W+$ jet multiplicity bin. All other soft lepton tags found in the Monte Carlo simulation are treated as fakes. For example, soft lepton muon tags from decays in flight of pions or fragmentation tracks passing the soft lepton requirements coming from a b-quark decay, appear also in genuine (non-heavy flavor) jets which,

as discussed in the next chapter, is the dominant source of fake SLT tags. Accepting these tags in the Monte Carlo would lead to an over-estimate of the true SLT tag rate. Based on equation 7.1 after track degradation the $t\bar{t}$ SLT event tagging efficiency is found to be:

$\epsilon^{\mathbf{event}}_{\geq \mathbf{1\ tag}} = \mathbf{0.157 \pm 0.016}$, for top mass of $M_{top} = 170\mathrm{GeV/c^2}$.

A variation of $\sim$12% is observed in the soft lepton tagging efficiency as a function of top mass. The gradual increase with increasing top mass is mainly due to the harder P_T spectrum of the emerging b-quarks, resulting in higher P_T soft leptons.

A conservative uncertainty of 10% [58] is assigned to the SLT tagging efficiency. It includes a 5% systematic uncertainty due to the modeling of the E_{had}/E_{em} requirement on the soft electron tags in the simulation and a 5% uncertainty on the knowledge of the branching fractions, $\mathrm{Br}(b/c \rightarrow \ell X)$, of the b and c-quark to leptons. Finally a 5% uncertainty is assigned, in the tracking efficiency for run 1B.

7.5.4 Tagging efficiency summary

Table 7.7 summarizes the $t\bar{t}$ tagging efficiency in $W+ \geq 3$ jet events for the SECVTX, Jetprobability and SLT tagging algorithms as a function of top mass. A small variation of $\sim$ 3% is observed as a function of top mass.

Tagger	$M_{top} = 150\ \mathrm{GeV/c^2}$	$M_{top} = 170\ \mathrm{GeV/c^2}$	$M_{top} = 190\ \mathrm{GeV/c^2}$
SECVTX	0.495±0.050	0.505±0.051	0.515±0.052
JPB	0.455±0.046	0.455±0.046	0.469±0.047
SLT	0.145±0.015	0.157±0.016	0.163±0.016

Table 7.7: Efficiency for tagging $t\bar{t}$ events in the $W+ \geq 3$ jet signal region, for the SECVTX, Jetprobability and SLT tagging algorithms as a function of top mass.

7.6 Summary of the $\mathcal{A}_{t\bar{t}}$ acceptance calculation

The total $t\bar{t}$ detection efficiency in the $W+ \geq 3$ jets events signal region, after the application of the tagging algorithms is summarized in Table 7.8, for top mass of $M_{top} = 170$ GeV/c^2.

Based on the theoretical predictions for the $t\bar{t}$ production cross section [78], [79] and [80], the number of $t\bar{t}$ events expected before and after tagging in $108pb^{-1}$ of data is shown in figure 7.6. The expected number of events is estimated using the calculation of Laenen *et al.* [80] for the $t\bar{t}$ production cross section. The error on the expectation includes an uncertainty of $\sim 16\%$ on the theoretical value of the $t\bar{t}$ cross section.

	$M_{top} = 170$GeV/c^2		
$\mathcal{A}_{\ell+jets}$	9.032±0.069%		
$\epsilon_{trigger}$	92.14±0.218%		
$\epsilon^{cor}_{\ell\ id}$	93.60±12.6%		
$\mathcal{A}_{t\bar{t}\ before\ tagging}$	7.789±1.105%		
Tag	SECVTX	JPB	SLT
ϵ_{tag}	50.5±5.1%	45.5±4.6%	15.7±1.6%
$\mathcal{A}_{t\bar{t}}$	3.93±0.68%	3.54±0.62%	1.22±0.21%

Table 7.8: The total $t\bar{t}$ event detection efficiency in $W+ \geq 3$ jet events, for top mass of 170 GeV/c^2.

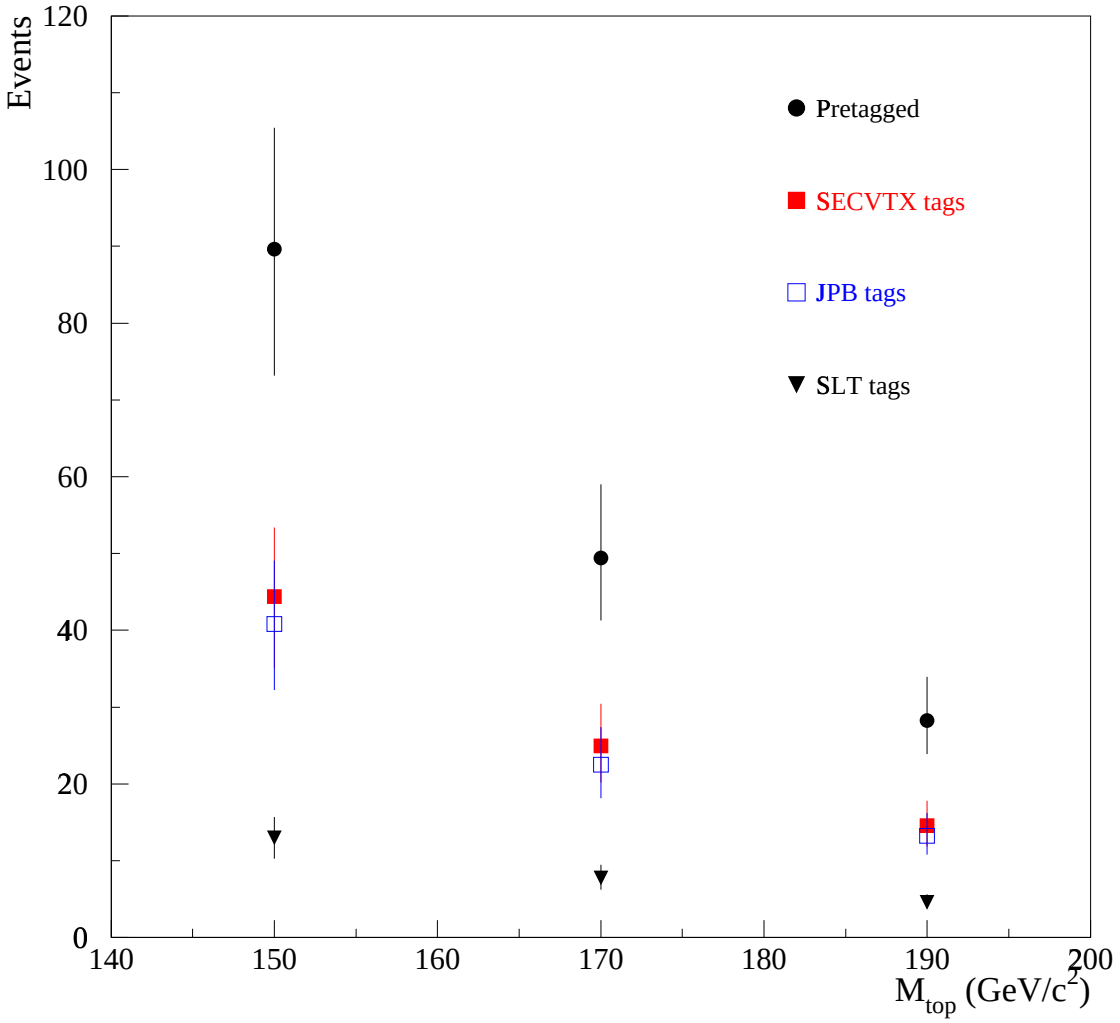

Figure 7.6: The expected number of $t\bar{t}$ events in 108.8 pb^{-1} of data before and after tagging with SECVTX, JetProbability and SLT, as a function of the top mass. The prediction is based on the $t\bar{t}$ cross section calculation of Laenen *et al.*[80]. A systematic error ($\sim$ 16%) based on the cross section calculation of Catani *et al.*[78] and Berger *et al.*[79] is assumed.

Chapter 8

Measurement of the $t\bar{t}$ production cross section

This chapter describes the measurement of the $t\bar{t}$ cross section using events in the $W+ \geq 3$ jet sample with at least one tagged jet. The cross section is calculated separately for each tagging algorithm used to identify heavy flavor candidates. The major sources of heavy flavor jets in the $W+ \geq 1$ jet sample are mistags and $b\bar{b}/c\bar{c}$ QCD heavy flavor production in association with a W and Z, along with a smaller contribution from physics processes like single top, WW, WZ, ZZ, $Z \rightarrow \tau\bar{\tau}$ production and fake W events. The contribution of all the above processes to the tagged $W+ \geq 1$ sample was estimated in Chapter 6. An excess of tagged events over the expectation is observed in the $W+ \geq 3$ jet event sample for all three tagging algorithms. In this chapter, the significance of the excess is calculated for each algorithm. Taking the excess observed in the $W+ \geq 3$ jet events due to events from $t\bar{t}$ production, an iterative procedure is used to determine the $t\bar{t}$ production cross section and the effect of the presence of $t\bar{t}$ events is propagated to the lower jet multiplicity. A

final cross check of the calculations for the heavy flavor contributions to the $W+\geq 1$ jet sample is performed by examining the rate of negative tags observed in the data and comparing it to the expectation from mistags and heavy flavor contributions to the negative tags.

8.1 Significance of the excess of tags in the $W+\geq 3$ jet region

The statistical significance of the excess of events in the $W+\geq 3$ jet region is quantified in terms of the probability, $(\mathcal{P})$, to observe the specific amount of data tags given the amount of expected tags from all examined physics processes.

The probability, $\mathcal{P}$, is estimated from a Poisson distribution with a Gaussian smeared mean M_{exp} according to the following expression:

$$\mathcal{P} = 1 - \sum_{N=0}^{N_{data}} \int_{-\infty}^{+\infty} G(N_{exp}, \sigma_{exp}; M_{exp}) P(M_{exp}; N) dM_{exp}$$

The term $G(N_{exp}, \sigma_{exp}; M_{exp})$ describes the distribution of the number of tagged events from all physics processes according to a Gaussian with mean N_{exp} and width σ_{exp} evaluated at M_{exp}. The term $P(M_{exp}; N)$ is the Poisson distribution of the mean number M_{exp} of expected tagged events evaluated for N actual events. It gives the probability of M_{exp} fluctuating to give N events. Therefore the expression $1 - \sum_{N=0}^{N_{data}} P(M_{exp}; N)$ represents the probability that M_{exp} expected events fluctuate to N_{data} observed tagged events.

The above procedure is repeated for a large number of Monte Carlo experiments, *pseudo-experiments*, in which the expected number of tagged events are drawn from

$W+\geq$3 jets	SECVTX	Jetprobability	SLT
Observed	29	41	25
Expected	9.18$\pm$1.09	15.04$\pm$1.57	14.27$\pm$1.27
$\mathcal{P}$	$1.03 \cdot 10^{-6}$	$3.41 \cdot 10^{-7}$	$9.19 \cdot 10^{-3}$
σ	4.88	5.09	2.60

Table 8.1: Significance of the excess of $W+\geq$3 jet events with a tagged jet. The probability is calculated performing 10M pseudo-experiments and counting the number of experiments with number of tagged events $\geq$ to the observed events.

a Gaussian distribution and the probability the extracted mean fluctuates to the observed number of events is calculated from the Poisson distribution.

Table 8.1 summarizes the observed and expected number of tagged events in the $W+\geq$3 jet sample for each tagging algorithm.

8.2 The measurement of the $t\bar{t}$ cross section

The observed excess of events with at least one b-tagged jet over the expectation in the $W+\geq$3 jet region is interpreted as originating from events due to $t\bar{t}$ production and it is used to determined the $t\bar{t}$ production.

The cross section is estimated from the following expression:

$$\sigma = \frac{N_{tag}^{obs} - N_{tag}^{bkg}}{\mathcal{A}_{t\bar{t}} \int \mathcal{L} dt}$$

where,

- N_{tag}^{obs} is the number of observed tagged events in the $W+\geq$3 jet region.
- N_{tag}^{bkg} is the number of estimated tagged events in the $W+\geq$3 jet region from all heavy flavor production processes other than top, described in Chapter 6.

- $\mathcal{A}_{t\bar{t}}$ is the efficiency for identifying $t\bar{t}$ events with at least one tagged b-jet.
- $\int \mathcal{L}dt$ is the total integrated luminosity of the data sample (108.8 ± 7.86 pb^{-1}).

The background estimates derived in Chapter 6 were calculated assuming no $t\bar{t}$ contribution. When estimating the the contribution from $W+$ heavy flavor production, the relative fractions of $W/Z+$ heavy flavor events are multiplied with the number of real W events in the sample before tagging. However, under the assumption that the excess of events in the $W+ \geq 3$ tagged jet sample is due to $t\bar{t}$ production, the estimated contribution due to $W+$ heavy flavor production is overestimated, since a fraction of the assumed real W events in the sample before tagging is actually due to $t\bar{t}$ events. In order to account for the $t\bar{t}$ contribution to the sample an iterative procedure is used. According to this procedure, the $t\bar{t}$ cross section is estimated from the observed excess of tagged events in the $W+ \geq 3$ jet region. The cross obtained cross section is used to calculate the expected amount of $t\bar{t}$ events before tagging and this contribution is subtracted from the number of real W events. The newly obtained number of W events is used to recalculate the amount of $W+$ heavy flavor events and a new excess of events is calculated resulting to a new $t\bar{t}$ cross section. The procedure is repeated until no change in the cross section is observed. The derived cross section is then used to calculate the $t\bar{t}$ contribution in the $W+1$ and $W+2$ jet multiplicity bins and adjust accordingly the $W+$ heavy flavor contribution.

8.3 The $t\bar{t}$ cross section with SECVTX tags

In the sample of the 252 $W+ \geq 3$ jet events, there are 29 events with at least one jet tagged by the SECVTX algorithm. Using the iteration procedure described in Section 8.2 and the background calculations described in Chapter 6, a background of

8.12 ± 0.99 events is estimated for the 29 tagged events. The resulting $t\bar{t}$ cross section for events tagged by SECVTX is $\sigma_{t\bar{t}} = 4.83 \pm 1.54$ pb and corresponds to acceptances using a top quark mass of 175 GeV/c^2.

The above derived cross section corresponds to the signal and background rates listed in Table 8.2. The errors on the various background estimates reflect the corresponding systematic uncertainties. The errors on the number of $t\bar{t}$ events include both the statistical and systematic uncertainties.

The last three rows of Table 8.2 list the difference between the number of observed tagged events and the expectation including the contribution of $t\bar{t}$ events. Very good agreement is observed in all jet multiplicity bins with a small excess of events in the $W + 2$ jet bin. The agreement in the $W+ \geq 3$ jet bin is expected because this region is saturated with $t\bar{t}$ events in the derivation of the $t\bar{t}$ cross section. The agreement in the $W + 1$ jet bin indicates that the method used for the background estimate reproduces very well the number of tagged events in the data.

Using the composition of the tagged $W+ \geq 1$ jet sample shown in Table 8.2 and unfolding the efficiencies for tagging the events of each process contributing tags to the sample, the composition of the $W+ \geq 1$ jet sample before tagging can be derived. A break-down of the pretagged $W+ \geq 1$ jet sample before tagging is shown in Table 8.3.

A comparison between the observed and predicted event tag yields is shown in Figure 8.1. The predicted rates are based on the contribution of the processes shown in Table 8.2.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
with single tags	66	35	10	11
with double tags		5	6	2
mistags	10.82± 1.08	3.80± 0.38	0.99± 0.10	0.35± 0.04
non-W	8.18± 0.78	1.49± 0.47	0.76± 0.38	0.31± 0.16
WW, WZ, ZZ	0.52± 0.14	1.38± 0.28	0.40± 0.13	0.00± 0.00
$Z \rightarrow \tau\tau$	0.96± 0.30	0.70± 0.25	0.17± 0.12	0.00± 0.00
single top	1.60± 0.41	2.74± 0.63	0.71± 0.16	0.15± 0.04
Wc	16.90± 5.39	3.95± 1.30	0.51± 0.17	0.08± 0.04
$Wc\bar{c}$ single tags	7.90± 2.17	3.59± 0.89	0.80± 0.25	0.16± 0.07
$Wc\bar{c}$ double tags		0.06± 0.04	0.00± 0.00	0.00± 0.00
$Wb\bar{b}$ single tags	17.09± 2.43	8.42± 1.76	1.64± 0.40	0.39± 0.14
$Wb\bar{b}$ double tags		1.51± 0.52	0.31± 0.13	0.06± 0.03
Zc	0.14± 0.04	0.03± 0.01	0.01± 0.00	0.00± 0.00
$Zc\bar{c}$ single tags	0.22± 0.06	0.10± 0.03	0.04± 0.02	0.00± 0.00
$Zc\bar{c}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$Zb\bar{b}$ single tags	0.93± 0.14	0.46± 0.12	0.17± 0.06	0.02± 0.02
$Zb\bar{b}$ double tags		0.08± 0.03	0.03± 0.02	0.00± 0.00
total bkg single tags	65.25± 6.46	26.75± 2.55	6.25± 0.69	1.50± 0.23
total bkg double tags		1.58± 0.52	0.31± 0.13	0.06± 0.03
$t\bar{t}$ single tags	0.48± 0.13	3.44± 0.90	6.94± 1.82	7.85± 2.06
$t\bar{t}$ double tags		0.78± 0.20	2.43± 0.64	3.67± 0.96
$t\bar{t}$ + bkg single tags	65.73± 6.46	30.18± 2.70	13.19± 1.95	9.35± 2.07
$t\bar{t}$ + bkg double tags		2.35± 0.56	2.74± 0.65	3.73± 0.96
excess with single tags	0.27±10.37	4.82± 6.12	-3.19± 3.17	1.65± 2.41
excess with double tags		2.65± 1.63	3.26± 0.86	-1.73± 0.99
excess of tagged events	0.27±10.37	7.46± 6.34	0.07± 3.28	-0.08± 2.60

Table 8.2: Summary of the predicted and observed number of W events with SECVTX tags as a function of the jet multiplicity. The expected background tag rate has been corrected for the $t\bar{t}$ contribution. The $t\bar{t}$ predictions correspond to the measured top production cross section of $\sigma^{t\bar{t}}_{SECVTX} = 4.83 \pm 1.54$ pb assuming a top mass of M_{top}=175 GeV/c^2.

number of events	$W+1$jet	$W+2$jet	$W+3$jet	$W+\geq 4$jet
initial sample	9460	1373	198	54
non-W	560.4± 14.9	71.3± 2.7	12.4± 2.0	5.1± 1.7
WW	31.2± 5.4	31.1± 5.4	5.2± 1.0	0.8± 0.2
WZ	4.4± 0.9	4.8± 1.0	0.9± 0.2	0.1± 0.0
ZZ	0.3± 0.1	0.4± 0.1	0.1± 0.0	0.0± 0.0
$Z \rightarrow \tau\tau$	35.2± 2.2	13.1± 1.2	1.6± 0.4	0.3± 0.2
$Z+$ jets	234.8± 14.5	38.5± 5.9	7.9± 2.4	0.7± 0.7
single top	15.1± 2.3	9.1± 2.0	2.0± 0.4	0.3± 0.1
Wc	413.4±124.0	87.0± 26.1	11.2± 3.4	1.9± 0.7
$Wb\bar{b}$	69.1± 9.5	29.7± 5.1	5.7± 1.1	1.5± 0.5
$Wc\bar{c}$	173.2± 46.3	62.0± 13.6	11.4± 2.6	2.2± 0.9
$W+$ non$-$h.f.	7956.4±133.6	1029.4± 31.1	121.4± 7.6	19.2± 6.4
$t\bar{t}$	1.7± 0.5	9.7± 2.7	19.7± 5.5	22.2± 6.2

Table 8.3: Breakdown of the composition of the $W+\geq 1$ jet sample before tagging as determined by the composition of the SECVTX tag rate.

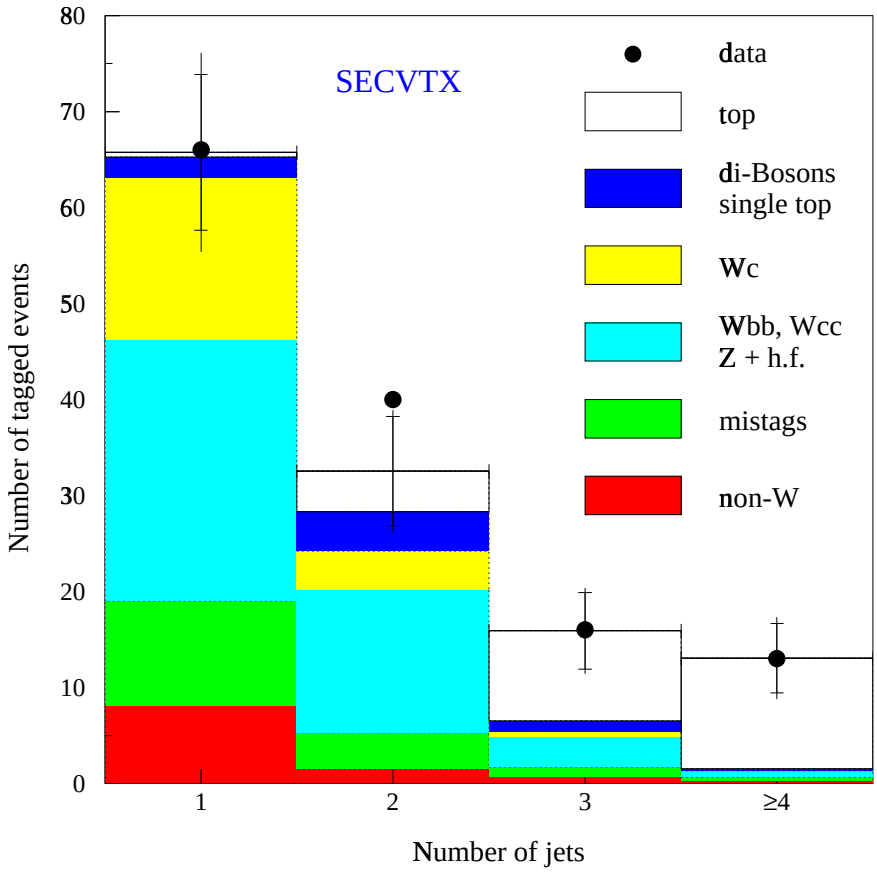

Figure 8.1: Comparison between the predicted and observed number of W events with SECVTX tags, as a function of the jet multiplicity. The vertical bars represent the overall uncertainty in the predictions and the horizontal ticks on the bars the contribution of the statistical uncertainty alone.

8.4 The $t\bar{t}$ cross section using Jetprobability tags

In the $W+ \geq 3$ jet signal region there are 41 events with at least one jet tagged by the jetprobability algorithm. Following the same technique as in the measurement of the $t\bar{t}$ cross section with SECVTX tags, it is found that 11.33 ± 1.36 background events are expected in the $W+ \geq 3$ jet signal region. The observed excess of events yields a $t\bar{t}$ cross section of $\sigma_{t\bar{t}} = 7.33 \pm 2.10$ pb, using acceptances for a top quark mass of 175 GeV/c^2.

Table 8.4 summarizes the observed and expected tag yields from all contributing processes corresponding to the above determined $t\bar{t}$ cross section. The errors on the various background estimates reflect the corresponding systematic uncertainties. The errors on the number of $t\bar{t}$ events include both the statistical and systematic uncertainties.

The composition of the $W+ \geq 1$ jet sample as a function of jet multiplicity is shown in Table 8.5. Figure 8.2 compares the observed rate of tags in the various jet bins with the corresponding expectations. Excellent agreement is found in the $W + 1$ jet bin and $W+ \geq 3$ jet bin while a small excess of events is observed in the $W + 2$ jet multiplicity bin.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
with single tags	125	62	21	12
with double tags		6	5	3
mistags	41.85± 4.24	12.78± 1.33	2.22± 0.27	0.24± 0.19
non-W	12.55± 0.95	2.53± 0.61	0.57± 0.33	0.24± 0.14
WW, WZ, ZZ	1.15± 0.26	2.39± 0.43	0.74± 0.19	0.05± 0.04
$Z \rightarrow \tau\tau$	2.35± 0.47	1.13± 0.32	0.17± 0.12	0.09± 0.09
single top	1.54± 0.38	2.53± 0.60	0.68± 0.16	0.12± 0.03
Wc	35.05±10.65	9.04± 2.84	1.70± 0.60	0.16± 0.11
$Wc\bar{c}$ single tags	17.14± 4.63	7.24± 1.72	1.73± 0.46	0.20± 0.14
$Wc\bar{c}$ double tags		0.47± 0.20	0.05± 0.03	0.01± 0.01
$Wb\bar{b}$ single tags	16.48± 2.33	7.53± 1.53	1.50± 0.35	0.21± 0.14
$Wb\bar{b}$ double tags		1.42± 0.48	0.25± 0.10	0.03± 0.02
Zc	0.29± 0.09	0.08± 0.03	0.03± 0.01	0.00± 0.00
$Zc\bar{c}$ single tags	0.47± 0.13	0.20± 0.06	0.09± 0.04	0.01± 0.01
$Zc\bar{c}$ double tags		0.01± 0.01	0.00± 0.00	0.00± 0.00
$Zb\bar{b}$ single tags	0.90± 0.14	0.42± 0.10	0.16± 0.06	0.02± 0.02
$Zb\bar{b}$ double tags		0.07± 0.03	0.03± 0.01	0.00± 0.00
total bkg single tags	129.77±12.64	45.96± 4.02	9.64± 0.98	1.35± 0.34
total bkg double tags		1.89± 0.52	0.31± 0.10	0.04± 0.02
$t\bar{t}$ single tags	0.64± 0.14	4.69± 1.04	9.90± 2.18	11.04± 2.44
$t\bar{t}$ double tags		1.13± 0.25	3.49± 0.77	5.23± 1.15
$t\bar{t}$ + bkg single tags	130.41±12.64	50.65± 4.16	19.54± 2.39	12.39± 2.46
$t\bar{t}$ + bkg double tags		3.03± 0.57	3.80± 0.78	5.26± 1.15
excess with single tags	-5.41±17.03	11.35± 8.24	1.46± 3.92	-0.39± 2.72
excess with double tags		2.97± 1.83	1.20± 0.95	-2.26± 1.17
excess of tagged events	-5.41±17.03	14.32± 8.44	2.67± 4.03	-2.66± 2.96

Table 8.4: Summary of the predicted and observed number of W events with Jet-probability tags as a function of the jet multiplicity. The expected background tag rate has been corrected for the $t\bar{t}$ contribution. The $t\bar{t}$ predictions correspond to the measured top production cross section of $\sigma^{t\bar{t}}_{JPB} = 7.33 \pm 2.10$ pb assuming a top mass of M_{top}=175 GeV/c^2.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
non-W	560.4± 14.9	71.3± 2.7	12.4± 2.0	5.1± 1.7
WW	31.2± 5.4	31.1± 5.4	5.2± 1.0	0.8± 0.2
WZ	4.4± 0.9	4.8± 1.0	0.9± 0.2	0.1± 0.0
ZZ	0.3± 0.1	0.4± 0.1	0.1± 0.0	0.0± 0.0
$Z \rightarrow \tau\tau$	35.2± 2.2	13.1± 1.2	1.6± 0.4	0.3± 0.2
$Z+$ jets	234.8± 14.5	38.5± 5.9	7.9± 2.4	0.7± 0.7
single top	15.1± 2.3	9.1± 2.0	2.0± 0.4	0.3± 0.1
Wc	413.3±124.0	86.6± 26.0	10.5± 3.2	1.0± 0.7
$Wb\bar{b}$	69.1± 9.5	29.6± 5.1	5.3± 1.0	0.8± 0.5
$Wc\bar{c}$	173.2± 46.3	61.7± 13.6	10.6± 2.5	1.2± 0.8
$W+$ non$-$h.f.	7955.6±133.6	1025.1± 31.1	113.2± 8.8	10.3± 8.3
$t\bar{t}$	2.5± 0.6	14.8± 3.6	29.9± 7.2	33.6± 8.2

Table 8.5: Breakdown of the composition of the $W+\geq 1$ jet sample before tagging as determined from the composition of the Jetprobability tag rate.

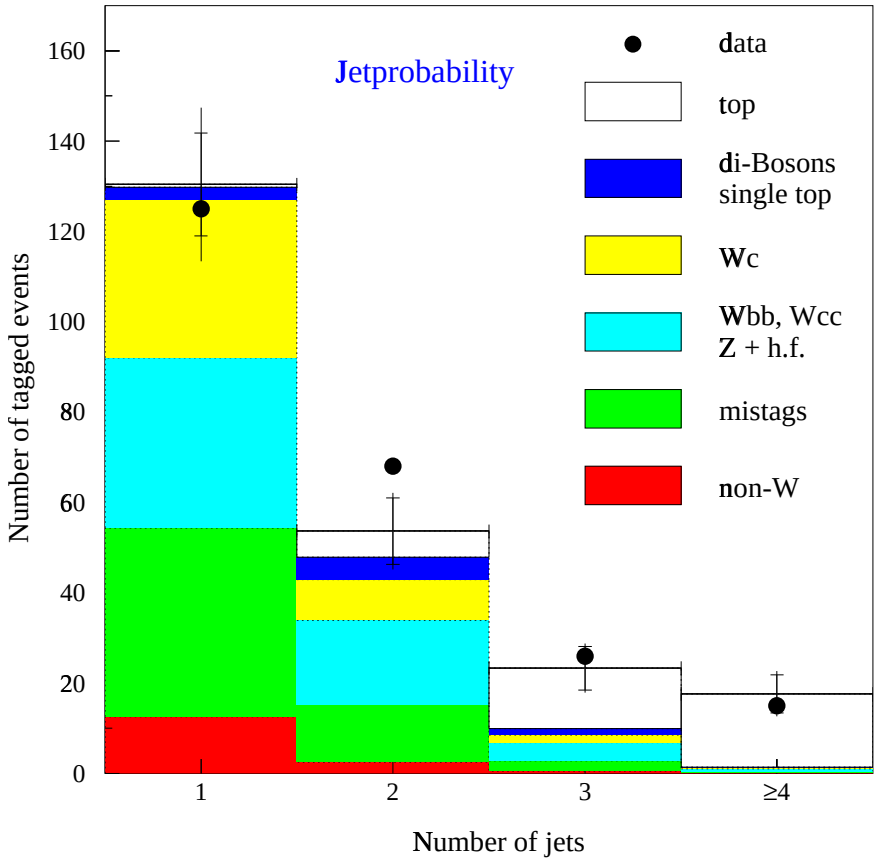

Figure 8.2: Comparison between the predicted and observed number of W events with Jetprobability tags, as a function of the jet multiplicity. The vertical bars represent the overall uncertainty in the predictions and the horizontal ticks on the bars the contribution of the statistical uncertainty alone.

8.5 The $t\bar{t}$ cross section with SLT tags

In the $W+ \geq 3$ jet signal region there are 25 events with at least one jet tagged by the SLT algorithm. As a reminder, soft lepton tags are considered only when are found within a cone of 0.4 around the axis of a jet with $E_T \geq 15$ GeV. Using the iteration procedure described in Section 8.2, the estimated background contribution in the $W+ \geq 3$ jet signal region corresponds to 13.27 ± 1.23 events. The observed excess of events yields a $t\bar{t}$ cross section for events tagged by the SLT algorithm of $\sigma_{t\bar{t}} = 8.37 \pm 3.98$ pb, using acceptances for a top quark mass of 175 GeV/c^2.

Table 8.6 summarizes the observed and expected tag yields from all contributing processes corresponding to the above determined $t\bar{t}$ cross section. The errors on the various background estimates reflect the corresponding systematic uncertainties. The errors on the number of $t\bar{t}$ events include both the statistical and systematic uncertainties.

The composition of the $W+ \geq 1$ jet sample as a function of jet multiplicity is shown in Table 8.7. Figure 8.3 compares the observed rate of tags in the various jet bins with the corresponding expectations. Excellent agreement is found in the $W+1$ jet bin and $W+ \geq 3$ jet bin while a small excess of events is observed in the $W+2$ jet multiplicity bin.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
with single tags	146	56	17	8
with double tags		0	0	0
mistags	102.00±10.20	31.10± 3.11	7.34± 0.73	3.01± 0.30
non-W	8.97± 0.84	2.09± 0.56	0.38± 0.27	0.16± 0.11
WW, WZ, ZZ	0.26± 0.10	0.94± 0.23	0.10± 0.05	0.00± 0.00
$Z \rightarrow \tau\tau$	0.54± 0.23	0.09± 0.09	0.09± 0.09	0.00± 0.00
single top	0.44± 0.12	0.75± 0.17	0.21± 0.05	0.05± 0.02
Wc	13.39± 4.35	3.79± 1.33	0.70± 0.31	0.04± 0.09
$Wc\bar{c}$ single tags	6.55± 1.93	2.57± 0.65	0.66± 0.22	0.05± 0.10
$Wc\bar{c}$ double tags		0.02± 0.02	0.00± 0.00	0.00± 0.00
$Wb\bar{b}$ single tags	5.26± 0.95	3.04± 0.70	0.33± 0.12	0.03± 0.06
$Wb\bar{b}$ double tags		0.05± 0.03	0.01± 0.01	0.00± 0.00
Zc	0.11± 0.04	0.03± 0.01	0.01± 0.01	0.00± 0.00
$Zc\bar{c}$ single tags	0.18± 0.05	0.07± 0.02	0.04± 0.02	0.00± 0.00
$Zc\bar{c}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$Zb\bar{b}$ single tags	0.29± 0.05	0.17± 0.05	0.04± 0.02	0.00± 0.01
$Zb\bar{b}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
total bkg single tags	137.98±11.33	44.64± 3.57	9.91± 0.88	3.35± 0.35
total bkg double tags		0.07± 0.04	0.01± 0.01	0.00± 0.00
$t\bar{t}$ single tags	0.24± 0.10	2.35± 1.03	4.87± 2.14	6.28± 2.76
$t\bar{t}$ double tags		0.07± 0.03	0.22± 0.10	0.36± 0.16
$t\bar{t}$ + bkg single tags	138.22±11.33	47.00± 3.71	14.78± 2.31	9.63± 2.78
$t\bar{t}$ + bkg double tags		0.14± 0.05	0.24± 0.10	0.36± 0.16
excess with single tags	7.78±16.33	9.00± 7.80	2.22± 3.91	-1.63± 3.33
excess with double tags		-0.14± 0.37	-0.24± 0.15	-0.36± 0.16
excess of tagged events	7.78±16.33	8.87± 7.81	1.99± 3.91	-1.99± 3.33

Table 8.6: Summary of the predicted and observed number of W events with SLT tags as a function of the jet multiplicity. The expected background tag rate has been corrected for the $t\bar{t}$ contribution. The $t\bar{t}$ predictions correspond to the measured top production cross section of $\sigma_{SLT}^{t\bar{t}} = 8.37 \pm 3.98$ pb assuming a top mass of M_{top}=175 GeV/c^2.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
non-W	560.4± 14.9	71.3± 2.7	12.4± 2.0	5.1± 1.7
WW	31.2± 5.4	31.1± 5.4	5.2± 1.0	0.8± 0.2
WZ	4.4± 0.9	4.8± 1.0	0.9± 0.2	0.1± 0.0
ZZ	0.3± 0.1	0.4± 0.1	0.1± 0.0	0.0± 0.0
$Z \rightarrow \tau\tau$	35.2± 2.2	13.1± 1.2	1.6± 0.4	0.3± 0.2
$Z+$ jets	234.8± 14.5	38.5± 5.9	7.9± 2.4	0.7± 0.7
single top	15.1± 2.3	9.1± 2.0	2.0± 0.4	0.3± 0.1
Wc	413.3±124.0	86.5± 26.0	10.2± 3.3	0.6± 1.3
$Wb\bar{b}$	69.1± 9.5	29.5± 5.1	5.2± 1.1	0.5± 1.0
$Wc\bar{c}$	173.2± 46.2	61.6± 13.5	10.3± 2.6	0.8± 1.6
$W+$ non−h.f.	7955.3±133.6	1023.3± 31.8	109.7± 16.2	6.6± 17.5
$t\bar{t}$	2.9± 1.3	16.9± 7.6	34.1± 15.4	38.4± 17.3

Table 8.7: Breakdown of the composition of the $W+\geq 1$ jet sample before tagging as determined from the composition of the SLT tag rate.

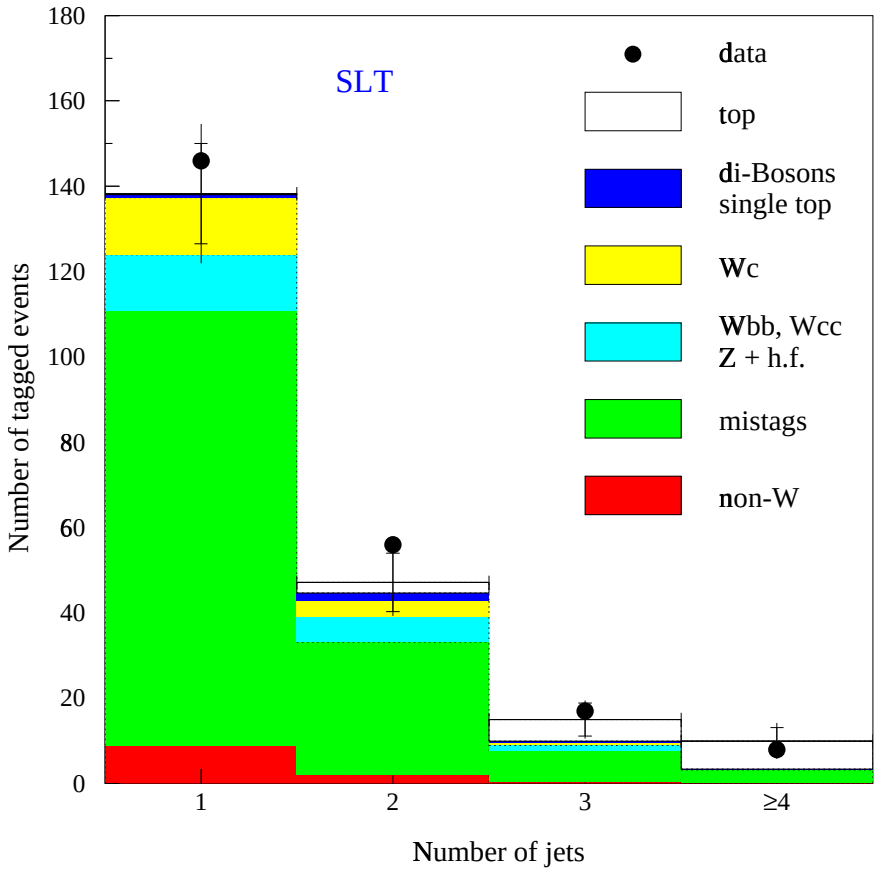

Figure 8.3: Comparison between the predicted and observed number of W events with SLT tags, as a function of the jet multiplicity. The vertical bars represent the overall uncertainty in the predictions and the horizontal ticks on the bars the contribution of the statistical uncertainty alone.

8.6 Negative tag rates in the $W+ \geq 1$ jet sample

As discussed in Chapter 5, heavy flavor decays can contribute also to the negative tagging rate. Special effort was put to find a reliable and easy way of estimating this contribution and subtracting it from the measured negative tagging rate. As presented in length in Chapter 5, the derived pure mistag rates plus the estimated heavy flavor contribution to the negative tags reproduces very well the observed negative tagging rates in all QCD samples examined.

In this Section, the same comparison is performed in the $W+ \geq 1$ jet sample in every jet multiplicity bin. This test offers an additional check of the mistag rate predictions and it also offers an alternative test of the method used to estimate the background contribution to the $t\bar{t}$ signal.

In order to perform this test, the negative tagging rate of each contributing process is calculated from the corresponding Monte Carlo simulations. For $Wb\bar{b}$ and $Wc\bar{c}$, the negative tagging efficiency is calculated in the same manner it was calculated for the positive tagging efficiency. The estimated negative tagging efficiencies are then multiplied with the appropriate event fraction and the number of $W+$ jet events in each jet bin as determined after the adjustement for the $t\bar{t}$ contribution. The contribution of all the other processes including $t\bar{t}$ production is scaled according to the $W+ \geq 1$ jet sample composition before tagging as determined by the $t\bar{t}$ cross section calculation. The sample composition listed in Tables 8.3 and 8.5 is used for SECVTX and jetprobability, respectively.

The results of this test are presented in the next two sections.

8.6.1 Negative SECVTX tags

As discussed before, the contribution of the various processes to the negative tags is estimated based on the composition of the $W+ \geq 1$ jet sample listed in Table 8.3. Table 8.8 and Figure 8.4 show the comparison of the observed and predicted SECVTX negative tagging rates as a function of jet multiplicity. The agreement is not excellent but within errors.

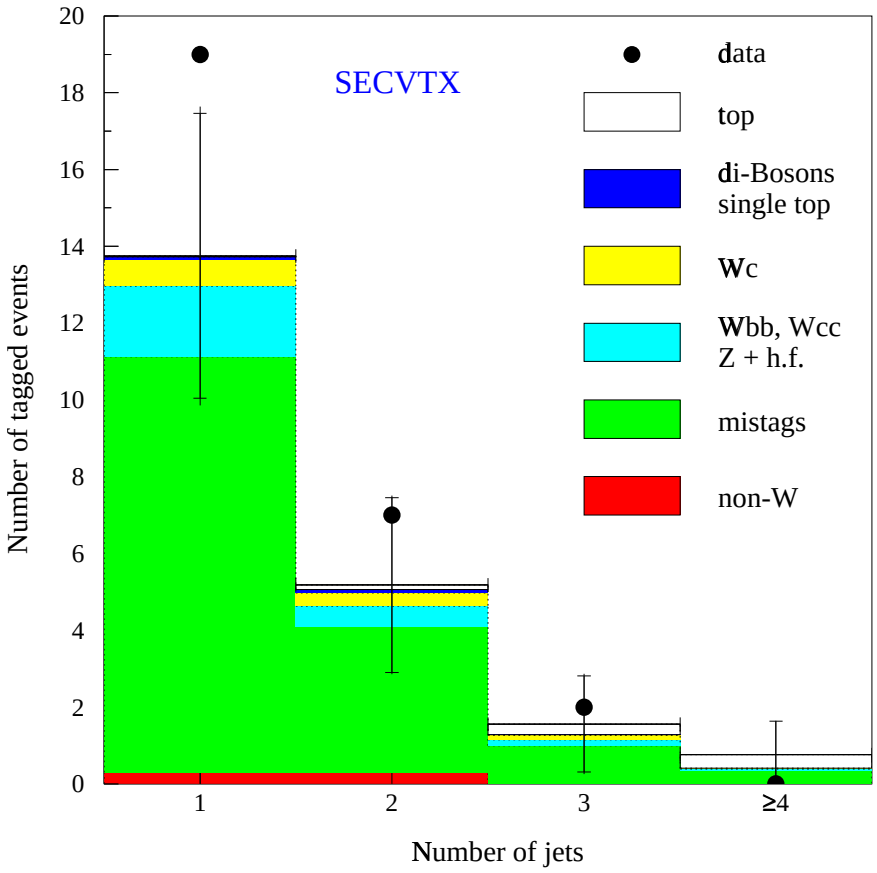

Figure 8.4: Comparison between the predicted and observed number of W events with negative SECVTX tags, as a function of the jet multiplicity. The vertical bars represent the overall uncertainty in the predictions and the horizontal ticks on the bars the contribution of the statistical uncertainty alone.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
with single tags	19	7	2	0
with double tags		0	0	0
mistags	10.82± 1.08	3.80± 0.38	0.99± 0.10	0.35± 0.04
non-W	0.30± 0.15	0.30± 0.21	0.00± 0.35	0.00± 0.14
WW, WZ, ZZ	0.00± 0.00	0.04± 0.04	0.00± 0.00	0.00± 0.00
$Z \to \tau\tau$	0.00± 0.00	0.00± 0.00	0.00± 0.00	0.00± 0.00
single top	0.07± 0.02	0.05± 0.02	0.01± 0.01	0.00± 0.00
Wc	0.69± 0.32	0.34± 0.15	0.12± 0.09	0.02± 0.02
$Wc\bar{c}$ single tags	0.34± 0.15	0.18± 0.07	0.07± 0.05	0.01± 0.01
$Wc\bar{c}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$Wb\bar{b}$ single tags	1.42± 0.26	0.32± 0.09	0.08± 0.05	0.02± 0.02
$Wb\bar{b}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
Zc	0.01± 0.00	0.00± 0.00	0.00± 0.00	0.00± 0.00
$Zc\bar{c}$ single tags	0.01± 0.00	0.00± 0.00	0.00± 0.00	0.00± 0.00
$Zc\bar{c}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$Zb\bar{b}$ single tags	0.08± 0.01	0.02± 0.01	0.01± 0.01	0.00± 0.00
$Zb\bar{b}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
total bkg single tags	13.73± 1.18	5.07± 0.48	1.29± 0.38	0.41± 0.15
total bkg double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$t\bar{t}$ single tags	0.02± 0.01	0.12± 0.03	0.28± 0.07	0.35± 0.09
$t\bar{t}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$t\bar{t}$ + bkg single tags	13.75± 1.18	5.18± 0.48	1.58± 0.39	0.76± 0.17
$t\bar{t}$ + bkg double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
excess with single tags	5.25± 3.89	1.82± 2.33	0.42± 1.20	-0.76± 0.67
excess with double tags		0.00± 0.02	0.00± 0.02	0.00± 0.01
excess of tagged events	5.25± 3.89	1.82± 2.33	0.42± 1.20	-0.76± 0.67

Table 8.8: Summary of the predicted and observed number of W events with negative SECVTX tags as a function of the jet multiplicity. The contribution of each individual process is derived according to the composition of the $W+\geq 1$ jet sample in Table 8.3.

8.6.2 Negative Jetprobability tags

The contribution of the various processes to the negative jetprobability tags is estimated based on the composition of the $W+ \geq 1$ jet sample as determined by the $t\bar{t}$ cross section measurement and listed in Table 8.5. Table 8.9 and Figure 8.5 show the comparison of the observed and predicted negative jetprobability tagging rates as a function of jet multiplicity. Very good agreement is observed in all jet multiplicity bins.

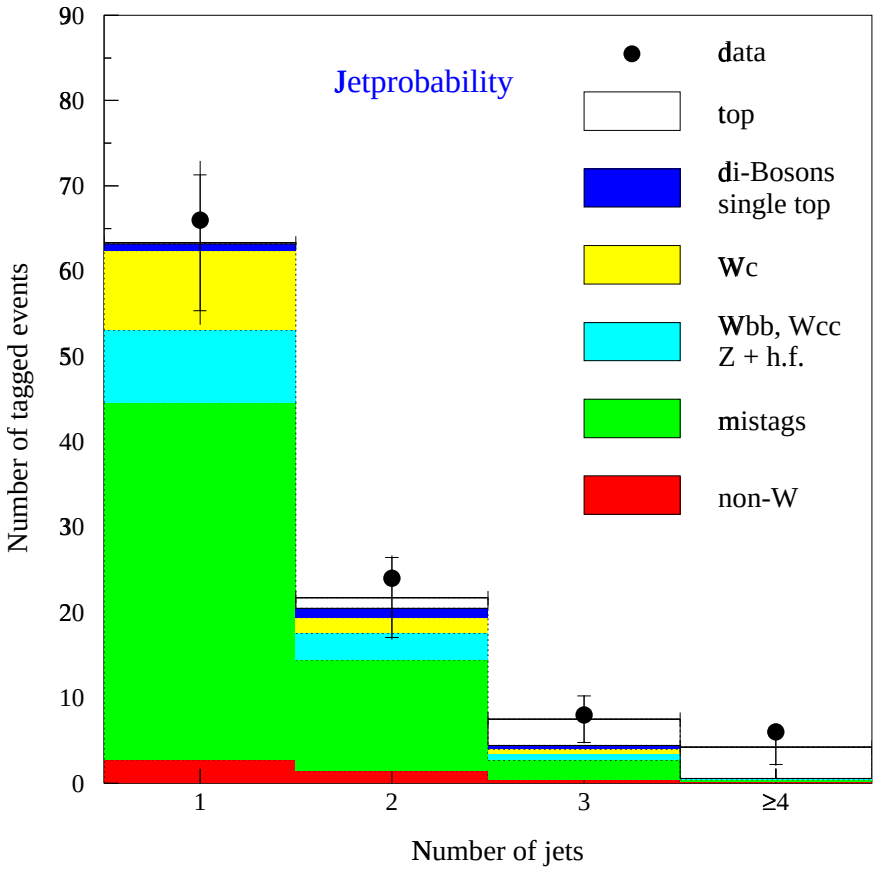

Figure 8.5: Comparison between the predicted and observed number of W events with negative Jetprobability tags, as a function of the jet multiplicity. The vertical bars represent the overall uncertainty in the predictions and the horizontal ticks on the bars the contribution of the statistical uncertainty alone.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
initial sample	9460	1373	198	54
with single tags	66	23	8	5
with double tags		1	0	1
mistags	41.85± 4.24	12.78± 1.33	2.22± 0.27	0.24± 0.19
non-W	2.75± 0.45	1.43± 0.43	0.39± 0.19	0.16± 0.08
WW, WZ, ZZ	0.50± 0.15	0.74± 0.19	0.36± 0.13	0.02± 0.01
$Z \to \tau\tau$	0.44± 0.20	0.52± 0.21	0.09± 0.09	0.00± 0.00
single top	0.26± 0.05	0.39± 0.09	0.11± 0.03	0.02± 0.01
Wc	9.31± 2.91	1.82± 0.67	0.47± 0.22	0.04± 0.03
$Wc\bar{c}$ single tags	4.55± 1.27	0.71± 0.24	0.26± 0.12	0.03± 0.02
$Wc\bar{c}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$Wb\bar{b}$ single tags	3.14± 0.50	1.77± 0.36	0.40± 0.12	0.06± 0.04
$Wb\bar{b}$ double tags		0.01± 0.01	0.01± 0.01	0.00± 0.00
Zc	0.08± 0.02	0.02± 0.01	0.01± 0.00	0.00± 0.00
$Zc\bar{c}$ single tags	0.12± 0.04	0.02± 0.01	0.01± 0.01	0.00± 0.00
$Zc\bar{c}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
$Zb\bar{b}$ single tags	0.17± 0.03	0.10± 0.02	0.04± 0.02	0.01± 0.01
$Zb\bar{b}$ double tags		0.00± 0.00	0.00± 0.00	0.00± 0.00
total bkg single tags	63.18± 5.35	20.50± 1.65	4.42± 0.46	0.59± 0.21
total bkg double tags		0.01± 0.01	0.01± 0.01	0.00± 0.00
$t\bar{t}$ single tags	0.15± 0.03	1.19± 0.26	2.95± 0.65	3.42± 0.75
$t\bar{t}$ double tags		0.05± 0.01	0.12± 0.03	0.24± 0.05
$t\bar{t}$ + bkg single tags	63.32± 5.35	21.69± 1.67	7.38± 0.79	4.00± 0.78
$t\bar{t}$ + bkg double tags		0.06± 0.02	0.14± 0.03	0.25± 0.05
excess with single tags	2.68± 9.59	1.31± 4.95	0.62± 2.25	1.00± 1.09
excess with double tags		0.94± 0.24	-0.14± 0.11	0.75± 0.07
excess of tagged events	2.68± 9.59	2.25± 4.95	0.49± 2.25	1.75± 1.09

Table 8.9: Summary of the predicted and observed number of W events with negative Jetprobability tags as a function of the jet multiplicity. The contribution of each individual process is derived according to the composition of the $W+\geq 1$ jet sample in Table 8.5.

8.7 Correlations between the taggers

It is interesting to study the expected rate of events tagged by more than one algorithms and see the correlations between the tagging algorithms. The fraction of events tagged by more than one algorithms and also some specific cases of jets with overlapping tags are given. The study is performed in terms of event fractions because the fractions can be applied directly to the composition of the $W+\geq 1$ jet sample determined in the cross section measurement.

8.7.1 Overlaps between Jetprobability and SECVTX

Using the same Monte Carlo samples used for the measurement of the $t\bar{t}$ cross section, the fraction of events with a jet tagged by jetprobability and SECVTX was determined. The results are shown in Table 8.10. The overlap between events tagged by

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
mistags	0.01	0.04	0.30	0.65
non-W	0.50	0.50	0.50	0.50
WW	0.32	0.32	0.32	0.32
WZ	0.67	0.67	0.67	0.67
ZZ	0.66	0.66	0.64	0.64
$Z\rightarrow\tau\tau$	0.58	0.58	0.58	0.58
single top	0.86	0.86	0.86	0.86
$t\bar{t}$	0.88	0.92	0.93	0.94
1b-jet	0.79	0.79	0.79	0.79
2b-jets		0.92	1.00	1.00
1c-jet	0.37	0.37	0.37	0.37
2c-jets		0.41	0.60	0.60

Table 8.10: Fraction of $W+\geq 1$ jet events with a Jetprobability that include also a SECVTX tagged jet.

both SECVTX and jetprobability is very high indicating the high degree of correlation between the two tagging algorithms. It is also evident that while the two tagging algorithms overlap almost 90for events with c-jets.

8.7.2 Overlaps between SLT and SECVTX

Repeating the same procedure as before, the fraction of events with a SLT tag that is confirmed by SECVTX is measured. The results are shown in Table 8.11.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
non-W	0.20	0.20	0.20	0.20
WW	0.10	0.10	0.10	0.10
WZ	0.30	0.30	0.30	0.30
ZZ	0.41	0.41	0.41	0.41
$Z \rightarrow \tau\tau$	0.00	0.00	0.00	0.00
single top	0.42	0.42	0.42	0.42
$t\bar{t}$	0.28	0.49	0.50	0.55
1b-jet	0.30	0.30	0.30	0.30
2b-jets		0.39	0.64	0.64
1c-jet	0.05	0.05	0.05	0.05
2c-jets		0.16	0.17	0.17

Table 8.11: Fraction of $W+\geq 1$ jet events with a SLT that include also a jet tagged by SECVTX.

For the SLT tags, it is important to calculate properly the amount of mistags estimated in Table 8.7. To do this, the total amount of mistags needs to be redistributed among the different processes according to their rate before tagging. For each process, the probability that a fake SLT tag is confirmed by SECVTX is measured with the corresponding Monte Carlo. Table 8.12 summarizes the results.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
W/Z+non h.f.	0.001	0.003	0.005	0.117
WW	0.03	0.03	0.03	0.03
WZ	0.10	0.10	0.10	0.10
ZZ	0.00	0.00	0.00	0.00
$Z \rightarrow \tau\tau$	0.00	0.00	0.00	0.00
single top	0.34	0.34	0.34	0.34
$t\bar{t}$	0.40	0.40	0.49	0.56
$1b$-jet	0.41	0.41	0.41	0.41
$2b$-jets		0.58	0.98	0.98
$1c$-jet	0.08	0.08	0.08	0.08
$2c$-jets		0.18	0.14	0.14

Table 8.12: Probability of a fake SLT tag in $W+\geq 1$ jet events to be confirmed by SECVTX.

8.7.3 Overlaps between SLT and SECVTX on the same jet

At this Section, the overlap between events with an SLT and SECVTX tag on the same jet is examined. This study can be used to study whether the jets tagged by SECVTX contain the correct amount of semileptonic decays. The results of this study are discussed in the following chapter. Here the fraction of events with a jet tagged by both the SLT and SECVTX are calculated for each contributing process. Tables 8.13 and Z8.14 show the fraction of events with a multi tagged jet and the probability for a fake SLT tag to be confirmed by SECVTX.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
non-W	0.08±0.03	0.07±0.07	0.00±0.00	0.00±0.00
WW	0.00±0.00	0.12±0.08	0.00±0.00	0.00±0.00
WZ	0.00±0.00	0.26±0.12	0.00±0.00	0.00±0.00
ZZ	0.30±0.30	0.27±0.19	0.61±0.43	0.00±0.00
$Z \rightarrow \tau\tau$	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
single top	0.29±0.13	0.34±0.09	0.32±0.17	0.52±0.30
$t\bar{t}$	0.28±0.09	0.36±0.05	0.36±0.04	0.38±0.05
1b-jet	0.29±0.05	0.31±0.09	0.30±0.30	0.00±0.00
2b-jets		0.38±0.07	0.38±0.18	0.00±0.00
1c-jet	0.05±0.02	0.04±0.03	0.18±0.13	0.00±0.00
2c-jets		0.13±0.05	0.13±0.09	0.00±0.00

Table 8.13: Fraction of $W+ \geq 1$ jet events with a jet tagged by both SLT and SECVTX.

number of events	$W+1$ jet	$W+2$ jet	$W+3$ jet	$W+\geq 4$ jet
WW	0.05±0.05	0.00±0.00	0.00±0.00	0.00±0.00
WZ	0.00±0.00	0.12±0.08	0.00±0.00	0.00±0.00
ZZ	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
$Z \rightarrow \tau\tau$	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
single top	0.06±0.04	0.20±0.09	0.40±0.20	0.20±0.20
$t\bar{t}$	0.52±0.16	0.29±0.05	0.26±0.07	0.26±0.04
1b-jet	0.41±0.11	0.16±0.12	0.00±0.00	0.00±0.00
2b-jets		0.48±0.13	0.46±0.27	0.24±0.24
1c-jet	0.08±0.03	0.04±0.03	0.08±0.08	0.00±0.00
2c-jets		0.09±0.05	0.10±0.10	0.00±0.00

Table 8.14: Fraction of $W+ \geq 1$ jet events with a fake SLT tagged jet confirmed by SECVTX.

Chapter 9

Summary of the results and Conclusions

The $t\bar{t}$ cross section is measured in the $W+ \geq 3$ jet sample by using three tagging techniques. The following results are obtained:

- $\sigma_{t\bar{t}} = 4.83 \pm 1.54$ pb using SECVTX tags.
- $\sigma_{t\bar{t}} = 7.33 \pm 2.10$ pb using Jetprobability tags.
- $\sigma_{t\bar{t}} = 8.37 \pm 3.98$ pb using SLT tags.

Figure 9.1 compares these results with the theoretical predictions and previous measurements of the $t\bar{t}$ cross section.

All measured cross sections in Figure 9.1 are in agreement with the theoretical calculation $\sigma_{t\bar{t}} = 4.7 \div 5.5$ pb for a top quark mass of $M_{top} = 175$ GeV/c^2 [78, 79, 80]. As shown in Figure 9.1, the $t\bar{t}$ cross section measured using SECVTX tags and the cross section measured by D0, using strict Kinematical selection criteria tuned on a top of mass 175 GeV/c^2, are in good agreement with the theory. On the other hand,

values of the $t\bar{t}$ cross section measured with Jetprobability or SLT (including the D0 value) are systematically higher. The errors associated to each cross section are large, but are also highly correlated (like uncertainties on the luminosity, acceptance, lepton identification efficiency).

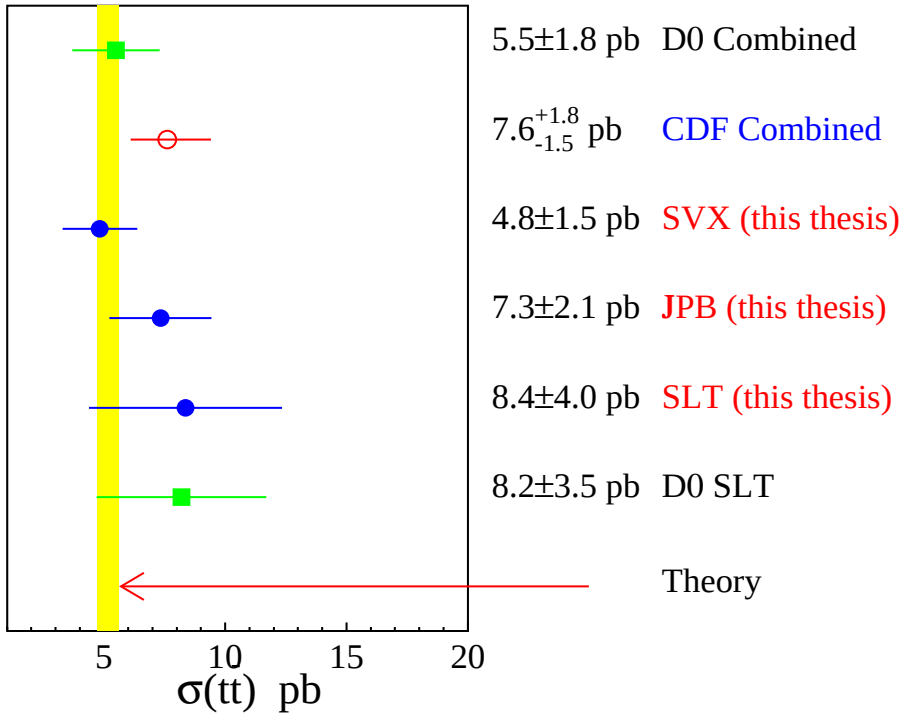

Figure 9.1: Comparison between measured and predicted $t\bar{t}$ production cross sections. The CDF value for the $t\bar{t}$ cross section is from Reference [23] and includes all $t\bar{t}$ decay channels. The vertical bar represents the central value and uncertainty of the theoretical calculation [78, 79, 80].

In comparing the SECVTX and jetprobability results, the only error not in common is the uncertainty in the tagging efficiency of each algorithm, which is of the order of 10%. These two cross sections differ by 2.5 pb with an error of 0.8.

In comparing the SLT and SECVTX results, the errors not in common are statistics and again the systematic uncertainty on the tagging efficiency. The observed discrepancy between the two cross sections is 3.5 pb with an error of 1.6 pb.

Compared correctly, these differences are statistically significant. Now, as shown in Section 6.7 and Table 6.16, the ratio of tagging efficiency for c-quark with respect to b-quark for the SLT and Jetprobability taggers is a factor two higher than for SECVTX.

The standard procedure for calculating the $t\bar{t}$ cross section assumes that the excess of tags in the data with respect to the predicted background is all due to b-quarks and $t\bar{t}$ production.

If part of the excess was due to c-quarks, then it would happen exactly what it is seen in the data.

It is interesting to study how the excess of events with Jetprobability and SLT tags is distributed with respect to events tagged by SECVTX. For this study, the W sample is divided in events with or without SECVTX tags. It is also assumed that the composition of these two samples is as determined by the measurement of the $t\bar{t}$ cross section with SECVTX (see Table 8.2).

Under this assumption, events with SECVTX tags are an almost pure b-sample. Events without SECVTX tags contain:

- Direct production of W+jets without heavy flavor.
- Most of the events due to $Wc\bar{c}$ and Wc production, since SECVTX is very inefficient for tagging c-quark jets.
- Events due to $Wb\bar{b}$ production when the b-jets are not taggable (SECVTX is more efficient for tagging b-jets than Jetprobability and SLT).

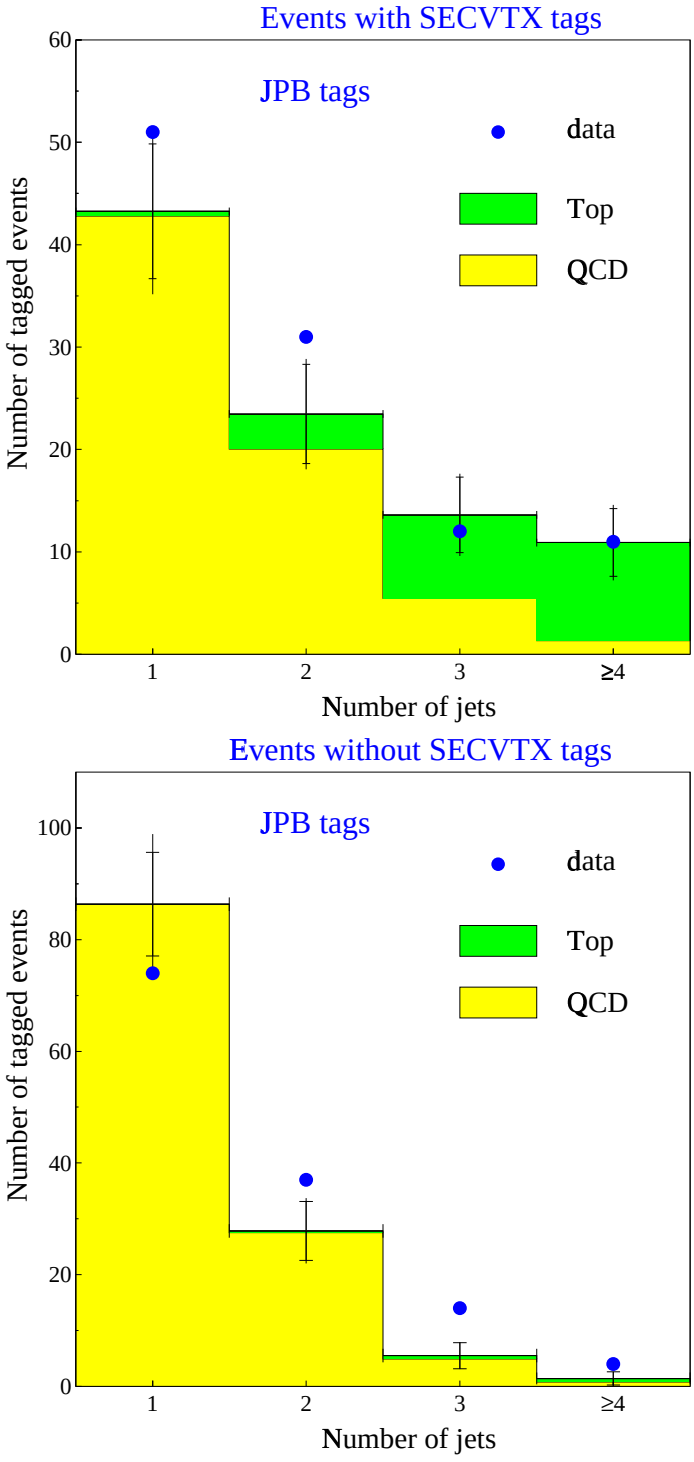

Figure 9.2: Comparison of the observed and predicted rates of Jetprobability tags in events with and without SECVTX tags. The predictions are based on the W sample composition listed in Table 8.2, with $t\bar{t}$ cross section of 4.83 pb.

Rates of observed and expected jetprobability tags in the two samples are shown in Figure 9.2.

In the $W+ \geq 2$ jet events tagged by SECVTX, the observed number of jetprobability tags agree with the expectations. There is an excess of 6.02 ± 7.67 events which, by the way, is the same excess measured by SECVTX (see Table 8.2).

In the remaining $W+ \geq 2$ jet events, there is an excess of Jetprobability tags of 20.2 ± 6.1 events. This excess is as large as the total number of $t\bar{t}$ events tagged by SECVTX (21.1 events). This excess can be explained either by a wrong calculation of the $Wc\bar{c}$ and Wc contributions or by a $W + c$-jets production process unaccounted by the Standard Model. On one hand, the good agreement between the observed and predicted rates of jetprobability tags in $W+1$ jet events (which are dominated by Wc and $Wc\bar{c}$ production) seems to exclude a mistake in the predictions. On the other hand, since the tagging efficiency of jetprobability for c-jets is a factor of two smaller than the tagging efficiency of SECVTX for b-jets, it implies that the new $W + c$-jets production mechanism will have a cross section a factor two larger than $t\bar{t}$ production. This is quite surprising.

For the SLT, it is useful to compare observed tags to expectations only in the sample tagged by SECVTX since the sample without SECVTX tags has too much background. The rates of observed and predicted SLT tags are shown in Figure 9.3. In the $W+ \geq 2$ jet sample tagged by SECVTX, there is an excess of 7.3 ± 3.3 events which has to be compared with the expectation of 5.3 top events. So, in the SLT case, the larger measured cross section does not relate to a new production of $W + c$ jet events as implied by the jetprobability results. The excess of SLT tags is not statistically very significant. An explanation can only come from semileptonic decays with branching ratios higher than in the Standard Model.

These anomalies of the data with respect to the Standard Model prediction are very interesting but, as such, will require a lot more work.

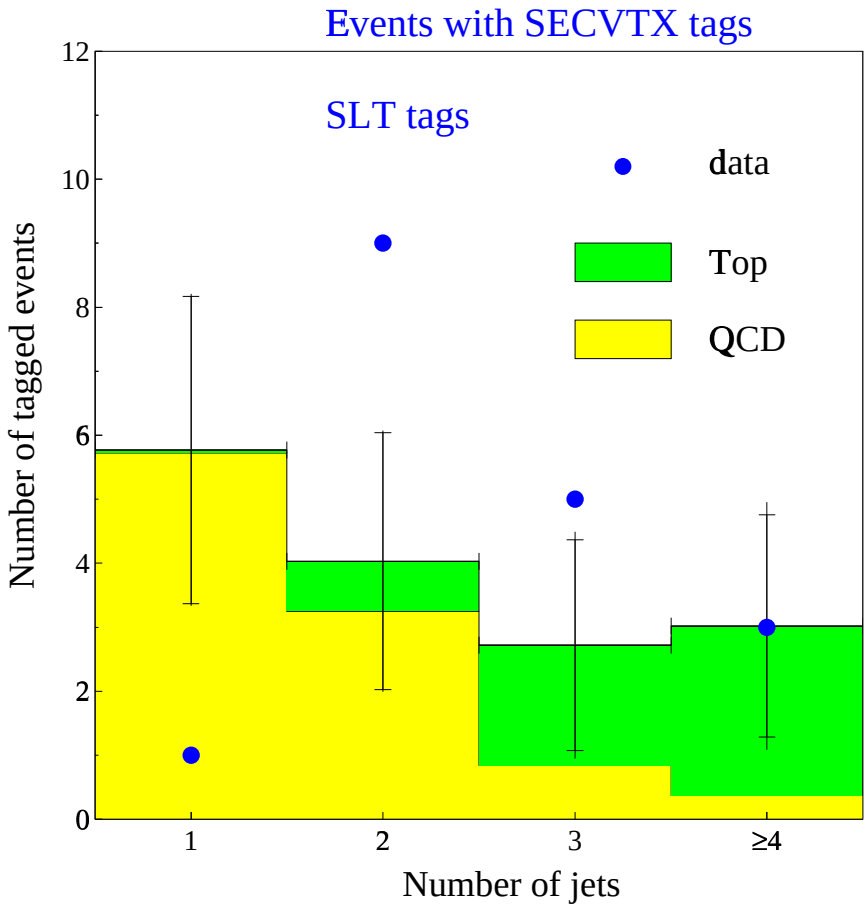

Figure 9.3: Comparison of the observed and predicted rates of SLT tags in events with SECVTX tags. The predictions are based on the W sample composition listed in Table 8.2, with $t\bar{t}$ cross section of 4.83 pb.

It's not over 'til the fat lady sings.

Appendix A

Probability matrices for SECVTX and Jetprobability tags

The positive and negative SECVTX and Jetprobability tagging probability matrices are parametrized as a function of the jet corrected E_T and track multiplicity. The matrices are constructed using the JET_20, JET_50, JET_70, JET_100 and $\sum E_T$_300 QCD samples. Well measured jets away from detector cracks or calorimeter regions with lower response are used and the jet corrected E_T is required to be above hardware trigger threshold of each sample. These jets consist the class of leading jets and they cover the E_T spectrum above 30 GeV. For the E_T region below 30 GeV, jets in events with at least two leading jets from the JET_20 trigger are used.

Each element of the probability matrix is calculated as the ratio of the number of tagged jets with corrected E_T and track multiplicity corresponding to the specific matrix element, to the total number of jets that belong to this matrix bin.

The derived negative tagging probability matrices include contribution from real

heavy flavor jets which is unwanted in the calculation of the mistag rates of the different processes. In order to obtain the true mistag rate of non-heavy flavor jets, the above derived negative tagging matrices are adjusted according to the following expression:

$$N_{mstg} = N_{neg}^{tag} - (N_{pos}^{tag} - N_{neg}^{tag}) \cdot R(E_T)$$

where N_{mstg} is the mistag rate, N_{neg}^{mstg} is the negative tagging probability, N_{pos}^{tag} is the positive tagging probability and $R(E_T)$ is the ratio of negative to positive tags yields due to quark and gluon jets as determined from Monte Carlo simulations. For the matrix elements corresponding to jet corrected E_T $\geq$30 GeV, the quark R parametrization is used while for matrix elements corresponding to jet corrected $E_T \leq 30$ GeV, the gluon R parametrization is used. The resulting matrices represent the true mistag probability matrices.

Both positive and mistag probability matrices for the SECVTX and Jetprobability tagging algorithms are shown in Tables A.1 and A.2. The errors for each matrix element are calculated assuming binomial statistics.

no. of	Jet E_t (GeV/c)									
SVX tracks	5 – 20	20 – 35	35 – 50	50 – 65	65 – 80	80 – 100	100 – 120	120 – 150	150 – 180	180 –∞
Positive SECVTX										
2	0.00000	0.00449	0.00436	0.00466	0.00658	0.00493	0.00459	0.00320	0.00450	0.00517
3	0.00000	0.00990	0.01178	0.01216	0.01151	0.01228	0.00972	0.00962	0.00911	0.01008
4	0.00000	0.01699	0.01734	0.01865	0.02065	0.01642	0.01552	0.01697	0.01216	0.01372
5	0.00000	0.02558	0.02664	0.02495	0.03108	0.02410	0.02285	0.02356	0.02054	0.02132
6-7	0.00000	0.02745	0.02906	0.04156	0.03628	0.03727	0.03423	0.03399	0.02939	0.03355
8-9	0.00000	0.04024	0.04073	0.03802	0.05109	0.04054	0.04665	0.04884	0.04560	0.04245
10-13	0.00000	0.05634	0.03846	0.04577	0.05354	0.05507	0.06290	0.07224	0.06805	0.05863
≥ 14	0.00000	0.00000	0.00000	0.00000	0.10345	0.08537	0.09478	0.07466	0.10064	0.12553
errors										
2	0.00000	0.00075	0.00063	0.00140	0.00134	0.00093	0.00094	0.00078	0.00120	0.00163
3	0.00000	0.00105	0.00093	0.00199	0.00156	0.00127	0.00119	0.00115	0.00147	0.00193
4	0.00000	0.00144	0.00116	0.00231	0.00191	0.00134	0.00136	0.00140	0.00156	0.00208
5	0.00000	0.00213	0.00159	0.00277	0.00237	0.00163	0.00163	0.00161	0.00195	0.00254
6-7	0.00000	0.00231	0.00160	0.00301	0.00207	0.00158	0.00155	0.00149	0.00182	0.00249
8-9	0.00000	0.00575	0.00353	0.00446	0.00347	0.00221	0.00232	0.00224	0.00281	0.00355
10-13	0.00000	0.01580	0.00700	0.00778	0.00516	0.00346	0.00342	0.00327	0.00403	0.00477
≥ 14	0.00000	0.90000	0.16364	0.08571	0.02828	0.01543	0.01288	0.00884	0.01202	0.01528
Negative SECVTX										
2	0.00000	0.00027	0.00109	0.00079	0.00252	0.00100	0.00234	0.00354	0.00091	0.00541
3	0.00000	0.00114	0.00156	0.00366	0.00294	0.00352	0.00332	0.00406	0.00592	0.00497
4	0.00000	0.00162	0.00163	0.00326	0.00363	0.00431	0.00475	0.00465	0.00597	0.00576
5	0.00000	0.00236	0.00280	0.00441	0.00445	0.00509	0.00523	0.00721	0.00931	0.00910
6-7	0.00000	0.00303	0.00366	0.00550	0.00528	0.00801	0.00900	0.01055	0.00880	0.01213
8-9	0.00000	0.00185	0.00373	0.01061	0.00959	0.01156	0.01417	0.01396	0.01570	0.02002
10-13	0.00000	0.00338	0.00835	0.00644	0.01787	0.01769	0.02219	0.02214	0.02940	0.03295
≥ 14	0.00000	0.00000	0.08182	0.08857	0.03172	0.03153	0.01454	0.04678	0.06071	0.05053
errors										
2	0.00000	0.00022	0.00033	0.00060	0.00087	0.00043	0.00069	0.00084	0.00056	0.00171
3	0.00000	0.00042	0.00036	0.00114	0.00082	0.00070	0.00072	0.00077	0.00122	0.00139
4	0.00000	0.00053	0.00038	0.00101	0.00084	0.00071	0.00078	0.00076	0.00112	0.00139
5	0.00000	0.00077	0.00055	0.00122	0.00095	0.00078	0.00081	0.00093	0.00135	0.00171
6-7	0.00000	0.00092	0.00061	0.00115	0.00084	0.00077	0.00083	0.00086	0.00103	0.00155
8-9	0.00000	0.00148	0.00114	0.00247	0.00160	0.00123	0.00134	0.00125	0.00172	0.00253
10-13	0.00000	0.00468	0.00349	0.00309	0.00316	0.00206	0.00214	0.00191	0.00277	0.00372
≥ 14	0.00000	0.90000	0.08668	0.06406	0.01694	0.00994	0.00543	0.00731	0.00978	0.01035

Table A.1: Probability matrix for SECVTX positive tags and mistags for leading jets in the QCD sample, as a function of the number of SVX tracks and the corrected jet E_T.

no. of	Jet E_t (GeV/c)									
SVX tracks	5 – 20	20 – 35	35 – 50	50 – 65	65 – 80	80 – 100	100 – 120	120 – 150	150 – 180	180 –∞
Positive Jetprobability										
2	0.06859	0.07388	0.06765	0.07313	0.07667	0.07834	0.07619	0.07976	0.07736	0.08860
3	0.06691	0.09647	0.08496	0.09673	0.09243	0.08546	0.08695	0.08666	0.08568	0.09188
4	0.09639	0.12033	0.11462	0.11000	0.11686	0.09967	0.10440	0.10466	0.10718	0.10375
5	0.33333	0.22143	0.13953	0.15517	0.16404	0.13341	0.12506	0.11433	0.11470	0.10969
6-7	0.00000	0.32000	0.26772	0.18519	0.13690	0.14621	0.16320	0.14234	0.14219	0.13591
8-9	0.00000	1.00000	0.00000	0.20000	0.13333	0.12281	0.06579	0.18354	0.17391	0.17241
10-13	0.00000	0.00000	0.00000	1.00000	1.00000	0.00000	0.33333	0.37500	0.25000	0.00000
≥ 14	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
errors										
2	0.00455	0.00253	0.00197	0.00366	0.00282	0.00217	0.00219	0.00215	0.00278	0.00387
3	0.01077	0.00490	0.00333	0.00564	0.00397	0.00284	0.00288	0.00276	0.00358	0.00466
4	0.03239	0.01112	0.00668	0.00943	0.00638	0.00442	0.00441	0.00414	0.00536	0.00678
5	0.19245	0.03509	0.01525	0.01941	0.01241	0.00795	0.00704	0.00630	0.00818	0.01020
6-7	0.00000	0.09330	0.03929	0.03738	0.01875	0.01277	0.01191	0.00939	0.01192	0.01404
8-9	0.00000	0.00000	1.80000	0.17889	0.08777	0.04347	0.02844	0.03080	0.03535	0.04960
10-13	0.00000	0.00000	0.00000	0.00000	0.00000	1.80000	0.27217	0.17116	0.12500	0.15000
≥ 14	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.90000
Negative Jetprobability										
2	0.04338	0.03106	0.02853	0.04071	0.03782	0.04338	0.04577	0.05621	0.05013	0.05639
3	0.02781	0.02611	0.02202	0.03105	0.03626	0.04278	0.04718	0.05172	0.05647	0.05006
4	0.05294	0.01714	0.02411	0.03363	0.03773	0.04049	0.04718	0.05836	0.04557	0.05899
5	0.09000	0.01328	0.02718	0.02601	0.03529	0.04138	0.05517	0.05393	0.05811	0.05313
6-7	0.00000	0.00000	0.02656	0.02071	0.02353	0.05259	0.05113	0.06625	0.06991	0.05088
8-9	0.00000	0.00000	0.00000	0.00000	0.00000	0.08800	0.02146	0.06971	0.04551	0.04250
10-13	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06923	0.17600	0.12143
≥ 14	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
errors										
2	0.00399	0.00193	0.00144	0.00301	0.00219	0.00176	0.00183	0.00194	0.00244	0.00335
3	0.00813	0.00305	0.00197	0.00360	0.00285	0.00223	0.00232	0.00231	0.00315	0.00389
4	0.02853	0.00523	0.00374	0.00607	0.00436	0.00318	0.00333	0.00335	0.00390	0.00572
5	0.09487	0.01150	0.00812	0.00982	0.00678	0.00507	0.00545	0.00478	0.00629	0.00795
6-7	1.80000	0.07200	0.01776	0.01663	0.00965	0.00864	0.00782	0.00731	0.00924	0.00995
8-9	0.00000	0.00000	1.80000	0.00000	0.09000	0.04243	0.01703	0.02243	0.02056	0.02814
10-13	0.00000	0.00000	0.00000	0.00000	0.90000	0.60000	0.36000	0.07391	0.17889	0.13226
≥ 14	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Table A.2: Probability matrix for Jetprobability positive tags and mistags for leading jets in the QCD sample, as a function of the number of SVX tracks and the corrected jet E_T.

Bibliography

[1] S. L. Glashow, Nucl. Phys. 22, 579(1961)

S. Weinberg, Phys. Rev. Lett. 19, 1264(1967)

A. Salam, *Elementary Particle Theory; Relativistic Groups and Analyticity (Nobel Symposium No. 8)*, ed. N. Svarthilm, Almqvist and Wiskell, Stockholm, (1968)

S. L. Glashow, J. Illiopoulos and L. Maiani, Phys. Rev. D 2, 1285(1970)

[2] F. Abe et al., Phys. Rev. Lett. 74, 2626(1995)

[3] S. Abachi et al., Phys. Rev. Lett. 74, 2632(1995)

[4] M. Kobayashi and M. Maskawa, Prog. Theor. Phys. 49, 652(1973)

[5] V. Berger and R. Phillips, *Collider Physics*, (Addison-Wesley Publishing Company, New York) (1987)

[6] F. Halzen and A. Martin, *Quarks and Leptons: An Introductory course in Modern Particle Physics*, (John Wiley & Sons, New York) (1984)

[7] I. J. R. Aitchison and A. J. G. Hey, *Gauge Theories in Particle Physics*, 2nd edition, IOP Publishing Ltd. (1989)

[8] D. Perkins, *Introduction to High Energy Physics*, (Addison-Wesley Publishing Company) (1987)

[9] S. Frixione et al., CERN-TH/97-16, HEP-PH/9702283

[10] A. D. Martin et al., Phys. Lett. B356, 89(1995), HEP-PH/9506423

[11] G. Kane, *Modern Elementary Paricle Physics*, (Addison-Wesley Publishing Company) (1993)

[12] P. Frazini, Phys. Reports, 173, 1(1991)

[13] F. Abe et al., FERMILAB-PUB-97/397-E, (1997) submitted to in Phys. Rev. D, Rapid Communications.

[14] The LEP Collaborations, CERN-PPE/97-154 (1997)

[15] I. Bigi et al., Phys. Lett. B181, 157(1986)

[16] S. Dawson et al., Nucl. Phys., B309, 607(1988)

[17] G. Altarelli et al., Nucl. Phys., B308, 725(1988)

[18] R. K. Ellis et al., Phys. Lett. B259, 492(1991)

[19] L. Orr, Phys. Rev. D44, 88(1991)

[20] F. Abe et al., Phys. Rev. Lett. 80, 2779(1998)

[21] F. Abe et al., Phys. Rev. Lett. 79, 3585(1997)

[22] F. Abe et al., Phys. Rev. Lett. 79, 1992(1997)

[23] F. Abe et al., Phys. Rev. Lett. 80, 2773(1998)

[24] F. Abe et al.,Phys. Rev. Lett. 80, 2767(1998)

[25] S. Abachi et al., Phys. Rev. Lett. 79, 1441(1997)

[26] S. Abachi et al., FERMILAB-PUB-98/031-E, (1998) submitted to Phys. Rev. D. S. Abachi et al., Phys. Rev. Lett. 79, 1197(1997) S. Abachi et al., FERMILAB PUB-97/172-E(1997) submitted to Phys. Rev. Lett.

[27] S. D. Holmes, Fermilab-conf-91/141, (1991) (unpublished)

[28] S. D. Holmes, Fermilab-conf-87/160, (1987) (unpublished) Book Co., Inc., New York, (1962)

[29] S. van de Meer, *An Introduction to Stochastic Cooling*, Proceedings of the SLAC and Fermilab Summer Schools on Particle Accelerators, edited by M. Dienes and M. Month, AIP Conference Proceedings 153, American Institute of Physics, New York, (1987)

[30] F. Abe et al., Nucl. Instrum. Methods Phys. Res. A 271, 387 (1988)

[31] S. Abachi et al., Nucl. Instrum. and Methods, A338, 185 (1994)

[32] D. Amidei et al., Nucl. Instrum. and Methods, A350, 73 (1994)

[33] G. Sciacca et al., Fermilab-Conf-94/205-E.

[34] F. Abe et al., Phys. Rev. Lett. 76, 3070(1996)

[35] F. Abe et al., Phys. Rev. D 50, 5518(1994) F. Abe et al., Phys. Rev. D 50, 5535(1994) F. Abe et al., Phys. Rev. D 50, 5550(1994)

[36] J. Lewis et al., CDF internal document 2858, (1995)

[37] M. Gold, CDF internal document 912, (1989)

[38] D. Gerdes, CDF internal document 2903, (1995)

[39] S. Belforte et al, CDF internal document 2361, (1994)

[40] C. Grosso-Pilcher et al., CDF internal document 3021, (1995)

[41] S. E. Kopp, C. Grosso-Pilcher, CDF internal document 1992, (1993)

[42] F. Abe et al., Phys. Rev. D 45, 1448(1992)

[43] R. M. Carey, Ph.D. dissertation, Harvard University (1989)

[44] J. Tonisson, CDF internal document 2868, (1994)
N. Eddy, CDF internal document 3534, (1996)
W. Yao et al., CDF internal document 3257, (1995)

[45] S. Kuhlmann, Fermilab-conf-94/148-E, (1994)

[46] S. Berhends, CDF internal document 3550, (1996)

[47] B. Winer, Ph.D. dissertation, University of California, Berkeley, (1991)

[48] F. Abe et al. Phys. Rev. Lett., 70, 4042(1993)

[49] F. A. Berends et al., Nucl. Phys. B357, 32(1991)
S. D. Ellis et al., Phys. Lett. B154, 435(1985)
F. A. Berends et al., Phys. Lett. B224, 237(1989)

[50] F. A. Berends et al., Nucl. Phys. B357, 32(1991)

[51] W Giele et al., Nucl. Phys B403, 633(1993)

[52] W. Yao and D. Glenzinski, CDF internal document 2989, (1995)

[53] D. Glenzinski, Ph.D. dissertation, The Johns Hopkins University, (1995)

[54] W. Yao et al., CDF internal document 2716, (1994)

[55] W. Yao et al., CDF internal document 2989, (1995)

[56] C. Miao and W. Yao, CDF internal document 3542, (1996)

[57] D. Buskulic et al., Phys. Lett. B313, 535(1993)

[58] D. Kestenbaum, Ph.D. dissertation, Harvard University, (1996)

[59] F. Abe et al., Phys. Rev. D50, 2966(1994)

[60] M. Mangano, Nucl. Phys. B405, 536(1993)
Z. Kunst, Nucl. Phys. B247, 339(1984)

[61] G. Unal, CDF internal document 3389, (1995)

[62] G. Unal, CDF internal document 3513, (1995)

[63] F. Abe et al., Phys. Rev. Lett. 75, 608(1995)

[64] F. Abe et al., Phys. Rev. D50, 5562(1995)

[65] Y. Cen et al., CDF internal document 2965, (1995)

[66] J. Gunion et al., *The Higgs Hunter's Guide* (Addison-Wesley Publishing Company, New York), (1990)

[67] E. Eichten and K. Lane, Phys. Lett. B388, 803(1996)
E. Eichten. K. Lane and J. Womersley, FERMILAB-PUB-97-116-T.

[68] G. Marchesini and B. R. Webber, *Nucl. Phys.* **B238** 1 (1984); G. Marchesini and B. R. Webber, *Nucl. Phys.* **B310** 461 (1988)

[69] F. A. Berends, W. T. Giele, H. Kuijf, and B. Tausk, *Nucl. Phys.* **B357** 32 (1991)

[70] T. Sjöstrand, Report No. CERN-TH-6488/92 (1992)
(URL: http://wwwcn.cern.ch/asdoc/pythia_html3/pythia57.html) PYTHIA V5.6 is used

[71] F. Paige and S. D. Protopopescu, BNL Report No. 38034, 1986(unpublished)

[72] F. Abe et al., Phys. Rev. D 52, 4784(1995)

[73] R. Field and R. Feynmann, Nucl. Phys. B 136, 1(1978)

[74] C. Peterson et al., Phys. Rev. D 27, 105(1983)

[75] F. Abe et al., Phys. Rev. Lett. 74, 850(1995)

[76] J. Lewis and P. Avery, CDF internal document 2724, (1994)

[77] M. Shapiro et al., CDF internal document 1810, (1992)

[78] S. Catani et al., Phys. Lett. B378, 329(1996)

[79] E. Berger & H. Contopanagos, Phys. Rev. D54, 3085(1996)

[80] E. Laenen et al., Phys. Lett. B321, 254 (1994)

[81] Particle Data Group, Phys. Rev. D 50, Part I (1994)

[82] J. Ohnemus and J. F. Owens, Phys. Rev. D43, 3477(1991)

J. Ohnemus, Phys. Rev. D, bf44, 1403 (1991)

J. Ohnemus, Phys. Rev. D 43, 3626 (1991)

[83] T. Stelzer and S. Willenbrock Phys. Lett. B357, 125 (1995)

M. Smith and S. Willenbrock, Phys. Rev. D54,6696(1996)