Fermilab FERMILAB-THESIS-1996-45

IDENTIFICATION OF TOP QUARK PRODUCTION USING KINEMATIC AND B-TAGGING TECHNIQUES IN THE LEPTON + JETS CHANNEL IN $p\overline{p}$ COLLISIONS AT $\sqrt{s} = 1.8$ TeV

Gareth M. Houk

A DISSERTATION

in

PHYSICS AND ASTRONOMY

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

1996

Supervisor of Dissertation

Graduate Group Chairperson

DEDICATION

This work is dedicated to Craig Lafferty. A great teacher is one whose example inspires his student in odd and distant moments and in endeavors unforeseeable and seemingly unrelated to his teachings.

ACKNOWLEDGMENTS

It is a commonplace within the academic community that theses are written by the score, read by a mere handful of experts, then filed under the category "too obscure for the laity" or "for accreditation purposes only." I hope that this text embodies something more. Since I began graduate level research, the cancellation of the SuperCollider has rendered the future of high energy physics palpably more uncertain. I have wondered, in the face of this disappointment, whether my art is indeed worthy of society's support. By our efforts, we scientists assert not only that there is an objective truth, but that it is worthwhile to seek it out. My epistemology professor in college, Dr. Paul Feyerabend, used to ask "what is knowledge, and what's so great about it?" This query circumscribes scientific inquiry yet challenges us continually, both as a consequence of and as a reward for our progress, to reinvent the measure of that inquiry's worth. I believe it is precisely this reinvention that makes science worthwhile: each scientific work, since it is reproducible, may stand as the foundation of its successors, even when that succession revolutionizes our understanding of the original work. Our eighteen year search for the top quark exemplifies this progression. In itself, the discovery of the top quark is not revolutionary, since its absence atop the Standard Model quark family hierarchy is as conspicuous as Mendeleev's periodic table gaps. Rather, the top quark is a stepping stone from which we may once again cast into the unknown. If we must ask what's so great about the discovery of top quark (and I believe it is our duty to ask that of all our scientific endeavors), I would assert that its value resides in what stands beyond it: the prospect of new and unforeseen horizons.

I have been very fortunate to participate in a research project which stands at the crossroads between the expected and the unknown. Such crossroads do exist, and students can contribute meaningfully to their passage. This account of that passage could not have been completed without a great deal of help and patience from many people. I owe a debt of gratitude to my family, Carolyn and Keelan Houk, who have given me the strength to persevere and without whom the effort would be sound and fury. I thank my parents, Ted and Carol Houk for culturing in me the love of intellect. Brig Williams, Guillaume Unal, Brian Harral, and Steve Hahn gave me incomparable guidance in this sometimes overwhelming field. I would be remiss to neglect my other friends and accomplices both at Penn and at CDF for the good times I will never forget and likely never recover from: Joe, Nir, Jim, John, Farrukh, Rick, Owen, Chris, Sarah, Doug, Richard, Dee. Finally, I thank two more of my teachers, Lynn Stevenson and George Trilling for their excellence, inspiration and direction.

ABSTRACT

IDENTIFICATION OF TOP QUARK PRODUCTION

USING KINEMATIC AND B-TAGGING TECHNIQUES

IN THE LEPTON + JETS CHANNEL

IN $p\overline{p}$ COLLISIONS AT $\sqrt{s} = 1.8$ TeV

Gareth M. Houk

Hugh H. Williams

We identify top quark production in $p\overline{p}$ collisions at $\sqrt{s} = 1.8$ TeV. We select events with a high energy lepton, evidence of a neutrino, and significant jet activity. Within this sample, $t\bar{t}$ events are isolated first by identifying jets arising from b-quark production via the long lifetime of their decay products and second by selecting events which are kinematically inconsistent with background expectations. We demonstrate a technique for combining these two approaches in order to isolate a significant $t\bar{t}$ signal and observe 11 events compared to a predicted background of 1.5±0.4 events. This observed signal excess of 4.5σ is more significant than the observed excess using either the b-tagging or kinematic technique alone.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Theory

We present a direct observation of top quark production in $p\overline{p}$ collisions with center-of-mass energy $\sqrt{s} = 1.8$ TeV at the Fermi National Accelerator Laboratory (FNAL) Tevatron. The data were collected between August, 1992 and February, 1995, and correspond to an integrated luminosity of $67\mathrm{pb}^{-1}$. The observation was made using the Collider Detector at Fermilab (CDF) by detecting the decay of a $t\bar{t}$ pair to a lepton (l), neutrino (ν), and several hadronic jets (j) through the decay chain $t \to Wb \to l\nu j$ and $t \to Wb \to jjj$. This signature by itself is insufficient to isolate the top quark signal; to this end, we apply two further techniques. First, we identify jets arising from b-quark production, since $t\bar{t}$ events contain two b-quarks and jets in background events are predominantly from light quarks and gluons rather than b-quarks. This technique is called *b-tagging*, and is accomplished here by searching for b-hadron decays within jets by identifying decay vertices stemming from long lived particles. The results of this analysis have been published previously [1]. Second, we separate $t\bar{t}$ from background events based on the distribution of jet energies and jet angles with respect to the $p\overline{p}$ beam, which we collectively refer to as the event *kinematics*. This second analysis is modified from that presented in [2] in order to facilitate combination with the b-tagging approach. Finally, we use both kinematics and b-tagging to select $t\bar{t}$ events. The results of this combined analysis and the demonstration of this technique are the primary focus of this paper.

This paper is organized as follows. In this chapter, we describe the Standard Model of particle interactions, the theoretical framework within which the top quark is one of the final unobserved fundamental particles. We use the Standard Model to understand why we expect the top quark to exist, what the top quark's known properties are, and how those properties may be exploited to search for its production in $p\overline{p}$ collisions. In chapter 2 we describe the FNAL Tevatron, which produces the $p\overline{p}$ collisions, and the CDF detector, with which we examine those collisions. Chapter 3 describes the selection of a sample of $p\overline{p}$ collision events with the characteristic $t\bar{t}$ signature of a high energy lepton, large energy imbalance transverse to the $p\overline{p}$ beam, and three or more high energy hadronic jets. In chapter 4 we apply the b-tagging technique to this sample and compare the observed number of b-tags to the number expected from background processes only. In chapter 5 we use the expected distributions of jet energy and jet angle with respect to the $p\overline{p}$ beam to separate $t\bar{t}$ and background lepton+jet events. We define a variable which parametrizes the likelihood that the kinematics of a given event is from background sources. We use this variable to select events which are more consistent with $t\bar{t}$ event kinematics than background event kinematics and compare the observed number of events to background expectations. Chapter 6 combines the b-tagging and kinematic approaches by selecting events which are kinematically inconsistent with background expectations and also b-tagged. We observe 11 such events with an estimated background of 1.5 ± 0.4. The probability that this observation could be due to an upwards fluctuation of the background is the *significance* of the observation. The significance is 4×10^{-6}, which corresponds to a 4.5σ discrepancy for a gaussian distributed variable. This observed signal excess is more significant than the observed excess using either the b-tagging or kinematic technique alone.

1.1 The Standard Model

Within our current experimental capability, all matter in nature is composed of two types of fundamental spin 1/2 fermions: quarks and leptons. Free quarks have not been observed in nature; instead quark-antiquark and three-quark bound states are observed. These are the mesons and baryons respectively, which we collectively refer to as *hadrons*. Hadrons include the pion and the proton. Leptons include the electron and neutrino. The Standard Model [3, 4] of particle physics is our current best understanding of the fundamental interactions of quarks and leptons. The search for the top quark began with the discovery of its partner, the bottom or b quark. Since quarks and leptons come in pairs, the discovery of this fifth quark strongly suggested the existence of the top quark. For our purposes we will concentrate on those aspects of the Standard Model which are most relevant to the top quark: its place in the quark family hierarchy, theoretical motivation and experimental evidence supporting its presence, experimental limits on its mass, its production mechanisms, decay mechanisms and experimental signatures.

1.1.1 The Electroweak Force and the Fermion Family Structure

The electroweak force is important to top physics because it mediates top quark decays ($t \to bW$) and because the classification of quark pairs into families and the instability of the higher mass families is dictated by the degree to which quarks are eigenstates of the electroweak force. In the Standard Model, electroweak interactions are described by a local gauge theory based on the symmetry group SU(2)$\times$U(1). This symmetry group generates four interaction mediators, which take on mass when the manifest symmetry is hidden via the Higgs mechanism. These mediators, or gauge bosons, are the massless photon and the massive $W^{\pm}$ and Z bosons. The weak angle $\sin\theta_W$ is an input parameter of the theory; it characterizes the mass non-degeneracy of the W and Z bosons and enters into electroweak transition coupling strengths. The matter fields in the Standard Model are spin 1/2 fermions, quarks and leptons which transform as specific representations under SU(2), and which are mass-ordered into three generations. The top quark belongs to the third and most massive quark generation. The separation of quark and lepton helicity states accounts for the observed parity violation of weak interactions. The left-handed helicity states take on the doublet representation of SU(2); we assign to them the isospin quantum numbers $t_L = 1/2, t_{3L} = \pm 1/2$. For clarity, we have used the subscript "L" on isospin to indicate left-handed states. As members of a doublet,

$$
\begin{array}{lcccccc}
 & t_{3L} & Q & & & \\
\text{quarks :} & \begin{array}{c}+1/2\\-1/2\end{array} & \begin{array}{c}+2/3\\-1/3\end{array} & \begin{pmatrix} u \\ d_c \end{pmatrix}_L & \begin{pmatrix} c \\ s_c \end{pmatrix}_L & \begin{pmatrix} t \\ b_c \end{pmatrix}_L \\
\text{leptons :} & \begin{array}{c}+1/2\\-1/2\end{array} & \begin{array}{c}0\\-1\end{array} & \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L & \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L & \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L
\end{array}
$$

Table 1.1: *The generations of left-handed quark and lepton doublets in the Standard Model. The quantities Q and t_{3L} are charge and the third component of isospin respectively. The $Q = -1/3$ quarks refer to the Cabibbo rotated states with respect to the eigenstates of the strong interactions, as indicated by the subscript c.*

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

Table 1.2: *The Cabibbo-Kobayashi-Maskawa rotation matrix.*

each left-handed quark has an *isospin partner*. The left-handed quark and lepton doublets are listed in table 1.1. The subscript c on the charge $Q = -1/3$ quarks indicates that the eigenstates with respect to weak transitions are not identical to the mass eigenstates. Rather, the weak eigenstates are rotated according to the Cabibbo-Kobayashi-Maskawa (CKM) matrix in table 1.2 in order to allow for generation-changing charged weak transitions. This rotation is not relevant for the leptons because they do not participate in strong interactions. The right-handed quark and charged lepton helicity states take on the singlet representation of SU(2); mathematically, we assign to these right-handed states the weak isospin quantum numbers $t = 0$,$t_{3R} = 0$. The quark states are u_R, $(d_c)_R$, c_R, $(s_c)_R$, t_R, $(b_c)_R$, where the $Q = -1/3$ states are Cabibbo-rotated, and the lepton states are e_R, μ_R, and τ_R,

c_R^f	$=$	$-Q^f \sin^2\theta_W$
c_L^f	$=$	$t_{3L}^f - Q^f \sin^2\theta_W$
c_V^f	$=$	$t_{3L}^f - 2Q^f \sin^2\theta_W$
c_A^f	$=$	t_{3L}^f

Table 1.3: *Coupling constants of the neutral electroweak interaction. The superscript f indicates fermion species. Q^f and t_{3L}^f are listed in table 1.1 for left-handed fermions. Right-handed fermions have $t_{3R}^f = 0$ and the same value of Q^f as the corresponding left-handed state.*

where the assumption of masslessness eliminates the right-handed neutrino states.

Top decay signatures will be discussed in section 1.4.1. Here we note that the top quark decays via the electroweak force predominantly to a b quark and a W, where the W is real for $M_{top} > M_b + M_W \simeq 86$ GeV/c^2.

In section 1.2.1, we will examine the implication of several Z boson-mediated interactions for the existence of top; in anticipation of this discussion, we examine the Standard Model neutral weak current couplings. Helicity states in neutral weak interactions are connected by right and left coupling constants c_R and c_L or alternately, vector and axial coupling constants c_V and c_A. These couplings are listed in table 1.3.

1.1.2 The Strong Force

We are primarily interested in the strong force for understanding top quark production mechanisms in $p\overline{p}$ collisions, and also in the formation of hadronic *jets* from the quarks produced by the decay of top quarks. The strong force affects quarks and governs the binding of baryons and mesons. It is described by a non-Abelian gauge theory based on the symmetry group SU(3). The degree of freedom associated to group transformations is called color, hence the theory is called Quantum Chromodynamics (QCD). The eight generators of SU(3) transformations correspond to the gauge quanta of the strong force; these are the gluons. Gluons are massless and the SU(3) symmetry is unbroken. Since gluons carry color, they may interact with other gluons. Quarks come in three colors which correspond to the three components of an SU(3) triplet. Leptons are colorless and do not participate in the strong force. All observable free particles are color singlets. Thus free quark-antiquark meson states and free three-quark baryon states are observed rather than free quarks. QCD accounts for the lack of free colored particles via *asymptotic freedom*, in which the coupling strength α_s decreases with increasing energy transfer Q, i.e. shorter distances. This property, $\lim_{Q^2\to\infty} \alpha_s(Q^2) = 0$, is described at leading order by equation 1.1, where f is the number of quark flavors ($f = 6$ including the top quark) and Λ is a characteristic energy scale of the order 100-200 MeV[5, p. 298]. Equation 1.1 is valid for $Q^2 \gg \Lambda^2$.

$$\alpha_s(Q^2) = \frac{12\pi}{(33 - 2f)\ln(Q^2/\Lambda^2)} \tag{1.1}$$

In the context of high energy collider interactions, asymptotic freedom allows us to treat strongly bound quarks (such as those in the colliding protons and antiprotons) as essentially free with respect to the weak interactions we will consider. Conversely, the coupling α_s increases with decreasing momentum transfer $Q^2 \to 0$. As colored particles are separated, energy density increases to the point at which the materialization of other partons is energetically favorable. The coalescence of free color singlets from this materialization yields final state hadrons rather than free quarks and gluons, which we collectively refer to as *partons.* As a parton propagates, it evolves into a shower of partons and finally into hadrons. These processes are called *parton shower evolution* and *hadronization.* At large collision energies, partons produce collimated *jets* of hadrons. We are interested in the jets produced in the top quark decay $t \rightarrow bW$; in this process jets arise from the hadronization of b quarks and from the hadronization of quarks produced in subsequent hadronic W decays. Owing to their large mass, top quarks themselves decay prior to hadronization rather than producing a single hadron jet.

1.2 Experimental Evidence in Support of the Top Quark

Prior to performing a direct search, we seek to understand why we expect to find the top quark and in what mass range we expect it. In the Standard Model, there is no limit *a priori* on the number of quark or lepton generations. The discovery of the third generation τ lepton in 1975 [6] and bottom quark in 1977 [7] constituted a plausible but theoretically nonessential experimental extension of the number of fermion generations. In contrast, the discovery of top constitutes the completion of the third quark generation; the presence or absence of the top quark affects the properties of its third-generation isospin partner, the b quark. Without direct experimental observation of the top quark, it is possible to determine whether it exists: we may measure the third component of weak isospin, t_{3L}, for the left-hand helicity state of the bottom quark and hence determine whether it is a singlet ($t_{3L} = 0$) or the SU(2) doublet partner ($t_{3L} = -1/2$) of another quark. By definition, the isospin partner of the b quark is the top quark. We determine the b quark isospin from the forward-backward asymmetry A_{FB} and the branching fraction R_b in section 1.2.1.

The top mass may also be inferred in the absence of direct evidence, as described in section 1.2.2. Since the top production cross section decreases with top mass, null direct searches for top decays can be inverted to give a lower limit on the top mass. In addition, precision electroweak measurements may be used in section 1.2.3 to set a decay mode-independent lower limit on the top mass and, further, to establish a top mass range.

1.2.1 Forward-Backward Asymmetry A^b_{FB} and Branching Ratio R_b in $e^+e^- \to Z \to b\bar{b}$: Implications for the Existence of Top

We examine the value of two experimental observables which depend directly on the existence of the top quark through the isospin properties of its isospin partner the b quark. These observables are the forward-backward asymmetry A^b_{FB} and the branching ratio R_b in $e^+e^- \to Z \to b\bar{b}$ events. The forward-backward asymmetry in the reaction $e^+e^- \to b\bar{b}$ is defined as

$$A^b_{FB} = \frac{\left[\int_{cos\theta>0} - \int_{cos\theta<0}\right] d\Omega \frac{d\sigma}{d\Omega}(e^+e^- \to b\bar{b})}{\left[\int_{cos\theta>0} + \int_{cos\theta<0}\right] d\Omega \frac{d\sigma}{d\Omega}(e^+e^- \to b\bar{b})} \tag{1.2}$$

where $\frac{d\sigma}{d\Omega}$ is the differential cross section, and θ is the angle between the electron beam and the final state b quark. At tree level, the process consists of the two diagrams in figure 1.1, one mediated by a photon and the other by a Z boson. We

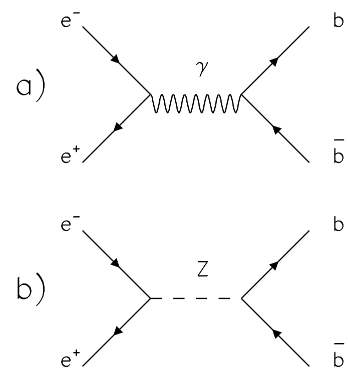


Figure 1.1: *Tree level diagrams for the process* $e^+e^- \to b\bar{b}$.

examine A^b_{FB} as measured at the Z-pole ($\sqrt{s} \simeq M_Z$), where the electromagnetic matrix element is very small compared to the neutral weak matrix element and also out of phase, canceling interference effects [8]. This asymmetry is given by:

$$A^b_{FB} = \frac{3}{4}\frac{2c^e_V c^e_A}{(c^e_V)^2 + (c^e_A)^2}\frac{2c^b_V c^b_A}{(c^b_V)^2 + (c^b_A)^2} \tag{1.3}$$

where the fermion axial and vector couplings c^f_V and c^f_A are given in table 1.3 with the Standard Model values of Q^f and t^f_{3L} given in table 1.1. If top does not exist, then the b quark is an isospin singlet ($t_{3L} = 0$), altering the b quark couplings from $c^b_V = -\frac{1}{2} + \frac{2}{3}\sin^2\theta_W$ and $c^b_A = -\frac{1}{2}$ to $c^b_V = \frac{2}{3}\sin^2\theta_W$ and $c^b_A = 0$. Without top, $A_{FB} = 0$ while in the presence of top (assuming $\sin^2\theta_W = 0.23$) $A^b_{FB} = 0.11$. Small corrections alter this tree-level calculation as described below.

Experimentally, A_{FB}^b is measured by searching for $Z \to jj$ hadronic jet events in which the jets may be identified as b-jets. The identification of b-jets is accomplished using b-tagging (lifetime-tagging) similar to that described in chapter 4 in the context of our top quark search or using the identification of leptons from b decays. In addition, one must discriminate between jets arising from b and $\bar{b}$ production; this discrimination is accomplished using the sign of the leptonic decay or using a momentum-weighted jet charge in the case of the lifetime tag. A third method involves reconstructing the decay of $D^{*\pm}$ mesons. Experimental results are corrected for the photon exchange amplitude, QED corrections, $\sqrt{s} \neq M_Z$, and $B^0\overline{B^0}$ mixing, which changes the sign of the tagged lepton or reconstructed $D^{*\pm}$. The result is $A_{FB}^{(0,b)} = 0.0997 \pm 0.0031$ [9]. For comparison to theoretical predictions based on the precisely measured Z mass, electroweak radiative corrections must be included. These corrections depend upon the renormalization scheme chosen and assumptions about the top and Higgs mass. Using $M_Z = 91.1884 \pm 0.0022$ GeV/c^2 [9], $M_{top} = 180$ GeV/c^2, and $M_H = 300$ GeV/c^2, the Standard Model predicts $A_{FB}^{(0,b)} = 0.101 \pm 0.001$ [10]. The agreement between the theoretical prediction and the experimental measurement is excellent and conclusively inconsistent with the Standard Model prediction in the absence of top.

Another Z-pole observable from which we may extract information about the b quark couplings is the $Z \to b\bar{b}$ hadronic branching ratio, $R_b = , (b\bar{b})/, (had)$. The Feynman diagrams involved are the same as in figure 1.1 where the $b\bar{b}$ pair may be replaced by a quark-antiquark ($q\overline{q}$) pair in the final state in determining the total hadronic width , (had). At the Z-pole, the electromagnetic term is small. The tree level partial width for each $q\overline{q}$ final state is:

$$, (q\overline{q}) \quad \alpha \quad (c_V^q)^2 + (c_A^q)^2$$

Here the charge $Q = +\frac{2}{3}$ $q\overline{q}$ final states may be $u\overline{u}$ or $c\overline{c}$, and the charge $Q = -\frac{1}{3}$ final states may be $d\overline{d}$, $s\overline{s}$, or $b\overline{b}$. The $t\overline{t}$ final state is not considered because the top mass is greater than half the Z mass, as we shall see in sections 1.2.2 and 1.2.3. Using $\sin^2\theta_W = 0.23$ and adding the contributions from all the possible final states in the case where $t_{3L}^b = -\frac{1}{2}$ yields $R_b = 0.220$. The singlet case, $t_{3L}^b = 0$, yields $R_b = 0.018$.

Experimentally, the measurement of R_b faces similar issues as the measurement of A_{FB}^b: one must identify $Z \to jj$ events and identify those events containing b-jet(s) using lepton or lifetime tagging. In this case, it is not necessary to identify the charge of the quark since we do not need to determine which jet arises from the b and which arises from the $\overline{b}$. Accounting for quark masses, QCD, QED, and electroweak radiative corrections, the Standard Model predicts $R_b = 0.2156 \pm 0.0004$ for a doublet and $R_b = 0.017$ for a singlet [11]. Recent experimental results from LEP yield $R_b = 0.2219 \pm 0.0017$ [9]. Although this value is 3.7σ larger than expected from the Standard Model, it is inconsistent with the absence of top.

Given the two observables A_{FB}^b and R_b, it is possible to determine the third component of weak isospin for left- and right-handed b-quarks. The results above lead to $t^b_{3L} = -0.500 \pm 0.005$ and $t^b_{3R} = -0.026 \pm 0.018$[1] [11], indicating that top must exist as the isospin partner of the b quark.

1.2.2 Top Mass Limits from Direct Searches

Despite indirect evidence that the top quark must exist, previous direct searches have not conclusively established its observation. Since the top production cross section falls as top mass increases (see section 1.3), a null direct search result may be used to set a lower limit on the top mass. The most recent limit from the D0 collaboration is $M_{top} > 131$ GeV/c^2 [12].

The CDF collaboration presented the first direct evidence for top quark production [13] using roughly a third of the data presented in the present analysis. The number of events observed exceeded background expectations by 2.8σ, insufficient to claim unequivocal discovery. Assuming that the excess resulted from top quark production, the mass indicated by that analysis was $M_{top} = 174 \pm 16$ GeV/c^2. Since we use a portion of the same data and some of the same search techniques in the present analysis (specifically b-tagging long-lived particles in jets), the results are correlated.

1.2.3 Indirect Top Mass Limits from Precision Electroweak Data

The D0 top mass limit presumes the Standard Model top decay mode $t \rightarrow Wb$. If top decays via a non-standard mode, then the limit may be invalid. A decay-

[1]This determination of t^b_{3L} uses the less recent values of $A^b_{FB} = 0.0967 \pm 0.0038$ $R_b = 0.2202 \pm 0.0020$.

independent mass limit may be obtained from the W boson width , $_W$, since the W decay channel $W \rightarrow \bar{b}t$ is available independent of the subsequent top decay chain and hence contributes to , $_W$. This limit requires $M_{top} + M_b \leq M_W$, and the corresponding limit is $M_{top} > 62$ GeV/c^2 [14].

The comparison of very precise experimental measurements to Standard Model electroweak predictions may be used to set stringent bounds on the top mass. Recent experiments at LEP have determined the Z mass to great precision (91.1884 ± 0.0022 GeV/c^2). In addition, many Z-pole observables have been measured, including the Z width , $_Z$, asymmetries (of which A^b_{FB} above is an example), and branching ratios (of which R_b above is an example) [9]. Polarized beams at SLC have made possible the measurement of the left-right asymmetry A_{LR} [15]. The UA2 [16], CDF [17] and D0 [18] collaborations have reported measurements of the W mass (the average of these measurements is 80.26 ± 0.16 GeV/c^2). These observables depend to varying degrees on the top mass M_{top}, the weak angle $\sin^2\theta_W$, the strong coupling α_s, and the Higgs mass M_H. Taking $M_H = 300$ GeV/c^2 as a central value, M_{top}, $\sin^2\theta_W$, and α_s may be simultaneously determined from a global fit to all precision electroweak observables. Using this technique, the top mass is estimated to be $M_{top} = 179 \pm 8^{+17}_{-20}$ GeV/c^2, where the first uncertainty accounts for precision electroweak experimental and theoretical error and the second uncertainty allows for Higgs mass variation between 60 GeV/c^2 $< M_H <$ 1 TeV/c^2 [10]. Since this top mass range is based upon the effect of the top mass on electroweak observables rather than a specific top quark decay mode, this determination constitutes an extension of the decay-independent top mass lower limits from Γ_W as described above. The bound 60 GeV/c^2 $< M_H$ is from LEP direct Higgs search [19]. Using a less recent determination of M_{top} using the global fit technique ($M_{top} = 175 \pm 11^{+17}_{-19}$ GeV/c^2)[11] leads to $M_{top} > 135$ GeV/c^2 at 95% CL. The bound $M_H < 600$ GeV/c^2 $- 1$ TeV/c^2 is based on theoretical arguments [20] and leads to $M_{top} < 209$ GeV/c^2 at 95% CL.

1.3 Top at the Tevatron, $\sqrt{s} = 1.8$ TeV

In $p\bar{p}$ collisions, the dominant production mechanism for top quarks with $M_{top} > M_W + M_b$ is through $t\bar{t}$ pair production, as opposed to single top quark production. Lowest order diagrams are shown in figure 1.2. For $M_{top} < M_W + M_b \simeq 86$ GeV/c^2,

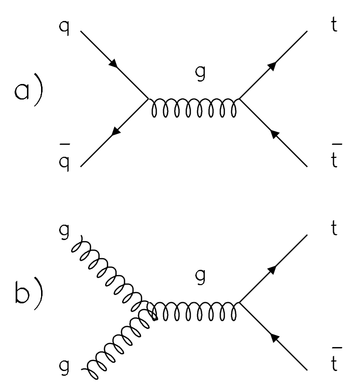


Figure 1.2: *Lowest order $t\bar{t}$ production diagrams.*

top quarks may also be produced through W decays to $t\bar{b}$, but as we have seen in sections 1.2.2 and 1.2.3, both direct ($M_{top} > 131$ GeV/c^2) and decay-independent indirect mass evidence ($M_{top} > 135$ GeV/c^2) indicate that the top mass is well above this threshold. Next-to-leading-order perturbative QCD cross section calculations are given in [21] and extended in [22]. The $t\bar{t}$ production cross section as calculated in [23] is shown in figure 1.3. Gluon fusion processes such as 1.2b comprise roughly 20% of the total cross section for $M_{top} > 150$ GeV/c^2.

1.4 Top Quark Signatures

1.4.1 $t\bar{t}$ Decay Modes and Branching Fractions

Within the Standard Model, the top quark decays predominantly to bW since the Cabibbo factors for the decay modes $t \to sW$ and $t \to dW$ are less than a few percent[24, p. 1315]. Top decays thus produce one b quark jet and a W boson, which may decay leptonically or hadronically, as shown in figure 1.4. All three ($e\nu_e$, $\mu\nu_\mu$, $\tau\nu_\tau$) leptonic W decays are kinematically allowed. Hadronic W decays produce first and second generation $q\overline{q\prime}$ pairs with a color factor of three for a subtotal of six potential decay paths. Together with the three leptonic W decays, there are nine possibilities, with the following branching ratios before higher order corrections:

decay mode			BR
t	$\to$	$be\overline{\nu}_e$	$\frac{1}{9}$
	$\to$	$b\mu\overline{\nu}_\mu$	$\frac{1}{9}$
	$\to$	$b\tau\overline{\nu}_\tau$	$\frac{1}{9}$
	$\to$	$q\overline{q\prime}$	$\frac{2}{3}$

In $t\bar{t}$ events, the t and $\bar{t}$ decay independently, yielding the following combined branching ratios before higher order corrections:

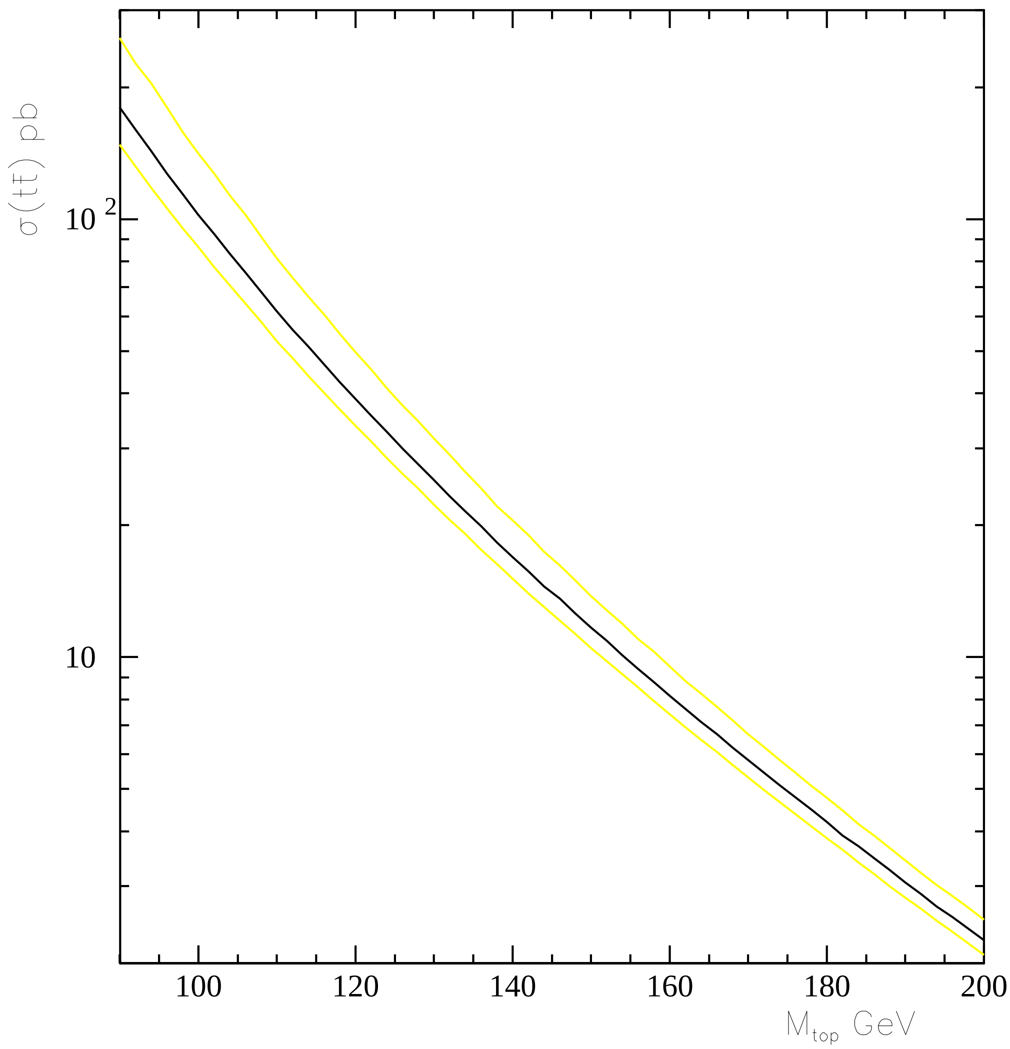


Figure 1.3: *Theoretical top-antitop production cross section* $\sigma_{t\bar{t}}$ *for* $\sqrt{s} = 1.8$ *TeV.*

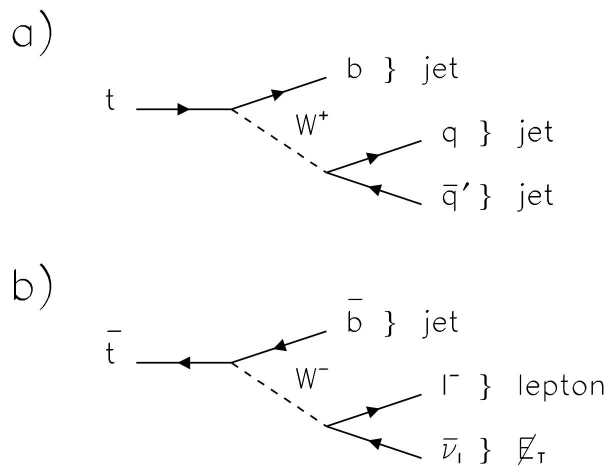


Figure 1.4: *Top quark decay diagrams.*

		decay mode	BR	
$t\bar{t}$	$\rightarrow$	$e\nu_e e\nu_e bb$	1/81	
	$\rightarrow$	$e\nu_e \mu\nu_\mu bb$	2/81	
	$\rightarrow$	$\mu\nu_\mu \mu\nu_\mu bb$	1/81	5%
	$\rightarrow$	$e\nu_e jjbb$	12/81	
	$\rightarrow$	$\mu\nu_\mu jjbb$	12/81	30%
	$\rightarrow$	$jjjjbb$	36/81	44%
	$\rightarrow$	τX	17/81	21%

where b indicates a jet from a b quark, j indicates a light quark jet from a hadronic W decay, and final state antiparticle designations have been omitted. In the final column, we have separated out the branching ratios for four classes of $t\bar{t}$ signatures: the first is the *dilepton* category, the second is the *lepton+jets* category, the third is the *all hadronic* category, and the fourth consists of the τ channels.

This paper involves only the lepton+jets, or equivalently the *lepton+multijet* channel, which has a relatively large branching fraction and reasonable background. The signature for $t\bar{t}$ events in the lepton+jets channel is shown in figure 1.5 and consists of two b-jets from the two $t \rightarrow Wb$ decays, a single lepton and a single neutrino from the leptonic W decay, and two more quark-jets from the hadronic W decay. Since the neutrino does not interact with the detector, we infer its presence from the calorimeter energy imbalance in the plane transverse to the beam, or missing transverse energy $\not{E}_t$. In practice, we require only three jets rather than four to account for inefficiency in reconstructing all four jets. The dominant background in this channel is QCD production of jets in association with W boson production,

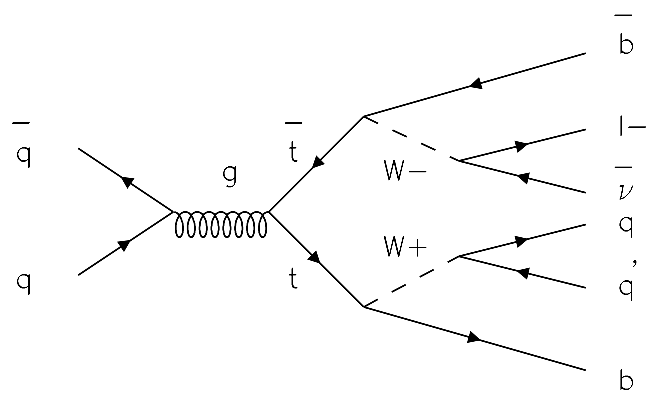


Figure 1.5: *The lepton+jets decay signature of $t\bar{t}$ events.*

as exemplified by figure 1.6. We refer to these background processes as *QCD W+jet* events.

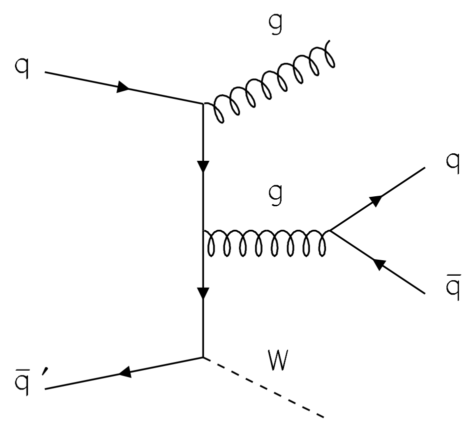


Figure 1.6: *An example of QCD W+jets production. Such processes are the dominant background to $t\bar{t}$ events in the lepton+jets channel.*

Top signals may also be isolated in the other decay modes. Despite its low branching ratio and hence small signal statistics, the dilepton channel search is very clean and characterized by low background. The results of such a search can be found in [1]. Recent work has extended the dilepton channel at CDF to include τ lepton signatures[25]. Although the QCD multijet background to the six-jet signature of the all-hadronic channel is very large, the combination of b-tagging, kinematics, and mass reconstruction may be used to isolate the $t\bar{t}$ signal[25, 26].

1.5 Objective and Regime of Validity

The objective of this work is the demonstration of a technique for isolating a very pure sample of $t\bar{t}$ events using both b-tagging and kinematic techniques. B-tagging entails identifying b-jets by searching for the decay vertices of long-lived b-hadrons within jets. Beginning with the lepton+multijet signature, we require at least one b-tagged jet. The kinematic technique distinguishes $t\bar{t}$ events from the dominant QCD W+jet background based on jet energies and and jet angles with respect to the $p\bar{p}$ beam. The isolation of a $t\bar{t}$ signal in this analysis assumes the Standard Model top decay mode $t \rightarrow Wb$. In the presence of a charged Higgs scalar H such that $M_H + M_b < M_{top}$, the decay $t \rightarrow H^+ b$ may be favored. Since the Higgs couples more strongly to massive particles, the dominant signature in this case would be $H \rightarrow \tau^+ \nu_\tau$ rather than the e and μ signature for W mediation. A search for such decays[27] resulted in exclusion of the entire $M_{top} - M_H$ plane at 95% confidence level for $M_{top} < M_W + M_b$, and an extension of this search has excluded additional regions of the $M_{top} - M_H$ plane[28]. The possibility of non-Standard Model decays is important to this analysis in that the signal may be suppressed or absent despite the presence of top.

Chapter 2

Experimental Design

The experimental apparatus used to search for $t\bar{t}$ events in $p\bar{p}$ collisions has two major components: the Tevatron, which produces and accelerates the protons and antiprotons to center-of-mass energy $\sqrt{s} = 1.8$ TeV, and the Collider Detector at Fermilab, which is a multipurpose particle detector.

2.1 The Tevatron

The Tevatron [29] at Fermi National Accelerator Laboratory is a superconducting synchrotron. As the highest energy particle accelerator in the world, the Tevatron sustains a rich program of numerous fixed target experiments and two collider experiments. A diagram of the Tevatron and its key components is provided in figure 2.1. For collider running, the Tevatron stores six counter-rotating bunches of protons and antiprotons with energy 0.9 TeV. Collisions occur at six interaction regions, labeled A0-F0 in figure 2.1. The protons and antiprotons are produced and accelerated in the following manner. Protons from the ionization of hydrogen are accelerated in stages in a Cockroft-Walton accelerator (to 0.75 MeV), a linear

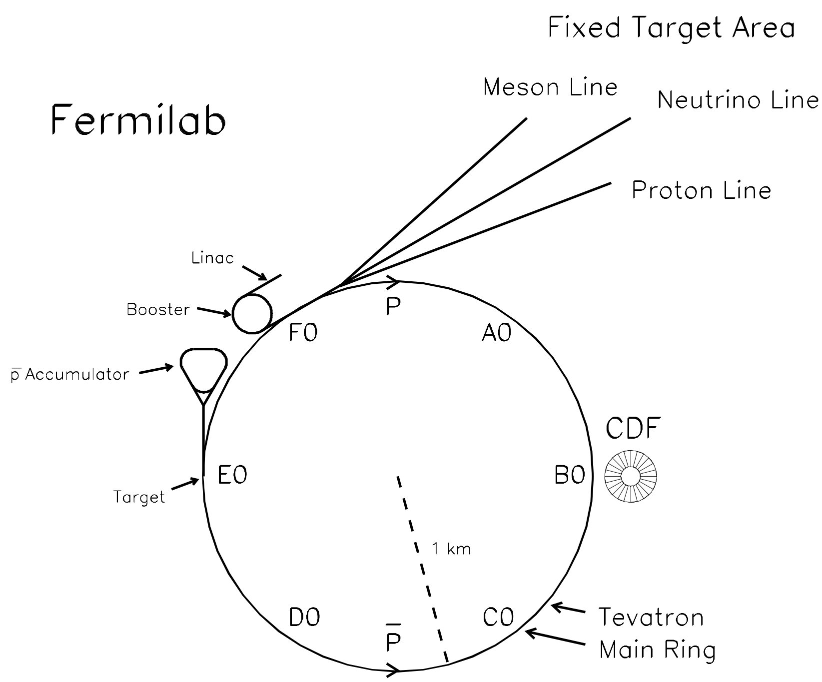


Figure 2.1: *Diagram of the Tevatron at the Fermi National Accelerator Laboratory. Elements are described in the text and are not necessarily drawn to scale.*

accelerator (to 200 MeV), a booster ring synchrotron (to 8 GeV), the Main Ring synchrotron (to 150 GeV), and finally the Tevatron (to 900 GeV). Antiprotons are created by extracting protons from the Main Ring and colliding them with a tungsten target, and the antiprotons are stored in a ring of magnets called the accumulator. Antiprotons are transferred from the accumulator to the Main Ring and then into the Tevatron for collisions. The injection of protons and antiprotons into the Tevatron is called a *shot*, and the following period of colliding beams is called a *store*, which lasts on the order of one day. Typical luminosity delivered to CDF during the course of data collection for this analysis was $5\times10^{30}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. As delivered to CDF, $p\overline{p}$ interactions occur within a gaussian region with transverse dimensions $\pm35\mu$m and longitudinal dimension ±30cm.

2.2 The Collider Detector at Fermilab

The Collider Detector at Fermilab (CDF), located at interaction region B0 in figure 2.1, is a particle detector designed to study a broad range of physics at the Tevatron. CDF is composed of many detector components, each designed for the detection of specific physics objects in specific spatial regions with respect to the nominal $p\overline{p}$ interaction point. This nominal interaction point defines the origin of the right-handed CDF coordinate system, with the positive $\hat{\mathbf{z}}$ direction specified by the proton beam direction, the positive $\hat{\mathbf{x}}$ direction defined by the vector from the center of the Tevatron to the CDF origin, and the positive $\hat{\mathbf{y}}$ direction being vertical. Coordinates may also be described by standard polar coordinates r (2d radius), θ (polar angle), and ϕ (azimuthal angle). Alternately, polar angle may be specified using *pseudorapidity*, defined as $\eta = -\ln(\tan\frac{\theta}{2})$. Quantities of interest are often defined in the plane transverse to the beam direction: transverse energy E_T and transverse momentum p_T are defined from energy E and momentum p as $E_T = E\sin\theta$ and $p_T = p\sin\theta$. Particle detection at CDF is approximately azimuthally and forward-backward symmetric: we may describe the general detector component arrangement in terms of r and η. Overall layout of these components is shown in figures 2.2 and 2.3. Detectors close to the interaction point in both r and η are designed to identify charged particle trajectories or *tracks* for the precise spatial determination of interaction and decay vertices. These include the Silicon Vertex Detector (SVX), which provides tracking information in the $r-\phi$ plane for $|z| < 25$cm and the Vertex Drift Chamber (VTX), which provides tracking information in the $r-z$ plane. Beyond these vertex detectors are tracking chambers in the central region, $|\eta| < 1$. For this analysis we use information from the Central Tracking Chamber (CTC), which occupies the spatial region to $r = 1.4$m. The CTC is surrounded by a superconducting solenoidal magnet of length 5m, radius 1.5m, and magnetic field $\vec{\mathbf{B}} = -1.4\mathrm{T}\hat{\mathbf{z}}$. The curvature of charged particles in this magnetic field provides for momentum determination by the CTC. Beyond the solenoid, at radii greater then 1.5m and $|z|$ coordinates greater than 2.5m in the central and forward regions, electromagnetic and hadronic calorimeters segmented in the $\eta-\phi$ plane and arranged in a projective tower geometry with respect to the interaction

Figure 2.2: *Three dimensional view of the CDF detector.*

Figure 2.3: *Two dimensional single quadrant view of the CDF detector in the* $y - z$ *plane. The* y *direction is vertical and the* z *direction is horizontal.*

point provide for energy determination and the identification of electrons, photons, and jets in the range $|\eta| < 4.2$. These are the Central, Plug, and Forward Electromagnetic Calorimeters (CEM, PEM, and FEM) and the Central, Wall, Plug, and Forward Hadronic Calorimeters (CHA, WHA, PHA, and FHA). Particles emanating from the origin which pass through the electromagnetic and hadronic calorimeters encounter drift chambers for muon identification and momentum measurement in the region $|\eta| < 1.0$. The muon chamber azimuthal coverage is η-dependent. Of interest to this analysis are the Central Muon, Central Muon Upgrade, and Central Muon Extension Chambers (CMU, CMP, and CMX).

The detector as constructed for data collection before 1992 has been described in detail in [30]. Data collected between August, 1992 and June, 1993 is described as *Run 1A* data; data collected between January, 1994 and February, 1995 comprises the *Run 1B* data set. The Run 1A (1B) dataset comprises 19pb^{-1} (48pb^{-1}) of the total 67pb^{-1} collected. The upgrades for the 1992–1995 data runs germane to this analysis are the Run 1A Silicon Vertex Detector (SVX) [31] and its replacement for Run 1B, the SVX′ [32]; upgraded muon chambers, the CMP and the CMX [33]; and the VTX.

We describe below in more detail those components used in this analysis. Specifically, we are interested in those components which are necessary to the collection and reconstruction of lepton+multijet events with evidence of a high energy neutrino and which are necessary to displaced vertex b-tagging and jet kinematic reconstruction within those events.

2.2.1 Tracking

Tracking refers to the determination of particle trajectories, or tracks, from the ionization paths of charge particles traversing the CTC, VTX, and SVX. Within our analysis, tracks will be used for a number of purposes, including the identification of and triggering on primary leptons (both muons and electrons), b-tagging in jets, Z boson removal, and removal of conversion electrons. Charged particle trajectories in the solenoid's uniform magnetic field are helices with axes parallel to the $\hat{\mathbf{z}}$ axis. These helices are mathematically described by five parameters and their correlated errors. These five tracking parameters are defined at the point of closest approach to the origin: (1) half-curvature C; (2) signed impact parameter d; (3) azimuthal angle ϕ; (4) coordinate z; (5) inclination with respect to the beam $\cot\theta$. The first three parameters constitute 2d information in the $x-y$ plane as shown in figure 2.4 and the final two comprise information in the third dimension. Transverse momentum p_T is related to the curvature C and the magnetic field strength B by $p_T = \frac{0.300B}{2C}$, where B is in Tesla, C is in m^{-1}, and p_T is in GeV/c.

Central Tracking Chamber

The Central Tracking Chamber or CTC is a cylindrical drift chamber of outer radius 1.38m, inner radius 0.28m and length 2m, with tracking coverage to $|\eta| < 1.0$ for the outer layer and $|\eta| < 2.0$ for the inner layer of the chamber. The CTC provides 2π azimuthal coverage. Sense and electric field wires within the chamber

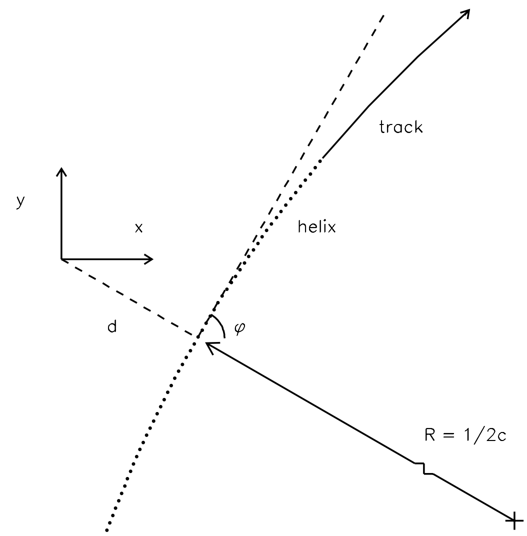


Figure 2.4: *A track in the plane transverse to the beam and its 2d tracking parameters: impact parameter d, azimuth ϕ, and half-curvature C. The cross indicates the helix axis.*

are grouped into cells, and these cells are grouped into nine superlayers as shown in the $r - \phi$ view of figure 2.5. Cells are tilted 45° with respect to the radial direction.

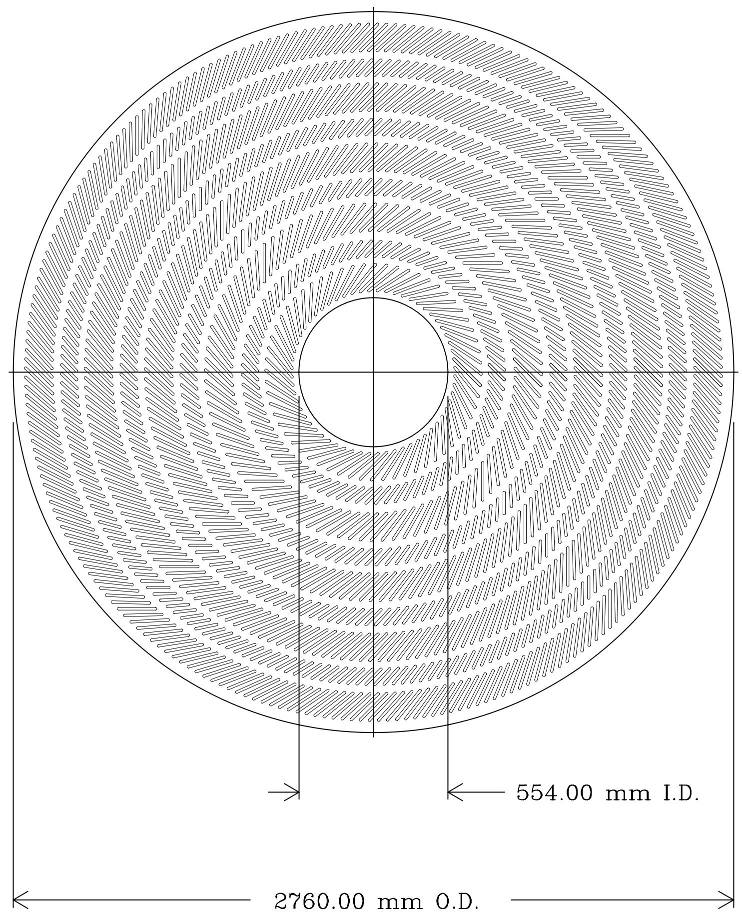


Figure 2.5: *The CTC in the $r - \phi$ view. Individual sense wires are not shown.*

This geometry compensates for the $\mathbf{E} \times \mathbf{B}$ component of the electron drift direction, and results in approximately azimuthal drift trajectories. The 45° cell tilt also maximizes cell overlap for triggering on high p_T tracks and resolves left-right ambiguity in track reconstruction. Five *axial* superlayers composed of 12-wire cells are parallel to the beam, and the remaining four interleaved 6-wire *stereo* superlayers are tilted with respect to the beam to provide tracking information in z in addition to $r-\phi$. Detector information from the CTC consists of drift times for individual wires within a cell. Drift times are converted to distances, which are grouped together using a pattern recognition algorithm. Within a single cell, a high p_T track produces drift distances consistent with two different particle trajectories; this is known as left-right ambiguity. However, the incorrect segment does not match segments in adjacent cells, and can be excluded. A helix is fitted to hit distances and their errors. The resolution of CTC tracks is as follows. The CTC transverse momentum resolution without constraining the track to the beam position is $\delta(p_T)/p_T = 0.0017 p_T$, where p_T has units of GeV/c. The $r - \phi$ resolution may be enhanced by linking the CTC track to SVX hits. In this analysis, we are particularly interested in the impact parameter resolution σ_d, which is primarily driven by the SVX because of its proximity to the beam and its fine resolution. However, CTC impact parameter resolution defines the search road for SVX track reconstruction and contributes to the overall CTC-SVX resolution. CTC impact parameter resolution is of the same order as CTC sense wire residuals, $\simeq 200\mu$m. Because of the the stereo layer tilt, a 2d projection of stereo segments appear shifted in ϕ with respect to axial segments, with the magnitude of the shift dependent on the z position of the trajectory. This z-dependence allows the helix fit to determine track z coordinate to $\pm$6mm.

Vertex Drift Chamber

The Vertex Drift Chamber (VTX) lies inside the CTC and provides $r-z$ information within $r < 22$cm and $|\eta| < 3.3$. It consists of octagonal time projection chamber modules with radial sense wires. It is similar in design to the Vertex Time Projection Chamber [34] which it replaced, except that it has shorter drift regions (56 rather than 16) and higher drift fields to allow for operation at higher luminosity. Tracking information is available in the $r-z$ plane for each of the eight azimuthal sections. Tracks in this $r-z$ projection appear as straight lines, from which a fit is performed. VTX tracks are combined in a simultaneous fit to determine the $p\overline{p}$ interaction point or *primary vertex z* coordinate to within ± 1mm. VTX information is also used in the rejection of electrons resulting from photon conversions, as described in section 3.2.

Silicon Vertex Detector

The Silicon Vertex Detector (SVX) and its Run 1B replacement, the SVX′ [1] are four layer dodecagonal silicon microstrip detectors located in the region between the beryllium beampipe at $r = 1.9$cm and the VTX. The layers have radii r=3.0cm, 4.2cm, 6.8cm and 7.9cm. There are three primary differences between the SVX and the SVX′ : (1) the SVX′ is radiation-hard to survive the larger integrated luminosity of Run 1B; (2) the SVX′ is ac-coupled, while the SVX is dc-coupled; (3) the SVX′ has improved signal to noise ratio compared to the SVX. There are two barrels, one on either side of $z = 0$ and spanning the region $|z| < 25.5$cm ($|\eta| < 2.0$). Given the

[1] Henceforth the term *SVX* will refer to both detectors unless otherwise noted.

longitudinal beam spread of ± 30cm, the SVX $p\bar{p}$ interaction acceptance is 60%. Figure 2.6 illustrates the SVX geometry. The pitch of the microstrips is 60μm for the three inner layers and 55μm for the outermost layer, providing excellent track resolution in the $r - \phi$ plane.

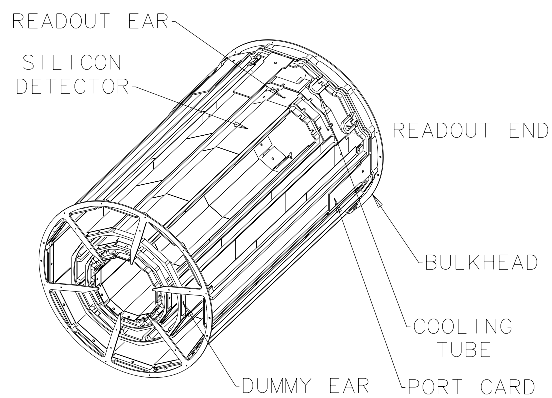


Figure 2.6: *Isometric view of the SVX.*

Particles passing within the fiducial volume of the SVX deposit ionization charge in up to four SVX layers. This charge deposition is Landau-distributed. The path-length through a layer depends on incident angle, and the particle may pass through more than one strip. Adjacent strips with significant charge deposition are grouped into *clusters*, from which a hit centroid may be calculated. The resolution of this centroid is a function of the number of strips in the cluster; cluster resolutions are listed below [35] [36].

Number of Strips	SVX Cluster Resolution	
	Run 1A	Run 1B
1	15μm	13μm
2	13μm	11μm
3	25μm	19μm

SVX tracking in this analysis requires the presence of a CTC track. A CTC track is extrapolated to the SVX, where track parameter errors and transverse uncertainty due to multiple scattering are used to define a 4σ "road" in ϕ in which to search for associated SVX hits. Within this road, there may be several hits on each layer; an ensemble of possible 4-, 3-, and 2-hit tracks is formed by selecting one or zero hits from each layer. An algorithm selects the best track according to the χ^2 of the fitted track, with preference for including more clusters. The combination of CTC and SVX tracking information yields a momentum resolution of $\delta(p_T)/p_T = [(0.0009p_T)^2 + (0.0066)^2]^{1/2}$ where p_T is in units of GeV/c. The $r - \phi$ resolution of SVX tracking yields impact parameter resolution $\sigma_d \simeq 10 - 35\mu$m and is discussed in more detail in chapter 4.

2.2.2 Calorimetry

Electromagnetic and hadronic calorimeters surround the tracking chambers in the central, endplug, and forward η regions. These calorimeters are segmented in η and ϕ within the range $|\eta| < 4.2$ and $0 < \phi < 2\pi$ as shown in figure 2.7. The calorimeters are arranged in a projective geometry in which $\eta - \phi$ segments or *towers* point back to the nominal interaction point. Electromagnetic calorimeters precede hadronic calorimeters with respect to the interaction point. The central region

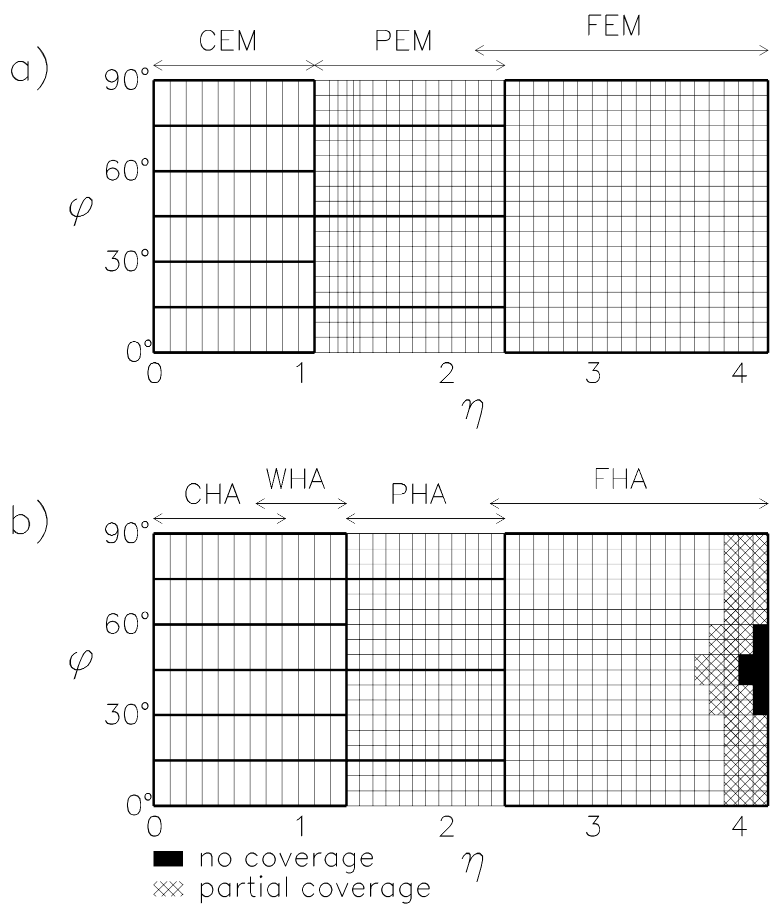


Figure 2.7: *Electromagnetic (a) and hadronic (b) calorimeter segmentation in η and ϕ.*

Component	η Range	$\Delta\eta \times \Delta\phi$	Energy Resolution
CEM	$\|\eta\| < 1.1$	$0.1 \times 15°$	$13.7\%/\sqrt{E_T} \oplus 2\%$
PEM	$1.1 < \|\eta\| < 2.4$	$0.09 \times 5°$	$22\%/\sqrt{E} \oplus 2\%$
FEM	$2.2 < \|\eta\| < 4.2$	$0.1 \times 5°$	$26\%/\sqrt{E} \oplus 2\%$
CHA	$\|\eta\| < 0.9$	$0.1 \times 15°$	$50\%/\sqrt{E_T} \oplus 3\%$
WHA	$0.7 < \|\eta\| < 1.3$	$0.1 \times 15°$	$75\%/\sqrt{E} \oplus 4\%$
PHA	$1.3 < \|\eta\| < 2.4$	$0.09 \times 5°$	$106\%/\sqrt{E} \oplus 6\%$
FHA	$2.4 < \|\eta\| < 4.2$	$0.1 \times 5°$	$137\%/\sqrt{E} \oplus 3\%$

Table 2.1: *Summary of CDF calorimeters. Tower segmentation is listed under the column $\Delta\eta \times \Delta\phi$. The symbol $\oplus$ signifies that the constant term is added in quadrature in the resolution. Energy resolutions for the electromagnetic calorimeters are for incident electrons and photons, and for the hadronic calorimeters are for incident isolated pions. Energy is given in GeV.*

calorimeters are scintillator/absorber sandwiches, while the endplug and forward region active material is gas for greater durability in these high-radiation locations. The electromagnetic absorber is lead and the hadronic absorber is iron. Electromagnetic calorimeter thicknesses are 18-25 radiation lengths and hadronic calorimeters thicknesses are 4.5-7.7 nuclear interaction lengths. Calorimeter segmentation and resolutions are summarized in table 2.1.

Central Electromagnetic Calorimeter

We focus on the Central Electromagnetic Calorimeter, or CEM, because of its role in identifying primary electrons in this analysis. The CEM, along with the Central Hadron Calorimeter (CHA) and Central Muon chambers (CMU) are contained in 48 wedge-shaped modules which surround the solenoid in $\Delta\phi$ =15° segments. Figure 2.8 shows an example of one such wedge, emphasizing its CEM elements.

Local coordinates z (global z) and x (global $r\hat{\phi}$) are as shown in figure 2.8. The

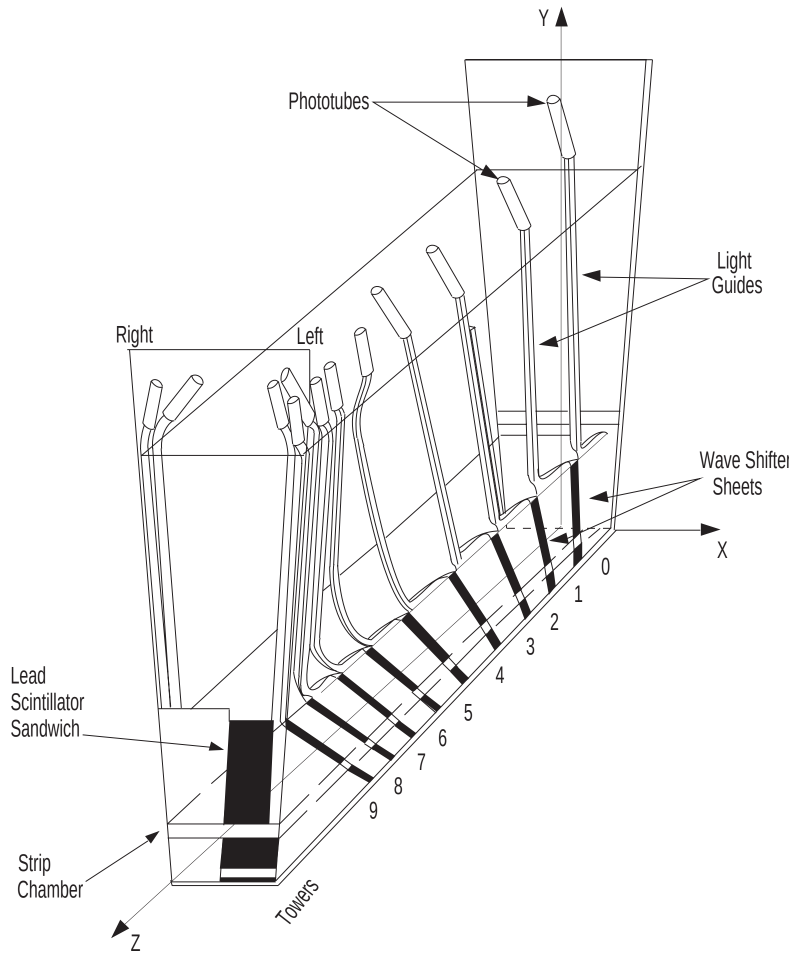


Figure 2.8: *One Central calorimeter wedge. The CEM system including light collection apparatus is shown in detail. The wedge also contains the CHA and CMU systems. Coordinates x,y,and z are local.*

CEM active volume consists of 31 layers of polystyrene scintillator and 30 layers of lead. Radiation length thickness is made uniform through the substitution of acrylic in place of lead in the high-η towers. Between the eighth and ninth layers, at the average shower maximum depth, is a proportional strip chamber, the Central Electromagnetic Strip chamber or CES, which obtains shower profiles in local

z and x for centroid and shape determination. The CES cathode strips run in the x (ϕ) direction and anode wires run in the z direction for profiles in the z and x (ϕ) directions respectively. The centroid determination is accurate to ± 2.5mm for 25 GeV electrons. Light from the CEM scintillators is collected via wave shifters at the wedge ϕ boundaries. Light pipes conduct the light to the outer edge of the wedge. The collected light creates photoelectrons in ten-stage photomultiplier tubes, of which there are two per tower, which are amplified to form a charge signal for fast trigger information and for integration in the front-end CDF electronics.

The initial energy response calibration of the CEM was performed for each module in a test beam of 50 GeV electrons. *In situ* at the B0 interaction region, there are three hardware gain calibration systems for the CEM. The primary of these is a calibration using the change in phototube dark current induced by 3 mCi ^{137}Cs sources. The source calibration differs from the 50 GeV electron calibration in that it depends on phototube current response rather than charge integration and samples only a few of the scintillator layers near shower maximum rather than the full depth of the detector. The two calibrations agree to within $\pm \simeq 1\%$. A set of ^{137}Cs calibrations performed simultaneously with the initial test beam calibration allows the detector response to be absolutely calibrated using later ^{137}Cs calibrations. Source calibrations were performed at the beginning of Run 1A and the beginning of Run 1B and used to re-scale the tower-by-tower energy response as it is read out of the detector. Additional source calibrations were performed several times during the data collection period for the purpose of monitoring CEM response. Two other calibration systems are used on a daily basis: the xenon and LED flasher systems. The xenon flash system injects the light from xenon-filled bulbs into the CEM waveshifters. The LED system illumines a transition guide immediately prior to the phototubes. Since they do not sample the scintillator response directly and because they are less reproducible than the source calibration, the flasher systems are used to monitor gain changes and identify gross problems on a short time scale. A fourth calibration of the CEM using collider data is performed offline by comparing the quantity E/p for electron candidates in the W sample and inclusive electron sample. In E/p, the quantity E is calorimeter energy and p is electron track momentum as measured in the CTC. The advantage of the E/p calibration is that it naturally integrates calorimeter variations for the entire dataset; electron energies in this analysis have been adjusted according to the corrections derived in the E/p analysis.

2.2.3 Muon System

In this analysis, we will use muon information from three muon chambers: the Central Muon chambers (CMU), the Central Muon Upgrade chambers (CMP), and the Central Muon Extension chambers (CMX). All three chambers are four-layer drift chambers, from which spatial and angular trajectory information may be extracted for comparison with matching track information from the CTC. The primary momentum measurement for muons is taken from the CTC. The CMU chambers are installed in the central wedges behind the CHA (see figure 2.9), with the 4.7 absorption lengths (λ_0) of steel in the CHA acting as a hadron absorber. The CMU covers 84% of the solid angle to $|\eta| < 0.6$. The muon coverage was extended prior to Run 1A with the installation of the CMP and the CMX. The CMP covers the same η region as the CMU, but is located behind 0.6m (3.5 λ_0) of additional steel for greater hadron punch-through rejection. The CMP covers 63% of the solid angle to $|\eta| < 0.6$, and the CMU-CMP coverage is 53%. Triggering on muons in the CMU and CMP is described in section 2.2.4. The CMX drift chambers are arranged as four conic frustum segments suspended from freestanding support structures. Scintillators on both faces of the CMX drift chambers provide trigger information. The CMX extends muon coverage to 71% of the solid angle in the region $0.6 < |\eta| < 1.0$, with hadron absorption of length $4.5\lambda_0$ provided by the central and endwall hadron calorimeter steel. The location of the muon chambers in relation to the rest of the detector is shown in figures 2.2 and 2.3. The $\eta - \phi$ coverage map of the muon system is illustrated in figure 2.10. Hits in the muon chambers are fit to line segments to form muon *stubs*. In the $r - \phi$ plane, this provides position resolution of $\sigma_x = .25$mm and segment slope information for matching to CTC tracks. Position in z is determined to $\sigma_z = 1.2$mm using charge division.

2.2.4 Trigger

A highly selective trigger system is necessary to reduce to a manageable level ($\sim$ 5Hz) the rate of events processed and recorded by the online data acquisi-

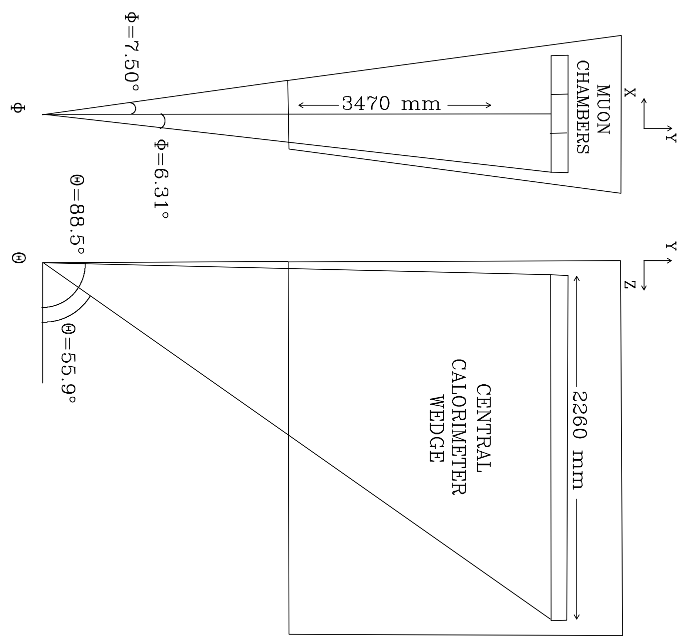


Figure 2.9: *Location of the CMU chambers in a central wedge.*

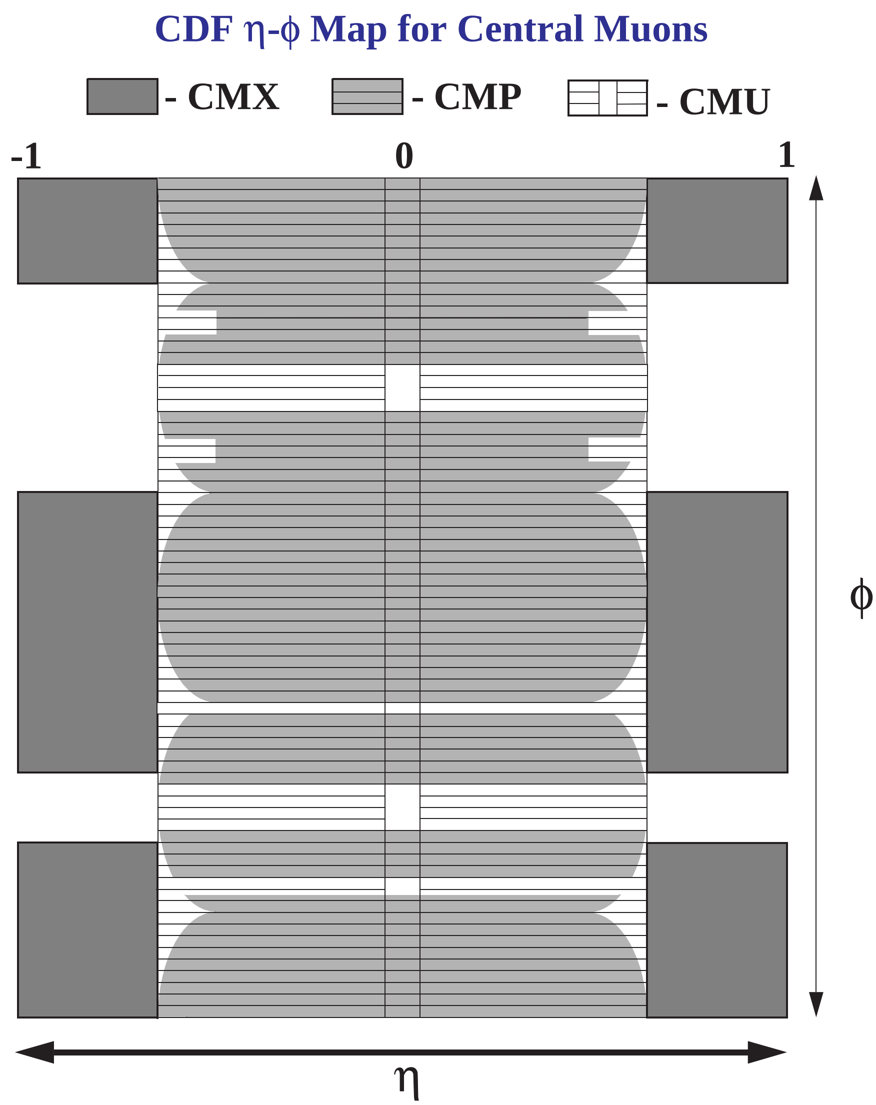


Figure 2.10: *Coverage of the three muon chambers used in this analysis. These are the CMU, the CMP, and the CMX.*

tion system while maintaining high efficiency for interesting physics. Scintillator planes called Beam-Beam Counters (BBCs) in the far forward and backward ($3.2 < |\eta| < 5.9$) regions provide a minimal trigger requirement for $p\overline{p}$ collisions. Coincidental signals from both sets of BBCs are required to fall within a 15ns window around the Tevatron bunch crossing signal. Such events constitute the Level 0 CDF trigger and occur with a cross section of 51.2±1.7 mb [37]. Such events are called *minimum bias* events. At an instantaneous luminosity of $5 \times 10^{30}\text{cm}^{-2}\text{s}^{-1}$, the Level 0 rate is 0.25 MHz, necessitating a trigger with a rejection factor of $\sim 5 \times 10^4$. This is accomplished using a three-stage trigger system referred to as Levels 1, 2, and 3.

Level 1 Trigger System

The Level 1 system is a FASTBUS-implemented hardware trigger which must form a decision during the 3.5μs between bunch crossings and must yield a sufficiently low output data rate that the higher level trigger systems may operate with somewhat longer cycles without introducing significant deadtime. Level 1 primarily selects events with significant calorimeter energy patterns for the electron and jet triggers, and also events with high-p_T muons in the central region. Since digitization of the analog calorimeter signals takes $\sim$ 1ms, the Level 1 calorimeter decision must be made directly from analog signals. For this reason, analog calorimeter signals are split into two paths; one path for digitization in the frontend crates on the detector and one path through differential cables directly to the Level 1 trigger modules.

The latter signals are called the "fast outs," and are installed for the entire EM-hadron calorimetry system. Fast out signals are summed into $0.2 \times 15^\circ$ $\eta - \phi$ trigger towers. Pedestal, gain and $\sin\theta$ adjustments standardize the signals and convert E to E_T for comparison with detector-specific analog thresholds. The analog sums $\sum E_T$, $\sum E_T \sin\phi$, and $\sum E_T \cos\phi$ for towers above threshold are formed for EM and hadronic calorimetry separately in five η regions corresponding to the $\pm\eta$ endplug and forward regions (4 regions), and to the entire central region (1 region). Four sets of thresholds may then be applied to each region for various trigger configurations. One simple trigger is the requirement of a single tower over threshold, i.e. the sum for the whole detector exceeds the single tower threshold. Level 1 muon triggers in the central region select high-p_T muons via a cut on the angle at which a track stub traverses the four drift chamber layers. A higher p_T muon is deflected less within the solenoid and hence has a smaller angle with respect to normal incidence when it reaches the muon chambers. This angle is calculated from the drift time differences between the layers. Another component of the Level 1 trigger includes CTC tracking information from the Central Fast Tracker (CFT) track processor, which identifies high-p_T CTC tracks from prompt ($t < 80$ns, compared to a maximum of $t = 800$ns) hit information. The CFT momentum resolution is $\delta p_T/p_T \sim 0.035 \times p_T$. The CFT is highly efficient for tracks with $p_T > 3$ GeV/c. In addition to calorimetry, muon and CFT pathways, Level 1 possesses pathways for other triggers, such as the selection of minimum bias events. The final Level 1 trigger decision is stored in a lookup table based on the decisions from the calorimetry overall sum logic, the muon triggers, track processor results and the additional trigger paths. Operation of several lookup tables in parallel provides for the option of *prescaling* high rate triggers, i.e. accepting only a fraction of such triggers. The Level 1 accept rate for an instantaneous luminosity of $5 \times 10^{30}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ is $\sim$ 1kHz.

Level 2 Trigger System

The Level 2 FASTBUS-based trigger system implements more sophisticated algorithms than Level 1. These algorithms are designed to identify potential physics objects of interest. A nearest-neighbor algorithm finds calorimeter clusters, their energies and $\eta - \phi$ moments from analog tower energies. Tracks from the CFT are matched to calorimeter clusters and also to muon candidates. Dedicated processor cards search for electrons, muons, and other interesting physics signatures. The muon card calculates a value of transverse vectoral energy imbalance, $\not{E}_t$. $\not{E}_t$ is corrected for muon candidates, whose full energies are not recorded in the EM-hadron calorimeter energy sum since muons are minimum ionizing particles. The final Level 2 accept/reject decision is made in a programmable module for flexibility. Prescaling in Level 2 allows for *dynamic* prescaling, in which high rate triggers are prescaled more at higher instantaneous luminosity than at low luminosity. The total Level 2 accept rate at $5{\times}10^{30}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ is $\sim$ 12Hz.

Level 3 Trigger System

The Level 3 trigger system is a software-based selection which runs on a farm of Silicon Graphics nodes. Level 3 executes an abbreviated version of offline event reconstruction (see section 2.2.6), including calorimeter clustering and tracking. At this level, the system is highly configurable. The Level 3 accept rate is ~5 Hz, and the resulting data stream is stored on 8mm tape.

2.2.5 Online Data Acquisition

The CDF data acquisition system (DAQ) is controlled through programs running on a microVAX cluster interfaced to a FASTBUS network. Digitization of tracking system detector signals is performed within the FASTBUS network. Programmable FASTBUS scanner modules (SSPs) control the readout and process the data. For calorimeter and muon data, digitization of detector signals is performed prior to entering the FASTBUS network in the frontend RABBIT [38] crates mounted on the detector. Programmable RABBIT scanners (MXs) are used to process and control the readout of calorimeter and muon data. The MXs pass their data to the FASTBUS network. Digitization of analog signals occurs after a Level 2 accept. From the scanners, the data for an entire event is assembled and reformatted for transfer to Level 3 and the VAX host. Events are written to 8mm tape for offline processing.

2.2.6 Offline Production

Offline production is a program which transforms large raw detector-level data structures with such information as hit times and individual cell energies into smaller analysis-level objects such as tracks, energy clusters, and physics objects like electrons, muons, and jets. The Run 1A offline production was run on a farm of eleven IBM RS-6000 workstations and processed 16 million events. Production for Run 1B was executed on two clusters, a Silicon Graphics farm with 64 nodes and an IBM RS-6000 farm with 37 nodes. For the 48pb^{-1} of the Run 1B data sample used in this analysis, production processed $\sim$23 million events. To expedite the top analysis, offline production is also run separately on a select data stream involving approximately one-tenth as many events. This analysis is performed using these latter samples.

2.2.7 Monte Carlo Programs

We use several Monte Carlo programs to generate event samples which will be used in the subsequent analysis for the evaluation of signal acceptances and for elements used in calculation of backgrounds. For the $t\bar{t}$ signal, we use three main Monte Carlo programs: PYTHIA, HERWIG, and ISAJET. The PYTHIA generator [39] is the primary Monte Carlo program from which we calculate $t\bar{t}$ acceptances, both for the lepton+multijet selection of section 3.11 and for the b-tag efficiency of section 4.2.1. We use the HERWIG [40] Monte Carlo program to describe $t\bar{t}$ kinematic distributions in chapter 5. The results obtained using PYTHIA and HERWIG were verified through comparison to one another and to the results obtained using a third Monte Carlo program, ISAJET [41]. All three Monte Carlo programs are based on the leading order QCD matrix elements for the hard-scattering sub-process. In PYTHIA, coherent parton shower evolution and string hadronization follow the hard scattering sub-process and an underlying event based on data. HERWIG uses coherent parton shower evolution and cluster hadronization. ISAJET is based on incoherent gluon emission and independent fragmentation of the outgoing partons. In chapter 4, HERWIG and PYTHIA are both used in the calculation of backgrounds to the b-tagging search. In chapters 4 and 5, we make use of the VECBOS Monte Carlo[42] in modeling W+jet events. VECBOS is based on tree-level matrix element calculations for the specified final state parton multiplicity. Since VECBOS is a parton-level Monte Carlo program, parton evolution was performed on these samples using the HERWIG program. The decay of b quarks in samples used for the b-tag analysis was modeled using the CLEO Monte Carlo program[43]. CDF detector response to the final state particles for each Monte Carlo sample is simulated. Physics objects are then reconstructed from these simulated samples using the offline reconstruction code. This allows us to use the same analysis selections on Monte Carlo events as we use on data events.

Chapter 3

Lepton+Jets Sample Event Selection

We describe in this chapter the criteria for electron, muon, and jet identification and energy measurement. We also describe the measurement of missing transverse energy, $\not\!\!E_t$. The goal of these criteria is the selection of a sample of events predominantly resulting from W production, including direct production and the production of W's from top quark decays. Such events are selected based on the decay signature $W \rightarrow e\nu$ or $W \rightarrow \mu\nu$ by requiring the presence of either an electron or a muon and evidence of a neutrino. In the following sections we describe these and several other event selection criteria:

1. a good, isolated lepton with large energy transverse to the $p\overline{p}$ beam; we use $E_T > 20$ GeV for electrons and $p_T > 20$ GeV/c for muons;

2. appropriate trigger path;

3. evidence of a high energy neutrino; we require $\not\!\!E_t > 20$ GeV;

4. absence of a Z boson candidate;

5. good data run quality;

6. absence of a $t\bar{t}$ dilepton candidate.

We refer to the sample obtained using these selections as the *high-p_T lepton* sample. This sample includes events resulting from electroweak processes such as $qq' \rightarrow W \rightarrow l\nu X$ and also events resulting from $t\bar{t}$ production. Other sources of events in the sample are described in section 3.10. Within the high-p_T lepton sample, we count the number of jets above a threshold transverse energy E_T. This number is referred to as *jet multiplicity.* The discussion of $t\bar{t}$ decay signatures in section 1.4.1 indicates that we expect $t\bar{t}$ events to populate the $\geq$ 3-jet multiplicity region. We refer to the $\geq$ 3-jet events within the high-p_T lepton sample as the *lepton+multijet* or *lepton+jets* sample. Events resulting from QCD production of jets in association with electroweak production of W bosons (see figure 1.6) also populate the W+multijet sample. We refer to these background processes as *QCD W+jet* events.

Event selections differ slightly for the Run 1A and Run 1B datasets in order to maintain consistency with the Run 1A analysis presented in [13] while allowing for improvements in the Run 1B selections. Such differences are noted below as appropriate. We reiterate that the Run 1A dataset comprises 19pb^{-1} of the total integrated luminosity of 67pb^{-1}.

3.1 Triggers

We make use of electron, muon and $\not\!\!E_t$ triggers. We begin with a description of the electron trigger. The Level 1 electron trigger requires a single trigger tower with E_T

	Level 0 Threshold (GeV)	
detector	Run 1A	Run 1B
CEM	5	8
PEM	8	11
FEM	8	51
CHA	8	12
PHA	25	51
FHA	25	51

Table 3.1: *Level 0 trigger threshold requirements.*

above the detector-dependent values listed in table 3.1. The Level 2 central electron trigger for Run 1A (1B) required that a CEM energy cluster with $E_T > 9$ GeV (16 GeV) be matched to a CFT track with $p_T > 9.2$ GeV/c (12 GeV/c), where a cluster consists of at least one seed trigger tower with $E_T > 9$ GeV (8 GeV) and adjacent contiguous towers with $E_T > 7$ GeV (7 GeV). The hadronic energy of this cluster is required to be less than 12.5% of the electromagnetic energy. Level 3 requires that central electron clusters have $E_T > 18$ GeV in association to a CTC track with $p_T > 13$ GeV/c. Here clusters and CTC tracks are reconstructed using the offline algorithms. For Run 1A data, the electron trigger efficiency is $\epsilon^{1A}_{e\text{trig}} = (92.8 \pm 0.2)\%$ for electrons with $E_T > 20$ GeV. For the Run 1B electron trigger efficiency, we combine lepton triggers with $\not\!\!E_t$ triggers, as described below.

For clarity at the price of redundancy, we describe the muon triggers for Run 1A and Run 1B separately. The muon trigger for Run 1A is as follows. The Level 1 muon trigger requires a 2-hit CMU (CMX) stub with $p_T > 6$ GeV/c (10 GeV/c) as calculated from muon chamber sense wire drift time differences. The CMU trigger requires confirmatory hits in the CMP, and the CMX requires confirmatory hits in the CMX scintillator pads. Time coincidence with a signal in the hadron calorimeter is required for the CMX. Since this CMX Level 1 trigger was operational for only 30% of the run, an additional trigger requiring a single calorimeter tower in Level 1 and a Level 2 CMX trigger was employed for 83% of the run. The Level 2 muon trigger requires a Level 1 CMU or CMX stub matched to a CFT track with $p_T > 9.2$ GeV/c. Level 3 requires a reconstructed CTC track of $p_T > 18$ GeV/c matched to a muon chamber track segment. The extrapolated CTC track must fall within 5 (10) cm of the CMU (CMX) stub at the muon chamber radius, and the corresponding hadron calorimeter tower must have $E_T < 6$ GeV. The CMU [CMX] trigger is measured to be (86.8±1.9)% [(54.4±5.5)%] efficient for muons with $P_T > 20$ GeV/c. The weighted average of the CMU and CMX trigger efficiencies is $\epsilon^{1A}_{\mu\mathrm{trig}} = (80 \pm 2)\%$.

The muon trigger for Run 1B is as follows. The Level 1 muon trigger requirement is the same as for Run 1A as described above except that both CMU and CMX triggers required coincidence with a hadron calorimeter signal. The Level 1 CMX trigger was fully operational during Run 1B. High efficiency for $t\bar{t}$ events was retained while reducing the overall muon trigger cross section by requiring calorimeter energy in muon triggers. Level 2 trigger paths for CMU and CMX muons required a CFT track with $p_T > 12$ GeV/c associated to the CMU stub. The CMU trigger also required the presence of a calorimeter cluster with $E_T > 15$ GeV anywhere in the event. The CMX trigger required a calorimeter cluster with $E_T > 15$ GeV anywhere in the event or any energy deposition in the calorimeter tower corresponding for the CMX muon. The Level 3 requirement is the same as for Run 1A except that CMP-only muons are permitted with the same conditions as CMX muons. For the Run 1B muon trigger efficiency, we combine lepton triggers with $\not{E}_t$ triggers, as described below.

Triggers for large $\not{E}_t$ are implemented at Level 2. These triggers were included for Run 1B data. The two $\not{E}_t$ trigger requirements at Level 2 were: (1) $\not{E}_t > 20$ GeV plus a CEM cluster with $E_T > 16$ GeV and Central Strip Chamber photon/electron profile; (2) $\not{E}_t > 35$ GeV with one or more calorimeter clusters in the central region. The first $\not{E}_t$ trigger above is accepted for events identified offline as electrons, while the second trigger is accepted for muon events. At Level 3, events with a Level 2 $\not{E}_t$ trigger are accepted through either the electron or muon path. Since Level 3 runs the same code as offline reconstruction, this constitutes no additional requirement beyond the analysis cuts described below.

We require that events with a good electron as described in section 3.2 pass the electron trigger. For Run 1B data the $\not{E}_t > 20$ GeV trigger was also permitted. We require that events with a good muon as described in section 3.3 pass the muon trigger appropriate to the muon detector region. For CMU muons with or without CMP confirmation, we allow CMU or CMP triggers, and for CMX muons, we allow CMX triggers. For Run 1B data events with any type of primary muon, the $\not{E}_t >$ 35 GeV trigger was permitted. For CMP muons without a corresponding CMU stub, only the $\not\!\!E_t > 35$ GeV trigger was accepted.

The inclusion of $\not\!\!E_t$ triggers in the electron and muon selections yields an overall Run 1B trigger efficiency of $\epsilon^{1B}_{e\mathrm{trig}} \sim 100\%$ for electrons from W decay and $\epsilon^{1B}_{\mu\mathrm{trig}} \sim 100\%$ for muons from W decay.

3.2 Electron Selection

Electrons are identified by associating an electromagnetic cluster in the Central Electromagnetic Calorimeter (CEM) to a CTC track. We select electrons in the central ($|\eta| < 1.0$) region only. We require the electron cluster to have large transverse energy, $E_T > 20$ GeV. A calorimeter cluster consists of three adjacent towers ($\Delta\eta \times \Delta\phi = 0.3 \times 15^\circ$) centered on a seed tower with $E_T > 3$ GeV. The shower centroid is measured using the Central Strip chamber (CES). Fiducial cuts require that this centroid be at least 2.5 cm away from the module ϕ boundaries, and at least 9 cm away from the $\theta = 90^\circ$ boundary. Electrons in the highest-$|\eta|$ tower, which spans $1.0 < |\eta| < 1.1$, are excluded because of this tower's geometrical irregularity. For comparison, at the depth of the strip chambers, individual towers have dimensions $\Delta x \times \Delta z \simeq 48\mathrm{cm} \times (12 - 18)\mathrm{cm}$. The highest-$p_T$ CTC track which matches the cluster is extrapolated to the calorimeter, where its x and z coordinates (see figure 2.8) must match the strip chamber shower position to within $|\Delta x| < 1.5$cm and $|\Delta z| < 3.0$cm. The Δz cut is looser to account for the larger CTC z uncertainty.

Charged hadrons which deposit energy in the CEM are rejected using several cuts. (1) The radiation length of the CEM is sufficient that very little energy from real electrons will appear in the hadron calorimeter. We require that the hadronic energy be less than 5% of the electromagnetic energy ($E_{had}/E_{em} < 0.05$). (2) For high energy electrons, the calorimeter cluster energy divided by the CTC track momentum (E/p) should be unity within detector resolution and absolute calibration. A tail on the upper edge of E/p results from Bremsstrahlung: photon emission reduces the electron momentum as measured in the tracking volume, while the electron and collimated photon strike the calorimeter together, resulting in a unreduced total energy measurement. We require $E/p < 1.5$ to avoid large track-cluster energy mismatches. (3) A parametrization of strip chamber profiles for electrons was obtained in a test beam. A χ^2 fit to this parametrization is performed using the profile in the z view. We require $\chi^2_{strip} < 10$. (4) The lateral shower profile across tower boundaries is measured using the variable $\mathcal{L}_{shr}$ [44]. $\mathcal{L}_{shr}$ is a measure of the significance of the cluster's non-seed tower energy deposition compared to that expected from the seed-tower energy, shower position and event vertex: hadron showers are typically broader than electron showers. We require $\mathcal{L}_{shr} < 0.2$.

We require that electron candidates be *isolated* from calorimeter energy deposition. One source of electron events which are not a result of W production is $b\bar{b}$ events where one of the b quarks decays semileptonically $b \to l\nu c$. Leptons from these decays are *non-isolated* because of the energy deposited in surrounding calorimeter towers by the daughter c quark. In contrast, leptons from W decay in $t \to Wb$ decays are relatively isolated because their direction is nearly uncorrelated to that of the associated b-quark. We define isolation I as the ratio of calorimeter tower transverse energy in the $\eta - \phi$ cone of $R < 0.4$ centered on the electron candidate to electron energy $E_T(e)$, where $R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$: $I = \sum_{R<0.4} \frac{E_T(i)}{E_T(e)}$. The sum runs over calorimeter towers surrounding the electron and excludes the electron energy. We require $I < 0.1$.

Two event vertex selection criteria are applied to electron events. We select as the event vertex z coordinate z_{vtx} as follows. The set of all high-quality primary vertices formed using VTX tracks (see section 2.2.1) are considered. From this set, we select the vertex with the largest $\sum p_T$ for all associated CTC tracks. We require $|z_{vtx}| < 60$cm. Given the primary vertex position, we require that the electron CTC track z match this vertex position to within 5 cm ($|z_{track} - z_{vtx}| < 5$cm), which is large in comparison to the CTC z resolution of 6 mm. The preceding high-p_T electron selection cuts are summarized in table 3.2. The efficiency of all cuts except for isolation was measured using $Z \to ee$ events and found to be $(84{\pm}2)\%$ for Run 1A data and $(81.3{\pm}0.9)\%$ for Run 1B data.

Photon conversions ($\gamma \to e^+e^-$), either from prompt photons or π^0 decays ($\pi^0 \to \gamma\gamma; \gamma \to e^+e^-$), occur when a photon interacts with matter in the detector before it reaches the EM calorimeter. We eliminate conversion electrons using the logical OR of two methods. (1) We search for a nearby oppositely signed CTC track. In the $r-\phi$ plane, we require the two helices point of two-dimensional closest approach to be $|d_{min}(r-\phi)| < 0.3$cm. The angular separation between the two tracks at the point of

Variable	cut
$\lvert\Delta x\rvert$	< 1.5 cm
$\lvert\Delta z\rvert$	< 3.0 cm
E_{had}/E_{em}	< 0.05
E/p	< 1.5
χ^2_{strip}	< 10
$\mathcal{L}_{shr}$	< 0.2
$\lvert z_{track} - z_{vtx}\rvert$	< 5.0 cm
Isolation I	< 0.1
Efficiency $\epsilon_{e\mathrm{id}}$	$(84 \pm 2)\%$ (Run 1A)
	$(81.3 \pm 0.9)\%$ (Run 1B)

Table 3.2: *Central electron selection requirements and efficiency. The efficiencies are for the combination of all the cuts, except the isolation requirement and does not include conversion removal inefficiency.*

tangency should both be almost zero; in the polar plane, we require $|\Delta(\cot\theta)| < 0.06$ between the two tracks. We take this point of tangency as the radius at which the potential conversion occured, R_c. For Run 1A, we required $R_c < 50$cm and for Run 1B we required -20cm< R_c <50cm. The invariant mass of the e^+e^- pair should be very small; for Run 1A data, we required $M_{e^+e^-} < 500\mathrm{MeV/c}^2$. In Run 1B this requirement was not used. (2) The second method for conversion identification uses the VTX to determine the presence or absence of the electron candidate close to the beam. A photon conversion beyond the VTX leaves no track in the VTX, while a prompt electron does. This is quantified using VTX occupancy f_{VTX}, the number of hits in the VTX divided by the number of hits expected. The Run 1B conversion algorithm considered an electron candidate to be a conversion if $f_{VTX} < 0.2$; for Run 1A this conversion selection was not used. The conversion identification algorithms, true conversion removal efficiencies, and overefficiencies on prompt electrons are summarized in table 3.3.

quanitity	Run 1A	Run 1B
$\|d_{min}(r-\phi)\|$	< 0.3cm	< 0.3cm
$\|\Delta(\cot\theta)\|$	< 0.06	< 0.06
Conv. Radius	$R_c < 50$cm	-20cm$< R_c <$50cm
M_{ee}	< 500MeV	–
		-OR-
VTX occupancy	–	< 0.2
conv. removal eff.	88±4%	88±4%
prompt electron eff., $\epsilon_{\rm conv}^{\rm prompt}$	95±3%	97.8±0.5%

Table 3.3: *Conversion removal algorithm summary. Prompt electron efficiency indicates the percentage of prompt electrons which are* not *identified as conversions.*

3.3 Muon Selection

Muons are selected by matching a CTC track to hits in the muon chambers. Stubs are required to have at least two hits. The primary measurement of muon transverse momentum p_T is from the matched CTC track, rather than from the momentum as measured using the muon chambers. We require the muon candidate track to have large transverse momentum, $p_T > 20$ GeV/c. We consider muon stubs in both the CMU/CMP and the CMX regions, $|\eta| < 1.0$. Muons are required to be separated from the detector boundaries. For the high-p_T muon selection, the matching CTC track is extrapolated to the muon chamber, where its local x (global ϕ, see figure 2.9) coordinate is required to match the position of muon chamber hits. The requirement is $|\Delta x| < 2.0$cm for CMU muons, $|\Delta x| < 5.0$cm for CMP muons, and $|\Delta x| < 5.0$cm for CMX muons.

The following selections reduce background from cosmic rays and hadrons which cause hits in the muon chamber after failing to be absorbed in the calorimeter. (1) A high-p_T muon is a minimum ionizing particle, which is expected to deposit ~ 1.5 GeV in the calorimeters as it passes through the absorbers. In the calorimeter tower corresponding to the muon candidate, we require the electromagnetic energy deposition to be $E_{em} < 2$ GeV and the hadronic energy deposition to be $E_{had} <$ 6 GeV. (2) Since we expect the muon to deposit *some* energy in the calorimeters, we also require $E_{em} + E_{had} > 0.1$ GeV. This requirement was used only for Run 1B data. (3) Cosmic rays which pass through the detector coincident with a beam crossing may produce muon segments along with very stiff CTC tracks. In order to veto most cosmic ray events, we require that the muon track come from the $p\overline{p}$ beam by cutting on its impact parameter, $|d_0| < 3$mm.

We reject non-isolated muon candidates in order to reduce backgrounds from $b\overline{b}$ production. Muon isolation is defined in the same way as electron isolation, but excluding the calorimeter energy due to the muon candidate; we require $I < 0.1$.

We apply the same event vertex requirements as in the case of electron events, $|z_{vtx}| < 60$cm and a match between the muon track z and the event vertex z of $|z_{track} - z_{vtx}| < 5$cm. The preceding high-$p_T$ muon selection cuts are summarized in table 3.4. The efficiency of all cuts except for isolation was measured using $Z \rightarrow \mu\mu$

events and found to be (90.6±1.4)% for Run 1A data and (92.4±1.5)% for Run 1B data.[1]

Variable	cut
η range	$\|\eta\| < 1.0$
$\|\Delta x\|$	< 2 cm (CMU),
	< 5 cm (CMP, CMX)
E_{em}	< 2 GeV
E_{had}	< 6 GeV
$E_{em} + E_{had}$	> 0.1 GeV (1B only)
Impact Parameter d_0	< 3 mm
Isolation I	< 0.1
$\|z_{track} - z_{vtx}\|$	< 5.0 cm
Efficiency $\epsilon_{\mu \mathrm{id}}$	$(90.6 \pm 1.4)\%$ (Run 1A)
	$(92.4 \pm 1.5)\%$ (Run 1B)

Table 3.4: *Selection criteria and efficiencies for high-p_T muons. The efficiencies are for the combination of all the cuts, except the isolation requirement.*

3.4 Jet Identification and Measurement

A fixed-cone calorimeter energy clustering algorithm is applied to the entire calorimeter. This analysis employs a cone of $R = 0.4$ units in $\eta - \phi$ space, where $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The clustering algorithm begins by grouping together pre-clusters of contiguous towers with $E_T > 1$ GeV. All towers with $E_T > 0.1$ GeV within the $R < 0.4$ cone centered on a given precluster are associated together as a cluster. An iteritive procedure of centroid calculation and tower inclusion/exclusion is performed

[1]For Run 1B data, the muon selection efficiency is a weighted average of the individual CMU, CMP, CMU+CMP, and CMX efficiencies, which are (93.1±1.4)%, (90.5±1.4)%, (92.3±1.4)%, and (95.2±1.1)% respectively. The fraction of $W \to \mu\nu$ events of each type are 11.3%, 0.0%, 66.6%, and 22.2% as measured in a Monte Carlo study and confirmed by the data.

until a stable cluster configuration is attained. Overlapping clusters are merged if more than 75% of the smaller cluster E_T falls in the overlap region between the two clusters. Otherwise the clusters are separated and the energy in the overlap region is divided between the two clusters. In this analysis, we will be concerned with calorimetry clusters resulting from hadronic jets. Jet energies may be corrected to reflect more accurately initial parton energy by accounting for several effects including: (1) calorimeter non-linearities; (2) loss of low p_T tracks due to large curvature in the magnetic field; (3) detector boundary effects; (4) energy contribution from the underlying event; (5) out-of-cone energy loss; (6) undetected energy from muons and neutrinos. For the purposes of this analysis, these corrections are not applied. Typically, the corrections result in 30% higher jet energies. Studies of events with a single photon recoiling against a jet indicate that the corrected jet energy scale due to uncertainty in final state QCD radiation is correct to within $\pm 10\%$. Including uncertainty due to detector resolution yields a total jet energy uncertainty given in equation 3.1, where jet E_T is in GeVand $\oplus$ indicates the sum in quadrature.

$$\frac{\Delta(E_T)}{E_T} = (2\% \oplus (2.2\% + 60\%/E_T)) \oplus 10\% \tag{3.1}$$

The uncertainty of the uncorrected jet energies used in this analysis is also described by this formula. Uncorrected jet energy does not correspond directly to parton energy. We may compare directly uncorrected jet energies in data and Monte Carlo samples by dint of having simulated the detector response in the Monte Carlo samples and having reconstructed physics objects using the same offline code in both data and Monte Carlo.

3.5 $\not{E}_t$ Measurement

The presence of a neutrino is signalled by missing transverse calorimeter energy, $\not{E}_t$, which is defined as the 2d vectoral sum over calorimeter towers:

$$\not{E}_t = -\sum_{|\eta|<3.6, E_T>\text{thresh}} E_T\hat{n}$$

The unit vector $\hat{n}$ indicates the direction from the origin to the tower and 'thresh' is a detector-dependent threshold. The thresholds are 100 MeV for the CEM, CHA, and WHA, 300 MeV in the PEM, 500 MeV in the PHA and FEM, and 800 MeV in the FHA. Since total muon energy is not recorded by the calorimeters, we correct $\not{E}_t$ for primary muons using the muon p_T as measured in the CTC, and accounting for the $\not{E}_t$ contribution from the minimum ionizing muon calorimeter energy deposition. $\not{E}_t$ is also corrected for non-primary muon candidates with $p_T > 10$ GeV/c. These candidates are identified using the soft muon tagging algorithm detailed in [13]. The $\not{E}_t$ resolution has been estimated from minimum bias events and is $0.7\sqrt{\sum E_T}$ where $\sum E_T$ is the scalar sum of calorimeter transverse energy within $|\eta| < 2.4$ in units of GeV.

3.6 Z removal

We remove $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ events from our high-p_T electron and muon samples by searching for an oppositely charged additional lepton or "second leg" in the event which reconstructs along with the primary lepton to an invariant mass close to M_Z. In order to remove Z boson events with high efficiency, the additional lepton identification requirements are less stringent than the primary lepton identification requirements. For events with a primary electron (muon), the second leg electron (muon) requirements are listed in table 3.5 (3.6). In the case of muons, a lower-quality category of muons, which we label as CMIO in table 3.6, is also allowed with requirements as indicated. Events with a second leg are rejected if M_{ll}, the invariant mass of the two leptons, is $70 < M_{ll} < 110$ GeV/c^2 (Run 1A) or $75 < M_{ll} < 105$ GeV/c^2 (Run 1B).

Variable	Cut
Detector	CEM or PEM or FEM
E_T	> 10 GeV
E/p	< 2.0 (CEM only)
E_{had}/E_{em}	< 0.12
Isolation, I	< 0.1 (1A) < 0.2 (1B)

Table 3.5: *Z removal requirements for additional electrons.*

3.7 Good Run Requirement

We require that high-p_T lepton events were collected at a time when the data quality was good. A *run* is a period of data collection during which experimental conditions are unchanged. A run may last no longer than a $p\overline{p}$ store. Typical runs last up to 20 hours and involve the collection of up to $\sim 10^6$ events. Electronic inhibit switches prevent the collection of data when any component of the detector is off. Several online data-quality assurance programs called *consumers* detect subtler detector

Variable	Cut	
	Run 1A	Run 1B
Detector	(CMU or CMU/CMP or CMP or CMX)	
P_T	> 15 GeV/c	> 10 GeV/c
E_{had}	-	< 10 GeV
		< 6 GeV (CMIO)
E_{em}	-	< 5 GeV
		< 2 GeV (CMIO)
$\|\Delta x\|$	< 5 cm (CMU)	< 5 cm
$\|\eta\|$	< 1.1	< 1.1 (CMIO)
Isolation, I	< 0.2	< 0.2

Table 3.6: *Z removal requirements for additional muons. CMIO indicates a lower quality muon.*

problems, from which the scientific coordinator and on-duty data acquisition "ace" make a judgement on the overall data quality. This judgement along with subsequent offline validation studies determines whether a run is considered good for all physics analyses, for only a subset of analyses, or for no analyses at all. We require that all events used in this analysis be in the first category.

3.8 Dilepton Removal

We remove by hand the six events which pass the dilepton top search requirements (see [13]) in order to avoid double-counting these events. Dilepton events from $t\bar{t}$ production would appear in our single lepton analysis predominantly as lepton+2-jet events. Since $t\bar{t}$ dilepton event contain b-jets, they could appear as b-tagged lepton+2-jet events in excess of the calculated b-tagged background.

Jet Multiplicity	Electron Events Run 1A	Electron Events Run 1B	Muon Events Run 1A	Muon Events Run 1B	Total
0 Jet	10663	27804	6264	18886	63617
1 Jet	1058	2772	655	2093	6578
2 Jet	191	395	90	350	1026
3 Jet	30	59	13	63	165
$\geq$4 Jet	7	11	2	19	39

Table 3.7: *Jet multiplicity for events passing all requirements.*

3.9 Lepton+Jets Selection and Application to Data

In summary, we select a sample of W and $t\bar{t}$ events by requiring a high energy electron or muon, evidence of a neutrino, appropriate trigger path and good data quality. In addition, we reject events containing either Z boson or $t\bar{t}$ dilepton candidates. We count jets by requiring $E_T > 15$ GeV and $|\eta| < 2.0$. For the top selection we will require $\geq$3 jets with these cuts.

After applying all of these cuts we find a total of 42,990 electron events and 28,435 muon events with the jet multiplicity distribution shown in table 3.7. Of these, 204 events pass the $\geq$3-jet requirement.

3.10 Backgrounds in the Lepton+Multijet Sample

There are several sources of lepton+multijet events other than single W production. We discuss the backgrounds from WW and WZ production separately for the b-tagging and kinematic searches in sections 4.3.1 and 5.6. The background from non-W sources includes events originating from several processes: (1) direct

Jet Multiplicity	Percentage Non-W Background
1 Jet	(7.2±2.0)%
2 Jet	(8.3±2.4)%
≥3 Jet	(9.5±3.8)%

Table 3.8: *Percentage non-W background in the lepton+multijet sample as a function of jet multiplicity.*

$b\bar{b}$ production in which a b quark decays seimleptonically ($b \to l\nu c$); (2) QCD jets misidentified as leptons; and (3) residual conversions. The fraction of events in the lepton+multijet sample due to these sources is estimated from the non-isolated lepton sample and the isolated lepton, low-$\not\!\!E_t$ sample, both of which are dominated by background. The ratio of non-isolated background events passing the $\not\!\!E_t > 20$ GeV selection to those in the non-isolated low-$\not\!\!E_t$ region is multiplied by the number of events observed in the low-$\not\!\!E_t$ region of the isolated lepton sample to find the number of background events expected in the W sample, i.e. the high-$\not\!\!E_t$ region of the isolated lepton sample. This method has been compared to explicit calculations of the individual sources of background and found to agree within errors (±30%). The non-W fraction as a function of jet multiplicity is shown in table 3.8, where the measured fraction for each lepton species and for each data-taking period have been determined separately and combined in a weighted average using the observed number of events in each category, and where uncertainties include statistical errors and a systematic uncertainty of ±30%.

3.11 Top Acceptance and Expected Top Content of the Lepton+Multijet Sample

We seek an overall $t\bar{t}$ acceptance $\mathcal{A}_{t\bar{t}}$ for the selection described in section 3.9. We combine the acceptances of both electrons and muons and both Run 1A and Run 1B datasets. The two datasets are combined by weighting the individual Run efficiencies by integrated luminosity. The acceptance $\mathcal{A}_e$ for electrons may be broken down into the following factors:

$$\mathcal{A}_e = \frac{\epsilon_{e\mathrm{id}}}{\epsilon_{e,MC}} * \epsilon_{\mathrm{conv}}^{\mathrm{prompt}} * \epsilon_{e\mathrm{trig}} * BR(t\bar{t} \rightarrow eX) * \epsilon_{iso,E_T,\not{E}_t,3j} \tag{3.2}$$

where $\epsilon_{e\mathrm{id}}$ is the Run 1A/1B-dependent electron identification efficiency from table 3.2 including good electron cuts except isolation; $\epsilon_{e,MC}$ is the same efficiency measured in Monte Carlo events; $\epsilon_{\mathrm{conv}}^{\mathrm{prompt}}$ is the efficiency from table 3.3 for prompt electrons to pass[2] the conversion removal algorithm; ϵ_{trig} is the Run 1A/1B-dependent trigger efficiency given in section 3.1; and the product $BR(t\bar{t} \rightarrow eX) * \epsilon_{iso,E_T,\not{E}_t,3j}$ is determined simultaneously by selecting from all generated $t\bar{t}$ Monte Carlo events those with an electron passing all the cuts described in section 3.9 including isolation, electron $E_T > 20$ GeV, $\not{E}_t > 20$ GeV, and requiring $\geq$3 jets. We normalize to the factor $\epsilon_{e,MC}$ in order to be able to apply the same lepton selection to Monte Carlo samples as to data with the stipulation that the conversion removal and trigger path requirements are not applied to the Monte Carlo samples. These stipulations thus require that we scale the total acceptance by the factors $\epsilon_{\mathrm{conv}}^{\mathrm{prompt}}$ and ϵ_{trig}

[2]Here *pass* means that the prompt electron is *not* identified as a conversion.

lepton l, Run	$\epsilon_{l\text{id}}$	$\epsilon_{\text{conv}}^{\text{prompt}}$	$\epsilon_{l\text{trig}}$	product, $\epsilon_{l\text{prod}}$
e, 1A	(0.84±0.02)	(0.95±0.03)	(0.919±0.004)	(0.73±0.04)
e, 1B	(0.813±0.009)	(0.978±0.005)	($\sim$ 1)	(0.80±0.01)
μ, 1A	(0.906±0.014)	-	(0.80±0.02)	(0.72±0.03)
μ, 1B	(0.924±0.015)	-	($\sim$ 1)	(0.92±0.015)

Table 3.9: *Lepton identification efficiencies, conversion removal efficiencies, and trigger efficiencies for electrons and muons.*

as shown. The acceptance $\mathcal{A}_\mu$ for muons may similarly be broken down into the following factors:

$$\mathcal{A}_\mu = \frac{\epsilon_{\mu\text{id}}}{\epsilon_{\mu,MC}} * \epsilon_{\mu\text{trig}} * BR(t\bar{t} \to \mu X) * \epsilon_{iso,E_T,\not{E}_t,3j} \tag{3.3}$$

where all factors are analogous to the electron acceptance calculation with $\epsilon_{\mu\text{id}}$ obtained from table 3.4, and where we have omitted the conversion removal efficiency, since the conversion removal algorithm is not applied to primary muons. We reiterate in table 3.9 the lepton identification efficiencies and their product $\epsilon_{l\text{prod}}$ necessary to the acceptance calculation above.

Weighting by the Run 1A and Run 1B integrated luminosities of 19.3pb^{-1} and 48pb^{-1}, we derive total lepton efficiencies of $\epsilon_{e\text{prod}} = 0.78 \pm 0.01$ and $\epsilon_{\mu\text{prod}} = 0.87 \pm 0.02$. Monte Carlo Lepton identification efficiencies have been measured in $W \to l\nu$ events generated using the PYTHIA Monte Carlo and found to be $\epsilon_{e,MC} = 0.924 \pm 0.005$ and $\epsilon_{\mu,MC} = 1.0$.

The quantity $BR(t\bar{t} \to lX)\epsilon_{iso,E_T,\not{E}_t,3j}$ for $M_{top} = 170$ GeV/c^2 has been measured in $t\bar{t}$ Monte Carlo events generated using the PYTHIA Monte Carlo program and found to be 0.046$\pm$0.002 for electrons and 0.053$\pm$0.002 for muons. The acceptance for $t\bar{t}$ events is then $\mathcal{A}_{t\bar{t}} = \mathcal{A}_e + \mathcal{A}_\mu = 0.085\pm0.003$. Top acceptance varies somewhat with M_{top}, but the variation is small compared to the variation in $t\bar{t}$ cross section. Given this total $t\bar{t}$ acceptance $\mathcal{A}_{t\bar{t}}$ and the $t\bar{t}$ production cross section $\sigma_{t\bar{t}}$ from figure 1.3 we may predict the expected top signal within the top mass range of interest. For M_{top} = 170 GeV/c^2, $\sigma_{t\bar{t}} = 5.8^{+0.9}_{-0.5}$pb, corresponding to a top signal in our lepton+$\geq$3-jet sample of 33^{+6}_{-4} events, including uncertainties on $\sigma_{t\bar{t}}$, the acceptance, and $\pm$10% on the luminosity. This top content varies by a factor of ~ 2 (~ 0.5) for M_{top} = 150 GeV/c^2 (M_{top} = 190 GeV/c^2). Assuming that the remainder of observed 204 lepton+$\geq$3-jet events in section 3.9 are W+jets background events, then the top signal to W+jets background ratio is expected to be between 1:10 and 1:2, necessitating further techniques for the purpose of isolating the top signal.

Chapter 4

Search for Top Using Secondary Vertex B-Tagging

Our first method for isolating a top signal within the lepton+multijet sample is the selection of events with at least one jet arising from b quark production. The selection of a lepton+$\geq$3 jet sample is expected to yield a top purity of $\sim$10-50% (see section 3.11). Further preferential selection of top events over background events is necessary within this sample to provide a clear top signal. Two b quarks are present in $t\bar{t}$ events while b quarks are not present in the majority of background W+multijet events. Given the relatively long b quark lifetime, we may select jets arising from b quark hadronization by searching for, or *tagging* long-lived hadrons within jets. The decay of a long-lived hadron produces a signature of several charged particle trajectories emanating from a point, or secondary vertex, separated from the $p\overline{p}$ interaction, or primary vertex. We describe below this secondary vertex b-tagging technique and verify that it selects b jets in control samples. We then calculate the number of tagged events expected in the lepton+multijet sample from background sources. We compare this background calculation to the observed number of tagged events, and from the excess over background we calculate a probability that the excess could be due entirely to an upwards fluctuation of the background.

4.1 Secondary Vertex Signature and Resolution

Typical mean b hadron decay lengths in $t\bar{t}$ events are of the order of several millimeters. This scale is set by the b hadron momentum in $t\bar{t}$ events, which is shown in figure 4.1a for $M_{top} = 160$ GeV/c^2. The b hadron momentum is large compared to the b mass, leading to the decay length distribution shown in figure 4.1b. In these plots, we have shown the b hadron momentum and decay length transverse to the beam (in the $x - y$ plane) rather than the total momentum in anticipation of using of transverse decay length in the selection of b hadron decays. We choose transverse decay length, which we henceforth refer to as L_{xy}, because we measure decay lengths using the SVX (see section 2.2.1), which only supplies information in the $x - y$ or $r - \phi$ plane. The mean L_{xy} in figure 4.1a is ~ 0.3 cm.

In order to identify b hadron decays within a jet, we search for its decay products, particles which appear as charged tracks originating from a point separated from the $p\bar{p}$ interaction by a significant distance, as indicated schematically in figure 4.2. The $p\bar{p}$ interaction point is called the *primary vertex* and the b hadron decay point is called the secondary vertex. Most particles in an event, even in the presence of real displaced secondary vertices, will emanate from the primary vertex. In order to distinguish a secondary vertex from the primary vertex, we must understand the spatial uncertainty on both vertices.

The method and precision of determining the primary vertex position is as fol-

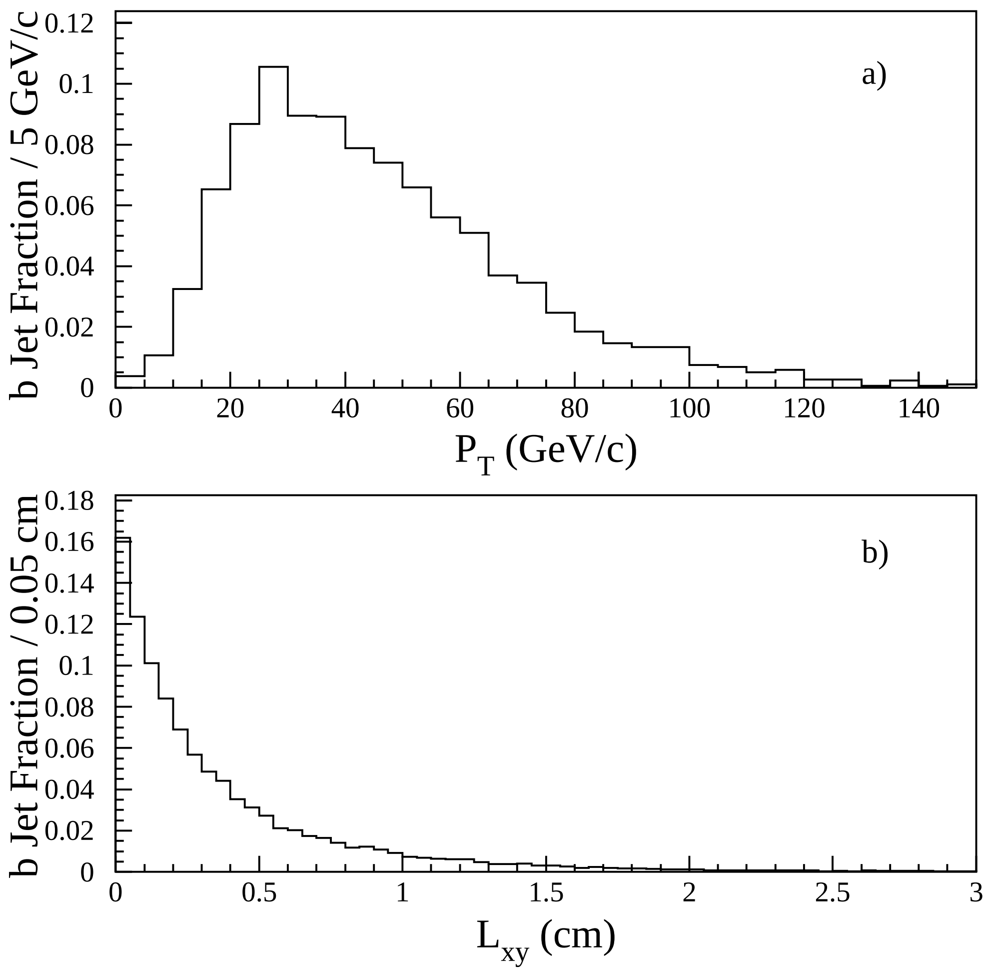


Figure 4.1: *a) The b hadron transverse momentum distribution and b) transverse decay length (L_{xy}) distribution from $t\bar{t}$ Monte Carlo events with M_{top} = 160 GeV/c^2.*

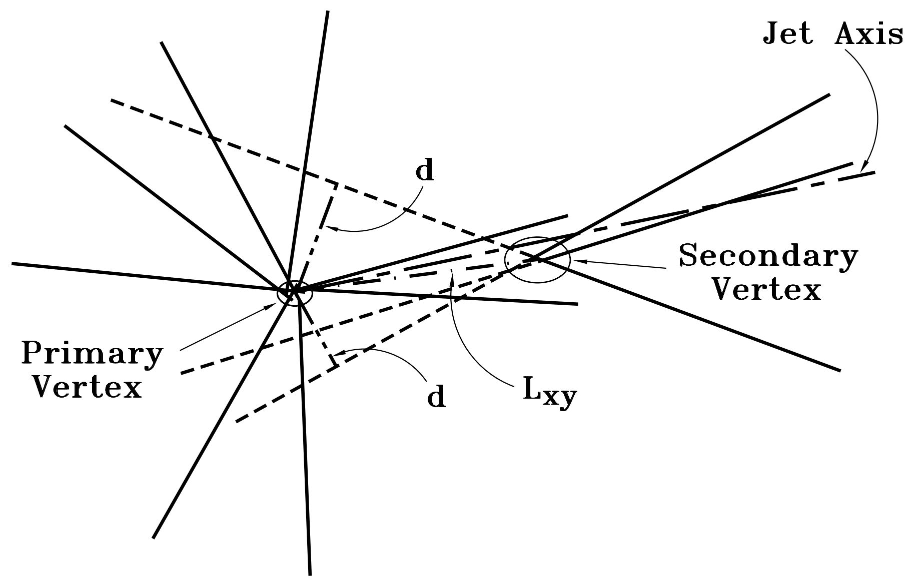


Figure 4.2: *Simplified 2d projection of an event containing a jet with a secondary vertex.*

lows. Since the $p\overline{p}$ beam is slightly tilted with respect to the CDF coordinate system, we must determine the primary vertex z coordinate before we determine the x and y coordinates. The tilt of the beam $\frac{dx}{dz}$ and $\frac{dy}{dz}$ is several microns per centimeter. As a consequence, interactions which occur at the $z = \pm 25$cm ends of the SVX may be displaced from one another by several hundred microns when viewed in the $x - y$ plane. This apparent displacement is large compared to the transverse dimensions of the beam, which is well described by a gaussian of width 36μm. Several good quality primary vertex z positions, may be present in an event due to the presence of one or more minimum bias $p\overline{p}$ collisions in addition to the hard collision. At a luminosity of $5 \times 10^{30}\text{cm}^{-2}\text{s}^{-1}$ there are 0.9 additional interactions per beam crossing.[1] Since

[1]Given an inelastic cross section of $\sim$ 50mb and 3.5μs beam crossing time.

these minimum bias interactions tend to be considerably less energetic than $t\bar{t}$ interactions, we select the good quality vertex with the largest $\sum p_T$ for all associated CTC tracks. The z coordinate of this primary vertex is used along with the slope and offset of the beam to determine the nominal primary vertex x and y coordinates. The beam slope and offset are determined on a run-by-run basis. During a store, the mean $p\bar{p}$ beam position is stable to within $\sim 10\mu$m. On an event-by-event basis, we improve the $x - y$ position determination of the primary vertex ($\sim 36\mu$m) by including SVX information. All good quality SVX tracks are constrained to come from a single point within the envelope of the beam spot. Tracks inconsistent with (i.e. displaced from) the true vertex position are excluded based on their contribution to the overall fit χ^2; after this exclusion the procedure is iterated. We expect the track multiplicity of $t\bar{t}$ events to be large; in such events the primary vertex resolution may be improved to $\sigma_x \sim \sigma_y \sim 10\mu$m. We note that since the primary vertex is determined partially from tracking information with some ϕ structure, σ_x and σ_y are in general unequal and correlated; we thus represent the primary vertex as an ellipse in figure 4.2.

Our ability to measure the separation of a secondary from the primary vertex depends on the spatial resolution of the individual tracks associated to that secondary vertex and also on the primary vertex resolution. Secondary vertices are identified by finding the common point of origin of several candidate tracks which are significantly displaced from the primary vertex; we defer the specific description of this selection to section 4.2. Track displacement with respect to the primary vertex is measured using the perpendicular distance at the point of closest approach, or impact parameter d (see figure 2.4). The impact parameter resolution σ_d has several components, including a p_T-dependent multiple scattering term [$(60/p_T)\mu$m, p_T in GeV/c] and an intrinsic SVX resolution term [$\sim 13\mu$m, see table on page 35]. In practice we may include the primary vertex uncertainties in the calculation of σ_d for individual tracks.[2] Studies in both data and Monte Carlo events indicate that this model for σ_d provides a proper measure of the actual error on d: the distribution of impact parameter significance, d/σ_d, is a gaussian of width ~ 1 with non-gaussian tails from heavy flavor content and tracking errors. Candidate tracks are constrained to a common point to determine the secondary vertex displacement L_{xy}. Like the primary vertex, the secondary vertex is represented by an error ellipse with different, correlated errors in x and y. Typical values of the displacement error are $\sigma_{Lxy} \sim 130\mu$m.

4.2 The Secondary Vertex Algorithm

The secondary vertex finding algorithm consists of three components: candidate track selection, attempted formulation of a secondary vertex, and quality selection on the formed vertex. The algorithm repeats the three steps upon failing to find a secondary vertex.

[2]This procedure produces a correlation between the values of σ_d for individual tracks. For this reason, primary vertex uncertainties are typically included in σ_d for the purposes of individual track studies and for displaced track selection. For the purposes of forming a secondary vertex from several tracks, this correlation is effectively removed.

Selection of candidate tracks is based upon studies of $t\bar{t}$ Monte Carlo b jets[45]. At the generator level, PYTHIA Monte Carlo events with $M_{top} = 160$ GeV/c^2 indicate that b decays produce on average 3.9 detectable ($p_T > 0.5$ GeV/c^2) tracks including those which emanate from tertiary cascade c quark decays. We select displaced tracks using the impact parameter significance; at least two tracks are significantly displaced ($|d/\sigma_d| > 2.5$) 60% of the time. For high efficiency we use a two-pass algorithm which attempts to take advantage of the high track multiplicities typical of $t\bar{t}$ b jets. First, the algorithm attempts a first pass to find a secondary vertex with ≥ 3 tracks using loose track quality cuts. If this pass fails to find a vertex, it attempts to find a 2-track vertex with tighter track quality cuts. The tighter quality cuts of the second pass are necessary to keep tags due to tracking mistakes (*mistags*) at a reasonable level. For consideration in either pass, all tracks are required to fall within a $\Delta R < 0.4$ ($\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$) cone centered on the jet axis, match the primary vertex z coordinate within 5 cm, pass a maximum impact parameter cut to reject tracks from photon conversions and from K_s and Λ decays, have a good combined CTC-SVX track fit, pass minimum CTC track quality cuts, and pass an additional K_s and Λ decay removal algorithm which rejects pairs of tracks within narrow invariant mass windows centered on the K_s and Λ masses. Initially, we require $|d/\sigma_d| > 2.5$ for tracks be included in the first pass. All tracks used in the second pass must pass this cut. The track quality cuts in the two passes vary in minimum p_T, and SVX cluster quality cuts. All tracks are required to have $p_T > 0.5$ GeV/c for the first pass[3] and $p_T > 1.5$ GeV/c for the second pass. Cluster quality cuts depend on the number of SVX hits associated to the track. Tracks containing SVX clusters shared by two or more tracks are prone to pattern recognition errors and are used with particular caution. These track quality requirements, absent the $|d/\sigma_d|$ cut, define the number of *good* tracks in a jet. Specifically, if there are at least two tracks within a jet which satisfy all first pass requirements with the exception of $|d/\sigma_d| > 2.5$, the jet is considered to be *taggable.*

Having selected a set of candidate tracks, the secondary vertex algorithm attempts to find a common vertex for those tracks. The algorithm used here is referred to as the *seed vertexing* algorithm because the first pass attempts to find a vertex from the best two candidate tracks and then attach other tracks to this seed vertex. Candidate tracks are ranked according to p_T, $|d/\sigma_d|$, and SVX cluster quality. The first pass constrains the best two tracks to a common vertex and requires that at least one has $p_T > 2$ GeV/c. Additional tracks which point to this seed vertex are attached to it. These additional tracks are not required to be significantly displaced from the primary vertex. A 3d vertex position is calculated and tracks with a large contribution to the overall vertex χ^2 are excluded. If no additional tracks point to the seed vertex, the seed vertexing process is repeated with another pair of tracks which are chosen based on their rank. Iteration stops upon finding a tag. If no tags with $\geq$3 tracks are found using the first pass, the second pass is executed to find

[3]Tracks with only 2 SVX clusters are also required to have $p_T > 1.5$ GeV/c.

two-track vertices with tighter track quality requirements. At least one of the tracks must have $p_T > 2$ GeV/c. If no tags are found in either the first or second pass, then the jet is untagged.

Two final selections are applied to vertices found using the secondary vertex algorithm. The quantity L_{xy} (see figure 4.2) and its error σ_{Lxy} are calculated; L_{xy} is defined as the displacement of the secondary vertex from the primary vertex in the plane transverse to the beam. The algorithm requires that the vertex be inside the innermost layer of the SVX, $|L_{xy}| < 2.5$ cm, and that the displacement be large with respect to its error, $|L_{xy}/\sigma_{Lxy}| > 3.0$. The sign of L_{xy} is determined by the inner product of the vector from the primary vertex to the the secondary vertex and the vector pointing to the jet centroid as calculated using calorimeter information. Tags with $L_{xy} < 0$ are called *negative* or $-L_{xy}$ tags; tags with $L_{xy} > 0$ are called *positive* or $+L_{xy}$ tags. No requirement is placed on the sign of L_{xy} at this point, although we will make use of it later, since tags due to heavy flavor are predominantly expected to be $+L_{xy}$ tags. One additional variable of interest is the effective lifetime $c\tau_{eff}$. It is calculated using the formula $c\tau_{eff} = L_{xy}\frac{M}{p_T F}$ where M is the total invariant mass of the tracks associated to the secondary vertex and F is a factor used to correct for particles from the b decay which are not identified as part of the secondary vertex. This factor has been determined from Monte Carlo to be $F = 0.7$.

4.2.1 Algorithm Performance and Efficiency

We seek to understand the performance of the secondary vertex algorithm: whether it properly tags real b jets in data and how well we can predict its efficiency using Monte Carlo samples; how often it mistakenly tags jets in which there is no heavy flavor; and finally, its efficiency on top events as measured using Monte Carlo samples.

We check the performance of the algorithm in tagging real b-jets using the b enriched inclusive electron sample [46]. This sample is enriched because it selects semileptonic b decays ($b \to e\nu X$) in $b\bar{b}$ events, for which the cross section is large. Inclusive electron events from $b\bar{b}$ contain two b quarks; the decay of one b produces the trigger electron, and the other usually produces a hadronic jet. The electron is non-isolated owing to the proximity of cascade c quark decay hadrons. We refer to the electron and the surrounding decay products as the *electron jet*, in which we search for the secondary vertex tags using the seed vertexing algorithm. We may also use the algorithm to search for b-tags in the other (hadronic) b jet in the event. The sample requires an electron with $p_T > 10$ GeV/c and a jet with $E_T > 15$ GeV in the ϕ hemisphere opposite the electron. For this comparison, we use the inclusive electron sample from the Run 1B dataset. One may measure the tagging efficiency using either the semileptonic b-jet or the hadronic b-jet. The former measurement is dubbed the *single* tag rate; the latter is the *double* tag rate because we first ask that the hadronic jet is tagged, then measure the electron jet tag rate in this subsample of events. In the single tag rate measurement, knowledge of the b content of the sample is required at first order. The b content is measured by searching for a muon from c quark cascade decay near the electron jet. The muon is required to have opposite charge. The rate of $b \to e\mu X$ events is compared to that in $b\bar{b}$ Monte Carlo events. The ratio between the rate in the data and the rate in the Monte Carlo gives the $b\bar{b}$ content of the data sample. The b content derived using this method is 43$\pm$8%. Using the b content, we rescale the observed electron jet tag rate in data (30$\pm$6% per taggable jet after rescaling) for comparison to the electron jet tag rate in $b\bar{b}$ Monte Carlo. The ratio of the data to Monte Carlo rates using this single tag method is 0.88$\pm$0.17[46]. To verify that the algorithm tags real b hadrons, the effective lifetime $c\tau_{eff}$ distributions for data and $b\bar{b}$ Monte Carlo events are compared and show very good agreement. The double tag calculation is nearly a direct measure of the tagging efficiency, since requiring a tag in the hadronic jet selects a nearly pure $b\bar{b}$ sample. Comparison to Monte Carlo indicates that the double tag rate in data (33$\pm$2% per taggable jet) and Monte Carlo events agree within a factor of 0.98$\pm$0.07[46]. Combining the single tag and double tag measurements yields a data/Monte Carlo tag rate scale factor of $F_{tag}^{1B} = 0.95 \pm 0.07$. Studies in the inclusive muon sample confirm this scale factor to within 2%. For Run 1A data, this scale factor was measured to be $F_{tag}^{1A} = 0.72 \pm 0.21$, giving a luminosity weighted scale factor of $F_{tag} = 0.88 \pm 0.08$. The inclusive electron and muon studies indicate not only that the secondary vertex algorithm does tag real b-jets with an efficiency of

$\sim 32\%$, but that our Monte Carlo models its performance well.

It is crucial that the algorithm not only tag real b jets with high efficiency, but also that it rejects the majority of jets which have no heavy flavor content. Tags which are found in such non-heavy flavor jets are called *mistags* and typically occur as a result of tracking errors and detector resolution effects. These effects cause tracks which actually originate at the the primary vertex to appear displaced. Mistags are expected to be distributed symmetrically about $L_{xy} = 0$. This was verified using a earlier vertexing algorithm similar to the second pass of the current algorithm by observing the L_{xy} distribution for two-track tags from two different jets in inclusive jet trigger events. Such tags are mistags by definition, and their L_{xy} distribution is symmetric.

Inclusive jet trigger samples such as that used above to verify the symmetry of mistags form one of our primary control samples, both for understanding b-tagging efficiency and for the calculation of the mistag background. Positive tags from heavy flavor in the inclusive jet sample are expected from three sources: direct production, final state gluon splitting and initial state gluon splitting, or flavor excitation. Schematic representations of these three processes are shown in figure 4.3. Figure 4.4 shows the observed $c\tau_{eff}$ distribution in 50 GeV inclusive jet trigger events. There are more positive tags in this sample than negative tags, presumedly due to the presence of heavy flavor. The percentage of taggable jets with a $+L_{xy}$ tag in this sample is 2.5%. This rate is lower than the 33% b-jet tag rate in the

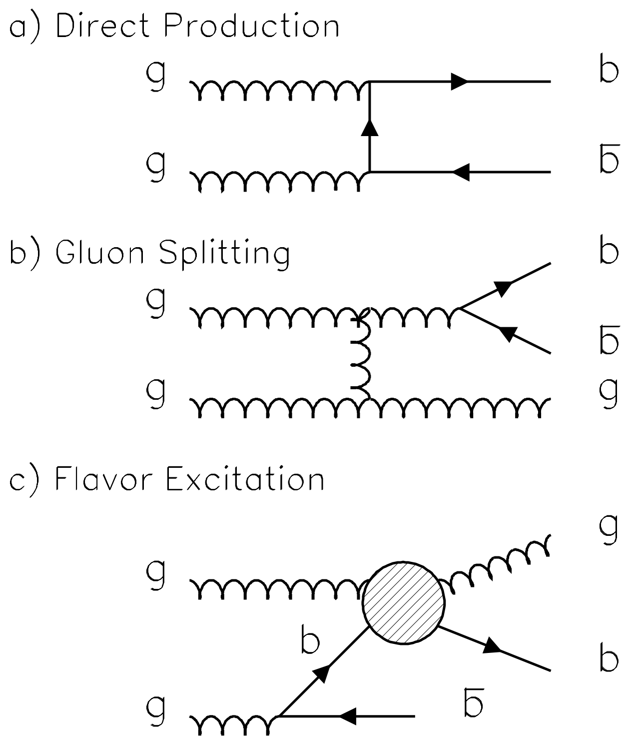


Figure 4.3: *Three sources of heavy flavor in inclusive jets: a) direct production; b) gluon splitting; c) flavor excitation (initial state gluon splitting).*

inclusive electron sample, indicating that inclusive jet events have lower heavy flavor content. The $-L_{xy}$ tag rate in the inclusive jet sample is 0.6%. Tags at negative values of L_{xy} are predominantly mistags. Fitting Monte Carlo $c\tau_{eff}$ distributions for b decays, c decays and mistags to the observed distribution in data indicates that mistags comprise ~65% of the negative tags, with the remainder coming from heavy flavor. Given that mistags are symmetric about $L_{xy} = 0$, we may estimate the true heavy flavor tag rate in a given sample from the excess of positive tags, i.e. the number of positive L_{xy} tags minus the number of negative L_{xy} tags. We verify this assumption by correlating the positive excess tag rate to the heavy flavor content as measured using the soft lepton heavy flavor tagging method. The two measurements agree to within ±35% in various samples. In section 4.3, we will describe how the two rates in these samples are used to estimate the tagged background in the W+multijet sample. Here we observe that the mistag rate, as judged by the $-L_{yx}$ rate in inclusive jet trigger events, is small compared to the real b-tagging rate.

Having established that the secondary vertex algorithm tags b-jets in data with high efficiency and has a low mistag rate in the inclusive jet control sample, we estimate its efficiency in $t\bar{t}$ events passing all selections including the requirement of ≥3 jets. For $t\bar{t}$ events generated with the PYTHIA Monte Carlo with $M_{top} =$ 170 GeV/c^2, this efficiency is 42±5%, including the scale factor F_{tag}.

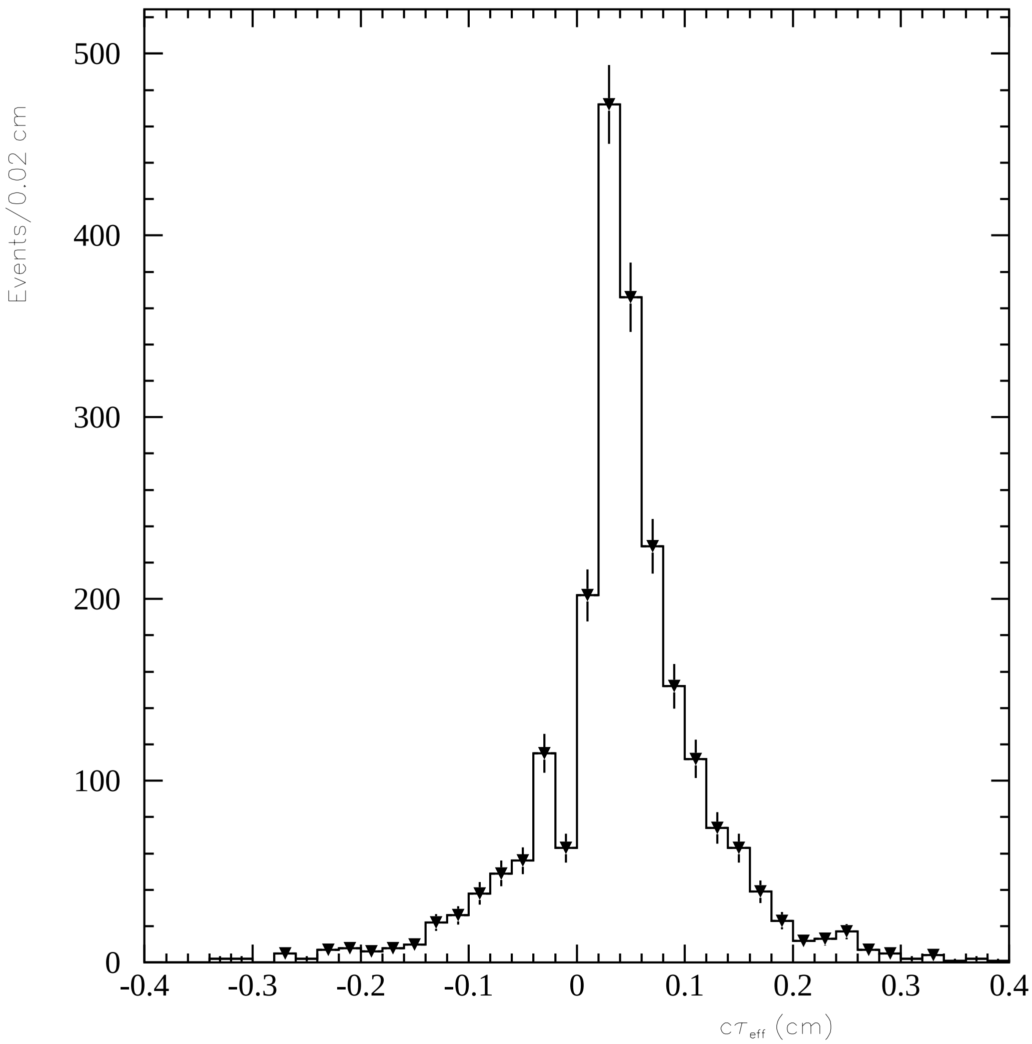


Figure 4.4: *The $c\tau_{eff}$ distribution of tags in the 50 GeV jet sample. The excess of events on the positive side of the distribution is predominantly from heavy flavor.*

4.3 Tagged Background in the Lepton+Jets Sample

In order to understand whether the tags we observe in the W+multijet sample are from $t\bar{t}$ events, we must understand how many tagged events we would expect in this sample from sources other than top. There are several sources of these tagged backgrounds which we will investigate: mistagged jets, heavy quark production ($b\bar{b}$ or $c\bar{c}$) in W events, flavor excitation Wc production, direct $b\bar{b}$ production, WW, WZ, and $Z \to \tau\bar{\tau}$ production. We shall calculate each of these backgrounds in the following sections and summarize them in tables 4.9 and 4.10. The overall strategy is as follows. We calculate the largest backgrounds, those due to $Wb\bar{b}$ and $Wc\bar{c}$, from theoretical estimates, but normalized to the observed rate of W events for each jet multiplicity. We add these contributions to the background due to mistagged jets as estimated using the $-L_{xy}$ tag rate in inclusive jet samples. The smaller contributions due to Wc, WW, WZ, and $Z \to \tau\bar{\tau}$ production are calculated using Monte Carlo samples. The tagged background from non-W sources including $b\bar{b}$ production is estimated from the data.

The larger backgrounds calculated using Monte Carlo models ($Wb\bar{b}$, $Wc\bar{c}$, Wc) are normalized to the observed data in the following manner. The pre-tagged background due to these processes are calculated by finding the fraction of W events as a function of jet multiplicity which correspond to that process and multiplying this fraction by the number of W events with that jet multiplicity. Here the number of W events is taken as the number of events observed after subtracting the expected number of non-W events as estimated in section 3.10. To find the number of tagged events, we apply a tag efficiency found in Monte Carlo events scaled according to the data-Monte Carlo tagging scale factor ($F_{tag} = 0.88 \pm 0.09$). In practice, the normalization to the observed number of events, and the calculation of non-W fractions and tagging scale factor are done separately for the two data sets (Run 1A and Run 1B). The calculations are then combined taking proper account of the correlated systematic errors. Several smaller backgrounds (WW, WZ, $Z \rightarrow \tau\overline{\tau}$) comprising less than 5% of the total are calculated from their absolute cross section.

Since events with more than one jet may have more than one tagged jet, we calculate the total expected number of tagged events and tagged jets from background sources. To this end, we calculate separately the *event* and *jet* tagging rates for the various backgrounds. The event tagging rate is the number of tagged events divided by the total number of events, while the jet tagging rate is defined as the number of tagged jets divided by the total number of events. Thus the jet tagging rate may be slightly higher than the event tagging rate due to the presence of two or more tagged jets in an event. Since $t\overline{t}$ events are more likely to be multiply tagged than background events, the number of tagged jets is somewhat more sensitive than the number of tagged events. In this chapter, we will compare the observed number of tagged jets to the number expected from background. However, we present the total background calculated in both ways in order to use the number of tagged events in chapter 6.

We also perform a confirmatory background calculation which is expected to overestimate the true tagged background. We use the $+L_{xy}$ tag rate in inclusive jet samples as an upper bound on the rate of tags from mistags, $Wb\bar{b}$, and $Wc\bar{c}$, since this $+L_{xy}$ rate includes contributions from mistags and heavy flavor. The other smaller background sources are calculated in the same way as in the primary background calculation. Theoretical tools indicate that this method overestimates the actual background when applied to the W+multijet sample. This confirmatory calculation is summarized in tables 4.11 and 4.12.

4.3.1 Primary Background Calculation

Mistag Background

We estimate the mistag background by measuring the rate of $-L_{xy}$ tags in inclusive jet events and applying this rate to the W+multijet sample. The mistag rate depends on jet characteristics: it is a strong function of SVX track multiplicity and a somewhat weaker function of jet E_T, as shown in figure 4.5. Here SVX track multiplicity is defined as the number of SVX tracks passing the requirements of the secondary vertexing algorithm with the exception of the d/σ_d cut. The rate is defined as the number of tags divided by the number of taggable jets with $E_T > 15$ GeV, where a taggable jet is defined as a jet with at least two good SVX tracks (see page 79). Since the W+multijet sample may have different SVX multiplicity and jet E_T spectra than the inclusive jet sample, we parametrize the mistag rate by these two variables. In order to attain good statistics for a broad range of jet E_T, this

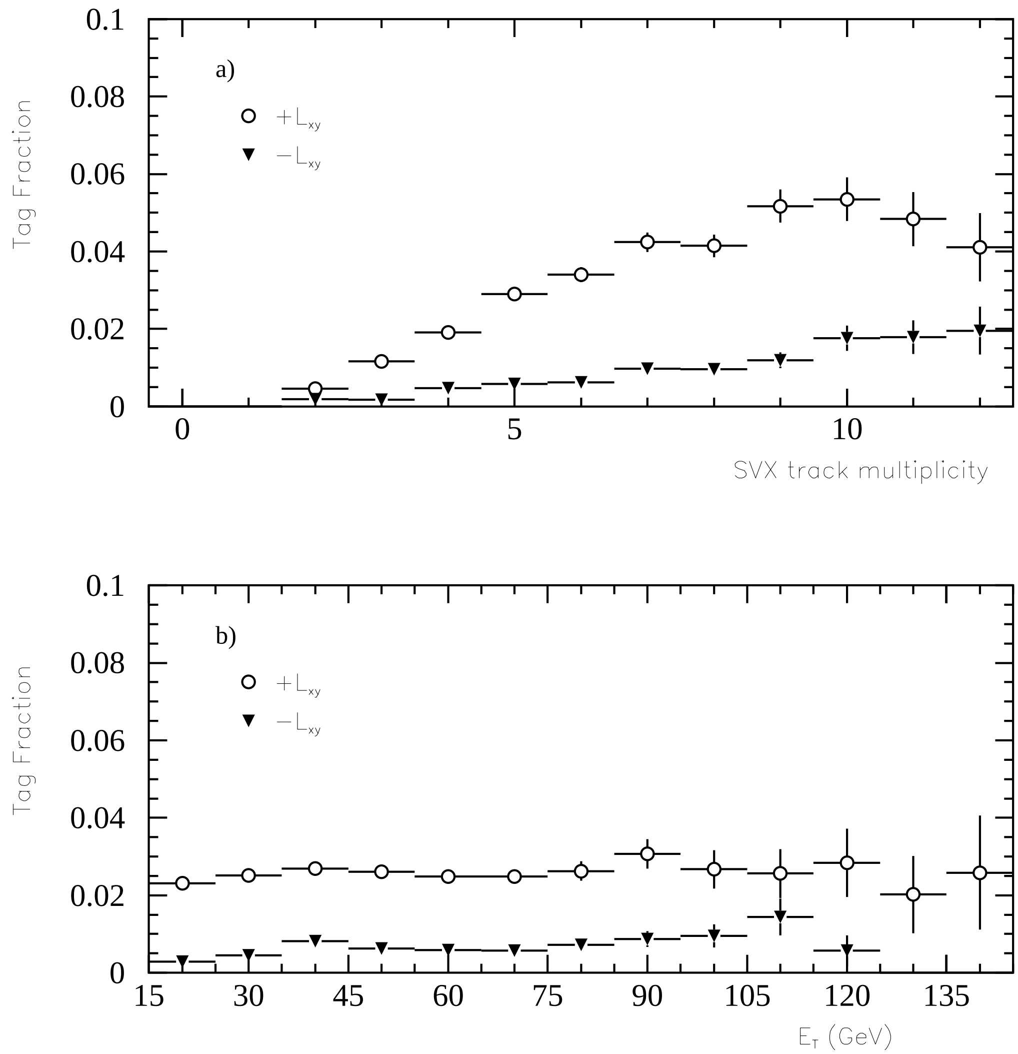


Figure 4.5: *The positive and negative L_{xy} tagging rate as a function of a) good SVX track multiplicity and b) jet E_T for all taggable jets in the inclusive 50 GeV jet sample.*

parametrization uses three samples: these are the 20-, 50- and 100 GeV inclusive jet trigger samples. Also shown in figure 4.5 is the $+L_{xy}$ tagging rate in inclusive jets, which is a mixture of real heavy flavor tagging and mistagging. This rate is also parametrized by SVX multiplicity and jet E_T. We check that these parametrizations accurately predict the positive and negative tag rates by applying them to various samples not used in the parametrization. The results of this comparison in the 70 GeV jet sample, the 140 GeV jet sample, a sample with scalar $\sum E_T$ greater than 300 GeV, a sample of electron conversions, and a sample of Z boson events is shown in table 4.1. The agreement is very good. Comparisons are also shown in the 20-, 50-, and 100 GeV inclusive jet samples. Since the parametrization for a given E_T and SVX track multiplicity is extracted from all three of these samples, the prediction in a single sample may be different from the actual number of tags in that sample. In addition, we may use $+L_{xy}$ and $-L_{xy}$ parametrizations to predict the expected distribution of tagged events in variables other than those used in the parametrization. Figure 4.6 compares the predicted and observed distributions for four such variables in the 140 GeV jet sample: jet multiplicity, scalar event $\sum E_T$, minimum ϕ separation between the tagged and next adjacent jet and the instantaneous luminosity. The good agreement indicates that any dependence of the tag rate on these four variables is well described by the parametrization we have chosen. Based on the maximum differences between predicted and observed tags in the various samples in table 4.1, we assign systematic errors of $\pm 15\%$ and $\pm 20\%$

Sample	positive tags		negative tags	
	Observed	Predicted	Observed	Predicted
Jet 20	795	814	108	120
Jet 50	2307	2414	541	564
Jet 100	1895	1767	538	500
Jet 70	1341	1296	371	365
Jet 140	739	665	270	272
$\sum E_T > 300$ GeV	1947	1767	691	697
Conversions	52	41	4	6
Z+jets	7	7.8	2	1.6

Table 4.1: *Observed and predicted positive and negative tags in various jet samples.*

respectively to the tag rates predicted by $+L_{xy}$ and $-L_{xy}$ parametrizations. The result of applying the $-L_{xy}$ parametrization to the W+multijet sample is listed as a function of jet multiplicity in line 1 of tables 4.9 and 4.10. Note that the negative mistag rate for jets (table 4.9) and events (table 4.10) are indistinguishable because the *double* negative mistag probability is very small. The result applying the $+L_{xy}$ parametrization to the W+multijet sample is listed as a function of jet multiplicity in line 1 of tables 4.11 and 4.12.

$Wb\bar{b}$ and $Wc\bar{c}$ Backgrounds

The background from $Wb\bar{b}$ and $Wc\bar{c}$ production is calculated using the HERWIG Monte Carlo program and the matrix element calculation [47] for the leading order diagram shown in figure 4.7a. The absolute cross section as calculated in [47] varies by $\pm 40\%$ with reasonable variations in the factorization/normalization scale. The absolute cross section for heavy flavor production in W+1 jet HERWIG Monte Carlo

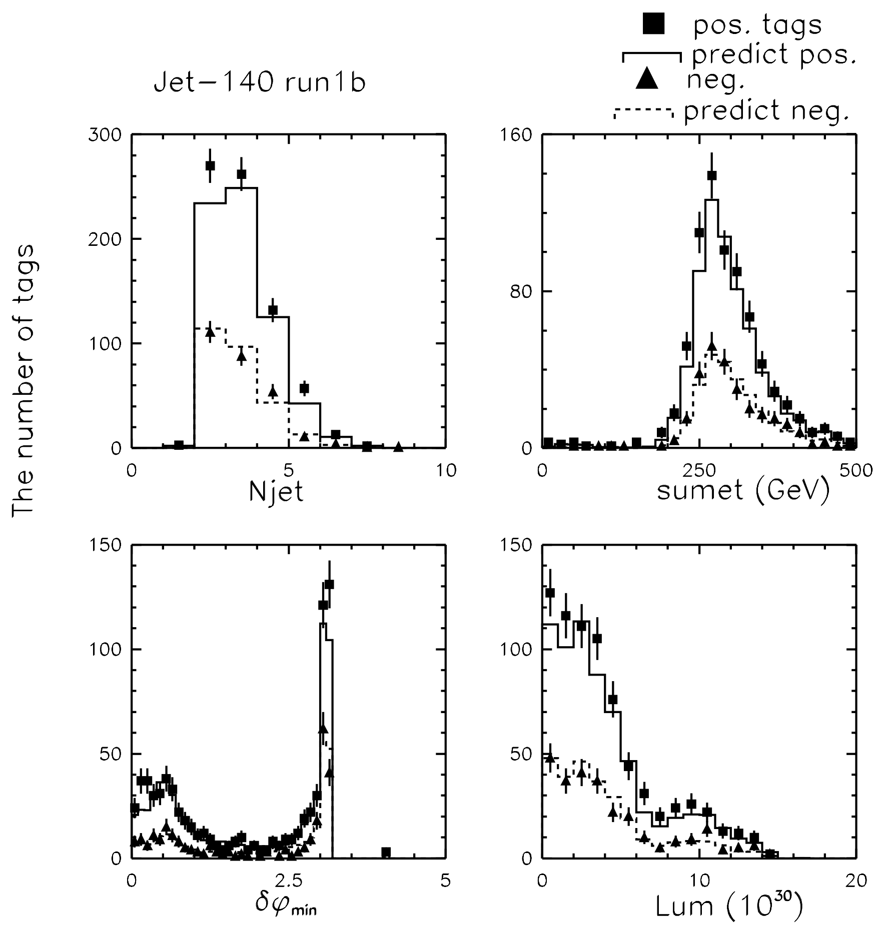


Figure 4.6: *The predicted number of positive (solid histogram) and negative (dashed histogram) tags versus the observed number of positive (filled squares) and negative (filled triangles) tags as a function of four variables in the 140 GeV inclusive jet sample. The prediction is based on the parametrization of positive and negative tag rates by SVX track multiplicity and jet E_T in three other inclusive jet samples. The four variables compared are the jet multiplicity (Njet), the event scalar $\sum E_T$ (sumet) in units of GeV, the minimum ϕ separation between the tagged jet and its nearest neighbor ($\delta\phi_{min}$), and the instantaneous luminosity in units of $10^{30} cm^{-2} s^{-1}$.*

events must be scaled up by a factor of 1.4 in order to match the upper part of this range. HERWIG Monte Carlo events are used to calculate the fraction of events containing a heavy quark pair at each jet multiplicity, as shown in tables 4.2 and 4.3, and we scale this fraction up by the factor of 1.4. We use the HERWIG Monte Carlo program rather than the leading order calculation because the HERWIG parton shower approximation accounts for higher order diagrams in addition to the leading order process shown in figure 4.7a. Especially in the $\geq$ 3-jet signal region, such contributions are non-negligible and lead to larger heavy flavor content than the leading order calculation alone. The number of W events with a $b\bar{b}$ or $c\bar{c}$ pair is determined as a function of jet multiplicity by multiplying this fraction by the observed number of events in data. Since the observed number of events is the sum of the number of W events and the number of non-W events, we must also correct for the non-W fraction. We then multiply by the tagging efficiency of $Wb\bar{b}$ or $Wc\bar{c}$ events as calculated for each jet multiplicity from a sample generated using the leading order matrix calculation with parton evolution performed by HERWIG. For the tagging efficiency, we elect to use the leading order calculation because such events more often have two hard jets arising from the two b-quarks than would be the case in the HERWIG calculation. The use of the leading order calculation thus leads to a somewhat larger or more conservative tagging efficiency than the HERWIG calculation. These tag efficiencies are listed in tables 4.2 and 4.3. The tag rate scale factors F_{tag} (see section 4.2.1) are applied to these tag efficiencies separately for application to data from Run 1A and Run 1B. The total $Wb\bar{b}$ and $Wc\bar{c}$ tagged jet and tagged event backgrounds are listed in line 2 of tables 4.9 and 4.10 respectively with uncertainties as described below.

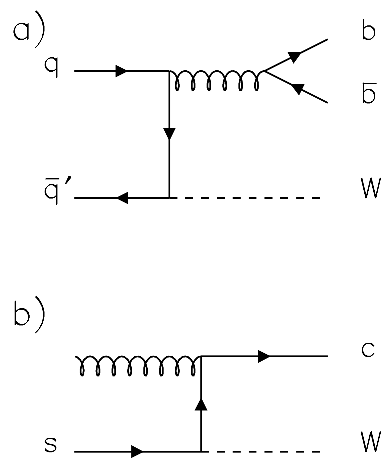


Figure 4.7: *Leading diagrams for (a) $Wb\bar{b}$ and (b) Wc production. $Wc\bar{c}$ production is the same as (a) but with a final state $c\bar{c}$ pair.*

The uncertainty on the $Wb\bar{b}$ and $Wc\bar{c}$ background is estimated as follows. We check HERWIG's ability to replicate heavy flavor production rates by comparing the tag rates in QCD jet events generated by HERWIG to tag rates in the inclusive jet samples in data. For this purpose we compare the positive tag excess rates $((+L_{xy}) - (-L_{xy}))$ in the two samples. We use the positive excess rather than the

Jet Multiplicity	$Wb\bar{b}$ (1) % $b\bar{b}$	(2) tag efficiency
1 Jet	0.74±0.03%	22.8±0.5%
2 Jet	1.5±0.1%	35±2% (42±2%)
≥3 Jet	3.0±0.4%	32±5% (36±5%)

Table 4.2: *Column 1: Fraction of W events as a function of jet multiplicity which contain a $b\bar{b}$ pair. A factor of 1.4 as described in the text has already been applied. Column 2: Event tag efficiency for $Wb\bar{b}$ events as a function of jet multiplicity. Where the jet tag rate differs from the event tag rate, the jet tag efficiency is listed in parentheses. The scale factor F_{tag} has not been applied.*

Jet Multiplicity	$Wc\bar{c}$ (1) % $c\bar{c}$	(2) tag efficiency
1 Jet	1.66±0.05%	5.1±0.4%
2 Jet	3.4±0.2%	9±1
≥3 Jet	4.8±0.7%	9±3

Table 4.3: *Column 1: Fraction of W events as a function of jet multiplicity which contain a $c\bar{c}$ pair. A factor of 1.4 as described in the text has already been applied. Column 2: Event tag efficiency for $Wc\bar{c}$ events as a function of jet multiplicity. Given the relatively low c-jet tag probability and thus the low double tag probability, jet tag rate and event tag rate are the same in this case. The scale factor F_{tag} has not been applied.*

raw positive tag rate in order to compare as directly as possible the true heavy flavor tagging rates and also to minimize dependence upon the Monte Carlo model of mistagging. Figure 4.8 shows agreement in the positive excess tag rate for data and HERWIG Monte Carlo events as a function of jet multiplicity and jet E_T. The scale factor of 1.4 has not been applied to the Monte Carlo. Using a previous tagging algorithm, the positive excess rate in the 20 GeV jet sample after the requirement of at least two good SVX tracks was larger in data than in Monte Carlo jets by a factor of 1.5. The systematic uncertainty on the the HERWIG heavy flavor content prediction was set by computing by what factor the gluon splitting to $b\bar{b}$ and $c\bar{c}$ must be increased in HERWIG in order to cause the positive excess rate in Monte Carlo to exceed the rate in data by one standard deviation. The necessary factor was 2.2, leading to the adoption of an uncertainty of ± 0.8 (60%) on the previously determined HERWIG heavy flavor content prediction scale factor of 1.4. We perform the same comparison with the current algorithm, using the same Monte Carlo/data scale factor (0.72, see section 4.2.1) as in the previous analysis and find that for the 20 GeV jet sample, one would need to increase the gluon splitting in HERWIG by a factor of about 2.1. Using the luminosity-weighted scale factor of 0.88, the gluon splitting increase factor is 1.5. For consistency, we assign the same 60% uncertainty to the final $Wb\bar{b}$ and $Wc\bar{c}$ background estimates. We assign an additional uncertainty of $\pm 40\%$ for the dependence on the number of jets, as shown in figure 4.8. We also add the uncertainty on the heavy flavor purity of the positive excess ($\pm 35\%$) and the uncertainty on the Monte Carlo tag efficiency ($\pm 9\%$) as described in section 4.2.1 for a total uncertainty of $\pm 80\%$ on the $Wb\bar{b}$ and $Wc\bar{c}$ prediction. Note that this error is correlated between $Wb\bar{b}$ and $Wc\bar{c}$ predictions.

Wc Background

We consider the number of events expected from the single charm production, or Wc processes. At leading order, the largest contribution is from the $gs \rightarrow Wc$ diagram shown in figure 4.7b, with the Cabibbo-suppressed process $gd \rightarrow Wc$ contributing at the 10-15% level. The fraction of W events from single charm production, F_{Wc}, is determined at the parton level from the HERWIG and VECBOS W+1 jet matrix element calculation and found to be $(5.3 \pm 1.3)\%$, where the uncertainty allows for variation in the strange sea content of the proton according to a variety of structure functions. The variation of the Wc fraction with jet multiplicity is measured using the VECBOS W+n-jet matrix element calculations at parton level, and the uncertainty due to strange sea content is not found to vary substantially with jet multiplicity. To this point, F_{Wc} has been determined at the parton level rather than at the fully simulated and reconstructed jet cluster level. Since c-quark jets are reconstructed more efficiently than light quark jets, an additional factor of 1.11 ± 0.03, as measured in fully simulated HERWIG W+1-jet events, is applied to the parton level values of F_{Wc}. We determine the tagging efficiency for Wc events from PYTHIA Monte Carlo events. The Wc fraction F_{Wc} and the tagging efficiencies are listed as a function of jet multiplicity in table 4.4. As in the case of $Wb\bar{b}$ and

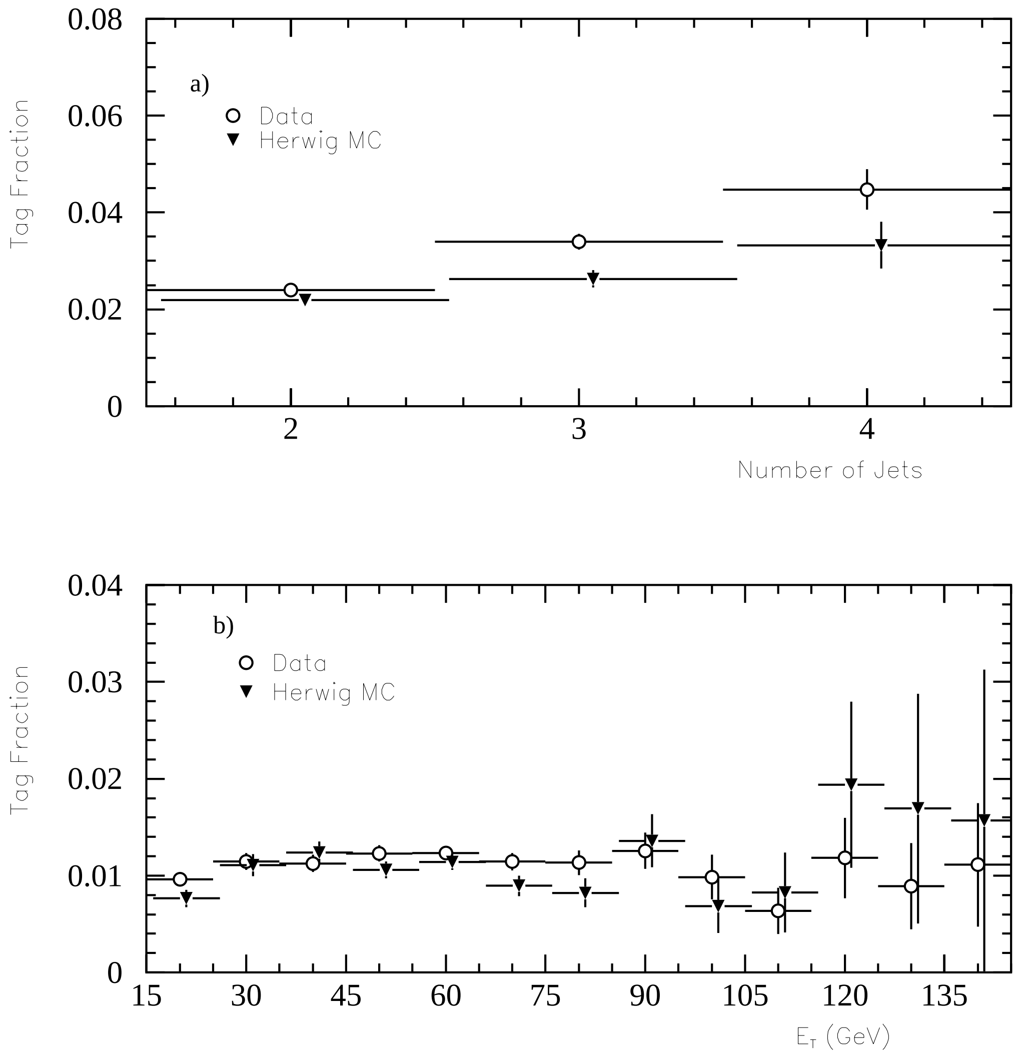


Figure 4.8: *Positive excess tag rate as a function of (a) jet multiplicity and (b) jet E_T for a HERWIG Monte Carlo simulation and the 50 GeV jet sample.*

Jet Multiplicity	(1) % Wc, F_{Wc}	(2) tag efficiency
1 Jet	5.9±1.4%	4.8±0.5%
2 Jet	8.3±1.7%	6.1±1.0%
≥3 Jet	8.9±1.7%	7.8±3.0%

Table 4.4: *Column 1: Fraction of W events as a function of jet multiplicity which contain a single c quark. A factor of 1.11 has already been applied to the parton-level values to account for c jet reconstruction bias at simulation level. Column 2: Event tag efficiency for Wc events as a function of jet multiplicity. Jet tag rate and event tag rate are identical. The scale factor F_{tag} has not been applied.*

$Wc\bar{c}$, the final contribution due to Wc is calculated by applying to the observed number of events (corrected for the non-W fraction) of a given multiplicity the Wc fraction and the tagging efficiency. The total Wc tagged background is listed in line 3 of tables 4.9 and 4.10. We use a $\pm 30\%$ uncertainty in the 1- and 2- jet bins and a $\pm 45\%$ uncertainty in the 3- and 4- jet bins based on the uncertainties shown in table 4.4.

Non-W Background

The $b\bar{b}$ component of the pre-tagged non-W background described in section 3.10 is a potential source of tagged background events. The tag rate per taggable jet for such events is determined from low-$\not{E}_t$, isolated lepton events. For 1-jet events, the tagging rate is (2.4±0.3)% per taggable jet and for ≥2-jet events, the tagging rate is (2.6±0.6)%. This rate is multiplied by the observed number of taggable jets for a given jet multiplicity and by the previously determined non-W fractions to estimate the number of tagged non-W events. A second method of estimating the tagged background involves selecting a sample dominated by $b\bar{b}$ events and using the distribution of such events in the isolation variable I to estimate the $b\bar{b}$ background in the low isolation ($I < 0.1$) region of the high-$\not\!\!E_t$ lepton sample. Such a $b\bar{b}$-dominated sample is selected by requiring a b-tag in low-$\not\!\!E_t$ events. This method yields the same result as the first within errors. The resulting backgrounds due to these non-W sources are listed on line 5 of tables 4.9 and 4.10. The uncertainties include statistical errors and an additional 30% systematic error on the isolation *vs.* $\not\!\!E_t$ method of estimating the pre-tagged non-W background.

WW, WZ, $Z \rightarrow \tau\bar{\tau}$ Backgrounds

The remaining backgrounds are from WW, WZ, and $Z \rightarrow \tau\bar{\tau}$ production and constitute less than 5% of the total background in the signal region. Diboson production may produce tags through the decays $W \rightarrow c\bar{s}$ and $Z \rightarrow b\bar{b}$. The tagged background due to diboson events is estimated from their theoretical cross sections, $\sigma_{WW} = 9.5 \pm 0.7$pb [48] and $\sigma_{WZ} = 2.6 \pm 0.4$pb [49], the branching ratios to heavy flavor processes (BR($WW \rightarrow e\nu c\bar{s}$) $\simeq \frac{2}{27}$, BR($WZ \rightarrow e\nu b\bar{b}$) $\simeq \frac{1}{54}$), the total integrated luminosity of 67 pb^{-1}, and lepton+jet selection acceptances and tag efficiencies as calculated using the PYTHIA Monte Carlo generator. These values are summarized in tables 4.5 and 4.6. Since the acceptances are calculated only using Monte Carlo events with electrons, the rate for muons is derived from the relative rate of muon and electron events observed in the data. Tags in the $Z \rightarrow \tau\bar{\tau}$ process arise through tagging multiprong hadronic decays of the long-lived ($c\tau = 90\mu$m) τ

	WZ	
Jet Multiplicity	(1) % acceptance	(2) % tag efficiency
1 Jet	11.0±0.5%	4.0±0.4%
2 Jet	11.0±0.5%	5.2±0.5%
≥3 Jet	2.3±0.2%	7.1±1.0%

Table 4.5: *Acceptance (Column 1) and tag efficiency (Column 2) of* $WW \rightarrow e\nu c\bar{s}$ *events as a function of jet multiplicity. Jet tag rate and event tag rate are identical. The scale factor* F_{tag} *has not been applied.*

	WZ	
Jet Multiplicity	(1) % acceptance	(2) % tag efficiency
1 Jet	9.4±0.5%	24±3%
2 Jet	9.9±0.5%	41±5 (49±6)
≥3 Jet	2.2±0.2%	42±5 (51±6)

Table 4.6: *Acceptance (Column 1) and tag efficiency (Column 2) of* $WZ \rightarrow e\nu b\bar{b}$ *events as a function of jet multiplicity. Where the jet tag rate differs from the event tag rate, the jet tag efficiency is listed in parentheses. The scale factor* F_{tag} *has not been applied.*

	$Z \to \tau\overline{\tau}$	
Jet Multiplicity	(1) % acceptance	(2) % tag efficiency
1 Jet	0.29±0.03%	11±2%
2 Jet	0.08±0.01%	11±2%
≥3 Jet	0.010±0.005%	11±2%

Table 4.7: *Acceptance (Column 1) and tag efficiency (Column 2) of $Z \to \tau\overline{\tau}$ events as a function of jet multiplicity. The scale factor F_{tag} has not been applied.*

lepton. The background due to this process is the product of: (1) the production cross section $\sigma(Z \to \tau\overline{\tau}) = 0.2nb$; (2) the total integrated luminosity $\mathcal{L} = 67\text{pb}^{-1}$; (3) the branching ratio for one τ to decay to an electron or muon and the other τ to decay to three charged pions, $BR(\tau \to e\nu_e\nu_\tau, \tau \to \pi\pi\pi) = 0.05$; (4) the lepton+jets selection acceptance; (5) the tag efficiency for the accepted events. The latter two values are determined from a sample of $Z \to \tau\overline{\tau}$ events generated using the PYTHIA Monte Carlo generator and are listed in table 4.7. The rate for muon events is normalized to the observed relative number of electrons and muon in data.

The total tagged backgrounds expected from WW, WZ, and $Z \to \tau\overline{\tau}$ events is listed in table 4.8. The uncertainties shown are from the luminosity, production cross sections, acceptances, tag efficiencies, and Monte Carlo/data tag rate scale factor and are of the order of 20-30%. The total of these three processes is shown in line 4 of tables 4.9 and 4.10. The uncertainty on the total is taken to have a systematic error of ±40% to account for jet multiplicity scaling uncertainty on the production cross sections.

Jet Multiplicity	(1) WW	(2) WZ	(3) $Z \rightarrow \tau\overline{\tau}$
1 Jet	0.32±0.06	0.12±0.03	0.33±0.09
2 Jet	0.44±0.08	0.22±0.05 (0.26±0.07)	0.09±0.03
≥3 Jet	0.12±0.03	0.049±0.013 (0.060±0.017)	0.011±0.006

Table 4.8: *Estimated tagged backgrounds from WW, WZ, and $Z \rightarrow \tau\overline{\tau}$ production. Uncertainties reflect uncertainties on the luminosity, acceptance, and tag efficiency. Where the expected number of tagged jets differs from the expected number of tagged events, the number of tagged jets is listed in parentheses.*

Total Tagged Background

Tables 4.9 and 4.10 list the individual backgrounds described above and also the total. In general, the uncertainties for individual sources are highly correlated between jet multiplicities. The total number of tagged jets (events) expected purely from background in the ≥3-jet multiplicity region is 6.7±2.1 (6.4±1.9). We defer discussion of the observed tags until after the presentation of the confirmatory background calculation.

4.3.2 Confirmatory Background Calculation

For the confirmatory background calculation, we replace the explicit calculation of the $Wb\overline{b}$ and $Wc\overline{c}$ backgrounds with the $+L_{xy}$ tag parametrization extracted from inclusive jets. Since the positive tag parametrization also includes mistags, we also use it to represent mistags rather than the $-L_{xy}$ parametrization. The accuracy of this method depends on the extent to which the rate of heavy flavor production in inclusive jets replicates the rate in W events. We expect that the rate in inclusive jets exceeds that in W+jets events by a factor of between 2 and 4

Source	$W+1$ Jet	$W+2$ Jets	$W+3$ Jets	$W+\geq 4$ Jets
(1) Mistags	14.80± 2.96	5.30± 1.06	1.40± 0.28	0.52± 0.10
(2) $Wb\bar{b}$, $Wc\bar{c}$	13.81±11.05	7.79± 6.23	1.96± 1.57	0.50± 0.40
(3) Wc	15.34± 4.60	4.24± 1.27	0.92± 0.41	0.22± 0.10
(4) $Z \to \tau\bar{\tau}$, WW, WZ	0.77± 0.31	0.78± 0.31	0.15± 0.06	0.04± 0.01
(5) Non-W, including $b\bar{b}$	5.70± 1.44	2.96± 0.76	0.77± 0.19	0.18± 0.04
(6) **Total Background**	50.42±12.42	21.07± 6.50	5.21± 1.65	1.45± 0.43
(7) **Observed Tagged Jets**	40	34	17	10

Table 4.9: *Summary of the expected number of tagged jets from background sources in the lepton+jets sample as a function of jet multiplicity. Individual background sources are listed on lines (1) through (5) and the total is listed on line (6). The number of observed tagged jets is listed on line (7).*

Source	$W+1$ Jet	$W+2$ Jets	$W+3$ Jets	$W+\geq 4$ Jets
(1) Mistags	14.80± 2.96	5.30± 1.06	1.40± 0.28	0.52± 0.10
(2) $Wb\bar{b}$, $Wc\bar{c}$	13.81±11.05	6.92± 5.53	1.79± 1.43	0.45± 0.36
(3) Wc	15.34± 4.60	4.24± 1.27	0.92± 0.41	0.22± 0.10
(4) $Z \to \tau\tau$, WW, WZ	0.77± 0.31	0.74± 0.30	0.15± 0.06	0.03± 0.01
(5) Non-W, including $b\bar{b}$	5.70± 1.44	2.96± 0.76	0.77± 0.19	0.18± 0.04
(6) **Total Background**	50.42±12.42	20.15± 5.83	5.03± 1.52	1.40± 0.39
(7) **Observed Tagged Events**	40	30	12	9
(8) **Events Before Tag**	6578	1026	165	39

Table 4.10: *Summary of the expected number of tagged events from background sources in the lepton+jets sample as a function of jet multiplicity. Individual background sources are listed on lines (1) through (5) and the total is listed on line (6). The number of observed tagged events is listed on line (7). The number of pre-tagged events is listed on line (8).*

owing to: (1) the absence of the direct production and flavor excitation mechanisms (figures 4.3a and 4.3c) from W events; (2) a higher fraction of gluon jets (as opposed to light quark jets, in which heavy flavor production is suppressed) in the inclusive jet sample; (3) biases on gluon jet fraction and heavy flavor production probability in inclusive jets due to the high energy trigger jet and the overall event energy. The total background due to $Wb\bar{b}$, $Wc\bar{c}$, and mistags as calculated using the $+L_{xy}$ rate in inclusive jets is listed on line 1 of tables 4.11 and 4.12, where the raw value obtained by applying the $+L_{xy}$ parametrization has been rescaled by subtracting out the fractional contribution from non-W sources. This subtraction is done in order to avoid double counting the tagged background from non-W sources, which has been calculated separately. We complete the calculation of the expected background by adding to this value the contributions from Wc, $Z \to \tau\overline{\tau}$, WW, WZ and non-W sources as previously calculated (lines 3–5 of tables 4.9 and 4.10). The resulting totals are listed on line 2 of tables 4.11 and 4.12. In the signal region, we expect 9.2$\pm$1.2 tagged jets.

4.3.3 Application to the Data and Significance

Application of the SVX tagging algorithm to the W+multijet sample yields 27 tagged jets in 21 tagged events in the $W+ \geq$ 3-jet signal region, well in excess of the expected background of 6.7$\pm$2.1 tags. The number of observed tagged jets and tagged events for all jet multiplicities is listed on line 7 of tables 4.9 and 4.10. Figure 4.9 shows the jet multiplicity of observed tags in the W+multijet sample versus

Source	$W+1$ Jet	$W+2$ Jets	$W+3$ Jets	$W+\geq 4$ Jets
(1) $Wb\bar{b}$, $Wc\bar{c}$ + Mistags	58.19± 8.73	19.42± 2.91	5.11± 0.77	1.78± 0.27
(2) **Total Background**	80.00± 9.98	27.40± 3.28	6.96± 0.90	2.21± 0.29

Table 4.11: *Summary of the expected number of tagged jets from background sources in the lepton+jets sample as a function of jet multiplicity using the confirmatory background calculation, which is expected to be an overestimate. The total background listed on line (2) is the sum of the value listed on line (1) and the value listed on lines (3) through (5) of table 4.9, as described in the text.*

Source	$W+1$ Jet	$W+2$ Jets	$W+3$ Jets	$W+\geq 4$ Jets
(1) $Wb\bar{b}$, $Wc\bar{c}$ + Mistags	58.19± 8.73	19.30± 2.89	5.06± 0.76	1.73± 0.26
(2) **Total Background**	80.00± 9.98	27.24± 3.27	6.90± 0.89	2.16± 0.28

Table 4.12: *Summary of the expected number of tagged events from background sources in the lepton+jets sample as a function of jet multiplicity using the confirmatory background calculation, which is expected to be an overestimate. The total background listed on line (2) is the sum of the value listed on line (1) and the value listed on lines (3) through (5) of table 4.10, as described in the text.*

the total background. Also shown is the confirmatory background calculation. Note that in the background-dominated W+1-jet region, the confirmatory background exceeds the observed number of tags by a factor of two (several standard deviations), while the primary background calculation agrees to within errors. In the W+2 jet bin, the observed number of tags exceeds both background calculations. However, since $t\bar{t}$ events may contribute tags to this region, we defer a discussion until after estimating the top content. The predicted mistag rate is in good agreement with the observed negative L_{xy} tags: there are 18 (5) negative L_{xy} tags in the W+1 jet (W+2 jet) bin compared to an expected mistag rate of 14.8±3.0 (5.3±1.1). In the $W+\geq 3$ jet region, no negative L_{xy} tags are observed compared to 1.9±0.3 expected. Figure 4.10 shows the lifetime distribution of the observed tags in the $W+\geq 3$ jet sample compared to the expected distribution of tagged b-jets in PYTHIA $t\bar{t}$ Monte Carlo events with $M_{top} = 170$ GeV/c^2. The lifetime distributions are consistent.

We now proceed to determine the *significance* of the observation, i.e. the probability that the observed excess in the signal region is due to background only. In the case of counting events with no error on the background estimate, this corresponds to the probability that a Poisson-distributed variable with mean given by the expected number of background events could yield greater than or equal to the observed number of events. We may account for the uncertainty on the background estimate using a Monte Carlo program to smear the Poisson distribution mean. However, in this case, we are interested in the number of *tags* observed, not the

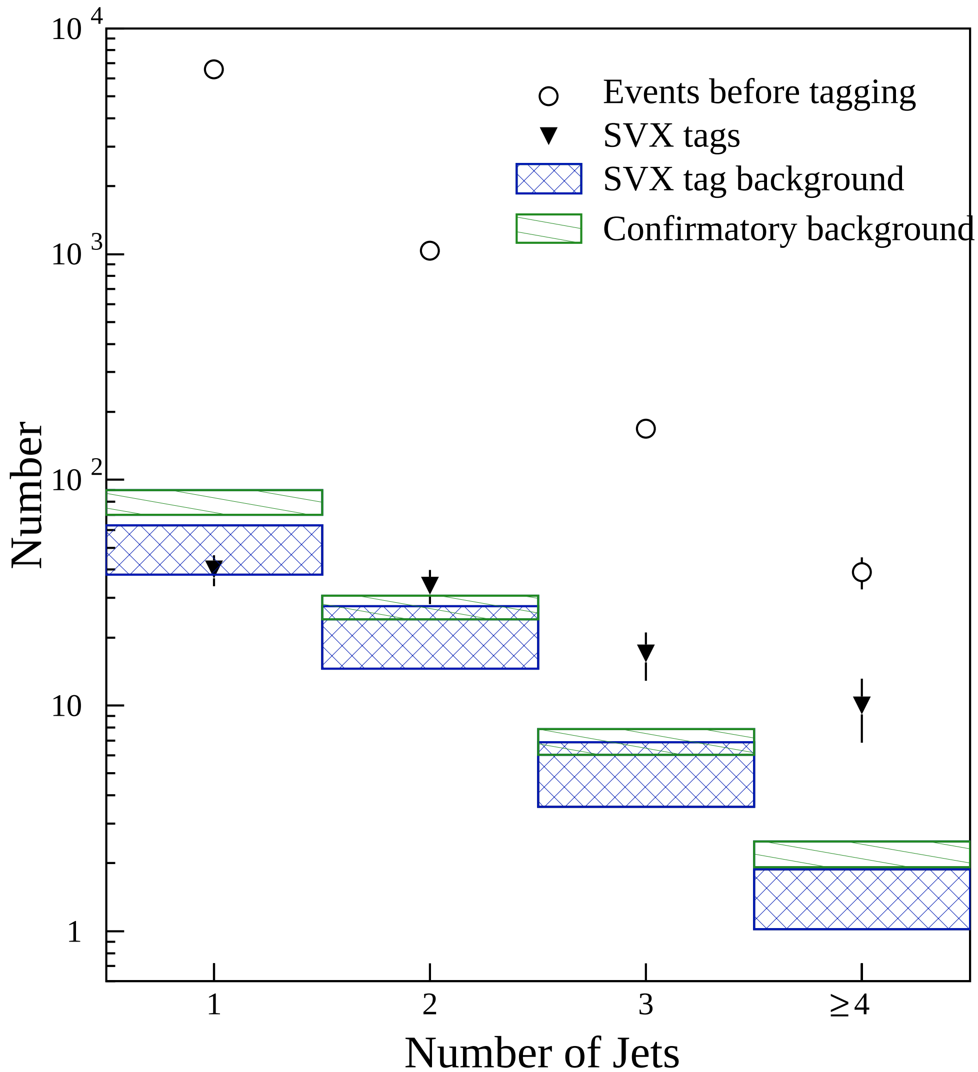


Figure 4.9: *The number of tagged jets (inverted triangles) in the W+multijet sample versus jet multiplicity. The expected background is bounded by the cross-hatched areas. The confirmatory background, which we expect to be an overestimate, is shown as the single hatched areas. The pre-tagged jet multiplicity distribution of all W+multijet events is shown as open circles. Top events are expected to appear as an excess above background in the $\geq$ 3-jet region.*

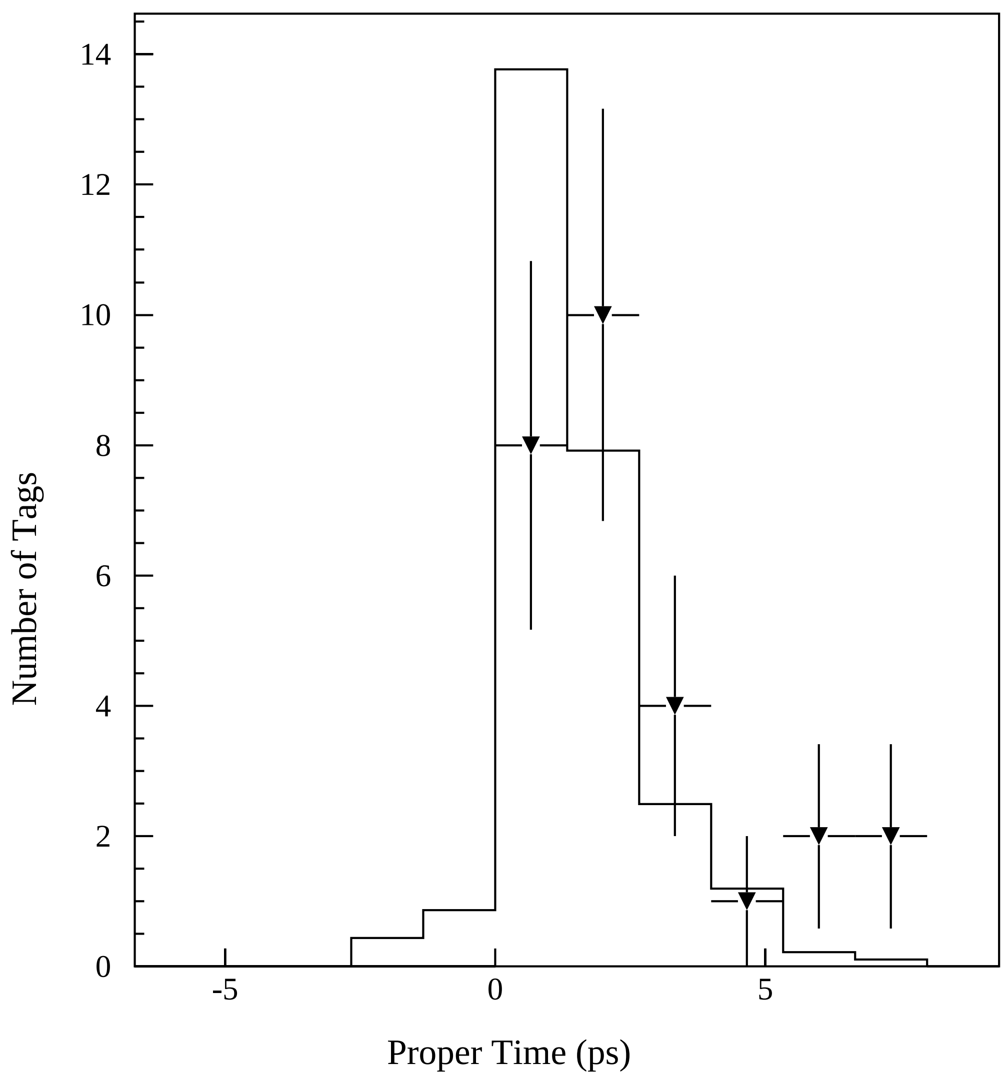


Figure 4.10: *The lifetime distribution of tags in the $W + \geq 3$ jet sample (inverted triangles) versus the lifetime distribution for tags in $t\bar{t}$ Monte Carlo events.*

number of *events* observed. The reason for this is that $t\bar{t}$ events are more likely than background events to contain more than one tag, hence our sensitivity to top is enhanced by counting tags. Since the number of tags in a background-only experiment is not a Poisson variable,[4] we may no longer apply Poisson statistics. We implement the correlations between tag multiplicities using a Monte Carlo program which performs a large number of background-only experiments with 204 pre-tagged events. The Monte Carlo program simulates the number of tags from each type of background ($Wb\bar{b}$, $Wc\bar{c}$, Wc, mistags, etc.). Multiple tag efficiencies are used to generate multiple tags. The significance is then the number of experiments with ≥ 27 tags divided by the number of experiments in the ensemble. The significance found using this technique is 2×10^{-5}, about 5 times larger than the pure Poisson probability. For a gaussian distributed variable, this probability corresponds to a 4σ deviation.

We now return to investigate whether the excess of tagged events over background in the background-dominated 2-jet bin is consistent with expectations from a mixture of top and background. Since the background calculation was normalized to the number of observed pre-tagged events under the assumption of no top content, it is an overestimate if top is present. Adjusting the background to account for this yields an excess of 15.5 events in the $\geq$3 jet due to $t\bar{t}$ production. Given this excess, we expect an additional 3.6 tagged events in the 2-jet bin from $t\bar{t}$ pro-

[4] The number of single tagged (N_1) and double tagged (N_2) events are both Poisson variables, but the weighted sum $N_1 + 2N_2$ is not. Thanks to G. Unal for this clarification.

duction. Also, single top production, an example of which is shown in figure 4.11, is expected to contribute ~ 0.6 tagged events. Such single top events are more likely to produce a lepton+2-jet signature than a lepton+3-jet signature because one b-jet tends to be of low energy and at small angle with respect to the beam [50]. The double tag rate in $t\bar{t}$ events is approximately 32%, bringing the expected number of tags from top to ~ 5 in the 2-jet bin. Figure 4.12 shows the expected distribution of tags from the sum of top and background versus the observed tags. The errors on the total include contributions from the uncertainty on the rescaled background and on the statistical error on the top content, both of which are shown individually to facilitate comparison with the the observed number of tags and to illustrate the bin-to-bin uncertainty correlations. In the 2-jet bin, the number of observed and expected tags from top and background agree to within errors.

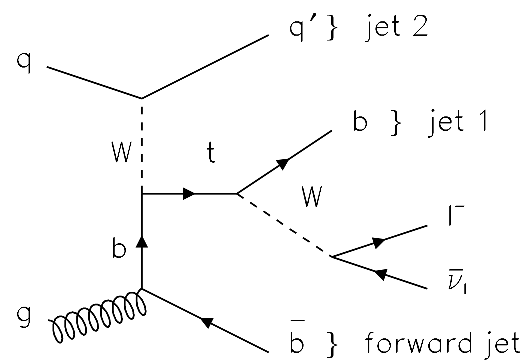


Figure 4.11: *An example of single top quark production.*

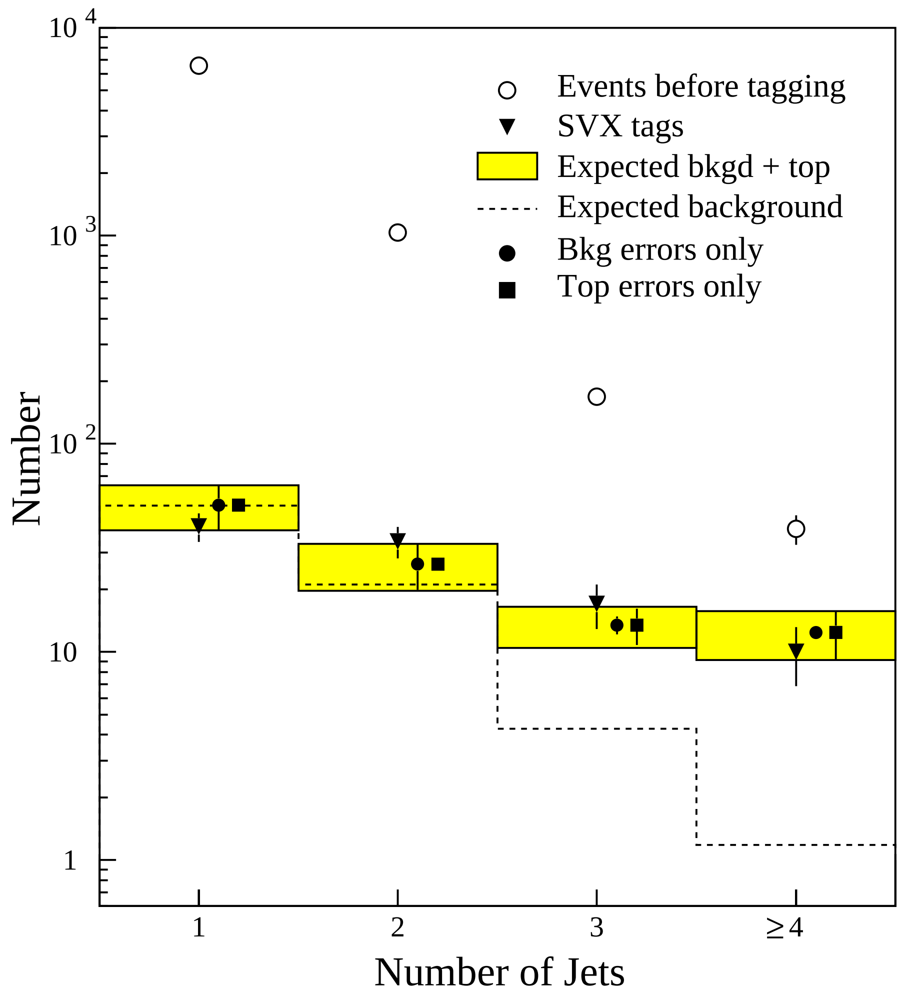


Figure 4.12: *The number of tagged jets (inverted triangles) in the W+multijet sample versus jet multiplicity. The shaded regions indicate the number of tags expected from background and top events. The top content is extracted from the excess of events in the $\geq$ 3-jet region. The errors on the filled circles include only the uncertainty on the background estimate. The errors on the filled squares include only the uncertainty on the top content. This uncertainty corresponds to the statistical uncertainty on the excess of observed events above background in the $\geq$ 3-jet region. The shaded bands include both uncertainties.*

Chapter 5

Identification of Top Events in the Lepton+Multijet Sample Using Kinematic Information

In chapter 4, we isolated $t\bar{t}$ events in the lepton+multijet sample by selecting events containing b quarks by identifying displaced b-hadron decay vertices within jets. In the present chapter, we isolate the $t\bar{t}$ signal in the lepton+multijet sample using solely the *kinematic* features of the events. By *kinematics* we mean the directions and energies of the final state particles and physics objects such as jets. There are many potential kinematic variables which effectively separate $t\bar{t}$ and background events [51]. The kinematic variables we use for this purpose are $cos\theta^*$, the inner product of the jet axis and the proton beam direction as measured in the center-of-momentum frame of the hard-scattering subprocess,[1] and the transverse energies E_T of the jets, where $E_T = E \sin\theta$. Here E is the jet energy and θ is the angle of the jet with respect to the $p\bar{p}$ beam. Figure 5.1 compares the predicted $cos\theta^*$ and jet E_T distributions of $t\bar{t}$ events and background QCD W+jet events. For this comparison,

[1]The hard subprocess is defined as the four-vector sum of the lepton momentum, jet momenta for jets with $E_T \geq 10$ GeV, and the vector $\not{E}_t$ assuming that the z-component of the neutrino momentum is zero.

we have selected lepton+$\geq$ 3-jet events with large $\not\!\!E_t$ from $t\bar{t}$ and QCD W+jet Monte Carlo samples using the event selection described in section 3.9. The $t\bar{t}$ sample was generated using the HERWIG Monte Carlo program with M_{top} = 170 GeV/c^2. The detector response was simulated using the full CDF detector simulation. The HERWIG Monte Carlo program is described in section 2.2.7. The QCD W+jet Monte Carlo sample is discussed in section 5.1. For the $cos\theta^*$ variable, figure 5.1 shows the maximum value of $|cos\theta^*|$ among the three leading jets; jets in $t\bar{t}$ events tend to have smaller $|cos\theta^*|$ values than jets in QCD background events. Since smaller values of $|cos\theta^*|$ correspond to the large angles with respect to the beam, we describe jets in $t\bar{t}$ events as more *central* than those in QCD W+jet events. We may understand somewhat more intuitively the $|cos\theta^*|$ distributions shown in figure 5.1: in $t\bar{t}$ events, we expect jet production to be approximately isotropic, leading to a flat distribution versus the solid angle $cos\theta^*$, whereas we expect jets produced via QCD to be approximately flat in η (where $\eta \sim -\ln(\tan\frac{\theta^*}{2})$) leading to a distribution peaked in the forward ($cos\theta^* \sim 1$) direction. Figure 5.1 compares the E_T distributions for the leading three jets ($E_T(1)$, $E_T(2)$, and $E_T(3)$); jets in $t\bar{t}$ events are harder, i.e. have larger E_T than those in QCD W+jet events. The scale for jet E_Ts in $t\bar{t}$ events is set by the top mass, and jet E_Ts are accordingly large. Figure 5.2 shows the separation between $t\bar{t}$ and background kinematics two-dimensionally by comparing the $E_T(2)$ versus $E_T(3)$ distributions for the same samples as in figure 5.1. Figures 5.1 and 5.2 motivate the following analysis, in which we use solely kinematic

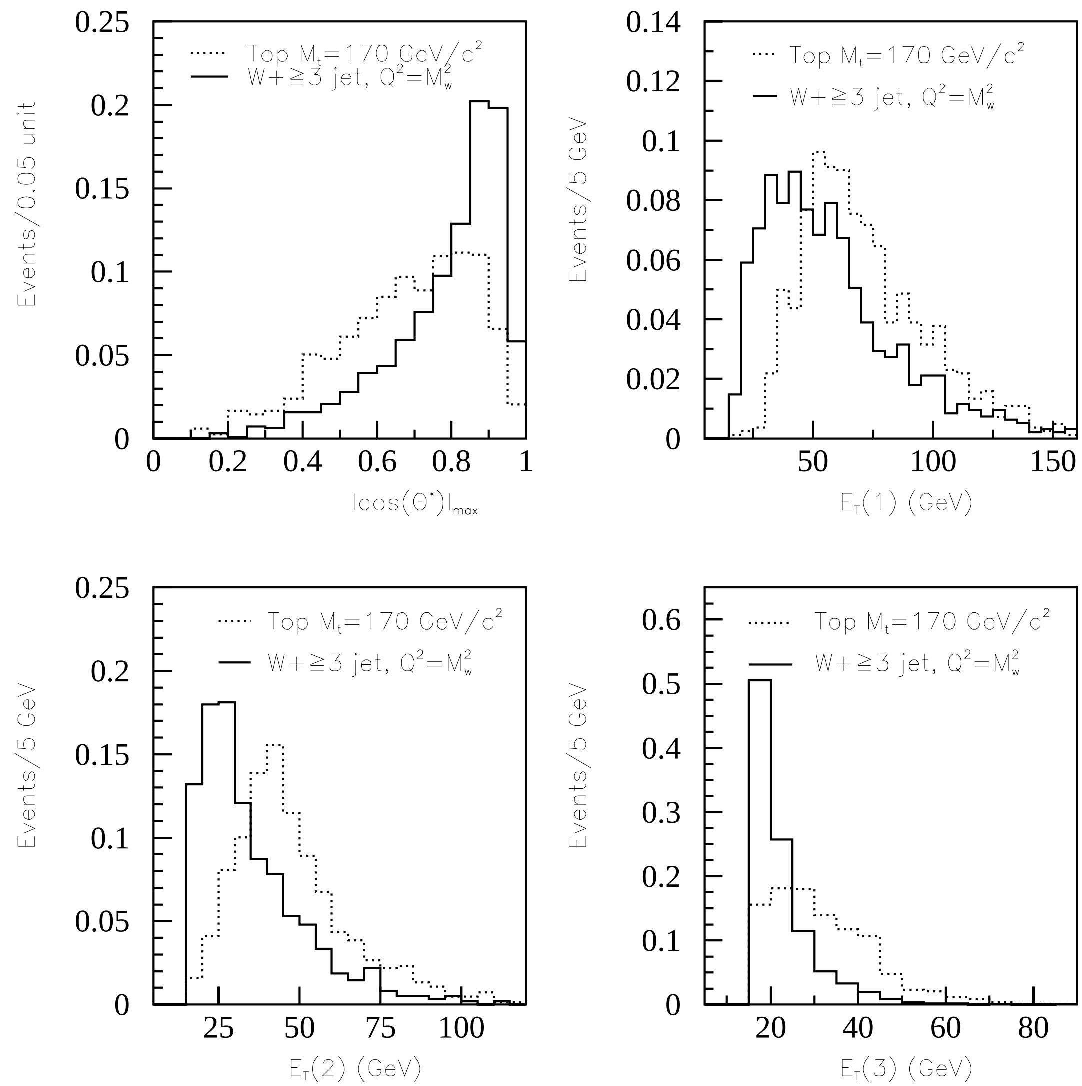


Figure 5.1: *Comparison of Monte Carlo predictions for kinematic variables used to separate top and background events. For events with ≥ 3 jets, distributions of $|cos\theta^*|_{max}$ and the transverse energies $E_T(1)$, $E_T(2)$, and $E_T(3)$ of the three leading jets are shown for $t\bar{t}$ (dashed, $M_{top} = 170\ GeV/c^2$) and QCD W+multijet events (solid).*

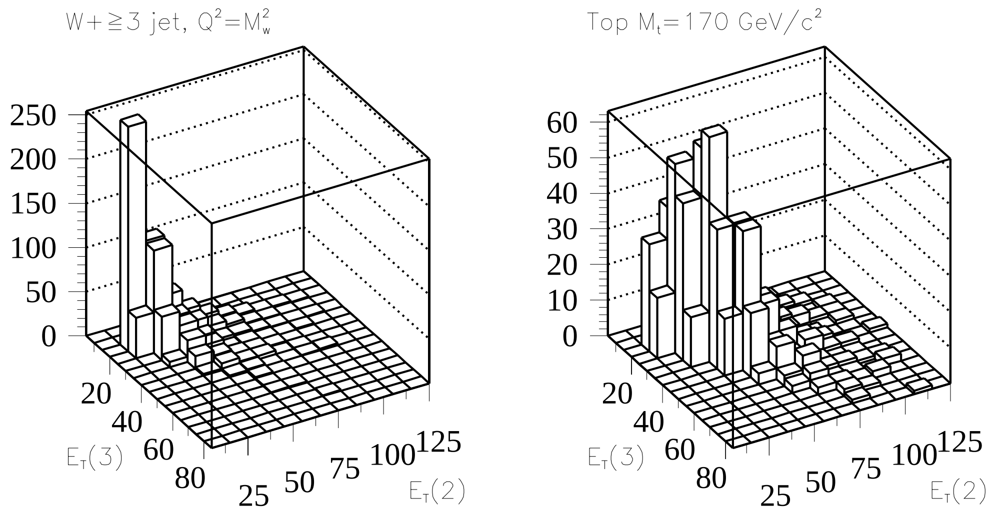


Figure 5.2: *Distributions of $E_T(3)$ versus $E_T(2)$ for QCD $W+\geq 3$ jet Monte Carlo events (left) and for top Monte Carlo events with $M_{top} = 170\ GeV/c^2$ (right).*

information to identify an enriched sample of $t\bar{t}$ events within the lepton+multijet sample.

Our approach is as follows. We verify that the Monte Carlo model we use to describe the dominant QCD W+jet background accurately describes the observed kinematic distributions of control samples which are expected to have low top content. We proceed to define a variable ($-\log L$) which separates the $t\bar{t}$ signal from the background based on jet energies. We use this variable in a shape analysis. The distribution of events in this variable is compared to the distribution expected from background events normalized to the observed number of events. If top is not present, these shapes should agree. We examine a region of this distribution in which $t\bar{t}$ events are expected to dominate, and compare the number of observed events in this region to the number expected from the shape of the background distribution. An excess in this region indicates the presence of top events in the sample. The choice of this region is based on a study of our expected sensitivity to the top signal using the $-\log L$ shapes in $t\bar{t}$ and background Monte Carlo events. This sensitivity study is performed prior to the examination of the data. The study accounts for both systematic uncertainties on the background shape and the effectiveness of using the $cos\theta^*$ variable in addition to jet energy information.

5.1 QCD W+jets Background Model

In order to separate the $t\bar{t}$ signal from background events using kinematic variables, we must have a reliable model for the kinematic characteristics of the dominant QCD

W+jet background. We model the kinematics of QCD W+jet events using event samples generated using the VECBOS Monte Carlo program [42] and simulated using the full CDF detector simulation. The VECBOS program is a parton-level Monte Carlo generator based on the leading order QCD matrix elements for production of a given multiplicity of final state partons in association with a W boson. In modeling the QCD $W+\geq$ 3-jet background, we use the W+3 parton matrix element calculation. The choice of momentum transfer scale $Q^2 = M_W^2$ in $\alpha_{\rm s}$ is discussed below. The formation of jets from partons is modeled using coherent parton shower evolution and cluster hadronization as implemented in the HERWIG Monte Carlo program [40]. The generation of events with jet multiplicities higher than the final state parton multiplicity (in this case three) is accomplished through the HERWIG parton evolution. Since the VECBOS matrix element calculation assumes that quarks are massless, it is necessary to specify a minimum $\Delta(R) = \sqrt{\Delta(\phi)^2 + \Delta(\eta)^2}$ separation between the final state partons and a minimum parton energy in order to avoid collinear and infrared divergences. We required a parton-parton separation of $\Delta(R) > 0.4$ at the generator level. Our jet clustering algorithm (see section 3.4) applies a clustering cone of $R = 0.4$ to the fully simulated events. To avoid biases in jet merging and jet separation introduced by using the same cone at generator level and simulated event reconstruction level, we additionally require $\Delta(R) > 0.7$ for the three leading jets in the event after detector simulation. This requirement is also applied to the data samples and other Monte Carlo samples used in this chapter and chapter 6.

We verify that the VECBOS model for QCD W+jet production is in agreement with data by comparing its predictions to observed kinematic distributions in data for samples expected to be dominated by QCD W+jet production and in which $t\bar{t}$ production is not expected to contribute at significant levels. We make this comparison in two samples. First, figure 5.3 compares the VECBOS prediction to data for lepton+jet events with exactly two jets. We show the predicted and observed distributions of the two leading jet E_Ts and $|cos\theta^*|$ for all jets. The agreement in this sample is excellent. For our second comparison, we select a lepton+ 3-jet sample in which events with a fourth jet with $E_T(4) > 8$ GeV have been excluded. The efficiency for these cuts on $t\bar{t}$ Monte Carlo events is 31.1$\pm$1.6%, while the efficiency for QCD W+jet Monte Carlo events is 78.4$\pm$1.3% (statistical errors only). Figure 5.4 compares the VECBOS prediction to data for this sample. We show the predicted and observed distributions of the three leading jet E_Ts and $|cos\theta^*|$ for all jets. The agreement in this sample is good.

5.2 The Likelihood Variable L

Based on figure 5.1, we observe that jets in $t\bar{t}$ events have larger values of E_T and smaller values of $|cos\theta^*|$ than jets in QCD W+jet events. We defer to section 5.3 the question of how to use the $cos\theta^*$ variable to separate $t\bar{t}$ and QCD W+jet events. In this section, we devise a single variable which incorporates information from the three leading jet E_Ts. Our motivation for devising a single variable is twofold.

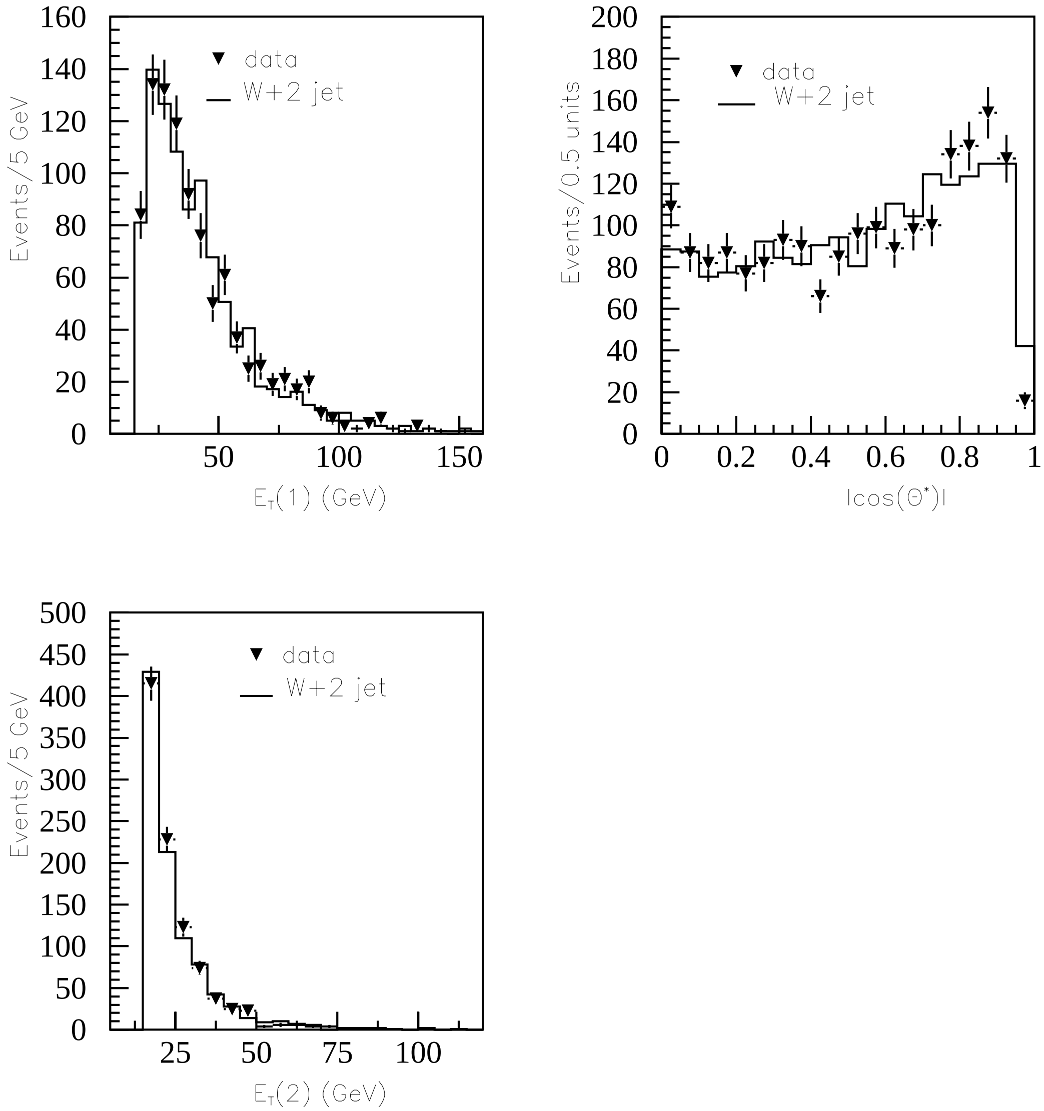


Figure 5.3: *Comparison of QCD W+2-jet Monte Carlo kinematic predictions to data in the lepton+2-jet sample. The distributions shown are the first (upper left) and second (lower left) leading jet transverse energies and the $|cos\theta^*|$ distribution for all jets (upper right). The Monte Carlo prediction is shown as the solid histogram and the data is shown as points. This sample is expected to be dominated by QCD W+jet production and the agreement is excellent.*

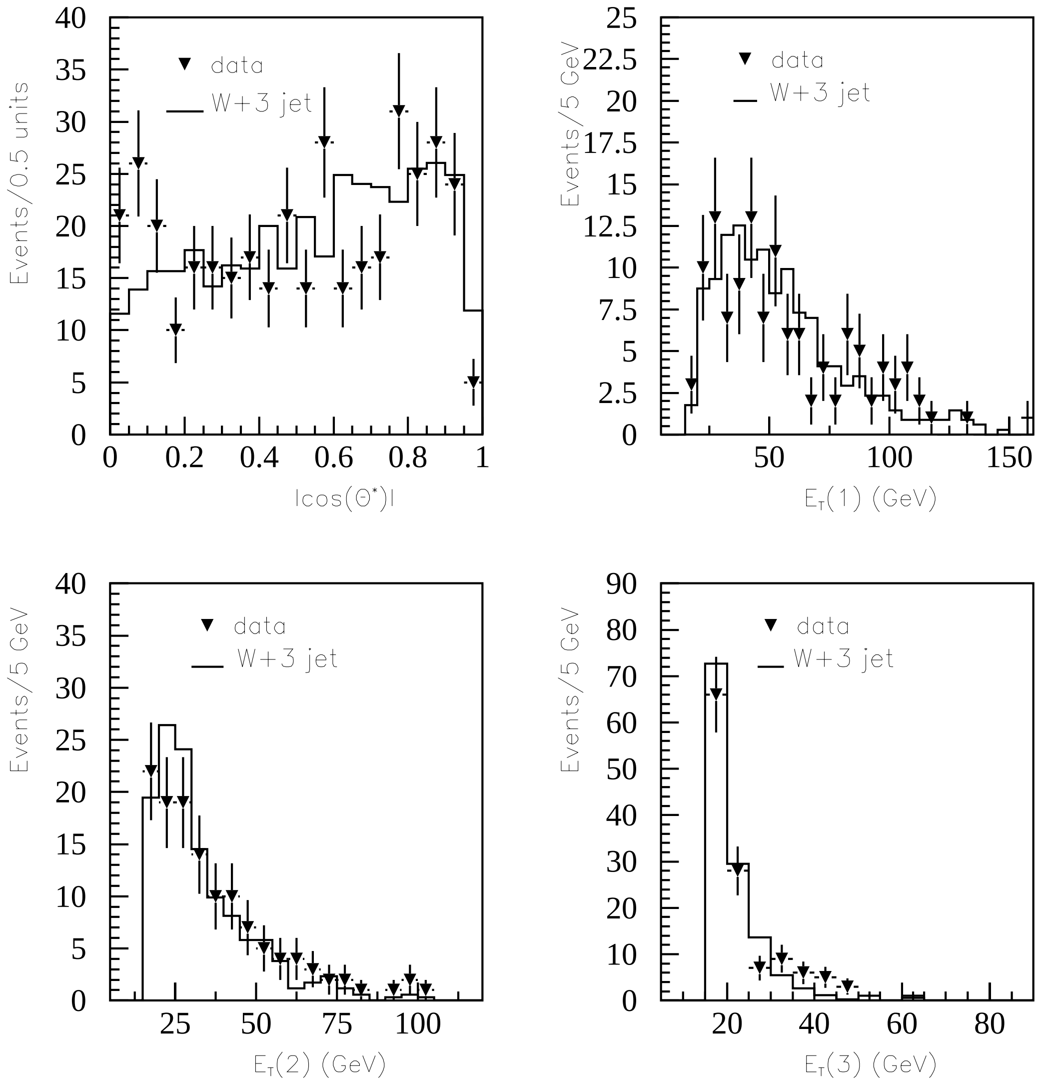


Figure 5.4: *Comparison of QCD W+3-jet Monte Carlo kinematic predictions to data in a subset of the lepton+3-jet sample in which we require that there be no fourth jet with $E_T > 8$ GeV. The distributions shown are the $|cos\theta^*|$ distribution for all jets (upper left) and the first (upper right), second (lower left) and third (lower right) leading jet transverse energies. The Monte Carlo prediction is shown as the solid histogram and the data is shown as points. This sample is expected to have low $t\bar{t}$ content, and the agreement is good.*

First, it is evident from figure 5.1 that there is potential for distinguishing the $t\bar{t}$ signal from background using each of the three jet E_T distributions, and it is to our advantage to use as much of this potential as possible. Second, in chapter 6, we shall use this variable to preserve much of the b-tagged $t\bar{t}$ signal while reducing the the background for b-tagged events.

The variable we adopt is the *background likelihood* L. For an observed jet with a given E_T, the likelihood represents the probability that the theoretical background E_T distribution would yield the observed value of E_T. We define the likelihood L_i for the three ordered E_T distributions $E_T(i)$ where $i = 1, 2, 3$ as follows:

$$L_i = \frac{1}{\sigma}\frac{d\sigma}{dE_T(i)} \tag{5.1}$$

where $\frac{1}{\sigma}\frac{d\sigma}{dE_T(i)}$ is ideally given by the normalized theoretical function which gives rise to the QCD W+jet E_T distributions shown in figure 5.1. Since we do not have analytic expressions for these functions, we choose a model for them and check that this model reasonably describes the E_T distributions derived from the Monte Carlo simulation. One reasonable model is a parent distribution of three uncorrelated, unordered jets (jets A, B, and C) of $E_T(A, B, C) \geq E_0$ each of which is described by a normalized exponential characterized by the single positive parameter α:

$$\frac{1}{\sigma}\frac{d\sigma}{dE_T(A,B,C)} = \alpha e^{-\alpha(E_T(A,B,C)-E_0)} \tag{5.2}$$

If we then order the density function

$$\frac{1}{\sigma}\frac{d^3\sigma}{dE_T(A)dE_T(B)dE_T(C)} = \alpha^3 e^{-\alpha(E_T(A)-E_0)} e^{-\alpha(E_T(B)-E_0)} e^{-\alpha(E_T(C)-E_0)} \tag{5.3}$$

we find that the following equations describe $E_T(1)$, $E_T(2)$, and $E_T(3)$:[2]

$$L_1 = \frac{1}{\sigma}\frac{d\sigma}{dE_T(1)} = 3\alpha\left(e^{-\alpha(E_T(1)-E_0)} - 2e^{-2\alpha(E_T(1)-E_0)} + e^{-3\alpha(E_T(1)-E_0)}\right) \quad (5.4)$$

$$L_2 = \frac{1}{\sigma}\frac{d\sigma}{dE_T(2)} = 6\alpha\left(e^{-2\alpha(E_T(2)-E_0)} - e^{-3\alpha(E_T(2)-E_0)}\right) \quad (5.5)$$

$$L_3 = \frac{1}{\sigma}\frac{d\sigma}{dE_T(3)} = 3\alpha\left(e^{-3\alpha(E_T(3)-E_0)}\right) \quad (5.6)$$

Given the parametrizations in equations 5.4 through 5.6, we may find the value of α which best fits the QCD W+jet E_T distributions shown in figure 5.1. We may variously fit to any single E_T spectrum or to any simultaneous combination of the three. The results for such fits are given in table 5.1, and are consistent, indicating that our simple parametrizations constitute a reasonable model of the E_T distributions predicted by the Monte Carlo simulation. We choose as our primary value for α the result fitting simultaneously to the leading three jet E_Ts, $\alpha = 0.0440$ GeV^{-1}. Figure 5.5 shows the QCD W+jet E_T spectra versus the parametrizations indicated by equations 5.4 through 5.6 using this value of α.

Equations 5.4 through 5.6 along with the value $\alpha = 0.0440$ determine the values

---

[2]The coefficients of exponential terms in the ordered distributions of an N-dimensional exponential density function are given by Pascal's triangle. For example, for four ordered variables, we have:

$$\frac{1}{\sigma}\frac{d\sigma}{dE_T(1)} = 144\alpha\left(e^{-\alpha(E_T(1)-E_0)} - 3e^{-2\alpha(E_T(1)-E_0)} + 3e^{-3\alpha(E_T(1)-E_0)} - e^{-4\alpha(E_T(1)-E_0)}\right)$$

$$\frac{1}{\sigma}\frac{d\sigma}{dE_T(2)} = 48\alpha\left(e^{-2\alpha(E_T(2)-E_0)} - 2e^{-3\alpha(E_T(2)-E_0)} + e^{-4\alpha(E_T(2)-E_0)}\right)$$

$$\frac{1}{\sigma}\frac{d\sigma}{dE_T(3)} = 48\alpha\left(e^{-3\alpha(E_T(3)-E_0)} - e^{-4\alpha(E_T(3)-E_0)}\right)$$

$$\frac{1}{\sigma}\frac{d\sigma}{dE_T(4)} = 4\alpha\left(e^{-4\alpha(E_T(4)-E_0)}\right)$$

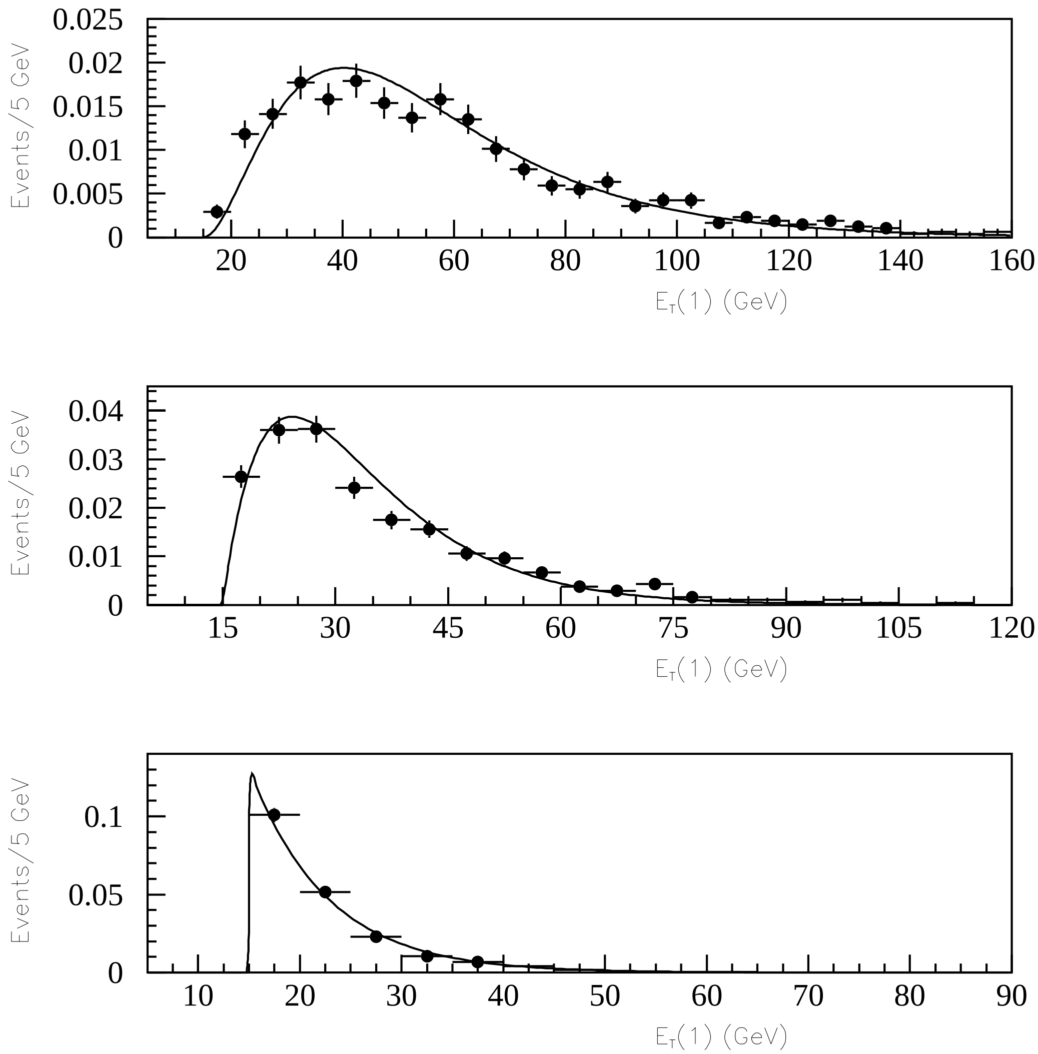


Figure 5.5: *Comparison of the QCD W+jet Monte Carlo E_T distributions for $W+\geq 3$ jet events (points with errors) to the parametrizations (solid curve) given by equations 5.4 through 5.6 calculated using the value $\alpha = 0.0440$. This value of α best describes all three E_T distributions simultaneously .*

Distributions Used In Fit	α, GeV^{-1}
$E_T(1)$	0.0431±0.0011
$E_T(2)$	0.0419±0.0013
$E_T(3)$	0.0469±0.0017
$E_T(2)$,$E_T(3)$	0.0436±0.0007
$E_T(1)$,$E_T(2)$,$E_T(3)$	0.0440±0.0010

Table 5.1: *Results of fitting various combinations of E_T distributions for QCD W+jet Monte Carlo events to the parametrizations in equations 5.4 through 5.6. The results are consistent for all combinations.*

of the individual jet likelihoods L_i (i=1,2,3) for the three leading jets in an event. Jets in $t\bar{t}$ events populate E_T regions in which the fraction of QCD W+jet events is small. Thus typical L_i values for jets in $t\bar{t}$ events are expected to be smaller than those found in QCD W+jet events. Rather than using L_i directly, we henceforth use the negative logarithms $-\log L_i$ so that $t\bar{t}$ events continue to appear at "harder" values than background QCD W+jet events. Figure 5.6 shows the $-\log L_i$ distributions for the same QCD W+jet and the $t\bar{t}$ Monte Carlo samples as shown in figure 5.1.

We now seek to use information from all three leading jet E_Ts in the definition of the single likelihood variable L. We generalize the logarithm of equation 5.1 as follows:

$$-\log L = -w_1 \log L_1 - w_2 \log L_2 - w_3 \log L_3 \qquad (5.7)$$

where the weights w_i allow us to give more weight to the information from one jet E_T over another. Choosing a set of weights w_i specifies a direction in $\log L_1 - \log L_2 - \log L_3$ space, which we characterize by the two angles ϕ_{23} and θ_1 in equations 5.8

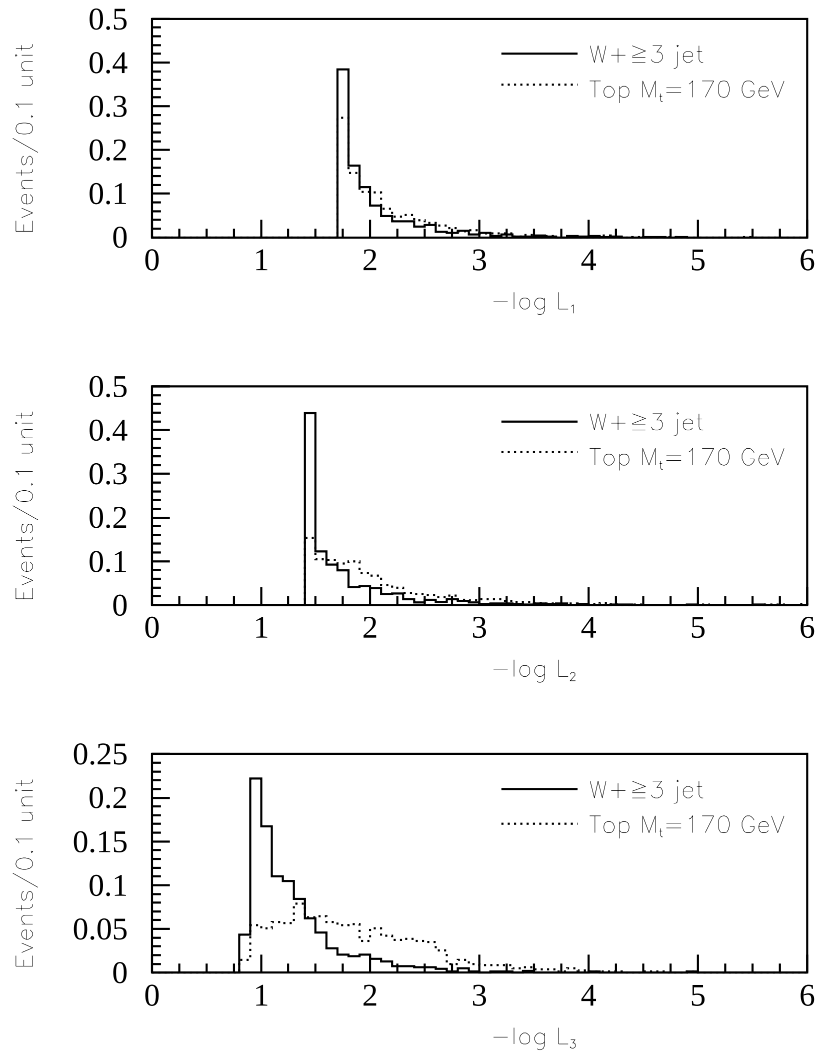


Figure 5.6: *Comparison of the predicted distributions of the negative logarithm of individual jet likelihoods $-logL_i$ for QCD W+jet and $t\bar{t}$ Monte Carlo events. For example, $-\log L_3$ is determined from the third leading jet E_T ($E_T(3)$) via equation 5.6 with* $\alpha = 0.0440$.

through 5.10.

$$
\begin{aligned}
w_1 &= \sin\theta_1 && (5.8)\\
w_2 &= \cos\theta_1 \sin\phi_{23} && (5.9)\\
w_3 &= \cos\theta_1 \cos\phi_{23} && (5.10)
\end{aligned}
$$

We choose the angles ϕ_{23} and θ_1 as follows. A given choice of angles specifies a likelihood variable $-\log L(\phi_{23}, \theta_1)$, for which we find distributions in $t\bar{t}$ and QCD W+jet Monte Carlo events. Using these distributions, we calculate the fraction S of $t\bar{t}$ events and the fraction B of QCD W+jet events which remain after applying a cut in $-\log L(\phi_{23}, \theta_1)$. We find the maximum value of $S/\sqrt{B}$ for this set of angles by successively increasing the cut. We choose to maximize the quantity $S/\sqrt{B}$ because it measures the expected statistical significance of an observation of a signal in the absence of systematic uncertainties. From the distributions of L_i, we expect that $w_3 > w_2 > w_1$ since the separation between $t\bar{t}$ and QCD W+jet distributions is greatest for L_3 and least for L_1. We begin by choosing to use information only from L_3 ($\phi_{23} = 0$,$\theta_1 = 0$), then scan in ϕ_{23} with fixed $\theta_1 = 0$. The maximum value of $S/\sqrt{B}$ as a function of ϕ_{23} is shown in figure 5.7 and peaks at $\phi_{23} = 0.1 \times \frac{\pi}{2}$. With this value of ϕ_{23} fixed, we perform a similar maximization scanning in the angle θ_1, as shown in figure 5.8. Since there is little or no gain in the maximum $S/\sqrt{B}$ as we increase θ_1, we choose $\theta_1 = 0$ in the definition of the likelihood variable.

We note that it is possible to select the vector w_i in other ways. We confirm our results using one such method, the Fischer discriminant technique. This technique

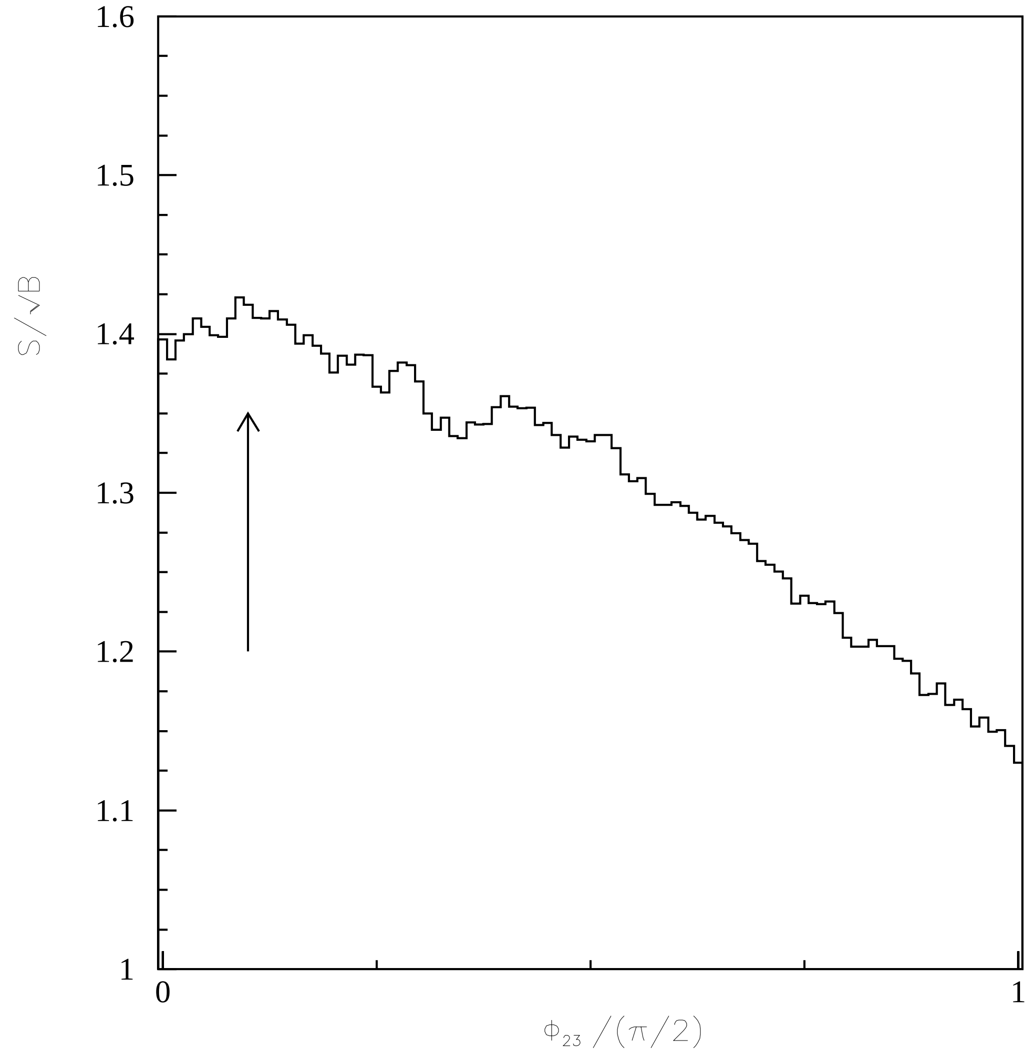


Figure 5.7: *Determination of the relative weights w_3 and w_2 given to information from $E_T(3)$ and $E_T(2)$ in the composition of the likelihood variable $-\log L$. The angle ϕ_{23} parametrizes the weights via equations 5.9 and 5.10. We have taken the weight for information from $E_T(1)$ to be zero ($\theta_1 = 0, w_1 = 0$). Figure shows the maximum value of $\frac{S}{\sqrt{B}}$ as described in section 5.2 as function of ϕ_{23}. The choice of $\phi_{23} = 0.1 \times \frac{\pi}{2}$ is motivated by the peak of this variable as indicated by the arrow.*

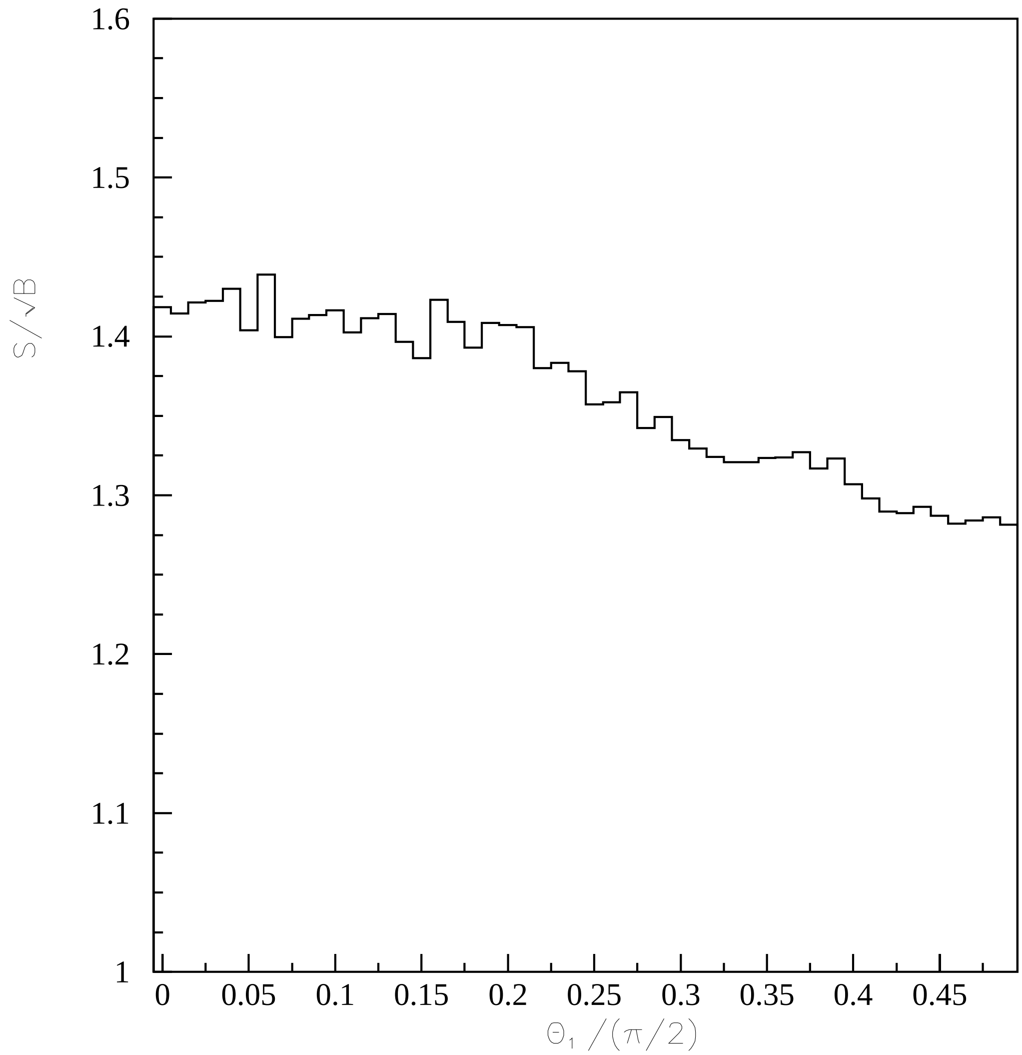


Figure 5.8: *Determination of the weight* w_1 *given to information from* $E_T(1)$ *in the composition of the likelihood variable* $-\log L$ *with the angle* ϕ_{23} *fixed at* $0.1 \times \frac{\pi}{2}$*. The angle* θ_1 *parametrizes the weight* w_1 *via equation 5.8. Figure shows the maximum value of* $\frac{S}{\sqrt{B}}$ *as described in section 5.2 as function of* θ_1*. We choose* $\theta_1 = 0$ *based on this figure.*

finds the vector w_i which maximizes the significance of the separation of the means of two distributions. The vector derived using this technique is $w_1 = 0.18$, $w_2 = 0.46$, $w_3 = 0.87$. Although the numerical solution is not identical, the Fischer discriminant technique confirms the general feature $w_3 > w_2 > w_1$.

In summary, we have chosen a single variable, $-\log L$, which is derived from the three leading jet transverse energies in lepton+$\geq$3 jet events. The relative weight of the three individual jet likelihoods L_i has been selected to maximize the signal over background statistical significance $S/\sqrt{B}$ in $t\bar{t}$ and background Monte Carlo samples. Equation 5.11 summarizes the $-\log L$ variable derived via this maximization. The individual jet likelihoods and numerical values necessary to calculating $-\log L$ are reiterated in equations 5.12 through 5.16. Figure 5.9 shows the distribution of this variable for QCD W+jet and top Monte Carlo events.

$$
\begin{aligned}
-\log L &= -\sin(\phi_{23})\log(L_2) - \cos(\phi_{23})\log(L_3) && (5.11)\\
L_2 &= 6\alpha\left(e^{-2\alpha(E_T(2)-E_0)} - e^{-3\alpha(E_T(2)-E_0)}\right) && (5.12)\\
L_3 &= 3\alpha\left(e^{-3\alpha(E_T(3)-E_0)}\right) && (5.13)\\
\phi_{23} &= 0.1 \times \frac{\pi}{2} && (5.14)\\
\alpha &= 0.0440\ \mathrm{GeV}^{-1} && (5.15)\\
E_0 &= 15\ \mathrm{GeV} && (5.16)
\end{aligned}
$$

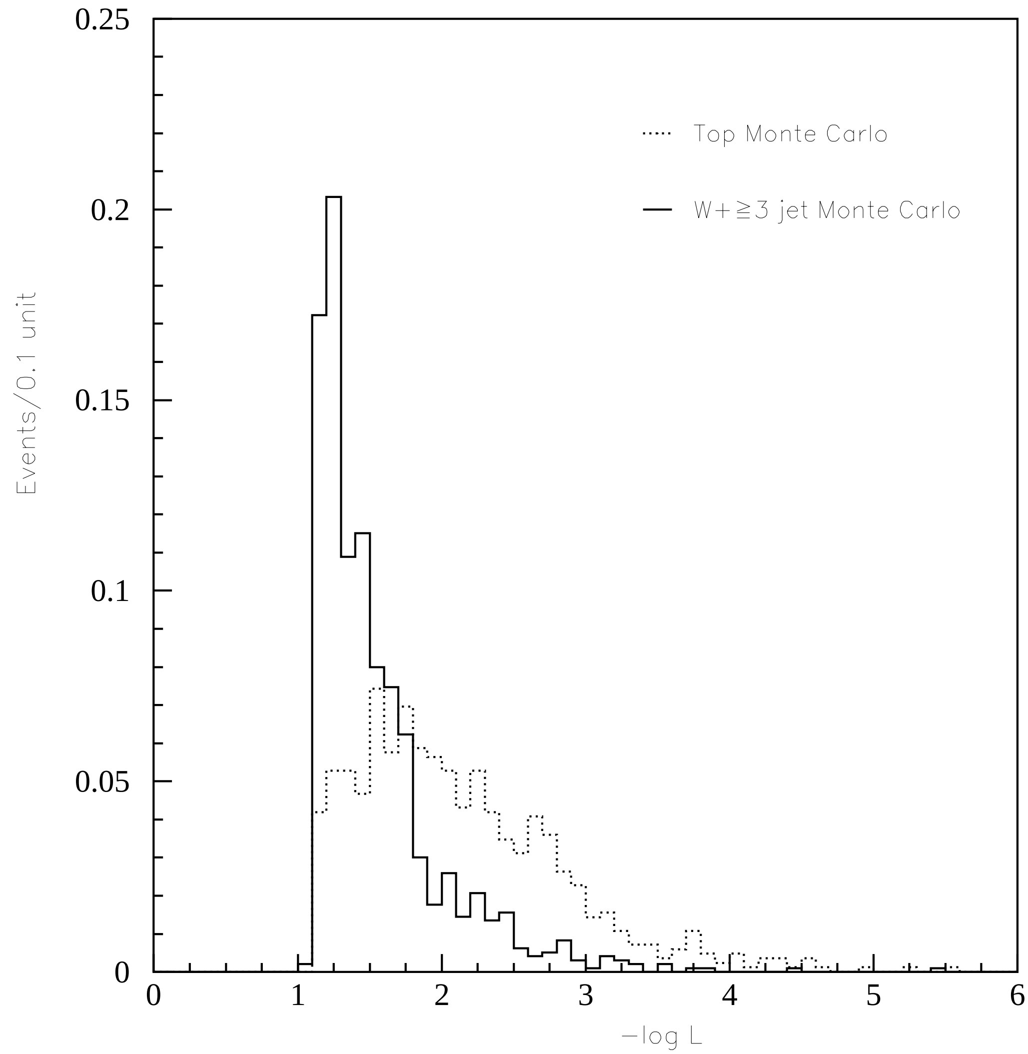


Figure 5.9: $-\log L$ *distribution for* $t\bar{t}$ *(dashed) and QCD* W*+jet (solid) Monte Carlo events passing the lepton+multijet requirements.*

5.3 The Central Sample

The likelihood variable $-\log L$ uses jet energies to distinguish $t\bar{t}$ and background events; we may also use the $cos\theta^*$ variable for this purpose, as previously illustrated in figure 5.1. Based on the distributions of $cos\theta^*$ in $t\bar{t}$ and QCD W+jet events, we select a top-enriched subset of the lepton+$\geq$ 3-jet sample by requiring that all three leading jets have $|cos\theta^*| < 0.7$. Since this requirement corresponds to selecting events in which the three leading jets are generally at large angles with respect to the beam, we refer to this top-enriched subsample as the *central sample.*[3] The efficiency of the central sample selection is (23.8$\pm$ 1.4)% for QCD W+jet Monte Carlo events and (49.4$\pm$ 1.7)% for $t\bar{t}$ Monte Carlo events (statistical errors only). Although this selection does not significantly increase $S/\sqrt{B}$ (defined in the same way as in section 5.2), it does increase S/B. The quantity S/B is the significance of an observation of a signal when the systematic error on the background ($f_{syst} * B$) is much larger than the statistical error ($\sqrt{B}$) on the background. Here f_{syst} is the fractional systematic error on the background. It is reasonable to consider the possibility that systematic errors on the background are non-negligible and that therefore the central sample may be a better sample for identifying a $t\bar{t}$ signal than the full lepton+$\geq$3 jet sample.[4] In section 5.5 we choose a cut in $-\log L$ based on statistical and systematic errors on the background and compare the performance

[3] We adopt the name "central sample" rather than "signal sample" in order to avoid confusion with the use of the word "signal" in other contexts.

[4] In our case, the systematic error on the background and the $\sqrt{B}$ statistical error on the background are comparable, as will be shown in section 5.4.

of the central sample to the full lepton+$\geq$ 3-jet sample in deciding which to use as our primary selection. Figure 5.10 shows the distribution of $-\log L$ for $t\bar{t}$ and background events passing the central sample requirements.

5.4 Sytematic Uncertainties on $-\log L$ in QCD W+jet Events

We consider sources of systematic uncertainty in the shape of the background QCD W+jet $-\log L$ distributions predicted by the VECBOS Monte Carlo and shown in figures 5.9 and 5.10. The variable we use to quantify the shape of the background distribution is ϵ_L, the fraction of events passing the lepton+multijet selection which are also expected to pass a given $-\log L$ cut. We may also define ϵ_L for the central subsample. In our shape analysis, ϵ_L is used to estimate the background in regions dominated by top events. First, we gauge uncertainty in event kinematics due to the lack of higher order contributions in the VECBOS calculation by varying the momentum transfer scale Q^2 in α_s, which is not uniquely defined in the fixed-order VECBOS calculation. Reasonable choices for Q^2 are M_W^2 and $< p_T^2 >$, the mean parton p_T^2. For the former choice, Q^2, and hence α_s is independent of the event kinematics. For $Q^2 =< p_T >^2$, events with larger parton energies have a larger Q^2 and thus a smaller coupling strength α_s (see equation 1.1). In general the population of events decreases faster with jet E_T for the choice $Q^2 =< p_T >^2$ than for the choice $Q^2 = M_W^2$; the former choice yields a "softer" energy distribution. Figures 5.11a and 5.11d compare the $-\log L$ distributions for the two choices of Q^2 in the lepton+$\geq$ 3-jet sample and the central sample respectively. We have

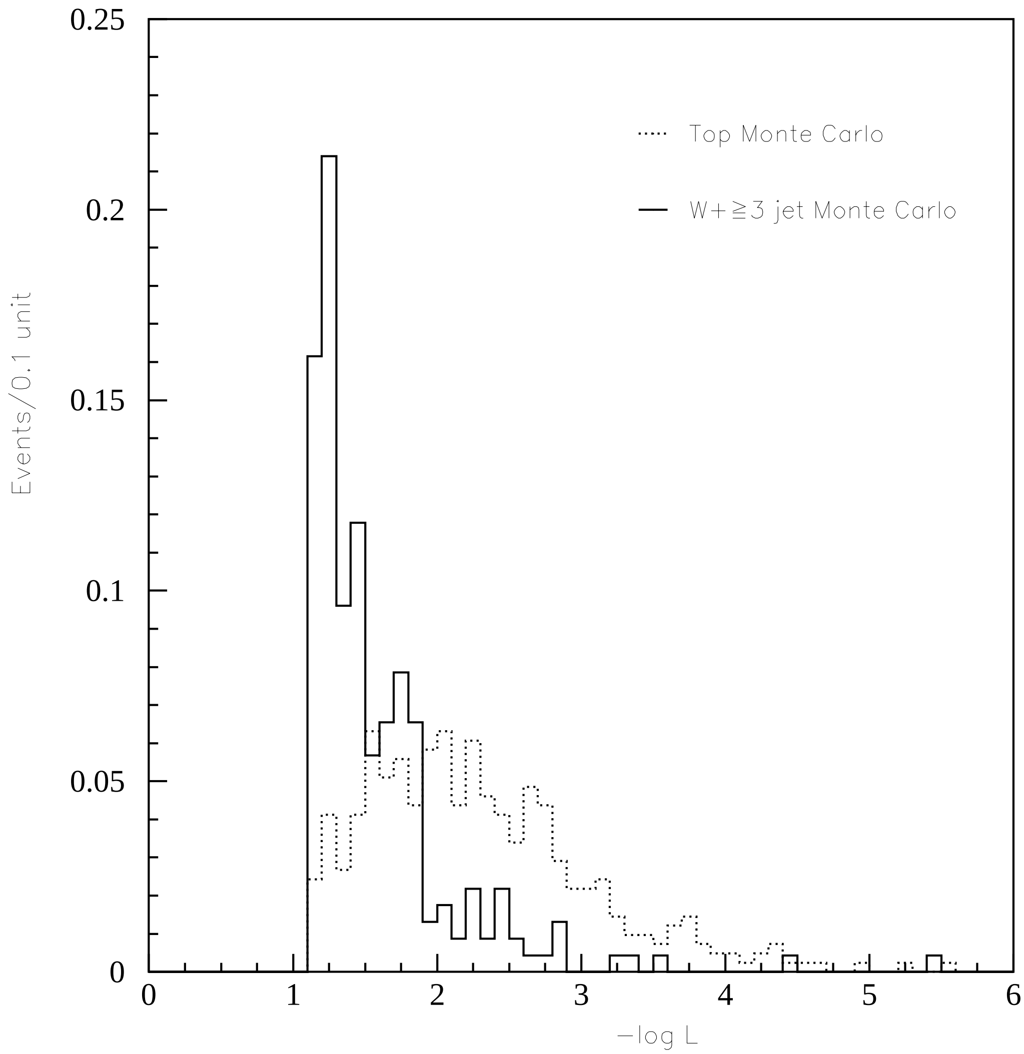


Figure 5.10: $-\log L$ *distribution for* $t\bar{t}$ *(dashed) and* QCD W*+jet (solid) events passing the central sample requirements* ($|cos\theta^*| < 0.7$ *for the three leading jets) in addition to the lepton+multijet selection.*

chosen the more conservative $Q^2 = M_W^2$ for our default model of QCD W+jet kinematics. The quantity ϵ_L for the default choice $Q^2 = M_W^2$ is calculated from figure 5.9 for the lepton+multijet requirement and from figure 5.10 for the central sample selection. The quantity ϵ_L for $Q^2 =< p_T >^2$ is calculated from the dashed curves in figures 5.11a and 5.11d. We estimate the systematic uncertainty on ϵ_L due to the lack of higher order terms in the VECBOS calculation to be the fractional difference $(\epsilon_L(Q^2 =< P_T >^2) - \epsilon_L(Q^2 = M_W^2))/\epsilon_L(Q^2 = M_W^2)$ for each value of $-\log L$. This fractional difference is shown as the points in figure 5.12a and 5.12d for the lepton+$\geq$ 3-jet and central samples respectively. Note that the fractional difference is always negative, as expected from the softer kinematics obtained using $Q^2 =< p_T >^2$.

The second systematic uncertainty we consider is due to uncertainty in the jet energy scale. As discussed in section 3.4, the jet energy scale uncertainty is described by equation 3.1, which we repeat here as equation 5.17.

$$\frac{\Delta(E_T)}{E_T} = (2\% \oplus (2.2\% + 60\%/E_T)) \oplus 10\% \tag{5.17}$$

In order to measure the effect of jet energy scale variations, we scale each jet in Monte Carlo events by $(1 \pm \frac{\Delta(E_T)}{E_T})$. Since this variation may alter jet multiplicity, it is applied before jet counting. Events passing the lepton+$\geq$ 3-jet and central sample requirements are selected. The normalized $-\log L$ distributions for the resulting samples are shown in figures 5.11b and 5.11e for increased $(+\frac{\Delta(E_T)}{E_T} \sim +10\%)$ jet energies and in figures 5.11c and 5.11f for reduced $(-\frac{\Delta(E_T)}{E_T} \sim -10\%)$ jet energies.

The systematic uncertainty on ϵ_L is taken to be the fractional difference between ϵ_L in these scaled plots and the default calculation. These fractional differences are shown in figure 5.12b, c, e, and f. Increasing jet energies leads to larger values of ϵ_L while reducing jet energies leads to smaller values of ϵ_L.

The systematic uncertainties on ϵ_L shown as points in figure 5.12 exhibit statistical variations. In sections 5.5 and 5.7, we use these systematic uncertainties in determining a $-\log L$ requirement designed to render a significant $t\bar{t}$ excess over background and in estimating the uncertainty on the QCD W+jet background passing this $-\log L$ requirement. For these purposes, we reduce the effect of statistical variations by using the "smoothed" curves shown as the shaded regions in figure 5.12 to represent the size of the systematic uncertainty as a function of $-\log L$.

5.5 Choosing a Likelihood Cut

We seek to define a cut in the likelihood variable $-\log L$ which can best isolate a $t\bar{t}$ signal in excess of the QCD W+jet background. We define a quantity, *Signif*, which estimates our expected sensitivity in distinguishing a $t\bar{t}$ signal from the background for a given cut in $-\log L$. *Signif* is calculated from Monte Carlo samples of $t\bar{t}$ and QCD W+jet events in a "model experiment." The $-\log L$ cut which maximizes *Signif* is the optimal selection. For this maximization, the normalization for the $t\bar{t}$ signal is taken from the $t\bar{t}$ cross section and the normalization of the background is derived from the observed number of lepton+$\geq$ 3-jet events. Of the 204 observed events in the lepton+$\geq$ 3-jet sample (see section 3.9), 162 events pass the jet-jet

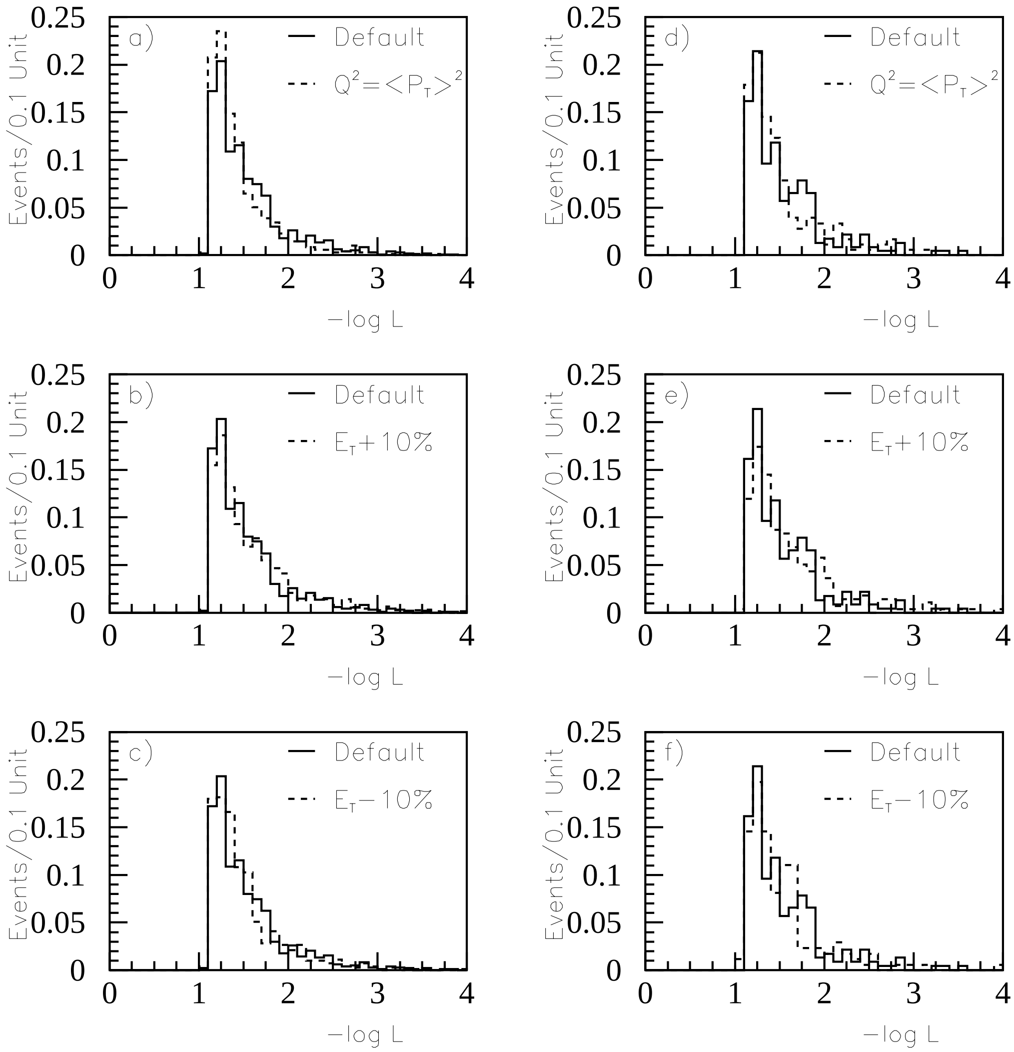


Figure 5.11: *Comparison of the $-\log L$ distribution in QCD W+jet Monte Carlo events for reasonable variation in systematic uncertainties. Figures in the left column are for the full lepton+$\geq$ 3-jet sample and figures in the right column are for the central sample. The "default" Monte Carlo sample indicated by the solid histogram is generated with $Q^2 = M_W^2$ and the no scaling of jet energies. The dashed curves indicate variations on these values as follows. Figures (a) and (d) compare the default choice of $Q^2 = M_W^2$ to the softer distribution obtained using $Q^2 =< p_T >^2$. Figures (b) and (e) show the effect of scaling jet energies by $\sim +10\%$ according to equation 5.17. Figures (c) and (f) show the effect of scaling jet energies by $\sim -10\%$. All histograms have been normalized to unit area.*

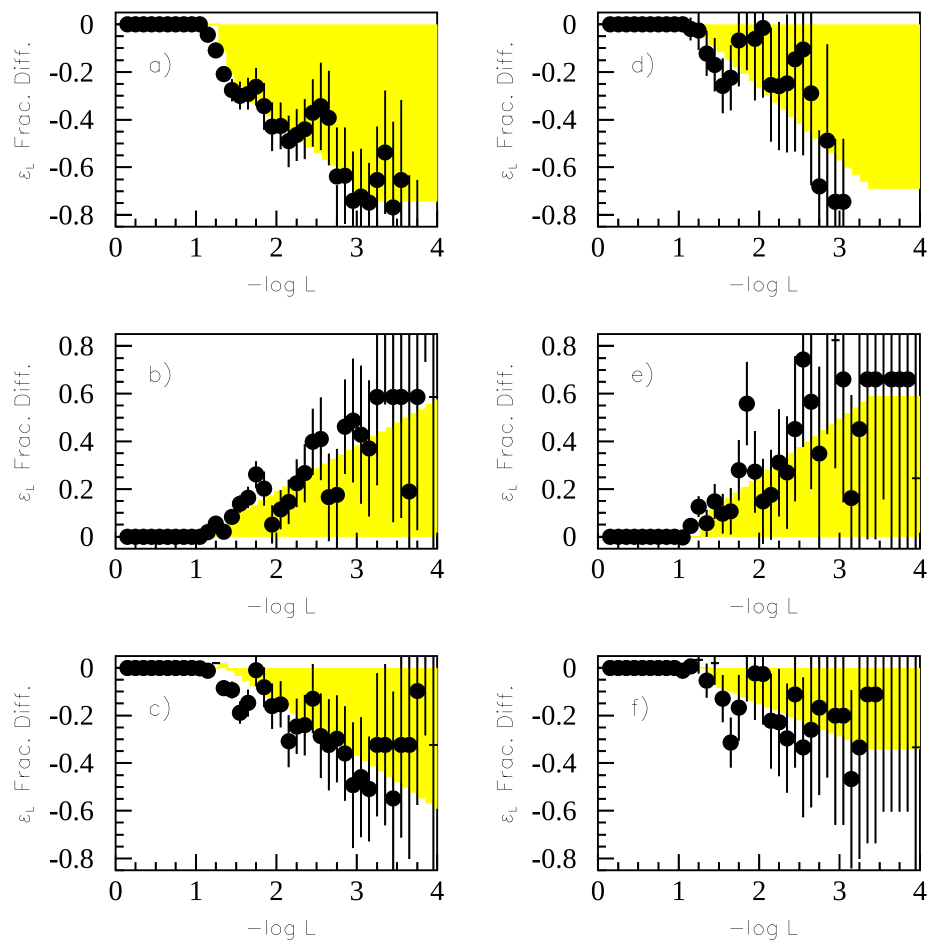


Figure 5.12: *We estimate the systematic errors on ϵ_L, the fraction of background events with $-\log L$ greater than a given value, from the curves shown. The value of ϵ_L is calculated using the default QCD W+jet background Monte Carlo and also using a variant of the default. These ϵ_L calculations are made using the distributions in figure 5.11. Points in the plots indicate fractional differences between the variant and default ϵ_L values as a function of $-\log L$. Errors on the points are from Monte Carlo statistics. The shaded areas are "smoothed" from these points. Figures in the left column are for the full lepton+$\geq$ 3-jet sample and figures in the right column are for the central sample. Fractional difference in ϵ_L with respect to the default is shown for: (a) and (d) $Q^2 = < p_T >^2$; (b) and (e) scaling jet energies by $\sim +10\%$; (c) and (f) scaling jet energies by $\sim -10\%$.*

$\Delta(R) > 0.7$ requirement. The quantity $N_W = 162$ sets the overall normalization: we choose a top content of $N_{top} = 40$ events based on the results of section 3.11, and ascribe the remaining events to QCD W+jet production: $N_{QCD\ W+jets} = N_W - N_{top} = 122$. The remainder of this discussion uses only quantities determined from Monte Carlo samples. The mean number of $t\bar{t}$ and QCD W+jet events which pass a given cut in $-\log L$ are derived from these normalizations and the predicted $-\log L$ distributions for $t\bar{t}$ and background events shown in figures 5.9 and 5.10. The number of background events passing the $-\log L$ cut is given by $\epsilon_L(N_{QCD\ W+jets})$. We define the quantity ϵ_L^{top}, the fraction of $t\bar{t}$ events in the lepton+ $\geq$ 3-jet sample passing a given $-\log L$ cut. The mean absolute number of $t\bar{t}$ events we expect to pass a given cut is $\epsilon_L^{top} * N_{top}$. The mean expected number of background plus $t\bar{t}$ events passing this cut is Nsig:

$$\text{Nsig} = \epsilon_L * N_{QCD\ W+jets} + \epsilon_L^{top} * N_{top} \tag{5.18}$$

For the background calculation in this model experiment, we assume that all of the events in the lepton+$\geq$ 3-jet sample (i.e. N_W events) are background events, and predict from the QCD W+jet $-\log L$ spectrum the expected number of background events passing the chosen $-\log L$ cut, Nsig:[5]

$$\text{Nbkg} = \epsilon_L * N_W \tag{5.19}$$

[5]Nbkg is not the "true background," which is given by $\epsilon_L N_{QCD\ W+jets}$, but its calculation emulates the calculation we will use in section 5.7, in which we normalize the background estimation to the total observed number of events, N_W.

In the presence of both statistical ($\sqrt{Nbkg}$) and systematic ($f_{syst} * \mathrm{Nbkg}$) errors on the calculated background, *Signif* as defined in equation 5.20 reasonably models the significance of a signal excess, where f_{syst} is the fractional systematic error on the estimated background.

$$Signif = \frac{\mathrm{Nsig} - \mathrm{Nbkg}}{\sqrt{\mathrm{Nbkg} + (f_{syst} * \mathrm{Nbkg})^2}} \tag{5.20}$$

For the calculation of *Signif*, f_{syst} is the sum in quadrature of the smoothed systematic uncertainties shown in figure 5.12 and the fractional uncertainty in ϵ_L due to QCD W+jet Monte Carlo statistics. The total uncertainty thus derived is asymmetric, where the negative uncertainty exceeds the positive uncertainty. We treat f_{syst} as symmetric with its magnitude given by the positive uncertainty. We adopt this approach for two reasons. First, it is unreasonable to expect positive deviations in ϵ_L associated to our choice of Q^2 scale since we have already taken as the default the most conservative reasonable choice of scale ($Q^2 = M_W^2$). This uncertainty is thus purely negative. Second, we are chiefly concerned with the maximum reasonable positive deviation of ϵ_L from its nominal value, since this positive uncertainty quantifies the upper limit of our expected background. The shaded curve in figure 5.13 shows the quantity *Signif* as a function of $-\log L$ for the lepton+ $\geq$ 3-jet requirement. The unshaded curve shows *Signif* if we neglect the systematic error term (i.e. $f_{syst} = 0$). A reasonable choice of cut is $-\log L > 2.5$.

We may repeat the procedure of calculating *Signif* and choosing a cut in $-\log L$ using the central sample. For the $t\bar{t}$ and QCD W+jet normalization, we apply to the normalizations selected above ($N_{top} = 40$, $N_{QCD\ W+jets} = 122$) the efficiency of the central sample $cos\theta^*$ requirement for $t\bar{t}$ and QCD W+jet events, 49% and 24% respectively (see section 5.3). We obtain $N_{top} = 20$ and $N_{QCDW+jets} = 29$. The number of observed events in the central sample is 45. For the calculation of *Signif* in the central sample, f_{syst} is the sum in quadrature of the smoothed systematic uncertainty shown in figure 5.12 and the fractional uncertainty in ϵ_L due to QCD W+jet Monte Carlo statistics. As above, we treat f_{syst} as symmetric, with its magnitude given by the positive uncertainty. The shaded curve in figure 5.14 shows the quantity *Signif* as a function of $-\log L$ for the central sample. The unshaded curve shows *Signif* if we neglect the systematic error term (i.e. $f_{syst} = 0$). A reasonable choice of cut is $-\log L > 2.0$. In both samples, inclusion of the systematic errors decreases the expected sensitivity *Signif*.

In figure 5.15 we compare *Signif* in the full lepton+ $\geq$ 3-jet sample to *Signif* in the central sample for a variety of top masses and top Monte Carlo generator programs. We do so in order to determine which sample is optimal, and whether our choice of $-\log L$ cut is seriously dependent on the choice of top mass. For three different top Monte Carlo generators (HERWIG, PYTHIA, and ISAJET) over the top mass range $M_{top} = 150, 170, 190$ GeV/c^2, our choices of $-\log L$ cut (>2.5 and >2.0) are reasonable. In all cases we expect more sensitivity to the top signal in the central sample (albeit only slightly). For this reason, we select a cut of $-\log L > 2.0$ in the central sample as our primary kinematic top sample selection. For QCD W+jet

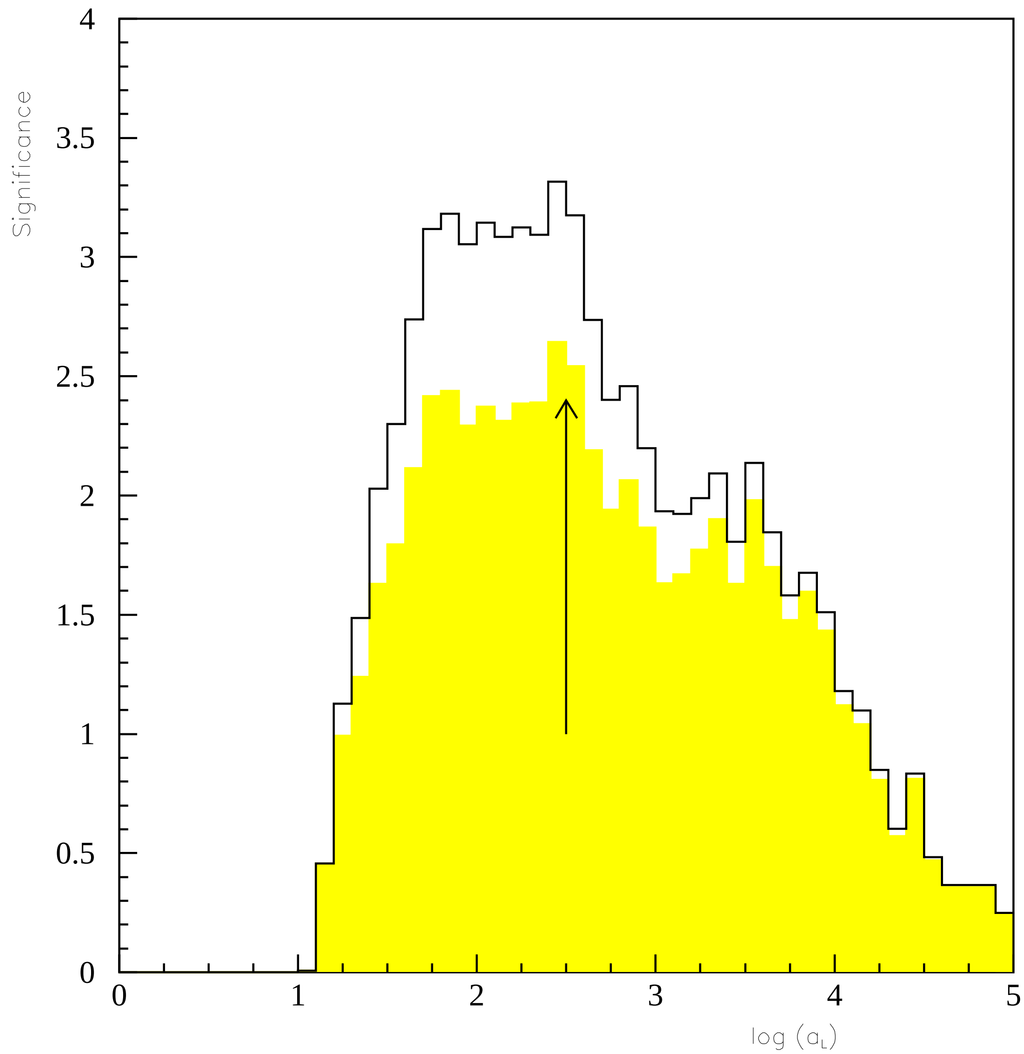


Figure 5.13: *Predicted signal excess significance (*Signif*) versus* $-\log L$ *for Monte Carlo* $t\bar{t}$ *and QCD* W*+jet predictions in the lepton+*$\geq$*3-jet sample. The shaded curve includes the effect of systematic uncertainties on* Signif*; the unshaded curve only includes statistical uncertainty on the background. Arrow indicates the optimized* $-\log L$ *selection.*

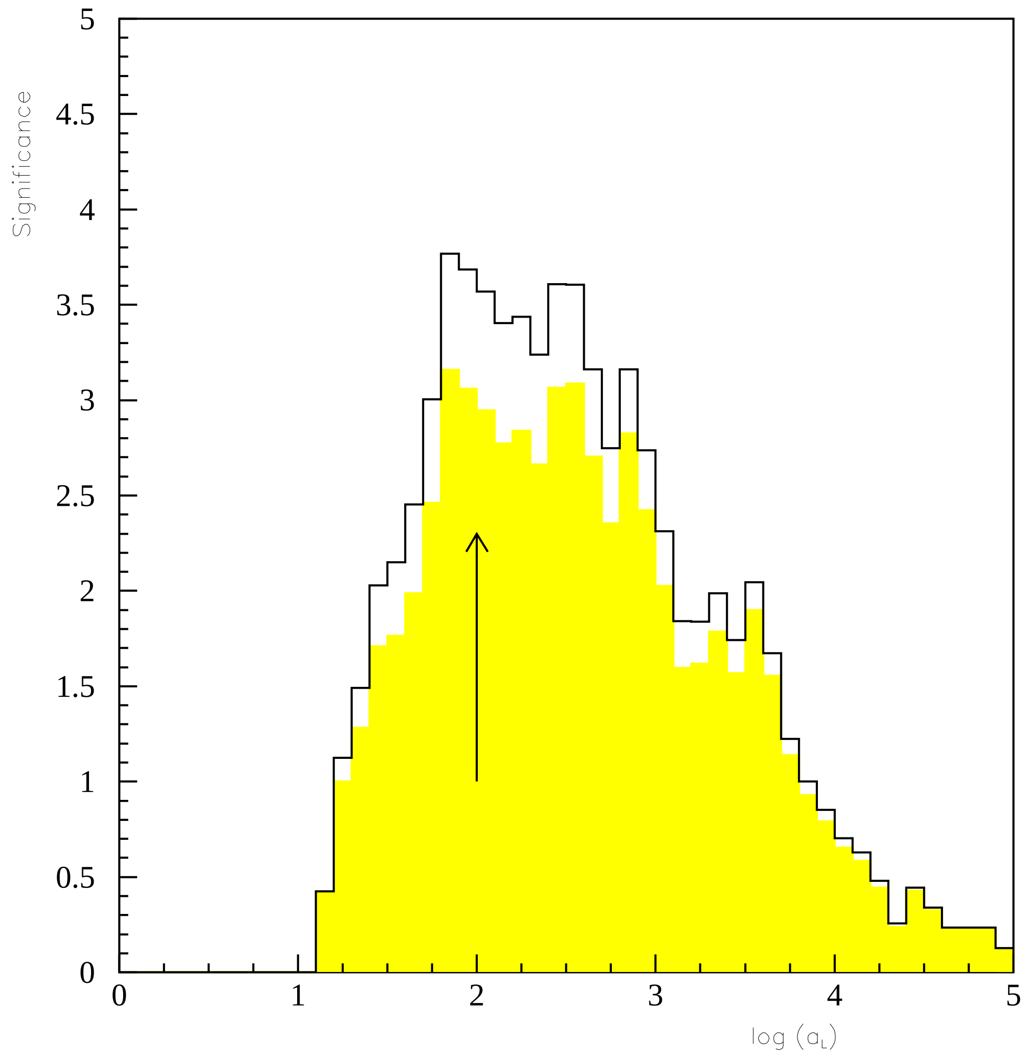


Figure 5.14: *Predicted signal excess significance (*Signif*) versus* $-\log L$ *for Monte Carlo* $t\bar{t}$ *and QCD* W*+jet predictions in the central sample. The shaded curve includes the effect of systematic uncertainties on* Signif*; the unshaded curve only includes statistical uncertainty on the background. Arrow indicates the optimized* $-\log L$ *selection.*

events passing the central sample requirements, the efficiency of the additional requirement $-\log L > 2.0$ is $\epsilon_L(-\log L > 2.0) = 0.13^{+0.03}_{-0.04}$ including systematic uncertainties. The central value for this efficiency is derived from figure 5.10. The systematic uncertainty is derived from the sum in quadrature of the values of the smoothed curves at $-\log L = 2.0$ shown in figures 5.12d, e, and f (see section 5.4) and the statistical uncertainty in our QCD W+jet Monte Carlo calculation of ϵ_L. The calculated and smoothed systematic errors for this cut are listed below. We have chosen to use the smoothed values for the systematic uncertainty in order to avoid statistical variations in the calculated values. These smoothed uncertainties are either larger than or very close to the measured uncertainties.

Systematic Errors on ϵ_L for central sample $-\log L > 2.0$		
Source	Measured	Smoothed
Q^2	-6.2%	-28%
E_T $+10\%$	$+27\%$	$+26\%$
E_T -10%	-2%	-13%
QCD W+jet Monte Carlo Stat.	$\pm 2\%$	
Total	$^{+27}_{-6}\%$	$^{+26}_{-31}\%$

5.6 Other Backgrounds

We now consider background processes other than QCD W+jet events which may produce the lepton+multijet signature. The processes which contribute at significant levels are non-W QCD background events, WW production, and WZ production. We must estimate the rate of these sources of background and also the fraction

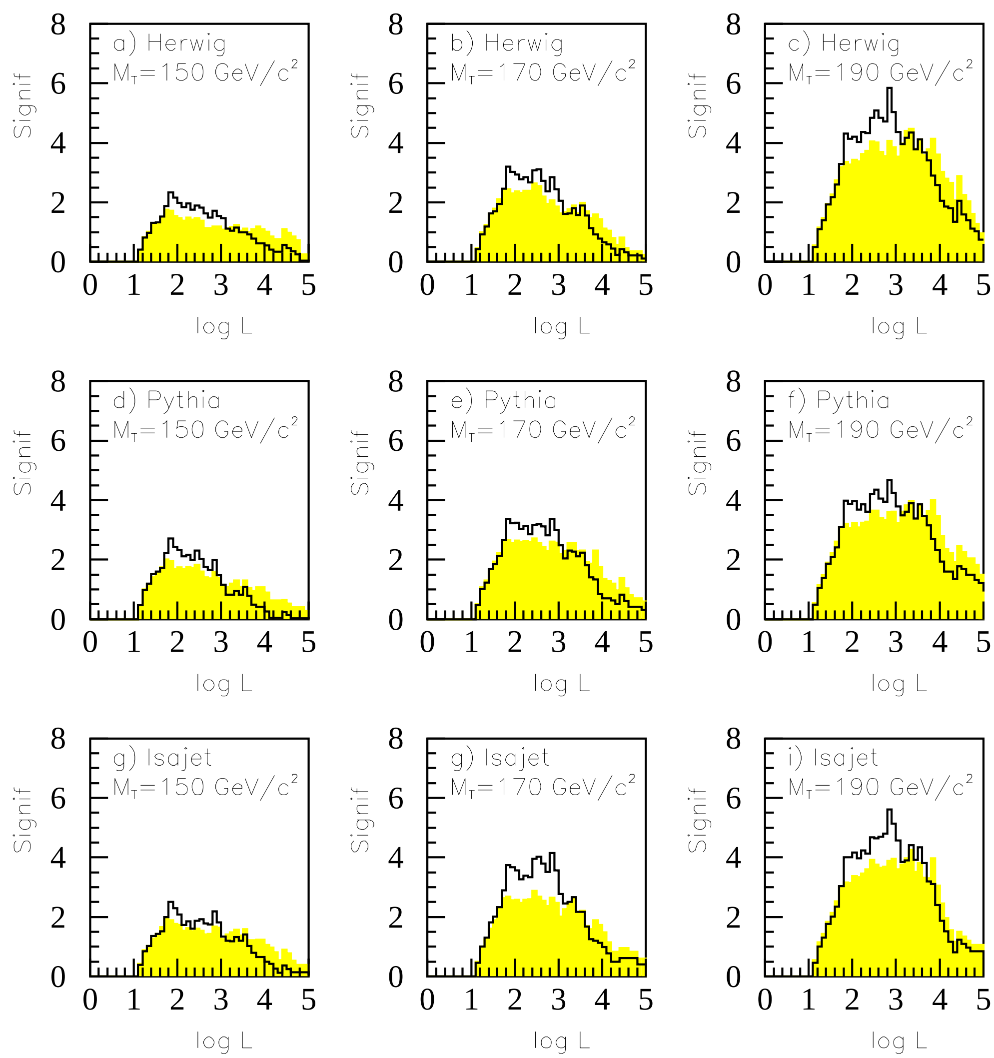


Figure 5.15: *Comparison of the predicted significance* Signif *in the entire lepton+$\geq$3-jet sample (shaded curves) and in the central sample (unshaded curves). Systematic uncertainties have been included in the estimation of* Signif. *Results using three top Monte Carlo generators (HERWIG, PYTHIA, and ISAJET) are shown for three top masses in the range* 150 $GeV/c^2 \leq M_{top} \leq$ 190 GeV/c^2. *In all cases, the use of the central sample selection with a cut of* $-\log L > 2.0$ *is reasonable, i.e. yields a near-maximal value of* Signif.

of such events which appear in the central sample in the region $-\log L > 2.0$. We denote the estimated numbers of events in the signal sample by N_{non-W}, N_{WW}, and N_{WZ}. We denote the fraction of events at $-\log L > 2.0$ by $\epsilon_L^{non-W}(-\log L > 2.0)$, $\epsilon_L^{WW}(-\log L > 2.0)$, and $\epsilon_L^{WZ}(-\log L > 2.0)$.

First, we look at non-W QCD backgrounds. In chapter 4, we estimated that non-W events comprise $(9.4\pm3.2)\%$ of events in the $\geq$3-jet sample. Normalizing to the 45 observed events in the central sample, we estimate $N_{non-W} = 4.2\pm1.4$ events. We measure the fraction of non-W events at $-\log L > 2.0$ by selecting events which pass all of our central sample cuts except the isolation cut. To select against W events, we require a non-isolated lepton, $I > 0.2$. In this sample, the fraction of events at $-\log L > 2.0$ is $\epsilon_L^{non-W}(-\log L > 2.0) = 0.27 \pm 0.13$. The total non-$W$ background at $-\log L > 2.0$ is thus 1.1 ± 0.6 events.

We now examine the diboson backgrounds. WW events appear in our sample when one W decays leptonically, the other W decays hadronically and an additional jet is present in the event. WZ events appear in our sample when the W decays leptonically and the Z decays to $q\overline{q}$. We estimate the WW and WZ signal sample acceptances and $-\log L$ distribution from ISAJET Monte Carlo events. Given the theoretical WW and WZ cross sections $\sigma_{WW} = 9.5 \pm 0.7$pb [48] and $\sigma_{WZ} = 2.6 \pm 0.4$pb [49], we find $N_{WW} = 3.2 \pm 0.7$ and $N_{WZ} = 0.39 \pm 0.09$ events. The ISAJET $-\log L$ distributions yield $\epsilon_L^{WW}(-\log L > 2.0) = 0.17 \pm 0.06$ and $\epsilon_L^{WZ}(-\log L > 2.0) = 0.21\pm0.06$. The total WW and WZ backgrounds in the region $-\log L > 2.0$

of the central sample are thus 0.5 ± 0.2 and 0.08 ± 0.03 events.

5.7 Application to Data and Significance

Figure 5.16 shows the $-\log L$ distribution for lepton+$\geq$3-jet events in the central sample. For completeness, we show in Figure 5.17 the $-\log L$ distribution for the entire lepton+$\geq$3-jet data sample. In the central sample, there are 14 events which satisfy the requirement $-\log L > 2.0$. We calculate the number of W+jet background events expected to satisfy this requirement by assuming that all of the forty-five events in the central sample are W+jet events. Including the systematic uncertainties described in section 5.5, the fraction of W+jet events expected to pass our $-\log L > 2.0$ requirement is $0.13^{+0.03}_{-0.04}$ yielding a background of $5.9^{+1.5}_{-1.8}$. In the absence of the uncertainties on the background estimate, we would calculate the significance using the binomial probability that a fractional probability of $\epsilon_L = 0.13$ applied to 45 events could produce 14 observed events. We account for the uncertainties on the background using a Monte Carlo program which simulates a large ensemble of background-only experiments with 45 events. We also use this Monte Carlo program to account for the contribution of non-W, WW, and WZ background events to the signal region. Each event may be selected from one of four categories: QCD W+jets, non-W, WW, or WZ. For the non-W, WW, and WZ events, we generate a contribution from each process according to a Poisson with mean given by the rates N_{WW}, N_{WZ}, and N_{non-W}. In each experiment, these means are subjected to gaussian smearing within the given errors. Having selected an integer number of events for each process, we then assume the remaining events (of 45) to be QCD W+jet events. We randomly assign the events to either side of $-\log L = 2.0$ according to the binomial probability ϵ_L. The value of ϵ_L we use for this assignment is appropriate to the category of that event: we use $\epsilon_L = 0.13$ for QCD W+jet events, ϵ_L^{non-W} for non-W events, etc. The ϵ_L's are smeared according to a gaussian distribution within errors. For the QCD W+jet events, we treat the error as symmetric with its magnitude given by the positive error, since the use of both the positive and negative errors would lead to a mean ϵ_L smaller than 0.13 over the ensemble of Monte Carlo experiments. The probability that the observed signal of 14 events is a result of a background fluctuation is 1.1×10^{-2}, which corresponds to a gaussian probability of 2.3σ. This is the primary result of this kinematic search: the use of kinematic information alone is effective at isolating the $t\bar{t}$ signal.

Another reasonable method of estimating the background is to assume that all 162 lepton+multijet events are background events and calculate the background in the central sample from the efficiency of the $cos\theta^*$ cut on QCD W+jets Monte Carlo events in addition to the factor $\epsilon_L(-\log L > 2.0)$. The predicted efficiency of the central sample $|cos\theta^*| < 0.7$ requirement is 23.8±1.4% (stat. error only) on QCD W+jet Monte Carlo events. We set a systematic error of ±4% on the $cos\theta^*$ efficiency by comparing this value to the efficiency observed in data, which is 28±4% ($\frac{45}{162}$).[6] The background is then:

[6]We expect $t\bar{t}$ events in the data to increase the measured efficiency with respect to the value predicted using QCD W+jet Monte Carlo events.

$$
\begin{aligned}
\text{bkg} &= (\text{Number of lepton}+\geq 3 \text{ jet events}) \\
&\quad *(cos\theta^{*} \text{ efficiency}) * \epsilon_L \\
&= 162 * (0.24 \pm 0.01(stat.) \pm 0.04(syst.)) * (0.13^{+0.03}_{-0.04}) \\
&= 5.1^{+1.6}_{-1.8}
\end{aligned}
\tag{5.21}
$$

The background calculated in this manner agrees well with our primary calculation of $5.9^{+1.5}_{-1.8}$.

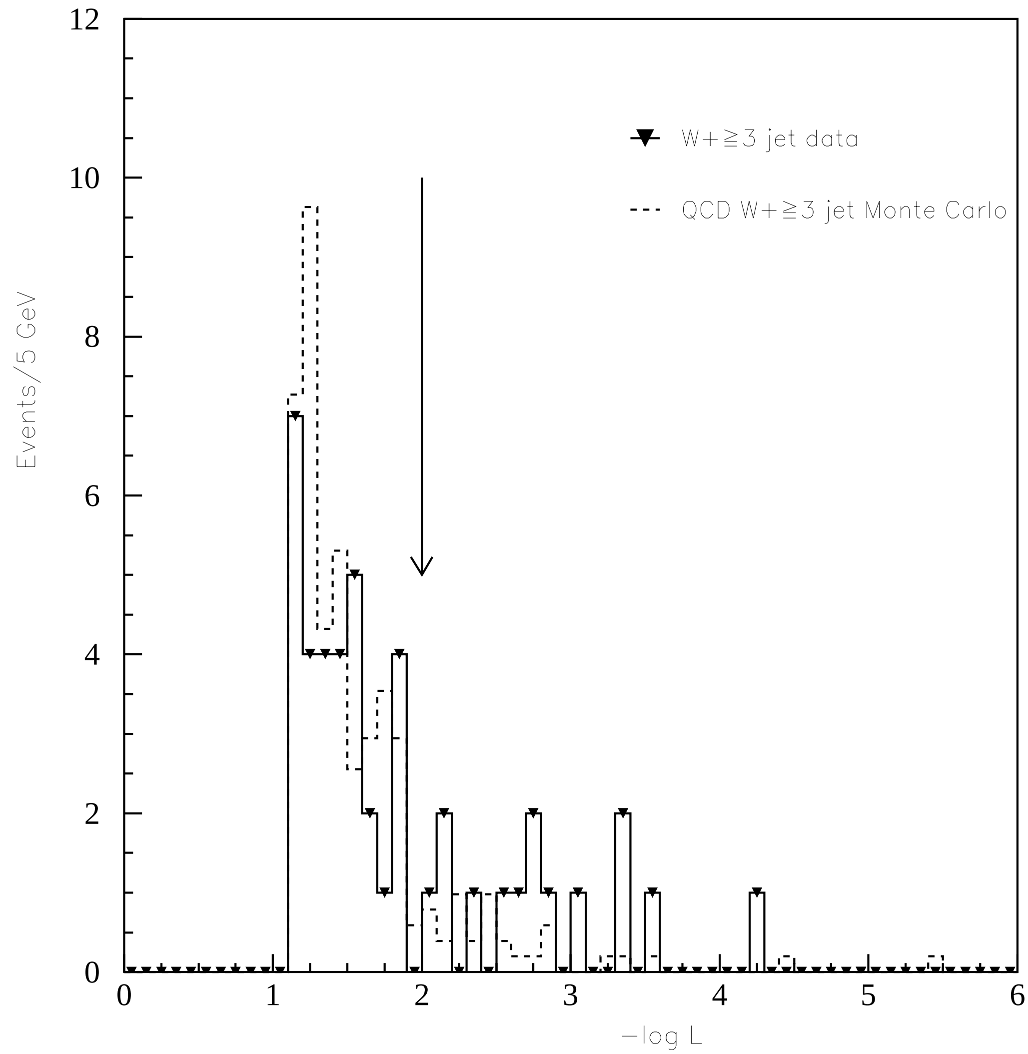


Figure 5.16: *The observed* $-\log L$ *distribution of events in data (solid histogram) passing the central sample requirements. The dashed histogram shows the* $-\log L$ *distribution expected for QCD W+jet background events, normalized to the observed 45 events. Arrow indicates the* $-\log L > 2.0$ *selection, yielding 14 events.*

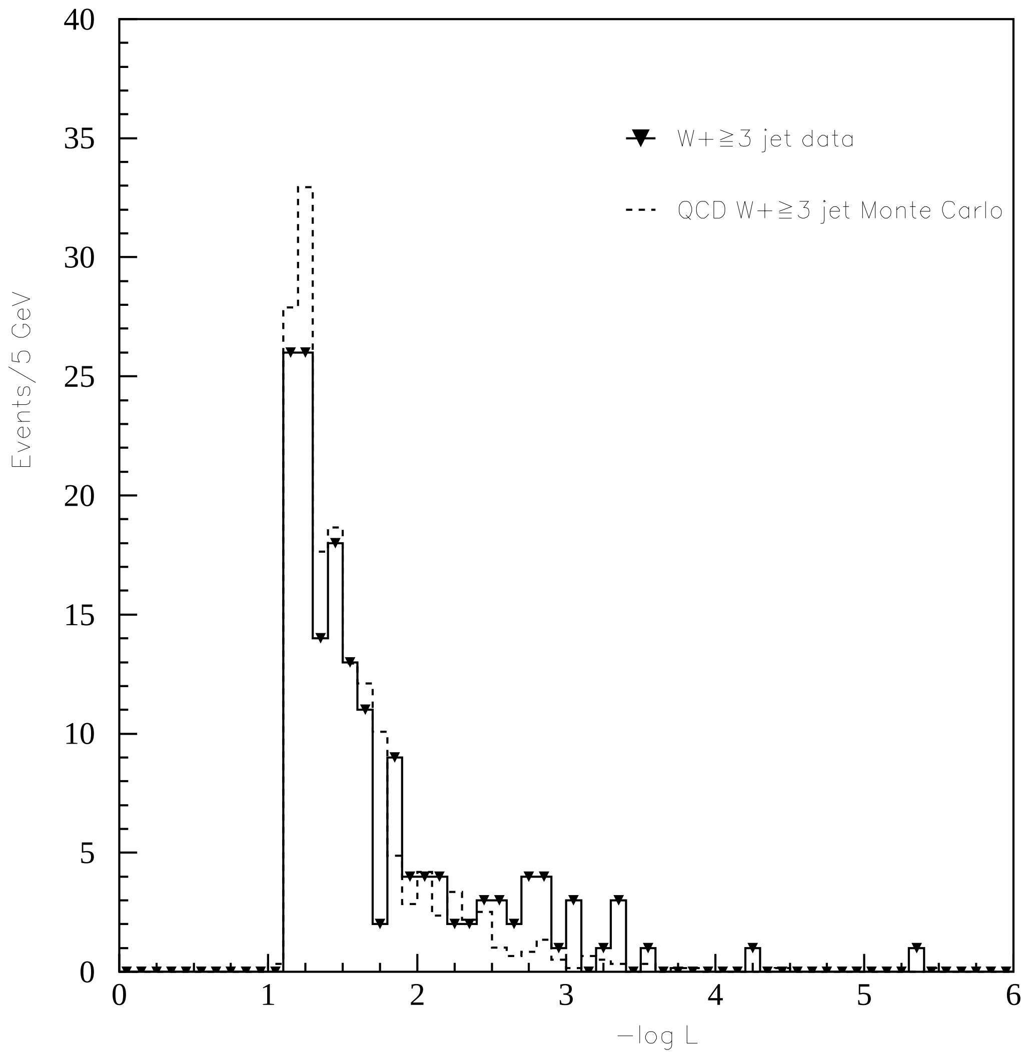


Figure 5.17: *The observed* $-\log L$ *distribution of all lepton+*$\geq$ *3-jet events in data (solid histogram). The dashed histogram shows the* $-\log L$ *distribution expected for QCD* W*+jet background events, normalized to the observed 162 events.*

Chapter 6

A Combined Kinematic and B-Tag Search for Top

In chapters 4 and 5, we established that both b-tagging and event kinematics may be used to identify the $t\bar{t}$ signal in the lepton+jets channel. In this chapter, we demonstrate a technique for combining the two approaches in the context of searching for the $t\bar{t}$ signal. We use the likelihood variable $-\log L$, as described in section 5.2, to characterize the kinematics of lepton+$\geq$ 3-jet events. We search for b-tagged events at large values of $-\log L$. This kinematic region is dominated by $t\bar{t}$ events. We compare the observed number of tagged events to the estimated number of tagged events expected in this region from background sources. Using both kinematics and b-tagging in the lepton+jets channel, we observe a $t\bar{t}$ signal which is more significant than that observed using either the b-tag or kinematic technique alone.

6.1 Choosing a $-\log L$ Cut for B-Tagged Events

We seek a $-\log L$ cut which is expected to reject most of the b-tagged background events while retaining as much b-tagged $t\bar{t}$ signal as possible. As in the case of choosing a $-\log L$ cut before requiring a b-tag (see section 5.5) this choice is based on maximizing a quantity $Signif_{tag}$[1], as defined below, which estimates our expected ability to distinguish a $t\bar{t}$ signal from the background for a given cut in $-\log L$. In the present context, the quantity $Signif_{tag}$ is calculated in a model experiment from the expected $-\log L$ distributions of $t\bar{t}$ and background events after requiring a b-tag. The $-\log L$ distribution for $t\bar{t}$ events in which at least one jet was tagged is shown as the dashed curve in the upper plot of figure 6.1; this $t\bar{t}$ sample was generated using the PYTHIA Monte Carlo program with detector response modeled using the full detector simulation. The $\log L$ distribution for the same $t\bar{t}$ sample but without requiring a b-tag is shown as the solid curve in the upper plot of figure 6.1. The $-\log L$ distribution for b-tagged $t\bar{t}$ events is slightly harder than the distribution for $t\bar{t}$ events in which we do not require a b-tag. We refer to this effect as the *tagging bias* which the b-tag requirement introduces on the $-\log L$ distribution. The lower plot of figure 6.1 shows the b-tag rate as a function of $-\log L$ in $t\bar{t}$ Monte Carlo events. For the purposes of our model experiment, the absolute normalization of the $t\bar{t}$ signal ($N_{top} = 40$) is taken from the $t\bar{t}$ cross section as in section 5.5, but here, we must account for the $t\bar{t}$ event b-tag efficiency of 42% (see section 4.2.1): $N_{top,tag} = 40 * 0.42 = 17$ events. We approximate the $-\log L$ distribution of b-tagged background events from QCD W+jet Monte Carlo events. This distribution was shown in figure 5.9 without the b-tag requirement. The leading order QCD W+3 parton matrix element calculation as implemented in the VECBOS Monte

[1]In the present context we use the subscript *tag* to indicate that a b-tag is required and to distinguish this and other quantities from analogous quantities defined in chapter 5.

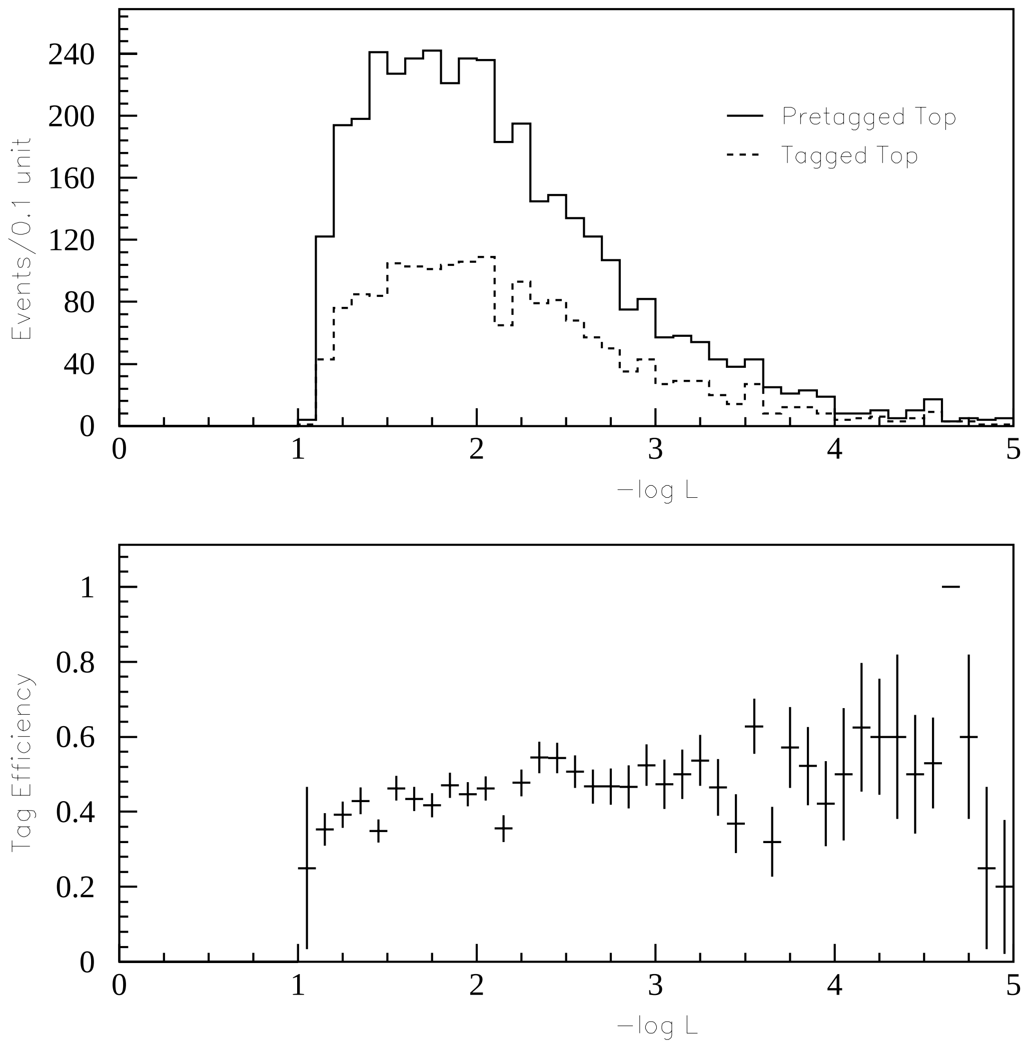


Figure 6.1: *Upper plot: the pre-tagged (solid) and tagged (dashed)* $-\log L$ *distribution for* $t\bar{t}$ *events passing the lepton+*$\geq$ *3-jet requirement. Lower plot: the tag rate as a function of* $-\log L$ *in the same sample.*

Carlo program is used to generate these events, with parton evolution performed using the HERWIG Monte Carlo program and detector response simulated with the full CDF detector simulation (see section 5.1). We expect the $-\log L$ distribution for background events to be somewhat harder after requiring a b-tag, just as the $-\log L$ distribution for $t\bar{t}$ events is harder after requiring a b-tag. We account for this effect on the background by multiplying the distribution in figure 5.9 by the b-tag bias shown in figure 6.1 for $t\bar{t}$ events.[2] The shape of the resulting distribution is shown as the solid histogram in the upper plot of figure 6.2. The normalization of the background $N_{bkg,tag}$ in this model experiment is taken from the number of b-tagged events expected from background as calculated in table 4.10 (6.5 events), rescaled by the factor $\epsilon_{\Delta R} = \frac{162}{204}$ to account for the jet-jet $\Delta(R)$ cut, which was not used in the b-tag analysis but which is applied when we are considering kinematics. Since the background calculation is largely normalized to the observed data in chapter 4, it is actually an overestimate if top is present. We account for this overestimate in our model experiment by rescaling the number of background events by the factor $\frac{N_W - N_{top}}{N_W}$, where $N_W = 162$ as in section 5.5. Thus $N_{bkg,tag} = 3.9$ events.

For our model experiment, we compute the mean number of $t\bar{t}$ and background events expected to pass a given $-\log L$ requirement. From the $-\log L$ distribu-

[2] We note that this tagging bias may not be correct for each of the components of the b-tagged background ($Wb\bar{b}$, $Wc\bar{c}$, Wc, mistags, etc.), since the b decays in top and background events have different characteristics, and also because a considerable proportion of the tagged background is from c-quark decays, which may also have different characteristics. In section 6.2, we calculate each of these backgrounds separately for the background estimate and find 1.46 ± 0.43 events. For comparison, the background in the region $-\log L > 1.8$ as computed using the QCD W+jets distribution weighted by the $t\bar{t}$ tag bias (i.e. the solid histogram in the upper plot of figure 6.2) is $\simeq 1.1$ events.

tions for tagged $t\bar{t}$ and background events shown in figure 6.2, we can extract the integrated fractions of tagged top and background events expected above a given $-\log L$ cut; we denote these fractions as $\epsilon_{L,tag}^{bkg}$ and $\epsilon_{L,tag}^{top}$ in analogy to ϵ_L and ϵ_L^{top} of sections 5.4 and 5.5. It is to be understood that $\epsilon_{L,tag}^{bkg}$ is meant to represent the total tagged background including sources other than W production. Nsig_{tag} is the number of top and background events expected to pass the $-\log L$ cut:

$$\text{Nsig}_{tag} = \epsilon_{L,tag}^{bkg} * N_{bkg,tag} + \epsilon_{L,tag}^{top} * N_{top,tag} \tag{6.1}$$

The background estimate in this region is then Nbkg_{tag}:

$$\text{Nbkg}_{tag} = \epsilon_{L,tag}^{bkg} * (\epsilon_{\Delta R}) * 6.5 \text{ events} \tag{6.2}$$

From Nsig_{tag}, the mean number of top and background events expected to pass a given $-\log L$ cut, and the background estimate Nbkg_{tag}, we compute the expected signal excess significance Signif_{tag} using equation 6.3.

$$Signif_{tag} = \frac{\text{Nsig}_{tag}}{\sqrt{\text{Nbkg}_{tag} + (f_{syst} \times \text{Nbkg}_{tag})^2}} \tag{6.3}$$

The definition of Signif_{tag} accounts for both statistical and systematic uncertainty in the background calculation. In this case the fractional systematic error f_{syst} on the background has two components: the systematic errors on the purely kinematic quantity ϵ_L, and the global $\pm 30\%$ errors on the total tagged background. The quantity Signif_{tag} of equation 6.3 is shown in the lower plot of figure 6.2 and leads us to choose $-\log L > 1.8$ as a reasonable kinematic selection for b-tagged events.

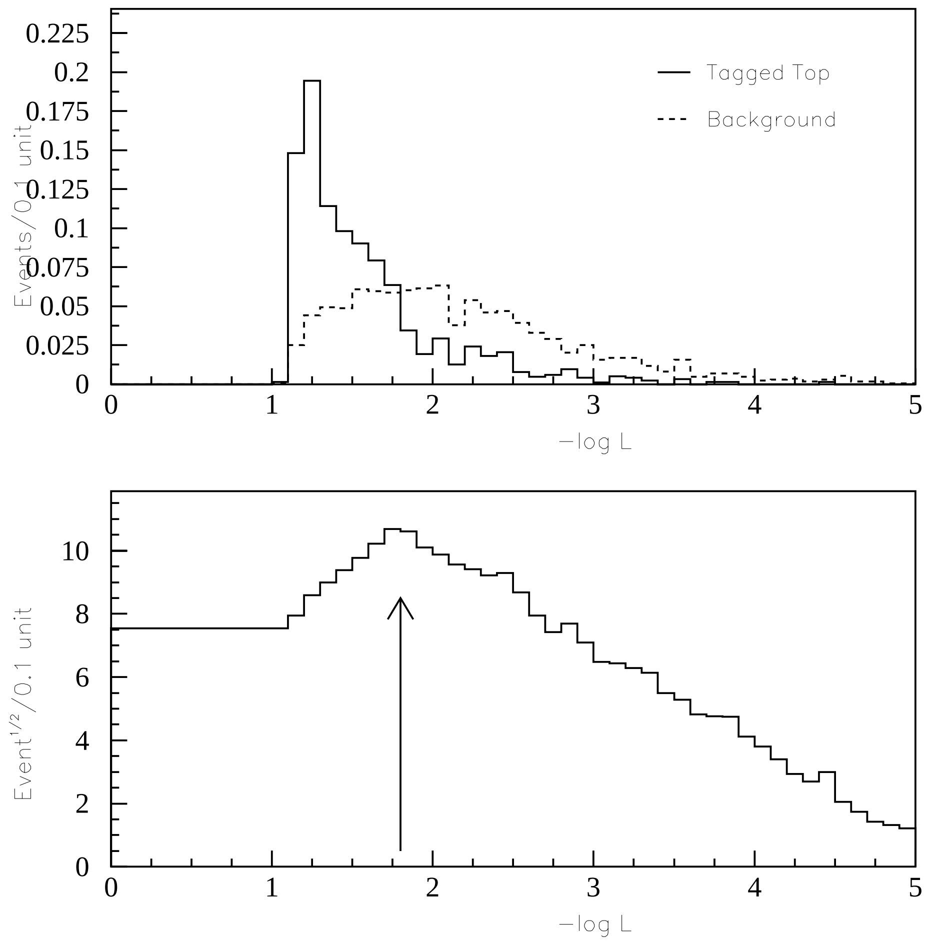


Figure 6.2: *Optimization of the* $-\log L$ *cut after b-tagging. The upper plot shows the* $-\log L$ *distribution for the expected background (solid histogram) and for* $t\bar{t}$ *Monte Carlo events (dashed histogram). The background curve is derived from QCD W+jet Monte Carlo events. The effect of requiring a b-tag on this distribution is approximated by applying the tagging bias in* $t\bar{t}$ *Monte Carlo events (figure 6.1) to the QCD W+jet events. The quantity* Signif_{tag} *is shown in the lower figure. The arrow indicates the maximum value of* Signif_{tag}*, at which we set the* $-\log L$ *cut.*

For the purposes of this combined kinematic and b-tag search for $t\bar{t}$ production, we elect not to use $cos\theta^*$ in selecting the signal region. In chapter 5, $cos\theta^*$ is defined as the cosine of the angle between the jet axis and the $p\bar{p}$ beam as calculated in the center of momentum frame of the hard scattering process. The $t\bar{t}$-enriched *central sample* is defined as those lepton+$\geq$ 3-jet events in which the leading three jets satisfy the requirement $|cos\theta^*| < 0.7$. In the kinematic-only search of chapter 5, the $cos\theta^*$ requirement is useful in reducing the background while retaining about half of the $t\bar{t}$ signal. However, in the present context, a $cos\theta^*$ requirement in addition to the b-tag requirement is unnecessary, since the b-tag requirement by itself reduces the background significantly. In this low background situation, we opt for signal efficiency over further background rejection, and thus perform the kinematic and b-tag search in the entire lepton+$\geq$ 3-jet sample.

6.2 Backgrounds in the Kinematic and B-Tag Search

Having selected a kinematic region of the lepton+$\geq$ 3-jet sample in which to search for b-tagged $t\bar{t}$ events, we now estimate the number of b-tagged events we expect in this region from sources other than $t\bar{t}$ events. Our primary background calculation in the lepton+jet sample prior to the kinematic requirement $-\log L > 1.8$ consists of several components, the details of which are given in section 4.3, and which we briefly review here. Background sources of tagged events are mistagged jets, $Wb\bar{b}$, $Wc\bar{c}$, Wc, non-W sources including direct $b\bar{b}$ production, WW, WZ, and $Z \to \tau\bar{\tau}$ production. Explicit calculations of the $Wb\bar{b}$, $Wc\bar{c}$, and Wc contributions are made using Monte Carlo models for the tag rate and heavy flavor content as a function of jet multiplicity. The $Wb\bar{b}$, $Wc\bar{c}$, and Wc computations are normalized to the observed number of W events as a function of jet multiplicity. The mistag rate is computed by applying the negative decay length $(-L_{xy})$ tag rate measured in a control sample of inclusive jet triggers to the events in the lepton+jet sample. By negative decay length, we mean that the inner product of the jet direction and the direction from the primary vertex to the secondary vertex tag is negative. This mistag rate is parametrized by track multiplicity and jet E_T. The number of tagged events expected from non-W sources including direct $b\bar{b}$ production is estimated from the tag rate observed in data samples dominated by non-W background processes. Finally, the small WW, WZ, and $Z \rightarrow \tau\bar{\tau}$ backgrounds are calculated from theoretical cross sections for these processes and Monte Carlo predictions for their acceptance and tag efficiencies. In section 4.3 we calculate the background in terms of the expected number of tagged events and also the expected number of tagged jets as a function of jet multiplicity; for jet multiplicities greater than one, these numbers may be different owing to the presence of two or more tagged jets in an event. In the present context, it is reasonable to expect a correlation between event kinematics and the probability that a background event contains two or more tagged jets. Thus in order to compare the observed number of tagged jets to the number expected from background sources in the kinematic region $-\log L > 1.8$, it would be necessary to calculate this correlation for each source of tagged background. In order to avoid this complication, we examine the number of tagged events rather than total number of tagged jets, and thus compute a background in terms of tagged events. The total tagged background in the lepton+$\geq$ 3-jet sample from all of these processes is 6.4$\pm$1.9 events (see table 4.10).

An additional method of calculating the tagged background is presented in section 4.3.2. This confirmatory computation uses the positive decay length ($+L_{xy}$) tag rate in inclusive jet trigger events to account for the background for mistags, $Wb\bar{b}$, and $Wc\bar{c}$. The remaining backgrounds are calculated in the same way as in the primary background computation. This method is expected to overestimate the actual background. The total background computed in this manner is 9.1$\pm$1.2 events.

Beginning with the total number of tagged events expected for each of the background sources in the entire lepton+$\geq$ 3-jet sample, we must estimate the number of events from each source expected to satisfy the kinematic requirement $-\log L > 1.8$ in addition to the b-tag requirement.

6.2.1 Mistags

The procedure used to calculate the mistag background is identical to the procedure used for the entire lepton+$\geq$ 3-jet sample except that the $-L_{xy}$ mistag parametrization is applied only to those pre-tagged events which satisfy $-\log L > 1.8$. In this region, the mistag rate is 0.55$\pm$0.11 tagged events. The systematic uncertainty of $\pm$20% is estimated in section 4.3.1. The upper plot in figure 6.3 shows the $-\log L$

distribution of events in the lepton+$\geq$ 3-jet sample weighted by the $-L_{xy}$ mistag parametrization. We shall use this distribution in estimating the distribution of the mistag background over the full $-\log L$ range rather than only in the region $-\log L > 1.8$. For completeness, the lower plot in figure 6.3 shows the same distribution weighted instead by the $+L_{xy}$ mistag parametrization. Recall that in section 4.3.2 we use the $+L_{xy}$ mistag parametrization as an overestimate of the background from mistags, $Wb\bar{b}$, and $Wc\bar{c}$ for the purpose of confirming the primary background calculation. We shall follow a similar confirmatory approach in section 6.4 after requiring $-\log L > 1.8$.

6.2.2 $Wb\bar{b}$, $Wc\bar{c}$, and Wc

Our objective here is the computation of how many tagged events we expect from $Wb\bar{b}$, $Wc\bar{c}$, and Wc production in the lepton+$\geq$ 3-jet sample in the kinematic region $-\log L > 1.8$. Equivalently, we may determine for each of these sources the ratio of the number of tagged events with $-\log L > 1.8$ to the number of tagged events expected without a cut on $-\log L$. For these three background sources, this ratio is referred to as $\epsilon_{L,tag}^{Wb\bar{b}}$, $\epsilon_{L,tag}^{Wc\bar{c}}$, and $\epsilon_{L,tag}^{Wc}$ for a general value of $-\log L$ cut and, for example, $\epsilon_{L>1.8,tag}^{Wb\bar{b}}$ for the $-\log L > 1.8$ cut. We may then multiply $\epsilon_{L,tag}^{bkg}$ by the total backgrounds for each process in the $W+\geq$ 3-jet sample as previously determined in section 4.3 ($N_{Wb\bar{b}} = 1.58 \pm 1.27$, $N_{Wc\bar{c}} = 0.65 \pm 0.52$, and $N_{Wc} = 1.14 \pm 0.51$).

For each of these processes, $\epsilon_{L>1.8,tag}$ is ideally calculated using a sample of tagged Monte Carlo events which pass the lepton+ $\geq$ 3-jet requirements. However, there

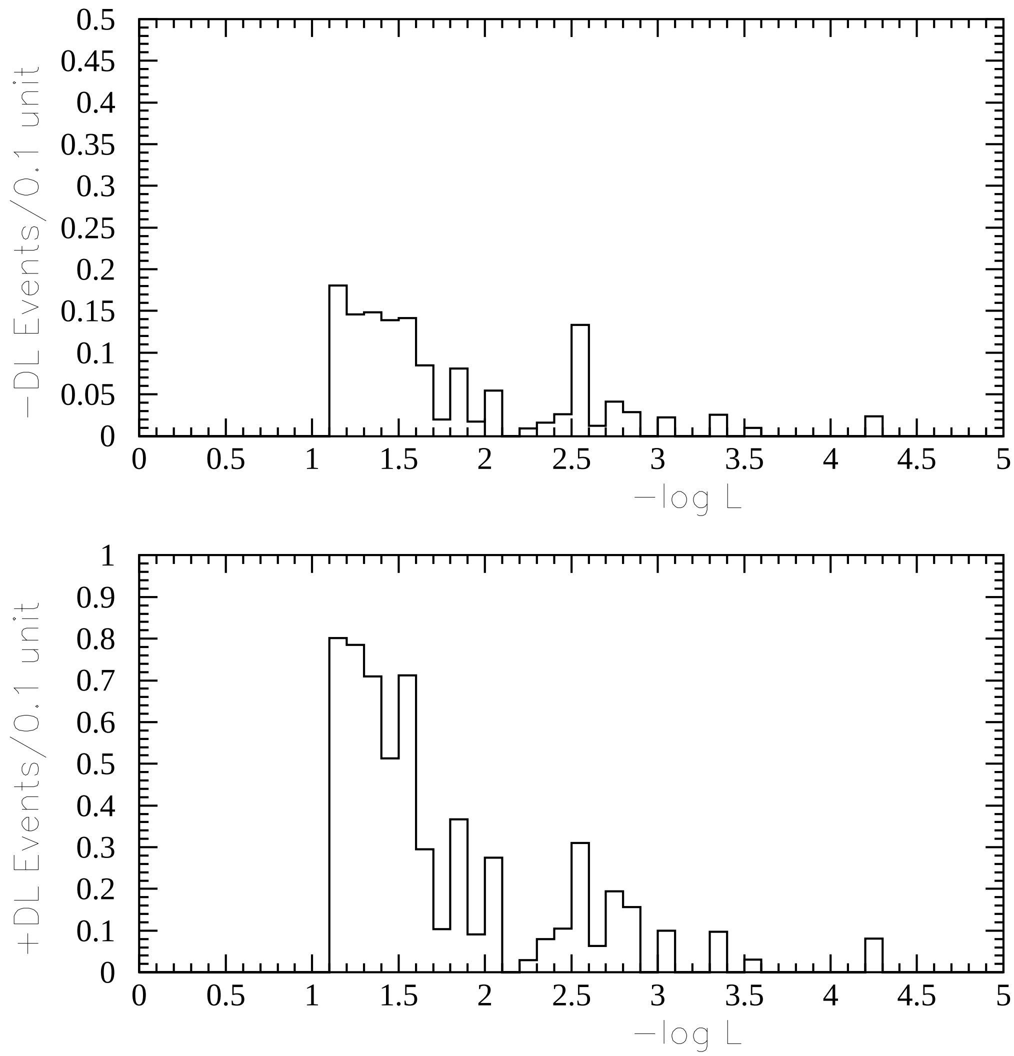


Figure 6.3: *Distributions in* $-\log L$ *of events in the lepton+*$\geq$*3-jet data sample weighted by the* $-L_{xy}$ *(upper plot) and* $+L_{xy}$ *(lower plot) mistag parametrizations as applied to each event.*

are two complications in this computation which we will address in the following sections. First, we must verify that the Monte Carlo model we are using accurately predicts the purely kinematic features of lepton+ $\geq$ 3-jet events. For the background prior to requiring a b-tag, we verified the kinematic features of the QCD W+jets background as implemented in the VECBOS Monte Carlo program (see section 5.1). However, we have not verified the kinematic features of our model for the sources of b-tagged background events. We use the HERWIG Monte Carlo program as our primary model for the $Wb\bar{b}$, $Wc\bar{c}$, and Wc processes. If this Monte Carlo model underestimates the number of pretagged events in the region $-\log L > 1.8$, then it will also underestimate the ratio $\epsilon_{L>1.8,tag}$. We shall address this issue of *kinematic modeling* below by comparing the pre-tagged $-\log L$ distributions to the pre-tagged $-\log L$ distribution from the leading order QCD W+jets matrix element calculation and rescaling the $Wb\bar{b}$, $Wc\bar{c}$, and Wc rates accordingly. The second issue we must address is the statistical reliability of the ratio $\epsilon_{L>1.8,tag}$ as calculated directly from finite Monte Carlo samples of tagged events. This issue is particularly critical for the $Wc\bar{c}$ and Wc Monte Carlo samples, in which the low c-jet tagging rate ($\sim 5\%$) results in small numbers of tagged events even when the pre-tagged samples are reasonably large. We shall address this issue below by developing a technique which predicts the number of tags expected from the characteristics of the jets in the pre-tagged Monte Carlo samples. We refer to this as accounting for the *tag bias* on pre-tagged events.

Kinematic Modeling

From the kinematic features of our Monte Carlo models for $Wb\bar{b}$, $Wc\bar{c}$, and Wc production, we establish correction factors f_{kin} by which we scale the b-tagged background estimates in the region $-\log L > 1.8$. Our primary Monte Carlo model for $Wb\bar{b}$ and $Wc\bar{c}$ events is the leading order matrix calculation [47] for the diagram shown in figure 4.7a. Additional jets are generated through parton evolution as implemented in the HERWIG Monte Carlo program. We use the W+1-jet matrix element of the HERWIG Monte Carlo program to generate Wc events, with jet multiplicities larger than one generated through HERWIG parton evolution. The pretagged $-\log L$ distributions for $Wb\bar{b}$, $Wc\bar{c}$, and Wc lepton+$\geq$ 3-jet events are shown as the dashed curves in figures 6.4a, 6.4b. and 6.4c. Also shown as the solid curve in each of these figures is the $-\log L$ distribution for QCD W+jet events. This distribution is used in chapter 5 to model the kinematic features of the pretagged background. Uncertainty due to the absence of higher order diagrams in the VECBOS calculation is estimated by varying the momentum transfer scale Q^2 in α_s. $Q^2 = M_W^2$ is chosen as the primary scale because it results in "harder" (more conservative) kinematic distributions than other reasonable choices. The $-\log L$ distributions for the lepton+$\geq$ 3-jet samples of $Wb\bar{b}$, $Wc\bar{c}$, and Wc Monte Carlo events shown in figure 6.4 are all softer than the QCD W+jet distribution. The QCD W+jet Monte Carlo calculates the leading order diagrams for W+3 parton final states, while the $Wb\bar{b}$, $Wc\bar{c}$, and Wc Monte Carlo programs depend on a par-

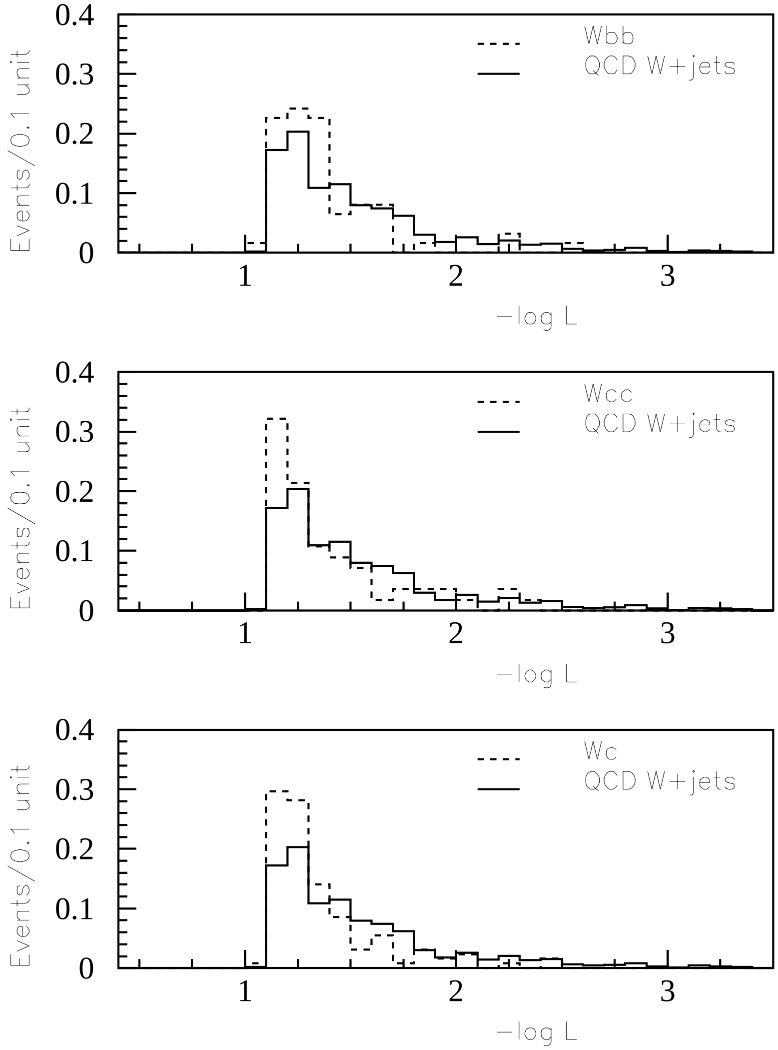


Figure 6.4: *Comparison of the pre-tagged* $-\log L$ *distribution for lepton+*$\geq$*3-jet* $Wb\bar{b}$ *(top, dashed),* $Wc\bar{c}$ *(middle, dashed), and* Wc *(bottom, dashed) Monte Carlo samples to the QCD* W*+jet Monte Carlo distribution (solid). In all three cases, the QCD* W*+jet distribution is harder.*

Source	untagged fraction, $\epsilon_{L>1.8}$	scale factor, f_{kin}
QCD W+jets	$0.182^{+0.027}_{-0.051}$	–
$Wb\bar{b}$	0.065 ± 0.031	2.8 ± 1.3
$Wc\bar{c}$	0.143 ± 0.047	1.3 ± 0.4
Wc	0.094 ± 0.036	1.9 ± 0.7

Table 6.1: *Fraction of untagged events passing $-\log L \geq 1.8$ for lepton+3-jet $Wb\bar{b}$, $Wc\bar{c}$, Wc, and QCD W+jet Monte Carlo samples. Errors are statistical on all except the QCD W+jet fraction, which are systematic errors as described in section 5.4.*

ton evolution model to generate $\geq$ 3-jet final states. We therefore take the QCD W+jet calculation of VECBOS to be a more accurate model of the kinematics of lepton+$\geq$ 3-jet events.[3] We adopt the $-\log L$ shape from the QCD W+jet Monte Carlo as the pre-tagged shape of the $Wb\bar{b}$, $Wc\bar{c}$, and Wc distribution in the following manner. Since we are specifically interested in the region $-\log L > 1.8$, we measure for each Monte Carlo sample the fraction of pre-tagged lepton+$\geq$ 3-jet events which lie in this region. These fractions, $\epsilon^{Wb\bar{b}}_{L>1.8}$, etc., are listed in table 6.1; in each case, $\epsilon_{L>1.8}$ is smaller than the same fraction as measured in QCD W+jets Monte Carlo events. This indicates that the $Wb\bar{b}$, $Wc\bar{c}$, and Wc Monte Carlo models underestimate the number of events in this region. To account for this underestimate, in table 6.1 we calculate a kinematic factor f_{kin} for each process by which we will scale the b-tagged background estimate. For example, the scale factor for $Wb\bar{b}$ is $f^{Wb\bar{b}}_{kin} = \frac{\epsilon^{QCD\ W+jets}_{L>1.8}}{\epsilon^{Wb\bar{b}}_{L>1.8}}$.

[3]The accuracy of the kinematic modeling of $Wb\bar{b}$, $Wc\bar{c}$, and Wc events in chapter 4 is not as critical as in the present context. In chapter 4, the absolute backgrounds are normalized to the observed data rather than calculated from a kinematically-determined acceptance.

The kinematic scale factors f_{kin} are intended to correct for kinematic differences resulting from the final state parton generation method (i.e. matrix element calculation versus parton shower evolution) rather than heavy flavor content; we must therefore check that the selection of W events with heavy quark jets does not introduce a bias on these scale factors. In deriving f_{kin}, we have compared jets from all sources ("generic jets") in the QCD W+jets calculation to heavy quark jets in HERWIG Monte Carlo events, when ideally we would have compared generic jets in QCD W+jet events to generic jets in HERWIG Monte Carlo events. A positive correlation between heavy flavor content and jet energy would cause events containing heavy flavor jets to have higher energy jets than generic W events generated using the same Monte Carlo program. In this case, the selection of heavy flavor events would then overestimate $\epsilon^{bkg}_{L>1.8}$ and thus underestimate f_{kin}. We check the correlation between heavy flavor and jet energy by examining the ratio of b- and c-jets to all jets as a function of jet energy. For this comparison, we use the W+1 jet matrix element calculation in the HERWIG Monte Carlo. Recall that the HERWIG Monte Carlo is used in section 4.3 to calculate the overall b and c content of W events. Figures 6.5a and 6.5b show the fraction of jets containing a b or c quark as a function of jet energy in W+1-jet and W+2-jet events. No strong correlation between jet energy and heavy flavor content is observed. We may also check for correlation by comparing the kinematics of events generated using the standard VECBOS Monte Carlo QCD W+2-parton and QCD W+3-parton matrix element calculations to VECBOS Monte Carlo events in which we require a final state $b\bar{b}$ or $c\bar{c}$ pair. Since this latter calculation does not account for quark mass effects, the rate of b and c quark production is nearly equal.[4] Figures 6.5c and 6.5d show the fraction of jets containing a b or c quark as a function of $E_T(2)$ for W+2-jet events and as a function of $-\log L$ for W+3-jet events. No strong correlation is observed. We may also compare the quantities $\epsilon_{L>1.8}$ for W+3-jet events with no heavy flavor selection to those in which we have required a b or c quark. For events generated with the HERWIG Monte Carlo, $\frac{\epsilon^{Wb\bar{b}}_{L>1.8}}{\epsilon^{W+jets}_{L>1.8}} = 1.2 \pm 0.3$ and $\frac{\epsilon^{Wc\bar{c}}_{L>1.8}}{\epsilon^{W+jets}_{L>1.8}} = 0.8 \pm 0.2$. For events generated with the VECBOS Monte Carlo, $\frac{\epsilon^{Wb\bar{b}}_{L>1.8}}{\epsilon^{W+jets}_{L>1.8}} = 1.0 \pm 0.2$ and $\frac{\epsilon^{Wc\bar{c}}_{L>1.8}}{\epsilon^{W+jets}_{L>1.8}} = 0.9 \pm 0.2$. We conclude that there is no significant bias on event kinematics due to the selection of events containing heavy flavor, and thus we use the values of f_{kin} derived above.

Tag Bias Modeling

In this section, we use a technique for predicting the kinematic distribution of tagged $Wb\bar{b}$, $Wc\bar{c}$, and Wc Monte Carlo events from pre-tagged Monte Carlo events. We determine $\epsilon^{bkg}_{L,tag}$ from the pre-tagged Monte Carlo lepton+$\geq$ 3-jet samples rather than the tagged samples for improved statistical reliability. We extract from W+1-jet events the tag rate for heavy flavor jets as a function of the track multiplicity of the jet. Figure 6.6 shows the tag rate for b-jets in $Wb\bar{b}$ events and for c-jets in $Wc\bar{c}$ and Wc events. Given this parametrization, we may predict the probability that a given b- or c-jet is tagged. We may test the ability of this parametrization to predict

[4]The rate of c jets is slightly larger than the rate of b jets due to energy lost to semileptonic decays, which are more prevalent for b quarks.

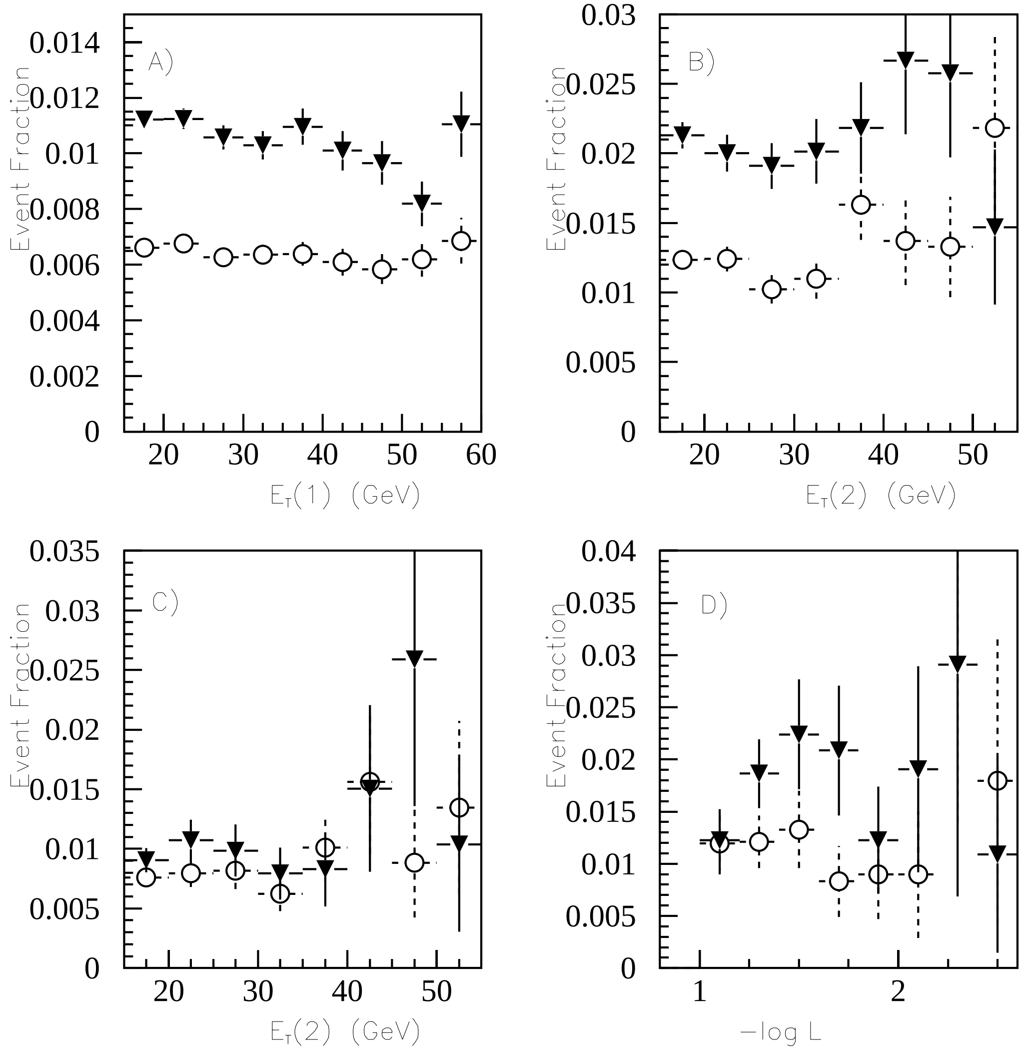


Figure 6.5: *Ratio of events containing a b quark (open circles) or c quark (filled triangles) to all events as a function of jet E_T in (a) HERWIG Monte Carlo $W+1$-jet events; (b) HERWIG Monte Carlo $W+2$-jet events; (c) VECBOS Monte Carlo QCD $W+2$ jet events. Figure (d) shows the same ratio as a function of $-\log L$ for VECBOS Monte Carlo QCD $W+3$ jet events. Quark mass effects are neglected in the VECBOS Monte Carlo, leading to similar b and c rates in (c) and (d).*

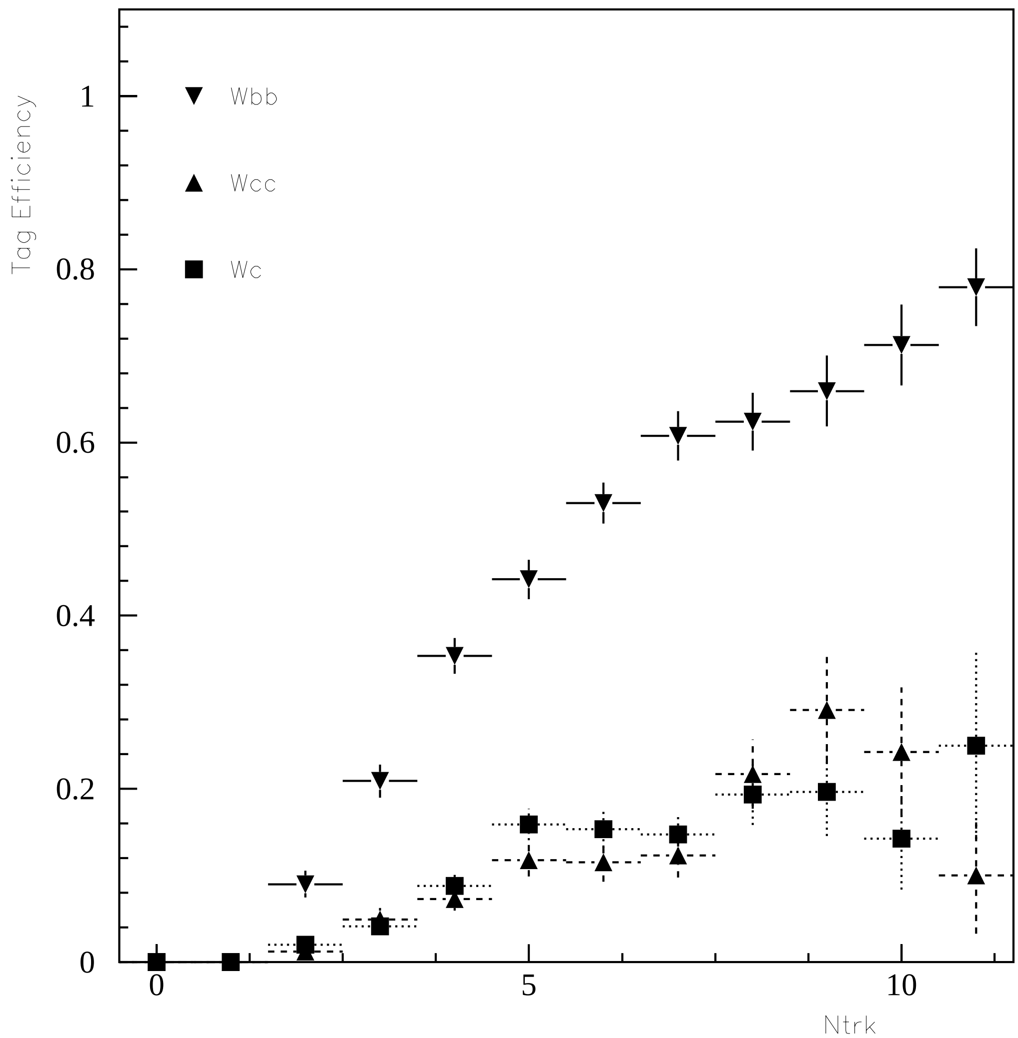


Figure 6.6: *Tag rate as a function of track multiplicity for heavy flavor jets in W+1-jet Monte Carlo samples. For the $Wb\bar{b}$ Monte Carlo we compute this tag bias for b-jets and for the $Wc\bar{c}$ and Wc Monte Carlo, we compute this tag bias for c-jets.*

the number of tagged events by summing the tagging probability for all the jets in a given Monte Carlo sample and comparing the resulting total to the actual number of tagged events in the sample.[5] For W+2 jet events, the ratios of the number of tagged events to the predicted number of tags are 1.04±0.06 ($Wb\bar{b}$), 1.18±0.17 ($Wc\bar{c}$) , and 1.13±0.14 (Wc) for the three Monte Carlo samples. Uncertainties are statistical errors on the number of tagged events. These values indicate that our tag bias parametrizations reproduce the absolute tag rate to within 20%. For the background calculation in the ≥3-jet sample, we will *not* use the predicted absolute value, using instead the total absolute prediction of chapter 4. Nonetheless, it is reassuring that the absolute agreement is good.

The tag bias may also be used to predict kinematic distributions of tagged events from pre-tagged samples; from this prediction we will estimate $\epsilon_{L,tag}^{bkg}$ for lepton+ $\geq$ 3-jet events. The prediction of kinematic distributions is accomplished in the same way as the prediction of the $-\log L$ shape for mistags in figure 6.3. A given event has a well-defined tag bias weight which we calculate from the track multiplicities of the heavy flavor jets in the event. For a pre-tagged sample of events, we plot the values of a given kinematic variable (e.g. jet E_T, $-\log L$), weighting each entry by the event's tag bias weight. We test this method by applying the tag bias extracted from W+1-jet events to W+2-jet events and comparing the predicted $E_T(2)$ distribution to the $E_T(2)$ distribution of tagged Monte Carlo events. Figure 6.7 shows this comparison

[5]This is essentially the same technique used to estimate the total mistag background. If there are several heavy flavor jets in a given event we must subtract the probability that the event is multiply tagged.

for $Wb\bar{b}$, $Wc\bar{c}$ and Wc Monte Carlo events where we have plotted the predicted (solid histogram) and actual (dashed points, statistical errors only) $E_T(2)$ distributions in the left hand plots. The histograms are *not* normalized to equal area. The plots on the right hand side of figure 6.7 show the ratio of the integrated actual tags (in the sense of integrated area above a given $E_T(2)$ cut) to integrated predicted tags. The errors in the integrated plots are from the statistical errors on actual (integrated) tags only. Perfect agreement between the prediction and the actual tags would appear as a flat ratio as a function of $E_T(2)$. In all three cases, the ratio is flat within statistical errors. From this plot, we estimate that the tag bias method predicts the fraction of tagged events above a given cut to within $\pm 20\%$.

Figure 6.8 shows a comparison between the $-\log L$ distribution of tagged $W + \geq 3$-jet events and the $-\log L$ distribution predicted by the tag bias parametrization method. The shapes of the pre-tagged samples are repeated for comparison to the predicted tagged shapes. The pre-tagged curves are rescaled for clarity. Although the tagged event statistics are limited (especially in the charm Monte Carlos), the agreement is reasonable. We estimate the fraction $\epsilon_{L>1.8,tag}$ of each tagged background lying in the region $-\log L \geq 1.8$ from the predicted tagged $-\log L$ distribution shown in figure 6.8. For Wc, the predicted background fraction above $-\log L \geq 1.8$ is zero. In light of this circumstance, we require $\epsilon^{bkg}_{L,tag} \geq \epsilon_L$, where ϵ_L is the fraction of pre-tagged events in the region $-\log L > 1.8$, since we expect the tagging bias to increase the fraction of events at large $-\log L$. In the case of Wc,

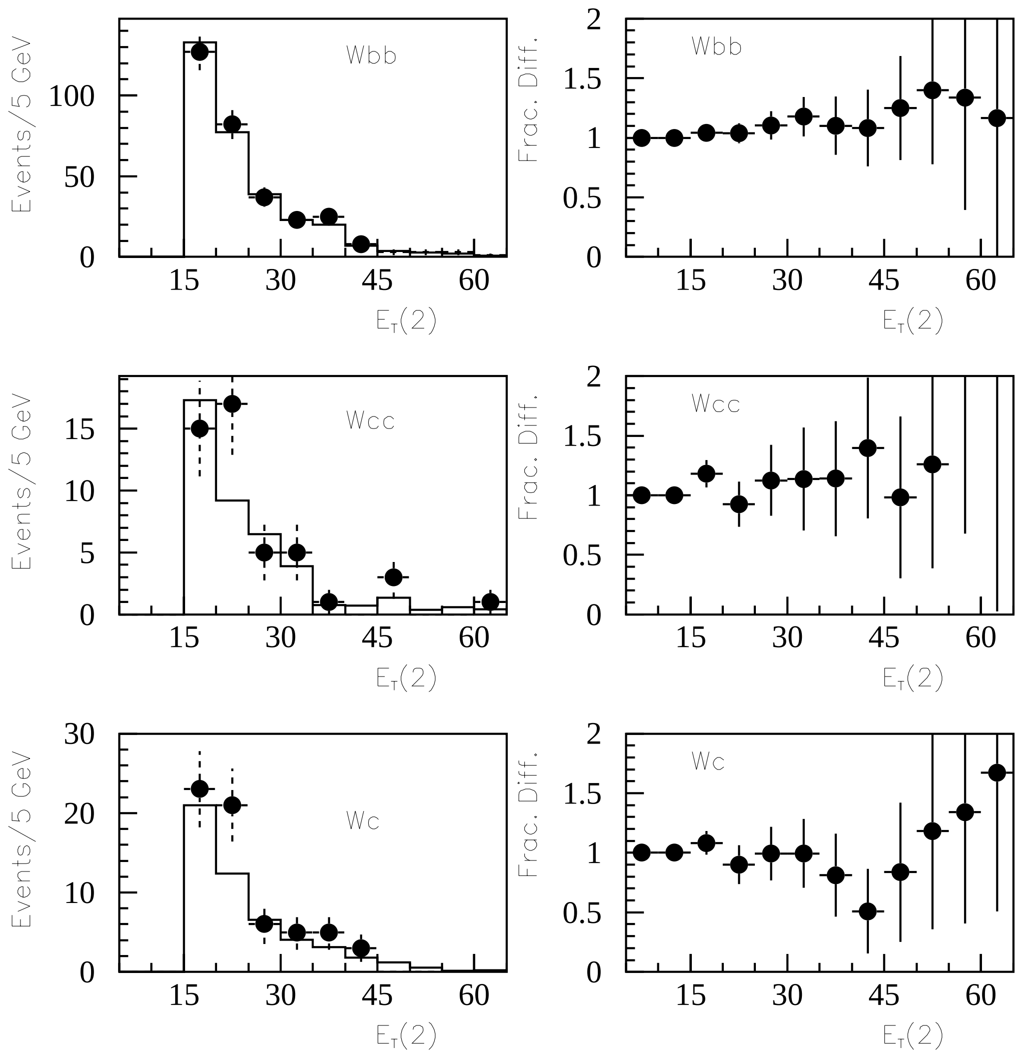


Figure 6.7: *Comparison of the actual (dashed points) to predicted (solid histogram) tagged $E_T(2)$ distribution for W+2-jet $Wb\bar{b}$, $Wc\bar{c}$ and Wc Monte Carlo samples. The prediction is extracted from W+1-jet Monte Carlo samples. The right hand side plots show the ratio of the integrated distributions and indicate that both the absolute prediction and the shape of the prediction are reliable at the 20% level.*

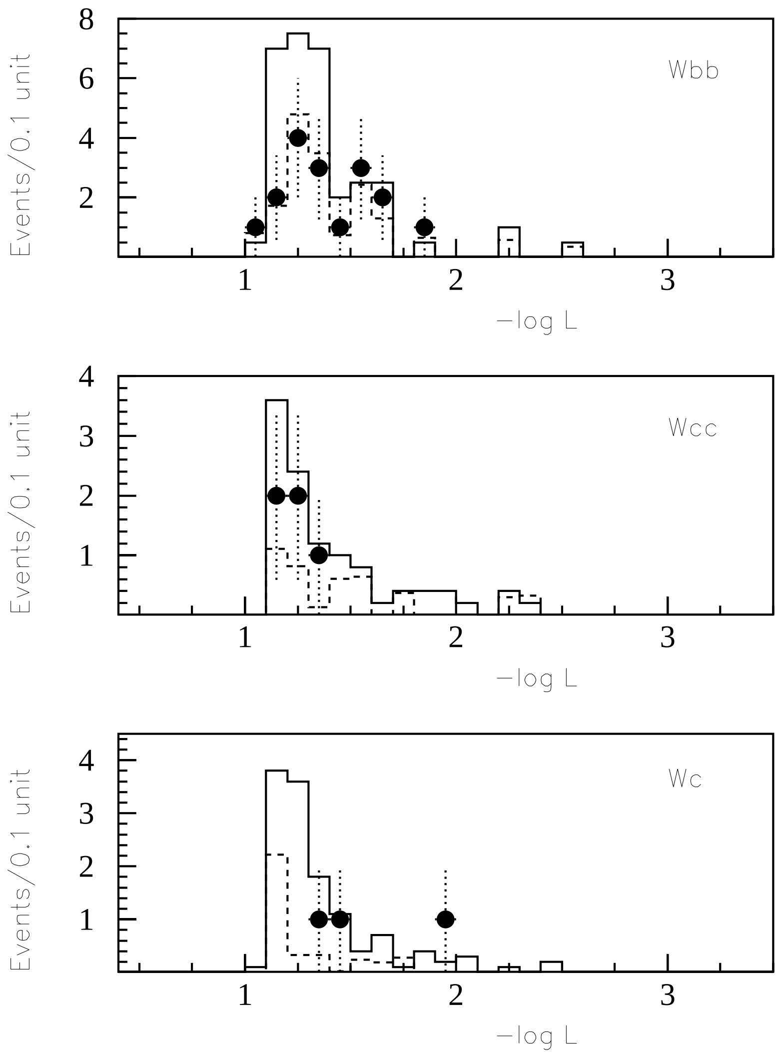


Figure 6.8: *Comparison of the tagged (points) to predicted (dashed histogram) $-\log L$ distribution for lepton+$\geq$3-jet $Wb\bar{b}$, $Wc\bar{c}$ and Wc Monte Carlo samples. The prediction is extracted from W+1-jet samples. The pre-tagged $-\log L$ shapes in these samples are also shown for comparison. The pre-tagged curves have been rescaled for clarity.*

Source	tagged fraction, $\epsilon_{L>1.8,tag}$
$Wb\bar{b}$	0.094 ± 0.019
$Wc\bar{c}$	0.15 ± 0.03
Wc	0.094 ± 0.019

Table 6.2: *Fraction of predicted tagged events passing the requirement* $-\log L \geq 1.8$ *for lepton+*$\geq$ *3-jet* $Wb\bar{b}$, $Wc\bar{c}$, *and* Wc *Monte Carlo samples. Errors are given by the* $\pm 20\%$ *systematic error.*

we replace $\epsilon^{Wc}_{L,tag}$ by ϵ^{Wc}_{L}. These fractions are listed in table 6.2.

Background Estimate for $Wb\bar{b}$, $Wc\bar{c}$, and Wc

We are now in a position to calculate the background from $Wb\bar{b}$, $Wc\bar{c}$, Wc in the region $-\log L \geq 1.8$. Equation 6.4 describes this calculation.

$$N_{bkg}(-\log L > 1.8) = N_{bkg} \times (\epsilon_{\Delta R}) \times f_{kin} \times \epsilon_{L>1.8,tag} \tag{6.4}$$

The constituent factors in equation 6.4 are as follows:

- N_{bkg} is the total background as described in chapter 4.
- $\epsilon_{\Delta R} = \frac{162}{204}$ scales N_{bkg} as normalized to the number of lepton+$\geq$ 3-jet events in chapter 4 to the number of lepton+$\geq$ 3-jet events passing the $\Delta(R) > 0.7$ cut in this kinematic analysis.
- f_{kin} accounts for our possible underestimation of the pre-tagged background Monte Carlo $-\log L$ distribution as described above and in table 6.1.
- $\epsilon_{L>1.8,tag}$ is our estimate of the fraction of tagged background events passing $-\log L \geq 1.8$, where we require that it exceed $\epsilon_{L>1.8}$.

Using equation 6.4, we calculate the $Wb\bar{b}$, $Wc\bar{c}$, and Wc backgrounds in equations 6.5, 6.6, and 6.7. The uncertainties are from (pre-tagged) Monte Carlo statistics, tag bias uncertainty ($\pm 20\%$, correlated between all three sources), and absolute normalization from chapter 4 (correlated $\pm 80\%$ for $Wb\bar{b}$, $Wc\bar{c}$; uncorrelated $\pm 50\%$ for Wc) respectively.

$$N_{Wb\bar{b}}(-\log L \geq 1.8) = 0.33 \pm 0.16 \pm 0.07 \pm 0.26 \tag{6.5}$$

$$N_{Wc\bar{c}}(-\log L \geq 1.8) = 0.10 \pm 0.03 \pm 0.02 \pm 0.08 \tag{6.6}$$

$$N_{Wc}(-\log L \geq 1.8) = 0.16 \pm 0.06 \pm 0.03 \pm 0.08 \tag{6.7}$$

In order to derive an expected $-\log L$ distribution for the background due to $Wb\bar{b}$, $Wc\bar{c}$, and Wc production, we may repeat the procedure described above for each value of $-\log L$. Figure 6.9 shows the expected $-\log L$ distribution of tagged background events resulting from $Wb\bar{b}$, $Wc\bar{c}$, and Wc production as calculated in this manner. This procedure breaks down at large values of $-\log L$, where there are no pre-tagged $Wb\bar{b}$, $Wc\bar{c}$, and Wc events. For completeness at such values, we use as our background estimate the total background times $\epsilon_L^{QCD\ W+jets}$. This quantity does not account for a tagging bias; a more realistic background calculation in this region would be necessary for analyses using a large $-\log L$ selection. The present analysis is not affected by this prescription, since it selects events at moderate values of $-\log L$, and accounts for statistical errors in the pre-tagged Monte Carlo samples.

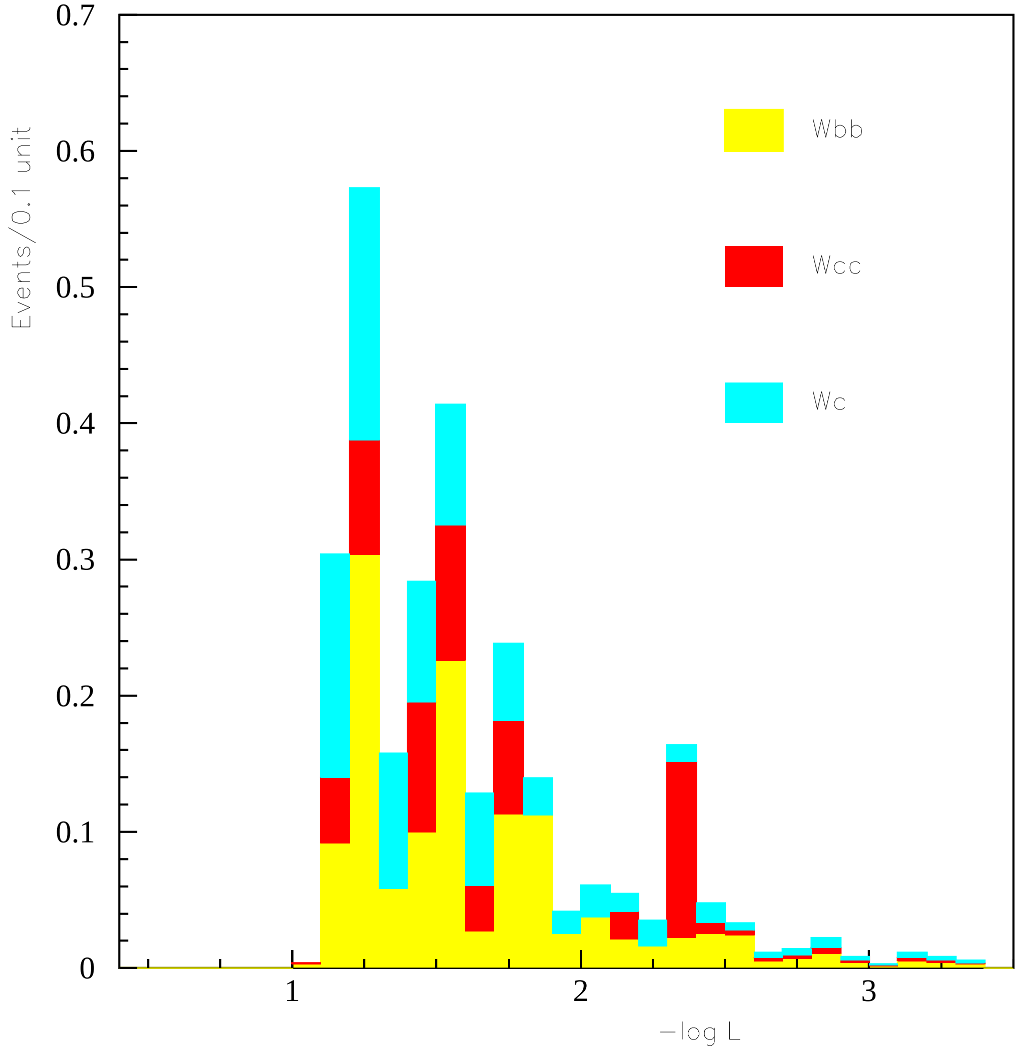


Figure 6.9: *Expected* $-\log L$ *distributions for the tagged background due to* $Wb\bar{b}$, $Wc\bar{c}$, *and* Wc *events.*

Effect of $\Delta(R)$ on Heavy Flavor Content of Jets

In equation 6.4, the factor $\epsilon_{\Delta R} = \frac{162}{204}$ is used to normalize the background to the observed number of $W+ \geq 3$-jet events after requiring a jet-jet separation of $\Delta(R) > 0.7$ for the three leading jets. This assumes that the $Wb\bar{b}$, $Wc\bar{c}$, and Wc rates calculated in chapter 4 are independent of the $\Delta(R)$ variable. Specifically, we are concerned with the effect of the $\Delta(R)$ selection on the fraction of $W+ \geq$3-jet events containing a $b\bar{b}$ or $c\bar{c}$ pair, since this value enters directly in the $Wb\bar{b}$ and $Wc\bar{c}$ background computation in chapter 4. Figure 6.10 shows the $b\bar{b}$ and $c\bar{c}$ fractions as a function of minimum $\Delta(R)$ separation in W+2-jet and W+3-jet HERWIG Monte Carlo events. These fractions differ slightly from those in tables 4.2 and 4.3 because the rescaling factor of 1.4 (see section 4.3.1) was not used and because of differences in the quark masses and structure functions used to generate these Monte Carlo samples. For the W+2-jet events, the $\Delta(R)$ separation is calculated between the two jets and for W+3-jet events, the $\Delta(R)$ separation is the minimum separation between the three leading jets. The $b\bar{b}$ and $c\bar{c}$ fractions display some structure as a function of $\Delta(R)$: generally, the fractions decrease at larger values of $\Delta(R)$. In the W+3-jet sample, requiring $\Delta(R) > 0.7$ reduces the $b\bar{b}$ fraction by 15$\pm$4% and the $c\bar{c}$ fraction by 11$\pm$4%. We do not include the effect of the $\Delta(R)$ cut on the $b\bar{b}$ and $c\bar{c}$ fractions. However, we note that this omission tends to overestimate the background slightly, although this overestimation is well within the absolute uncertainties on the background.

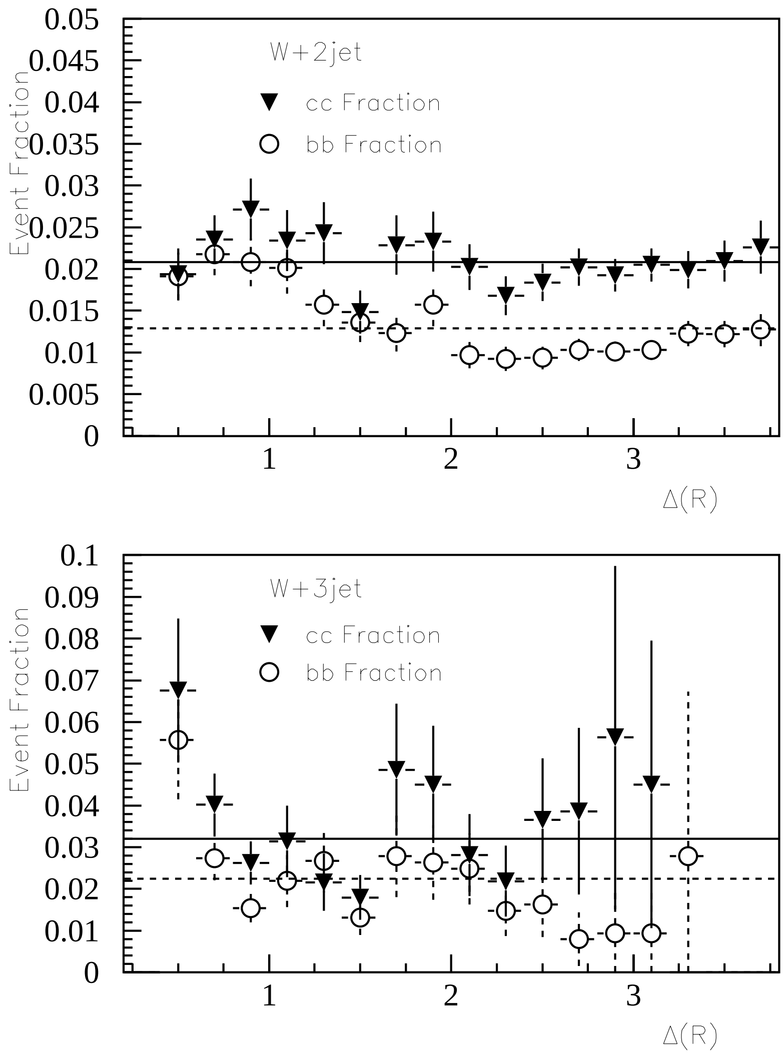


Figure 6.10: *Fraction of events containing a $b\bar{b}$ (open circles) or $c\bar{c}$ (filled triangles) pair as a function of jet-jet separation $\Delta(R)$ for W+2-jet (upper plot) and W+3-jet (lower plot) events generated using the HERWIG Monte Carlo program. The solid lines indicate the average $b\bar{b}$ fraction and the dashed lines indicate the average $c\bar{c}$ fraction.*

Additional Checks

We have calculated the $Wb\bar{b}$ and $Wc\bar{c}$ backgrounds in the region $-\log L > 1.8$ using the leading order matrix calculation to model the kinematics of these processes. For the Wc background, we use the HERWIG Monte Carlo to model the event kinematics. In both cases, we rescale the calculated backgrounds so that the pre-tagged event kinematics correspond to those of the QCD W+jet matrix element calculation via the factors f_{kin}. As a check of the reliability of this method, we may model the event kinematics using other Monte Carlo generators. For the $Wb\bar{b}$ and $Wc\bar{c}$ processes, we select two alternate calculations. We select W+3-jet events which contain a $b\bar{b}$ or $c\bar{c}$ pair. Two Monte Carlo programs are used to generate these samples. First, we use the W+1-jet matrix element and subsequent parton shower model of the HERWIG Monte Carlo program. Second, we use the W+3-parton matrix element calculation in the VECBOS Monte Carlo program. For the Wc background, we perform an alternate computation using events generated with the PYTHIA Monte Carlo W+1-jet matrix element. From these samples, we recalculate the quantities f_{kin} and $\epsilon_{L>1.8,tag}$. The results, including the corresponding background rates N_{bkg}, are listed in table 6.3. The uncertainties on N_{bkg} are from (pre-tagged) Monte Carlo statistics, tag bias uncertainty ($\pm 20\%$), and absolute normalization respectively. The backgrounds calculated using these models agree with the backgrounds calculated in equations 6.5 through 6.7 to within errors.

Process	Monte Carlo	f_{kin}	$\epsilon_{L>1.8,tag}$	N_{bkg} $-\log L > 1.8$
$Wb\bar{b}$	HERWIG	1.1±0.2	0.15±0.03	0.21±0.03±0.04±0.17
	VECBOS	1.0±0.2	0.32±0.06	0.40±0.07±0.08±0.32
$Wc\bar{c}$	HERWIG	1.5±0.2	0.13±0.03	0.10±0.01±0.02±0.08
	VECBOS	1.0±0.2	0.19±0.07	0.11±0.02±0.02±0.08
Wc	PYTHIA	1.8±0.4	0.10±0.02	0.17±0.03±0.03±0.08

Table 6.3: *Estimation of the $Wb\bar{b}$, $Wc\bar{c}$, and Wc backgrounds using alternate Monte Carlo models.*

6.2.3 Non-W, $Z \to \tau\overline{\tau}$, WW, and WZ Backgrounds

For simplicity, we model the $-\log L$ distributions for the small non-W, WW, WZ and $Z \to \tau\overline{\tau}$ background using the positive decay length $+L_{xy}$ mistag background shape. This distribution is normalized to the total background for the sum of these processes. The fraction of the $+L_{xy}$ distribution which passes the $-\log L > 1.8$ cut is 35%, which exceeds the effective tagged fractions we encountered in calculating the $Wb\bar{b}$, $Wc\bar{c}$, and Wc calculations. The background due to $Z \to \tau\overline{\tau}$, WW, and WZ is then 0.06±0.03 events, where we have included the factor $\epsilon_{\Delta R} = \frac{162}{204}$ to account for the efficiency of the $\Delta(R)$ cut, and where the uncertainty is 50% as described in chapter 4. The background due to non-W sources including direct $b\bar{b}$ production in the region $-\log L > 1.8$ is 0.26±0.07 events.

6.3 Total Background and Observed Tags

We compare the observed distribution of tagged events in $-\log L$ to the expected distribution of tagged background events in figure 6.11a. There are a total of seventeen tagged events.[6] The background distribution is the sum of the distributions shown in figures 6.9 and 6.3 with appropriate normalizations. In the figure 6.11b, this background distribution is compared to the expected $-\log L$ distribution for tagged $t\bar{t}$ events. The observed $-\log L$ shape (figure 6.11a) is consistent with a mixture of $t\bar{t}$ and background events. We observe a large excess of tagged events over the expected background in the region dominated by $t\bar{t}$ events (i.e. large $-\log L$): the data are inconsistent with the background-only hypothesis. We compare the number of observed tagged events in the region $-\log L > 1.8$ to the expected background in order to quantify this excess and its significance. The tagged backgrounds in the region $-\log L > 1.8$ are summarized in table 6.4. Of the 162 pre-tagged events, 52 lie in the region $-\log L > 1.8$. Of these, eleven are tagged. There are six tagged events in the region $-\log L < 1.8$. The total background in the region $-\log L > 1.8$ is 1.46 ± 0.43 events, where the uncertainty accounts for the correlated errors between the sources. The probability that the observed eleven tags are due to a Poisson fluctuation of the background is 4.0×10^{-6}. For a gaussian distributed variable, this probability is equivalent to 4.5σ. We have demonstrated a method of combining kinematic and b-tagging techniques in a search for $t\bar{t}$ events. This technique allows us to identify a very pure sample of $t\bar{t}$ events in which the $t\bar{t}$ signal is more significant than in either the kinematic search or the b-tagging search alone. This observation, along with figure 6.11, are the primary results of this paper; figure 6.11 demonstrates that the distribution of b-tagged events in the kinematic variable $-\log L$ is

[6] Four of the twenty-one tagged events shown in chapter 4 fail the $\Delta(R)$ selection.

Source	Events
(1) Mistags	0.55±0.11
(2) $Wb\bar{b}$, $Wc\bar{c}$	0.43±0.39
(3) Wc	0.16±0.10
(4) $Z \rightarrow \tau\tau$, WW, WZ	0.06±0.03
(5) Non-W, including $b\bar{b}$	0.26±0.07
(6) **Total Background**	1.46±0.43
(7) **Observed Tagged Events**	11
(8) **Events Before Tagging**	52

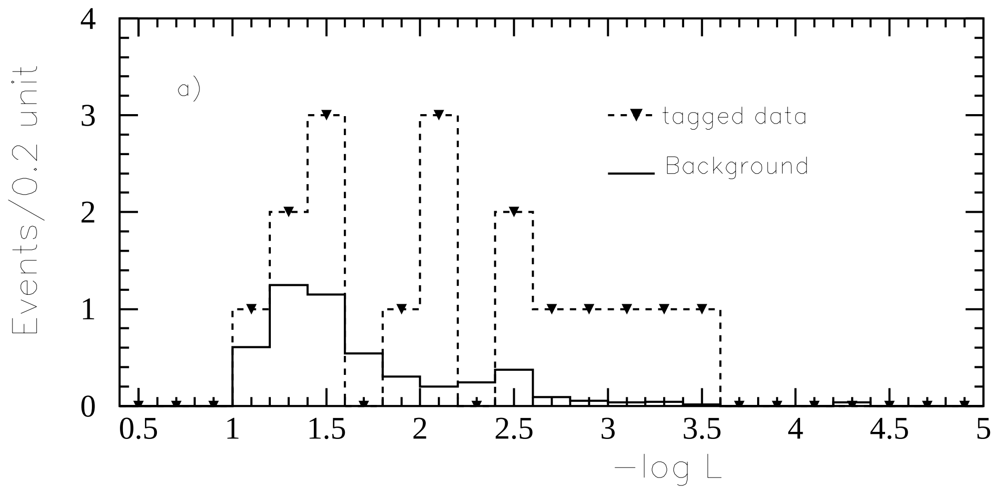
Table 6.4: *Summary of the expected number of tagged events from background sources in the lepton+jets sample after requiring* $-\log L > 1.8$. *Individual background sources are listed on lines (1) through (5) and the total is listed on line (6). The number of observed tagged events is listed on line (7). The number of pre-tagged events is listed on line (8).*

consistent with a mixture of $t\bar{t}$ and background events.

6.4 Confirmatory Background Calculation

We may also estimate the background in the region $-\log L > 1.8$ using the confirmatory background calculation method of section 4.3.2. In this method, we use the $+L_{xy}$ fake tag parametrization to estimate the contribution from $Wb\bar{b}$, $Wc\bar{c}$, and mistags. This method is expected to overestimate the actual background from these source by a factor of two or more. From the lower plot in figure 6.3, we find that the $+L_{xy}$ rate after requiring $-\log L > 1.8$ is 1.86±0.28 events,[7] where the 15% systematic uncertainty is estimated in section 4.3.1. Adding to this the background

[7]We have corrected the raw $+L_{xy}$ rate of 2.11 events to account for the fraction of events expected to be from sources other than W production, which are calculated separately.



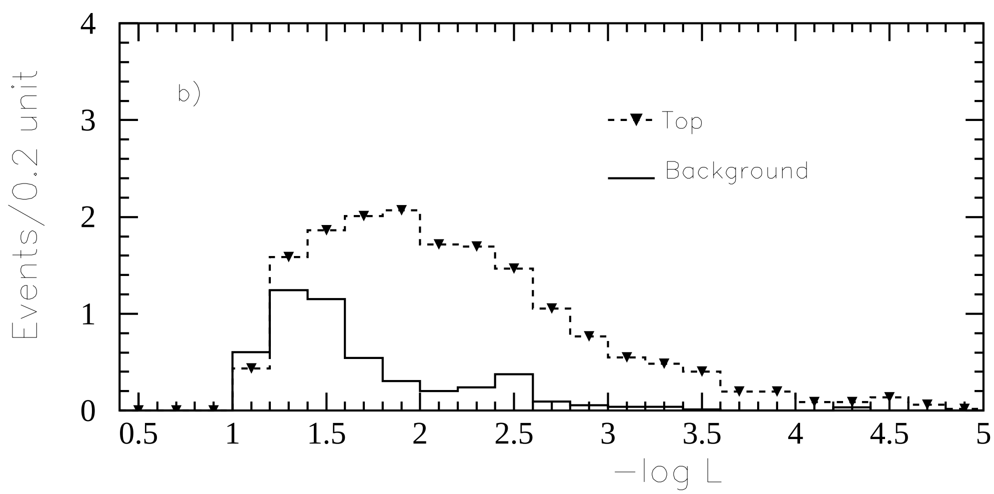


Figure 6.11: **a)** *The distribution of tagged events in* $-\log L$ *(points) versus the expected distribution of background events (solid histogram). There are 11 tagged events in the signal region* $-\log L > 1.8$ *compared to an estimated background of* 1.5 ± 0.4. **b)** *The expected distribution of tagged ttbar events in* $-\log L$ *(points), normalized to the observed 17 tagged events, versus the expected background distribution (solid histogram).*

contributions from Wc, $Z \rightarrow \tau\tau$, WW, WZ, and Non-W sources listed in table 6.11 yields a total estimated background of 2.34$\pm$0.31 in the region $-\log L > 1.8$.

Although the lepton+$\geq$3-jet sample was specifically chosen to have a large top contribution, it is possible to find regions within it which are relatively poor in top content. These regions may be used as a check on the tagged event background calculation in addition to the checks we have already made in the lepton+1-jet and lepton+2-jet events in chapter 4. In the region $-\log L < 1.8$, the background is $\sim 3.5 \pm 1.1$ events. From the lower plot in figure 6.11, we find that the fraction of tagged top events at $-\log L < 1.8$ is 34$\pm$2%. If we assume that all of the $\simeq$9.5 excess tags at $-\log L \geq 1.8$ are from top, then we expect $\simeq$5 tags from top in the region $-\log L < 1.8$. In this region, we observe six tagged events and expect 8.5 events from background and top. Given that the number of tags expected in this region from top is on the same order as the number of tags expected from background sources, we are unable to use this region to set an upper bound on the background estimate. However, the excess of tags in this region is smaller than the excess observed at $-\log L > 1.8$ and is consistent with the expectations from the sum of top and background.

Chapter 7

Conclusion

We have isolated a highly enriched sample of $t\bar{t}$ events by combining secondary vertex b-tagging and kinematic techniques in $p\bar{p}$ collision events with a lepton, evidence of a neutrino and significant jet activity. The b-tagging approach had previously been used as one component of the top discovery by CDF [1], while the kinematic approach had been used as supporting evidence for the discovery and as a consistency check for the $t\bar{t}$ hypothesis [2]. In the present paper, we have demonstrated that combining the kinematic and b-tagging approaches in the context of a search for $t\bar{t}$ production yields a more significant signal in the single lepton+multijet channel than either approach alone.

Beyond the discovery of the top quark, the kinematic and b-tagging approaches may be used in combination in analyzing the physics of top events. One example of this use is a recent CDF measurement of the CKM matrix element $|V_{tb}|$[52]. This analysis uses the kinematic likelihood technique to select a $t\bar{t}$ event sample with low background prior to the b-tag requirement. Within this sample, the relative fraction of events with two, one, or no b-tagged jets is used to extract the branching ratio

$BR(t \rightarrow bW)$ and hence $|V_{tb}|$. Beyond top physics, the combination of kinematic and b-tagging techniques may be useful in distinguishing top events as a background to events arising from more exotic phenomena such as supersymmetry or Higgs boson production. Supersymmetric top squark decay to a charged gaugino and b quark leads to the b-tag signature. The b-tag signature is also present in Higgs boson production, since the mass-generating Higgs couples most strongly to b quarks. For certain values of Higgs mass and supersymmetric particle mass the jet energies would be significantly lower leading to population of a different region of likelihood.

7.1 Prospects for Top Quark Physics

In [1], the top mass was measured in 67pb^{-1} of $p\overline{p}$ collision data using a constrained kinematic fit to the $t\bar{t}$ hypothesis for the candidate events in the b-tagged lepton+multijet sample described in chapter 4 with the additional requirement of a fourth jet. The top mass was found to be M_{top} =176±8 (stat.)±10 (syst.) GeV/c^2; updated results[26] using 109pb^{-1} of data find M_{top} =175.6±5.7±7.1 GeV/c^2. Using the single lepton and dilepton channels, the cross section for $t\bar{t}$ production was measured to be $\sigma_{t\bar{t}} = 7.6^{+2.4}_{-2.0}$pb using 67pb^{-1} of data; updated results[25] using 109pb^{-1} of data find $\sigma_{t\bar{t}} = 7.5 \pm 1.8$pb. A simultaneous publication by the D0 collaboration [53] found $M_{top} = 199^{+19}_{-21}(stat.) \pm 22(syst.)$ GeV/c^2 and $\sigma_{t\bar{t}} = 6.4 \pm 2.2$pb using 50pb^{-1} of data. These results have been recently updated[54] using 100pb^{-1} of data to $M_{top} = 170 \pm 15(stat.) \pm 10(syst.)$ GeV/c^2 and $\sigma_{t\bar{t}} = 4.7 \pm 1.6$pb. For the foreseeable future, these and other measurements relating to the properties of the top quark will be studied at the Tevatron and at the Large Hadron Collider (LHC)[55]. The Tevatron will be upgraded for data collection beginning in 1999 (Run II) with the addition of the Main Injector and with an increase of the center-of-mass energy from $\sqrt{s} = 1.8$ TeV to 2.0 TeV. Each collider experiment is expected to accumulate between 2fb^{-1} and 10fb^{-1} of data. The CDF dilepton and tagged single lepton+jet $t\bar{t}$ samples are expected to consist of over one hundred and over one thousand events, respectively. With such samples, the systematic error is expected to dominate the statistical error and uncertainties in the range $\sigma_{M_{top}} \sim 2-4$ GeV/c^2 and $\sigma_{\sigma_{t\bar{t}}} \sim 5-10\%$ are anticipated. Run II data will provide for the measurement of $|V_{tb}|$ to $\pm 10\%$, and the $V-A$ couplings of top via the W boson polarization in top decays. Rare decay modes such as $t \to \gamma c$ and $t \to Zc$ will be accessible to Run II data only if they occur at rates many orders of magnitude above Standard Model predictions. Beyond the Tevatron, the top quark will be studied at the LHC. At $\sqrt{s} = 14$ TeV and instantaneous luminosity of $10^{32} - 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, the LHC will produce about 100 tagged single lepton and 20 dilepton $t\bar{t}$ events per day, leading to $t\bar{t}$ samples 10-100 times larger than the Tevatron yield within a year of running. The top mass and cross section are expected to have uncertainties of $\sigma_{M_{top}} \sim 1-2$ GeV/c^2 and $\sigma_{\sigma_{t\bar{t}}} \sim 5-10\%$. Rare decay searches will be concomitantly more sensitive. The LHC will be sensitive to top decay to charged Higgs ($t \to H^+b$) up to $M_{Higgs} \sim 150$ GeV/c^2.

7.2 The Fourth Generation of Quarks

The observation of the top quark naturally leads to the question of whether there are further generations of quarks beyond the third, since the Standard Model does not predict the number of quark and lepton generations. However, the number of light neutrino generations may be determined from the width of the Z boson (, (Z)) by subtracting from the total width the fractional widths to hadrons and charged leptons. By "light," we mean that the mass of the neutrino is less than $M_Z/2$. The measurement of , (Z) from LEP yields $N_\nu = 2.991 \pm 0.016$ neutrino generations [9]. In the absence of a heavy fourth-generation neutrino, the top is thus the final quark if the number of quark and lepton generations are equal. Such an equality is reasonable: chiral anomalies in the Standard Model cancel if the number of quark and lepton generations are equal[56] and a natural goal of theories beyond the Standard Model is the unification of leptons and quarks into common multiplets. A direct search for the fourth generation b' quark with charge $-1/3$ found a lower mass limit of $M_{b'} > 85$ GeV/c^2[57].

7.3 Higgs Boson Mass Constraint from M_{top}

Since precision electroweak observables depend on both the top mass M_{top} and the Higgs boson mass M_H, we may use the experimental determination of M_{top} to place a constraint on M_H. Electroweak data is consistent with a top mass of $M_{top} = 179 \pm 8^{+17}_{-20}$ GeV/c^2 (see section 1.2.3). This indirect determination of M_{top} constrains the Higgs mass M_H to the lower region of the range 60 GeV/c^2 $< M_H <$ 1 TeV/c^2 with $M_H <$ 320 GeV/c^2 at the 90% C.L. This constraint is driven by the two precision electroweak measurements, R_B and A^0_{LR}, which deviate from the Standard Model prediction by $+3.7\sigma$ and $+2.3\sigma$ respectively; if the discrepancies are statistical fluctuations or are due to new physics, the constraint on M_H vanishes. Since the direct and indirect top mass measurements are very similar in both central value and precision, the constraint on M_H is not significantly affected by the inclusion of the direct value of M_{top}.

Since the Higgs mass enters into the top mass and the W boson mass M_W though radiative corrections, the determination of both M_{top} and M_W may be used as an indirect determination of M_H. Currently, the experimental uncertainties on $M_{top}$and M_W do not constrain M_H. However, these uncertainties are expected to be reduced to $\sigma_{M_{top}} \approx 2$ GeV/c^2 and $\sigma_{M_W} \approx 20$MeV/c^2 using 10fb^{-1} of Run II data at the Tevatron. Figure 7.1 indicates that uncertainties of this magnitude should be capable of placing a constraint on M_H[58]. By way of demonstration, the constraint on M_H using M_{top}=176$\pm$2 GeV/c^2 and $M_W = 80.330 \pm 0.020$ GeV/c^2 would be $M_H = 285^{+105}_{-80}$.

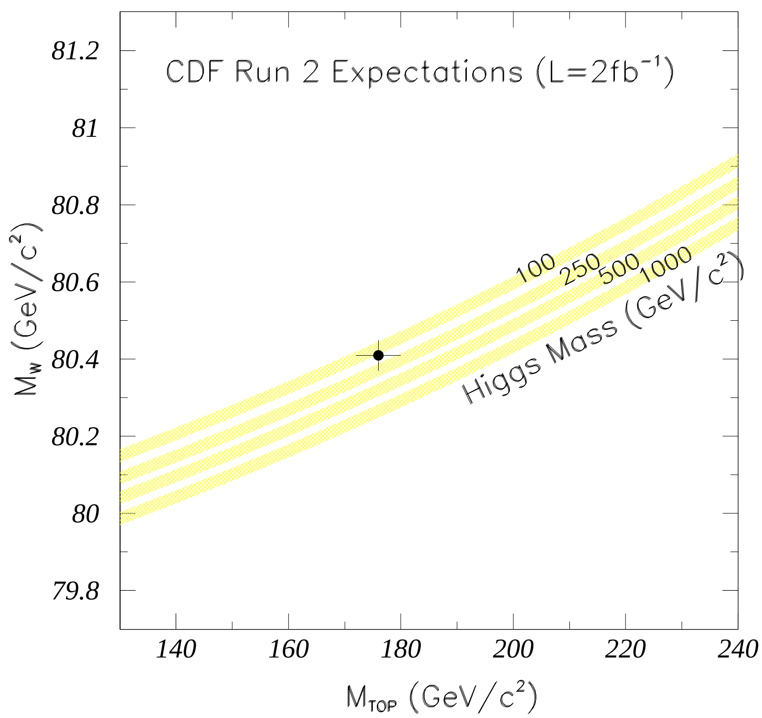


Figure 7.1: *M_W versus M_{top} for various values of M_H. The data point indicates the uncertainties on M_W and M_{top} using the Run II Tevatron dataset; such precise determinations may be used to constrain M_H.*

Bibliography

[1] F. Abe, *et al.*, CDF collab. Phys. Rev. Lett. **74**, 2626-2631 (1995).

[2] F. Abe, *et al.*, CDF collab. Phys. Rev. D **52**, 2605-2609 (1995).

[3] S. L. Glashow, Nucl. Phys. **22**, 579 (1961);
S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967);
A. Salaam, *Elementary Particle Theory: Relativistic Groups and Analyticity (Nobel Symposium No. 8)*, ed. N. Svartholm (Almqvist and Wiksell, Sweden, 1968), p. 367.

[4] S. L. Glashow, J. Illiopoulos and L. Maini, Phys. Rev. D **2**, 1285 (1970);
M. Kobayashi and M. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).

[5] I. J. R. Aitchison and A. J. G Hey, *Gauge Theories in Particle Physics, 2nd ed.* Adam Hilger, Bristol and Philadelphia, 1989.

[6] M. Perl, *et al.*, Phys. Rev. Lett. **35**, 1489 (1975).

[7] S. W. Herb, *et al.*, Phys. Rev. Lett. **39**, 252 (1977).

[8] P. G. Langacker, M. Luo, and A. L. Mann, Rev. Mod. Phys. **64**, No. 1, p. 87 (1992).

[9] A. Olshevsky, invited talk at the International Europhysics Conference on High-Energy Physics, Brussels, Belgium, Jul 27 - Aug 2, 1995.

[10] P. G. Langacker, NSF-ITP-95-140, Oct. 1995, talk given at International Workshop on Supersymmetry and Unification of Fundamental Interactions (SUSY 95), Palaiseau, France, 15-19 May 1995.

[11] P. G. Langacker, *Precision Tests of the the Standard Electroweak Model*, ed. P. G. Langacker (World, Singapore, 1994).

[12] S Abachi, *et al.*, D0 collab., Phys. Rev. Lett. **72**, 2138 (1994).

[13] F. Abe, *et al.*, CDF collab. Phys. Rev. D **50**, 2966 (1994);
F. Abe, *et al.*, CDF collab. Phys. Rev. Lett. **73**, 225 (1994).

[14] F. Abe, *et al.*, CDF collab. Phys. Rev. Lett. **73**, 220 (1994).

[15] K. Abe, *et al.*, SLD collab., Phys. Rev. Lett. **73**, 25 (1994); and SLAC-PUB-95-6979, Aug. 1995, submitted to the International Europhysics Conference on High-Energy Physics, Brussels, Belgium, Jul. 27 - Aug. 2, 1995.

[16] F. Alitti, *et al.*, UA2 collab., Phys. Lett. **B276**, 354 (1992).

[17] F. Abe, *et al.*, CDF collab. Phys. Rev. Lett. **75**, 11 (1995).

[18] C. K. Jung for the D0 collab., presented at ICHEP, Glasgow, 20-27 Jul 1994.

[19] P. Janot, invited talk at Neutrino 94, Eilat, Israel, May 1994.

[20] J. Guinion, *et al.*, *The Higgs Hunter's Guide*, Addison-Wesley, Redwood City, 1990;
M. Sher, Phys. Rep. **179**, 273 (1989).

[21] P. Nason, S. Dawson, R. K. Ellis, Nucl. Phys. **B369**, 607 (1988);
W. Beenakker, H. Kuijf, W. L. van Neerven, and J. Smith, Phys. Rev. D **40**, 54 (1989);
G. Alterelli, M. Diemoz, G. Martinelli, and P. Nason, Nucl. Phys. **B308**, 724 (1988).

[22] E. Laenen, J. Smith, and W. L. van Neerven, Nucl. Phys. **B369**, 543 (1992).

[23] E. Laenen, J. Smith, and W. L. van Neerven, Phys. Lett. **B321**, 254-258 (1994).

[24] L. Montanet, *et al.*, Phys. Rev. D **50**, 1173 (1994); and 1995 off-year partial update for the 1996 edition available on the PDG WWW pages (URL: http://pdg.lbl.gov/).

[25] P. Azzi for the CDF collab., invited talk at the Moriond Hadron Session, Moriond, France, March 23-30, 1996.

[26] L. Galtieri for the CDF collab., invited talk at the Moriond Hadron Session, Moriond, France, March 23-30, 1996.

[27] F. Abe, *et al.*, CDF collab. Phys. Rev. Lett. **73**, 2667-71 (1994).

[28] F. Abe, *et al.*, CDF collab. FERMILAB-PUB-96-004-E, Dec. 1995, submitted to Phys. Rev. D.

[29] S. D. Holmes, " A Practical Guide to Modern High Energy Accelerators", Santa Fe TASI-87;

S. D. Holmes, " Achieving High Luminosity in the Fermilab Tevatron", Fermilab CONF-91-341-E.

[30] F. Abe, *et al.*, CDF collab. Nucl. Instrumen. Meth. Phys. Res. A **271**, 387 (1988).

[31] D. Amidei, *et al.*, Nucl. Instrumen. Meth. Phys. Res. A **350**, 73 (1994).

[32] S. Cihangir, *et al.*, Nucl. Instrumen. Meth. Phys. Res. A **360**, 137-140 (1995).

[33] J. Lewis, *et al.*, internal CDF document 2858 (1994), intended for submission to Nucl. Instrumen. Meth. Phys. Res.

[34] F. Snider, *et al.*, Nucl. Instrumen. Meth. Phys. Res. A **268**, 75 (1988).

[35] F. Bedeschi, *et al.*, internal CDF document 1939 (1993).

[36] C. H. Wang, *et al.*, internal CDF document 3002 (1995).

[37] F. Abe, *et al.*, CDF collab. Phys. Rev. D **50**, 5518-5534 (1994).

[38] G. Drake, *et al.*, Nucl. Instrumen. Meth. Phys. Res. A **269**, 68 (1988).

[39] H. Bengtsson and T. Sjostrand, Procedings of the UCLA SSC Workshop 1986, p. 228 (QCD 184:U2:1986).

[40] G. Marchesini and B. R. Webber, Nucl. Phys. **B310**, 461 (1988); G. Marchesini, *et al.*, Comput. Phys. Comm. **67**, 465 (1992).

[41] F. Paige and S. D. Protopopescu, BNL Report No. 38034, 1986 (unpublished).

[42] F. A. Berends, W. T. Giele, H. Kuijf and B. Tausk, Nucl. Phys. **B357**, 32 (1991).

[43] P. Avery, K. Read, G. Trahern, Cornell Internal Note CSN-212, March 25, 1985 (unpublished).

[44] F. Abe, *et al.*, CDF collab. Phys. Rev. D **44**, 29 (1991).

[45] G. Unal, *et al.*, internal CDF document 2568 (1994).

[46] W. Yao, *et al.*, internal CDF document 2989 (1995).

[47] M. L. Mangano, Nucl. Phys. **B405**, 536 (1993).

[48] J. Ohnemus, *et al.*, Phys. Rev. D **44**, 1403 (1991);
S. Frixione, Nucl. Phys. **B410**, 280 (1993).

[49] J. Ohnemus, Phys. Rev. D **44**, 3477 (1991);

S. Frixione, P. Nason, and G. Ridolfi, Nucl. Phys. **B383**, 3 (1992).

[50] D. Gerdes, internal CDF document 2865 (1994).

[51] A. Beretvas and M. Binkley, internal CDF document 3652 (1996).

[52] F. Bedeschi *et al.*, internal CDF document 3400 (1995).

[53] S Abachi, *et al.*, D0 collab., Phys. Rev. Lett. **74**, 2632 (1995).

[54] R. Partridge and P. Bhat for the D0 collab., invited talk at the Moriond Hadron Session, Moriond, France, March 23-30, 1996.

[55] T. Liss, "The Future of Top Physics at the Tevatron and LHC," HEPPH-9510274, Jul 1995, talk given at International Europhysics Conference on High Energy Physics (HEP 95), Brussels, Belgium, 27 Jul - 2 Aug, 1995.

[56] C. Bouchiat, J. Iliopoulos, P. Meyer, Phys. Lett. **B38**, 519-523 (1972).

[57] F. Abe, *et al.*, CDF collab. Phys. Rev. Lett. **68**, 447 (1992).

[58] S. Errede, internal CDF document 3372 (1995).