Fermilab

FERMILAB-THESIS-2001-08

Upper Limits on Rare Decays of the Top Quark Using Jet-Flavor Tagging in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV

by

David Scott Wolinski

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in The University of Michigan
2001

Doctoral Committee:

Professor Dante E. Amidei, Chair
Assistant Professor David W. Gerdes
Professor Gordon L. Kane
Associate Professor John K. Riles
Professor Chien-Fu J. Wu

"So I say to you, ask and it will be given to you; seek and you will find; knock and the door will be opened to you."

Jesus of Nazareth, Luke 11:9

"...Scripture and Nature are both emanations from the divine word..."

Galileo Galilei

To my Uncle Dick, who was always so interested in everything

Acknowledgments

This thesis, as well as my entire graduate study, has benefited greatly from the collaboration and input of many people. First and foremost, I would like to thank my advisor, Dan Amidei, for his guidance throughout my entire time here: from persuading me to join the Michigan CDF group, to enduring multiple iterations of my almost complete analysis. His genuine concern for my progress and well being was reflected in his careful instruction, willingness to advocate my cause, and support over the years. I value his scholarly example and friendship.

I also wish to thank Gordy Kane, Keith Riles, and Jeff Wu, for serving on my thesis committee and making useful suggestions. I am especially grateful to Dave Gerdes, who originally showed me the ropes at CDF and finally served on my committee, providing advice and examples for my analysis in the interim. I was privileged to participate in the muon trigger upgrade with J. Chapman, someone who was patient to explain and quick to share his vast electronics expertise. I have greatly enjoyed working with Myron Campbell on many different projects. Myron was also partly responsible for bringing me to Michigan, and I have since learned much from his uncanny experimenter's ability to find clever solutions to hard problems.

I enjoyed working with Regina Demina, who contributed to the early part of this analysis. Paul Derwent supplied much technical advice and expertise on tracking problems, and I appreciate his willingness to answer thoroughly my questions on short notice.

I am thankful for tutoring from the Michigan CDF post-docs, who always entertained my questions. I enjoyed designing boards with Eric James, and I appreciate his skilled help in completing a well-defined and interesting portion of the upgrade. Stephen Miller and

Monica Tecchio provided advice and further resources for many physics problems. Ken Bloom deserves special recognition because, in addition to mentoring the latter part of my analysis, he regularly spurred me on towards completing my thesis. I thank Kevin Burkett for helping with my early analysis questions. I also thank Dave Winn for many entertaining office discussions and for demonstrating the feasibility of graduation.

I am fortunate to have an older sister and brother who have completed physics Ph.D.'s themselves. From graduate school application to thesis writing, Karen has consistently given me practical advice which has served me well. As a former CDF collaborator, Jeff greatly facilitated my indoctrination into HEP and could relate to the challenges I faced in pursuit of the degree.

I wish to express my sincere gratitude to my parents for enabling me to achieve this milestone. I appreciate their continual support for my education and their high confidence in me. Their solid backing throughout my life has been invaluable to me.

Last but not least, I want to thank my wife, Sarah, for her faithful companionship throughout graduate school. From the days of all-nighter problem sets to pseudo-experiment discussions around the dinner table, I have highly valued her keen intellect and kindness towards me. Sarah has patiently listened to more of my complaints than anyone else, but as a fellow collaborator she has also been able to offer remedies. So much of this thesis has been accelerated and improved because of her input. I count myself among the lucky physicists who can claim an excellent wife, and a Ph.D. besides, from the same experiment.

This work was supported by the U.S. Department of Energy and the Air Force Office of Scientific Research.

Table of Contents

List of Tables

Table

List of Figures

Figure

List of Appendices

Appendix

Chapter 1

Introduction

The Standard Model (SM) is a highly successful description of the fundamental constituents of matter and their interactions. Over the past several decades, the theory has been tested and refined by the results of high-energy particle scattering experiments. Among other particles, experiments have revealed the existence of six different flavors of quarks which are included among the fundamental particles in the SM. The SM makes testable predictions regarding the behavior of these quarks, leading most recently in 1995 to the discovery of the long-awaited top quark by observing its predicted decay to a bottom quark.

Although the SM theory explains the existing data remarkably well, it is nevertheless incomplete. It requires as input 19 independent parameters[1] whose presence is not explained by the SM, but rather suggests the need for an even deeper description of nature [1]. To this end, we search the experimental results for possible discrepancies with the SM as pointers to new physics. But this means interpreting the observed data in a certain way, a task which requires the identification of particles emerging from each scattering event.

We detect quarks as jets of particles, and the ability to separately identify the originating-quark flavor in a jet is of great interest in high-energy physics. Bottom-jet identification made possible the discovery of the top-quark mentioned above. Possible top-quark decay modes involving charm jets arise in a variety of SM extensions.

We present a new jet-flavor identification (tagging) technique and use it to search for

[1] These are the particle masses, coupling constants, mixing angles, and the QCD vacuum angle.

flavor-changing decays of the top-quark which are forbidden in the SM (FCNC): $t \to cg$ and $t \to cZ$. This thesis is organized as follows. In the remainder of this chapter, we describe the SM and motivate the search for FCNC decays, which is fundamentally a problem of jet-flavor identification. We then discuss the state-of-the-art of particle identification, and the ways that bottom and charm jets are produced, at a $p\bar{p}$ collider. In Chapter 2 we describe the accelerator complex necessary for the colliding $p\bar{p}$ beams and the detector apparatus used to observe the collision aftermath. We introduce our jet-flavor tagging technique in Chapter 3 with a model algorithm for combining jet-tracking information using a neural network. In Chapter 4 we derive and describe the measurement of a set of corrections which characterize the tagging-performance differences between simulation and data. In Chapter 5 we describe the experimental signatures and the selection of a search sample of candidate top-quark events. The expected composition of the search sample is given in Chapter 6. In Chapter 7, we describe the application of an optimized version of the flavor tagger to the search sample. We describe the measurement and statistical limit on the FCNC branching ratios in Chapter 8. We summarize the analysis and make some final conclusions in Chapter 9.

1.1 The Standard Model

The beginnings of modern particle physics are marked by two events earlier this century. In 1935, Yukawa hypothesized that the nuclear binding force could be the result of the exchange of a massive particle (now called a meson) between nuclear constituents. Then in 1946, Powell discovered these π-mesons in cosmic radiation, with interactions consistent with the Yukawa theory [2]. Subsequent experimental and theoretical work has culminated in the Standard Model (SM), which describes quarks, leptons, and their electromagnetic, strong, and weak interactions. The SM is a self-contained description of nature up to energies of several hundred GeV, or equivalently, down to lengths of order 10^{-18} m. The SM does not include gravity, whose effects at this scale are assume to be completely negligible.

The formalism of the SM theory is similar to classical mechanics, where the equations

of motion of a system are determined from the action of the Lagrangian,

$$\mathcal{L} = T - V$$

where T is the kinetic energy and V is the potential energy of the system. The Lagrangian completely describes the dynamics of the system. In high energy physics, particles are properly described by quantum mechanics and special relativity, the combination of which leads to quantum fields and associated particles [3]. The SM contains the field theoretic Lagrangians which embody our knowledge of interacting relativistic quanta.

Symmetry plays the central role in determining the dynamics of the SM [1]. By symmetry, we mean some change in the variables of a system which leave the essential physics unchanged. In field theory, the dynamical variables are the particle fields, and a symmetry is a transformation of the fields under which the Lagrangian, and therefore the resulting physics, is invariant. A gauge theory is a quantum field theory which includes an invariance principle, requiring the existence of both interactions among the particles and a conserved quantity called a charge [3]. This gauge invariance also allows for the vital property of renormalizability, in which various unphysical infinite contributions that routinely arise can be eliminated [4]. The SM theories of electromagnetic, strong, and weak interactions are all gauge theories, and their specific Lagrangians are known.

1.1.1 Fundamental Particles in the Standard Model

All particles in nature can be divided into two categories, fermions and bosons, based on their intrinsic angular momentum. The fermions, also called matter particles, are the spin one-half quarks and leptons, which collectively obey Fermi-Dirac statistics because each has spin one-half. The gauge bosons transmit the forces, have integer spin, and obey Bose-Einstein statistics. At least one final and yet-to-be-observed particle, the spin-zero, neutral Higgs boson, is required for a consistent theory which allows massive particles. All of these particles are truly taken as elementary; neither their number nor properties can be derived from the SM.

Quarks			Leptons		
Flavor	Mass	Charge	Type	Mass	Charge
u	0.001-0.005	$\frac{2}{3}$	ν_e	$< 3 \times 10^{-9}$	0
d	0.003-0.009	$-\frac{1}{3}$	e	5.11×10^{-4}	-1
c	1.15-1.35	$\frac{2}{3}$	ν_μ	$< 1.9 \times 10^{-4}$	0
s	0.075-0.170	$-\frac{1}{3}$	μ	0.106	-1
t	174	$\frac{2}{3}$	ν_τ	< 0.182	0
b	4.0-4.4	$-\frac{1}{3}$	τ	1.78	-1

Table 1.1: Mass (measured, in GeV) and electric charge (defined, in units of the proton charge) of quarks and leptons, the fundamental fermions in the Standard Model. The quark masses (except for t, which has been observed directly) correspond to "current masses". There is convincing evidence from the apparent observation of neutrino oscillations that the neutrinos have nonzero masses [5].

1.1.1.1 Quarks and Leptons

A quark can be defined as a fermion that carries color charge, while a lepton is a fermion without color charge. Consequently, quarks participate in all four fundamental interactions, while the colorless leptons do not feel the strong force. Table 1.1 lists the quarks and leptons, and gives each particle's mass and electric charge. Note that for each particle listed there exists an antiparticle with opposite quantum numbers and the same mass and spin.

There currently appear to be six different kinds, or flavors, of quarks in nature. The quark flavors are called: up, down, charm, strange, top, bottom, and are denoted by the first letters of their names. The quarks fall into two classes according to their electric charge: the u, c, t quarks have charge $+\frac{2}{3}$, and the d, s, b quarks have charge $-\frac{1}{3}$. Each quark flavor comes in three colors called red, green, and blue. As a consequence of the theory of the strong interaction, quarks are believed to be permanently confined inside colorless bound states. We observe two kinds of these states, called hadrons. In a meson, one quark of a certain color is bound to a second quark of that anti-color. In a baryon, three quarks of color red, green, and blue form a white (colorless) bound state. The familiar proton and neutron are examples of baryons, having valence quark contents uud and udd, respectively.

Force	Range (m)	Relative Strength[2]	Gauge Boson	Mass	Charge	Spin
Gravity	∞	10^{-41}	graviton	0	0	2
Electromagnetism	∞	1	photon (γ)	0	0	1
Weak	$\sim 10^{-18}$	0.8	$W^{\pm}$	80.4	± 1	1
			Z	91.2	0	1
Strong	$\sim 10^{-15}$	~ 25	8 gluons	0	0	1

Table 1.2: Range and strength (relative to electromagnetism) of the four basic interactions, and mass (in GeV), electric charge (in units of the proton charge), and spin of the gauge boson force carriers [5].

For reasons to be discussed in § 1.1.2.1, the quarks also naturally fall into three generations of doublets:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} \tag{1.1}$$

There are also three types of leptons: electron (e, ν_e), muon (μ, ν_μ), tau (τ, ν_τ). The leptons fall into two classes according to electric charge: the neutral neutrinos ν_e, ν_μ, ν_τ, and the negatively charged e, μ, τ. The leptons may exist freely in nature, but like the quarks, they fall into three generations of doublets:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{1.2}$$

1.1.1.2 Gauge Bosons

Because the SM is a quantum field theory, we interpret the four basic interactions among particles as the exchange of discrete quanta. These exchanged quanta are the gauge bosons, each of which is associated with a particular force. Properties of the gauge bosons are listed in Table 1.2. The graviton is the only gauge boson which has not been observed; its existence and properties are inferred from the structure of the theory [3].

1.1.2 Interactions in the Standard Model

On a cosmological scale, gravity is the dominant force which, through general relativity, determines the structure of space and time. On the nuclear scale, however, gravity is far weaker than the other forces, and is currently excluded from the SM quantum field theory. Electromagnetic forces act between electrically charged particles like electrons and nuclei, and is necessary for building the rich variety of atoms and molecules and their chemistry. The weak force causes, for example, the β-decay of the neutron and the muon, and is vital in the fusion reactions of stars. The strong force binds together quarks to make nucleons (protons and neutrons), and a residual strong force attracts the nucleons to form nuclei [6] [4].

Taking the fundamental fermions and gauge bosons as inputs to the SM, we can now discuss its dynamics. As a gauge theory, the SM involves two kinds of particles, those which carry charge and those which mediate interactions between charged particles. The former class includes the fundamental fermions and certain gauge bosons, while the latter consists only of the gauge bosons. Three kinds of charge appear in the SM: color, weak isospin, and weak hypercharge (the latter two are related to electric charge, as described in § 1.1.2.1) [1].

As mentioned earlier, the SM is a gauge theory whose dynamics reflect the internal symmetries possessed by the Lagrangian. The SM contains just three symmetries (described with group theory), and all experiments to date support the idea that these three adequately describe the interactions (besides gravity) of known particles [3].

1.1.2.1 The Electroweak Interaction

Historically, electromagnetism (EM) was the first force to be described by a gauge field theory. In 1961, Glashow unified the electromagnetic and weak interactions, and in 1968 Weinberg and Salam showed how massive gauge bosons could arise without spoiling renormalizability. The SM theory of electroweak (EW) interactions is based on the $SU(2) \times U(1)$ gauge group and is known as the Glashow-Weinberg-Salam model [6].

The gauge group $SU(2) \times U(1)$ has four associated fields, of which four combinations

appear in nature as gauge bosons. The first, the photon (γ), is the quantum of the EM field and is the carrier of the EM force. The infinite range of this force requires the photon to be massless. The other three are the $W^{\pm}$ and the Z, the mediators of weak interactions. The corresponding $SU(2)\times U(1)$ invariant piece of the SM Lagrangian describing EW interactions for first generation fermions is [3]

$$\begin{aligned}\mathcal{L}_{\rm EW} \;\approx\; & \sqrt{\alpha}\, Q_f \left(\bar{f}\gamma^\mu f\right) A^\mu \\ & + \tfrac{\sqrt{\alpha}}{\sin\theta_W \cos\theta_W} \sum_{f=\nu_e,e,u,d} \left[\bar{f}_L\gamma^\mu f_L \left(T_f^3 - Q_f \sin^2\theta_W\right) + \bar{f}_R\gamma^\mu f_R \left(-Q_f \sin^2\theta_W\right)\right] Z_\mu \\ & + \tfrac{\sqrt{\alpha}}{\sqrt{2}\sin\theta_W} \left[\left(\bar{u}_L\gamma^\mu d_L + \bar{\nu_e}_L\gamma^\mu e_L\right) W_\mu^+ + \text{Hermitian conjug.}\right] \quad .\end{aligned} \tag{1.3}$$

where Q_f is the fermion's electric charge in units of e (proton charge), f (or u, d, ν_e, e) is the fermion spin-wave function[3], γ are matrices associated with solutions to the Dirac equation, and T_f^3 is the weak isospin of the fermion, defined as $\frac{1}{2}$ for upper members of the doublets (1.1,1.2), and $-\frac{1}{2}$ for lower members. The angle θ_W is a parameter which gives the amount of mixing between the EM and weak interactions. The gauge boson fields are the photon A^μ of EM, and $W_\mu^{\pm}$,Z_μ for weak interactions. The fine structure constant α represents the strength of the EW coupling.

The three rows of (1.3) define, in order, the EM, weak neutral current, and weak charged current interactions for one fermion generation. The interactions are mediated by the photon (γ), $W^{\pm}$, and Z gauge bosons, respectively. Examples of each interaction are shown in the Feynman diagrams of Figure 1.1. In addition to fermion-fermion-gauge boson interactions, terms in the full EW Lagrangian allow gauge boson self-couplings.

The EW coupling strengths depend on the fine structure constant α and the charge of the participating fermion. This charge is the familiar electric charge for EM interactions, and the electroweak charge for weak neutral current interactions. However, α is not really a constant. Because of a screening effect due to vacuum polarization[4], the coupling strength varies with distance, or equivalently, with the square of the momentum transfer q^2 of the

[3]The subscript 'L' refers to the left-handed projection of the fermion wave function in spin space.

[4]The creation of virtual particle-antiparticle pairs, permitted by the Heisenberg Uncertainty Principle.

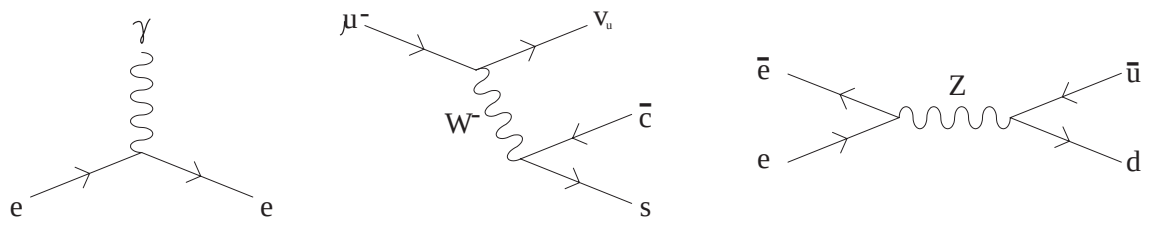


Figure 1.1: Feynman diagrams showing examples of EM, weak charged current, and weak neutral current interactions allowed in the SM.

interaction as

$$\alpha(q^2) \approx \frac{\alpha(\mu^2)}{1 + \frac{\alpha(\mu^2)}{3\pi} \log \frac{\mu^2}{q^2}} \quad ,$$

where $\alpha(\mu^2)$ is the value at some reference momentum μ^2. The EM coupling strength evidently becomes very large at high momentum transfer (short distances), at which point perturbation theory fails [6] [3].

Note that left-handed and right-handed fermions are treated differently in electroweak interactions. Only the left-handed quarks and leptons transform under electroweak $SU(2)$, and are therefore put into doublets. Right-handed fermions do not couple to the $W^{\pm}$ and occupy $SU(2)$ singlets. This helicity asymmetry of weak interactions is called parity (P) violation. Furthermore, we assume that right-handed neutrinos do not exist, an example of the weak interaction's violation of charge conjugation (C) symmetry, in which both particle and anti-particles have allowed interactions. The weak interaction is thus known to not conserve CP symmetry [3].

Remarkably, the same theory, with the substitutions $(\nu_e, e, u, d) \rightarrow \left(\nu_\mu, \mu, c, s\right) \rightarrow (\nu_\tau, \tau, t, b)$, describes the EW interactions of all three generations of particles. One feature of the Lagrangain (1.3) is that it appears to prohibit mixing between different generations of quarks. This is contradicted experimentally, since we observe, for example, b quarks which decay to c quarks. The situation is explained by the fact that the weak interaction eigenstates, appearing in (1.3), are not the same as the mass (or strong) eigenstates. The Kobayashi-Maskawa (KM) matrix relates the down-type quark mass eigenstates to their

corresponding weak eigenstates,

$$\begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\rm EW} = \begin{pmatrix} c_1 & -s_1c_3 & -s_1s_3 \\ s_1c_2 & c_1c_2c_3 - s_2s_3e^{i\delta} & c_1c_2s_3 + s_2c_3e^{i\delta} \\ s_1s_2 & c_1s_2c_3 + c_2s_3e^{i\delta} & c_1s_2s_3 - c_2c_3e^{i\delta} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\rm mass} ,$$

where $s_i = \sin\theta_i$, $c_i = \cos\theta_i$ $(i = 1, 2, 3)$. This unitary matrix has a general representation which includes three mixing angles θ_i and a complex phase δ. This phase incorporates into the SM the CP violation mentioned above [3]. The magnitude squared of a KM matrix element gives the strength of the electroweak coupling between two quarks. The KM matrix allows charged weak current interactions between quarks of different generations. Current measurements show that transitions between adjacent generations are small, and transitions across two generations are very small [5].

One feature of the SM is the absence of flavor-changing-neutral-currents (FCNC); quark flavor can only change in charged-current weak interactions. The absence of any qZq' vertex in the EW theory is provided by the Glashow-Iliopoulos-Maiani (GIM) mechanism. The mixing among different quark generations described above introduces no FCNC interactions only if each quark has a partner in its generation, with the left-handed components in $SU(2)$ doublets, and the right-handed components in $SU(2)$ singlets [4]. FCNC are almost completely absent experimentally [5]; their observation would be an interesting probe of new physics. Experimental searches for forbidden decays place very low limits ($\sim 10^{-10}$) on the ratios of FCNC-to-SM decay rates for light quarks (u, d, s), and somewhat looser limits ($\sim 10^{-6} - 10^{-7}$) for c and b quarks. Possible FCNC top-quark decays are discussed in § 1.2.

Finally, we note that the electroweak theory, which contains massive fermions and gauge bosons, would normally not be renormalizable. This difficulty is avoided by introducing a scalar Higgs boson which has a non-zero vacuum expectation value, effectively breaking the exact $SU(2) \times U(1)$ symmetries. In this process of spontaneous symmetry breaking, the fermions and the W and Z bosons acquire non-zero masses, and renormalizability is retained. The Higgs boson has not yet been observed, but is expected to have a mass of order 100 GeV/c^2 [6].

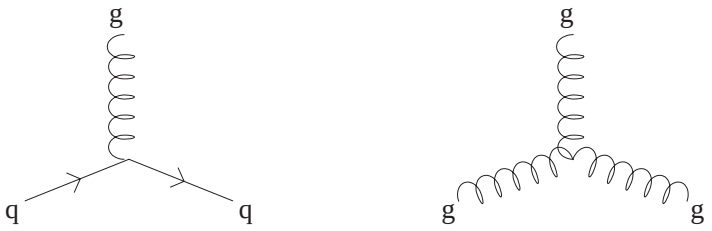


Figure 1.2: Feynman diagrams showing the basic QCD interactions among quarks and gluons (color indices are omitted), the colored particles of the SM.

1.1.2.2 The Strong Interaction

Quantum chromodynamics (QCD), developed in the 1970s, is the $SU(3)$ nonabelian gauge theory of the strong interaction of particles. Only particles which carry color charge participate in the strong interaction, which is invariant under transformations among the three colors (red, green, and blue). The quarks are the only colored fermions. The gauge bosons of QCD are the gluons, which also carry color charge. The gluons generate transitions from one quark color to another as given by terms of the QCD Lagrangian

$$\mathcal{L}_{\rm QCD} \;=\; \frac{\sqrt{\alpha_{\rm s}}}{2} \sum_{q=u,d,c,s,t,b} \bar{q}_\alpha \gamma^\mu \lambda^a_{\alpha\beta} q_\beta G^a_\mu \quad , \tag{1.4}$$

where $\sqrt{\alpha_{\rm s}}$ is the QCD coupling strength, q_α is the quark spin-wave function, the γ and λ are matrices necessary for the $SU(3)$ gauge invariance, and G_μ is the gluon field. There are eight different color combinations of gluon states. Therefore, in addition to quark-gluon interactions, QCD allows gluon-gluon interactions, as shown in Figure 1.2.

The gluon self-interactions provide an anti-screening effect which has profound consequences in QCD (and opposite to the screening effects in QED). The QCD coupling strength varies with momentum transfer q^2 as

$$\alpha_{\rm s} \;\approx\; \frac{\alpha_{\rm s}}{1 + \frac{\alpha_{\rm s}(\mu^2)}{12\pi}\left(33 - 2n_f\right)\log\left(\frac{q^2}{\mu^2}\right)} \quad ,$$

where n_f is the number of quark flavors (presently = 6), and $\alpha_{\rm s}$ is the strong coupling measured at some reference momentum μ^2. Therefore, the coupling decreases at small distances ($q^2 \to \infty$), giving rise to the so-called "asymptotic freedom", the fact that quarks and gluons

inside a hadron (generally called partons) behave like free particles. One the other hand, α_s diverges at large distances, signaling the breakdown of perturbation theory. Nevertheless, we expect that the strength of the color force between particles increases as they are pulled apart. As a result, quarks and gluons are confined inside bound states as described in § 1.1.1.1.

1.2 FCNC Decays of the Top Quark

The top quark is the heaviest known fundamental particle, and as such, it may provide a unique window to physics beyond the SM. The SM predicts that top quarks will decay almost exclusively to Wb, and its discovery in 1995 was made by interpreting candidate top events using this decay topology. However, the rare decays of the top quark are of great theoretical interest because they are sensitive to certain SM parameters and the effects of possible new physics.

Figure 1.3 shows example Feynman diagrams for flavor-changing decay modes of the top quark to a charm quark and a gauge boson (γ, Z, org) which are allowed in the SM. These allowed decays, which have the same final state as the forbidden FCNC decays, are exceedingly rare[5]. The branching ratios have been shown to be $\sim 10^{-12}$ for $t \to c\gamma$, $\sim 10^{-13}$ for $t \to cZ$, and $\sim 10^{-10}$ for $t \to cg$ [11]. The flavor-changing decay rates can be increased by several orders of magnitude in certain extensions to the SM such as multi-Higgs-doublet models, and to $\sim 10^{-6}$ in the Minimal Supersymmetric Standard Model [12]. Even in the most optimistic models, however, the flavor-changing top-quark decay rates are unobservably small, and any such observation would challenge the short-distance structure of the electroweak and QCD interactions of the SM.

Direct experimental limits on the rate of top-quark FCNC decays come from the

[5]Each one-loop process is suppressed by a W propagator and the GIM mechanism.

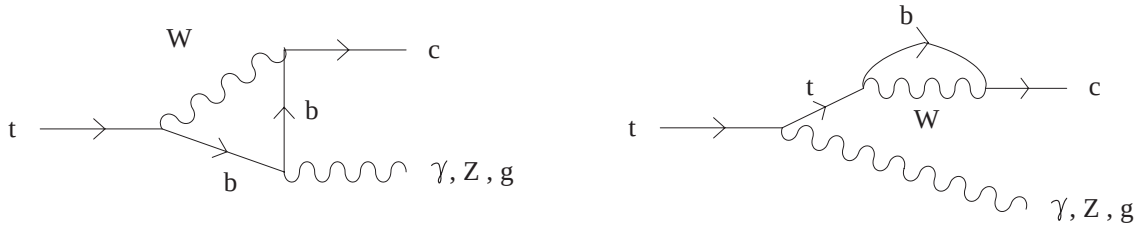


Figure 1.3: Examples of allowed flavor-changing decays of the top quark to a charm quark and gauge boson. Note that the internal particles may by off-shell.

Collider Detector at Fermilab (CDF) [13]:

$$\begin{aligned} B(t \to q\gamma) &< 3.2\% \\ B(t \to qZ) &< 33\% \end{aligned} \tag{1.5}$$

at 95% confidence level, where q represents a u or c quark. The limits are set by measuring the ratios of detected events with FCNC and SM signatures. A better limit on the second rate is obtained indirectly from a limit on the cross section for single top-quark production in e^+e^- collisions [14]:

$$B(t \to qZ) < 22\% \quad . \tag{1.6}$$

Currently, there are no published experimental limits on the rate of $t \to cg$, presumably because of the difficulty in distinguishing between charm and gluon jets.

In this thesis we set limits on the FCNC processes $t \to cg$ and $t \to cZ$ in $t\bar{t}$ events. In both cases, our technique requires the identification of the originating-quark flavor of jets appearing in the detector. This test of the SM therefore becomes a problem of particle identification, which we discuss in general terms in the next section.

1.3 Particle Identification

Particle collisions (hard scattering) are used to probe matter and the forces within. During such a collision, many new particles can appear. For example, an interaction of two initial energetic particles can create heavier short-lived particles which one wishes to study. These unstable particles then decay, perhaps successively, into lighter, stable particles.

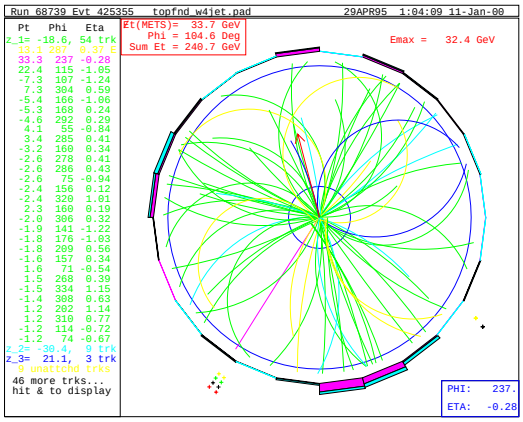


Figure 1.4: Reconstructed charged particles in a $t\bar{t}$ event.

The aftermath of a typical $\bar{p}p$ collision at Fermilab can include several hundred particles. Although the experimenter's goal is to measure some properties of an interesting intermediate particle, one is left with a messy final state (Fig. 1.4).

The first step in any particle physics analysis is, therefore, the identification of final-state particles. These particles, which are sufficiently long-lived to be detected, include the lighter leptons and hadrons.

Figure 1.5 shows a typical arrangement of a detector used to observe high-energy collisions, along with the behaviors of different particles in various parts of the detector. The collision occurs inside the beam pipe, whose axis is perpendicular to the plane of the page. Detectors contain several subsystems, each of which is designed to gather information for certain types of particles. Detection methods for each particle type are described in the next sections.

1.3.1 Bosons

$W^{\pm}$ and Z bosons have a lifetime of about 10^{-25} sec [5]. At typical collider energies, the boson will travel about 10^{-14} cm in this time, which is far too small to detect. These bosons are instead identified by their decay products.

Photons are stable but will lose energy by interacting with electrons in matter. Since they are not charged, they leave no track in the tracking chambers, but eventually deposit

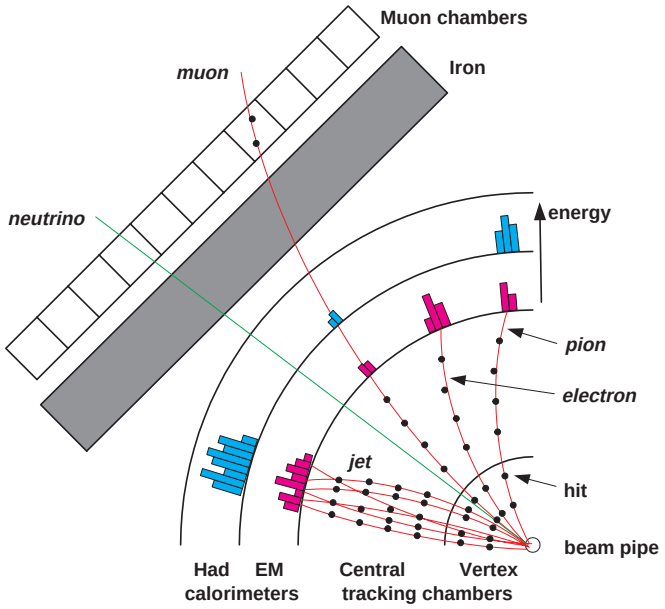


Figure 1.5: Elements of a typical particle detector in a high-energy physics experiment, and the behavior of certain particles [15].

their energy in the EM calorimeter, forming a shower of electrons. The signature of a photon is EM calorimeter energy with no associated track.

Gluons are not directly observable, but can be inferred by reconstruction of associated hadrons, discussed below.

1.3.2 Leptons

Neutrinos only interact weakly, so they can be detected only indirectly by energy imbalance. τ leptons decay into lighter particles before entering the detector.

Electrons have electric charge, and can be detected by electromagnetic (EM) interactions. Moving outward from the collision, an electron first encounters tracking chambers which record its path. Because these chambers reside inside a magnetic field, the electron path is helical and the curvature gives the momentum and sign of the charge. Next, an electron enters the EM calorimeter where it undergoes multiple Coulomb scattering and rapidly loses energy by bremsstrahlung. Thus the signature of an electron is a reconstructed track pointing at a deposit of energy in the EM calorimeter.

Muons have electric charge but are about 200 times heavier than electrons (and they live long enough to be detected because of relativistic boost). Like electrons, they leave a curved track. But in passing through matter they instead lose energy primarily by ionization (not bremsstrahlung[6]). Muons at Fermilab are usually sufficiently energetic to be minimum ionizing particles (their energy loss by ionization is at its minimum). Therefore muons are the only particles likely to survive out to the muon chambers. The signature of a muon is a central track pointing at hits in the muon chambers and minimal calorimeter energy.

1.3.3 Hadrons

The unique identification of hadrons is more difficult. Tracking chambers are used to measure the momentum of charged hadrons. All hadrons will scatter strongly off of protons (resident in the calorimeter) which will then recoil and cause detectable ionization. As a result, they are likely to penetrate the EM calorimeter and eventually be absorbed in the hadronic calorimeter.

Unlike electrons and muons, which have dedicated detector systems, no unambiguous identification can be made for hadrons. As mentioned earlier, there is a wide variety of hadrons, and unique identification would be equivalent to determining the mass of the hadron. Since the momentum p of a charged hadron is measured, its mass m could be determined by an additional measurement of its velocity v via the relation:

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}}$$

One velocity measurement technique, called time-of-flight, allows a measurement of the time interval, and therefore the velocity, of a particle traveling a known distance. Another technique, called dE/dx, measures the rate of energy loss of the particle which is related to its velocity [7].

Fortunately, the precise identification of each hadron in the final state is often not necessary. The process of an initial parton (quark or gluon) forming bound hadronic states

[6] Radiative energy loss is proportional to the inverse square of the particle mass

[7] Unfortunately, neither velocity measuring technique can be done well at CDF (see Chapter 2).

is called fragmentation. During fragmentation, many light hadrons are created until the original parton is embedded inside a (roughly collinear) *jet* of hadrons. Therefore a jet consists of a collection of particle tracks pointing at a deposit of energy in the hadronic calorimeter. One usually wishes to know the flavor of the parton which originated the jet, not a detailed accounting of every hadron in the final jet.

The identification of the flavor of an initial quark becomes the following problem. First, a hadronic jet of particles must be detected. Secondly, we must discriminate different original quark flavors (called jet flavors). Note that the former is accomplished by a dedicated hardware detection system (the hadronic calorimeter), but the latter must be accomplished in subtler ways using information from a variety of detector systems.

Finally, we need not distinguish all parton flavors. Our physics goals require discrimination between the following three jet flavors: b, c, and "prompt". Prompt jets originate with u,d,s quarks or gluons. Note that we do not distinguish quark-antiquark flavors.

1.4 Heavy Flavor Production at the Tevatron

In this thesis we are particularly interested in identifying b and c jets in the environment of $p\bar{p}$ collisions at Fermilab. The characteristics of these jets depend on the way in which they are produced, which is therefore an important consideration when measuring the performance of a jet-flavor identification technique (Chapter 4). Here we briefly describe the various production mechanisms for heavy-flavor (b, c) quarks. Note that here we focus on the primary production of heavy flavor. Top-quark decays to heavy flavor are discussed in Chapter 5.

The Fermilab Tevatron is a proton-antiproton ($p\bar{p}$) collider (see Chapter 2). The parton model describes a proton as a bound state of quarks and gluons (§ 1.1.1.1), and a $p\bar{p}$ collision is really an interaction among some of the partons, with the remaining partons as non-participating "spectators". Partons emerging from the interaction, which must also form bound states, are detected as jets of particles. Most $p\bar{p}$ collisions are low-momentum-transfer processes with probabilities (cross sections) that cannot presently be calculated. However,

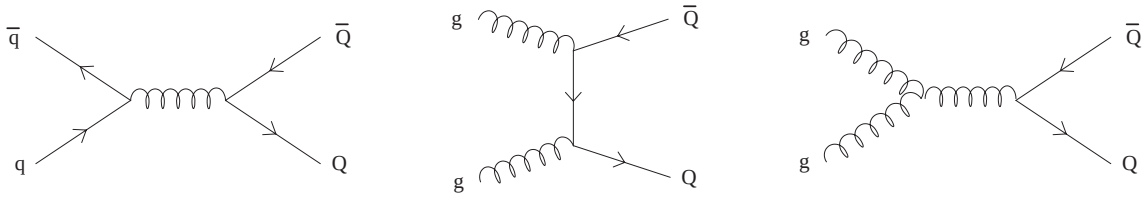


Figure 1.6: Feynman diagrams showing the leading-order QCD direct production (dp) of heavy-flavor quark-antiquark pairs.

certain parton-parton scattering cross sections do have perturbative QCD predictions (below).

The QCD cross section for b and c jet production is given by the convolution of the momentum distribution of two interacting partons i (from the p) and j (from the $\bar{p}$) with the short-distance partonic (or constituent) cross section,

$$\sigma\left(p\bar{p} \;\rightarrow\; Q\bar{Q}+X\right) \;=\; \sum_{ij}\int dx_i\, dx_j\, F_i^p(x_i)\, F_j^{\bar{p}}(x_j)\, \hat{\sigma}\left(ij \;\rightarrow\; Q\bar{Q}+X\right) \quad ,$$

where Q denotes a b or c quark, X refers to spectators and other possible final-state particles, and the structure function $F_{i/j}^{p/\bar{p}}(x)$ gives the probability that parton i/j carries momentum fraction $x_{i/j}$ in the proton/antiproton. The structure functions are determined empirically, but typically quarks and gluons carry less than 20% of the proton momentum [4]. The sum is over all possible interacting parton combinations ij. The constituent cross section $\hat{\sigma}$, which also depends on the i/j momenta, is calculated as a perturbative series in powers of α_s [3].

The leading-order (LO) term of the cross-section is of order α_s^2, and gives the contribution to heavy-flavor production from quark-antiquark annihilation and gluon fusion. Feynman diagrams for these processes are shown in Figure 1.6. These processes are well-understood at QCD leading-order given a choice of structure functions [7] [8] [9]. We refer to these $2 \rightarrow 2$ processes as the direct production (DP) of heavy flavor since the scattering results in a $Q\bar{Q}$ pair only.

The next-to-leading-order (NLO) processes are of order α_s^3, and include the $2 \rightarrow 3$ diagrams shown in Figure 1.7. In flavor excitation (fe) events, an initial-state (e.g. in the proton) gluon splits into a $Q\bar{Q}$ pair, of which one quark scatters with a gluon (in the antiproton). In gluon splitting (gs) events, a final-state gluon splits into a heavy-quark pair.

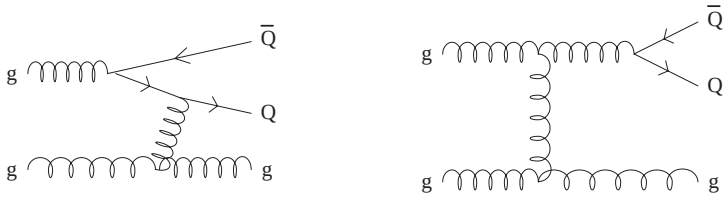


Figure 1.7: Feynman diagrams showing the next-to-leading-order QCD production of heavy quarks: flavor excitation (fe) (left), and final-state gluon splitting (gs) (right).

Not shown, but also of order α_s^3, are the diagrams of Figure 1.6 in which one of the final-state quarks radiates a gluon.

The NLO processes are additional production mechanisms, not corrections to, the LO processes. Even though the NLO diagrams are higher order in α_s, they are dominant over the LO diagrams by a factor of ~ 10 at the Tevatron. This is because at large momentum-transfer, the gluon structure functions are large, and the underlying $gg \to gg$ cross section exceeds the fusion process $gg \to Q\bar{Q}$ by over two orders of magnitude [10].

The kinematic configuration of heavy-quark pairs produced in $2 \to 2$ and $2 \to 3$ interactions are quite different. Quarks which are directly produced ($2 \to 2$) tend to have momenta in opposite directions, and are therefore likely to be resolved as separate jets in the detector. Quark from $2 \to 3$ processes tend to have aligned momenta, and as a result, may often merge into a single observed jet. As shown in Chapter 4, our ability to separately identify b and c jets depends on their production mechanism.

Chapter 2

Experimental Apparatus

The high-energy frontier of particle physics is at the Fermilab Tevatron Collider. Completed in 1984 on the site of Fermi National Accelerator Laboratory (FNAL) in Batavia, Illinois, the Tevatron has provided the highest-energy collisions ever achieved by an accelerator, and will continue to do so through much of the next decade. The Tevatron collides counter-rotating beams of protons (p) and antiprotons ($\bar{p}$) at a center-of-mass energy $\sqrt{s} = 1.8$ TeV. The aftermath of collisions is studied at two general-purpose detector facilities, the Collider Detector at Fermilab (CDF) and the D0 detector.

The operations leading up to the study of $p\bar{p}$ collisions can be broken into two categories: the creation, acceleration, and collision of the $p\bar{p}$ pairs, and the observation of the particles emerging from the collision. It is beyond the scope of this thesis to describe in detail all necessary accelerator and detector components. Instead, we provide a broad overview of the major systems most relevant for collecting data from the 1994–1996 period which is studied in this analysis.

2.1 FNAL Accelerator Complex

In this section we follow the protons and antiprotons through their successive stages of preparation for high-energy collisions. Figure 2.1 shows a schematic diagram of the FNAL accelerator complex [16].

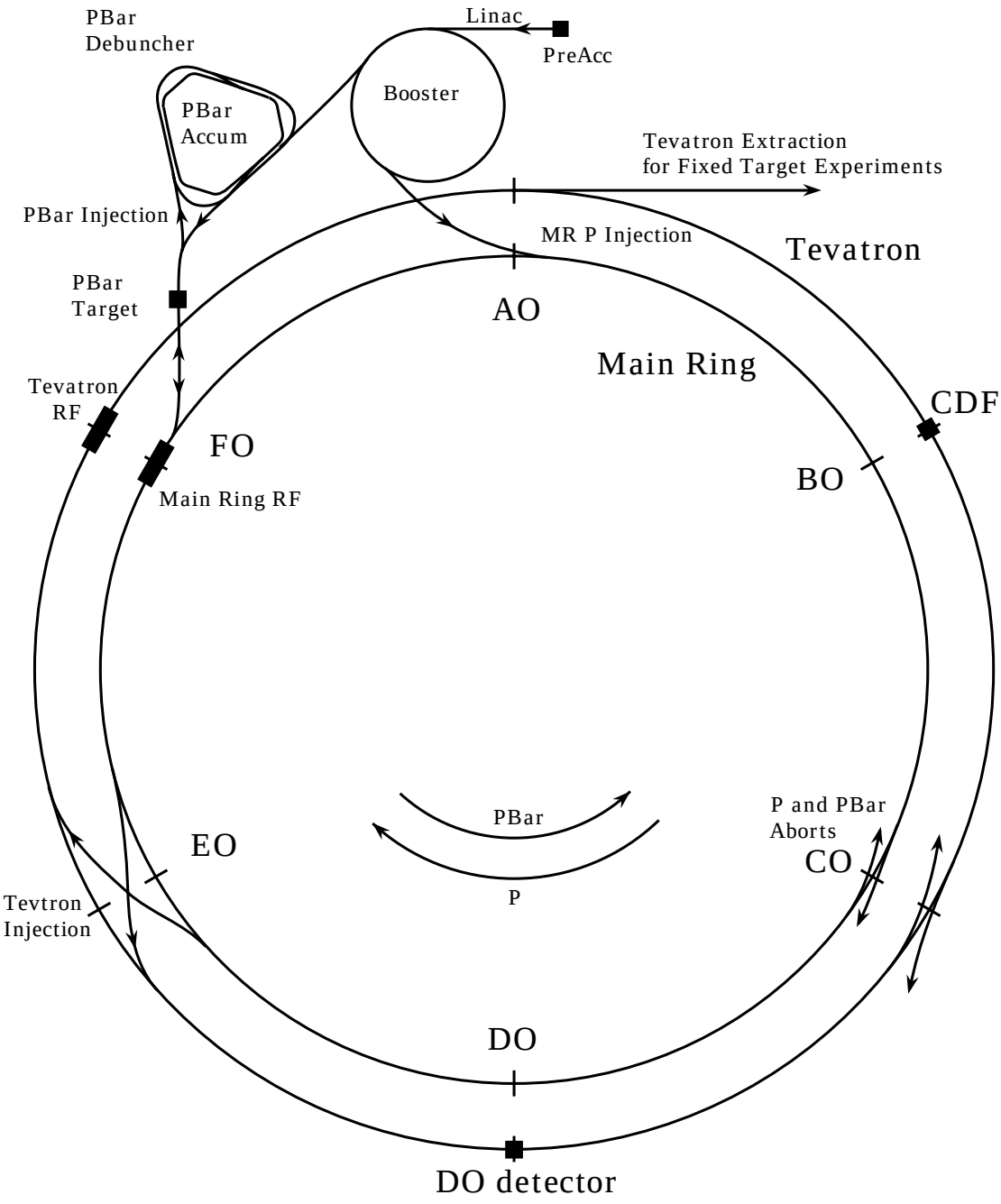


Figure 2.1: Schematic diagram of the FNAL accelerator complex. The Tevatron and Main Ring share the same tunnel which follows a circle of radius 1 km. The $p\bar{p}$ collisions occur at the sites of the two detector facilities, CDF and D0.

The protons which are eventually annihilated in $p\bar{p}$ collisions begin their life at FNAL inside hydrogen gas atoms. Electrons are added to the hydrogen atoms to yield ionized hydrogen (H^-). Inside the Preacc (a.k.a. Cockcroft-Walton), the ions are electrostatically accelerated to an energy of 750 keV.

The hydrogen ions next enter a 150 m linear accelerator, called the Linac, which accelerates them to 400 MeV. Upon their exit, the ions pass through a carbon foil which removes their electrons, leaving only the protons.

The protons then enter the Booster, a circular accelerator, or synchrotron, of radius 75 m. Magnets placed around the Booster bend the path of the protons in a circle as they experience repeated acceleration in radio-frequency (RF) cavitities. After an average of 20000 revolutions, the proton energy is raised to 8 GeV, and the protons move on to the Main Ring.

The Main Ring is a synchrotron of radius 1 km, with a tunnel 10 feet in diameter, buried 20 feet underground. It can accelerate protons from the Booster to 120 GeV (for the Antiproton Source), or to 150 GeV (for the Tevatron). It can also accelerate antiprotons from the Antiproton Source to 150 GeV. During their average acceleration time of 2.4 seconds, the (anti)protons are continually bent and focused by 1000 magnets around the ring.

The Antiproton Source is comprised of a target source, two rings called the Debuncher and Accumulator, and the transfer lines which connect them to the Main Ring. The 120 GeV protons from the Main Ring collide with a tungsten target, producing a host of secondary particles, some of which are antiprotons. A portion of the secondaries are magnetically collimated by passing through a low-density lithium lens. The pulses of 8 GeV antiprotons then go to the Debuncher, a rounded triangular-shaped synchrotron with mean radius of 90 m. Its main purpose is to reduce the momentum spread and to damp oscillations (perpendicular to the orbit) of the antiprotons by RF cooling. This necessary step improves the Debuncher-to-Accumulator transfer rate, which is limited by the momentum aperture of the Accumulator. The antiprotons are then passed to the Accumulator, another synchrotron in the same enclosure as the Debuncher. Here they are stored and cooled at 8

GeV until about 10^{12} have been collected, when they are then passed to the Main Ring. The antiproton production process is inherently inefficient: for every 10^5 protons on the target, only 1-2 antiprotons are stored.

The Tevatron provides the final stage of acceleration, accepting protons and antiprotons from the Main Ring and increasing their energy to 900 GeV. Sharing the same tunnel as the Main Ring, the Tevatron is also a 1 km radius synchrotron with RF accelerating cavities[1], but has the distinction of containing 1000 cryogenically-cooled superconducting magnets. The large magnetic fields produced by these magnets are necessary to steer and focus the (anti)protons which are traveling at 99.9999% of the speed of light. When operating in collider mode, the Tevatron circulates six bunches each of protons and antiprotons in counter-rotating helical paths in the same beam pipe. The beam paths intersect, allowing $p\bar{p}$ collisions, at just two locations on the ring, inside the CDF and D0 detectors. The colliding beams have an approximate Gaussian profile in the plane transverse to the orbit, with typical widths of 35 μm. The interaction point is also approximately Gaussian along the beam direction, with a characteristic width of about 30 cm. Collisions occur every 3.5 μs, and the peak instantaneous luminosity achieved at the Tevatron was more than 2×10^{32} cm^{-2} s^{-1}.

2.2 Collider Detector at Fermilab

CDF is a general-purpose detector system designed to study $p\bar{p}$ collisions at the Tevatron. This 5000-ton device, built in the 1980s and since upgraded several times, is located at the B0 collision point. Figure 2.2 shows an isometric view of CDF as it envelopes the 1.5-inch Tevatron beam pipe. The detector is approximately cylindrically and forward-backward symmetric, reflecting the symmetry of the $p\bar{p}$ collisions. Individual detector subsystems have been documented in detail elsewhere [17] [18]. Here we will briefly describe the major components with particular emphasis on those most relevant to this analysis.

Figure 2.3 shows a schematic cross-section view of one quadrant of CDF. The figure

[1]Synchrotrons cannot use an electro*static* field to accelerate particles because of the Maxwell equation $\oint E \cdot dl = 0$. The work done on a particle by such a field is zero after each revolution.

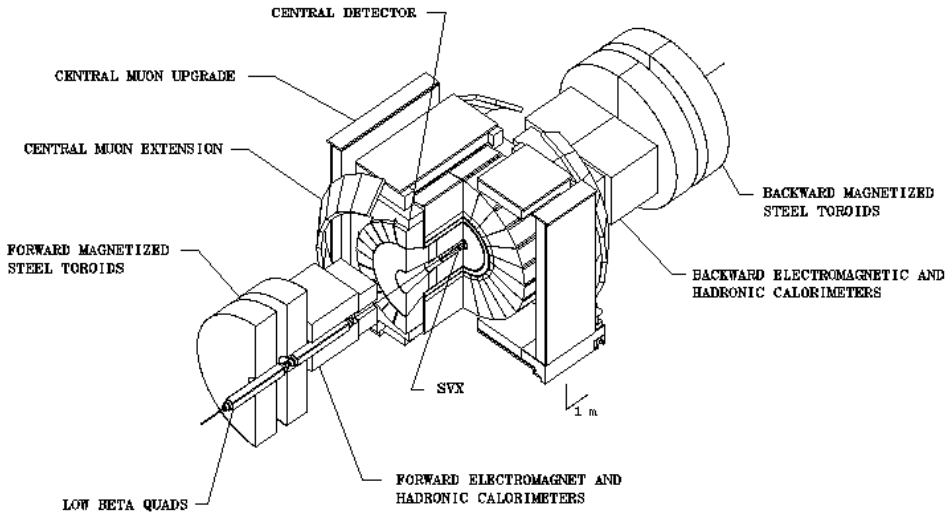


Figure 2.2: Isometric view of the CDF detector surrounding beam pipe. The detector is cylindrically and forward-backward symmetric about the interaction region at its center.

defines the detector Cartesian (x, y, z) and spherical (θ, ϕ) coordinates systems, whose origin is at the center of the detector at the nominal $p\bar{p}$ interaction point. The z-direction is along the beam pipe in the proton direction, while the x-y plane is transverse to the beam pipe. The azimuthal angle ϕ is measured in the transverse plane about the beam pipe axis, and the polar angle θ is measured from the z-axis. The pseudorapidity $\eta = -\ln\left(\tan\frac{\theta}{2}\right)$ is a convenient alternate coordinate[2], and several detector subsystems are segmented in the η-ϕ plane.

CDF is designed to detect particles with high momentum in the transverse plane, which tend to originate from hard scatterings, as opposed to particles from elastic scatterings which are more forward. The detector covers three different rapidity regions: central ($|\eta| < 1.1$), plug ($1.1 < |\eta| < 2.$), and forward ($2.2 < |\eta| < 4.2$). The central region is instrumented with the highest-resolution subsystems, described in the following sections.

[2]For light particles, $\eta \simeq y = \frac{1}{2}\ln\left(\frac{E+P_z}{E-P_z}\right)$, where y is the rapidity. It can be shown that a Lorentz boost in the z-direction produces an additive change in the rapidity, leaving any rapidity distribution dN/dy unchanged and therefore independent of the z-velocity of the reference frame [5].

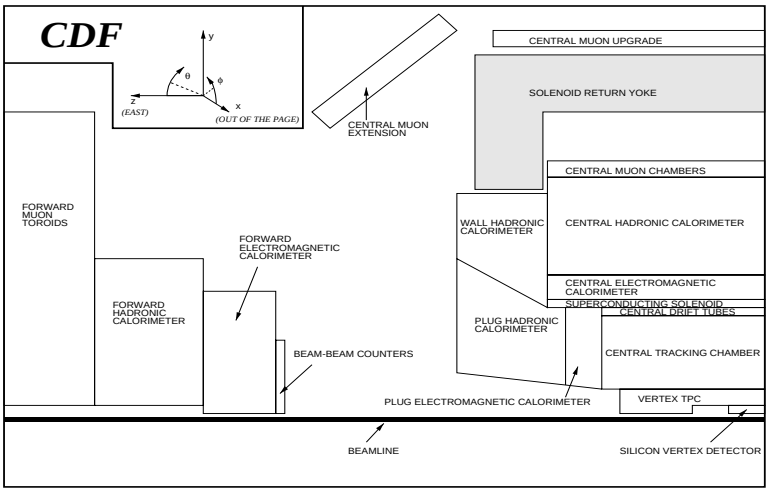


Figure 2.3: Schematic side view of a CDF quadrant.

2.2.1 Tracking Systems

In order of increasing distance from the beam pipe, the Silicon Vertex Detector (SVX), the Vertex Time Projection Chamber (VTX), and the Central Tracking Chamber (CTC) provide information on charged-particle trajectories at CDF. These devices are enclosed in a superconducting magnet which provides a 1.4 T axial field (parallel to the beam pipe), and each contains an argon/ethane gas mixture. In such a field, a charged particle tracks a helix with axis parallel to the field. The goal of the tracking systems is to allow measurement of the five parameters which completely describe the track helix [19]:

1. Polar angle θ at point of track's closest approach to origin
2. Track curvature ($\propto$ momentum^{-1}) in transverse plane
3. z-position of point of track's closest approach to origin
4. Impact parameter (distance between track and origin at point of closest approach)
5. Direction in transverse plane of track at point of closest approach

Each tracking system provides necessary information for different parameters, reflective of the design goals of each device.

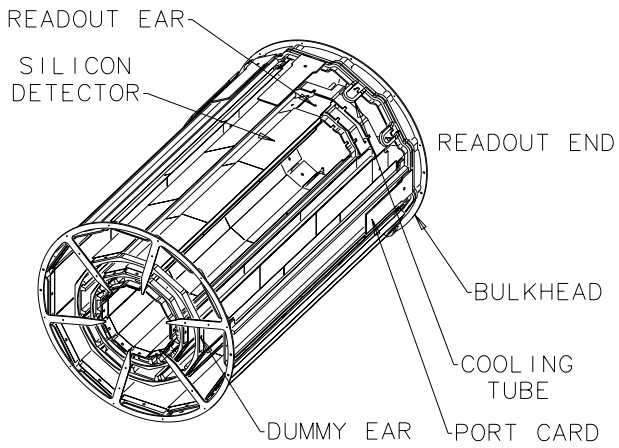


Figure 2.4: Schematic view of one barrel of the SVX.

2.2.1.1 Silicon Vertex Detector

The SVX is designed to track particles in the transverse plane very near the B0 interaction point [20]. It consists of two independent 25.5 cm barrels placed end-to-end with axes coincident with the beam axis. Each barrel contains four concentric layers (3.0-7.9 cm in radius) of silicon strip detectors. Figure 2.4 shows an isometric view of one SVX barrel. The SVX covers the pseudorapidity range $|\eta| < 1.9$ and accepts about 60% of $p\bar{p}$ interactions.

The strips are arranged along twelve faces in each radial layer, and are bonded to each other along the beam axis in groups of three, called ladders. As charged particles pass through the SVX, they create ionization (electron/hole pairs) which is collected onto the strips and read out by circuits at the end of each ladder. The three inner layers have strips with 60 μm pitch, and the outer-layer strips have 55 μm pitch. The SVX has a hit efficiency of $\sim$90% per layer, and a single hit resolution of 13 μm in the transverse plane. Combining SVX and CTC tracking information gives a resolution of the track impact parameter (i.p.) in the transverse plane of $\sim$ 17 μm for high momentum tracks. The SVX also provides the track direction at its point of closest approach to the origin, but no z information.

A long-lived particle produced in the primary collision can move several millimeters

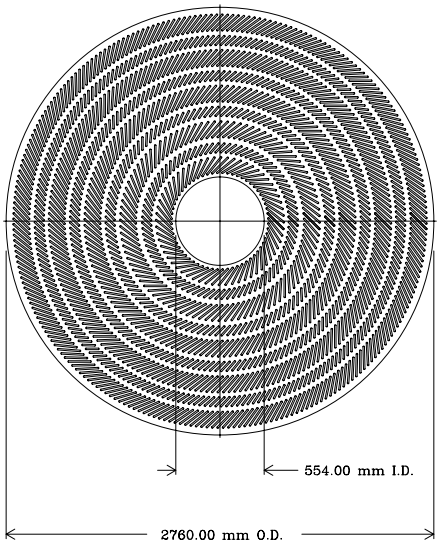


Figure 2.5: CTC endplate view. Each thin box contains a wire supercell. Short boxes in the stereo superlayers have six sense wires; long boxes in the axial superlayers have 12 wires.

before decaying into other particles. Tracks coming from such a decay can have i.p. $\sim 100\mu$m (significantly greater than the SVX resolution), allowing for their identification. The SVX is therefore a powerful tool to distinguish secondary vertices of long-lived particles from the dozens of tracks originating from the primary interaction vertex.

2.2.1.2 Vertex Time Projection Chamber

The VTX is primarily used to measure the z position of the primary interaction vertex. Located between the SVX and CTC, it consists of octagonal modules along the beam direction, each containing layers of sense wires oriented along ϕ. It is similar in design to the previous version described in [21]. It provides tracking information in the r-z plane in the pseudorapidity range $|\eta| < 3.5$, yielding a primary-interaction vertex z-position resolution of 1 mm.

2.2.1.3 Central Tracking Chamber

The CTC is a cylindrical drift chamber which gives precise transverse momentum and spatial resolution for charged particles in the pseudorapidity range $|\eta| < 1.0$ [22]. The

chamber contains a total of 36,504 sense and field-shaping wires arranged in alternating axial and stereo superlayers, as shown in Figure 2.5. The axial superlayers have wires parallel to the beam axis. The stereo superlayers have wires canted at $\pm 3^\circ$ with respect to the beam axis in order to obtain z tracking information. The wire positions in each supercell are tilted at 45° with respect to the radial direction. This tilt corrects for the Lorentz angle of the drift electrons in the crossed electric and magnetic fields, resulting in an approximately azimuthal drift trajectory.

The CTC has a transverse position resolution of approximately 200 μm per sense wire. Additionally, the CTC is sufficiently large to measure the curvature of charged particles as they bend in the magnetic field. From the track curvature the particle's transverse momentum P_T is determined with a resolution of $\frac{\sigma_{P_T}}{P_T} = 0.0011 \cdot P_T \cdot (\mathrm{GeV}/c)^{-1}$.

2.2.2 Calorimeters

CDF uses electromagnetic and hadronic calorimeters to sample the energy of particles and jets [23] [24]. Relevant to this analysis is the central region, in which outgoing particles first encounter the electromagnetic calorimeter (CEM) followed by the hadronic calorimeter (CHA). Both calorimeters have a cylindrical geometry divided into projective towers (with respect to the interaction point) of size 0.1 in η and 15° in ϕ. Each 15° calorimeter segment is called a wedge, and each calorimeter has 24 wedges covering the full azimuthal range. Figure 2.6 shows a CEM wedge.

The CEM covers the pseudorapidity range $|\eta| < 1.1$. The CEM uses lead sheets interspersed with scintillator for a total thickness of 18 radiation lengths and one pion absorption length. The CEM has an energy resolution $\frac{\sigma_E}{E} = \frac{13.5\%}{\sqrt{(E/\mathrm{GeV}\sin\theta)}} \oplus 2\%$.

Located at approximately the maximum electron shower position of six radiation lengths inside the CEM is a set of strip chambers (CES). The CES gives shower position and profile measurements is both z and r-ϕ coordinates, which is used to distinguish electrons from hadronic showers. The transverse shower position resolution σ_x is about 2 mm.

The CHA is located radially outward from the CEM, and covers the pseudorapidity

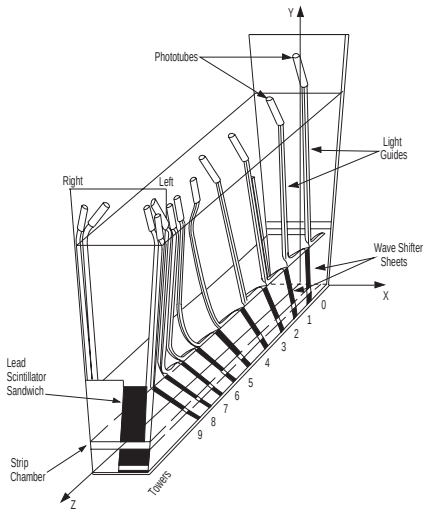


Figure 2.6: Diagram of a CEM wedge. Each wedge contains 10 towers, each spanning the η-ϕ range of $0.1 \times 15°$. The scintillator from each tower is read out at the ends by phototubes.

range $|\eta| < 0.9$. It contains alternating layers of iron absorber and scintillator for a total thickness of 4.7 absorption lengths. The CHA has energy resolution $\frac{\sigma_E}{E} = \frac{33.0\%}{\sqrt{(E/\mathrm{GeV}\sin\theta)}} \oplus 4\%$.

2.2.3 Muon Chambers

Three different drift-chamber systems identify minimum ionizing particles in the central region: central muon (CMU) [25], central muon upgrade (CMP), and central muon extension (CMX) [26]. Figure 2.7 shows the coverage provided by the muon detectors, which contains gaps in both η and ϕ. Since these chambers lie outside the magnetic field, particle tracks are straight, and their angle with respect to radial allows for a crude transverse momentum measurement.

The CMU has a cylindrical geometry divided into 24 wedges of 15° in ϕ, and provides pseudorapidity coverage $|\eta| < 0.6$. Each CMU wedge contains 48 drift chambers with axial sense wires in argon/ethane gas. Each chamber is divided into east/west halves at $\theta = 90°$. Figure 2.8 shows a transverse view of a CMU section with a candidate muon-track stub. Typical transverse position resolution is about 250 μm. Sense wires from adjacent cells in a layer are ganged in such a way to allow for charge-division readout, allowing for a crude z-position measurement.

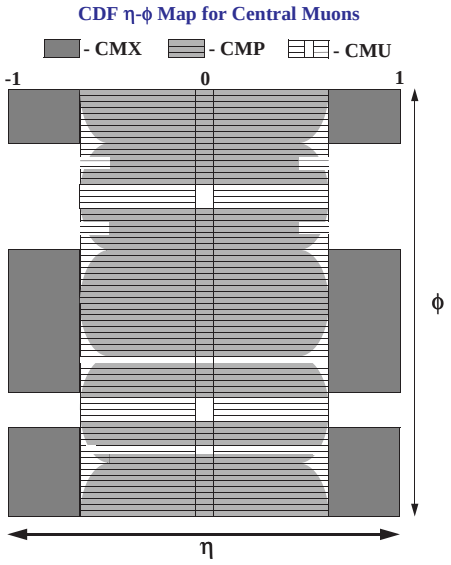


Figure 2.7: Central muon chamber coverage in the η-ϕ plane.

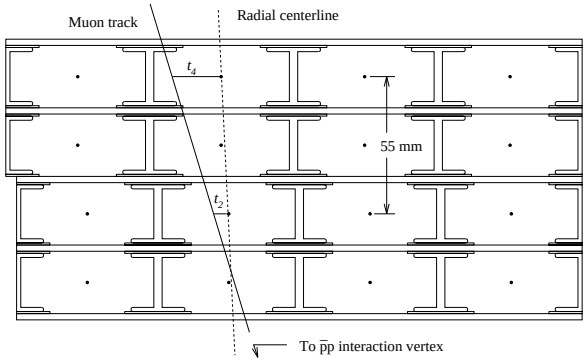


Figure 2.8: Transverse plane view of 5° section of CMU drift chambers. The passing muon creates ionization electrons which collect on the sense wire in each cell.

The CMP chambers are arranged in a rectangular geometry around the central detector. On the north and south sides of the detector, the chambers are mounted on steel walls; on the top and bottom, the chambers are mounted on the steel solenoid field return yoke. CMP cells have the same transverse size and axial alignment as the CMU, but adjacent layers are half-cell staggered. The chambers are not east/west divided, so no z information is possible. Typical transverse position resolution is about 300 μm. The steel absorbing layers inside the CMP add an additional three pion interaction lengths, greatly reducing the fake rate from hadronic calorimeter punch-through compared to the CMU. For this reason, the CMP is mainly used to confirm CMU muon-track stubs in the 53% of solid angle over which the two systems overlap.

The CMX chambers extend the coverage of the system to $|\eta| < 1.0$. The chambers lie on free-standing conical arches outside the central barrel region. Muon track stubs can be reconstructed as in the CMU and CMP, with a transverse position resolution of 250 μm. Scintillators on the front and back faces of the CMP and CMX chambers provide trigger information.

2.2.4 Beam-beam Counters

The vast majority of $p\bar{p}$ events are so-called minimum bias events in which a soft scattering produces a spray of particles close to the beam axis. The minimum bias event rate is monitored with scintillation planes called beam-beam counters (BBC) which are located on both the far forward east and west sides of the interaction point. The counters have high timing resolution (< 200 ps), and cover the pseudorapidity range $3.2 < \eta < 5.9$. A coincidence (within 15 ns) of hits from both BBCs signals a $p\bar{p}$ event. The instantaneous luminosity at CDF is calculated from this rate:

$$\mathcal{L} = \frac{R_{\mathrm{BBC}}}{\sigma_{\mathrm{BBC}}} \quad ,$$

where R_{BBC} is the BBC event rate, and $\sigma_{\mathrm{BBC}} = 51 \pm 2$ mb is the effective BBC cross section [27].

2.2.5 Data Acquisition System

CDF contains over 100,000 electronic channels of phototube, wire and strip outputs. Calorimeter phototube pulses are sampled by analog circuits on RABBIT cards [28]. The signals are then digitized and sent to MX processors, which are interfaced to the Fastbus data acquisition system (DAQ) [29] . Drift chamber signals are shaped at the detector and sent to commercial Fastbus TDCs, which are then read out by Fastbus scanners called SSPs. The silicon strip signals are digitized at the SVX and then read out by SSPs. The DAQ can read out the complete detector in ~3 ms, producing ~200 kB per event. Events which pass the multistage trigger (next section) are recorded on permanent media for later analysis.

2.2.6 Trigger

A hadron collider provides an operating environment with simultaneous features and challenges. The Tevatron's $p\bar{p}$ collisions provide a rich spectrum of interesting physics processes buried under an enormous background rate. The total $p\bar{p}$ cross section at $\sqrt{s} = 1.8$ TeV is on the order of 80 mb, while the $t\bar{t}$ production cross section is on the order of several pb. CDF uses a three-level trigger system to provide the necessary rejection while selecting events of interest [30]. The Tevatron provides $p\bar{p}$ collisions every 3.5 μs, corresponding to a rate of about 280 kHz, while the DAQ records events at only several Hz. It is the task of the trigger to choose one out of every 50–100 thousand events to save.

2.2.6.1 Level 1

The Level 1 trigger decision is made by hardware in 2 μs, incurring no dead time. An event is passed if it contains calorimeter energy above a threshold (using fast outputs from calorimeter towers of $\Delta\eta \times \Delta\phi = 0.2 \times 15^\circ$), a transverse energy imbalance in the calorimeters, or a pair of hits in the muon chambers. Level 1 selects and passes events to Level 2 at a rate of a few kHz.

2.2.6.2 Level 2

More information from the calorimeter, tracking, and muon systems is made available to the Level 2 trigger for a more sophisticated decision. A hardware calorimeter-cluster finder forms clusters of towers with significant energy. It first searches for seed towers above a threshold and then adds adjacent towers over a lower threshold. The Central Fast Tracker (CFT) uses CTC hits to reconstruct high-momentum tracks in r-ϕ with a resolution of $\frac{\sigma_{P_T}}{P_T} \simeq 0.035 P_T\,(\mathrm{GeV})^{-1}$. Muon-track-stub position and transverse momentum are also available. CFT tracks can be matched to CEM clusters to form an electron candidate, or to muon stubs to form a muon candidate. The final Level 2 trigger decision is made by a programmable processor in about 20 μs. During this decision time the trigger ignores further collisions, typically incurring about 4% deadtime. Level 2 accepts and passes events to Level 3 at an approximate rate of 20 Hz.

2.2.6.3 Level 3

Upon a Level 2 accept, the entire detector is read out and sent to the Level 3 trigger. Level 3 is a software trigger which runs on a farm of 48 processors, plus associated hardware to transfer the event data into the trigger [31]. The event is filtered for physics objects such as electrons, muons, jets, etc., using a simplified version of the offline event reconstruction code. After the detector is read out and before Level 3 has made a decision, the system may continue to examine subsequent beam crossings for Level 1 and Level 2 triggers. Level 3 passes events at a typical rate of 5 Hz. The detector readout and Level 3 decision times incur about 10% deadtime.

Chapter 3

Jet Flavor Identification Using a Neural Network

Our goal was to construct an efficient and accurate jet-by-jet identification of b,c and p (prompt) jets, starting with tracking information, and constructing discriminant variables which exploit their expected differences in mass, lifetime, and fragmentation. The most powerful flavor discrimination handles come from the daughters of a decaying hadron in the jet. Previous b-tagging algorithms basically work by using the charged daughters to form a single variable related to decay distance or impact parameters. We will generalize this technique, using the decay tracks to form a number of flavor-related variables, and then combine them with a neural net to optimize the discrimination. An analysis of this kind has been used to identify charm in hadronic W decays at LEP [32].

This chapter is organized as follows. We first discuss the observable differences between b,c, and p jets based on underlying physical properties. Next we describe the Monte Carlo sample which we use to study the separate characteristics of the jets. Then we outline how we select decay tracks and use them to form a set of jet-flavor discrimination variables. Finally, we describe the construction and performance of a jet-flavor tagger, based on a neural network, from these variables.

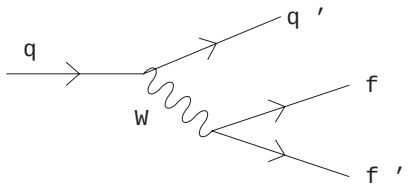


Figure 3.1: Quark q decaying to q' via W, which then decays to fermions.

3.1 Expected Differences Between Jets

3.1.1 Fragmentation

When a hard scattering transfers energy to a quark, it moves away from other quark(s) to which it is bound. QCD confinement forbids free colored particles, so the energy in the color field grows with the separation distance. Eventually it becomes energetically favorable to create a light quark pair. The fragmenting quark is most likely to form a meson with a light quark when the quarks have similar velocities. A heavy quark loses a small fraction of its energy in materializing enough light quark pairs with comparable velocity [4]. The resulting meson will therefore contain a large fraction of the original quark energy. Then the fragmentation function, $z = E_{hadron}/E_{quark}$, should have a harder distribution (peaked at a greater value) for heavy quarks than for light quarks.

3.1.2 Mass and Lifetime

Differences in bare quark masses and electroweak coupling strengths lead to different lifetimes among the quark flavors. The lifetime of a quark which undergoes the three-body electroweak decay shown in Figure 3.1 is proportional to [3]: [1]

$$\tau_q \;\sim\; \left(V_{qq\prime}^2\, M_q^5\right)^{-1} \quad ,$$

where $V_{qq\prime}$ are CKM matrix elements and M_q is the bare quark mass. The lifetimes of the quark flavors should then be in the ratio

$$b : c : \{s, u, d\} \;\sim\; 1 : 0.7 : 10^5 \quad .$$

[1] For quarks with $M_q < M_W + M_{q'}$ (decay via virtual, not real, W), i.e. not top quarks.

Meson	Mass (GeV)	$c\tau$ (μm)
B^0	5.27	468
$B^{\pm}$	5.28	495
D^0	1.86	124
$D^{\pm}$	1.87	317

Table 3.1: Meson mass and decay distance $c\tau$ [5].

The decaying quark in a hadron is strongly bound, but in the spectator approximation, this decay occurs independently of the other partons in the hadron. This implies that all hadrons of a given flavor should have the same lifetime. This approximation is more valid for heavier quarks whose decay releases more energy compared to the typical hadronic binding energy [4]. Table 3.1 lists masses and decay distances of some common heavy mesons.

3.1.3 Electric Charge

Although $b, \bar{b}, c, \bar{c}$ quarks have different electric charge, the utility of this quantum number is lost after fragmentation. Hadrons of a given flavor, having integer charge, are equally likely to have either positive or negative charge. A variable such as jet charge works only for distinguishing hadrons built from quarks from those built from antiquarks of the same flavor; the correlation between original quark charge and jet charge is preserved in this case. For the generalized flavor tagger (but not q vs. $\bar{q}$), electric charge is not useful on a per-jet basis.

3.2 Monte Carlo

To assess the feasibility of a potential jet-flavor tagger, we need a pure sample of each flavor jet (b,c,p). Since no such samples can be made from actual data without introducing significant bias, we use a Monte Carlo (MC) simulation to generate them. More detailed descriptions of the event simulation and reconstruction steps are given in § 5.2.2 and § 5.3.1.2.

3.2.1 Generator

PYTHIA 5.6 generated 20,000 $b\bar{b}$, 40,000 $c\bar{c}$, and 77,554 QCD 2→2 events with minimum $\hat{p}_T$ of 20 GeV/c. Bottom and charm hadrons are redecayed using updated CLEO tables.

3.2.2 Detector Simulation

CDFSIM is used to simulate response in the tracking chambers, including the SVX′. QFL then simulates the calorimeter and muon chambers.

3.2.3 Reconstruction

The primary-interaction vertex position is fit in each event using the beam position as a seed. Track finding efficiency is degraded according to the CTC environment around the track to mimic the performance in real data [33]. The following jet and tracking cuts are then made:

- Calorimeter clustering η-ϕ cone size 0.4
- Uncorrected jet $E_T \geq$15 GeV
- Track-jet association η-ϕ cone size 0.4
- Jet has at least one SVX-fiducial track

After these cuts, 10,895 b, 25,011 c, and 39,756 p jets remain.

3.3 Track and Jet Selection

3.3.1 Decay Track Identification

3.3.1.1 SVX Track Cuts

All tracks in a jet fall into three categories: decay tracks from a heavy hadron, fragmentation tracks, and tracks coming from other interactions in the event. Our hypothesis is that decay tracks will contain most of the heavy flavor information (mass, lifetime, and fragmentation). Therefore the first step in jet flavor tagging is the identification of the decay tracks.

We begin by applying *good_svx* track cuts:

- $P_T \geq 1.0$ GeV/c
- Impact parameter (i.p.) $|d| < 0.1$ cm
- $|z| < 5.0$ cm [2]
- ≥ 2 unshared SVX hits
- Displaced K_{short}, prompt Λ veto

After these cuts, the track i.p. d and associated error σ_d are used to select decay tracks. In Appendix B we show that

$$< d > \sim \tau_{hadron} ,$$

where the average is over all tracks from the decay of a hadron. The i.p. is then signed according to the position of the point of closest approach with respect to the jet axis and primary vertex [34], as shown in Figure 3.2.

To normalize to our resolution, we define the track significance $s = d/\sigma_d$. Tracks from prompt jets are equally likely to have positive or negative small s. Tracks from secondary

[2] z is the distance (in the z-direction) between the track's point of closest approach to the track helix origin and the primary vertex.

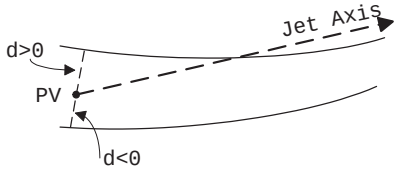


Figure 3.2: Signed track impact parameter with respect to the primary interaction vertex (PV) in the event.

vertices can have large positive s, while tracks from tertiary vertices can have large positive or negative s.

3.3.1.2 Track Probability

In order to understand the discriminating power of track s, one must first characterize the resolution of the SVX using a prompt sample. An ideal prompt track sample in an ideal detector would have identically zero i.p. In an imperfect detector, the i.p. distribution of a prompt track sample will be distributed symmetrically about zero. If the errors are Gaussian and calculated correctly, then the corresponding track s distribution will be Gaussian of unit width. Finally, the presence of heavy flavor in the sample will produce (mostly positive) tails in the i.p. resolution.

For this MC-based study it is possible to generate a pure prompt sample, but to get an idea of what to expect from the data, we allow a mix of heavy flavor as predicted by PYTHIA. To make the resolution function $R(s)$ (density function of s) shown in Figure 3.3, we used *good_svx* tracks in jets with $E_T > 10$ GeV from a MC sample of inclusive jets which is described in § 4.2.4. The central Gaussian width of 0.82 shows that the i.p. errors are approximately Gaussian and somewhat understood. Since heavy-flavor tracks contaminate mostly the positive side of this distribution, the negative side more accurately characterizes the i.p. resolution.

For a track with i.p. significance s, the track probability is calculated by integrating

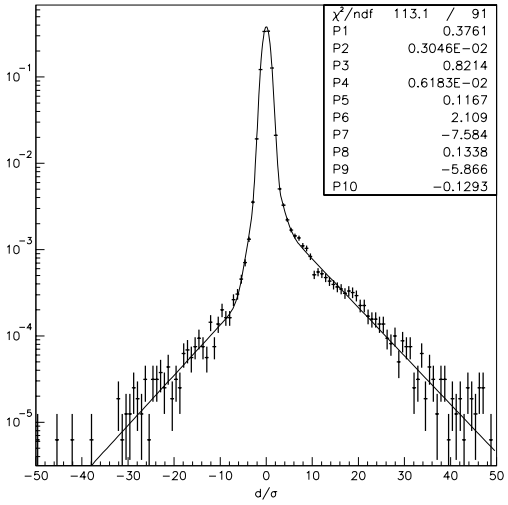


Figure 3.3: MC track i.p. resolution function fitted to two Gaussians plus two exponentials.

the negative side of the resolution function to $-|s|$: [3]

$$P_{\text{track}}(s, N_{\text{hits}}, N_{\text{shared}}, P_T) = \int_{-\infty}^{-|s|} R(s', N_{\text{hits}}, N_{\text{shared}}, P_T)\, ds'$$

Track probability is the probability that a track with a given s is consistent with resolution effects alone. For the resolution function sample of tracks, the distribution of track probabilities is flat by construction. Tracks coming from the decay of a long-lived hadron will tend to be improbable, although the exact shape of their track probability distribution is impossible to calculate analytically[4].

By dividing the i.p. by its error, we intend to remove most functional dependencies (such as multiple scattering, track fit quality, etc.) from the significance s. This method is not entirely successful, as we still observe different resolution shapes depending on the P_T and number of SVX hits associated with the track. To remove these remaining dependencies, we parameterize the resolution function by the number of SVX hits N_{hits} and shared hits N_{shared} on the track, and the track P_T ($1 \leq P_T < 2, 2 \leq P_T < 5, 5 \leq P_T$). We also studied

[3]Note that when the track probability is calculated for tracks in the data, a separate resolution function, made from an inclusive jets data sample described in § 4.2.4, is used.

[4]One can convolute a falling exponential distribution with a Gaussian resolution to yield an expected distribution of track d. But the input to track probability is d/σ_d, for which there is no analytic model.

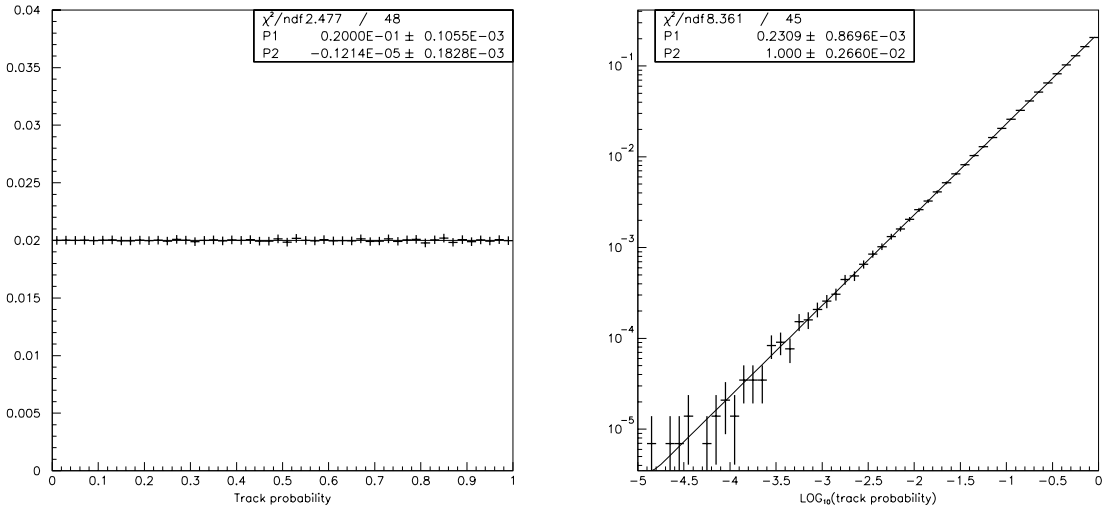


Figure 3.4: Track probability for tracks used to make the resolution function, shown in lin-lin and log-log plots with line fits. It can be shown that a flat distribution on a lin-lin plot is a line of unit slope on a log-log plot.

possible correlations between the i.p. and track z, χ^2, and primary vertex quality, but none noteworthy were found. There is a well-known bias in track i.p. which depends on whether or not the track was used in the primary vertex fit. Instead of trying to correct this bias, we reasonably assume that its effects are similar for data and MC.

The track probability for a given track is computed by numerically integrating the resolution function (for $0 < |s| \leq 50$, we use 10,000 bins with linear interpolation between bins). For (highly improbable) tracks with $|s| > 50$, track probability is calculated by integrating an exponential fit to the tail of the resolution function. This numerical technique does not rely on the imperfect fit to the resolution function in Figure 3.3, which actually has a more complicated shape than a sum of Gaussians and exponentials. Figure 3.4 shows the necessarily flat track probability distribution for track used to make the resolution function.

3.3.1.3 Decay Track Definition

We require a decay track to pass *good_svx* quality cuts and have positive and significant i.p. ($d > 0$, $P_{\rm track} < P_{\rm cut}$). Requiring positive i.p. tracks immediately rejects half of

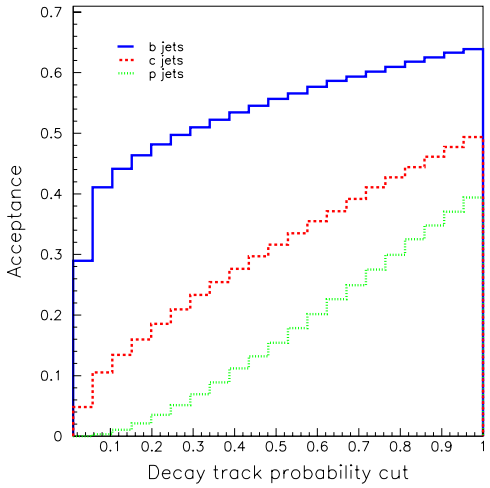


Figure 3.5: Taggable jet acceptances for b,c,p jets versus decay track probability cut.

the prompt tracks. The subsequent track probability cut is, by definition, $P_{\rm cut}$ efficient for the remaining prompt tracks. Tracks from b and c hadron decays have higher efficiencies (discussed in next subsection) for a given value of $P_{\rm cut}$. At best, we will still miss about one-third of the decay daughters to neutrals.

3.3.2 Taggable Jet Acceptances

Most of the jet discrimination variables we use are well defined for jets with at least two decay tracks. We therefore define a taggable jet as having $E_T > 15$ GeV and to contain ≥ 2 decay tracks. The value of $P_{\rm cut}$ used to select decay tracks is a matter of choice. The goal is to select true decay tracks while rejecting fragmentation or prompt tracks. For a given jet flavor, we define the taggable jet acceptance as

$$A \;=\; \frac{\#\text{ jets with } E_T \geq 15\text{ GeV},\ \geq 2\text{ decay tracks}}{\#\text{ jets with } E_T \geq 15\text{ GeV},\ \geq 1\text{ SVX track}}\ . \tag{3.1}$$

Figure 3.5 shows the acceptances for b, c, p jets as a function of $P_{\rm cut}$. The acceptance does not rise to one because *two* decay tracks are required. For this study, we chose $P_{\rm cut} = 0.3$, which biases against prompt jets and selects decay tracks with reasonable efficiency and purity, and gives the taggable jet acceptances $A_b = 0.51$, $A_c = 0.23$, $A_p = 0.07$.

3.3.3 Jet Variables

We now describe 14 different jet variables, most of which are constructed using decay tracks. In the subsequent plots, all distributions have unit normalization.

1. Hadron mass approximated by the reconstructed mass of decay tracks:

$$M_{\rm trks} = \sqrt{\left(\sum_{\rm decay\ trks} E_{\rm trk}\right)^2 - \left(\sum_{\rm decay\ trks} \vec{p}_{\rm trk}\right)^2} \quad , \text{ with } M_{\rm trk} = M_\pi$$

2. Calorimeter cluster mass: [5]

$$M_{\rm clus} = \sqrt{E_{\rm clus}^2 - \vec{P}_{\rm clus}^2} \quad , \text{ where}$$

$$E_{\rm clus} = \sum_{\substack{\text{towers in}\\ \text{cluster}}} E_{\rm tower} \quad , \quad \vec{P}_{\rm clus} = \sum_{\substack{\text{towers in}\\ \text{cluster}}} \vec{P}_{\rm tower} = \sum_{\substack{\text{towers in}\\ \text{cluster}}} E_{\rm tower}\hat{P}_{\rm tower}$$

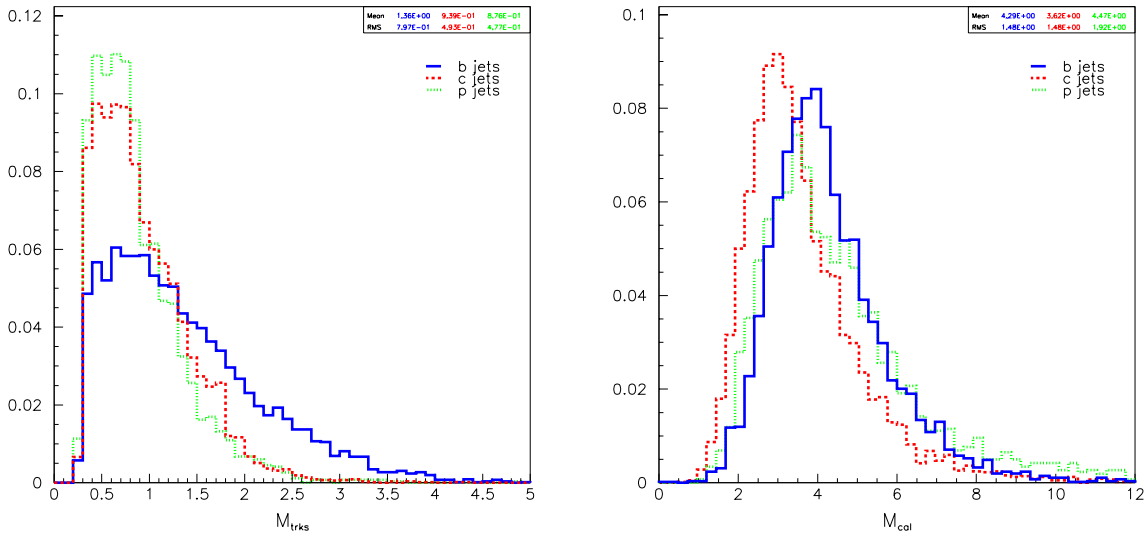


Figure 3.6: Reconstructed mass of decay tracks $M_{\rm trks}$ and calorimeter cluster mass $M_{\rm cal}$.

[5] Cluster refers to an offline reconstructed EM+HAD jet, not a trigger cluster.

3. Fragmentation function

$$Z \equiv \frac{E_{\text{hadron}}}{E_{\text{quark}}} \simeq \frac{\sum_{\text{decay trks}} E_{\text{trk}}}{E_{\text{jet}}} \quad , \text{ with } M_{\text{trk}} = M_{\pi}$$

B hadrons have the hardest fragmentation, followed by c, but the discrimination power is far reduced from that for reconstructed exclusive decays.

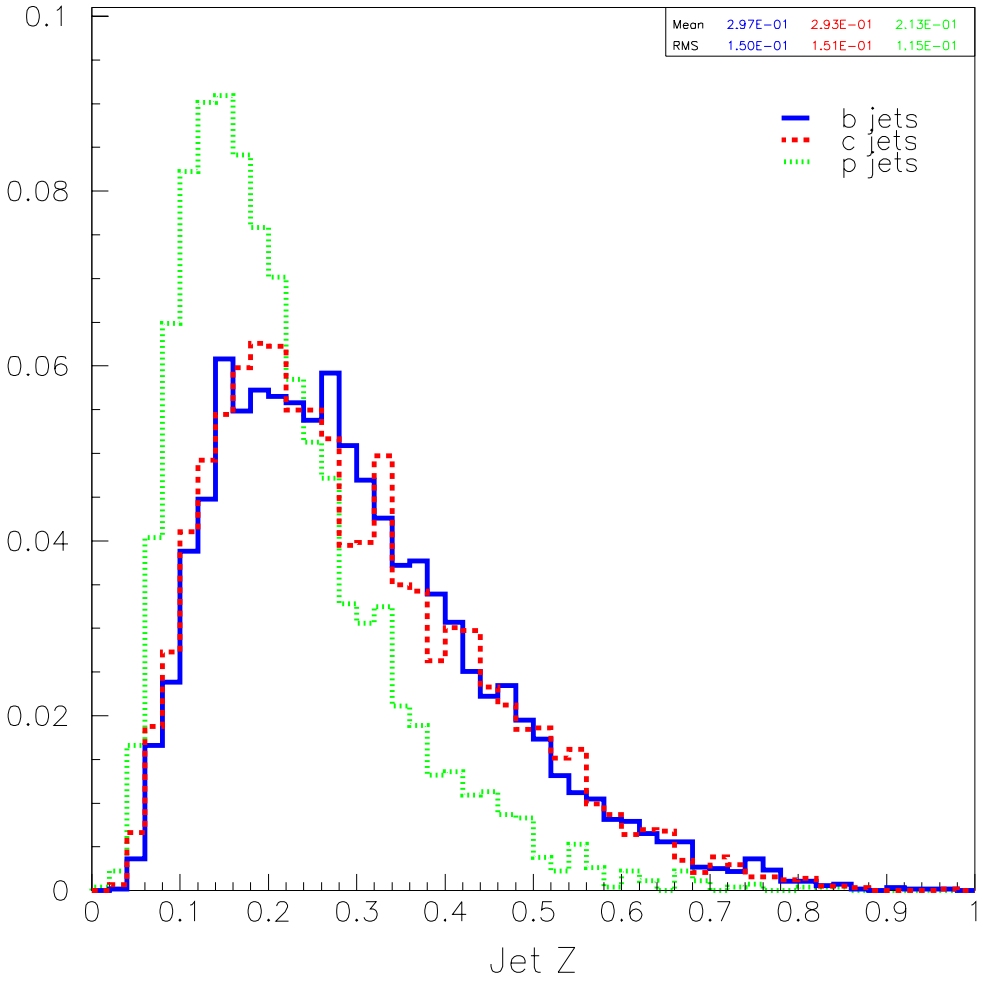


Figure 3.7: Fragmentation function Z.

Track multiplicities:

4. N_{decay} = number of positive i.p. tracks passing track probability cut.

   b jets contain more decay tracks than c jets, followed by p jets (≥ 2 are required for taggable jets).

5. N_{frag} = number of *good_svx* tracks in jet $- N_{\text{decay}}$.

   Prompt jets contain more fragmentation tracks, i.e., non-decay tracks.

6. $N^{\text{positive}}_{\text{good\_svx}} - N^{\text{negative}}_{\text{good\_svx}}$.

   Prompt jets are equally likely to contain positive or negative tracks, while decaying b and c mesons generate more positive tracks. The difference distributions below all have positive means because taggable jets require at least *two positive* decay tracks.

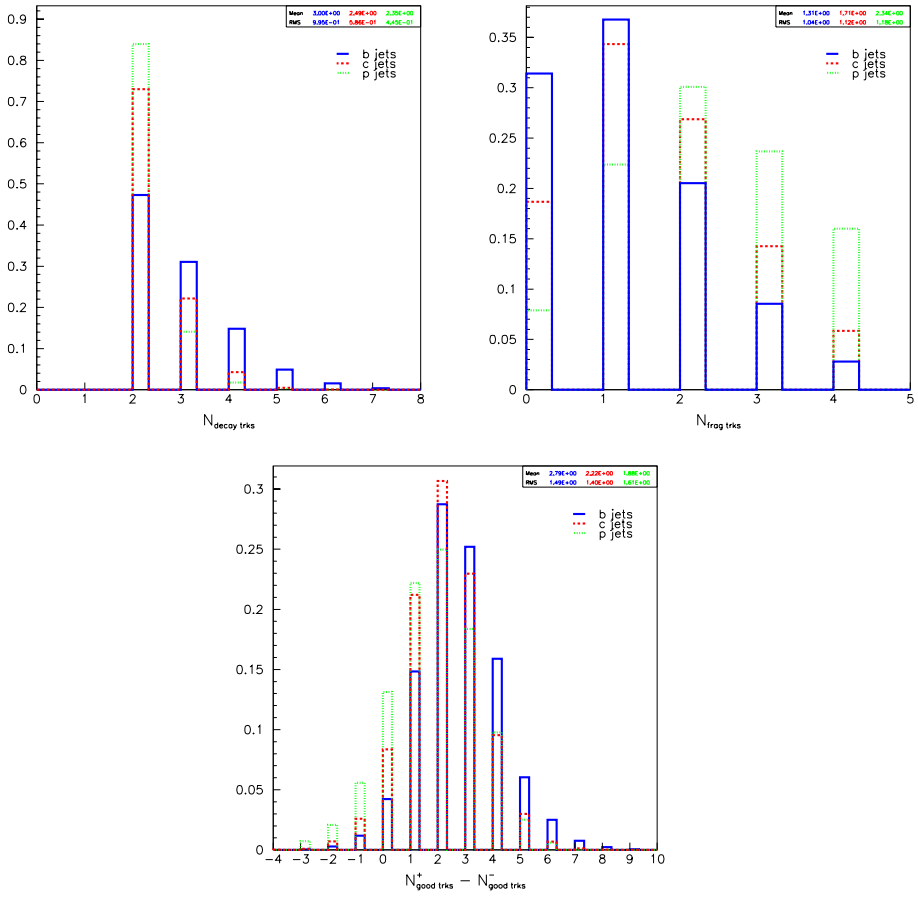


Figure 3.8: Track multiplicities.

7. Jet probability, $P_{\rm jet}$, is the probability that the set of tracks in the jet is consistent with coming from the primary vertex[6]. It is formed from the track probabilities of all positive i.p. *good_svx* tracks in the jet (not just decay tracks). The $P_{\rm jet}$ distribution is flat by construction for the resolution function jets; all $P_{\rm jet}$ shapes below are unpopulated near unit probability because of the taggable criteria.

$$P_{\rm jet} \;=\; \Pi \sum_{k=0}^{N-1} \frac{(-\ln \Pi)^k}{k!} \quad , \quad \Pi \;=\; P_1 P_2 \cdots P_N \quad , \quad N = \#\ good\_svx \text{ trks}$$

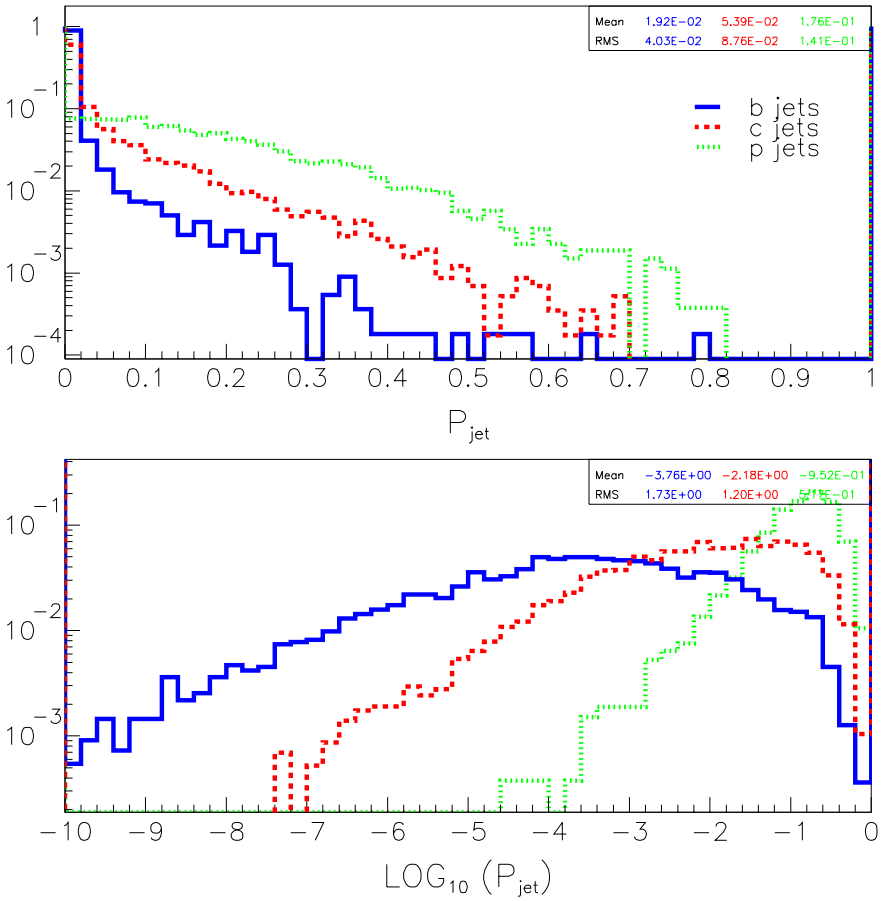


Figure 3.9: Jet probability $P_{\rm jet}$ shown on lin-log (top) and log-log (bottom) scales.

[6]The precise definition of $P_{\rm jet}$ is the probability that a prompt (no-life) jet would have tracks with the observed set of i.p. (due to resolution effects alone), or any combination less probable [34]. For a given track probability product Π, $P_{\rm jet}$ decreases with the number of tracks. This means that additional probable tracks ($P_{\rm track} \sim 1$) in the jet serve to *increase* $P_{\rm jet}$, a situation more likely in prompt jets.

8. Decay distance significance, $L_{xy}/\sigma_{L_{xy}}$, of secondary vertex. The least probable *good_svx* (positive or negative i.p.) tracks (three at most) are vertexed[7]. The two-dimensional decay distance is L_{xy} with error $\sigma_{L_{xy}}$.

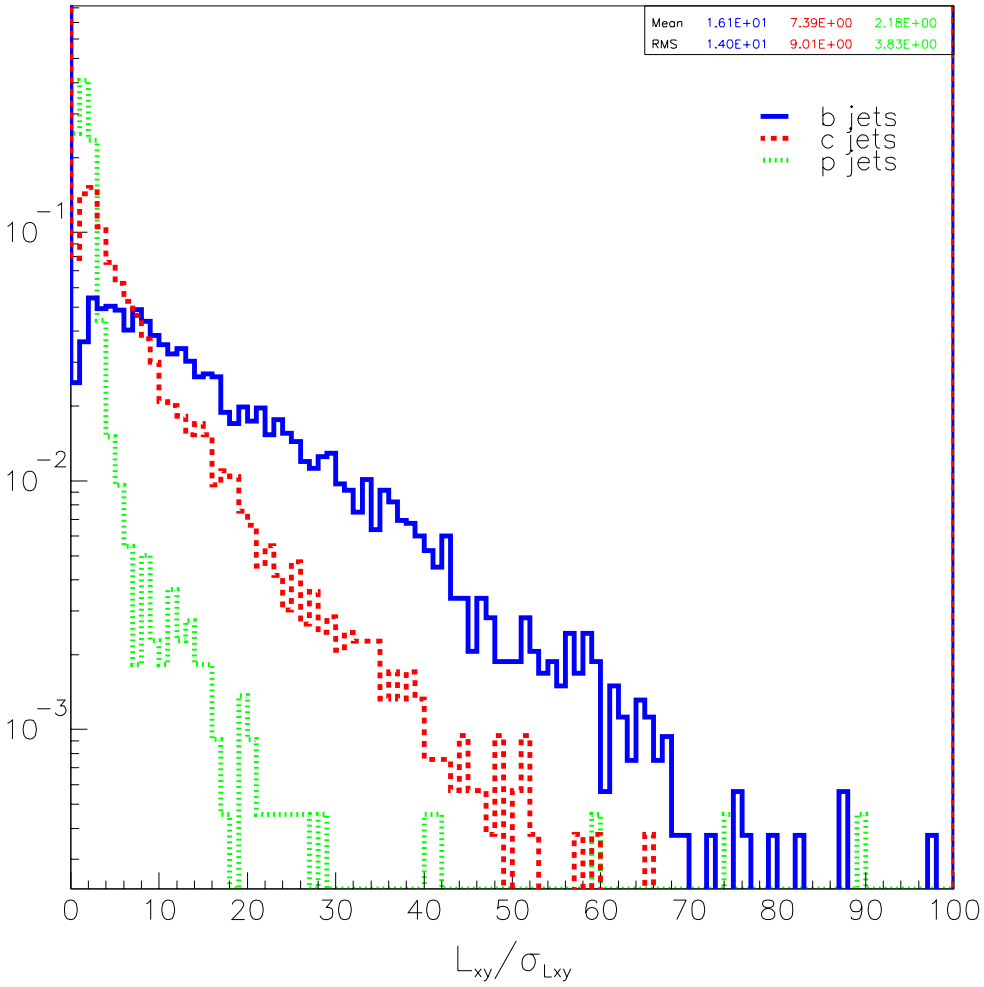


Figure 3.10: Decay distance significance $L_{xy}/\sigma_{L_{xy}}$ of vertexed tracks.

[7]By the CDF offline routine CTVMFT

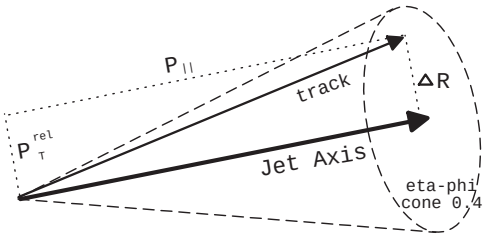


Figure 3.11: Track momentum components: $P_T^{\rm rel}$ is perpendicular to, $P_{||}$ is along the jet axis.

9. $\sum_{\text{decay trks}} |\,\Delta R\,| \;=\; \sum_{\text{decay trks}} \sqrt{[\Delta\eta(\text{track-jet})]^2 + [\Delta\phi(\text{track-jet})]^2}$

10. $\sum_{\text{decay trks}} |\,P_T^{\rm rel}\,|$,

where the track quantities ΔR and $P_T^{\rm rel}$ are defined in Figure 8. These jet-track topology variables exploit two features that grow with the mass of the decaying hadron:

(a) Track multiplicity

(b) Decay products diverge from the jet axis (more massive hadron is less boosted and releases more energy in decay)

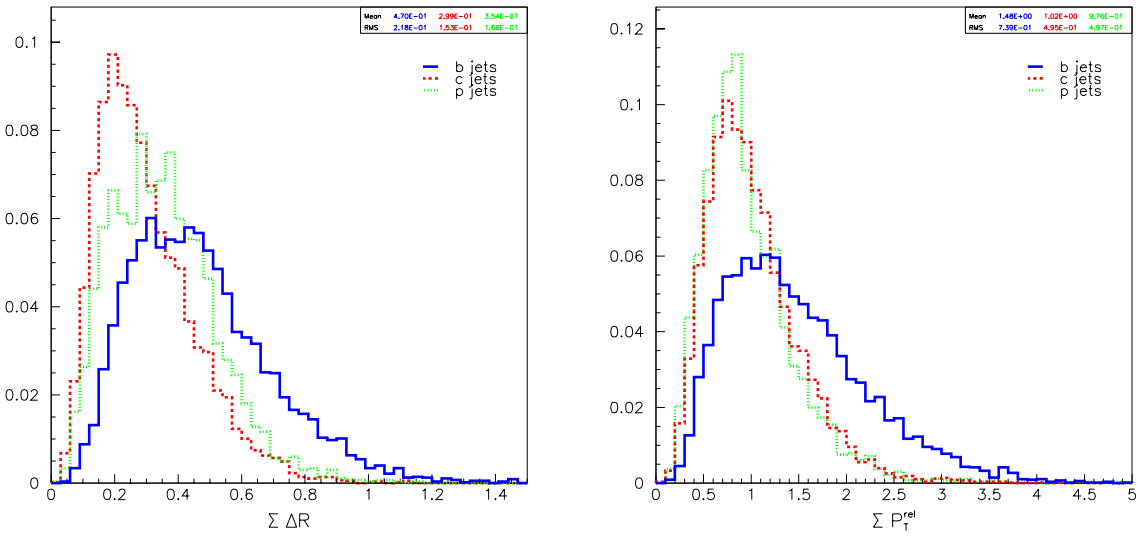


Figure 3.12: Jet topology variables $\Sigma\,\Delta R$ and $\Sigma\,P_T^{\rm rel}$.

11. Jet rapidity: $Y_{\text{jet}} = \frac{1}{2} \ln \left(\frac{E^{\text{sum}} + p_{\parallel}^{\text{sum}}}{E^{\text{sum}} - p_{\parallel}^{\text{sum}}} \right)$

$$E^{\text{sum}} = \sum_{\text{decay trks}} E_{\text{trk}} \quad , \quad p_{\parallel}^{\text{sum}} = \sum_{\text{decay trks}} p_{\parallel}$$

Analogous to event rapidity, jet rapidity grows as energetic tracks are collimated along the jet axis. Tracks in c jets are more forward (more boost, less energetic decay), so c jets have higher jet rapidity than b jets, which have more energetic tracks spread out.

12. Track energy fraction in half-cone: $\Psi = \dfrac{\sum_{trks, \Delta R \leq 0.2} E_{trk}}{\sum_{trks, \Delta R \leq 0.4} E_{trk}}$

Similar to the previous jet topology variables, Ψ helps discriminate charm jets because they concentrate a greater proportion of their decay products close to the jet axis.

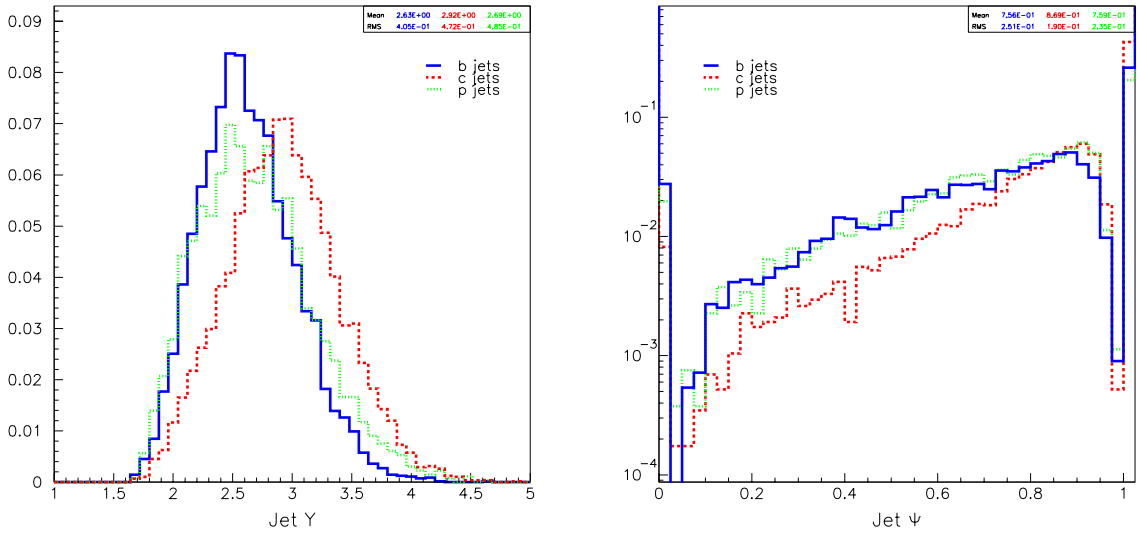


Figure 3.13: Track energy distribution variables Y_{jet} and Ψ.

13. The directed sphericity S of the jet attempts to exploit the more isotropic meson decay in b jets compared to more boosted c jets and the longitudinal fragmentation of hadrons in prompt jets[8]:

$$S = \frac{\sum_{\text{decay trks}} \left(P_T^{\text{rel}}\right)^2}{\sum_{\text{decay trks}} \left(P_{\text{boost}}\right)^2} \quad ,$$

where P_T^{rel} is defined in Figure 8, and P_{boost} is the track momentum boosted along the jet axis[9].

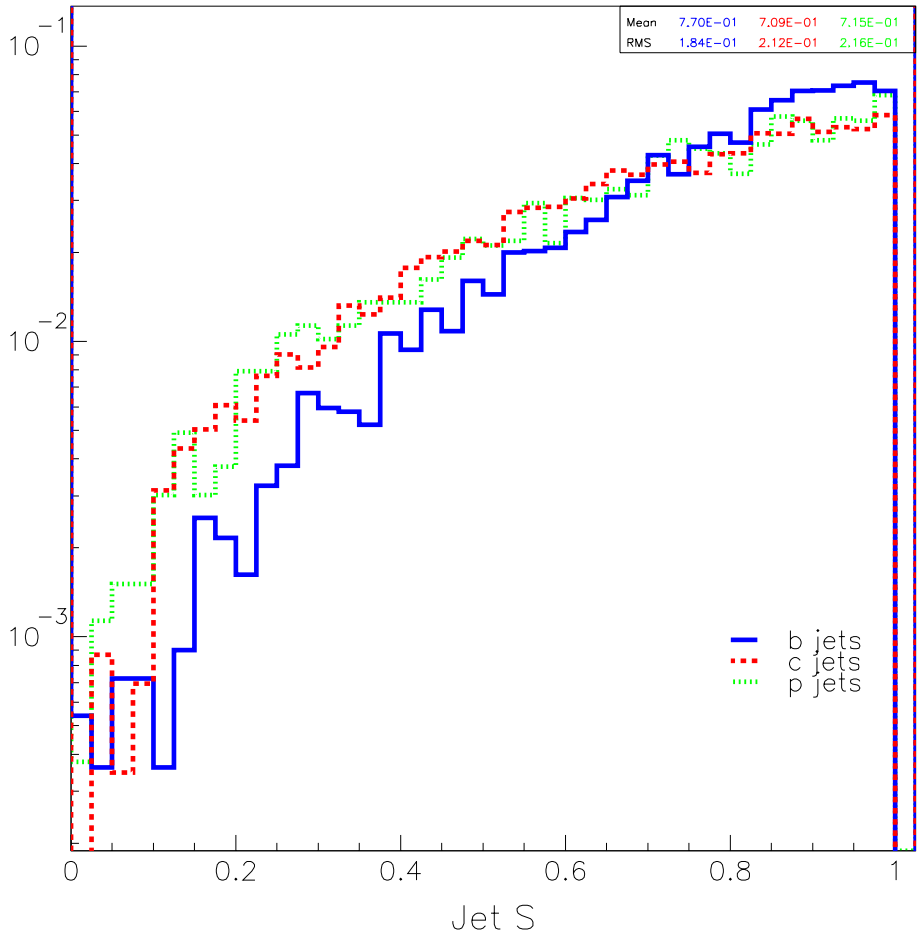


Figure 3.14: Directed sphericity S of the jet.

[8]For a spherically symmetric distribution of decay tracks, $S = \frac{2}{3}$.

[9]To calculate the boost factor $\gamma = E/M$, we use the corrected jet energy and the cluster mass.

14. Corrected jet E_T. The raw jet energy is corrected for various effects[10]. If the jet contains a muon, its P_T is added. This quantity does not provide any jet flavor discrimination by itself, although we expect it to be correlated with other jet variables[11].

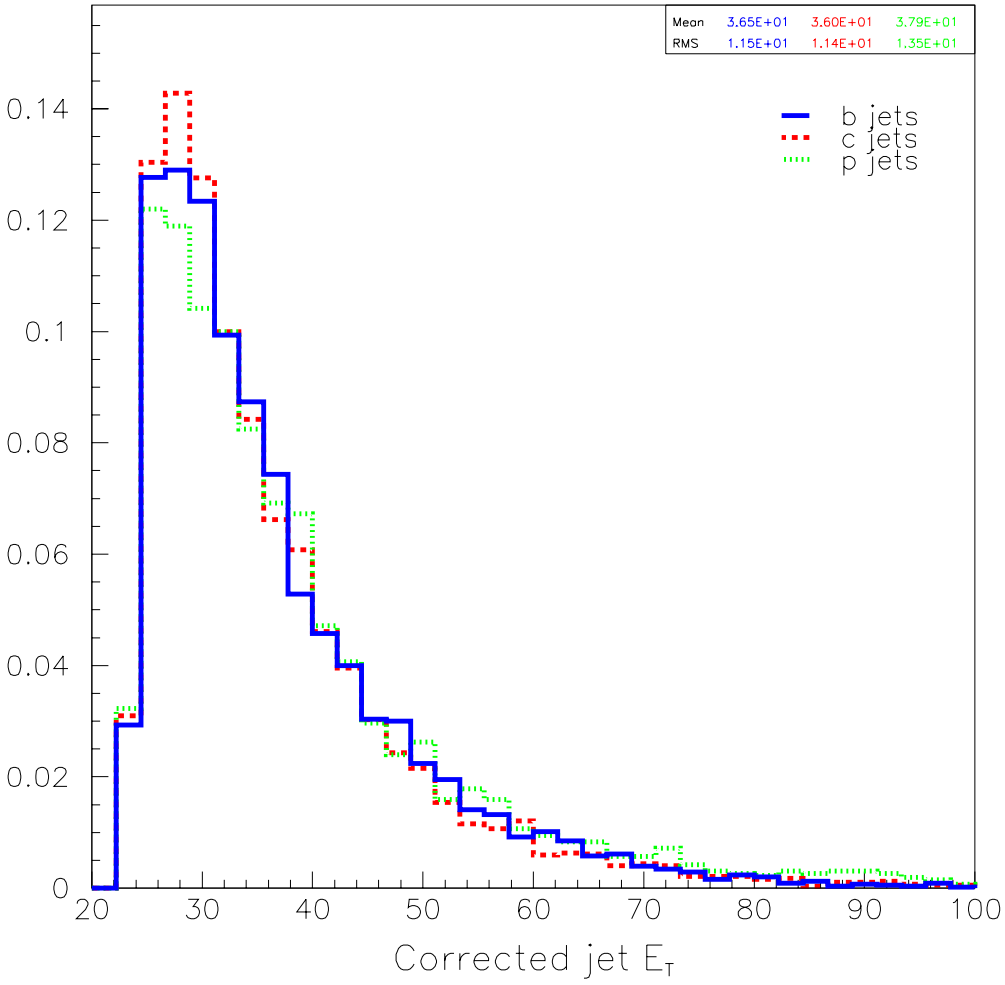


Figure 3.15: Corrected jet E_T.

[10]The jet energy is intended to measure the energy of the quark originating the jet, but several factors can decrease its accuracy. Gluon radiation out of the jet will incorrectly lower the measured jet energy, while particles entering the jet from elsewhere in the event will increase it. Out-of-cone and underlying-event jet energy corrections are applied to reduce these effects. An additional correction, parameterized by jet η and ϕ, is made to account for non-uniform calorimeter response.

[11]Including jet E_T helps the jet flavor discrimination problem to the extent to which the correlations between jet E_T and the other jet variables differ depending on the jet flavor.

3.4 Neural Network

3.4.1 Multivariate Classification Techniques

Jet flavor tagging is an example of a classification problem. Because no single variable provides sufficient discrimination, a multivariate approach is necessary. Generally, objects to be classified are characterized by an N-dimensional vector vector $\vec{x}$ in pattern space, and each is assigned to a specific class. In this application, $N = 14$ and there are three classes: b,c, and p. Because the characteristic distributions of different classes overlap, it is in principle no longer possible to classify with 100% efficiency[12][35].

There are many multivariate techniques which could be used for this problem. We have chosen to combine the information from the jet variables using a neural network (NN). NN's are an extension of conventional methods with the following additional features [36]:

- NN's make no assumption about underlying probability distributions.
- NN's exploit complicated correlations between many variables.
- NN's learn to classify from examples (useful when no classification algorithm is known).
- NN's do not rely on projections of the data onto particular axes for optimal classification (principle component analysis unnecessary).
- NN's are more efficient than linear classifiers (e.g. Fisher Discriminant) since they form *non-linear* decision boundaries.
- NN outputs are direct estimators of probability densities.

[12]Optimal classification is then limited by the so-called Bayesian discriminator,

$$y_i(\vec{x}) = \frac{N_i\, p_i(\vec{x})}{\sum_{\text{classes } k} N_k\, p_k(\vec{x})} \quad , \qquad \begin{array}{rcl} y_i(\vec{x}) & = & \text{probability that object } \vec{x} \text{ belongs to class } i \\ N_i & = & \text{number of objects of class } i \\ p_i(\vec{x}) & = & \text{probability density of class } i \end{array} \quad .$$

Classification accuracy is limited by knowledge of the probability densities $p_i(\vec{x})$. For the jet classification problem, just a coarse estimate of $p_i(\vec{x})$, using 10 bins for each of the 14 variables, would require an infeasible look-up table of 10^{14} elements. Clearly a different discriminator is needed.

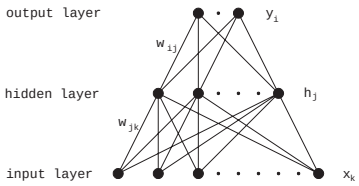


Figure 3.16: Feed-forward network architecture.

- NN's are easily generalized to discriminate > 2 classes.
- NN's learn to ignore useless information.

3.4.2 NN Architecture

A NN is made up of nodes, weights, and thresholds (Fig. 3.4.2). A value [0,1] is associated with each node. Nodes are organized into layers, and each node is "connected" to nodes in adjacent layers. In a feed-forward network, the value at each node is a non-linear weighted function of all nodes in the previous layer:

$$y_i(\vec{x}) = g\left[\sum_j w_{ij}\, g\left(\sum_k w_{jk}x_k + \theta_j\right) + \theta_i\right]$$

weights w, thresholds θ

activation function $g(x)$

Each input node is assigned the value of an input variable. There can be one or more output nodes.

3.4.3 NN Training

Once the NN architecture has been prescribed, the weights and thresholds are fitted using the input distributions during a procedure called training. There are several different training methods; we use the simplest, called "back-propagation". During training, an object of a known class is presented to the net which generates outputs y_i using its current set of weights and thresholds. The desired outputs t_i, corresponding to the class, are known. After

some number of training patterns, N_p, the mean square error

$$E = \frac{1}{2N_p} \sum_{p=1}^{N_p} \sum_i \left(y_i^{(p)} - t_i^{(p)} \right)^2$$

is minimized by varying the weights and thresholds. The minimization is an iterative procedure which uses error gradient descent:

$$w_{t+1} = w_t + \Delta w_t \quad , \text{ where } \quad \Delta w_t = -\eta \, \nabla E_t + \alpha \, \Delta w_{t-1} \quad ,$$

which is controlled by the "learning rate" η and "momentum parameter" α.

The training sample must be large enough to represent the statistical diversity of each class. The NN performance is evaluated on a statistically-independent testing sample, since the net is, by construction, optimized on the training sample.

3.4.4 Implementation

We use the package JETNET 3.5 [38]. The 14 jet variables described in the § 3.3.3 are input to a 14-6-3 feed-forward NN: 14 nodes in the input layer, six nodes in a single hidden layer, and three output nodes. The input variables are rescaled to unit variance so that no one variable dominates during training. The network outputs are defined as flavor outputs; one output is trained to be one for the correct flavor and zero for the others. Other internal NN parameters are:

- Non-linear activation function $g(x) = \frac{1}{2}\left[1 + \tanh(x)\right] = \frac{1}{1+e^{-2x}}$
- Training by back propagation
- Mean-log-squared-error ($\overline{\log E^2}$) minimization by gradient descent
- Momentum parameter $\alpha = 0.5$
- Learning rate during epoch t (defined below) $\eta = \eta_0 \, \gamma^{t-1}$, where the initial learning rate $\eta_0 = 2.0$, and decay rate $\gamma = 0.99$ per epoch.

During training, one jet of each flavor is presented to the NN. The training jet flavor cycles over b,c,p and each jet is chosen randomly from a training sample. One complete cycle over the training sample is called an epoch.

After each training epoch, the NN performance is tested on an independent sample of taggable jets[13]. An equal number of b, c, and p jets are presented to the NN. For each input jet, the NN outputs must be interpreted in order to the assign jet flavor; we make the simplest interpretation by assigning the jet flavor by the greatest NN output. Using this output interpretation rule, the NN tagging efficiencies can be measured. Figure 3.17 shows a tagging figure of merit[14], the network error function, and the learning rate as a function of epoch. The network error and figure of merit initially have large fluctuations which are damped out as the learning rate decreases. The NN settles into a stable internal configuration by the end of training, which we define as 1000 epochs.

The distribution of the three output variables of the trained NN for the sample of test jets is shown in Figure 3.18. The plots show the pairwise correlations between the output variables. As expected, b and p jets are more easily separated than c jets, which appear as an intermediate smear.

3.4.5 NN Performance on Test Sample

Table 3.2 characterizes the performance of the trained NN on the sample of test jets. Taggable jet acceptances (§ 3.3.2) are shown at the left. Next are two 3×3 matrices of tagging efficiencies, for taggable jets (center) and all SVX-fiducial jets (right). The matrix element ϵ_{ij} is the efficiency for a jet of true flavor i to be identified as j. The entries on (off)

[13] An independent test sample is necessary because the NN is optimized for performance on the training sample

[14] The figure of merit is the product of b and c jet signal significances after tagging. For a given signal S (correctly identified jets) in the presence of background B (misidentified jets), the significance is the ratio

$$\text{significance} \equiv \frac{S}{\sqrt{S+B}} = \sqrt{\frac{S \cdot S}{S+B}} = \sqrt{S^{pre} \cdot \epsilon \cdot pur} \quad ,$$

where ϵ is the signal tagging efficiency, pur is the signal purity after tagging (assuming equal numbers of b,c,p jets before tagging), and S^{pre} is the (constant) signal quantity before tagging.

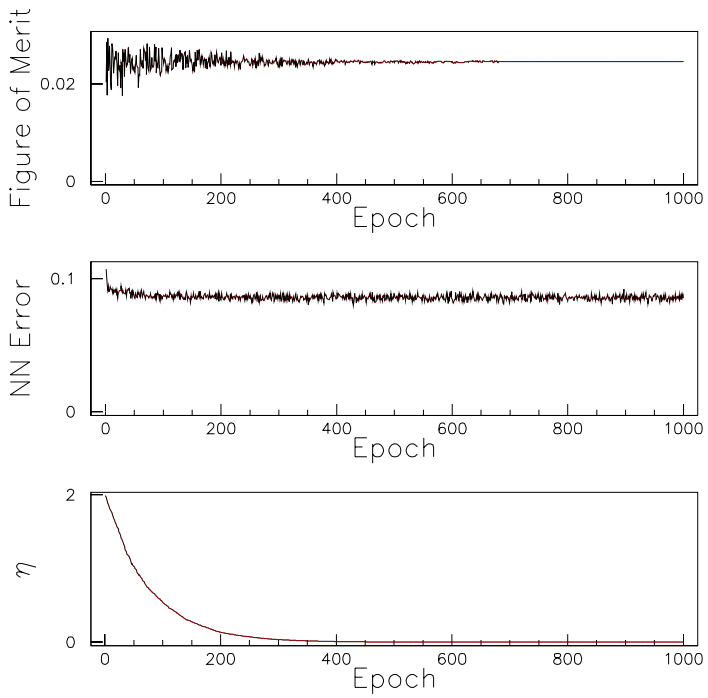


Figure 3.17: Tagging figure of merit $\epsilon_b \cdot pur_b \cdot \epsilon_c \cdot pur_c$, network error function, and learning rate η as a function of training epoch.

Input Flavor	Acceptance	Identified Flavor Efficiency					
		Taggable Jets			All SVX-fiducial Jets		
		b	c	p	b	c	p
b	0.51	0.74	0.20	0.06	0.38	0.10	0.03
c	0.23	0.19	0.58	0.24	0.04	0.13	0.05
p	0.07	0.04	0.11	0.85	0.002	0.01	0.06
		Correct flavor purity			0.89	0.55	0.40

Table 3.2: NN tagging efficiencies for jets with $E_T \geq 15$ GeV and ≥ 1 SVX track. Taggable jets are also required to have ≥ 2 positive i.p. decay tracks with $P_{\text{track}} < 0.3$. The flavor is assigned by the greatest NN output.

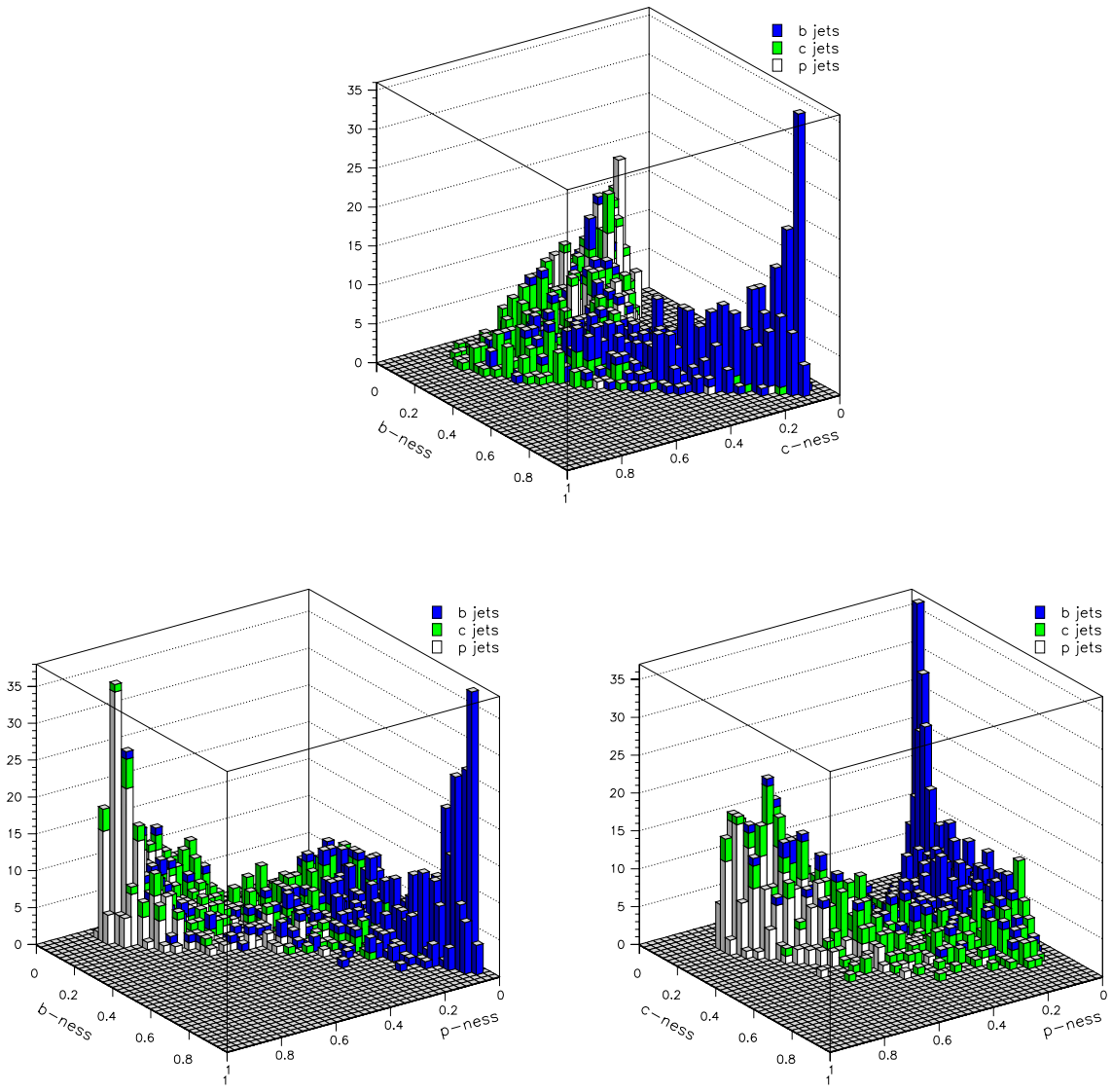


Figure 3.18: 2-dimensional projections of output variables for a trained NN tested on an independent sample of taggable jets. The populations are restricted to the region b-ness + c-ness + p-ness < 1 because of our choice of unit target outputs during training.

the diagonal are correct-tagging (mistagging) efficiencies. Each element in the right matrix is the product of the taggable jet acceptance and the taggable jet identification efficiency. The "correct flavor purity" row shows the resulting purity for each identified flavor assuming an initial sample containing an equal mix of b,c,p jets:

$$pur_i = \frac{N_{ii}}{\sum_{j=b,c,p} N_{ji}} = \frac{N_i^{\text{taggable}} \epsilon_{ii}}{\sum_{j=b,c,p} N_j^{\text{taggable}} \epsilon_{ji}} = \frac{N_i^{\text{pre}} A_i \epsilon_{ii}}{\sum_{j=b,c,p} N_j^{\text{pre}} A_j \epsilon_{ji}} = \frac{f_i A_i \epsilon_{ii}}{\sum_{j=b,c,p} f_j A_j \epsilon_{ji}}$$

where
N_{ij} = Number of jets of flavor i identified as flavor j
N_i^{taggable} = Number of taggable jets of flavor i
ϵ_{ij} = Efficiency for taggable jet of flavor i to be identified as flavor j
N_i^{pre} = Number of SVX-fiducial jets of flavor i
A_i = Taggable acceptance for jets of flavor i
f_i = Fraction ($= \frac{1}{3}$) of SVX-fiducial jets which are flavor i.

We conclude from Table 3.2 that we can separately discriminate b, c, and p jets on a jet-by-jet basis. The correct b-tagging efficiency of 0.38 is slightly less than the CDF standard SECVTX tagger of 0.45 [37], an expected result for an algorithm which is optimized for general flavor tagging, not just for b jets. In addition, this tagger yields the new possibility of relatively pure samples of tagged b and c jets. In the end, this flavor tagging algorithm is just a set of cuts, and therefore incurs an efficiency/purity trade-off. By relaxing the taggable jet criteria, we increase the identification efficiencies but lower the purities. Conversely, by tightening the taggable jet criteria, or more strictly interpreting the network outputs, we can increase the purities at the cost of lower efficiencies. The best choices of these parameters depend on the analysis in which the tagger is used.

Chapter 4

Scale Factors For Data-to-MC Jet Flavor Tagging Efficiencies

In this chapter we discuss the corrections we make to the jet-tagging efficiencies, which we measure in the MC, before their application to jets in the data. First we motivate the need for these corrections, called scale factors, and derive their defining equations. Next we describe three different sets of data and MC events which serve as different jet flavor control samples in the scale-factor measurement. The measurement is made by an over-constrained fit of tagged dijet events in the MC to those in the data. The procedure is first tested on pure MC control samples, and then the fit results to actual data are discussed. Finally, we present some NN discrimination variable distributions which support the overall consistency of the scale-factor measurement.

4.1 Scale Factor Definitions

4.1.1 The Case For Scale Factors

The performance of the NN tagger described in Chapter 3 is characterized by a tagging efficiency matrix as shown in Table 4.1. The tagging efficiencies are measured in MC event samples, and vary depending on the particular physics process being simulated. Our intention is to apply the tagger to actual data. But deficiencies in the MC simulation (including, but not limited to: uncertainties in the fragmentation and decay of heavy quarks, tracking resolution, and calorimeter response [39]) mean that ideally, these efficiencies should

True Flavor	Identification			
	b	c	p	untaggable
b	ϵ_{bb}	ϵ_{bc}	ϵ_{bp}	ϵ_{bu}
c	ϵ_{cb}	ϵ_{cc}	ϵ_{cp}	ϵ_{cu}
p	ϵ_{pb}	ϵ_{pc}	ϵ_{pp}	ϵ_{pu}

Table 4.1: NN jet flavor tagging efficiencies. Since the efficiencies in each row sum to one, there are nine independent parameters.

be measured in the data. However, unlike MC events, where we know the originating quark flavor of each jet, we have no pure control samples of each jet flavor in the data.

To measure the tagging efficiencies in the data, we could try counting tagged jets in a data sample of N jets with a known flavor mixture. Then the numbers of tags are given by

$$N_{j-\mathrm{tag}} \;=\; \sum_i N_i\,\epsilon_{ij} \;=\; \sum_i N\,f_i\,\epsilon_{ij} \tag{4.1}$$

where the sum is over true jet flavors (b, c, p), $f_i = N_i/N$ is the fraction of jets of flavor i, and ϵ_{ij} is the efficiency for a jet of true flavor i to be identified as j. Even if the flavor fractions f_i are known, the four equations (4.1) cannot be solved for the nine unknown efficiencies $\{\epsilon_{ij}\}$.

The tagging efficiencies for a given jet flavor can depend on other features of the jet and event. Table 4.2 shows sample tagging efficiencies, measured in MC, for jets which may contain a muon track for different production processes. There is significant variation across the correct tagging efficiencies for b and c jets. The possibility of separately measuring each of these efficiencies in the data is even more remote than in the basic case outlined above.

Because of the difficulties in directly measuring the tagging efficiencies in the data, we use a different strategy. We define a set of scale factors which give the ratios of various tagging-related efficiencies in the data to MC. Each scale factor corresponds to a quantity for which we posit a significant difference between its value in the data and MC (i.e. a quantity which the MC may model relatively poorly). While various tagging efficiencies depend on the type of events in a given sample (as shown in Table 4.2), the scale factors, by definition, do not. Once the scale factors for a tagger are measured, they can be used to

True Flavor	Lepton Track	Event Type	Identified Flavor Efficiency (%)		
			b	c	p
b	-	gs	23.0 ± 4.9	9.5 ± 3.4	9.5 ± 3.4
b	μ	gs	38.1 ± 1.2	13.1 ± 0.8	2.2 ± 0.4
b	-	fe	32.8 ± 4.1	10.4 ± 2.6	3.0 ± 1.5
b	μ	fe	32.7 ± 1.2	18.3 ± 1.0	1.6 ± 0.3
b	-	dp	38.1 ± 1.2	10.8 ± 0.8	4.2 ± 0.5
b	μ	dp	31.6 ± 1.1	17.8 ± 0.9	1.4 ± 0.3
c	-	gs	4.1 ± 1.7	5.5 ± 1.9	10.3 ± 2.5
c	μ	gs	5.9 ± 0.6	13.5 ± 0.9	5.5 ± 0.6
c	-	fe	5.1 ± 1.7	9.7 ± 2.2	6.3 ± 1.8
c	μ	fe	2.4 ± 0.4	14.7 ± 0.9	3.2 ± 0.4
c	-	dp	4.8 ± 0.8	13.8 ± 1.3	5.6 ± 0.8
c	μ	dp	1.5 ± 0.4	15.5 ± 1.3	3.6 ± 0.7
p	-	-	0.3 ± 0.1	0.9 ± 0.1	6.6 ± 0.3
p	μ	-	0.3 ± 0.1	2.1 ± 0.2	5.4 ± 0.2

Table 4.2: Jet-tagging efficiencies measured in MC low-P_T muon dijet events (discussed in § 4.2.3). The symbol μ indicates that the jet contains a reconstructed muon track with $P_T > 8$ GeV/c. The heavy-flavor production process is given in the "event type" column: gluon splitting (gs), flavor excitation (fe), or direct production (dp).

calculate the tagging efficiencies expected in any data sample of jets given the efficiencies in a corresponding MC sample.

The success of this strategy depends on several factors. First, we must define the scale factors for the correct quantities; this is discussed in the next section. The second linchpin is the assumption that the scale factors are invariant across different event samples. For systematic uncertainties in the MC associated with, for example, the decay of a heavy meson, it is a reasonable assumption. Although the scale factors may still have residual dependencies which we ignore[1], we nevertheless assume their invariance because it allows us to proceed. Furthermore, we expect that any such additional variance of the scale factors is small compared to the uncertainty with which we measure them. This leads to the final success point, which has to do with how well we can measure the scale factors; this is discussed in the remainder of this chapter.

4.1.2 Scale Factor Parameterization

Recall that the jet flavor tag is made in two steps: first, a jet must be taggable, then second, the NN tagger identifies a taggable jet as b,c, or p. We therefore factorize the jet-tagging efficiencies (for either data or MC) into pieces as follows:

$$\epsilon_{ij} = \begin{cases} \epsilon_{it} \cdot \epsilon_{ij}^{\mathrm{NN}} & j = b,c,p \\ 1 - \left(\epsilon_{ib} + \epsilon_{ic} + \epsilon_{ip}\right) & j = u \end{cases} , \tag{4.2}$$

$$i = b,c,p \text{ (true flavor)} , \quad j = b,c,p,u \text{ (tag outcome)}$$

where $\epsilon_{it} = 1 - \epsilon_{iu}$ = efficiency for jet of flavor i to be taggable

$\epsilon_{ij}^{\mathrm{NN}}$ = efficiency for taggable jet of flavor i to be identified as j by NN

Note that the L.H.S. of (4.2) represents 12 efficiencies, of which nine are independent. The parameterization on the R.H.S. uses a different set of 12 efficiencies, but we can then

[1] By "residual dependence", we mean dependence on a variable which is poorly modeled by the MC. For example, although we expect jet-tagging efficiency scale factors to vary with the jet E_T, we can ignore this natural dependence because the MC models the jet E_T distribution adequately well.

use the fact that the NN makes a flavor identification for each taggable jet to write three new relations:

$$\epsilon_{ib}^{\rm NN} + \epsilon_{ic}^{\rm NN} + \epsilon_{ip}^{\rm NN} \;=\; 1 \quad , \quad i = b, c, p \text{ (true flavor)}.$$

Therefore the R.H.S. of (4.2) also contains nine independent parameters, but in a more physical representation. For jets of true flavor i, our convention is that the efficiencies $\epsilon_{it}, \epsilon_{ib}^{\rm NN}, \epsilon_{ic}^{\rm NN}$ are independent, and $\epsilon_{ip}^{\rm NN} = 1 - \epsilon_{ib}^{\rm NN} - \epsilon_{ic}^{\rm NN}$. Then in terms of the independent efficiencies, (4.2) becomes (for either data or MC):

$$\begin{aligned} \epsilon_{ib} &= \epsilon_{it} \cdot \epsilon_{ib}^{\rm NN} \\ \epsilon_{ic} &= \epsilon_{it} \cdot \epsilon_{ic}^{\rm NN} \\ \epsilon_{ip} &= \epsilon_{it} \cdot \left(1 - \epsilon_{ib}^{\rm NN} - \epsilon_{ic}^{\rm NN}\right) \\ \epsilon_{iu} &= 1 - \epsilon_{it} \end{aligned} \qquad , \qquad i \;=\; b, c, p \text{ (true flavor)} \tag{4.3}$$

We now introduce scale factors which relate the data and MC tagging efficiencies in the following way,

$$\epsilon_{it}^{\rm data} \;=\; s_{it}\, s_l\, \epsilon_{it}^{\rm MC} \qquad \epsilon_{ib}^{\rm NN,data} \;=\; s_{ib}\, \epsilon_{ib}^{\rm NN,MC} \qquad \epsilon_{ic}^{\rm NN,data} \;=\; s_{ic}\, \epsilon_{ic}^{\rm NN,MC} \tag{4.4}$$

where the extra superscripts denote data or MC efficiencies. For each jet flavor i, there are three scale factors: s_{it} for the taggable efficiency, and $s_{ib,ic}$ for the NN b, c identification efficiencies for taggable jets. An additional scale factor, s_l, modifies the taggable efficiency for any flavor jet which contains a lepton track with $P_T > 8$ GeV/c (jets without a lepton track have $s_l = 1$). By substituting the scale factor definitions (4.4) into the jet-tagging efficiency expressions (4.3, written for data jets), we relate the data and MC tagging efficiencies via

scale factors:

$$
\begin{aligned}
\epsilon_{ib}^{\rm data} &= \epsilon_{it}^{\rm data} \cdot \epsilon_{ib}^{\rm NN,data} \\
&= s_{it}\, s_l\, \epsilon_{it}^{\rm MC}\, s_{ib}\, \epsilon_{ib}^{\rm NN,MC} \\
&= s_{it}\, s_l\, s_{ib}\, \epsilon_{ib}^{\rm MC} \\
\\
\epsilon_{ic}^{\rm data} &= \epsilon_{it}^{\rm data} \cdot \epsilon_{ic}^{\rm NN,data} \\
&= s_{it}\, s_l\, \epsilon_{it}^{\rm MC}\, s_{ic}\, \epsilon_{ic}^{\rm NN,MC} \\
&= s_{it}\, s_l\, s_{ic}\, \epsilon_{ic}^{\rm MC} \\
\\
\epsilon_{ip}^{\rm data} &= \epsilon_{it}^{\rm data} \cdot \left(1 - \epsilon_{ib}^{\rm NN,data} - \epsilon_{ic}^{\rm NN,data}\right) \\
&= s_{it}\, s_l\, \epsilon_{it}^{\rm MC} \cdot \left(1 - s_{ib}\, \epsilon_{ib}^{\rm NN,MC} - s_{ic}\, \epsilon_{ic}^{\rm NN,MC}\right) \\
&= s_{it}\, s_l \left(1 - \epsilon_{iu}^{\rm MC} - s_{ib}\, \epsilon_{ib}^{\rm MC} - s_{ic}\, \epsilon_{ic}^{\rm MC}\right) \\
\\
\epsilon_{iu}^{\rm data} &= 1 - \epsilon_{it}^{\rm data} \\
&= 1 - s_{it}\, s_l\, \epsilon_{it}^{\rm MC} \\
&= 1 - s_{it}\, s_l \left(1 - \epsilon_{iu}^{\rm MC}\right) \qquad , \qquad i \;=\; b, c, p \text{ (true flavor).}
\end{aligned}
\tag{4.5}
$$

There are 10 scale factors in the equations above, and we discuss their measurement in the next section.

4.2 Scale-Factor Measurement Technique

4.2.1 Double-Tag Combinations In Dijet Events

As explained in § 4.1.1, the jet-tagging efficiency scale factors for a given tagger are defined to be invariant across different event samples. In this section, we introduce our method of measuring these scale factors for the NN tagger described in Chapter 3.

We begin by considering dijet event samples of MC and data events. Given four possible tag outcomes for each jet (b,c,p,u), there are 16 possible double-tag combinations for distinguishable jets (to be explained in the following sections). We then count the number

of events, separately for data and MC, with each double-tag combination,

$$N_{pq-\mathrm{tag}} = \text{number of events in which the first jet was } p\text{-tagged and the second jet was } q\text{-tagged,}$$

$$p, q = \text{tag outcome of first, second jets } (b, c, p, u)$$

The scale factors can then be measured, as described in § 4.2.5, by comparing the numbers of double tags observed in data with those expected from MC.

The power of this measurement technique depends on the double-tagged event statistics of the event sample used. For example, the measurement accuracy of the scale factors involving true c jets improves with the number of events containing c jets. With the assumption of scale factor invariance, we proceed to measure them by simultaneously using three different event samples, reconstructed D^*, low-P_T muon, and inclusive jets events. These three sets of events serve as separate control samples for each true jet flavor, and both a MC and data sample is required for each. The event samples are described in the following sections.

4.2.2 Reconstructed D^* Events

In these events, we seek to reconstruct the charm meson decay,

$$D^{*\pm} \to D^0\ \pi_s^{\pm} \quad , \quad D^0 \to K^{\mp}\ \mu^{\pm}\ \nu_\mu$$

where π_s denotes a soft (low-P_T) pion[2]. These events serve mostly as the charm control sample for the scale-factor measurement.

The data sample begins with events from the Run 1b trigger stream MULB_5B, which requires the presence of at least one trigger-level central muon candidate with $P_T > 8$ GeV/c.

For our MC sample, we used PYTHIA to generate 2 parton $\to$ 2 parton events[3] (see § 5.2.2 for more detail on MC programs mentioned here). We then processed events with

[2]Because of the small (6 MeV/c^2) mass difference between the D^* and its daughters, $D^0\ \pi_s$, the daughters have very little kinematic energy available in the rest frame of the decay.

[3]The physics processes were $f_i f_j \to f_i f_j$, $f_i \bar{f}_i \to f_k \bar{f}_k$, $f_i \bar{f}_i \to gg$, $f_i g \to f_i g$, $gg \to f_k \bar{f}_k$, $gg \to gg$, generated with $\hat{p_t}^{\min} = 25$ GeV/c and $|\eta| < 3.0$. For this and future references to 2 $\to$ 2 processes, we note that the simulation includes the possibility of QCD initial and final-state radiation. Therefore multijet events (such as the 2 $\to$ 3 heavy-flavor production processes of § 1.4) may result from these core processes.

QQ for two reasons. First, the decays of b and c hadrons are replaced with a better model. Second, to reduce event generation time, if the event contained a D^* meson, it was forced to have the desired decay above[4]. Next, we only kept events in which a D^* decayed to daughters including a muon with $P_T > 8$ GeV/c and $|\eta| < 1$. Finally, the generated events were put through the CDFSIM and QFL detector simulations. Our D^* MC sample contained 76,000 events at this point.

We then applied event selection criteria to both data and MC events. The first set of criteria follows that used for a charm quark analysis [40]. These criteria are meant to select $D^{*\pm}$ events with the above decays by exploiting a characteristic reconstructed mass difference between the D^* and D^0: [5]

- Event contains ≥ 3 SVX tracks which can be associated with the μ, K, π_s, with charges $Q(\mu) = -Q(K)$
- μ candidate has CMU and CMP detector hits
- $P_T^{\mu} \geq 8$ GeV/c, $P_T^K \geq 1$ GeV/c
- η-ϕ separation $\Delta R\,(\mu - K) < 1.0$
- z-coordinate match $\Delta z(\mu, K, \pi_s) < 5$ cm
- I.p. significance $\pi_s(d/\sigma_d) < 5$
- Reconstructed D^0 mass $1.3 \leq M(\mu, k) < 1.8$ GeV/c^2
- Event has only one combination of tracks satisfying above criteria
- Reconstructed D^*-D^0 mass difference $\Delta M = M(\mu, K, \pi_s) - M(\mu, K) < 0.23$ GeV/c^2

We define "right sign" (RS) events to have $Q(\mu) = Q(\pi_s)$, otherwise they are "wrong sign" (WS). The combinatoric background to the above selection criteria are equally likely to be

[4] In one-half of the events we forced the D^{*+} decay and did not force the D^{*-} decay, and vice-versa in the other half. Note that we did not force the decay of the D^0.

[5] The difference in the reconstructed masses of the D^* and D^0 peaks near $M_\pi + 6$ MeV/c^2 (the Q value of the decay). The peak is distorted, but still present, because the neutrino goes undetected.

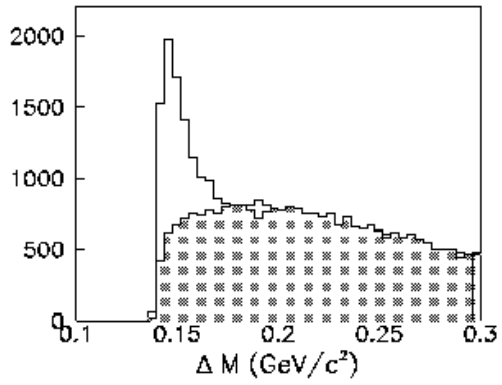


Figure 4.1: The ΔM distribution for D^* candidate events with RS (open) and WS (shaded) combinations in the data. The number of RS and WS events outside the D^* peak are about equal; there is an excess of 5913 RS events in the peak.

RS or WS. After all cuts in the data, we have 9866 RS and 3953 WS events. Note that all MC events are RS.

The mass difference distribution ΔM for D^* candidate events (with all but the last cuts above applied) in the data is shown in Figure 4.1. The D^* signal is shown by the mass difference peak for RS events. The events outside of the peak are combinatoric background. Since the RS and WS combinatoric backgrounds have very similar shapes outside the peak, we assume their shapes are also the same under the peak. Therefore we can use the WS events to model the combinatoric background. Then the D^* signal is isolated by subtracting the WS events from the RS events, giving 5913 D^* events.

After applying the D^* selection, both data and MC events were processed by the CDF offline code described in § 5.3.1.2. We then applied the following event selection criteria for the scale-factor measurement:

- Event contains two jets only
- First jet is identified as "μ jet" with:
  - Jet $E_T \geq 15$ GeV
  - Contains ≥ 1 SVX track
  - Contains muon track, η-ϕ separation $\Delta R(\mu - \text{jet}) < 0.4$

- $P_T^{\mu} \geq 9$ GeV/c

- Second jet is identified as "away jet" with:
  - Jet $E_T \geq 15$ GeV
  - Contains ≥ 1 SVX track
  - Back-to-back with μ jet, $\Delta\phi(\mu$ jet - away jet$) > \frac{3}{4}\pi$

- Event primary vertex z-coordinate $|\mathrm{PV}(z)| < 30$ cm

These cuts leave 608 RS and 271 WS data events and 1243 MC events. According to the MC, 11/89% of muon jets are b/c, and 39/37/24% of events have the muon jet heavy flavor produced by gs/fe/dp processes.

4.2.3 Low-P_T Inclusive Muon Events

This sample contains events containing muon candidates with $P_T > 9$ GeV/c, regardless of their source. Events containing b or c quarks which decay semi-leptonically to a muon serve as control samples for the scale-factor measurement. However, in the data, muon candidates also come from calorimeter punch-through and pion decay-in-flight, which together we call "fakes". Note that we do not know *a priori* the ratio of muon candidates from heavy flavor to fakes.

The data sample consists of 146,162 events from the Run 1b trigger stream MULB_5B, which require the presence of at least one trigger-level central muon candidate with $P_T > 8$ GeV/c.

The MC sample has two parts, both of which were generated similarly as the D^* sample described in § 4.2.2, from PYTHIA $2 \to 2$ processes and QQ (with no forced decays).

The first MC sample contains events with muons from semi-leptonic b and c decays. After the initial event generation, we required the presence of a generator level muon with $P_T^{\mu} > 8$ GeV/c and $|\eta| < 1.0$. This muon signal sample has 43,785 events.

The second MC sample models the fake muon candidates. After the initial event generation, we excluded events which contained a b or c quark. We then required each event to contain, at the generator level, a final state charged particle with $P_T > 8$ GeV/c, which was then changed to a muon (at the generator level). We therefore model the fake muon component of the data by these transformed charged particles (usually pions) which are necessarily prompt (not originating from a secondary vertex of a decaying meson). The latter assumption is reasonably true for the punch-through fakes, but less obvious for the decay-in-flight fakes, whose tracks need not point back to the primary vertex. Despite this shortcoming, this is the most reasonable MC model of the fake muon events available. This fake muon sample has 38,520 events.

Both samples of MC events were processed by the CDFSIM and QFL detector simulations. Next, events from data and both MC samples which pass the D^* selection criteria listed in § 4.2.2 were rejected, leaving 139,355 data, 36,926 MC signal, and 28,758 MC fake events. It was necessary to reject the D^* candidate events from these samples to maintain independent control samples for the scale-factor measurement. Finally, both data and MC events were processed by the CDF offline code, and the following additional event selection criteria were applied:

- Event contains two jets only
- First jet is identified as "μ jet" with:
  - Jet $E_T \geq 15$ GeV
  - Contains ≥ 1 SVX track
  - Contains muon track, η-ϕ separation $\Delta R(\mu - \text{jet}) < 0.4$
  - $P_T^{\mu} \geq 9$ GeV/c
- Second jet is identified as "away jet" with:
  - Jet $E_T \geq 15$ GeV
  - Contains ≥ 1 SVX track

– Back-to-back with μ jet, $\Delta\phi(\mu$ jet - away jet$) > \frac{3}{4}\pi$

- Event primary vertex z-coordinate $|\mathrm{PV}(z)| < 30$ cm

These cuts leave 44,687 data, 8744 MC signal, and 9054 MC fake events. According to the MC signal sample, 55/45% of muon jets are b/c, and 36/35/29% of events have the muon jet heavy flavor produced by gs/fe/dp processes.

4.2.4 Inclusive Jets Events

These QCD jet events make up the majority of our prompt control sample for the scale-factor measurement.

The data sample begins with the Run 1b trigger stream QJ2B_5P, which require the presence of at least one trigger-level jet with $E_T > 20$ GeV.

The MC sample was generated from PYTHIA $2 \to 2$ processes ($\hat{p_T}^{\min} = 25$ GeV) and QQ (no forced decays), followed by the CDFSIM and QFL detector simulations.

Both data and MC events were processed by the CDF offline code. To obviate the need for a trigger simulation, we required both data and MC events to contain at least one reconstructed jet with $E_T > 30$ GeV (well above above trigger threshold effects), leaving 184,609 data and 80,995 MC events. We then applied the following criteria to all events:

- Event contains two jets only
- First jet is defined as the "leading jet" (higher E_T), with $E_T > 30$ GeV
- Second jet is defined as "away jet", with $E_T > 15$ GeV
- Each jet contains ≥ 1 SVX track
- Leading and away jets are back-to-back, $\Delta\phi$(leading jet - away jet$) > \frac{3}{4}\pi$
- Event primary vertex z-coordinate $|\mathrm{PV}(z)| < 30$ cm

These cuts leave 22,276 data and 8,595 MC events. According to the MC sample, 2.4/5.0/92.6% of leading jets are $b/c/p$, and 34/47/19% of events with heavy-flavor leading jets are produced by gs/fe/dp processes.

4.2.5 Fitting the Scale Factors

In § 4.2.1 we introduced our scale factor-measurement strategy of counting double-tag combinations in dijet events with distinguishable jets. In the preceding three sections, we described the control samples to be used in this measurement. The dijet events in the D^* and low-P_T muon samples have distinguishable jets: the muon jet and the away jet. However, dijet events in the inclusive jets sample have indistinguishable jets because, although we label them as leading and away jets based on their E_T, they lack any other physical distinction. Therefore the first two control samples each give 16 double-tag combination measurements, while the latter gives just 10. [6] To simplify the following discussion, we will refer to the first and second jet as the muon and away jet, respectively, in all control samples.

For each control sample, the number of each combination of double-tagged events in the data can be expressed as the following sum,

$$N_{pq-\mathrm{tag}}^{\mathrm{data}} = \sum_{\substack{i,j\,=\,b,c,p \\ k\,=\,\mathrm{gs,fe,dp,nhf}}} N_{ijk}\ \epsilon_{ip}^{\mu,\,k,\,\mathrm{data}}\ \epsilon_{jq}^{\mathrm{away},\,k,\,\mathrm{data}} \quad , \tag{4.6}$$

where i,j = true flavor of μ, away jets (b,c,p)

k = production mechanism of event (gs,fe,dp,nhf=no heavy flavor)

p,q = tag outcome of μ, away jets (b,c,p,u)

N_{ijk} = number of type k events with μ, away jets of true flavor i,j

and the jet-tagging efficiencies in the data $\left\{\epsilon_{ij}^{\mathrm{data}}\right\}$ are related to those in the MC $\left\{\epsilon_{ij}^{\mathrm{MC}}\right\}$ by the scale factors $\{s\}$ through (4.5). From Table 4.2, we expect the jet-tagging efficiencies to depend on the presence of a muon and the production mechanism of the event. We therefore include separate terms for muon and away jet-tagging efficiencies, each of which

[6] For the indistinguishable inclusive jets, there are 4 double-tag combinations where both jets have the same tag, and $(16-4)/2 = 6$ combinations (ignoring order) where they have different tags.

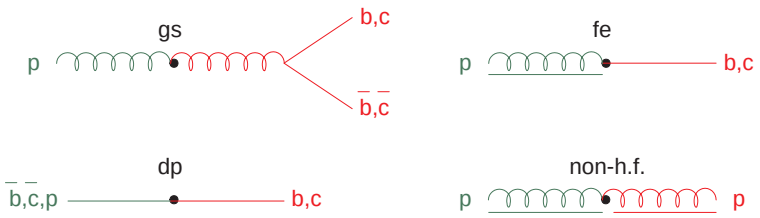


Figure 4.2: Schematic representation of the likely muon and away jet flavor combinations for each event production mechanism considered: gluon splitting (gs), flavor excitation (fe), direct production (dp), and no heavy flavor (nhf). In each diagram, the dijets emerge from the interaction point, with the muon jet to the right and the away jet to the left.

can also vary with the event type k[7]. Each term in the sum gives the double-tag contribution expected from dijet events of a particular production mechanism with a particular true jet flavor configuration.

The dijet flavor flavor event counts, N_{ijk}, are estimated from MC in the following way. Figure 4.2 shows the some possible dijet true flavor combinations for each event type considered here. In a control sample of N dijet events, we define

$$\begin{aligned} \beta_k &= \text{fraction of events of type } k \\ \alpha_{ijk} &= \text{fraction of type } k \text{ events in which } \mu\text{, away jets were flavor } i,j \end{aligned}$$

so that $N_{ijk} = N\,\beta_k\,\alpha_{ijk}$ can be substituted in (4.6). The fractions β_k and α_{ijk} are measured separately in each MC control sample, and were reported at the ends of § 4.2.2 – § 4.2.4.

We then calculate a χ^2 function from the observed and predicted mean numbers of events with each possible double-tag outcome in each control sample:

$$\chi^2 = \sum_{\substack{\text{control} \\ \text{samples}}} \sum_{\substack{pq\,=\,b,c,p,u \\ \text{double tag} \\ \text{outcomes}}} \left[\left(N^{\text{data}}_{pq-\text{tag}}\right)^{\text{obs}} - \left(N^{\text{data}}_{pq-\text{tag}}\right)^{\text{pred}} \right]^2 / \sigma^2 \qquad (4.7)$$

where the predicted means are given by (4.6)[8] and depend on the scale factors $\{s\}$, and σ^2 is the expected variance of the (unsquared) numerator term, and has two contributions. First,

[7]The parameterization by event type allows the possibility of changing the relative mix of event types from the MC. In practice, we varied only the non-heavy-flavor fraction of events (which could not be derived from the MC); the scale factor fit does not converge if the others vary too.

[8]For the D^* sample, the predicted mean is the sum of the D^* signal (4.6) and the WS background counts in the data, and the observed number is the RS counts in the data.

if we predict a mean of $\left(N_{pq-\mathrm{tag}}^{\mathrm{data}}\right)^{\mathrm{pred}}$ events with pq-tag outcome in a control sample of N events, the binomial variance of the observed number of such events is

$$\left(\sigma^2\right)^{\mathrm{binom}} \;=\; \left(N_{pq-\mathrm{tag}}^{\mathrm{data}}\right)^{\mathrm{pred}} \left(1 - \frac{\left(N_{pq-\mathrm{tag}}^{\mathrm{data}}\right)^{\mathrm{pred}}}{N}\right) \quad .$$

But the predicted mean itself, because of uncertainties in efficiencies and event fractions measured in the MC, has additional variance given by

$$\left(\sigma^2\right)^{\mathrm{stat}} \;=\; \sum_{v \,=\, \{\epsilon^{\mathrm{MC}}\},\{\alpha\},\{\beta\}} \left(\frac{\partial \left(N_{pq-\mathrm{tag}}^{\mathrm{data}}\right)^{\mathrm{pred}}}{\partial v}\right)^2 \sigma_v^2 \quad ,$$

where v denotes each independent variable in (4.6), and σ_v its statistical error. Since the statistical variance of the predicted mean and the natural binomial variance about it are independent, the net variance is given by their sum in quadrature,

$$\sigma^2 \;=\; \left(\sigma^2\right)^{\mathrm{binom}} + \left(\sigma^2\right)^{\mathrm{stat}} \quad ,$$

which we put in (4.7), thereby incorporating the effect of finite MC statistics into the χ^2 function.

We use the MINUIT function minimization program [41] to find the 10 scale factor values $\{s\}$ which minimize the χ^2, a procedure we call the "scale factor fit". One additional fit parameter is needed, β_{nhf}, for the low-P_T muon sample only, because we do not have a MC estimate of the fake muon fraction in that control sample, as explained in § 4.2.3.[9] We then have 11 parameters (which vary freely over non-negative values) to fit using 16 (D^*) + 16 (low-P_T μ) + 10 (jets) = 42 double-tag measurements. Since the fit also requires as input the total number of dijet data events in each control sample, we are left with $42-11-3 = 28$ degrees of freedom (d.o.f.) in the measurement.

[9] We tried letting each event fraction β_k vary as independent parameters, but the fit would not converge. Instead, we fix the ratios $\beta_{\mathrm{gs}} : \beta_{\mathrm{fe}} : \beta_{\mathrm{dp}}$ to their MC values and constrain their sum to $1 - \beta_{\mathrm{nhf}}$.

Fit Parameter	Input Value	Fitted Value and Error	
		Test	Full
$\beta_{\rm nhf}$	0.51	0.52 ± 0.03	0.51 ± 0.02
s_l	1	0.98 ± 0.07	1.00 ± 0.05
s_{bt}	1	1.02 ± 0.09	1.00 ± 0.06
s_{bb}	1	0.94 ± 0.05	1.00 ± 0.04
s_{bc}	1	1.12 ± 0.11	1.00 ± 0.07
s_{ct}	1	1.01 ± 0.13	1.00 ± 0.09
s_{cb}	1	1.07 ± 0.28	1.00 ± 0.21
s_{cc}	1	1.06 ± 0.11	1.00 ± 0.08
s_{pt}	1	1.02 ± 0.06	1.00 ± 0.04
s_{pb}	1	1.13 ± 0.68	1.00 ± 0.42
s_{pc}	1	0.88 ± 0.18	1.00 ± 0.13

Table 4.3: Input and fitted values of parameters with errors in the scale factor fit tested on pure MC control samples. The "test" column shows results for the case where the MC was divided into independent halves for the MC and "data" portions of the control samples (fit has χ^2/d.o.f. = 0.49), while the "full" column shows results when the same full set of MC events was used for both the MC and "data" portions (fit has χ^2/d.o.f. = 0.0).

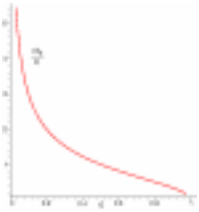
4.3 Scale Factor Fit Results

4.3.1 Test on MC

To verify that the scale factor fit converges reasonably and to gauge its sensitivity, we performed the following check. We divided each MC control sample in half, using the first half as the MC portion of the sample and the second as the "data" portion.

The results of the test fit are shown in Table 4.3 under the "test" column. Since we used only MC events, the input value of each scale factor is one, to which most fit within error. The additional parameter for the fake muon event fraction, $\beta_{\rm nhf}$, also shows good agreement with its input value. Note that the absolute values of the fit parameter errors decrease as the number of events in either the data or MC control samples increase (the variance in the denominator of the χ^2 decreases relative to the fluctuations of the numerator, thereby "tightening up" the fit). The relative sizes of the errors indicate our relative sensitivity to each. The scale factors with the largest errors tend to be those corresponding to the smallest

tagging efficiencies (e.g. s_{cb}, s_{pb}, s_{pc}). Since a scale factor is a ratio of efficiencies, its error, loosely speaking, is like a relative error on an efficiency. But the relative error σ_ϵ on an efficiency ϵ after N trials is



$$\frac{\sigma_\epsilon}{\epsilon} \;=\; \frac{\sqrt{N\,\epsilon\,(1-\epsilon)}/N}{\epsilon} \;=\; \sqrt{\frac{1-\epsilon}{N\epsilon}} \quad ,$$

which diverges for small values of ϵ. We therefore we expect to measure scale factors for small efficiencies with less precision than those for large efficiencies.

We cannot judge whether or not the fit has a systematic bias from just this one test. To correctly do so would require repeating the test fit procedure with a large number of independent MC samples (each of adequate size to ensure fit convergence), but is prohibited by the large MC generation time required. We did, however, confirm that the fit is free of any sort of gross bias by the following test. For each control sample, the full MC sample was used as both the MC and "data" portions during the fit. For this case of identical MC and "data" samples, all fit parameters converged to their true values within ~0.1%, as shown in Table 4.3 under the "full" column. It is reasonable to conclude from this and the previous test that the scale factor fit procedure works as intended.

4.3.2 Results on Data

Table 4.4 shows the results of the scale factor fit (for the NN tagger of Chapter 3) using control samples containing the full set of MC and data events described in § 4.2.2 – § 4.2.4. We defer the discussion of the fitted non-h.f. fraction in the low-P_T muon control sample, $\beta_{\rm nhf}$, until § 4.3.3. The remaining parameters are discussed here.

Since $s_l = 1.10$, we expect data jets with a lepton track to be taggable 10% more often than corresponding MC jets. Either lepton tracks in the data are more likely to pass the decay track selection criteria, or they are correlated with tracks which do so. Another

Fit Parameter	Fitted Value and Error
$\beta_{\rm nhf}$	0.56 ± 0.02
s_l	1.10 ± 0.05
s_{bt}	0.78 ± 0.05
s_{bb}	1.04 ± 0.04
s_{bc}	0.78 ± 0.08
s_{ct}	1.17 ± 0.11
s_{cb}	0.82 ± 0.23
s_{cc}	0.76 ± 0.08
s_{pt}	0.89 ± 0.03
s_{pb}	1.76 ± 0.40
s_{pc}	0.84 ± 0.10

Table 4.4: Parameter values and errors returned by the scale factor fit using data and MC control samples. The fit has $\chi^2/\text{d.o.f.} = 0.30$.

interesting result of the fit is that for taggable jets, the mis-b-tag efficiency from prompt jets is almost double its MC value. The scale factor errors are in ratios which agree with the tests of § 4.3.1.[10]

Table 4.5 shows the fitted and observed double-tag counts and their contributions to the fit χ^2 in the three control samples. The agreement between the fit and data is generally good, and the largest discrepancies come from tag combinations with low statistics.

We perform an additional cross check of the scale factors for prompt jets, s_{p*}, using the Jet20 control sample. First we count the numbers of taggable, b-, and c-tagged jets in the data. Although Jet20 contains mostly prompt jets, some of these taggable and tagged jets are b or c. These heavy-flavor contributions must be subtracted off before measuring the prompt jet-tagging efficiencies in the data. To estimate these contributions, we measure jet flavor fractions ($f_b \simeq 0.02,\ f_c \simeq 0.05$) and efficiencies ($\epsilon_{b*},\ \epsilon_{c*}$) in the Jet20 MC. The efficiencies are then corrected by the scale factors in Table 4.4. Using these flavor fractions and corrected efficiencies, we estimate the numbers of taggable and tagged b or c jets in the data, and then subtract them from our original counts to yield the numbers of taggable and tagged prompt

[10]The absolute values of the fit errors are smaller here than in the fit tests because the control samples used in the data fit are larger than for the fit tests (where the control samples were MC only).

μ - away	D^*			Low-p_T inclusive μ			Jet20		
jet tags	Fit + WS	RS	χ^2	Fit	Data	χ^2	Fit	Data	χ^2
u - u	291.3	296	0.1	31976.6	31920	0.1	18653.1	18695	0.1
u - b	6.6	6	0.0	924.5	930	0.0	515.6	530	0.2
u - c	10.9	8	0.7	539.5	559	0.2	458.5	415	2.0
u - p	25.1	17	2.0	2020.5	2096	0.6	2411.4	2421	0.0
b - u	56.4	70	2.9	3070.5	2991	0.7			
b - b	3.3	10	4.5	352.1	365	0.2	8.5	16	2.4
b - c	1.2	5	1.4	111.7	106	0.1	17.2	13	0.5
b - p	8.8	5	1.3	210.5	222	0.1	45.4	41	0.1
c - u	88.5	90	0.0	2331.1	2419	1.1			
c - b	5.3	1	1.9	173.6	156	0.8			
c - c	4.4	1	1.2	93.5	87	0.2	1.5	4	1.8
c - p	9.9	11	0.1	239.4	209	1.1	33.3	43	0.9
p - u	87.2	77	0.7	2294.4	2288	0.0			
p - b	2.0	3	0.1	72.7	72	0.0			
p - c	3.4	0	1.2	56.8	50	0.1			
p - p	3.6	8	1.6	219.5	217	0.0	131.5	98	2.2

Table 4.5: Dijet double-tag counts, and their contribution to the χ^2, in each control sample for the converged scale factor fit. Note that the Jet20 sample has only 10 distinguishable combinations.

True Flavor	Lepton Track	Event Type	Identified Flavor Efficiency (%)		
			b	c	p
b	-	gs	18.7 ± 2.5	5.8 ± 1.9	8.3 ± 3.9
b	μ	gs	34.2 ± 2.1	8.8 ± 1.1	3.0 ± 1.8
b	-	fe	26.7 ± 2.2	6.4 ± 1.5	3.1 ± 3.0
b	μ	fe	29.3 ± 1.8	12.4 ± 1.4	3.7 ± 1.9
b	-	dp	30.9 ± 1.7	6.6 ± 0.8	3.9 ± 1.7
b	μ	dp	28.3 ± 1.8	12.0 ± 1.4	3.5 ± 1.8
c	-	gs	4.0 ± 2.4	4.9 ± 1.9	14.5 ± 4.0
c	μ	gs	6.3 ± 2.1	13.3 ± 1.6	12.7 ± 2.3
c	-	fe	4.9 ± 2.4	8.7 ± 2.0	11.2 ± 3.7
c	μ	fe	2.5 ± 1.0	14.4 ± 1.7	9.2 ± 2.0
c	-	dp	4.6 ± 1.7	12.3 ± 1.7	11.5 ± 2.4
c	μ	dp	1.6 ± 1.1	15.2 ± 1.9	9.8 ± 2.5
p	-	-	0.4 ± 0.4	0.7 ± 0.2	5.8 ± 0.5
p	μ	-	0.5 ± 0.5	1.7 ± 0.3	5.4 ± 0.7

Table 4.6: Jet-tagging efficiencies expected in low-P_T muon dijet events (discussed in § 4.2.3) with data-to-MC scale factors. The symbol μ indicates that the jet contains a reconstructed muon track with $P_T > 8$ GeV/c. The heavy-flavor production process is given in the "event type" column: gluon splitting (gs), flavor excitation (fe), or direct production (dp).

jets. These last numbers give the taggable and tagging efficiencies for prompt jets which lead directly to their scale factors: $s_{pt} = 0.87 \pm 0.03$, $s_{pb} = 1.97 \pm 0.49$, $s_{pc} = 0.78 \pm 0.21$. [11] These scale factor values agree, within error, with their fitted counterparts in Table 4.4.

Table 4.6 shows the tagging efficiencies expected in the data for jets which may contain a muon track for different production processes. These efficiencies were calculated by applying the scale factors in Table 4.4 to the MC efficiencies in Table 4.2. The errors include both the statistical error on the original efficiencies and the scale factor errors and correlations.

[11] Errors include MC statistics and b, c jet scale factor errors, but no correlations.

4.3.3 Jet Flavor Fractions in the Low-P_T Muon Control Sample

The scale factor fit returns, in addition to scale factors, the non-h.f. event fraction in the low-P_T muon control sample, $\beta_{\rm nhf} = 0.56 \pm 0.02$. This additional parameter gives the fraction of muon events which do not contain heavy flavor, and is required because we lack for it an *a priori* estimate, as mentioned in § 4.2.3.

Using the fitted $\beta_{\rm nhf}$ and jet flavor fractions measured in the MC, we calculate the flavor fractions for the pretag (before requiring taggable criteria) jets in the data low-P_T muon control sample:

$$\begin{array}{rcc} & \underline{\quad\mu \text{ jets}\quad} & \underline{\quad\text{Away jets}\quad} \\ f_b \;=\; & (1-\beta_{\rm nhf})\; f_b^{\mu,\ \rm sig} & (1-\beta_{\rm nhf})\; f_b^{\rm away,\ sig} \\ f_c \;=\; & (1-\beta_{\rm nhf})\; f_c^{\mu,\ \rm sig} & (1-\beta_{\rm nhf})\; f_c^{\rm away,\ sig} \\ f_p \;=\; & \beta_{\rm nhf} & \beta_{\rm nhf} \;+\; (1-\beta_{\rm nhf})\; f_p^{\rm away,\ sig} \end{array} \tag{4.8}$$

where $f_i^{\{\mu,\rm away\},\ \rm sig}$ is the fraction of muon or away jets which have flavor i, as measured in the MC muon signal sample. The prompt fraction in muon jets is $\beta_{\rm nhf}$, by definition. Each non-heavy-flavor event gives a prompt away jet, but the prompt fraction has an additional contribution from prompt jets recoiling against heavy-flavor muon jets. The b and c fractions, for both muon and away jets, are given by their respective fractions in MC muon signal events, normalized by the fraction of heavy-flavor events $(1-\beta_{\rm nhf})$.

Similarly, we calculate the flavor fractions for taggable jets in the data using $\beta_{\rm nhf}$, taggable jet flavor fractions measured in MC, and appropriate scale factors:

$$\begin{array}{rcc} & \underline{\quad\text{Taggable } \mu \text{ jets}\quad} & \underline{\quad\text{Taggable away jets}\quad} \\ f_b \;=\; & K^\mu\;(1-\beta_{\rm nhf})\; f_{b,\ \rm taggable}^{\mu,\ \rm sig}\; s_{bt}\; s_l & K^{\rm away}\;(1-\beta_{\rm nhf})\; f_{b,\ \rm taggable}^{\rm away,\ sig} \\ f_c \;=\; & K^\mu\;(1-\beta_{\rm nhf})\; f_{c,\ \rm taggable}^{\mu,\ \rm sig}\; s_{ct}\; s_l & K^{\rm away}\;(1-\beta_{\rm nhf})\; f_{c,\ \rm taggable}^{\rm away,\ sig} \\ f_p \;=\; & K^\mu\;\beta_{\rm nhf}\; f_{p,\ \rm taggable}^{\mu,\ \rm fake}\; s_{pt}\; s_l & K^{\rm away}\;\Big[\beta_{\rm nhf}\; f_{p,\ \rm taggable}^{\rm away,\ fake}\; s_{pt} + \\ & & (1-\beta_{\rm nhf})\; f_{p,\ \rm taggable}^{\rm away,\ sig}\; s_{pt}\Big] \end{array} \tag{4.9}$$

where $f_{i,\ \rm taggable}^{\{\mu,\rm away\},\ \{\rm sig,fake\}}$ is the fraction of muon or away jets which have flavor i and are taggable, as measured in the MC muon signal or fake sample. The contributions to the

	Fraction of jets (%)			
Jet	Pretag		Taggable	
flavor	μ jets	away jets	μ jets	away jets
b	24.1 ± 1.2	9.3 ± 0.5	52.4 ± 2.5	33.9 ± 2.5
c	19.5 ± 1.0	5.3 ± 0.3	26.8 ± 2.5	12.9 ± 1.7
p	56.4 ± 2.1	85.3 ± 0.8	20.8 ± 2.2	53.2 ± 1.7

Table 4.7: Estimated pretag and taggable jet flavor fractions for muon and away jets in the low-P_T muon data control sample.

flavor fractions have a similar structure as in (4.8), with the difference here that the MC fractions also require taggable jets. As a result, the MC fractions implicitly include the taggable jet efficiencies, and are therefore multiplied by appropriate scale factors: s_{it} for jets of flavor i, and additionally s_l for muon jets. The normalization constants $K^{\{\mu,\text{away}\}}$ are chosen so that the flavor fractions sum to one.

Table 4.7 shows the values of the jet flavor fractions in the low-P_T muon data sample, estimated using (4.8) and (4.9). The errors on the fractions are calculated from the statistical error on MC event fractions, the errors on the fitted values of β_{nhf} and the scale factors, and their correlations[12]. The heavy-flavor fractions are greater in muon jets than away jets (as expected when muons come from semi-leptonic b or c decays), and they also increase for taggable jets (heavy-flavor jets are more likely to contain the required decay tracks).

We perform the following consistency check of the estimated taggable jet flavor fractions. Starting with the NN discrimination variable distributions for taggable b, c, and p jets in the low-P_T muon MC samples, we add them in the ratio given in Table 4.7. Figure 4.3 shows these distributions, along with their sums compared to the data, for the variables: jet probability P_{jet}, $\log_{10} P_{\text{jet}}$, decay track multiplicity N_{decay}, and decay tracks mass M_{tracks}. There is good agreement between the summed flavor components and the data in both muon and away jets[13]. Since the low-P_T muon control sample contains a mix of b,c, and p jets,

[12]The scale factor fit returns the covariance matrix of the fit parameters.

[13]The excess of low jet probability jets in the MC compared to data is a known shortcoming of the MC, and is largely responsible for the poor goodness-of-fit of these distributions.

and we find agreement between the data and MC in distributions which depend on parameters measured in the scale factor fit ($s_{*t}, \beta_{\mathrm{nhf}}$), we conclude that tagging in the MC + scale factors approximates tagging in the data reasonably well.

Similar comparison plots for all NN input (discrimination) and output (flavor) variables, for all control samples, are found in Appendix C.

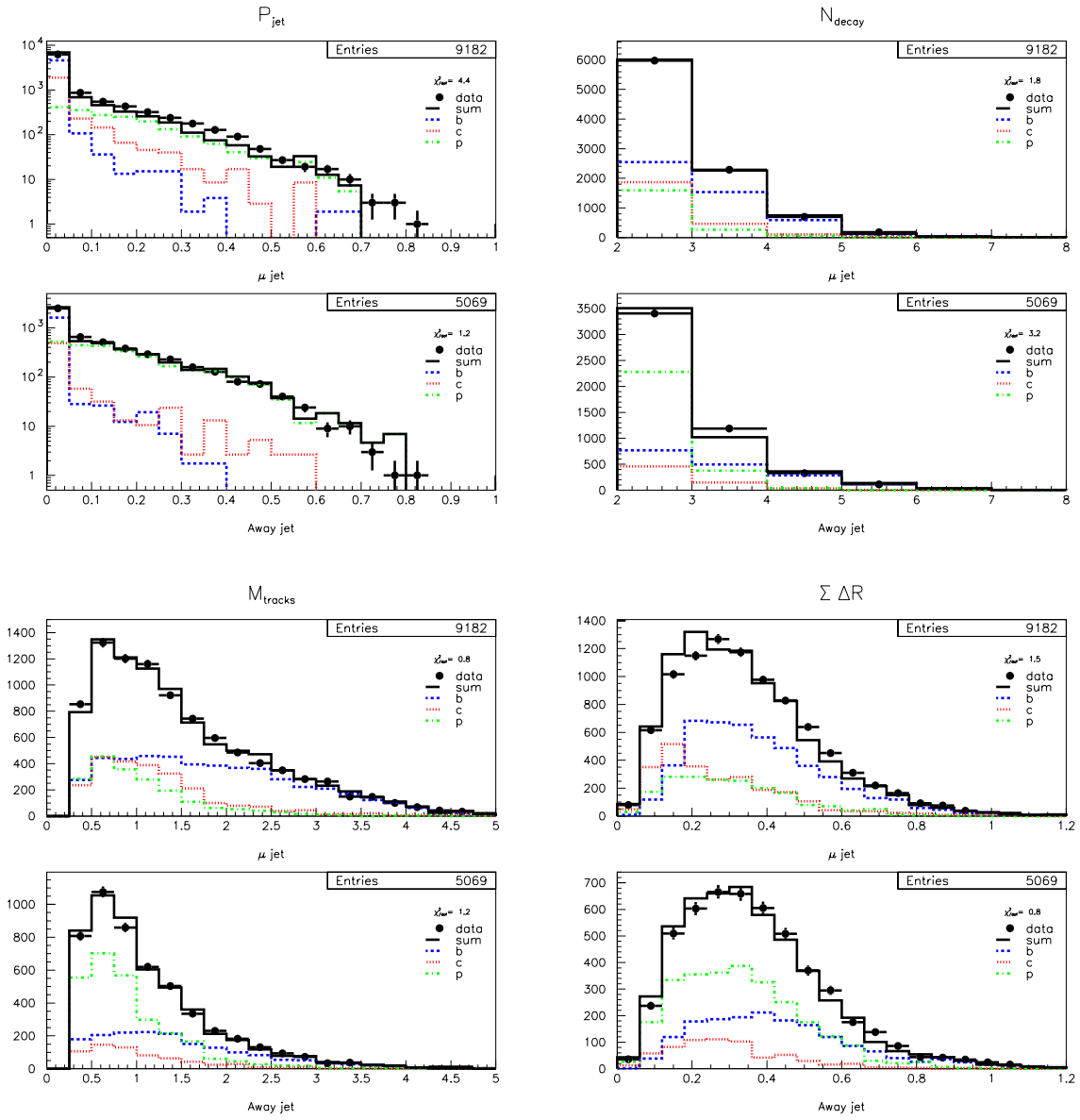


Figure 4.3: Distributions of NN discrimination variables for taggable muon and away jets in the low-P_T muon control samples: jet probability $P_{\rm jet}$, decay track multiplicity $N_{\rm decay}$, decay tracks mass $M_{\rm tracks}$, and decay track spread $\sum \Delta R$. Separate MC distributions are shown for b,c,p jets in the ratio given in Table 4.7, and their sum is normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

Chapter 5

Event Samples and Selection

In this chapter we postpone the jet flavor identification discussion in order to discuss the non-SM top decays for which we search. We first describe SM and non-SM top-quark decay signatures. Next, we introduce the data and MC samples used in the search. Finally, we describe the W + jets event selection criteria and efficiencies which target our top-quark search channels.

5.1 Top Quark Production and Decays

In $p\bar{p}$ collisions at the Tevatron, top quarks pairs are produced by both gluon-gluon fusion and quark-antiquark annihilation [42]. At $\sqrt{s} = 1.8$ TeV and $m_t = 175$ GeV/c^2, the latter is dominant[43]. Figure 5.1 (left) shows the production of a $t\bar{t}$ pair followed by the SM decay chain.

We search for the FCNC decay of a top quark to a charm quark and a gluon. In $t\bar{t}$ events we consider mixed SM-FCNC top quark decays, where one top quark decays to cg and the other undergoes SM decay (or vice-versa), shown in Figure 5.1 (right). We neglect the dual FCNC decay signature for two reasons: we expect the branching fraction $B(t \to cg) \ll 1$, and the dual FCNC final-state lacks a lepton which is required by the event trigger (described in § 5.2.1). SM top-quark decays include two real W bosons, while the

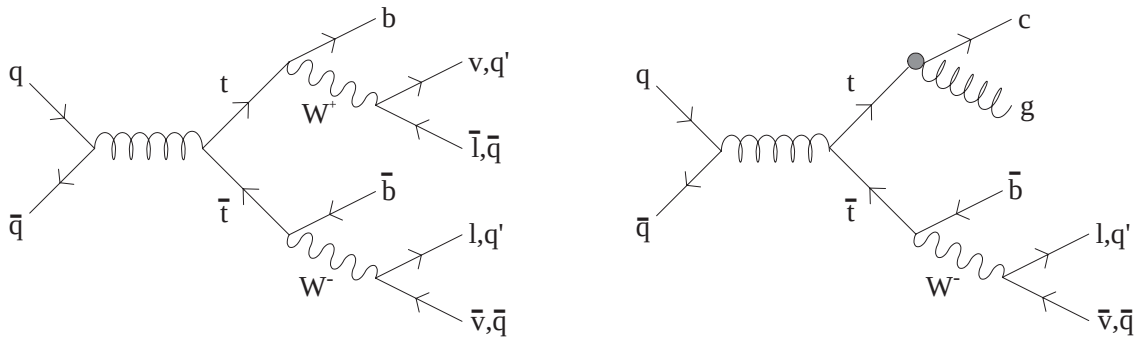


Figure 5.1: Tree level top quark pair production by $q\bar{q}$ annihilation followed by SM (left) or hypothetical mixed SM-FCNC (right) decay chain.

$t\bar{t} \to (Wb)(Wb)$		$t\bar{t} \to (Wb)(cg)$	
Final State	Relative Branching Ratio	Final State	Relative Branching Ratio
$(q\bar{q}'b)(q\bar{q}b)$	36/81	$(q\bar{q}'b)(cg)$	6/9
$(q\bar{q}'b)(e\nu b)$	12/81	$(e\nu b)(cg)$	1/9
$(q\bar{q}'b)(\mu\nu b)$	12/81	$(\mu\nu b)(cg)$	1/9
$(q\bar{q}'b)(\tau\nu b)$	12/81	$(\tau\nu b)(cg)$	1/9
$(e\nu b)(\mu\nu b)$	2/81		
$(e\nu b)(\tau\nu b)$	2/81		
$(e\nu b)(e\nu b)$	1/81		
$(\mu\nu b)(\tau\nu b)$	2/81		
$(\mu\nu b)(\mu\nu b)$	1/81		
$(\tau\nu b)(\tau\nu b)$	1/81		

Table 5.1: Final states (disregarding order) of a $t\bar{t}$ pair undergoing SM and mixed SM-FCNC decays and approximate branching ratios (q denotes a light quark: u, d, s, c).

mixed SM-FCNC mode has only one. In either case, the final state is characterized by the W decay(s), summarized in Table 5.1.

The W boson will most often decay to a quark-antiquark pair, but an enormous QCD multijet background makes isolation of these fully hadronic final states unfeasible. This background is substantially reduced if we require a leptonic W decay to an e or μ (backgrounds to hadronic τ decays are difficult to reduce). Approximately $\frac{24}{81}$ (30%) of SM $t\bar{t}$ decays include one (and only one) e or μ, while only $\frac{2}{9}$ (22%) of mixed SM-FCNC decays do; the reduction for the mixed mode is due to the absence of a second W boson in

the event. This "lepton + jets" final state therefore includes a charged lepton with large transverse momentum, missing energy from the undetected neutrino, and a number of jets from the hadronized quarks: four in SM $t\bar{t}$ decays, three in mixed SM-FCNC $t\bar{t}$ decays.

In this analysis, we search for the top quark FCNC decay into cg resulting in the lepton + jets final state. This chapter introduces the data and MC samples used and describes the criteria used to select events for the analysis.

5.2 Event Samples

5.2.1 Data

The data used in this analysis were collected from $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV with the CDF detector from 1994 to 1995 (Run 1B), corresponding to an integrated luminosity of $\int \mathcal{L}\, dt = 86.34 \pm 3.52$ pb^{-1} [44]. The sample originates from electron and muon triggers[1]. From these data sets, inclusive electron and muon samples were made with tight electron and muon cuts, but no isolation or missing energy cuts. These inclusive samples are the input to this analysis: approximately 129,000 electron events and 91,000 muon events.

5.2.2 Monte Carlo

Three MC event generators were used in this analysis: ISAJET [45], PYTHIA [46], and VECBOS [47]. ISAJET 7.06 is a $p\bar{p}$ MC which uses perturbative QCD cross sections, initial and final state QCD radiative corrections, independent fragmentation and underlying event models. PYTHIA 5.6 is a general purpose MC which can generate a wide variety of processes using tree level matrix elements and parton distributions[2]. PYTHIA events were then processed by JETSET 7.3, which simulates radiation with parton showers, models jets

[1] To avoid double counting, events satisfying both electron and muon triggers were removed from the electron sample

[2] The region of phase space in which PYTHIA generates events was specified by the following kinematic cuts on the particles emerging from the hard scattering: $P_T > 5.0$ GeV/c and $|\eta| < 5.0$.

Sample	Generator	Number of Events
$t\bar{t} \to WbWb$	PYTHIA	48792
$t\bar{t} \to Wbcg$ (NN train)	ISAJET	80011
(analysis)		39997
$Wb\bar{b}$	VECBOS	90677
$Wc\bar{c}$	VECBOS	71044
Single Top (Wg)	PYTHIA	20085
Single Top (W^*)	PYTHIA	20047
Wc	PYTHIA	63862
W + non-h.f.	PYTHIA	185889
Z + jets	PYTHIA	124228
$Z \to \tau\tau$	PYTHIA	26671
Dibosons	PYTHIA	22680

Table 5.2: MC samples, generators, and numbers of events after generation cuts.

using string fragmentation, and decays particles. VECBOS is a leading order MC for inclusive production of a W or Z boson plus jets in hadron colliders.

After the initial event generation, the decays of bottom and charm hadrons were replaced with a better model using QQ 9.0 [48]. Similarly, τ lepton decays were redone by the TAUOLA module [49] using a consistent model. An updated particle property database and decay table were used for all particle decays [50].

To reduce the sample sizes (in anticipation of later cuts), we required that each event include, regardless of source, an electron or muon with $P_T > 15$ GeV/c and $|\eta| < 3.0$. These kinematic cuts are sufficiently relaxed from later analysis cuts to avoid bias.

MC events were then passed through a CDF detector simulation. First, CDFSIM [51] simulated the tracking chambers' response (including the SVX′) by tracing each particle through the detector, generating individual hits as inputs for pattern recognition. Next, QFL [52] simulated the calorimeter and muon chambers using a parameterized detector response.

Table 5.2 summarizes the MC samples, and specific details for each are given next.

5.2.2.1 Signal

To search for FCNC top quark decays in lepton + jets events, we need MC samples of both pure SM decay $t\bar{t}$ events and mixed SM-FCNC decay $t\bar{t}$ events.

The direct production of top quark pairs with SM decays was modeled by PYTHIA $q\bar{q} \to t\bar{t}$ and $gg \to t\bar{t}$ processes.

In mixed SM-FCNC decay $t\bar{t}$ events, one top quark decays to Wb and the other to cg. We use ISAJET to generate $t\bar{t}$ events in which one top quark has a SM decay and the other a FCNC decay. In the simulated FCNC decay, the top quark is changed into a charm quark and gluon. In the rest frame of the top quark, the two-body final state is back-to-back and the decay proceeds by 4-momentum conservation despite the lack of a model for such a matrix element. We generated equal numbers of the following types of events: $t\bar{t} \to W^+ b\ \bar{c}g$ and $t\bar{t} \to cg\ W^-\bar{b}$. Two separate signal samples were generated: one for NN training only, and another for use in the rest of the analysis.

5.2.2.2 Backgrounds

Significant backgrounds in lepton + jets events come from the higher-order production of a real W or Z boson which recoils against significant jet activity. There is an additional *non-W* component which we measure from the data, as discussed in § 6.3. The backgrounds that we simulate are listed below.

$Wb\bar{b}, Wc\bar{c}$ In these events, a W boson is produced in association with heavy-quark pairs (Figure 5.2). This background was modeled by VECBOS $W + n$ jet matrix elements (with $Q^2 = M_W^2 + P_{TW}^2$). Then the HERWIG [53] MC program added initial/final state radiation and provided parton shower fragmentation. We used only events containing a quark-antiquark pair; this pair was then changed into $b\bar{b}$ or $c\bar{c}$. With this approach, the heavy-quark mass does not enter into the matrix element, causing the cross section to diverge as the jets become collinear. However, for quark $P_T > 15$ GeV, the cross section error due to the omission of the quark mass is less than 5% [54]. Partons in events

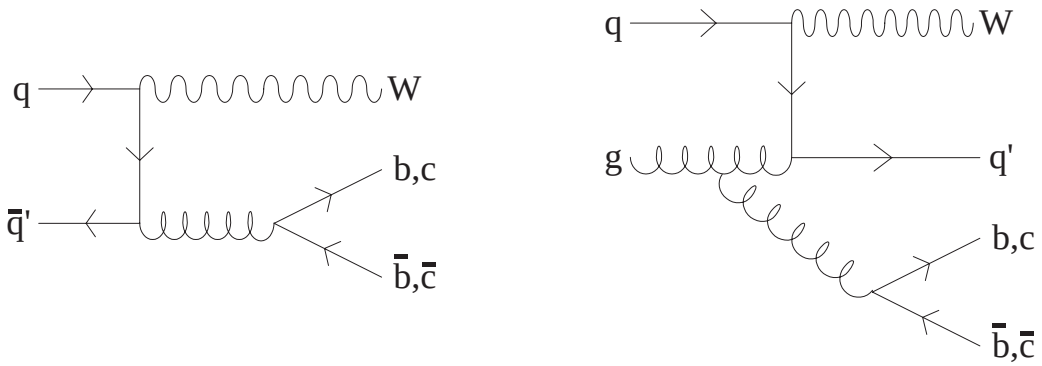


Figure 5.2: Leading and next-to-leading order diagrams for $Wb\bar{b}/Wc\bar{c}$ production.

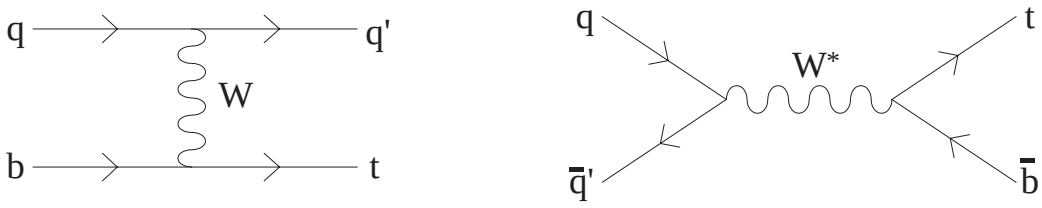


Figure 5.3: Wg fusion and W^* production of single top.

were required to have $P_T > 8$ GeV/c, $|\eta| < 3.5$, and η-ϕ separation < 0.4. Events were required to include a lepton with $P_T > 8$ GeV/c.

Single Top Single top quark production comes primarily from t-channel "Wg fusion" and s-channel W^* electroweak processes (Figure 5.3). Samples of each type were generated using PYTHIA.

Wc These events involve a $Wc\bar{s}$ vertex, where the s quark comes from the proton sea (Figure 5.4). This background was generated by the PYTHIA $qg \to q'W$ process. Events were required to have a c quark as a product of the hard scattering with $P_T > 10$ GeV/c and $|\eta| < 2.5$.

W + *non-h.f.* These events are W production in association with light quark or gluon (non-heavy-flavor) jets, and were generated by PYTHIA $qg \to q'W$ and $q\bar{q}' \to gW$ processes. Events were required to not include b or c quarks.

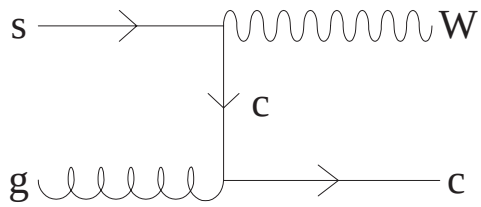


Figure 5.4: Leading order diagram for Wc production.

Z + jets These events are Z production in association with jets, and were generated by PYTHIA $q\bar{q} \to gZ$ and $qg \to qZ$ processes.

$Z \to \tau\tau$ The production of a single Z resonance was generated by PYTHIA $q\bar{q} \to Z$ process. We only used events in which the Z decayed to a τ lepton pair.

Dibosons These events are the production of electroweak boson pairs, and were generated by PYTHIA $q\bar{q} \to WW/ZZ$ and $q\bar{q}' \to WZ$ processes.

5.3 Event Selection Criteria

5.3.1 Analysis Path

5.3.1.1 Event Triggers

Data events were selected online by various central electron or muon and missing energy triggers; the exact trigger path is documented elsewhere [55].

In MC events, electron triggers were not required because these triggers have a combined efficiency of ∼100% for events with electrons passing cuts in §5.3.2.1 [56]. MC events lacking such electrons were required to contain a muon which fired an appropriate trigger[3].

[3]The muon trigger was simulated with the offline routine SIM_MUTRIG [57]

5.3.1.2 Event Reconstruction

Data and MC events are processed identically using Version_7_12 of the CDF offline code to reconstruct events[4]. The primary interaction vertex position in the transverse plane is found for each event and track parameters are calculated with respect to it[5]. A fixed cone size of 0.4 in η-ϕ space is used to cluster jets and associate tracks with jets.

Simulated events undergo some additional processing. Track finding efficiency is degraded using an algorithm which reproduces the efficiencies measured in the data according to the CTC environment around the track[33]. Jet flavor is determined by $\eta - \phi$ matching between the calorimeter jet axis and generator-level hadrons.

We use the same event selection criteria as other CDF top quark analyses using quantities calculated in the TOPFND module [55].

5.3.2 W Selection

The leptonic decay products of a W boson for which we search are an isolated, high momentum electron or muon and a neutrino. The neutrino escapes direct detection but its presence is inferred by energy imbalance in the event. We require one high P_T electron or muon passing appropriate triggers and missing transverse energy $\not\!\!E_T > 20$ GeV in the event. Details of these cuts, which are standard for CDF top analyses [43], are given below.

5.3.2.1 Electron Identification

Table 5.3 summarizes the cuts used to identify electron candidates. We use electrons in the central rapidity region ($|\eta| \leq 1.0$). Central electron candidates have a reconstructed CTC track pointing to a CEM cluster. The shower position and profile is measured in the CES. Fiducial cuts on the shower position ensure that the candidate is away from calorimeter

[4] Actually we use a version of the package which includes a modified TRCHOR routine to fix a bug in the calculation of the track impact parameter error.

[5] The VXPRIM module performs a fit of SVX tracks seeded by the beam position. Tracks with large impact parameter ($d/\sigma_d > 3$) are not used and the fit is iterated until a stable primary vertex is found.

boundaries and that the energy is well measured. This fiducial volume covers 84% of the solid angle for central electrons.

We reject electrons candidates which can be paired with an oppositely-charged CTC track to form a small mass since they might come from a photon conversion.

5.3.2.2 Muon Identification

Table 5.4 lists the cuts used to identify central muons ($|\eta| \leq 1.0$). For muon candidates, we require a match between a reconstructed CTC track and a track stub in the CMU, CMP, or CMX.

5.3.2.3 Missing Energy Measurement

The missing transverse energy ($\not\!\!E_{\rm T}$) in an event is defined as the negative vector sum of the transverse energy in all calorimeter towers with $|\eta| < 3.6$. Individual tower energies must exceed detector-dependent thresholds to be included in the sum [58]. For events with muon candidates, the missing energy is corrected by substituting the muon P_T (measured in the CTC) for the minimum-ionizing-muon energy in the calorimeter.

Figure 5.5 shows the $\not\!\!E_{\rm T}$ distribution for $t\bar{t}$ events with an isolated high P_T lepton. The SM and SM-FCNC distributions are quite similar, and we require $\not\!\!E_{\rm T}$> 20 GeV, as per the standard CDF top selection.

5.3.3 Jet Cuts

We select jets using standard CDF top-quark analysis criteria (referred to as TOPFND jets [55]): uncorrected jet $E_T > 15$ GeV and $|\eta| < 2.0$. Calorimeter clusters associated with isolated electrons or non-isolated muons are not used.

Figure 5.6 shows the jet multiplicity for both SM and SM-FCNC $t\bar{t}$ events with W candidates and a typical background, $Wc\bar{c}$. Fluctuations are caused by acceptance losses, merging of two adjacent jets into one, and gluon radiation. For the $B(t \to cg)$ measurement,

Quantity	Description	Cut
Rapidity	Rapidity with respect to detector origin	$\|\eta\| \leq 1.0$
E_T	Transverse energy in CEM cluster	> 20 GeV
HAD/EM	Ratio of hadronic to electromagnetic calorimeter energy of the cluster	< 0.05
E/P	Ratio of cluster energy to track momentum	< 1.8
L_{shr}	Comparison of the lateral shower profile with that of test beam electrons: $L_{shr} = 0.14 \sum_i \frac{E_i^{obs} - E_i^{test}}{\sqrt{(0.14\sqrt{E})^2 + \sigma^2_{E_i^{test}}}}$ where the sum is over towers adjacent to the seed tower, E_i^{obs} is the observed tower energy with CEM resolution $0.14\sqrt{E}$, E_i^{test} is the expected energy from test beam data with error $\sigma_{E_i^{test}}$	< 0.2
Strip χ^2	Comparison of the CES shower profile with that of test beam electrons	< 10
Track-strip match	Distance in position between the extrapolated track and CES shower in the r-ϕ (Δx) and r-z (Δz) views	$\|\Delta x\| < 1.5$ cm $\|\Delta z\| < 3.0$ cm
z-vertex match	Distance between interaction vertex and track origin in the z direction	$\|z_{vtx} - z_0\|$ < 5.0 cm
z-vertex	Track origin z coordinate	$\|z_0\| < 60.0$ cm
Isolation	Ratio of the sum of transverse energy in towers within an η-ϕ circle of radius 0.4 centered on (but excluding) the electron to the electron E_T	$I_{cal}/E_T(e)$ < 0.1

Table 5.3: Central electron candidate selection criteria.

Quantity	Description	Cut
Rapidity	Rapidity with respect to detector origin	$\|\eta\| \leq 1.0$
P_T	Transverse momentum	> 20 GeV/c
EM	EM tower energy	< 2.0 GeV
HAD	HAD tower energy	< 6.0 GeV
Track-segment match	Distance between extrapolated track and muon chamber track stub	$\|\Delta x\| < 2.0$ cm (CMU), 5.0 cm (CMP,CMX)
z-vertex match	Distance between interaction vertex and track origin in the z direction	$\|z_{vtx} - z_0\|$ < 5.0 cm
z-vertex	Track origin z coordinate	$\|z_0\| < 60.0$ cm
Impact parameter	Distance of closest approach of the track to the origin	$d < 0.3$ cm
Isolation	Ratio of the sum of transverse energy in towers within an η-ϕ circle of radius 0.4 centered on (but excluding) the muon to the muon P_T	$I_{cal}/P_T(\mu) < 0.1$

Table 5.4: Central muon candidate selection criteria.

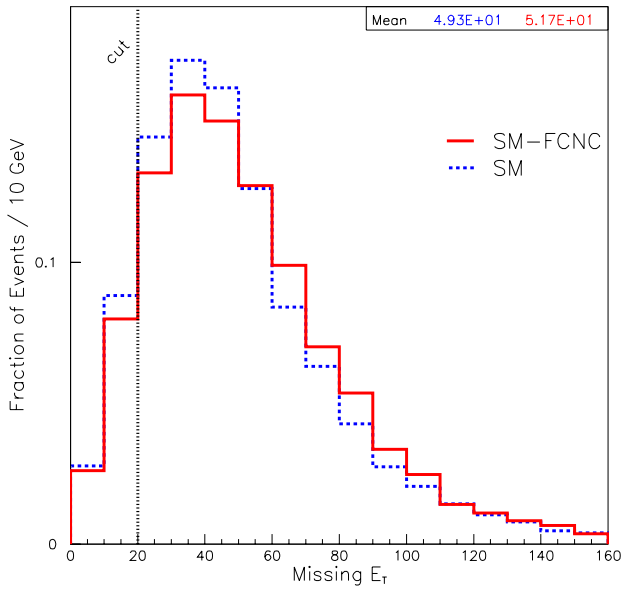


Figure 5.5: $\not\!\!E_{\rm T}$ for SM and SM-FCNC decay $t\bar{t}$ events containing an isolated high P_T lepton. The vertical dotted line indicates the cut $\not\!\!E_{\rm T}>$ 20 GeV.

we will cut at $N_{\rm jets} \geq 3$ to retain much of the top signal but reject most of the backgrounds. For the rest of this chapter we will, however, report quantities for all jet multiplicities.

5.3.4 Event Removals

Bad Run We exclude events from runs ($< 5\%$ of total) during which some detector systems (usually muon systems) were nonfunctional as recorded in the run-summary database.

Z **Candidate Veto** The decay $Z \rightarrow l\bar{l}$ can fake the $W \rightarrow l\nu$ signature if there is mismeasured energy elsewhere in the event. Therefore we exclude events which contain a second lepton with $P_T > 10$ GeV/c which forms an invariant mass with the primary lepton in the range $75 < M_Z < 105$ GeV/c^2. This cut removes $< 1\%$ of the W-candidate events.

Dilepton Removal Z decays to dileptons, in which one lepton fails selection cuts or misses the fiducial region of the calorimeter, are not rejected by the Z candidate veto above. In

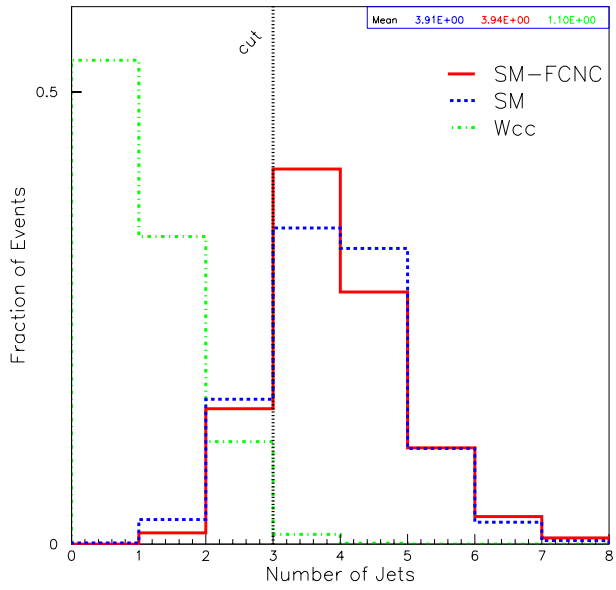


Figure 5.6: Jet multiplicities for SM and SM-FCNC decay $t\bar{t}$ events containing a W candidate, and a typical background, $Wc\bar{c}$. The vertical dotted line indicates the cut $N_{\text{jets}} \geq 3$.

addition, SM $t\bar{t}$ dilepton events (where both W's decay leptonically) can contribute to the lepton + jets signal. These kinds of dilepton events are removed in two passes. First, we reject events which satisfy the criteria of the CDF $t\bar{t}$ dilepton analysis: two leptons of opposite sign passing certain event topology cuts. Second, we use the extended dilepton removal of [59] to discriminate against $t\bar{t}$ dilepton events where one lepton escapes detection in the calorimeter. We reject events with an isolated, high momentum ($P_T > 15$ GeV/c) isolated track with charge opposite to that of the primary lepton. These cuts reject an additional 7% of the W-candidate events.

5.4 Efficiencies of the W + jets Selection Cuts For Top Events

A determination of the fraction of events of a given process which pass the W + jets cuts allows us to estimate its contribution to the W + jets sample. Because the contributions

Process	Efficiency (%) for W selection *and*					
	≥ 0 jets	1 jet	2 jets	3 jets	$\geq$3 jets	≥ 4 jets
$t\bar{t} \to WbWb$	9.0 ± 1.2	0.2 ± 0.1	1.5 ± 0.2	3.2 ± 0.4	7.2 ± 1.0	4.0 ± 0.5
$t\bar{t} \to Wbcg$	6.1 ± 0.8	0.1 ± 0.1	0.9 ± 0.1	2.5 ± 0.3	5.1 ± 0.7	2.6 ± 0.3
Single Top (Wg)	5.0 ± 0.7	1.5 ± 0.2	2.2 ± 0.3	0.9 ± 0.1	1.1 ± 0.2	0.2 ± 0.1
Single Top (W^*)	4.7 ± 0.6	1.1 ± 0.2	2.6 ± 0.4	0.7 ± 0.1	0.9 ± 0.1	0.2 ± 0.1

Table 5.5: Efficiencies for top events in data to pass W + jets cuts.

from most background processes will be estimated using another method (Chapter 6), we only need these efficiencies for top events.

The efficiency for a $t\bar{t}$ event to pass the W + jets selection can be expanded as a product of individual cut efficiencies:

$$\epsilon_{W+\text{jets}} \;=\; \epsilon_{W\,\text{selection}} \cdot \epsilon_{\text{jets}} \;=\; \epsilon_{\text{trigger}} \cdot \epsilon_{\text{lepton-ID}} \cdot \epsilon_{\not\!\!E_T} \cdot \epsilon_{\text{jets}}$$

We use MC to estimate these efficiencies for the data. A difference in the combined trigger and lepton-ID efficiencies between data and MC is accounted for by the scale factor:

$$s_{tl} \;=\; \frac{\epsilon^{\text{data}}_{\text{trigger}} \cdot \epsilon^{\text{data}}_{\text{lepton-ID}}}{\epsilon^{\text{MC}}_{\text{trigger}} \cdot \epsilon^{\text{MC}}_{\text{lepton-ID}}} \;=\; 0.942 \pm 0.126 \quad . \tag{5.1}$$

This scale factor was determined separately for electrons and muons using HERWIG simulation and Run 1B data [60]. The results were averaged according to the relative composition in simulated top events. The error includes systematic uncertainties in both the trigger and lepton-ID efficiencies. The W + jets efficiency in data is then related to the value in MC by:

$$\epsilon^{\text{data}}_{W+\text{jets}} \;=\; s_{tl} \cdot \epsilon^{\text{MC}}_{W+\text{jets}} \qquad \text{where} \quad \epsilon^{\text{MC}}_{W+\text{jets}} = \frac{N^{\text{MC}}_{W+\text{jets}}}{N^{\text{MC}}_{generated}} \tag{5.2}$$

is the ratio of MC events passing W + jets cuts to the number generated[6].

The expected W + jets efficiencies for top events in data are shown in Table 5.5. The efficiency is calculated separately for each process using relation 5.2. The error includes both the uncertainty on s_{tl} and the statistical uncertainty on $\epsilon^{\text{MC}}_{W+\text{jets}}$.

[6] $N^{MC}_{generated}$ is the number of MC events generated *before* the generator-level kinematic cuts in §5.2.2.

Chapter 6

Expected Composition of the W + Jets Sample

With the W + jets selection defined, we must now anticipate the makeup of the resulting sample which will be used throughout the rest of this analysis. We discuss here the techniques to estimate the composition of the W + jets data before applying jet flavor tagging criteria (i.e. the *pretag* sample). In the following discussion, we refer to all events except $t\bar{t}$ as backgrounds.

6.1 Single Top Events

We estimate the number of single top events in the W + jets data sample using production cross sections from theory [61] [62] and efficiencies from MC. The number of events in the data passing W + jets cuts can be expressed as the product

$$N_{W+\text{jets}} \;=\; \sigma \;\cdot\; \int \mathcal{L}\; dt \;\cdot\; \epsilon^{\text{data}}_{W+\text{jets}} \tag{6.1}$$

where σ is the cross section, $\int \mathcal{L}\; dt$ is the integrated luminosity of the data set (§ 5.2.1), and $\epsilon^{\text{data}}_{W+\text{jets}}$ is the efficiency for events in the data to pass the W + jets cuts (§ 5.4). The calculation is done separately for Wg and W^* processes and the results are shown in the upper portion of Table 6.1. Compared to other background sources (to be discussed in the next sections), we expect relatively few single top events in the pretag sample. But their

Process	Theory Cross Section (pb)	Number of events passing W selection with 1 jet	2 jets	3 jets	$\geq$3 jets	$\geq$ 4 jets
Single Top						
Wg	1.70 $\pm$ 0.26	2.2 $\pm$ 0.5	3.3 $\pm$ 0.7	1.3 $\pm$ 0.3	1.8 $\pm$ 0.4	0.3 $\pm$ 0.1
W^*	0.73 $\pm$ 0.07	0.7 $\pm$ 0.1	1.6 $\pm$ 0.3	0.5 $\pm$ 0.1	0.5 $\pm$ 0.1	0.1 $\pm$ 0.0
$t\bar{t} \rightarrow$						
$W^+bW^-\bar{b}$	5.06 $\pm$ 0.91	1.0 $\pm$ 0.2	6.7 $\pm$ 1.5	14.0 $\pm$ 3.2	31.0 $\pm$ 7.1	17.5 $\pm$ 4.0

Table 6.1: Single top and SM $t\bar{t}$ production cross sections and estimated numbers of events in the W + jets data. The errors on the number of events come from, in order of decreasing effect, errors on the theory cross sections (uncertainties in parton distribution functions, top quark mass, and scale dependence), integrated luminosity, and the statistical error on the efficiency.

inclusion is important because, to be shown in Chapter 7, they contain a top quark[1] which greatly increases their heavy-flavor tagging efficiencies relative to the other sources.

6.2 $t\bar{t}$ Events

The acceptance of $t\bar{t}$ events into the W + jets sample depends on the top quark decay mode ($t \rightarrow Wb$ or $t \rightarrow cg$) as shown in Table 5.5. Since both modes have comparable efficiencies, we assume (in this chapter) that the top quark has only SM decays in order to estimate the size of the $t\bar{t}$ contribution to the pretag sample. We use (6.1) with the theory $t\bar{t}$ production cross section [63] and W + jets efficiency for SM $t\bar{t}$ events. The estimated numbers of events are shown in the last row of Table 6.1.

6.3 Non-W Background

Physics processes and detector effects can lead to events which mimic the W signal of an isolated high-P_T lepton and missing transverse energy. The predominant type of non-W process is QCD light-quark or gluon multi-jet events, where one jet fragments such that it

[1] In these simulated single-top quark events, the top may only undergo the SM decay $t \rightarrow Wb$.

Region	Isolation	$\not\!\!E_T$
A	> 0.1	< 20
B	> 0.1	> 20
C	< 0.1	< 20
D	< 0.1	> 20

Table 6.2: Cuts defining four regions in isolation vs. $\not\!\!E_T$ space.

fakes a high-P_T electron or muon, while at least one other jet in the event is mismeasured to yield apparent missing energy. A second type of non-W process is heavy quark production (either b or c), where the semileptonic decay of the heavy quark can yield a real electron or muon while another jet in the event is mismeasured. In spite of their small event-selection probabilities, the much larger rate of QCD and heavy flavor processes compared to real W production at the Tevatron allows them to contribute to the W sample [64] [65].

We study the non-W background in the data with a technique used by CDF top analyses [42]. Figure 6.1 shows the correlation between lepton isolation (defined in Tables 5.3 and 5.4) and $\not\!\!E_T$ for electrons and muon events (§5.2.1), respectively. Each plot is divided into four regions, summarized in Table 6.2. The cluster of events with small isolation and large $\not\!\!E_T$ (Region D) constitute the W sample.

The non-W background is spread over the entire space of isolation vs. $\not\!\!E_T$, including the W signal region. In non-W events, we do not expect the isolation and $\not\!\!E_T$ to be correlated because the fluctuations by which a jet fakes a lepton are independent of those which result in jet energy mismeasurements. There is no visible correlation outside the W signal region in Figure 6.1, and this expectation has been validated [64]. Thus we assume that the ratio of the number of non-W events with high isolation to low isolation is independent of $\not\!\!E_T$:

$$\frac{N_A}{N_C} = \frac{N_B}{N_D}$$

By rearranging this relation, we estimate the number of non-W events in the W signal region:

$$N_D = N_B \cdot \frac{N_C}{N_A} \tag{6.2}$$

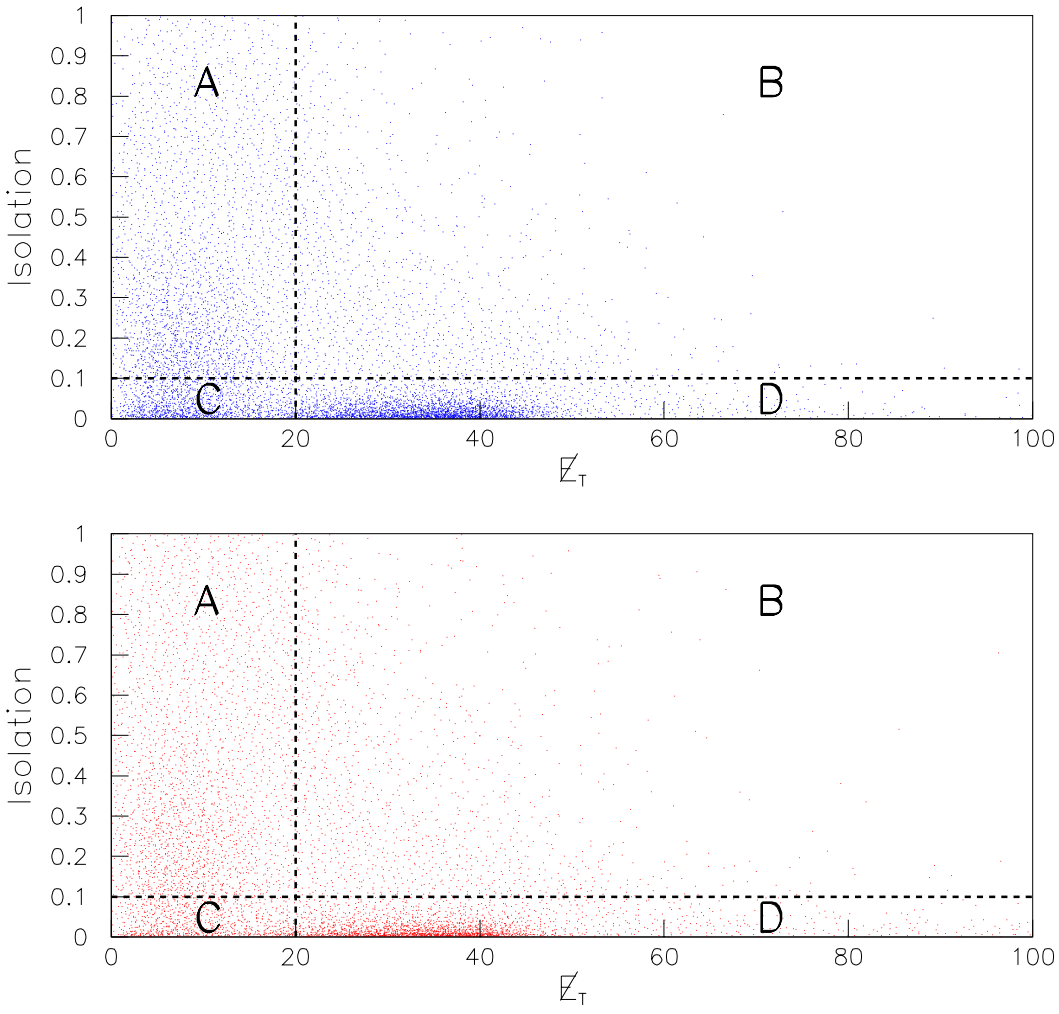


Figure 6.1: Isolation vs. $\not\!\!E_T$ for the electron (top) and muon (bottom) data samples. Regions A-D are defined by the cuts listed in Table 6.2.

Data Sample	Number of non-W events passing W selection with				
	1 jet	2 jets	3 jets	$\geq$3 jets	$\geq$ 4 jets
Electrons	367.6 $\pm$ 13.5	60.3 $\pm$ 5.3	10.9 $\pm$ 2.4	12.0 $\pm$ 2.4	1.1 $\pm$ 0.7
Muons	238.7 $\pm$ 9.2	47.9 $\pm$ 4.2	9.3 $\pm$ 2.1	11.7 $\pm$ 2.4	2.5 $\pm$ 1.3

Table 6.3: Non-W events in the W + jets data.

$$\text{and variance} \quad \sigma^2_{N_D} = \left(\frac{N_C}{N_A}\right)^2 \sigma^2_{N_B} + \left(\frac{N_B}{N_A}\right)^2 \sigma^2_{N_C} + \left(\frac{N_B N_C}{N_A^2}\right)^2 \sigma^2_{N_A}$$

We assume binomial variance for each of the event counts:

$$\sigma^2_{N_i} = N p_i (1 - p_i) = N_i (1 - N_i/N) \tag{6.3}$$

where N_i $(i = A, B, C)$ is the number of events in a region, N is the number of events in the inclusive sample (before isolation and $\not\!\!E_{\rm T}$ cuts), and $p_i = N_i/N$ is the selection probability.

We estimate the non-W background in the electron and muon samples by separately applying relation 6.2 to each[2]. We repeat this procedure for events with different jet multiplicities; the respective non-W background estimates are shown in Table 6.3.

6.4 Remaining Backgrounds

The remaining contributions to the W + jets sample have been estimated for a top-quark production cross section analysis [60]. These contributions were estimated using several methods which are briefly described here. The number of Wbb, Wcc, and diboson events were estimated directly from MC simulations. Z contributions were calculated by normalizing event fractions measured from MC to corresponding numbers of events in data control samples. After the $t\bar{t}$ cross section was measured, the remaining events were interpreted as Wc and W + non-h.f. events in a ratio measured from MC.

The background estimates were made for all of Run 1, but this analysis uses Run 1b data only. We therefore keep the relative fractions of these backgrounds but scale down

[2]Separate applications allow the possibility of a different ratio in relation 6.2 for non-W events which fake electrons than for those which fake muons.

Background	Number of events passing W selection with				
	1 jet	2 jets	3 jets	$\geq$3 jets	$\geq$ 4 jets
Dibosons	34.4 ± 5.3	28.8 ± 4.3	5.2 ± 0.8	6.3 ± 0.8	1.0 ± 0.2
$Z \to \tau\tau$	46.0 ± 2.5	15.9 ± 1.2	2.8 ± 0.4	3.6 ± 0.5	0.7 ± 0.3
Z + jets	325.6 ± 15.3	63.0 ± 7.3	11.5± 2.5	12.3 ± 2.6	0.7 ± 0.7
Wc	339.2 ±101.6	69.2 ± 20.8	9.9 ± 3.0	12.4 ± 3.1	2.5 ± 0.9
Wbb	56.7 ± 7.8	23.7 ± 4.1	5.0 ± 0.9	7.0 ± 1.1	2.0 ± 0.6
Wcc	142.1 ± 38.0	49.3 ± 10.9	10.1 ± 2.3	13.0 ± 2.6	3.0 ± 1.0
W + non-h.f.	6527.8 ±110.1	819.7 ± 25.5	106.9± 6.9	132.3± 9.9	25.4 ± 7.1

Table 6.4: Remaining backgrounds in the W + jets data.

their total for use with our data sample of less integrated luminosity. We scale them by the ratio of the numbers of W events (for each jet multiplicity) in the Run 1b data to the Run 1a+1b data. The number of W events is the W-sample size minus the non-W background (estimated directly from the data), single top and $t\bar{t}$ contributions (calculated from theory cross sections, shown in Table 6.1).

Table 6.4 shows these background estimates for each jet-multiplicity bin, where the errors are also scaled down from the cross section analysis.

6.5 W + Jets Pretag Sample Composition

We now synthesize the preceding sections to form a coherent picture of the W + jets data before tagging. The event estimates from Tables 6.1, 6.3, and 6.4 have been combined graphically in Figure 6.2. We have again assumed SM top-quark decays. The lower jet-multiplicity bins are dominated by W + non-h.f. events. The emergence of top events with higher jet multiplicities shows up in the log plot, but top is still overshadowed by the backgrounds. In our search channel of $W + \geq 3$ jets, the significance of the top signal is $S/\sqrt{B} \simeq 2.2$.

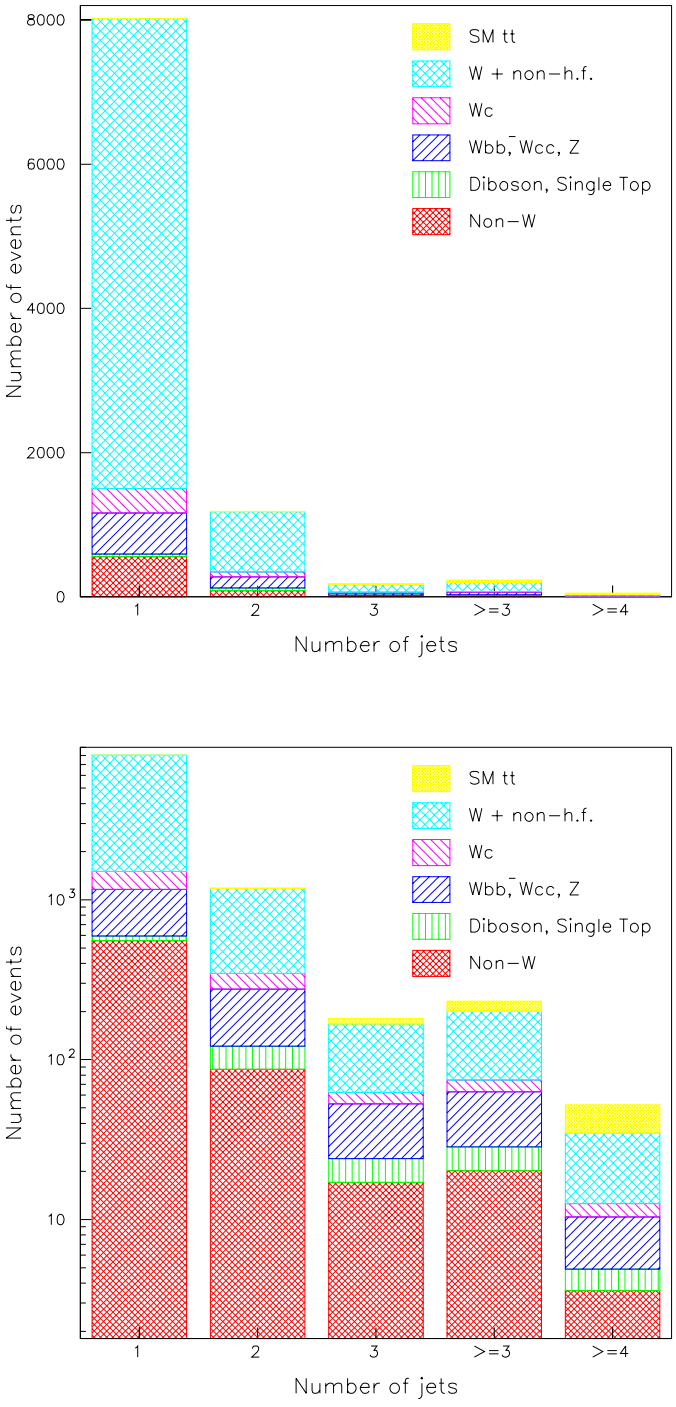


Figure 6.2: Estimated composition of the W + jets data before tagging, with linear (upper) and log (lower) vertical scales.

Chapter 7

Flavor-Tag Counting in the W + jets Sample

In Chapter 6, we saw that $t\bar{t}$ events make up a relatively small part of the W + jets data. We describe here the jet flavor tagging which we use to select events containing heavy quarks. The tagging criteria greatly enhance the top signal significance in the sample (compared to the pretag sample), allowing tagged top events to be resolved from background fluctuations.

We first motivate a jet-flavor tagger which is optimized for the high-P_T regime of jets in top events. This tagger is a modified version of the model studied in Chapter 3. Since our goal is to distinguish SM- from FCNC-decays of the top quark, we try different combinations of b- and c-jet tagging criteria in order to discriminate the two types of events. Next we estimate the expected numbers of tags from $t\bar{t}$ signal and background events by applying tagging rates measured in the MC samples of § 5.2.2 to the pretag composition given in Chapter 6. The expected numbers of tags, measured in MC, are converted to their data equivalents using the scale factor technique of Chapter 4. Finally, we compare the predicted and observed numbers of tags from $t\bar{t}$ and background events in the $W+ \geq 3$ jets sample. Tag contributions from $t\bar{t}$ events with both SM and FCNC decays are compared, but no conclusion about the consistency of either with the data is made. Instead, these tagging efficiencies and background estimates will serve as inputs to the statistical upper limit set in Chapter 8.

7.1 NN Tagger

7.1.1 Training

The NN tagger described in Chapter 3 was trained using MC jets in direct production $b\bar{b}$, $c\bar{c}$, and QCD light quark/gluon events. We initially tried using this tagger for this analysis, but discovered that the identification efficiencies in top events were prohibitively small. This is because the direct production training samples contain just a relatively small number of jets in the high-P_T range characteristic of top events[1]. The direct production training samples of limited size did not contain sufficient diversity to successfully generalize to jets in top events.

To resolve this problem, we instead trained the NN using jets from a sample of $t\bar{t}$ events with mixed SM-FCNC decays. By training on jets in events in which we are ultimately interested, we have effectively sped up the training process. These particular jets populate unique regions of the jet-variable space which would only be adequately represented after a very long MC generation time of direct production events[2].

To further tune the NN to the target application, training jets were required to pass the TOPFND cuts of § 5.3.3, and their events were required to pass the $W + \geq 3$ jets selection. Note that separate samples of SM-FCNC $t\bar{t}$ events were used for NN training than for the rest of the analysis.

7.1.2 Taggable-Jet Criteria

In addition to the TOPFND jet cuts, the taggable-jet definition includes several choices for decay track cuts. As in Chapter 3, we require positive impact parameter decay tracks which pass *good_svx* quality cuts. We did not know, *a priori*, what taggable-jet selection was optimal for setting a branching-fraction limit. We studied ten different taggable-jet defini-

[1] With (unreasonably) large statistics in the direct production samples, the features of energetic b and c jets from SM and FCNC top quark decays would be present for training.

[2] But there is always a tradeoff: a SM-FCNC trained NN is worse at identifying b and c jets in non-top events, but we want to exclude those events anyway.

Identifier	Decay Tracks in Jet # Required	P_{track}	Additional Requirements	# NN Inputs
1a	≥ 1	< 1		10
1b	≥ 1	< 1	≥ 1 track w/ $P_{\text{track}} < 0.1$	10
2a	≥ 2	< 1		14
2b	≥ 2	< 0.5		14
2c	≥ 2	< 1	≥ 1 track w/ $P_{\text{track}} < 0.1$	14
2d	≥ 2	< 0.3		14
2e	≥ 2	< 0.2		14
2f	≥ 2	< 0.6		14
2g	≥ 2	< 0.7		14
2h	≥ 2	< 0.1		14

Table 7.1: Various definitions of taggable jets and decay tracks. In addition to the criteria here, jets were required to pass TOPFND cuts and tracks were required to have positive impact parameter and pass *good_svx* quality cuts. Events containing these jets were required to pass the $W + \geq 3$ jets selection.

tions, summarized in Table 7.1. Note that for taggable jets which required only one track, we used just 10 of the 14 available jet variables[3]. For each taggable-jet definition, a separate NN tagger was trained and used in the subsequent analysis.

The discussion in the rest of this chapter pertains to NN tagger "2e" which, as described in Chapter 8, yields the best (lowest) expected branching-fraction upper limit.

7.1.3 Scale Factors

Using the method of Chapter 4, data-to-MC scale factors for jet-tagging efficiencies were measured separately for each NN tagger. The scale factors for tagger "2e" are shown in Table 7.2. For all taggers, most fitted scale factors were $\mathcal{O}(1) \pm 0.2$ and the fits had $\chi^2/\text{d.o.f.} \lesssim 1.0$.

The tagging efficiency parameterization adopted for the scale factor fit defines the

[3]The following jet variables are ill-defined for one-track jets: M_{trks}, ΣP_T^{rel}, Ψ, and $L_{xy}/\sigma_{L_{xy}}$.

scale factors as the set $\{s_{ij}\}$:

$$s_{ij} = \begin{cases} \frac{\epsilon_{it}^{data}}{\epsilon_{it}^{MC}} & j = t \text{ (for jet to be taggable)} \\ \frac{\epsilon_{ij}^{data,taggable}}{\epsilon_{ij}^{MC,taggable}} & j = b, c \text{ (for taggable jet to be identified as } b, c) \end{cases}$$

$$i = \text{true jet flavor} = b, c, p$$

Note that these scale factors are sample-independent. However, in order to count tagged jets, we will need a set of modified scale factors $\{s'_{ij}\}$ which are simply the ratios of tagging efficiencies in the data to MC[4]:

$$s'_{ij} = \frac{\epsilon_{ij}^{data}}{\epsilon_{ij}^{MC}}\ , \quad i = \text{true jet flavor} = b, c, p$$
$$j = \text{identified flavor} = b, c, p, \text{ or } u \text{ for untaggable}$$

The tagging efficiencies in the data are related to those in MC and the scale factors $\{s_{ij}\}$ through (4.5). Dividing (4.5) by the MC-tagging efficiencies gives the modified scale factors:

$$\begin{aligned} s'_{ib} &= s_{it} s_{ib} \\ s'_{ic} &= s_{it} s_{ic} \\ s'_{ip} &= s_{it} \left(1 - \epsilon_{iu}^{MC} - s_{ib}\epsilon_{ib}^{MC} - s_{ic}\epsilon_{ic}^{MC}\right) / \epsilon_{ip}^{MC} \\ &= s_{it} \left(\epsilon_{ib}^{MC}(1 - s_{ib}) + \epsilon_{ic}^{MC}(1 - s_{ic}) + \epsilon_{ip}^{MC}\right) / \epsilon_{ip}^{MC} \\ s'_{iu} &= \left[1 - s_{it}\left(1 - \epsilon_{iu}^{MC}\right)\right] / \epsilon_{iu}^{MC} \\ &= \left[1 - s_{it}\left(\epsilon_{ib}^{MC} + \epsilon_{ic}^{MC} + \epsilon_{ip}^{MC}\right)\right] / \left(1 - \epsilon_{ib}^{MC} - \epsilon_{ic}^{MC} - \epsilon_{ip}^{MC}\right) \end{aligned} \tag{7.1}$$

where we have used the identity

$$1 - \epsilon_{iu}^{MC} = \epsilon_{it}^{MC} = \epsilon_{ib}^{MC} + \epsilon_{ic}^{MC} + \epsilon_{ip}^{MC}$$

to minimize the number of variables in the transformation. Note that the last two modified scale factors depend on certain MC-tagging efficiencies, so they are sample-dependent.

[4]The modified scale factors $\{s'\}$ have a simpler definition than the original scale factors $\{s\}$ defined in Chapter 4. Either set can be measured using the fit technique of Chapter 4. We found that we could measure with significantly greater precision the original set $\{s\}$, which uses a more physical parameterization, than the modified set $\{s'\}$ needed here.

True Flavor	Scale Factor Types		
	Taggable	If Taggable, Tagged as	
		b	c
b	$s_{bt} = 0.81 \pm 0.04$	$s_{bb} = 0.96 \pm 0.03$	$s_{bc} = 0.99 \pm 0.14$
c	$s_{ct} = 1.14 \pm 0.15$	$s_{cb} = 0.83 \pm 0.18$	$s_{cc} = 0.75 \pm 0.10$
p	$s_{pt} = 0.87 \pm 0.06$	$s_{pb} = 2.42 \pm 0.79$	$s_{pc} = 0.86 \pm 0.16$

Table 7.2: Data-to-MC jet-tagging scale factors as returned from the scale factor fit (χ^2/d.o.f. = 0.66) for NN tagger "2e".

Therefore, the modified scale factors are calculated separately for each MC sample using 1) the sample independent scale factors $\{s_{ij}\}$ and 2) tagging efficiencies $\{\epsilon_{ij}^{MC}\}$ measured in events passing $W + \geq 3$ jets cuts in that MC sample.

The covariance matrix $\{\sigma_{s'_{ij}s'_{kl}}\}$ of the modified scale factors will be needed when counting tagged MC events in § 7.3. This covariance matrix is related to the original covariance matrix $\{\sigma_{s_{ij}s_{kl}}\}$ (obtained from the method of Chapter 4) and the covariance matrix of MC-tagging efficiencies $\{\sigma_{\epsilon_{ij}}\sigma_{\epsilon_{kl}}\}$ through Equations 7.1. The calculational details necessary for this transformation are found in Appendix E.

7.1.4 Tagging Efficiencies For Top Events

Table 7.3 shows the jet-tagging efficiencies expected in data (i.e. including scale factors) for $t\bar{t}$ events with both SM and SM-FCNC decays. The uncertainties include both the statistical error on the MC efficiencies and the errors on the scale factors described in the previous section. The correct b-tag efficiencies are 16-18%, and the correct c-tag efficiencies are 5-6%. These efficiencies seem surprisingly low, but this tagger was selected because its signal-to-background ratio admits the best branching-fraction limit[5].

[5]Other taggers with more relaxed taggable-jet criteria in Table 7.1 achieve significantly greater tagging efficiencies and better b/c discrimination.

$t\bar{t} \to WbWb$ events				
True	Tagging Outcome (%)			
Flavor	b	c	p	untaggable
b	19.3 ± 1.2	3.8 ± 0.6	2.7 ± 0.9	74.3 ± 1.5
c	3.4 ± 1.3	5.3 ± 1.3	6.6 ± 1.6	84.7 ± 2.9
p	0.5 ± 0.3	0.5 ± 0.1	2.0 ± 0.4	97.0 ± 0.4
$t\bar{t} \to Wbcg$ events				
True	Tagging Outcome (%)			
Flavor	b	c	p	untaggable
b	16.0 ± 1.2	3.9 ± 0.7	2.9 ± 0.9	77.2 ± 1.5
c	2.8 ± 0.9	5.2 ± 1.0	6.5 ± 1.5	85.5 ± 2.4
p	0.7 ± 0.4	0.4 ± 0.2	2.9 ± 0.6	95.9 ± 0.6

Table 7.3: Jet-tagging efficiencies (%) expected in data for a NN tagger "2e" in $t\bar{t}$ events with SM (upper) and SM-FCNC (lower) decays. The errors include the statistical errors on the scale factors, MC efficiencies, and their correlations; see the note at the end of Appendix E for more details on this error calculation.

7.2 Criteria For Flavor-Tagging Events

We separately count events containing b and c quarks to increase the $t\bar{t}$ signal-to-background ratio. We now define the two *event tag* types considered in this analysis:

b-tag An event which contains at least one b-tagged jet.

c-tag An event which contains at least one c-tagged jet which, in combination with another jet in the event, forms a dijet mass within a specified range centered on the top quark mass (see Figure 7.1). This additional dijet mass cut is meant to select $t \to cg$ decays where the two-body decay of the top quark is fully reconstructible, unlike SM top decays. The second jet in the dijet pair is selected as that jet in the event which, when combined with the c-tagged jet, forms a dijet mass closest to the top quark mass. Several dijet mass range widths were tried: 50, 80, 110, and 140 GeV/c^2. Lower widths select $t \to cg$ decays with higher purity; for example, the range width of 50 GeV/c^2 rejects about 75% more SM-SM than SM-FCNC events.

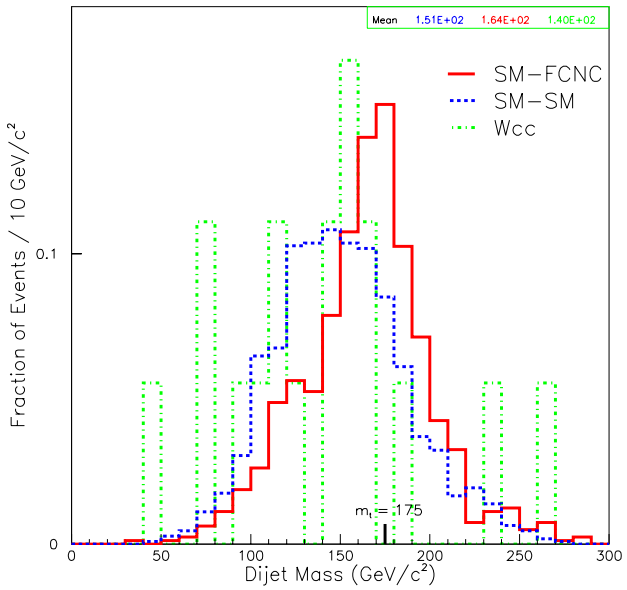


Figure 7.1: Dijet (including c-tagged jet) mass for SM-SM and SM-FCNC decay $t\bar{t}$ events and a typical background, $Wc\bar{c}$. Events pass the $W + \geq 3$ jet selection cuts.

***bc*-tag** An event which contains at least one b-tagged jet *and* at least one c-tagged jet. There is no dijet mass cut on the c-tagged jet.

7.3 Counting Tagged Events

7.3.1 Weighting MC Event Counts Using Tagging Scale Factors

A data event with a certain event-tag outcome contributes exactly one event to the corresponding sum of tagged events. The situation is different if we wish to use MC events to estimate the expected number of tagged events in the data because jet-tagging efficiencies differ between data and MC. In general, a tagged MC event contributes a fraction of an event to the corresponding sum; the derivation is shown below.

Consider an event (in a certain MC sample) with N_{jet} jets of true flavors $f_1, \ldots, f_{N_{jet}}$ ($f_i = b, c, p$) and jet-tagging outcomes $t_1, \ldots, t_{N_{jet}}$ ($t_i = b$-, c-, p-tagged, or untaggable). The total number of MC events with N_{jet} jets with true flavors $\{f_i\}$ and tag outcomes $\{t_i\}$ can

be expressed as:

$$N^{MC}_{\{f_i\},\{t_i\}} = N_{\{f_i\}}\, \epsilon^{MC}_{f_1 t_1} \cdots \epsilon^{MC}_{f_{N_{jet}} t_{N_{jet}}} = N_{\{f_i\}} \prod_{i=1}^{N_{jet}} \epsilon^{MC}_{f_i t_i}$$

where $\epsilon^{MC}_{f_i t_i}$ is the MC efficiency for a jet of true flavor f_i to be tagged as t_i

and $N_{\{f_i\}}$ is the number of events with N_{jet} jets with true flavors $f_1, \ldots, f_{N_{jet}}$

But in the data, the jet-tagging efficiencies, and therefore the expected number of corresponding events, are modified by the jet-tagging scale factors $\{s'_{ij}\}$:

$$\begin{aligned} N^{data}_{\{f_i\},\{t_i\}} &= N_{\{f_i\}} \prod_{i=1}^{N_{jet}} \epsilon^{data}_{f_i t_i} = N_{\{f_i\}} \prod_{i=1}^{N_{jet}} s'_{f_i t_i} \epsilon^{MC}_{f_i t_i} \\ &= \left(\prod_{i=1}^{N_{jet}} s'_{f_i t_i} \right) N_{\{f_i\}} \prod_{i=1}^{N_{jet}} \epsilon^{MC}_{f_i t_i} = \left(\prod_{i=1}^{N_{jet}} s'_{f_i t_i} \right) N^{MC}_{\{f_i\},\{t_i\}} \end{aligned}$$

The result is that the number of expected data events with a specific combination of jet flavors and tags is modified from the number of MC events by the product of jet-tagging scale factors for that combination. Because this holds for any such combination, the result is true in general: each MC event contributes to the data expectation a (real) number of events given by the product of jet-tagging scale factors for that MC event. Thus the number of tagged events expected in the data is a sum over tagged MC events of modified scale factor products for each event:

$$N^{data}_{tag} = \sum_{j=1}^{N^{MC}_{tag}} \left(\prod_{i=1}^{N^{j}_{jet}} s'_{f_i t_i} \right) \tag{7.2}$$

where N^{MC}_{tag} is the number of tagged MC events in the sample

and N^{j}_{jet} is the number of jets in each tagged MC event

The error on the number of tagged events expected in data ($\sigma_{N^{data}_{tag}}$) comes from two independent sources, the modified scale factors and finite MC statistics:

$$\sigma^2_{N^{data}_{tag}} = \left(\sigma_{N^{data}_{tag}, s'} \right)^2 + \left(\sigma_{N^{data}_{tag}, N^{MC}_{tag}} \right)^2$$

The first contribution is obtained by propagating the modified scale factor errors (including their correlations) through (7.2); calculational details are given in Appendix F. The second is calculated by:

$$\left(\sigma_{N_{tag}^{data},N_{tag}^{MC}}\right)^2 \;=\; \left(\frac{\partial N_{tag}^{data}}{\partial N_{tag}^{MC}}\right)^2 \cdot \sigma^2_{N_{tag}^{MC}} \;\simeq\; \langle s' \rangle^2 \cdot N_{tag}^{MC}(1 - N_{tag}^{MC}/N^{MC})$$

where $\frac{\partial N_{tag}^{data}}{\partial N_{tag}^{MC}} \simeq \frac{N_{tag}^{data}}{N_{tag}^{MC}} = \langle s' \rangle$ is the average value of the product of modified scale factors over tagged MC events and $\sigma^2_{N_{tag}^{MC}}$ is the binomial variance of the number of tagged MC events.

7.3.2 Tag Counting For the Non-W Background

The method of estimating the non-W background before tagging was described in § 6.3. Ideally, one would like to use *tagged* event counts in (6.2) to estimate the tagged non-W background. However, the method breaks down (both the estimate and error become undefined) in the low statistics of tagged non-W data events.

To estimate the tagged non-W background N_D^{tag}, we instead begin with the pretag estimate N_D of § 6.3 and apply an effective non-W tagging efficiency ϵ_{tag}:

$$N_D^{tag} \;=\; N_D \cdot \epsilon_{tag} \;=\; \frac{N_B\, N_C}{N_A} \cdot \epsilon_{tag} \tag{7.3}$$

We measure this efficiency (separately for b- and c-tagging) in the data outside of the W signal region (i.e. regions A,B,C):[6]

$$\epsilon_{tag} \;=\; \frac{N_A^{tag} + N_B^{tag} + N_C^{tag}}{N_A + N_B + N_C} \tag{7.4}$$

so that 7.3 becomes:

$$N_D^{tag} \;=\; \frac{N_B\, N_C}{N_A} \cdot \frac{N_A^{tag} + N_B^{tag} + N_C^{tag}}{N_A + N_B + N_C} \;=\; f(N_A, N_B, N_C) \cdot N_{ABC}^{tag} \tag{7.5}$$

[6]Since both the numerator *and* the denominator of (7.4) have binomial variance, there is no simple expression for the error on this effective efficiency; its need is obviated by regrouping terms in 7.3 as 7.5 and taking an upper limit on the resulting error.

$$
\begin{aligned}
\text{where} \quad f(N_A, N_B, N_C) &= \tfrac{N_B\, N_C}{N_A\,(N_A+N_B+N_C)} \\
\text{with variance} \quad \sigma_f^2 &= \left(\tfrac{N_B N_C (2N_A+N_B+N_C)}{(N_A(N_A+N_B+N_C))^2}\right)^2 \sigma^2_{N_A} + \\
&\quad \left(\tfrac{N_C(N_A+N_C)}{N_A(N_A+N_B+N_C)^2}\right)^2 \sigma^2_{N_B} + \\
&\quad \left(\tfrac{N_B(N_A+N_B)}{N_A(N_A+N_B+N_C)^2}\right)^2 \sigma^2_{N_C} \\
\text{and} \quad N^{tag}_{ABC} &= N^{tag}_A + N^{tag}_B + N^{tag}_C
\end{aligned}
$$

The event counts $N_{A,B,C}$ and N^{tag}_{ABC} have binomial variances of the form 6.3. The error on the number of tagged non-W events is then given by:[7]

$$\sigma_{N^{tag}_D} \;=\; N^{tag}_{ABC}\,\sigma_f + f\,\sigma_{N^{tag}_{ABC}}$$

Note that in this procedure, we have applied an average efficiency (over the non-W regions) to the W signal region. However, the tagging efficiency may depend on the isolation and $\not\!\!E_{\rm T}$ of the event. We believe that the error $\sigma_{N^{tag}_D}$ 1) brackets any such systematic effects, and 2) is small compared to the statistical error associated with a direct application of (6.2) with low statistics. An additional check of this method is described in Appendix I.

7.4 Tag Table

In Tables 6.1, 6.3, and 6.4 we listed the expected composition of the W + jets pretag sample. We are now ready to update those predictions to include the effects of tagging. Table 7.4 shows tagging statistics for $W + \geq 3$ jets events. Note that here we use NN tagger "2e" which requires taggable jets to have at least two tracks with $P_{\rm track} < 20\%$, and a c-tag dijet mass width of 140 GeV (this choice is justified in Chapter 8). Similar statistics for each NN tagger listed in Table 7.1 are found in Appendix D.

7.4.1 Organization and Description of Entries

Table 7.4 shows the pretag and event-tagging statistics of $W + \geq 3$ jets events for backgrounds, $t\bar{t}$ signal, and data. The table is organized as follows. Each event sample has

[7] Here we have used the *Schwarz inequality* $|\sigma_{f\,N^{tag}_{ABC}}| \leq \sigma_f\,\sigma_{N^{tag}_{ABC}}$ because we have insufficient statistics to measure the covariance between f and N^{tag}_{ABC}. We take the upper bound as the error.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2438	0.43	0.0147	0.03	0.0322	0.06
	±0.37	±0.0172	±0.09	±0.0035	±0.01	±0.0061	±0.02
Single top (W^*)	0.53	0.2853	0.15	0.0333	0.02	0.0677	0.04
	±0.09	±0.0222	±0.03	±0.0073	±0.00	±0.0124	±0.01
Dibosons	6.27	0.0274	0.17	0.0000	0.00	0.0000	0.00
	±0.82	±0.0163	±0.10	±0.0000	±0.00	±0.0000	±0.00
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0145	0.18	0.0000	0.00	0.0028	0.03
	±2.58	±0.0079	±0.10	±0.0000	±0.00	±0.0028	±0.04
Wc	12.38	0.0197	0.24	0.0000	0.00	0.0171	0.21
	±3.11	±0.0112	±0.15	±0.0000	±0.00	±0.0086	±0.12
Wbb	6.99	0.2719	1.90	0.0185	0.13	0.0324	0.23
	±1.11	±0.0323	±0.38	±0.0086	±0.06	±0.0128	±0.10
Wcc	13.04	0.0813	1.06	0.0043	0.06	0.0638	0.83
	±2.55	±0.0299	±0.44	±0.0044	±0.06	±0.0220	±0.33
W + non-h.f.	132.32	0.0000	0.00	0.0000	0.00	0.0031	0.41
	±9.89	±0.0000	±0.00	±0.0000	±0.00	±0.0032	±0.42
Non-W e	12.03	0.0756	0.91	0.0000	0.00	0.0025	0.03
	±2.43		±0.35		±0.00		±0.04
Non-W μ	11.72	0.0572	0.67	0.0000	0.00	0.0162	0.19
	±2.36		±0.28		±0.00		±0.11
$\sum$ Backgrounds	212.86		5.72		0.23		2.03
			±1.36		±0.63		±1.03
SM-SM $t\bar{t}$	31.00	0.2951	9.15	0.0287	0.89	0.0744	2.31
	±7.07	±0.0148	±2.14	±0.0030	±0.22	±0.0071	±0.57
SM-FCNC $t\bar{t}$		0.1847		0.0177		0.0842	
		±0.0095		±0.0025		±0.0072	
$\sum$ Backgrounds	243.86		14.87		1.12		4.34
+ SM-SM $t\bar{t}$			±2.54		±0.67		±1.18
Data	244.00		20.00		1.00		5.00
Tag excess			5.13		-0.12		0.66
			±4.62		±1.25		±2.39

Table 7.4: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "2e", where a taggable jet requires ≥ 2 positive i.p. decay tracks, each with track probability $< 20\%$. The dijet mass range associated with the c-tag is 140 GeV.

two rows of information. The top row shows the number of events before tagging (pretag), and the event-tagging efficiencies ($\epsilon_{b-,bc-,c-\mathrm{tag}}$) and numbers of tagged events for b-,bc-,and c-event tagging. The bottom row gives the errors on these quantities.

We now describe the calculation of each entry in the table, beginning with event samples at the top and moving down the rows of the table.

The pretag calculation for single top background events is described in § 6.1, and the pretag numbers for the W/Z backgrounds are described in § 6.4. The event-tagging efficiencies were measured separately for each of these backgrounds in the following way. For each background, we used a MC sample of events passing the $W + \geq 3$ jets selection and the scale-factor weighting technique described in § 7.3.1 to count the numbers of tagged events (and errors) expected in a data sample of equal size. Dividing these numbers (and errors) by the MC sample size yields the tagging efficiencies shown in the table. Each number of tags is the product of the tagging efficiency and the pretag event count, and the error on the number of tags comes from the errors on these two factors.

The pretag calculation for the non-W backgrounds is described in § 6.3. The effective event-tagging efficiencies and estimated numbers of tags are discussed in § 7.3.2.

The background sum row gives the sums of background events in each column. The pretag sum has no error because the overall pretag background was normalized to the expected number of non-$t\bar{t}$ pretag events in the data as described in § 6.4. The error on the tag sums includes the effects of correlations between the event-tagging efficiencies between background samples and non-W events, and is described in Appendix H.

The next two rows show information for $t\bar{t}$ events with both SM-SM and SM-FCNC decays. The SM-SM $t\bar{t}$ pretag calculation, which assumes $\mathrm{BR}(t \to Wb) = 1$, is described in § 6.2. No pretag count for SM-FCNC events is given since it depends on $\mathrm{BR}(t \to cg)$. Event-tagging efficiencies expected in the data were measured from weighted counts of tagged events in separate $W + \geq 3$ jet MC samples (the same method described above for the single top and W/Z backgrounds).

The final three rows are grouped together for comparison. The top row gives the

sums of background tags and SM-SM $t\bar{t}$ tags (each error is the quadrature sum of the errors on these terms). These sums are to be compared with the actual event counts observed in the data listed in the middle row. Finally, the tag excesses listed in the bottom row are the data $-\sum$(backgrounds + SM-SM $t\bar{t}$) differences for each tag type. When we compare the tag sums to the data, however, there is a subtlety associated with what quantity we use as the error on the sum. We first note that each sum S (or any other number of tags given in the table) is really the expected mean of a Poisson distribution. The error on each sum reported in the table, σ_S, is the error on each Poisson mean. But the variance of the *observed* numbers of tags has an additional statistical component, $\sigma^2_{S_{fluc}} = S$, from the natural Poisson fluctuations about the true mean. These are independent effects so we add them in quadrature to get the variance of the *observed* sum, which is also the reported variance of the tag excess.

The errors on the tag excesses are simply the errors on each tag sum since the data is fixed. When we compare to our observations in the data, however, there is a subtlety associated with what quantity we use as the error on the sum. We first note that each sum S (or any other expected number of tags given in the table) is really the expected mean of a Poisson distribution. The error on each sum reported in the table, σ_S, is the error on each Poisson mean. But the variance of the *observed* numbers of tags has an additional statistical component, $\sigma^2_{S_{fluc}} = S$, from the natural Poisson fluctuations about the true mean. These are independent effects so we add them in quadrature to get the variance of the *observed* sum, which is also the reported variance of the tag excess.

7.4.2 Background

We expect the most background b-tags to come from $Wb\bar{b}$,$Wc\bar{c}$ events because of their higher b-tag efficiencies and pretag statistics. In general, the former contributes events with correctly b-tagged b jets (with a higher efficiency), while the latter contributes incorrectly b-tagged c jets (with a lower efficiency). The next highest contributions come from non-W events, followed by Wc and W + non-h.f. events. Although the latter source has a

b-tag efficiency of only $\frac{1}{2}\%$ (a pure mistag efficiency), its large pretag presence results in a significant tag contribution.

We anticipate the most background c-tags to also come from $Wb\bar{b}$,$Wc\bar{c}$ events. Here we expect the former to contribute events with incorrectly c-tagged b jets (with a lower efficiency), and correctly c-tagged c jets from the latter. Note that all c-tag efficiencies are significantly less than those for b-tagging, an expected result of the inherently lower c-tagging vs. b-tagging efficiency of the NN tagger and the additional dijet mass cut. All other backgrounds have much smaller c-tag efficiencies.

We expect very few bc-tags (relative to the other tags) from any background, although W + h.f. events again contribute the most.

7.4.3 $t\bar{t}$ Signal

The b-tag efficiency is significantly greater in SM-SM than SM-FCNC events because the former events contain two b jets while the latter contain just one.

The c-tag efficiencies for both event types are the same within error. Although counterintuitive to our NN training on SM-FCNC events, several factors contribute to this result. First, SM-SM event have two b jets, each of which can be mistagged as c with an efficiency comparable to the efficiency to correctly c-tag a c jet (see Table 7.3). The same misidentification can occur in SM-FCNC events, but there is only one opportunity with the single b jet. Second, both types of events have comparable correct c-tagging efficiencies. Both factors (similar c-jet c-tag efficiencies and comparable mistag efficiencies) are results of the strict track probability cut (20%) of this NN tagger. Additionally, the dijet mass range required by the c-tag is sufficiently large to accept with high efficiency both SM-SM and SM-FCNC events.

Each bc-tag efficiency is approximately the product of the respective b- and c-tag efficiencies. This efficiency is again higher in SM-SM than SM-FCNC events, presumably because the b-tag efficiency is higher in the former while the c-tag efficiencies are comparable, for the reasons given above.

7.4.4 Data

We observe approximately a 1-σ excess of b-tags in the data over the sum of background and SM-SM $t\bar{t}$ tags, while the numbers of bc- and c-tags agree well.

Chapter 8

Branching Fraction Upper Limits

The previous chapter culminated with Table 7.4, which showed overall good agreement in the numbers of tagged events between the data and the predicted sum of backgrounds and SM-SM $t\bar{t}$ events. We also saw that in $t\bar{t}$ events, the event-tagging efficiencies depend on the top-quark decay modes. Therefore the expected numbers of tagged $t\bar{t}$ events depend on the branching fraction $B(t \to cg)$.

In the first section of this chapter, we describe a maximum-likelihood fit used to measure the branching fraction value that is most compatible with the observed numbers of tagged events in the data. This measured branching fraction does not differ significantly from the SM prediction of zero. Then we describe how we derive an upper limit on the branching fraction from this measured value. Both the measured branching fraction and upper limit are shown to be reasonable compared to their values expected from a statistical simulation of pseudo-experiments. The choice of tagger used throughout this analysis is justified as that which yields the lowest expected limit, as shown in Appendix J. In the last section of this chapter, we apply the same techniques to a different FCNC top-quark decay to set an upper limit on $B(t \to cZ)$.

8.1 Branching Fraction $B(t \to cg)$ Fit

8.1.1 Tags From $t\bar{t}$ Events

We begin with the expected number of $t\bar{t}$ events produced in the data, given by the product of the theory cross section (Table 6.1) and Run 1b integrated luminosity (§ 5.2.1):

$$N^{t\bar{t}} \;=\; \sigma(t\bar{t}) \cdot \int \mathcal{L}\; dt \;=\; 436.9 \pm 80.6 \text{ events} \tag{8.1}$$

where the error comes mostly from the uncertainty in the cross section and also from the independent uncertainty in the integrated luminosity.

Next we assume that the top quark has two possible decay modes, with branching fractions related by $B \equiv B(t \to cg) = 1 - B(t \to Wb)$. Then the number of $t\bar{t}$ events with SM-SM decays is $N^{t\bar{t}}(1-B)^2$ and the number with SM-FCNC decays is $2N^{t\bar{t}}B(1-B)$.

The efficiency for a $t\bar{t}$ event to pass the $W + \geq 3$ jets selection, $\epsilon_{W\text{j}}$, depends on the top-quark decays. These efficiencies were measured in MC:

$$\begin{aligned} \epsilon_{W\text{j}}^{\text{SM-SM}} &= 0.0753 \pm 0.0008 \\ \epsilon_{W\text{j}}^{\text{SM-FCNC}} &= 0.0543 \pm 0.0006 \end{aligned} \tag{8.2}$$

where the errors are statistical[1]. The event b-,bc-, and c-tagging efficiencies for $t\bar{t}$ events passing $W + \geq 3$ jets cuts, $\epsilon_{W\text{j},\{b,bc,c\}}$, also depend on the top-quark decays as shown in Table 7.4. The net event-tagging efficiency for $t\bar{t}$ events expected in the data (for each tag type, for both SM-SM and SM-FCNC decays) is given by the product:

$$\epsilon^i_{b,bc,c} \;=\; s_{tl} \;\cdot\; \epsilon^i_{W\text{j}} \;\cdot\; \epsilon^i_{W\text{j},\{b,bc,c\}} \quad , \; i = \text{SM-SM , SM-FCNC} \tag{8.3}$$

where s_{tl} is the data-to-MC scale factor for trigger and lepton-ID efficiencies discussed in § 5.4. Because we have completely factorized (8.3) into nine independent factors, the correlations between the six different efficiencies it represents are obvious.

[1]Note that, unlike the efficiencies listed in Table 5.5, these efficiencies are MC-only and do not include the scale factor for trigger and lepton-ID efficiencies, s_{tl}.

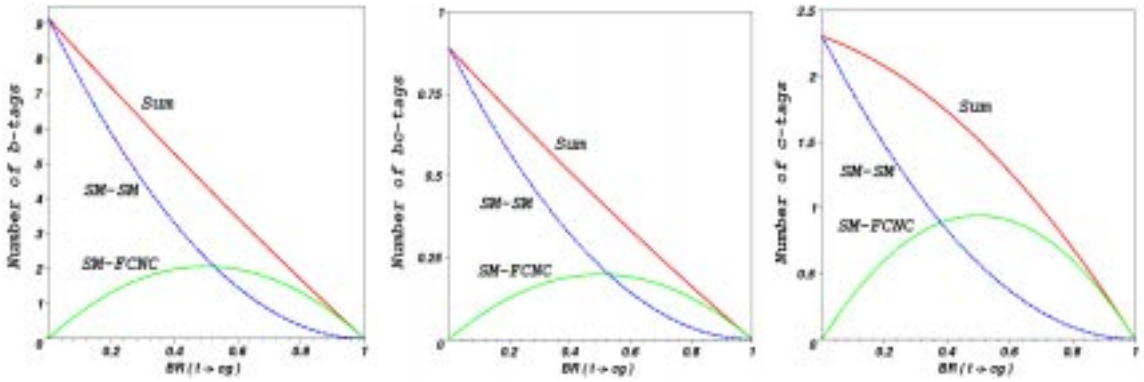


Figure 8.1: Numbers of b-,bc-, and c-tagged $t\bar{t}$ events versus $B(t \to cg)$. Contributions from SM-SM and SM-FCNC top-quark decay modes are shown separately beneath their sum.

The expected number of tagged $t\bar{t}$ events is then given as the following sum of tags from SM-SM and SM-FCNC events,

$$N^{t\bar{t}}_{b,bc,c}(B) \;=\; N^{t\bar{t}} \left[(1-B)^2 \epsilon^{\text{SM-SM}}_{b,bc,c} + 2B(1-B)\epsilon^{\text{SM-FCNC}}_{b,bc,c}\right] \tag{8.4}$$

where each term is the product of the number of $t\bar{t}$ events with a given decay mode, which depends quadratically on B, and the event-tagging efficiency for those events. Note that this number of events is really a Poisson mean, and actual observations of the number of tagged $t\bar{t}$ events are distributed about this mean.

Figure 8.1 shows the numbers of tagged $t\bar{t}$ events, separately for each decay mode, as a function of B in the physical range $[0, 1]$. For each tag type, as B increases from zero, the number of SM-SM tags decreases from its maximum (the SM-SM tag counts reported in Table 7.4), while the number of SM-FCNC tags increases from zero to its maximum. The numbers of tags from both modes necessarily vanish as $B \to 1$ since the $t\bar{t}$ pairs are lost to FCNC-FCNC decays to which we are not sensitive[2].

If the additional tags from background processes were negligible, we could measure B by comparing the numbers of observed tags in data with those predicted by the sum curves in Figure 8.1. Then the error on the measured B is inversely proportional to the slope of

[2]These $t\bar{t} \to cg\bar{c}g$ events fail the lepton trigger and W selection.

the sum curve at the measured value of B. Differentiating (8.4), we have

$$\left|\frac{dN^{t\bar{t}}_{b,bc,c}}{dB}\right|_{B=0} \propto \left|\epsilon^{\text{SM-SM}}_{b,bc,c} - \epsilon^{\text{SM-FCNC}}_{b,bc,c}\right| \tag{8.5}$$

which shows that, for small B (which is our region of interest), the precision of the B measurement increases with the difference of event-tagging efficiencies of the two decay modes. If this were our measurement technique, then we would have chosen one of several other taggers from Table 7.1 which have greater tagging efficiency differences in (8.5) than does our tagger choice "2e". As discussed in the next section, we must also contend with background tags. Our analysis strategy, and therefore tagger choice, is optimized for setting a limit instead of measuring B.

8.1.2 Tags From $t\bar{t}$ and Background Events Compared to Data

For each tag type, we can add (8.4) to the expected total number of background tags (independent of B) from Table 7.4 to get the expected total number of tags as a function of B, as shown in Figure 8.2. The figure extends the branching fraction beyond the physical range to show more completely the parabolic shape of each tag expectation curve. Note that the concavity of each tag expectation curve depends on the sign of the difference of the efficiencies in (8.4); b- and bc-tags are concave-up, and c-tags are concave-down. The numbers of each tag observed in the data are indicated by horizontal lines in the figure. Clearly there is no value of B for which the expected and observed numbers of tags agree simultaneously for all tag types. Since the expected numbers of tags are really Poisson means, we anticipate the data to fluctuate about these means instead of exact agreement. We find the value of best agreement, consistent with these fluctuations, in the next section.

8.1.3 Likelihood Maximization

We use the method of maximum likelihood to measure the value of B which is most compatible with the observed data. We define the likelihood function $\mathcal{L}$ as the joint probability of observing the numbers of tagged events in the data, with the number of $t\bar{t}$

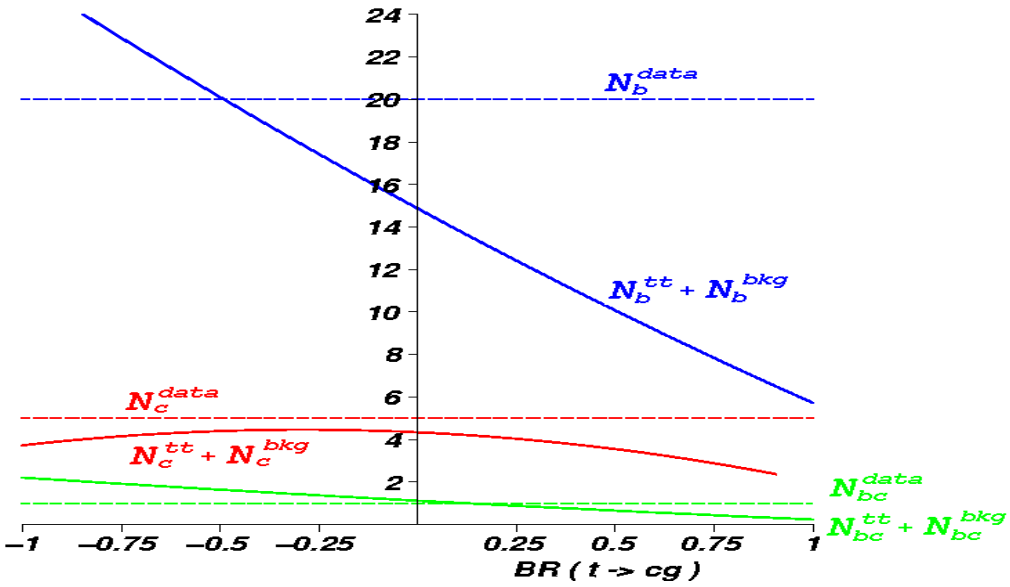


Figure 8.2: Numbers of b-,bc-, and c-tagged events from $t\bar{t}$ and background events versus $B(t \to cg)$. The dashed horizontal lines show the corresponding numbers of tags observed in the data.

events produced, tagging efficiencies, and numbers of tagged background events constrained near their estimated values:

$$\begin{aligned}
\mathcal{L} \;=\; & \prod_{k=b,bc,c} \Big[\, P\left(N_k^{t\bar{t}} + N_k^{\rm bkg}\,,\, N_k^{\rm data}\right) \cdot G\left(N_k^{\rm bkg}\,,\, \overline{N_k^{\rm bkg}}\,,\, \sigma_{N_k^{\rm bkg}}\right) \cdot \\
& \qquad G\left(\epsilon_{Wj,k}^{\text{SM-SM}}\,,\, \overline{\epsilon_{Wj,k}^{\text{SM-SM}}}\,,\, \sigma_{\epsilon_{Wj,k}^{\text{SM-SM}}}\right) \cdot G\left(\epsilon_{Wj,k}^{\text{SM-FCNC}}\,,\, \overline{\epsilon_{Wj,k}^{\text{SM-FCNC}}}\,,\, \sigma_{\epsilon_{Wj,k}^{\text{SM-FCNC}}}\right) \Big] \cdot \\
& G\left(N^{t\bar{t}}\,,\, \overline{N^{t\bar{t}}}\,,\, \sigma_{N^{t\bar{t}}}\right) \cdot g\left(s_{tl}\,,\, \overline{s_{tl}}\,,\, \sigma_{s_{tl}}\right) \cdot \\
& G\left(\epsilon_{Wj}^{\text{SM-SM}}\,,\, \overline{\epsilon_{Wj}^{\text{SM-SM}}}\,,\, \sigma_{\epsilon_{Wj}^{\text{SM-SM}}}\right) \cdot G\left(\epsilon_{Wj}^{\text{SM-FCNC}}\,,\, \overline{\epsilon_{Wj}^{\text{SM-FCNC}}}\,,\, \sigma_{\epsilon_{Wj}^{\text{SM-FCNC}}}\right)
\end{aligned} \tag{8.6}$$

$$\begin{aligned}
\text{with Poisson probability} \quad P(\mu,\nu) &= \frac{e^{-\mu}\mu^{\nu}}{\nu!} \\
\text{and Gaussian probability} \quad G(x,\overline{x},\sigma) &= \frac{e^{-(x-\overline{x})^2/2\sigma^2}}{\sigma\sqrt{2\pi}} \quad .
\end{aligned}$$

The likelihood depends on both variable and constant parameters. The constants represent our best estimates of the variables and their variances. The following 14 parameters, which represent quantities that we know only within some uncertainty, are variable:

$N^{t\bar{t}}_{b,bc,c}$ Expected numbers[3] of b-,bc-,and c-tagged $t\bar{t}$ events which are, through (8.4) and (8.3), functions of variables $N^{t\bar{t}}, B,\ s_{tl}\ ,\ \epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j}}\ ,\ \epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j},\{b,bc,c\}}$.[4]

$N^{t\bar{t}}$ Expected number of $t\bar{t}$ events produced.

B Branching fraction $B(t \to cg)$.

s_{tl} Scale factor for trigger and lepton-ID efficiencies.

$\epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j}}$ $W + \geq 3$ jets selection efficiencies for SM-SM, SM-FCNC $t\bar{t}$ events.

$\epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j},\{b,bc,c\}}$ Event b-,bc-,c-tagging efficiencies for SM-SM, SM-FCNC $t\bar{t}$ events which pass $W + \geq 3$ jets selection.

$N^{\text{bkg}}_{b,bc,c}$ Expected numbers of b-,bc-, and c-tagged background events.

The following 29 parameters are constant:

$N^{\text{data}}_{\{b,bc,c\}}$ Numbers of b-,bc-, and c-tagged events in the data.

$\overline{N^{\text{bkg}}_{\{b,bc,c\}}}$, $\sigma_{N^{\text{bkg}}_{\{b,bc,c\}}}$ Estimates of $N^{\text{bkg}}_{b,bc,c}$ and errors, given in Table 7.4.

$\overline{N^{t\bar{t}}}$, $\sigma_{N^{t\bar{t}}}$ Estimate of $N^{t\bar{t}}$ and its error, given by (8.1).

$\overline{s_{tl}}$, $\sigma_{s_{tl}}$ Estimate of s_{tl} and its error, given by (5.1).

$\overline{\epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j}}}$, $\sigma_{\epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j}}}$ Estimates of $\epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j}}$ and errors, given by (8.2).

$\overline{\epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j},\{\text{b,bc,c}\}}}$, $\sigma_{\epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j},\{\text{b,bc,c}\}}}$ Estimates of $\epsilon^{\{\text{SM-SM,SM-FCNC}\}}_{W\text{j},\{\text{b,bc,c}\}}$ and errors, given in Table 7.4.

[3] "Expected number" refers to the mean of a Poisson distribution.

[4] The numbers of tagged events given by (8.4) depend on the six net tagging-efficiencies on the L.H.S. of (8.3), but we instead use the nine factors on the R.H.S. of (8.3). This more complicated representation of the six net tagging-efficiencies is correct because the correlations between them are preserved, via (8.3), while the nine factors vary independently.

The factors in the likelihood product $\mathcal{L}$ (8.6) are organized as follows.

First there are three groups of four factors, where each group pertains only to b-,bc-, or c-tagged events. The first term in each group is the probability of observing the number of tagged events in the data given a Poisson mean of the sum of the expected numbers of tagged $t\bar{t}$ and background events. This is the only term in which B enters; all the other terms in the likelihood are present only because this term contains other parameters, besides B, which can vary. The second term is the probability of having the (variable) expected number of tagged background events given its (constant) estimate and error[5]. The third and fourth terms in each group are the probabilities of having the (variable) event-tagging efficiencies for SM-SM, SM-FCNC events given their (constant) estimates and errors.

Next come four terms for quantities which pertain commonly to b-,bc-, and c-tagged events: the expected number of $t\bar{t}$ events produced, the scale factor for trigger and lepton-ID efficiencies, and the efficiencies for SM-SM, SM-FCNC to pass the $W + \geq 3$ jets selection. Each term gives the probability of a quantity having its (variable) value given its (constant) estimate and error.

To find the variable parameter values (in particular, B) which maximize $\mathcal{L}$, we use the MINUIT function minimization program [41] to equivalently find the same values which minimize $-\log \mathcal{L}$. We refer to this procedure as "the branching fraction fit", where the constant parameters are inputs to the fit and the variable parameters are its outputs. With the exception that efficiencies and expected numbers of tags are required to be non-negative, all variable fit parameters are allowed to vary freely in the fit.

8.1.4 Branching Fraction Fit in Pseudo-experiments

We studied the performance of the branching fraction fit to understand the quality of the branching fraction estimator, $B_{\rm fit}$, and to check for possible biases. The fit was tested using an ensemble of pseudo-experiments. Recall that the fit takes as input a set of 29 fixed parameter inputs: three for the numbers of tagged events observed in the data, and

[5]This is appropriately a Gaussian probability because it is a Poisson *mean* which varies.

26 for variable parameter estimates and errors. The numbers of tagged events in the "data" will vary between pseudo-experiments, while the latter 26 parameters remain constant. For each pseudo-experiment, the numbers of tagged events in the data were generated according to the null hypothesis, $B = 0$. The generation procedure for each null hypothesis pseudo-experiment is outlined below:

- $N^{t\bar{t}}$, s_{tl}, $\epsilon_{W\mathrm{j}}^{\{\text{SM-SM,SM-FCNC}\}}$, and $\epsilon_{W\mathrm{j},\{\mathrm{b,bc,c}\}}^{\{\text{SM-SM,SM-FCNC}\}}$ are sampled from Gaussian distributions with respective means and widths given by (8.1), (5.1), (8.2), and Table 7.4.

- These fluctuated parameters are used in (8.4), with $B = 0$, to give the expected numbers of tagged $t\bar{t}$ events, $N_{b,bc,c}^{t\bar{t}}$. The actual numbers of tagged $t\bar{t}$ events are then sampled from Poisson distributions with these means.

- The expected numbers of tagged background events, $N_{b,bc,c}^{\mathrm{bkg}}$, are sampled from Gaussian distributions with means and widths given in Table 7.4. The actual numbers of tagged background events are then sampled from Poisson distributions with these means.

- The numbers of b-, bc-, and c-tagged events in the "data" are the sums of the actual numbers of tagged $t\bar{t}$ and background events.

Note that in the pseudo-experiment generation, each quantity with an uncertainty (and therefore with a corresponding variable parameter in the likelihood) varies.

Figure 8.3 (left) shows the distribution of the fitted branching fraction B_{fit} over a set of 20000 null hypothesis pseudo-experiments. The distribution has a mean of -0.018 ± 0.003 which is statistically different from the true input value of $B = 0$. The distribution is asymmetric about the mean, with a longer negative-side tail. It also shows some "chopiness" (independent of binning) because of the low statistics of tagged events: the numbers of tagged events in the "data" in each pseudo-experiment have fluctuating integer values, but the number of their different combinations, and therefore fitted branching fractions, is limited by the sizes of their Poisson means (especially the relatively small bc- and c-tag means shown in Figure 8.2).

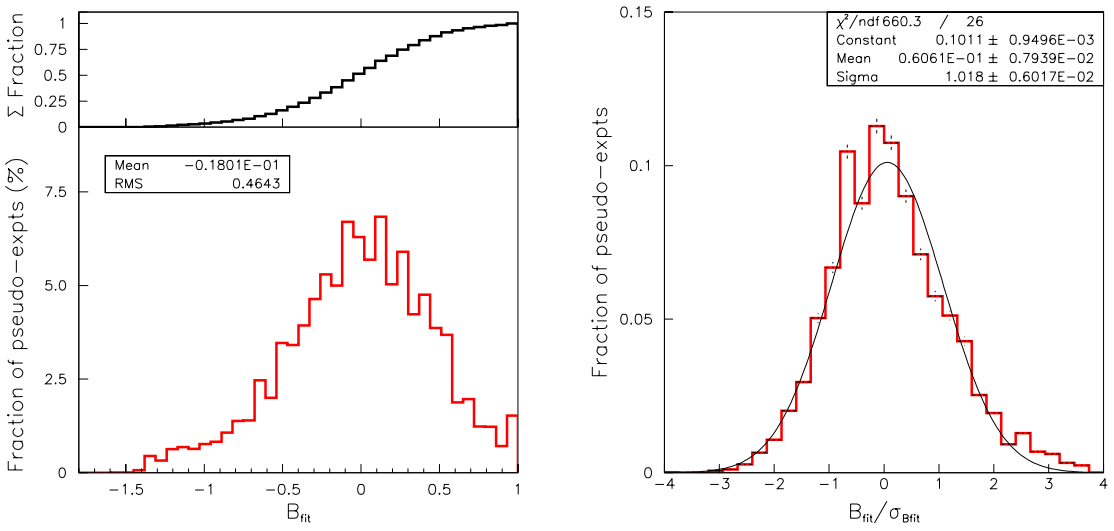


Figure 8.3: The distribution of fitted branching fraction (left, with integral inset) and pulls (right) for 20000 null hypothesis ($B = 0$) pseudo-experiments.

The upper-left plot of Figure 8.3 shows the integrated distribution of $B_{\rm fit}$ which will be used in § 8.1.5.

Figure 8.3 (right) shows the pull distribution over the pseudo-experiments, where the pull is defined for a particular pseudo-experiment as

$$\frac{B_{\rm fit} - B_{\rm input}}{\sigma_{B_{\rm fit}}} = \frac{B_{\rm fit}}{\sigma_{B_{\rm fit}}} \quad ,$$

where $B_{\rm input} = 0$ (null hypothesis), and the error $\sigma_{B_{\rm fit}}$ is returned by the fit. Overlaid on the pull distribution is its best fit to a Gaussian with mean 0.06 ± 0.008 and width 1.02 ± 0.006, both of which are statistically different from the ideal values of 0 and 1, respectively. Furthermore, there is poor goodness-of-fit to the Gaussian (χ^2/d.o.f. $\simeq 25$), a consequence of the "chopiness" mentioned above.

Both the distributions of fitted branching fractions and their pulls show the non-ideal features mentioned above. We argue that 1) these features are not shortcomings of the fit procedure and 2) they are inconsequential. First, in the limit of large $N^{t\bar{t}}$ (an effective increase in integrated luminosity), the non-ideal features vanish: $\overline{B_{\rm fit}} \to 0$ and the pull distribution approaches a Gaussian of unit width and zero mean. We therefore conclude

that all non-ideal features are the result of the low statistics of tagged events mentioned above. Second, even in our low-statistics regime, the fit procedure has no unreasonable biases on our branching fraction measurement. The branching fraction upper limit that we will set is independent of such biases, since the procedure used requires only that the estimator, $B_{\rm fit}$, has a known probability density function.

8.1.5 Branching Fraction Fit in Data

When run on the data, the fit converges to the branching fraction value

$$B_{\rm fit}^{\rm data} \;=\; -0.41\,^{+0.43}_{-0.52} \tag{8.7}$$

where the (one standard deviation) errors are defined as the change in B required to increase $-\log\mathcal{L}$ by 0.5 from its minimum. Figure 8.4 shows the shape of $-\log\mathcal{L}$ as a function of B (where $-\log\mathcal{L}$ is minimized with respect to the other variable parameters at each value of B). The likelihood shape is asymmetric about its minimum, and shows a preference for negative rather than positive values of B. The measured branching fraction (8.7) is consistent within error with the SM prediction of 0.

This particular value of the fitted branching fraction is reasonable according to the null hypothesis pseudo-experiments of § 8.1.4. If $B(t \to cg) = 0$, then we would expect to measure $B_{\rm fit} < B_{\rm fit}^{\rm data}$ 20% of the time, and $|B_{\rm fit}| > \left|B_{\rm fit}^{\rm data}\right|$ 37% of the time; see the upper-left integral plot in Figure 8.3.

The other interesting variable parameter value that the fit returns is

$$N^{t\bar{t}} \;=\; 440.7 \pm 78.3 \;.$$

This expected number of $t\bar{t}$ events produced is very consistent with the central value (8.1) to which it was Gaussian constrained. Therefore this fit suggests that the SM $t\bar{t}$ cross section agrees well with the observed numbers of tags in the data.

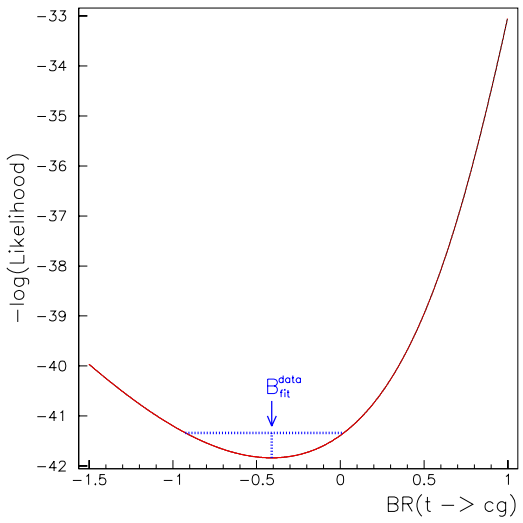


Figure 8.4: The $-\log \mathcal{L}$ vs. branching fraction B for the fit to data. The error determination on the value of B at its minimum, $B_{\rm fit} = -0.41\,^{+0.43}_{-0.52}$, is indicated by the dotted lines.

8.2 Branching Fraction $B(t \to cg)$ Upper Limit

Let us define $B_{\rm true}$ as the true but unknown value of $B(t \to cg)$. Next, let the fitted branching fraction, $B_{\rm fit}$, be an observable whose probability density function (pdf) depends on $B_{\rm true}$ in a known way, which we call $P(B_{\rm fit}|B_{\rm true})$. Then our situation is that of having made a single observation, $B_{\rm fit}$, with which, in addition to the pdf, we wish to make an inference about $B_{\rm true}$.

The first step is to measure the pdf, which is done using the pseudo-experiment technique of § 8.1.4. To measure $P(B_{\rm fit}|B_{\rm step})$, we sample the distribution of $B_{\rm fit}$ for pseudo-experiments generated with $B = B_{\rm step}$. To measure the entire pdf, we step $B_{\rm step}$ over the physical range $[0, 1]$. The accuracy of the pdf measurement increases with the number of pseudo-experiments. We defer discussion of the pdf shape until the following section.

High energy physicists have traditionally used classical (frequentist) confidence intervals, which are statements derived from the pdf, to report errors on experimental results [66]. One possible inference regarding $B_{\rm true}$ is given by the central confidence interval (8.7), whose proper interpretation is: in the limit of many experiments, such intervals will contain

the fixed unknown $B_{\rm true}$ in about 68% of the experiments[6]. However, since our result is near the lower physical boundary, it is natural to report a one-sided interval in the form of an upper limit on $B_{\rm true}$.

8.2.1 Confidence Belts For an Upper Limit

Here we follow the classical confidence interval construction outlined in [66]. We define a confidence interval $[B_1, B_2]$ as a member of a set, where the set has the property

$$P\left(B \in [B_1, B_2]\right) \;=\; \alpha \tag{8.8}$$

where B_1, B_2 are functions of the fitted branching fraction $B_{\rm fit}$ in an experiment, and the confidence intervals $[B_1, B_2]$ vary over an ensemble of experiments with fixed branching fraction B. Because (8.8) is true for every allowed B, the intervals contain the fixed $B_{\rm true}$ in a fraction α of experiments.

We use Neyman's construction of intervals which satisfy the above property[5]. For each value of the branching fraction B, we select an *acceptance* interval $\left[B_{\rm fit,1}\,, B_{\rm fit,2}\right]$ such that

$$P\left(B_{\rm fit} \in \left[B_{\rm fit,1}\,, B_{\rm fit,2}\right] \,\middle|\, B\right) \;=\; \alpha \quad ,$$

that is, given the fixed value of the branching fraction B, there is probability α of measuring a value of the fitted branching fraction $B_{\rm fit}$ which is included in the acceptance interval. To uniquely specify the acceptance intervals, we choose $B_{\rm fit,2} = \infty$,

$$P\left(B_{\rm fit} < B_{\rm fit,1} \,\middle|\, B\right) \;=\; 1-\alpha \quad , \tag{8.9}$$

which leads to upper confidence limits satisfying

$$P\left(B > B_2\right) \;=\; 1-\alpha \;, \tag{8.10}$$

where we choose the confidence level (CL) $\alpha = 0.95$. The confidence intervals are constructed by first determining the acceptance intervals as a function of B. For a given value of B, the

[6]Strictly speaking, (8.7) is such a confidence interval only to the extent that the likelihood function is Gaussian [67].

pdf $P(B_{\rm fit}|B)$ is sampled in a ensemble of 20000 pseudo-experiments. Then the lower edge of the acceptance interval, $B_{\rm fit,1}$, is measured using (8.9). Finally, this procedure is repeated for B in the physical range $[0,1]$ in steps of 1%.

The resulting set of acceptance intervals, called a confidence belt, is shown in Figure 8.5. These acceptance intervals give the region of $B_{\rm fit}$ vs. B space which contains 95% of the probability in the pdf, and they characterize the limit-setting power of our technique with a particular tagger. The horizontal lines show the acceptance intervals for each value of B. The lower edge of these one-sided intervals, $B_{\rm fit,1}(B)$, is a function of B. The inverse of this function gives the upper edge of the belt, $B_2(B_{\rm fit})$, as a function of $B_{\rm fit}$. The construction is completed by defining $B_2(B_{\rm fit})$ as the confidence interval upper limit for B corresponding to the measurement $B_{\rm fit}(B)$ in an experiment. Then by construction, (8.10) is satisfied for all B, including the unknown fixed value $B_{\rm true}$. Therefore these intervals have correct coverage at the stated 95% CL.

Note that the width of the confidence belt, and therefore the upper limits, increases with the width of the distribution of the fitted branching fraction $B_{\rm fit}$ shown in Figure 8.3 (left); i.e. the better we can measure the branching fraction, the lower the limit we can set.

8.2.2 Expected Upper Limit From Pseudo-experiments

Each null hypothesis pseudo-experiment described in § 8.1.4 returns a fitted branching fraction $B_{\rm fit}$ which, using the confidence intervals derived in § 8.2.1, defines a branching fraction upper limit B_{95}. The distribution of upper limits over the null hypothesis pseudo-experiments is shown in Figure 8.6. The pile-up in the zero-bin is from pseudo-experiments with values of $B_{\rm fit}$ less than the acceptance intervals for all physical B (Figure 8.5); in these cases the upper limit was assigned by the closest acceptance interval ($B = 0$)[7]. The pile-up in the one-bin is from pseudo-experiments with values of $B_{\rm fit} > 0.45$, which all have upper

[7]We tried an alternate confidence belt construction which uses an ordering principle to define the acceptance intervals [66]. This construction avoids the problem of empty confidence intervals mentioned in the text, and the pile-up at $B_{95} = 0$ spreads smoothly over small values of B_{95}. However, the mean of this distribution and the limit we set in the data do not change, so we continue to use the classical belt.

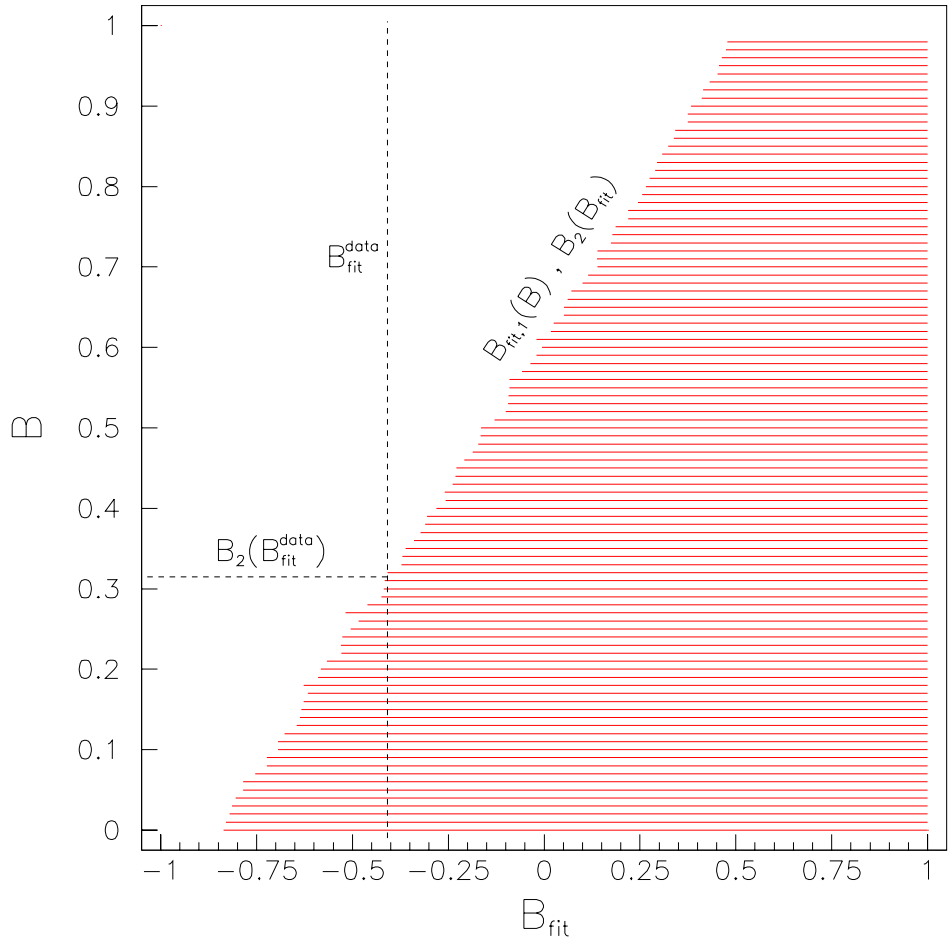


Figure 8.5: Confidence belt corresponding to 95% CL upper limits for unknown branching fraction B and estimator $B_{\rm fit}$. For each value of B, a horizontal line, with lower edge $B_{{\rm fit},1}(B)$, shows the acceptance interval satisfying (8.9). Then in an experiment with measured $B_{\rm fit}$, the confidence interval for B is the union of all values of B whose acceptance intervals intersect the vertical line through $B_{\rm fit}$. The upper edge of this union, $B_2(B_{\rm fit})$, gives the 95% CL upper limit for B. This construction is shown for the fitted branching fraction in the data by the dashed lines.

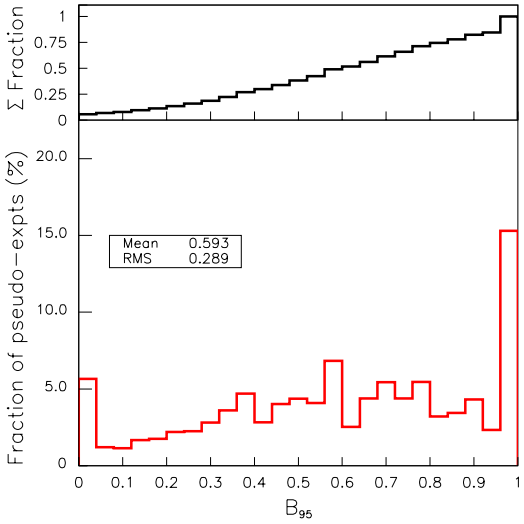


Figure 8.6: The distribution of 95% CL upper limits (with integral inset) for 20000 null hypothesis ($B = 0$) pseudo-experiments.

limits of one. The mean of the distribution of B_{95} gives the upper limit we would expect to set,

$$B_{95}^{\text{expected}} = 0.59 \ , \tag{8.11}$$

given the null hypothesis ($B = 0$). The integrated B_{95} is shown in the upper inset plot and is used in § 8.2.3.

8.2.3 Branching Fraction Upper Limit For the Data

The upper limit on the branching fraction B is determined by the fitted branching fraction in the data (8.7) and the confidence belt described in § 8.2.1. We set the limit[8]

$$B_{95}^{\text{data}} \ = \ 0.32 \ , \tag{8.12}$$

as shown by the dashed lines in Figure 8.5. The meaning of this limit is that if the true value of B were B_{95}^{data}, then we would expect to observe a fitted branching fraction $B_{\text{fit}} < B_{\text{fit}}^{\text{data}}$ just 5% of the time.

[8]This limit, checked with 10^6 pseudo-experiments, is slightly conservative ($\alpha = 0.951$).

This particular upper limit is reasonable according to our expectation from null hypothesis pseudo-experiments. From the integral inset plot in Figure 8.6, we would expect to measure $B_{95} < B_{95}^{\rm data}$ 19% of the time. The upper limit we set is less than the expected upper limit (8.11) because the upward fluctuations of the numbers of tags in the data (Table 7.4) favor the low fitted branching (8.7).

8.3 Branching Fraction $B(t \to cZ)$ Fit and Upper Limit

The techniques used in this analysis were also applied to a search for the FCNC top-quark decay $t \to cZ$. The same data and background event samples, event selection criteria, sample-composition estimates, and flavor-tag counting methods were used as in Chapters 5-7[9]. But this search requires a new simulated sample of events in which the top quark may SM-decay as $t \to Wb$ or FCNC-decay as $t \to cZ$. We again used ISAJET to generate 20,000 events with mixed SM-FCNC decays. Unlike events in the previous search for $t \to cg$, events with dual FCNC top-quark decays $t\bar{t} \to cZ\bar{c}Z$ can include a leptonic final-state, so they are modeled as well (with 10,000 ISAJET events).

Following the discussion of § 8.1.1, we assume that the top quark has two possible decay modes, with $B(t \to Wb) + B(t \to cZ) = 1$. The expected numbers of tagged $t\bar{t}$ events is given as the following sum of tags from SM-SM, SM-FCNC, and FCNC-FCNC events,

$$N_{b,bc,c}^{t\bar{t}}(B) \;=\; N^{t\bar{t}} \left[(1-B)^2 \epsilon_{b,bc,c}^{\text{SM-SM}} + 2B(1-B)\epsilon_{b,bc,c}^{\text{SM-FCNC}} + B^2 \epsilon_{b,bc,c}^{\text{FCNC-FCNC}}\right] \quad , \qquad (8.13)$$

where $B = B(t \to cZ)$, event-tagging efficiencies are analogous to those in described § 8.1.1, and there is an additional contribution from events with dual-FCNC top-quark decays. We again use NN Tagger "2e" because it yielded the lowest expected limit on $B(t \to cZ)$; the corresponding tagging efficiencies and tagged-event counts, for $t\bar{t}$ signals and backgrounds, are listed in Appendix K.

Figure 8.7 shows the numbers of tagged $t\bar{t}$ events, separately for each decay mode, as a function of B. For each tag type, as B increases from zero, the number of SM-SM

[9] One difference here is that for c-tagged events, we no longer make any dijet mass cuts.

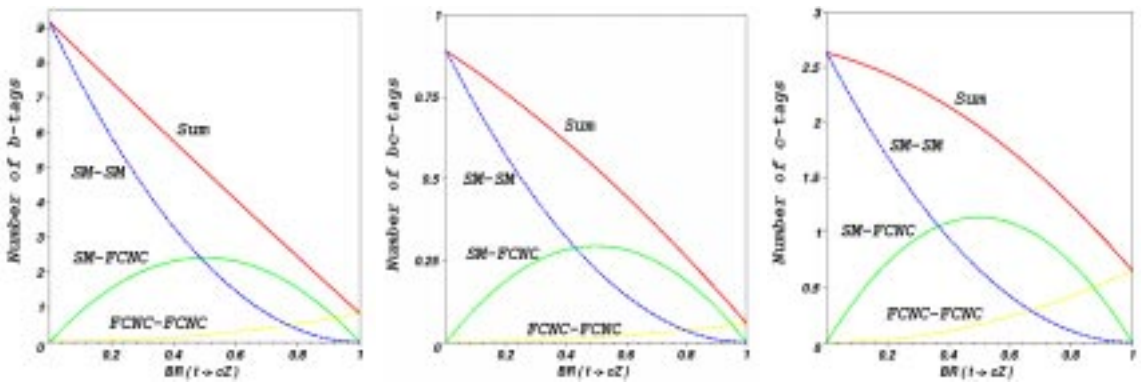


Figure 8.7: Numbers of b-,bc-, and c-tagged $t\bar{t}$ events versus $B(t \to cZ)$. Contributions from SM-SM, SM-FCNC, and FCNC-FCNC top-quark decay modes are shown separately beneath their sum.

tags decreases from its maximum, while the numbers of tags from FCNC events begin to increase. As $B \to 1$, contributions from the SM-SM and SM-FCNC events are lost to the FCNC-FCNC mode which increases to its maximum.

Figure 8.8 shows the expected numbers of tagged events, given by the sum of (8.13) and backgrounds, as a function of B. The numbers of each tag observed in the data are shown by horizontal lines in the figure.

We again perform a maximum-likelihood fit to determine the value of $B(t \to cZ)$ which is most compatible with the observed data. The fit is identical to that described in § 8.1.3 with the following exceptions: the expected numbers of tagged $t\bar{t}$ events are given by (8.13), and there are additional Gaussian terms in the likelihood function for the efficiencies associated with the FCNC-FCNC signature. We measured the branching fraction value

$$B(t \to cZ)^{\rm data}_{\rm fit} \;=\; -0.55\,^{+0.55}_{-0.71} \quad ,$$

which is consistent with the SM prediction of 0. Using the same procedure as described in § 8.2, this measured value translates into the following 95% CL upper limit on the branching fraction,

$$B(t \to cZ)^{\rm data}_{95} \;=\; 0.34 \; .$$

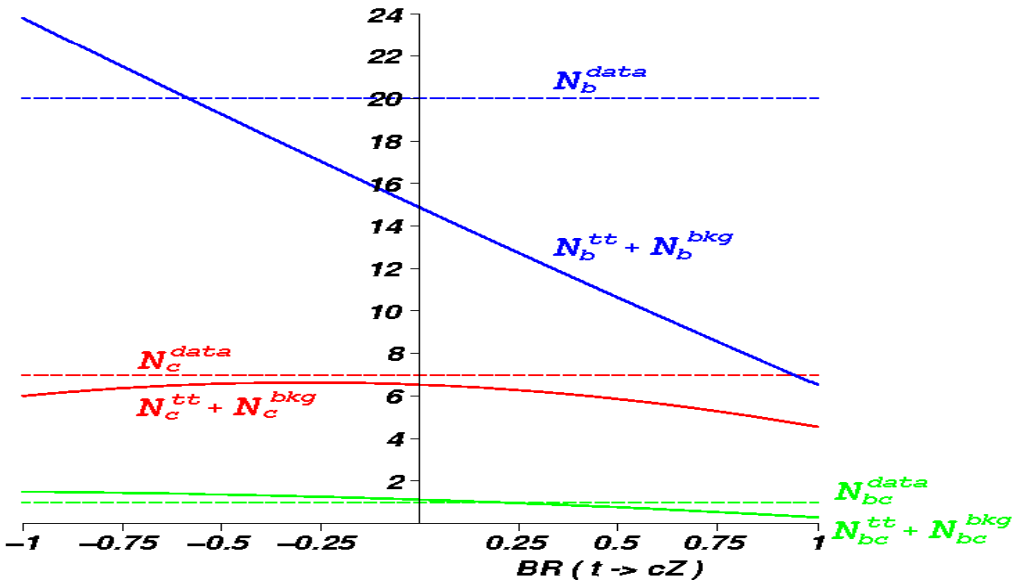


Figure 8.8: Numbers of b-,bc-, and c-tagged events from $t\bar{t}$ and background events versus $B(t \to cZ)$. The dashed horizontal lines show the corresponding numbers of tags observed in the data.

Chapter 9

Conclusions

The SM prohibits flavor-changing decays of quarks through neutral currents. We searched for two such decays of the top quark, $t \to cg$ and $t \to cZ$. Decays with these final states are highly suppressed in the SM, with expected rates on the order of 10^{-10}. Searches for these decays provide tests of the short-distance structure of the QCD and electroweak interactions at the largest fermion mass scale in the SM.

Using a new jet-flavor identification technique, we counted events with b-, c-, and bc-tags in a data sample of candidate $t\bar{t}$ events. In each search, we assumed that a top quark may have either a SM or FCNC decay. The tag counts in the data were compared to predictions given by a sum of tagged background events and tagged $t\bar{t}$ events, with the top-quark decay parameterized by the branching fraction to the FCNC mode. The predictions were obtained by measuring tagging rates in simulated events, and were then modified with a set of corrections to account for expected differences between the simulation and data. We then determined the values of the branching fractions which maximize the likelihood of observing the data to be

$$\begin{aligned} B(t \to cg) &= -0.41\,^{+0.43}_{-0.52} \\ B(t \to cZ) &= -0.55\,^{+0.55}_{-0.71} \quad . \end{aligned}$$

With standard statistical techniques, these measurements translate into the following upper

limits at 95% CL,

$$B(t \to cg)^{95} = 0.32$$
$$B(t \to cZ)^{95} = 0.34 \quad .$$

The errors on the above measurements reflect statistical uncertainties and the differences in tagging efficiencies between simulation and data. The limit we set on $t \to cZ$ is comparable to the existing limit, $B(t \to cZ) < 0.33$, achieved in another direct search[1]. Our limit on the flavor-changing QCD decay $t \to cg$ is new. In either case, we set significantly better limits than expected because of upward fluctuations in the number of b- and c-tags in the data.

The measured values of the branching fractions are both consistent, within error, with the SM predictions ~ 0. The limits, at 10 orders of magnitude greater than their expected values, also pose no challenge to the SM. The weakness of the limits is due to the low statistics of tagged events. In the data we observe 20 b-tags and about 5 c-tags in both searches. Meanwhile, the difference in the expected numbers of tags between pure-SM and mixed SM-FCNC decays is typically on the order of just several events (depending on the branching ratio). The limits we set above reflect the point at which statistical fluctuations, which are relatively large, can no longer account for the tag differences between events with SM and FCNC decays, which are relatively small.

Tagged events are in short supply because both the $t\bar{t}$ production cross section and our tagging efficiencies are small. In our search sample of 244 events before tagging, we expect only 31 from $t\bar{t}$. Flavor-tagging greatly improves the signal-to-background ratio, but the statistics suffer from b- and c- tagging efficiencies of about 30% and 8%, respectively. These efficiencies could be increased (significantly for c jets) by relaxing the taggable-jet criteria, but our choice was optimal for setting a limit. We also remark that for our small sample of tagged events, an additional kinematic cut on c-tagged jets (to help distinguish SM from FCNC events) only increased the expected limit.

Fermilab Run 2 will provide about 20 times more integrated luminosity than our Run 1 data set. By increasing the amounts of signal and backgrounds to Run 2 levels in pseudo-experiments, but otherwise keeping the analysis unchanged, we find that the expected

[1] However our search, which includes Z rejection cuts, was not optimized for this decay mode.

limits are halved by the increase of statistics alone. With this increased amount of data, stricter tagging criteria could be used to further extract the signal and set a lower limit.

Before this analysis, there was no efficient way to separately identify b, c, and prompt jets. Previous flavor taggers typically cut on one lifetime-related variable to identify b jets only. We used tracking information from charged daughters to form 14 flavor discrimination variables related to expected differences in mass, lifetime, and fragmentation characteristics of heavy-flavor jets. We chose to combine the discrimination variables with a neural network, and we interpret the network outputs as flavor amplitudes when making the tag. Our NN tagger includes a host of choices which could be made differently for other analyses. We found that our model algorithm yields reasonable flavor-tagging efficiencies and purities.

An interesting spin-off to this thesis is the technique to measure the tagging-efficiency scale factors. Deficiencies in the simulation have long been known, but a way to quantify their effects on a three-outcome tagger, separately for three input flavors, was invented for this analysis. Although application of the scale factors requires some assumptions about their invariance, their physical parameterization and measurement using flavor control-samples makes them a powerful and necessary first-order correction to the simulation.

Appendices

Appendix A

The CDF Collaboration

T. Affolder,[21] H. Akimoto,[42] A. Akopian,[35] M. G. Albrow,[10] P. Amaral,[7] S. R. Amendolia,[31] D. Amidei,[24] J. Antos,[1] G. Apollinari,[35] T. Arisawa,[42] T. Asakawa,[40] W. Ashmanskas,[7] M. Atac,[10] P. Azzi-Bacchetta,[29] N. Bacchetta,[29] M. W. Bailey,[26] S. Bailey,[14] P. de Barbaro,[34] A. Barbaro-Galtieri,[21] V. E. Barnes,[33] B. A. Barnett,[17] M. Barone,[12] G. Bauer,[22] F. Bedeschi,[31] S. Belforte,[39] G. Bellettini,[31] J. Bellinger,[43] D. Benjamin,[9] J. Bensinger,[4] A. Beretvas,[10] J. P. Berge,[10] J. Berryhill,[7] S. Bertolucci,[12] B. Bevensee,[30] A. Bhatti,[35] C. Bigongiari,[31] M. Binkley,[10] D. Bisello,[29] R. E. Blair,[2] C. Blocker,[4] K. Bloom,[24] B. Blumenfeld,[17] B. S. Blusk,[34] A. Bocci,[31] A. Bodek,[34] W. Bokhari,[30] G. Bolla,[33] Y. Bonushkin,[5] D. Bortoletto,[33] J. Boudreau,[32] A. Brandl,[26] S. van den Brink,[17] C. Bromberg,[25] N. Bruner,[26] E. Buckley-Geer,[10] J. Budagov,[8] H. S. Budd,[34] K. Burkett,[14] G. Busetto,[29] A. Byon-Wagner,[10] K. L. Byrum,[2] M. Campbell,[24] A. Caner,[31] W. Carithers,[21] J. Carlson,[24] D. Carlsmith,[43] J. Cassada,[34] A. Castro,[29] D. Cauz,[39] A. Cerri,[31] P. S. Chang,[1] P. T. Chang,[1] J. Chapman,[24] C. Chen,[30] Y. C. Chen,[1] M. -T. Cheng,[1] M. Chertok,[37] G. Chiarelli,[31] I. Chirikov-Zorin,[8] G. Chlachidze,[8] F. Chlebana,[10] L. Christofek,[16] M. L. Chu,[1] S. Cihangir,[10] C. I. Ciobanu,[27] A. G. Clark,[13] M. Cobal,[31] E. Cocca,[31] A. Connolly,[21] J. Conway,[36] J. Cooper,[10] M. Cordelli,[12] J. Guimaraes da Costa,[24] D. Costanzo,[31] J. Cranshaw,[38] D. Cronin-Hennessy,[9] R. Cropp,[23] R. Culbertson,[7] D. Dagenhart,[41] F. DeJongh,[10] S. Dell'Agnello,[12] M. Dell'Orso,[31] R. Demina,[10]

L. Demortier,[35] M. Deninno,[3] P. F. Derwent,[10] T. Devlin,[36] J. R. Dittmann,[10] S. Donati,[31] J. Done,[37] T. Dorigo,[14] N. Eddy,[16] K. Einsweiler,[21] J. E. Elias,[10] E. Engels, Jr.,[32] W. Erdmann,[10] D. Errede,[16] S. Errede,[16] Q. Fan,[34] R. G. Feild,[44] C. Ferretti,[31] I. Fiori,[3] B. Flaugher,[10] G. W. Foster,[10] M. Franklin,[14] J. Freeman,[10] J. Friedman,[22] Y. Fukui,[20] S. Gadomski,[23] S. Galeotti,[31] M. Gallinaro,[35] T. Gao,[30] M. Garcia-Sciveres,[21] A. F. Garfinkel,[33] P. Gatti,[29] C. Gay,[44] S. Geer,[10] D. W. Gerdes,[24] P. Giannetti,[31] P. Giromini,[12] V. Glagolev,[8] M. Gold,[26] J. Goldstein,[10] A. Gordon,[14] A. T. Goshaw,[9] Y. Gotra,[32] K. Goulianos,[35] H. Grassmann,[39] C. Green,[33] L. Groer,[36] C. Grosso-Pilcher,[7] M. Guenther,[33] G. Guillian,[24] R. S. Guo,[1] C. Haber,[21] E. Hafen,[22] S. R. Hahn,[10] C. Hall,[14] T. Handa,[15] R. Handler,[43] W. Hao,[38] F. Happacher,[12] K. Hara,[40] A. D. Hardman,[33] R. M. Harris,[10] F. Hartmann,[18] K. Hatakeyama,[35] J. Hauser,[5] J. Heinrich,[30] A. Heiss,[18] B. Hinrichsen,[23] K. D. Hoffman,[33] C. Holck,[30] R. Hollebeek,[30] L. Holloway,[16] R. Hughes,[27] J. Huston,[25] J. Huth,[14] H. Ikeda,[40] M. Incagli,[31] J. Incandela,[10] G. Introzzi,[31] J. Iwai,[42] Y. Iwata,[15] E. James,[24] H. Jensen,[10] M. Jones,[30] U. Joshi,[10] H. Kambara,[13] T. Kamon,[37] T. Kaneko,[40] K. Karr,[41] H. Kasha,[44] Y. Kato,[28] T. A. Keaffaber,[33] K. Kelley,[22] M. Kelly,[24] R. D. Kennedy,[10] R. Kephart,[10] D. Khazins,[9] T. Kikuchi,[40] M. Kirk,[4] B. J. Kim,[19] H. S. Kim,[23] S. H. Kim,[40] Y. K. Kim,[21] L. Kirsch,[4] S. Klimenko,[11] D. Knoblauch,[18] P. Koehn,[27] A. Köngeter,[18] K. Kondo,[42] J. Konigsberg,[11] K. Kordas,[23] A. Korytov,[11] E. Kovacs,[2] J. Kroll,[30] M. Kruse,[34] S. E. Kuhlmann,[2] K. Kurino,[15] T. Kuwabara,[40] A. T. Laasanen,[33] N. Lai,[7] S. Lami,[35] S. Lammel,[10] J. I. Lamoureux,[4] M. Lancaster,[21] G. Latino,[31] T. LeCompte,[2] A. M. Lee IV,[9] S. Leone,[31] J. D. Lewis,[10] M. Lindgren,[5] T. M. Liss,[16] J. B. Liu,[34] Y. C. Liu,[1] N. Lockyer,[30] M. Loreti,[29] D. Lucchesi,[29] P. Lukens,[10] S. Lusin,[43] J. Lys,[21] R. Madrak,[14] K. Maeshima,[10] P. Maksimovic,[14] L. Malferrari,[3] M. Mangano,[31] M. Mariotti,[29] G. Martignon,[29] A. Martin,[44] J. A. J. Matthews,[26] P. Mazzanti,[3] K. S. McFarland,[34] P. McIntyre,[37] E. McKigney,[30] M. Menguzzato,[29] A. Menzione,[31] E. Meschi,[31] C. Mesropian,[35] C. Miao,[24] T. Miao,[10] R. Miller,[25] J. S. Miller,[24] H. Minato,[40] S. Miscetti,[12] M. Mishina,[20] N. Moggi,[31] E. Moore,[26] R. Moore,[24] Y. Morita,[20] A. Mukherjee,[10] T. Muller,[18] A. Munar,[31] P. Murat,[31]

S. Murgia,[25] M. Musy,[39] J. Nachtman,[5] S. Nahn,[44] H. Nakada,[40] T. Nakaya,[7] I. Nakano,[15] C. Nelson,[10] D. Neuberger,[18] C. Newman-Holmes,[10] C.-Y. P. Ngan,[22] P. Nicolaidi,[39] H. Niu,[4] L. Nodulman,[2] A. Nomerotski,[11] S. H. Oh,[9] T. Ohmoto,[15] T. Ohsugi,[15] R. Oishi,[40] T. Okusawa,[28] J. Olsen,[43] C. Pagliarone,[31] F. Palmonari,[31] R. Paoletti,[31] V. Papadimitriou,[38] S. P. Pappas,[44] A. Parri,[12] D. Partos,[4] J. Patrick,[10] G. Pauletta,[39] M. Paulini,[21] A. Perazzo,[31] L. Pescara,[29] T. J. Phillips,[9] G. Piacentino,[31] K. T. Pitts,[10] R. Plunkett,[10] A. Pompos,[33] L. Pondrom,[43] G. Pope,[32] F. Prokoshin,[8] J. Proudfoot,[2] F. Ptohos,[12] G. Punzi,[31] K. Ragan,[23] D. Reher,[21] A. Ribon,[29] F. Rimondi,[3] L. Ristori,[31] W. J. Robertson,[9] A. Robinson,[23] T. Rodrigo,[6] S. Rolli,[41] L. Rosenson,[22] R. Roser,[10] R. Rossin,[29] W. K. Sakumoto,[34] D. Saltzberg,[5] A. Sansoni,[12] L. Santi,[39] H. Sato,[40] P. Savard,[23] P. Schlabach,[10] E. E. Schmidt,[10] M. P. Schmidt,[44] M. Schmitt,[14] L. Scodellaro,[29] A. Scott,[5] A. Scribano,[31] S. Segler,[10] S. Seidel,[26] Y. Seiya,[40] A. Semenov,[8] F. Semeria,[3] T. Shah,[22] M. D. Shapiro,[21] P. F. Shepard,[32] T. Shibayama,[40] M. Shimojima,[40] M. Shochet,[7] J. Siegrist,[21] G. Signorelli,[31] A. Sill,[38] P. Sinervo,[23] P. Singh,[16] A. J. Slaughter,[44] K. Sliwa,[41] C. Smith,[17] F. D. Snider,[10] A. Solodsky,[35] J. Spalding,[10] T. Speer,[13] P. Sphicas,[22] F. Spinella,[31] M. Spiropulu,[14] L. Spiegel,[10] L. Stanco,[29] J. Steele,[43] A. Stefanini,[31] J. Strologas,[16] F. Strumia, [13] D. Stuart,[10] K. Sumorok,[22] T. Suzuki,[40] R. Takashima,[15] K. Takikawa,[40] M. Tanaka,[40] T. Takano,[28] B. Tannenbaum,[5] W. Taylor,[23] M. Tecchio,[24] P. K. Teng,[1] K. Terashi,[40] S. Tether,[22] D. Theriot,[10] R. Thurman-Keup,[2] P. Tipton,[34] S. Tkaczyk,[10] K. Tollefson,[34] A. Tollestrup,[10] H. Toyoda,[28] W. Trischuk,[23] J. F. de Troconiz,[14] S. Truitt,[24] J. Tseng,[22] N. Turini,[31] F. Ukegawa,[40] J. Valls,[36] S. Vejcik III,[10] G. Velev,[31] R. Vidal,[10] R. Vilar,[6] I. Vologouev,[21] D. Vucinic,[22] R. G. Wagner,[2] R. L. Wagner,[10] J. Wahl,[7] N. B. Wallace,[36] A. M. Walsh,[36] C. Wang,[9] C. H. Wang,[1] M. J. Wang,[1] T. Watanabe,[40] T. Watts,[36] R. Webb,[37] H. Wenzel,[18] W. C. Wester III,[10] A. B. Wicklund,[2] E. Wicklund,[10] H. H. Williams,[30] P. Wilson,[10] B. L. Winer,[27] D. Winn,[24] S. Wolbers,[10] D. Wolinski,[24] J. Wolinski,[25] S. Worm,[26] X. Wu,[13] J. Wyss,[31] A. Yagil,[10] W. Yao,[21] G. P. Yeh,[10] P. Yeh,[1] J. Yoh,[10] C. Yosef,[25] T. Yoshida,[28] I. Yu,[19] S. Yu,[30] A. Zanetti,[39] F. Zetti,[21] and S. Zucchelli[3]

(CDF Collaboration)

[1] *Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China*

[2] *Argonne National Laboratory, Argonne, Illinois 60439*

[3] *Istituto Nazionale di Fisica Nucleare, University of Bologna, I-40127 Bologna, Italy*

[4] *Brandeis University, Waltham, Massachusetts 02254*

[5] *University of California at Los Angeles, Los Angeles, California 90024*

[6] *Instituto de Fisica de Cantabria, University of Cantabria, 39005 Santander, Spain*

[7] *Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637*

[8] *Joint Institute for Nuclear Research, RU-141980 Dubna, Russia*

[9] *Duke University, Durham, North Carolina 27708*

[10] *Fermi National Accelerator Laboratory, Batavia, Illinois 60510*

[11] *University of Florida, Gainesville, Florida 32611*

[12] *Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy*

[13] *University of Geneva, CH-1211 Geneva 4, Switzerland*

[14] *Harvard University, Cambridge, Massachusetts 02138*

[15] *Hiroshima University, Higashi-Hiroshima 724, Japan*

[16] *University of Illinois, Urbana, Illinois 61801*

[17] *The Johns Hopkins University, Baltimore, Maryland 21218*

[18] *Institut für Experimentelle Kernphysik, Universität Karlsruhe, 76128 Karlsruhe, Germany*

[19] *Korean Hadron Collider Laboratory: Kyungpook National University, Taegu 702-701; Seoul National University, Seoul 151-742; and SungKyunKwan University, Suwon 440-746; Korea*

[20] *High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305, Japan*

[21] *Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720*

[22] *Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

[23] *Institute of Particle Physics: McGill University, Montreal H3A 2T8; and University of Toronto, Toronto M5S 1A7; Canada*

[24] *University of Michigan, Ann Arbor, Michigan 48109*

[25] *Michigan State University, East Lansing, Michigan 48824*

[26] *University of New Mexico, Albuquerque, New Mexico 87131*

[27] *The Ohio State University, Columbus, Ohio 43210*

[28] *Osaka City University, Osaka 588, Japan*

[29] *Universita di Padova, Istituto Nazionale di Fisica Nucleare, Sezione di Padova, I-35131 Padova, Italy*

[30] *University of Pennsylvania, Philadelphia, Pennsylvania 19104*

[31] *Istituto Nazionale di Fisica Nucleare, University and Scuola Normale Superiore of Pisa, I-56100 Pisa, Italy*

[32] *University of Pittsburgh, Pittsburgh, Pennsylvania 15260*

[33] *Purdue University, West Lafayette, Indiana 47907*

[34] *University of Rochester, Rochester, New York 14627*

[35] *Rockefeller University, New York, New York 10021*

[36] *Rutgers University, Piscataway, New Jersey 08855*

[37] *Texas A&M University, College Station, Texas 77843*

[38] *Texas Tech University, Lubbock, Texas 79409*

[39] *Istituto Nazionale di Fisica Nucleare, University of Trieste/ Udine, Italy*

[40] *University of Tsukuba, Tsukuba, Ibaraki 305, Japan*

[41] *Tufts University, Medford, Massachusetts 02155*

[42] *Waseda University, Tokyo 169, Japan*

[43] *University of Wisconsin, Madison, Wisconsin 53706*

[44] *Yale University, New Haven, Connecticut 06520*

Appendix B

Dependence of Mean Impact-Parameter on Lifetime

In this appendix we show that the mean impact parameter of decay daughters is proportional to the lifetime of the parent particle. This solution is modeled on the problem posed on page 84 of [4].

Consider a particle X moving with Lorentz factor $\gamma = (1-\beta^2)^{-\frac{1}{2}}$ in the lab frame. Suppose that, in the rest frame of X, it lives for time τ before decaying into a secondary particle Y (other decay daughters are ignored here) with energy Y_0 and momentum components Y_L, Y_T along and perpendicular to the X direction as shown in Figure B.1.

The impact parameter d is the projection of the X decay length onto the Y direction:

$$d = \mid \gamma\beta c\tau \sin\theta^{lab} \mid$$

$$\text{where } \sin\theta^{lab} = \frac{Y_T^{lab}}{\sqrt{(Y_T^{lab})^2 + (Y_L^{lab})^2}} \quad .$$

To find the momentum components of Y in the lab frame we boost backwards from the rest frame of X:

$$\mathbf{Y^{lab}} = \begin{pmatrix} Y_0^{lab} \\ Y_L^{lab} \\ Y_T^{lab} \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} Y_0 \\ Y_L \\ Y_T \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma Y_0 + \gamma\beta Y_L \\ \gamma\beta Y_0 + \gamma Y_L \\ Y_T \\ 0 \end{pmatrix}$$

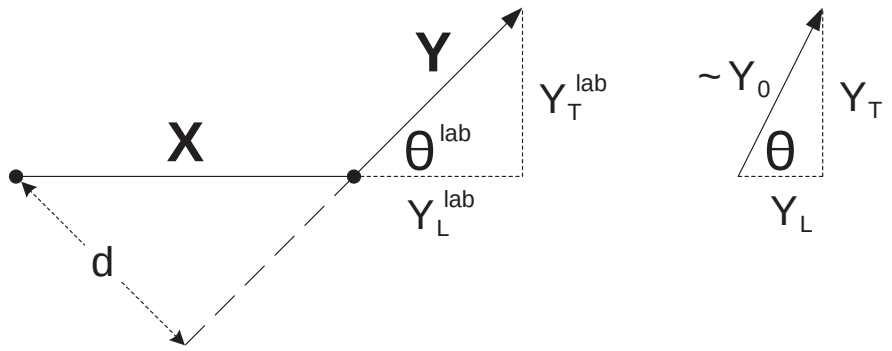


Figure B.1: The decay of particle X and a daughter Y in the lab frame (left), where the decay length of X is $\gamma\beta c\tau$. Also shown (right) are Y momenta in the rest frame of X.

Then for relativistic X ($\beta \simeq 1, \gamma \gg 1$), the impact parameter becomes

$$d \;=\; \gamma\beta c\tau \, \frac{Y_T}{\sqrt{Y_T^2 + (\gamma\beta Y_0 + \gamma Y_L)^2}} \;\simeq\; c\tau \, \frac{Y_T}{Y_0 + Y_L} \quad .$$

If X is spinless and unpolarized, then its decay is isotropic and the mean impact parameter is

$$\langle d \rangle \;=\; \langle c\tau \rangle \left\langle \frac{Y_T}{Y_0 + Y_L} \right\rangle \quad .$$

The momentum average is done in the rest frame of X, and we assume that particle Y has small mass ($Y_0^2 \simeq Y_T^2 + Y_L^2$),

$$\left\langle \frac{Y_T}{Y_0 + Y_L} \right\rangle \;=\; \left\langle \frac{Y_T/Y_0}{1 + Y_L/Y_0} \right\rangle \;=\; \left\langle \frac{\sin\theta}{1 + \cos\theta} \right\rangle \;=\; \frac{\pi}{2} \quad .$$

Therefore the mean impact parameter is proportional to the mean lifetime of X,

$$\langle d \rangle \;=\; \frac{\pi}{2} \, c \, \langle \tau \rangle \quad .$$

Note that for a 20 GeV b quark (corresponding to uncorrected jet E_T of 15 GeV),

$$\gamma \;=\; 20\mathrm{GeV}/m_b \;=\; 4.3 \quad ,$$

so the approximations made above are good. A similarly energetic c quark is boosted by a factor m_b/m_c more than a b quark, so the approximations are even better.

Appendix C

Comparisons of Data and MC Control Samples Used in the Scale-Factor Measurement

In this appendix we compare data and MC distributions of NN input (discrimination) variables, NN output (flavor) variables, and other event quantities.

C.1 Reconstructed D^* Events

Because the RS reconstructed D^* events in the data include combinatoric background (which we model with WS data events as described in § 4.2.2), we compare the data RS−WS difference and MC distributions in this section. Note that in most cases the WS shapes are very similar to the RS shapes, leading us to conclude that the WS events are also charm-like (although they do not necessarily contain the desired D^* decay).

C.1.1 Event Quantities

Figure C.1 shows the jet multiplicity (after $N_{\rm jets} \geq 2$, before $N_{\rm jets} = 2$ cuts) and dijet opening angle (after $N_{\rm jets} = 2$ and $\Delta\phi > \frac{3}{4}\pi$ cuts) distributions for data and MC events.

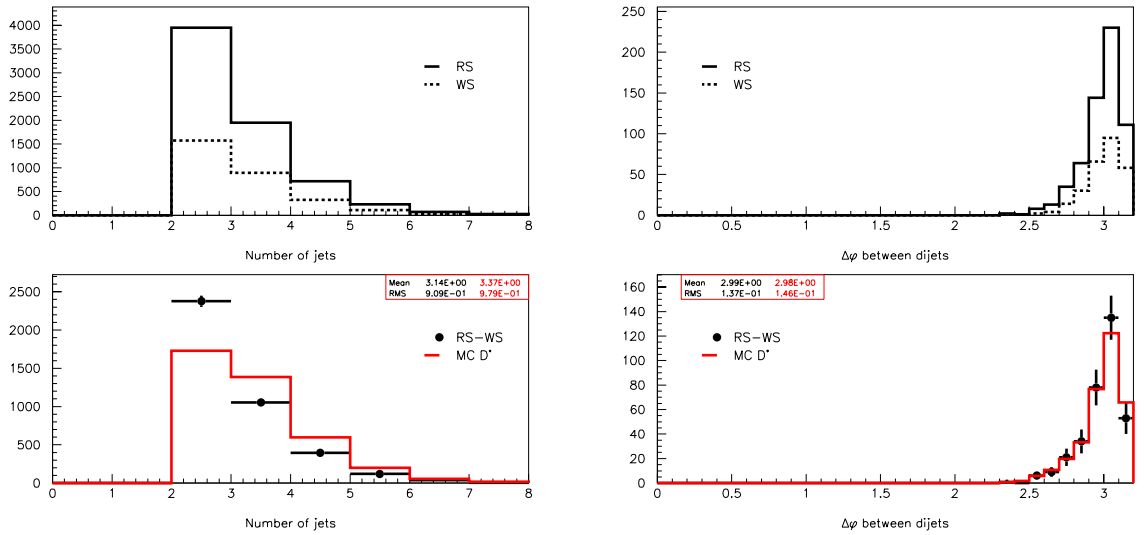


Figure C.1: Jet multiplicity and dijet opening angle distributions for events in the D^* control samples. Each upper plot shows distributions for RS and WS events in the data, and each lower plot shows the MC distribution compared and normalized to the data RS−WS difference distribution.

Jet flavor	Fraction (%) of taggable μ jets	away jets
b	14.6 ± 1.8	17.5 ± 3.2
c	85.4 ± 7.8	50.0 ± 7.9
p	0	32.5 ± 4.5

Table C.1: Estimated flavor fractions of taggable jets in the D^* data control sample.

C.1.2 Neural Network Variables

We multiply the flavor fractions of taggable jets measured in the D^* MC control sample by the appropriate taggable scale factors (s_{*t} from Table 4.4) to estimate these fractions in the data, shown in Table C.1. Then for each NN discrimination variable, we add the distributions for taggable b, c, and p MC jets in these ratios. Figures C.2 – C.5 show the summed MC distributions compared to data for the 14 NN discrimination variables. Figure C.6 shows similar distributions for the three NN output variables.

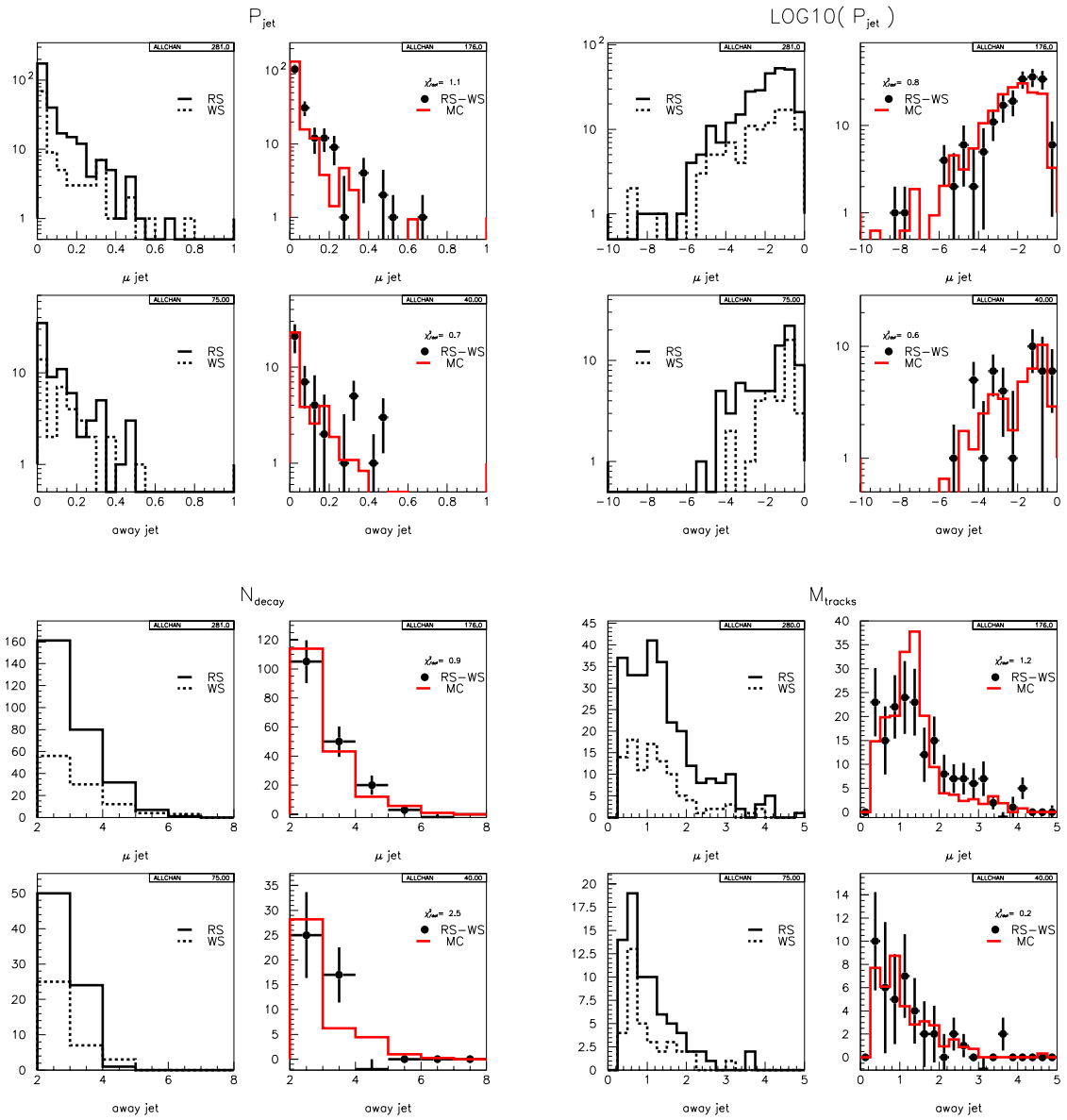


Figure C.2: Distributions of NN discrimination variables for taggable jets in the D^* control samples: jet probability $P_{\rm jet}$, $\log_{10} P_{\rm jet}$, decay track multiplicity $N_{\rm decay}$, and decay tracks mass $M_{\rm tracks}$. Each quantity has four plots organized as: muon jets (upper), away jets (lower), RS and WS data events (left), and the sum of MC b,c,p jet distributions added in the ratio of Table C.1, normalized to the RS−WS data (right). The plots include a reduced χ^2 measure of the goodness-of-fit between the RS data and the WS data + MC sum.

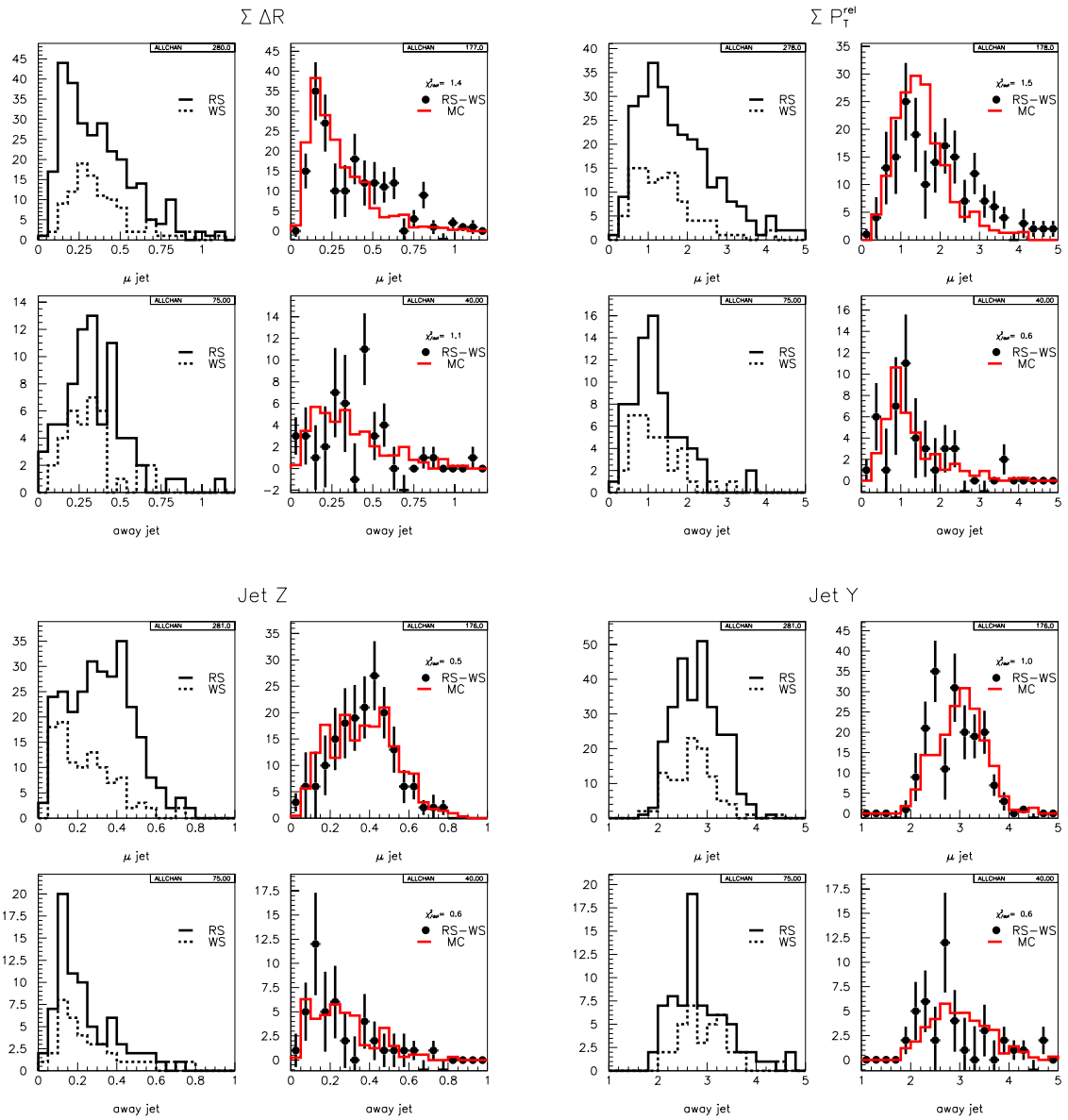


Figure C.3: Distributions of NN discrimination variables for taggable jets in the D^* control samples: decay track spread $\sum \Delta R$, decay track momentum $\perp$ jet axis $\sum P_T^{rel}$, fragmentation function Z, and jet rapidity Y. Each quantity has four plots organized as: muon jets (upper), away jets (lower), RS and WS data events (left), and the sum of MC b,c,p jet distributions added in the ratio of Table C.1, normalized to the RS$-$WS data (right). The plots include a reduced χ^2 measure of the goodness-of-fit between the RS data and the WS data + MC sum.

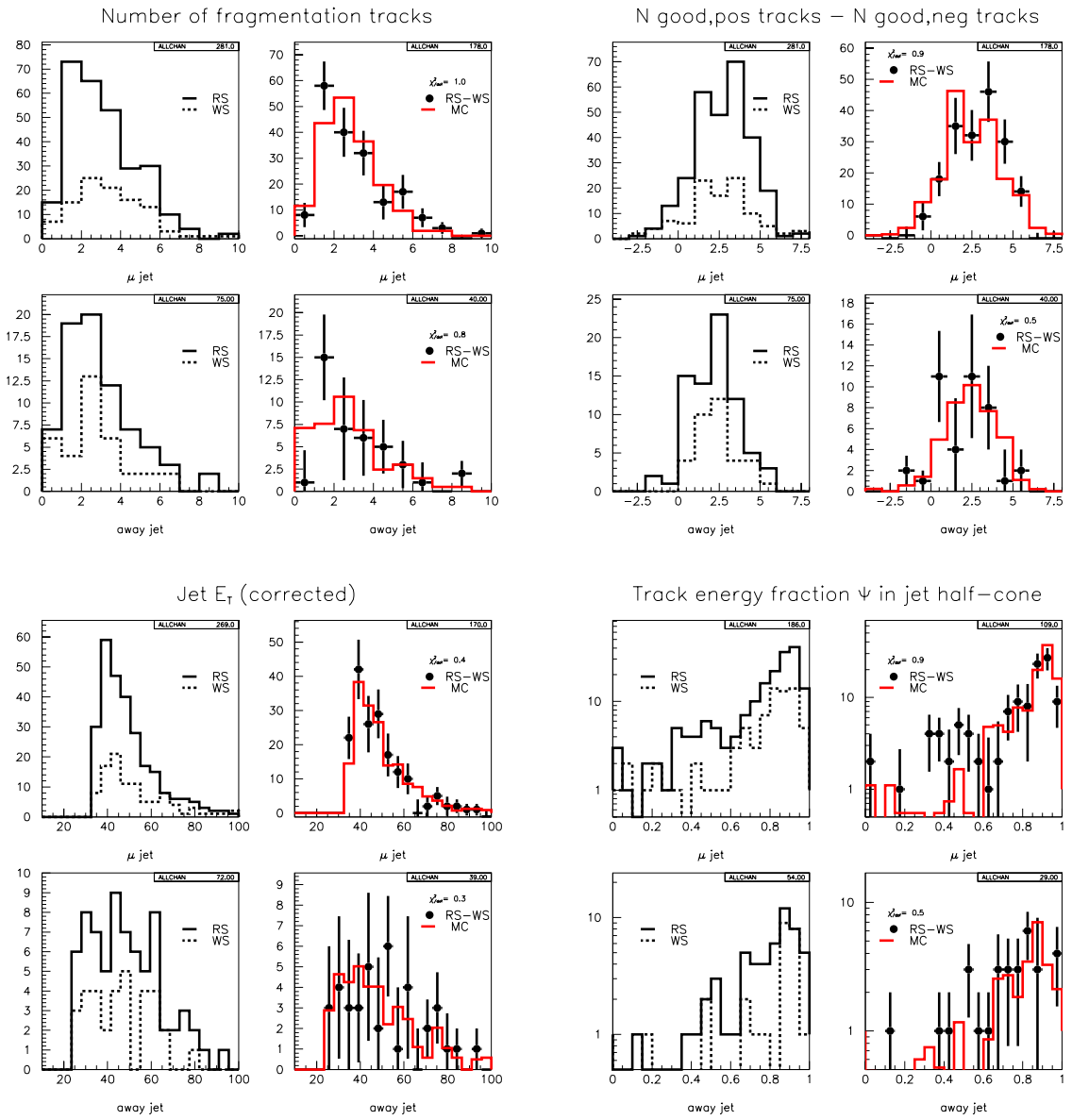


Figure C.4: Distributions of NN discrimination variables for taggable jets in the D^* control samples: number of fragmentation tracks, positive i.p. track excess, corrected jet E_T, and track energy fraction ψ in jet half-cone. Each quantity has four plots organized as: muon jets (upper), away jets (lower), RS and WS data events (left), and the sum of MC b,c,p jet distributions added in the ratio of Table C.1, normalized to the RS−WS data (right). The plots include a reduced χ^2 measure of the goodness-of-fit between the RS data and the WS data + MC sum.

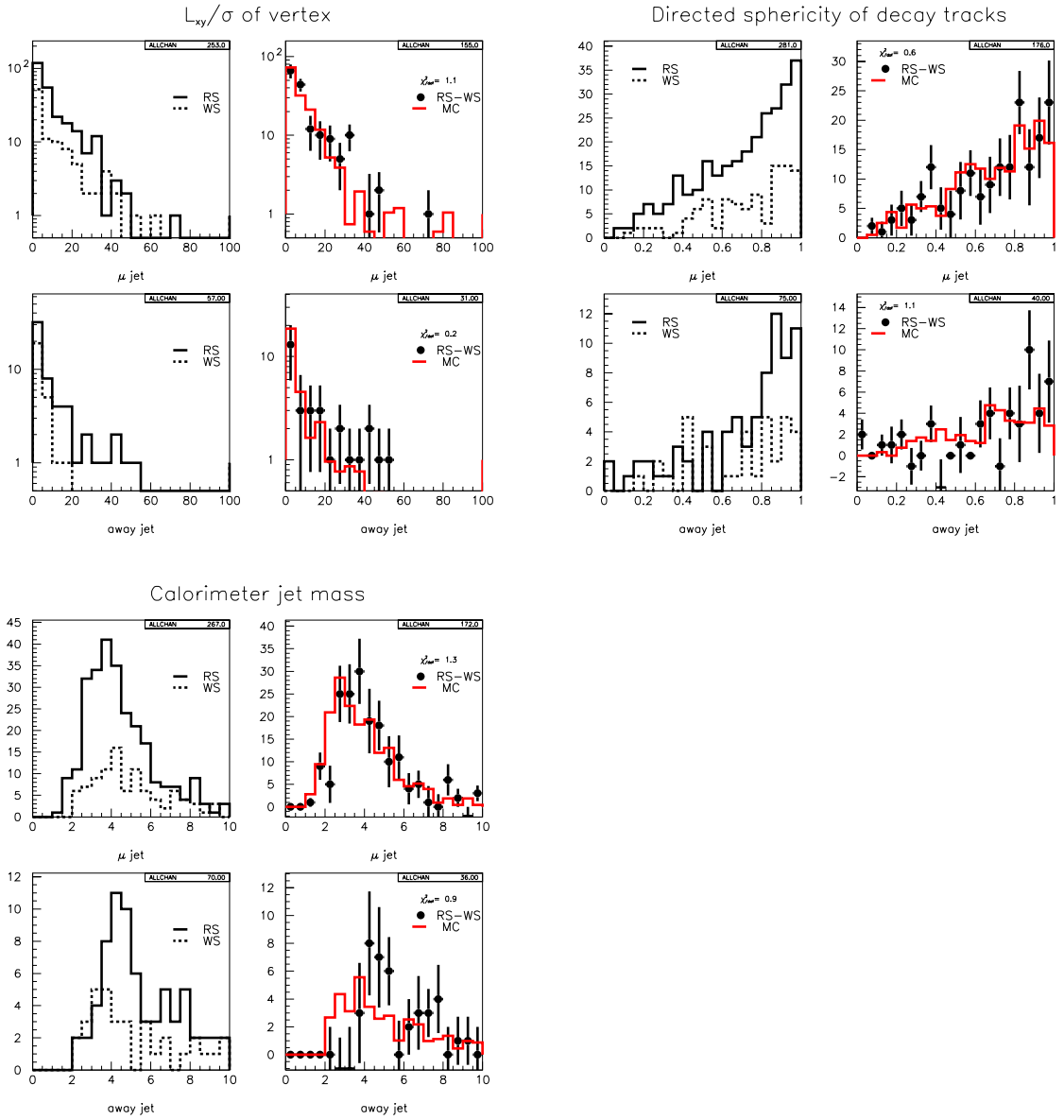


Figure C.5: Distributions of NN discrimination variables for taggable jets in the D^* control samples: secondary vertex significance $L_{xy}/\sigma_{L_{xy}}$, directed sphericity of decay tracks, and calorimeter jet mass. Each quantity has four plots organized as: muon jets (upper), away jets (lower), RS and WS data events (left), and the sum of MC b,c,p jet distributions added in the ratio of Table C.1, normalized to the RS−WS data (right). The plots include a reduced χ^2 measure of the goodness-of-fit between the RS data and the WS data + MC sum.

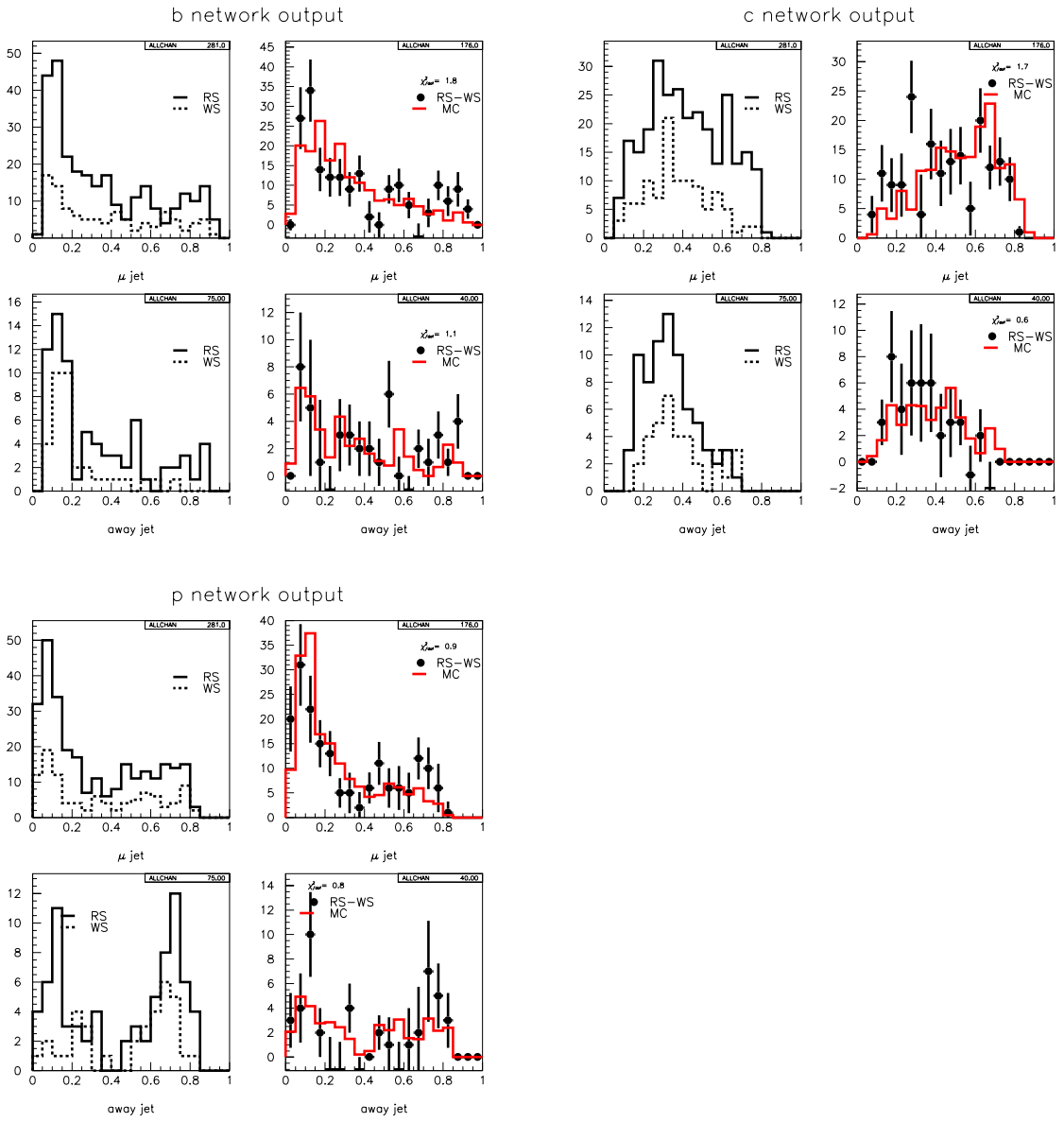


Figure C.6: Distributions of NN output variables for taggable jets in the D^* control samples: b-, c-, and p-ness. Each quantity has four plots organized as: muon jets (upper), away jets (lower), RS and WS data events (left), and the sum of MC b,c,p jet distributions added in the ratio of Table C.1, normalized to the RS$-$WS data (right). The plots include a reduced χ^2 measure of the goodness-of-fit between the RS data and the WS data + MC sum.

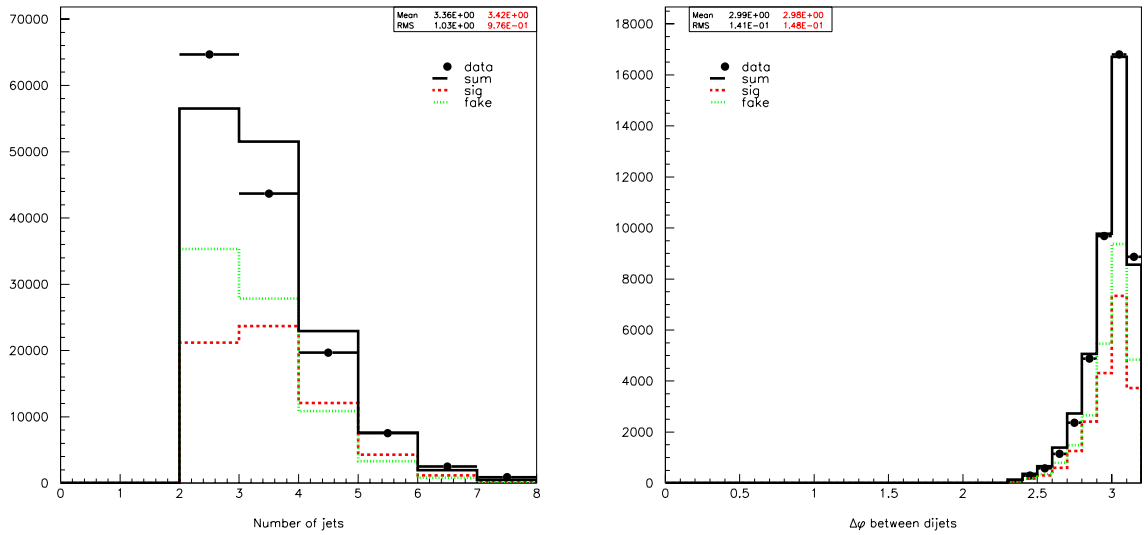


Figure C.7: Jet multiplicity and dijet opening angle distributions for data and MC events in the low-P_T muon control samples. Separate distributions for MC signal and MC fake muon events are shown in the ratio given by the fitted non-h.f. event fraction $\beta_{\rm nhf}$, and their sum is normalized to the data.

C.2 Low-P_T Inclusive Muon Events

As described in § 4.3.3, the scale factor fit returns both the scale factors and the non-h.f. event fraction in the low-P_T muon control sample, $\beta_{\rm nhf} = 0.56 \pm 0.02$. The parameter $\beta_{\rm nhf}$ gives the relative normalization of MC muon signal and MC fake muon events in § C.2.1, while the taggable jet scale factors are additionally required to determine the flavor mix of taggable jets in the NN variable plots of § C.2.2.

C.2.1 Event Quantities

Figure C.7 shows the jet multiplicity (after $N_{\rm jets} \geq 2$, before $N_{\rm jets} = 2$ cuts) and dijet opening angle (after $N_{\rm jets} = 2$ and $\Delta\phi > \frac{3}{4}\pi$ cuts) distributions for data and MC events. The MC sum has two components: MC muon signal and MC fake muon events, with relative weights $(1 - \beta_{\rm nhf})$ and $\beta_{\rm nhf}$, respectively.

C.2.2 Neural Network Variables

For each NN discrimination variable, we add the distributions for taggable b, c, and p MC jets in the ratio given in Table 4.7. Figures C.8 – C.11 show these distributions, along with their sums compared to data, for the 14 NN discrimination variables. Figure C.12 shows similar distributions for the three NN output variables.

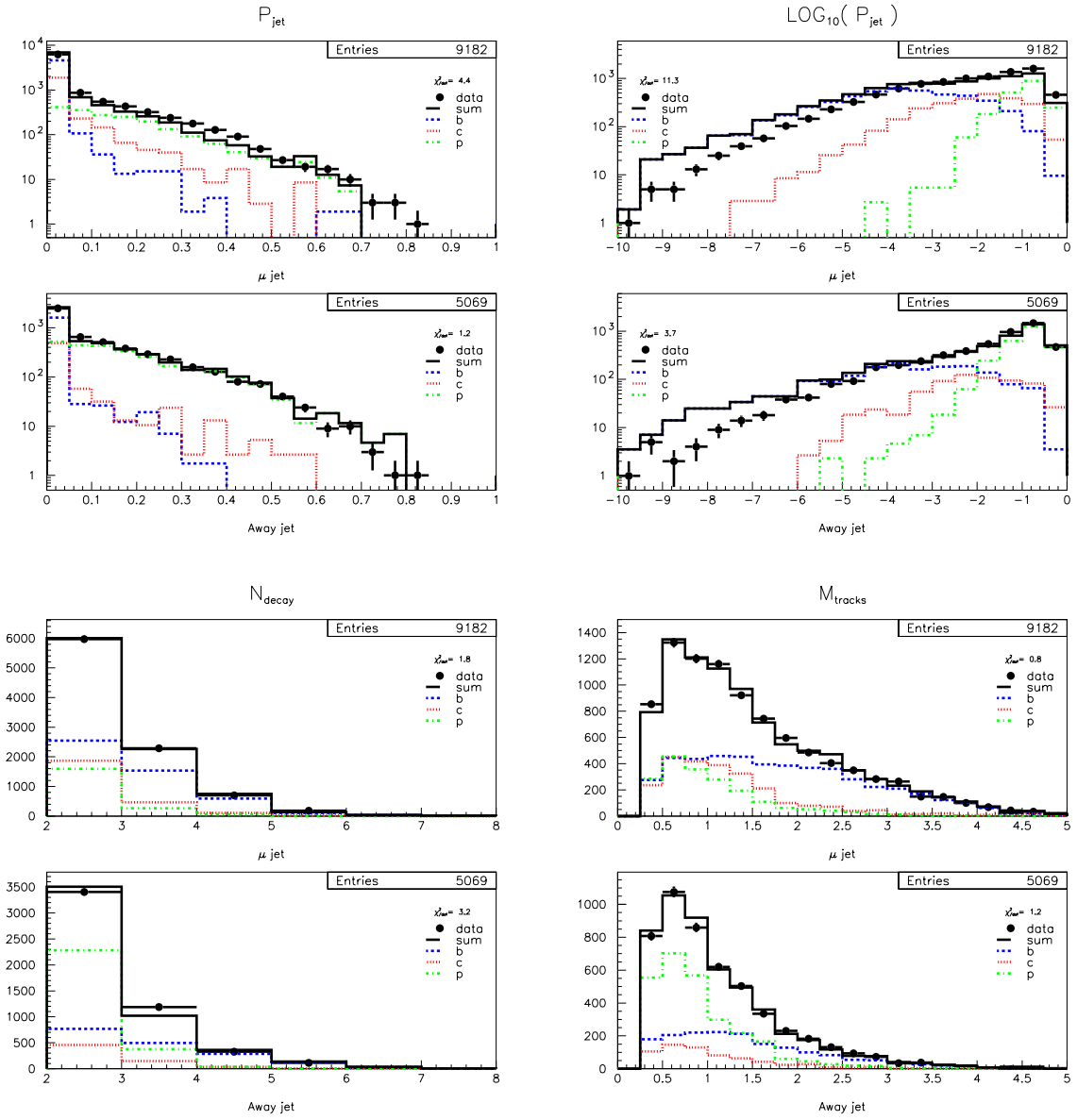


Figure C.8: Distributions of NN discrimination variables for taggable muon and away jets in the low-P_T muon control samples: jet probability $P_{\rm jet}$, $\log_{10} P_{\rm jet}$, decay track multiplicity $N_{\rm decay}$, and decay tracks mass $M_{\rm tracks}$. Separate MC distributions are shown for b,c,p jets in the ratio given in Table 4.7, and their sum is normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

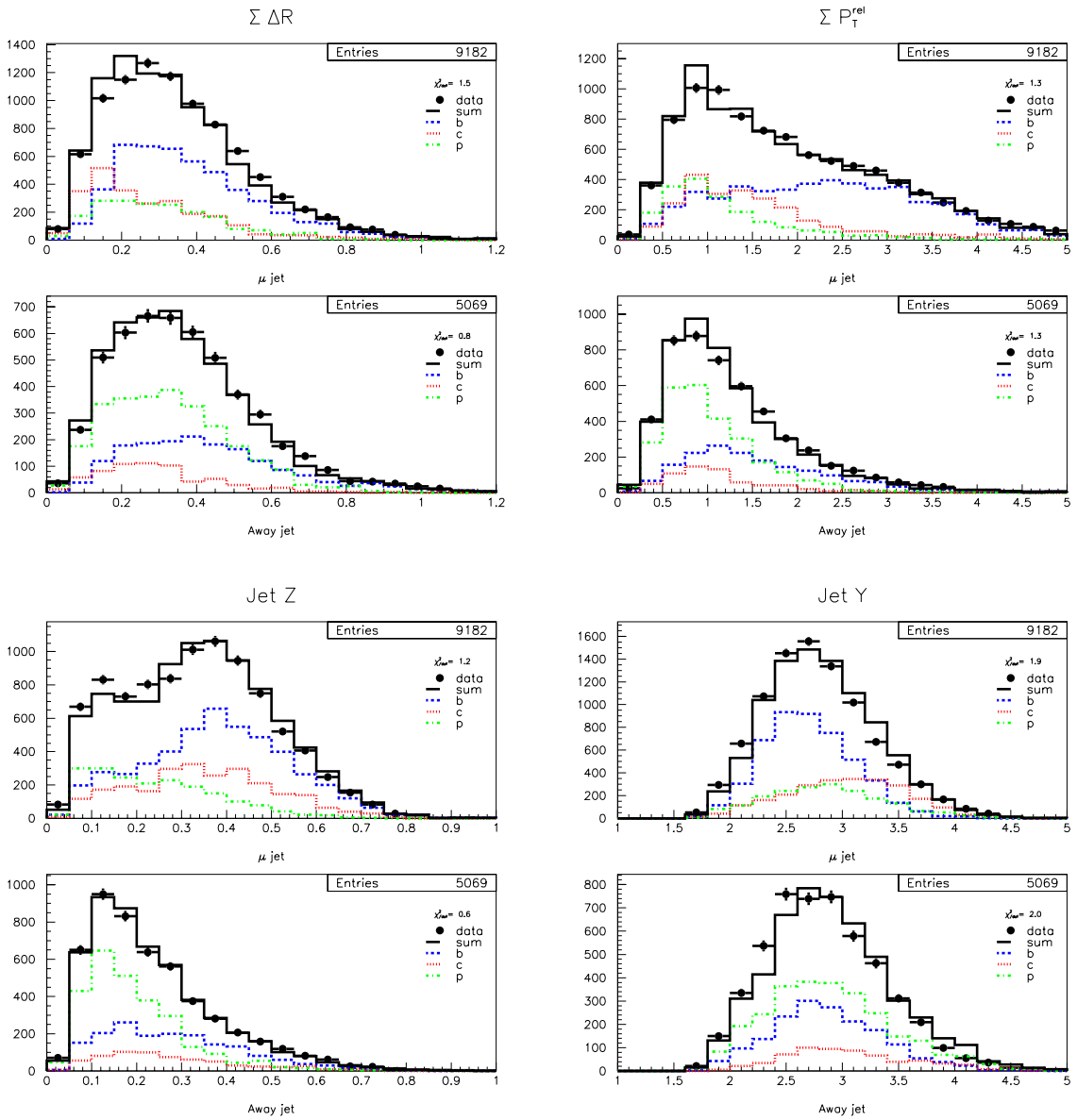


Figure C.9: Distributions of NN discrimination variables for taggable muon and away jets in the low-P_T muon control samples: decay track spread $\sum \Delta R$, decay track momentum $\perp$ jet axis $\sum P_T^{rel}$, fragmentation function Z, and jet rapidity Y. Separate MC distributions are shown for b,c,p jets in the ratio given in Table 4.7, and their sum is normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

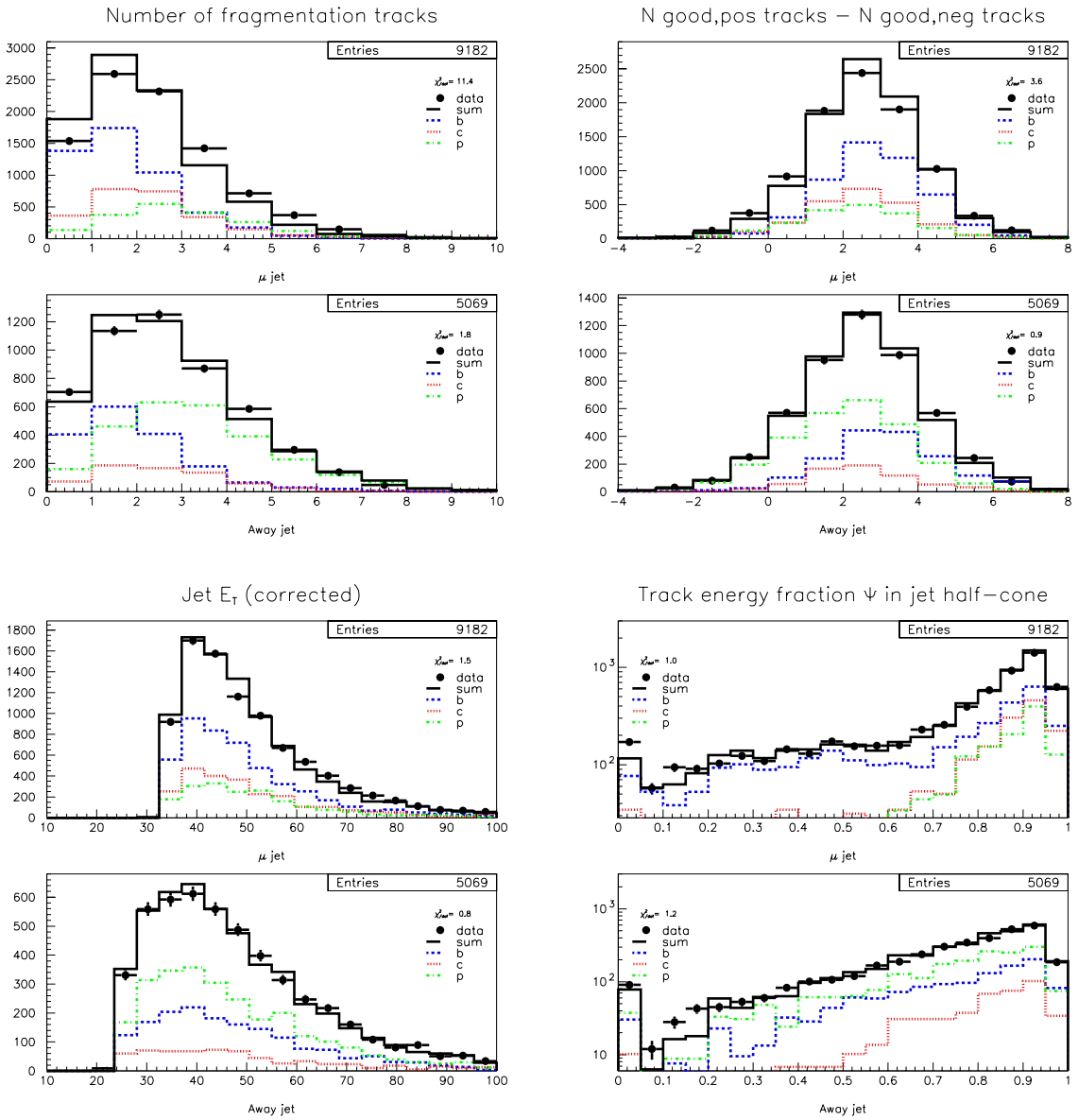


Figure C.10: Distributions of NN discrimination variables for taggable muon and away jets in the low-P_T muon control samples: number of fragmentation tracks, positive i.p. track excess, corrected jet E_T, and track energy fraction ψ in jet half-cone. Separate MC distributions are shown for b,c,p jets in the ratio given in Table 4.7, and their sum is normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

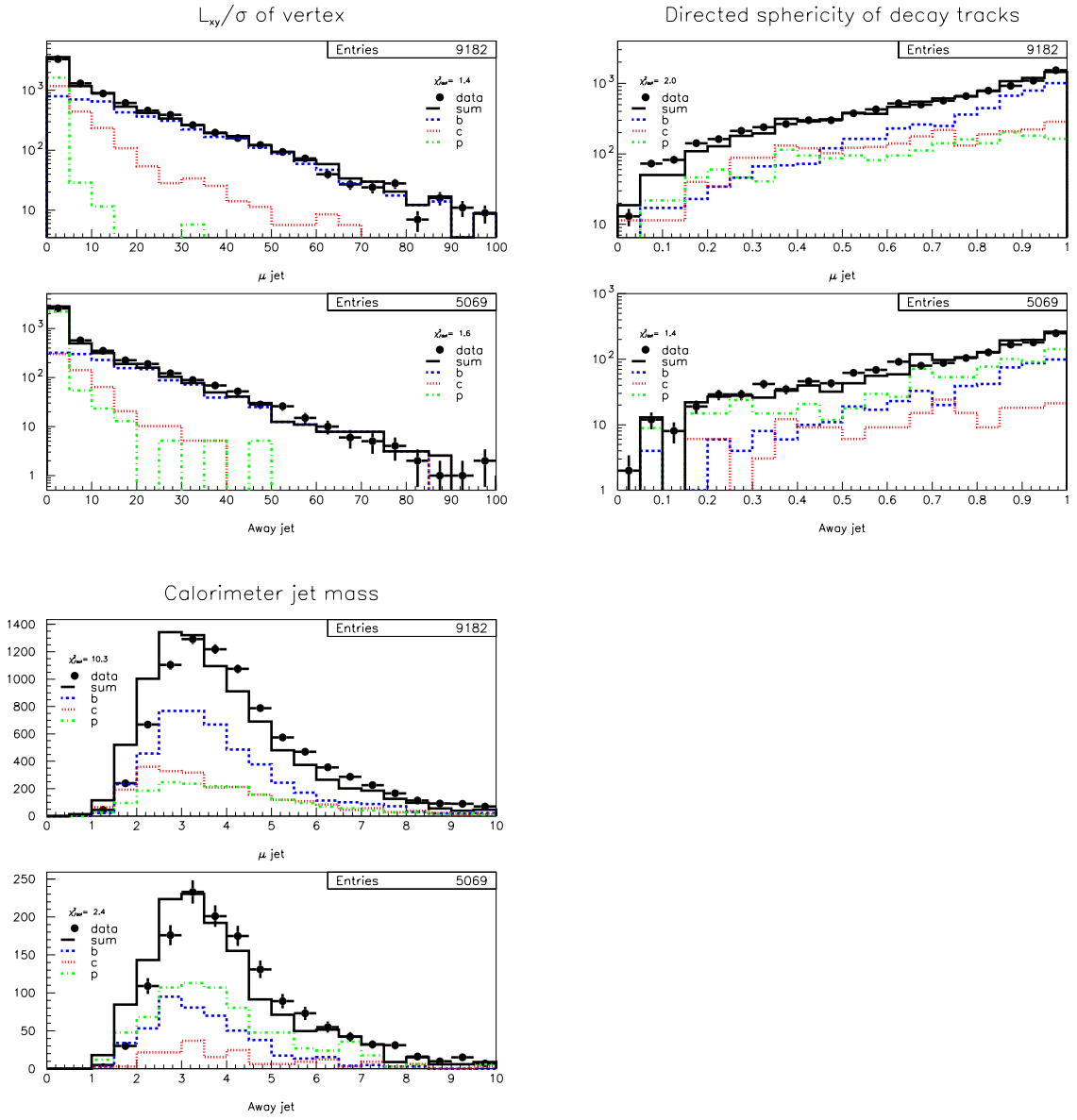


Figure C.11: Distributions of NN discrimination variables for taggable muon and away jets in the low-P_T muon control samples: secondary vertex significance $L_{xy}/\sigma_{L_{xy}}$, directed sphericity of decay tracks, and calorimeter jet mass. Separate MC distributions are shown for b,c,p jets in the ratio given in Table 4.7, and their sum is normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

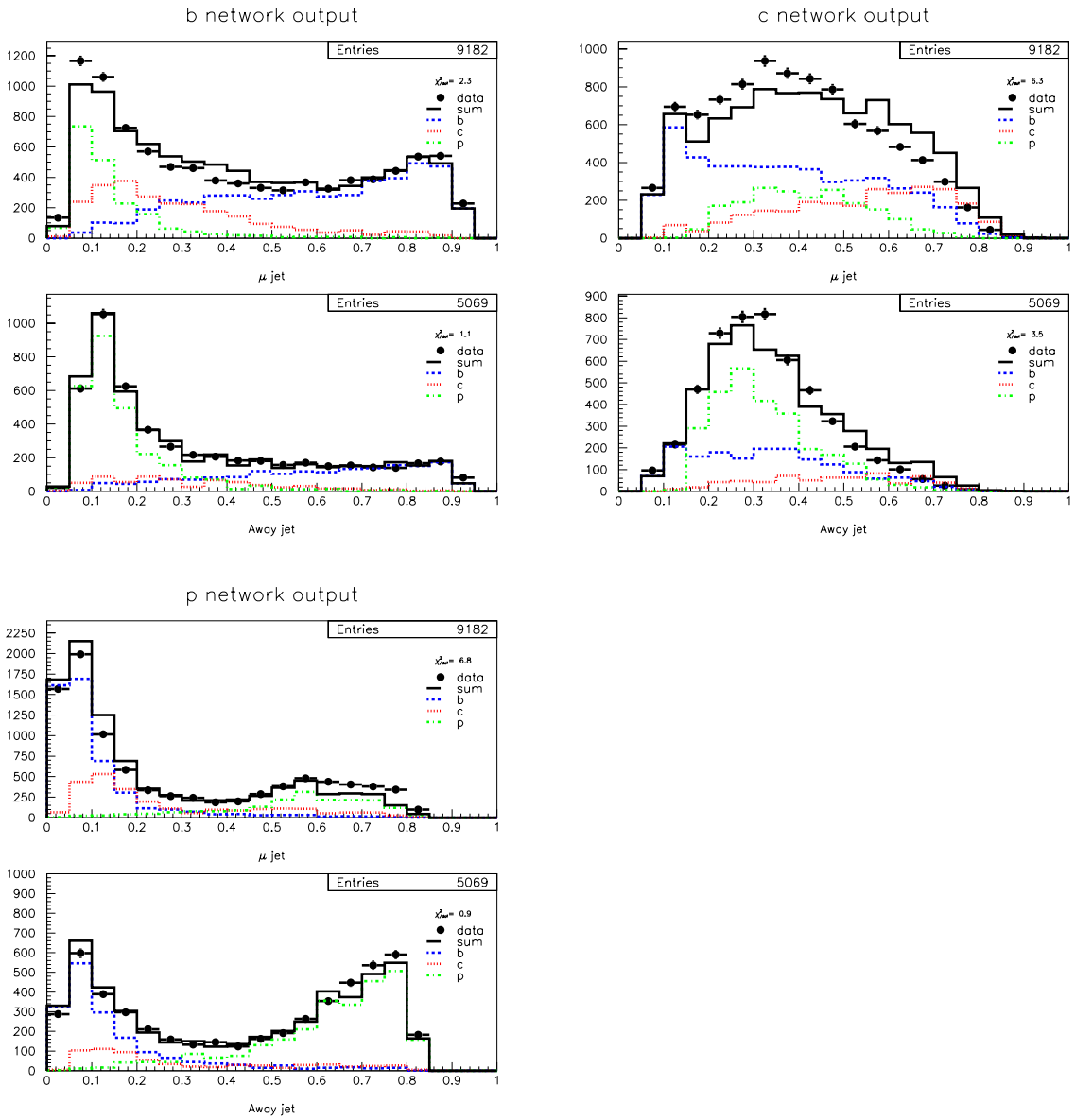


Figure C.12: Distributions of NN output variables for taggable muon and away jets in the low-P_T muon control samples: b-, c-, and p-ness. Separate MC distributions are shown for b,c,p jets in the ratio given in Table 4.7, and their sum is normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

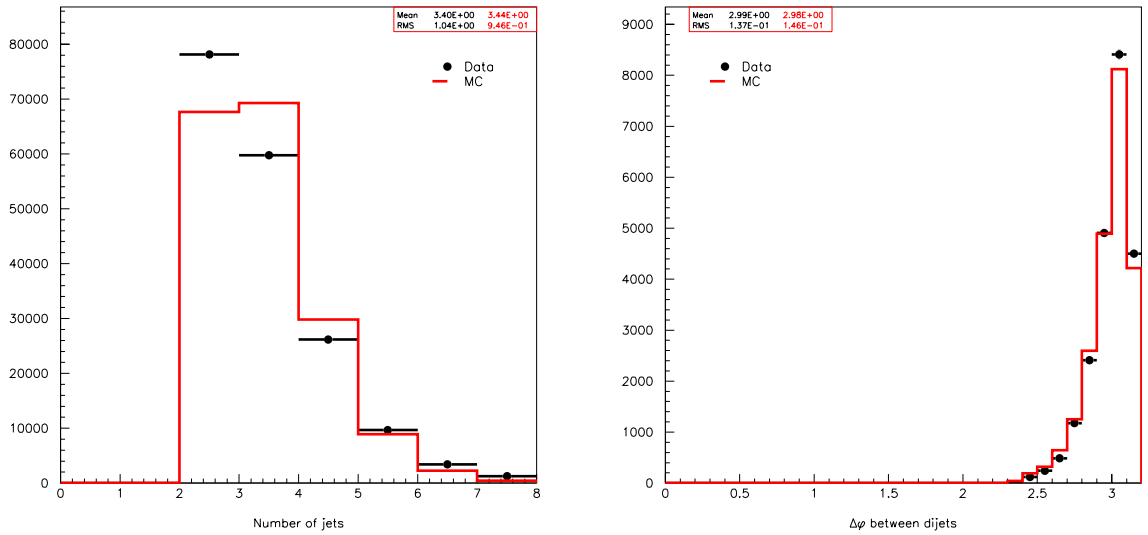


Figure C.13: Jet multiplicity and dijet opening angle distributions for data and MC events in the inclusive jets control samples. The MC distributions are normalized to the data.

Jet flavor	Fraction (%) of taggable jets
b	9.8 ± 0.9
c	14.7 ± 1.8
p	75.5 ± 3.4

Table C.2: Estimated flavor fractions of taggable jets in the inclusive jets data control sample.

C.3 Inclusive Jets Events

C.3.1 Event Quantities

Figure C.13 shows the jet multiplicity (after $N_{\text{jets}} \geq 2$, before $N_{\text{jets}} = 2$ cuts) and dijet opening angle (after $N_{\text{jets}} = 2$ and $\Delta\phi > \frac{3}{4}\pi$ cuts) distributions for data and MC events.

C.3.2 Neural Network Variables

We multiply the flavor fractions of taggable jets measured in the inclusive jets MC control sample by the appropriate taggable scale factors (s_{*t} from Table 4.4) to estimate

these fractions in the data, shown in Table C.2. Then for each NN discrimination variable, we add the distributions for taggable b, c, and p MC jets in these ratios. Figures C.14 – C.17 show the summed MC distributions compared to data for the 14 NN discrimination variables. Figure C.18 shows similar distributions for the three NN output variables.

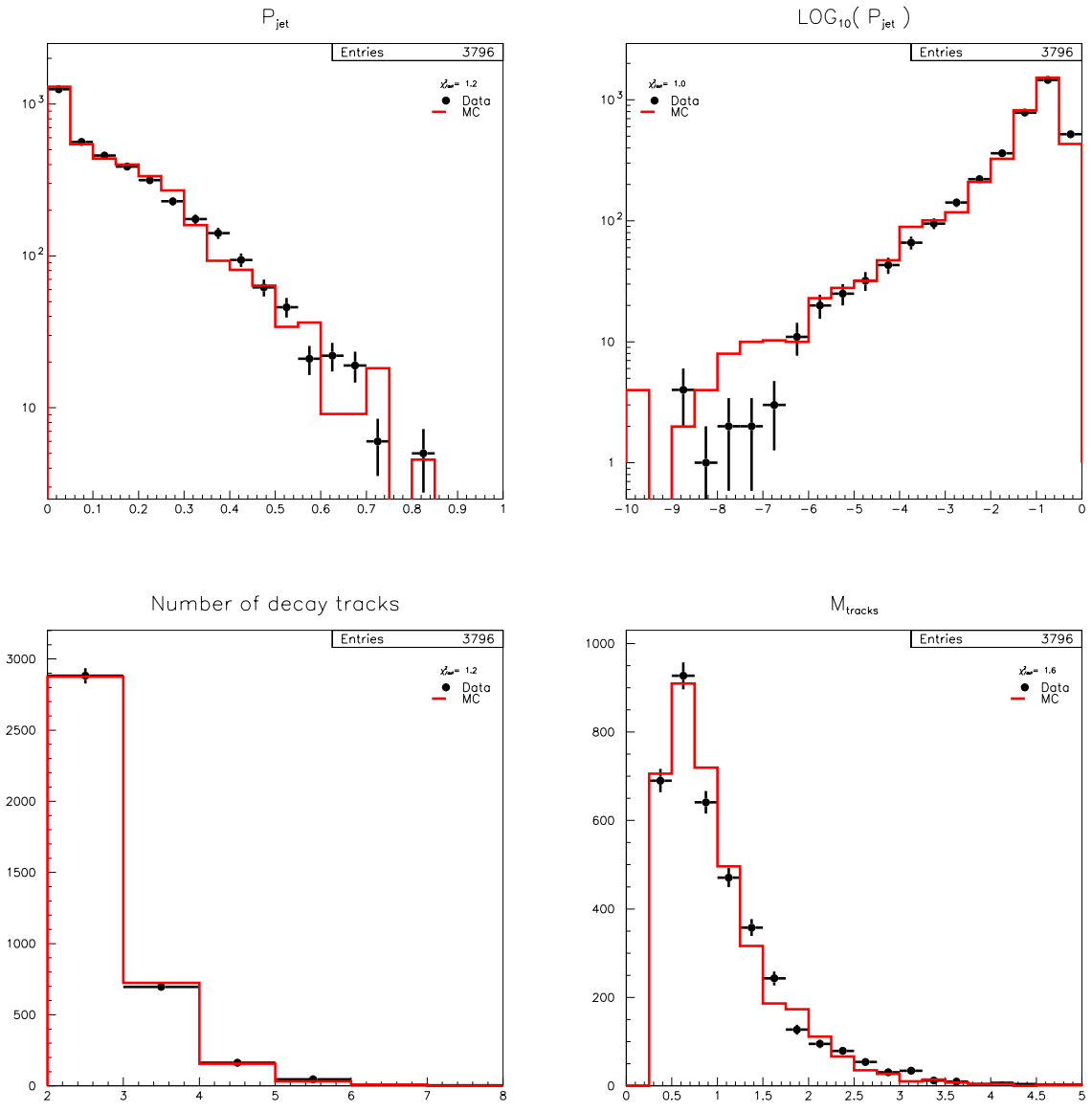


Figure C.14: Distributions of NN discrimination variables for taggable jets in the inclusive jets control samples: jet probability $P_{\rm jet}$, $\log_{10} P_{\rm jet}$, decay track multiplicity $N_{\rm decay}$, and decay tracks mass $M_{\rm tracks}$. Each MC distribution is the sum of b,c,p jet distributions added in the ratios of Table C.2, normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

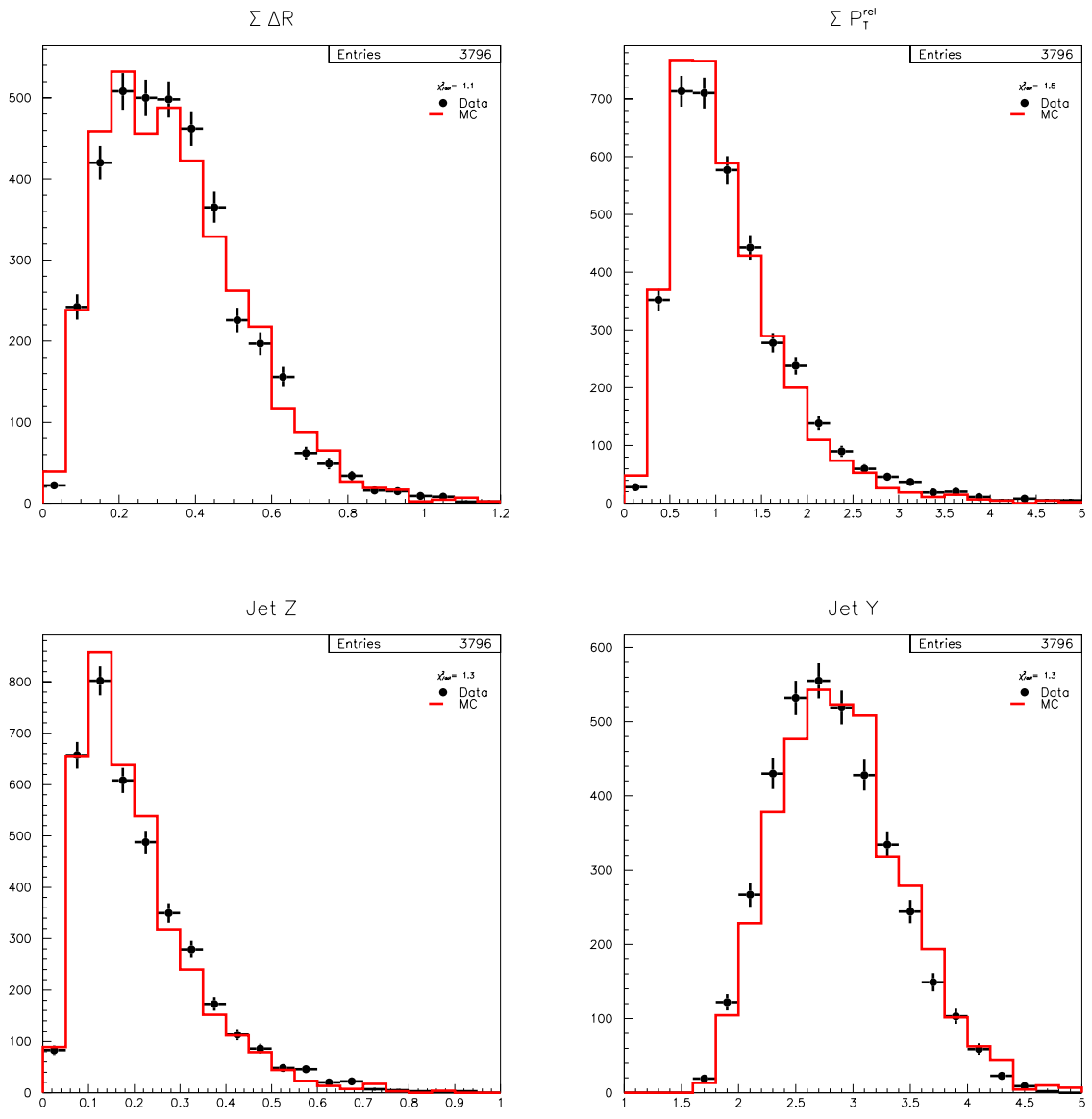


Figure C.15: Distributions of NN discrimination variables for taggable jets in the inclusive jets control samples: decay track spread $\sum \Delta R$, decay track momentum $\perp$ jet axis $\sum P_T^{rel}$, fragmentation function Z, and jet rapidity Y. Each MC distribution is the sum of b,c,p jet distributions added in the ratios of Table C.2, normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

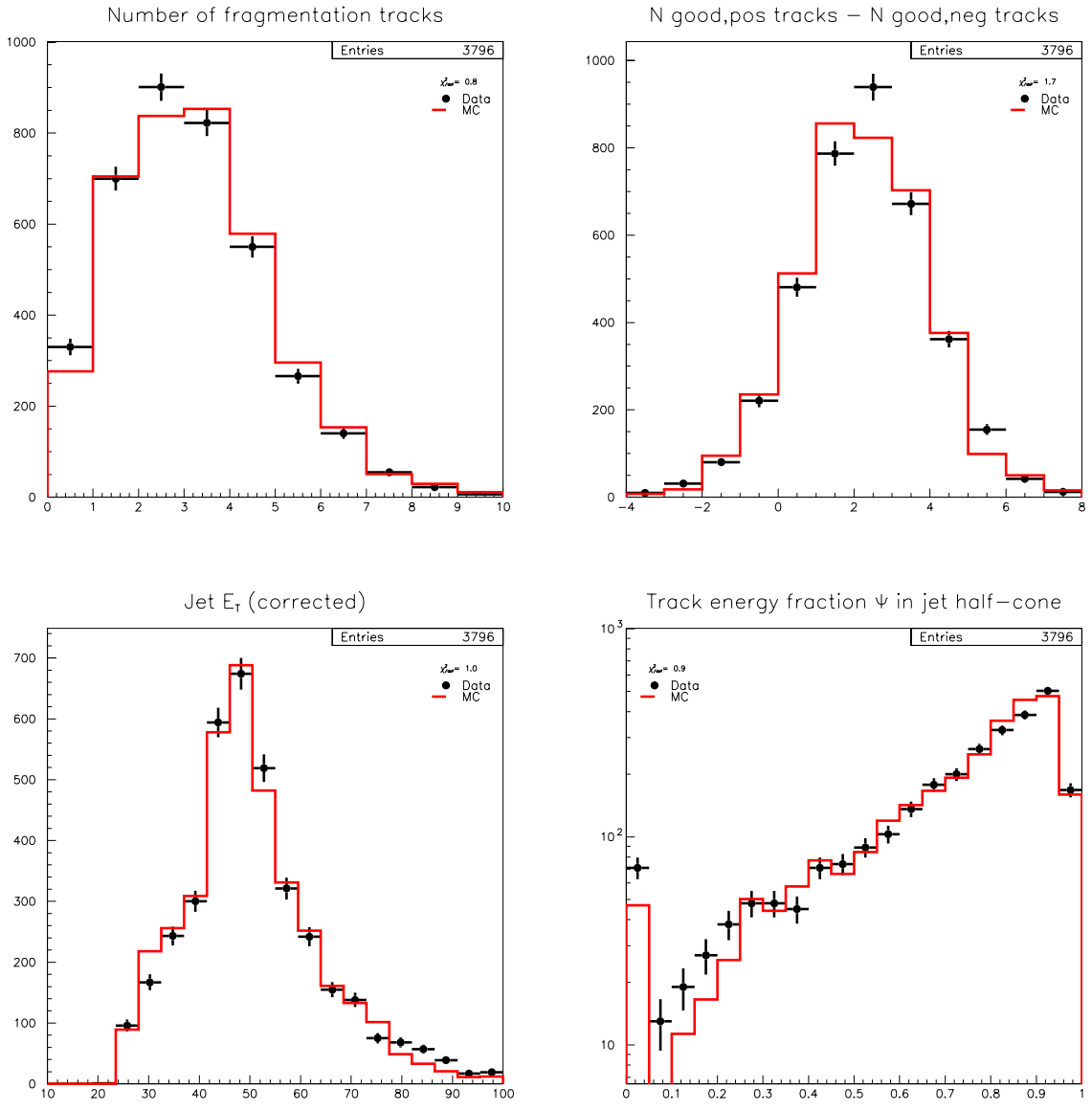


Figure C.16: Distributions of NN discrimination variables for taggable jets in the inclusive jets control samples: number of fragmentation tracks, positive i.p. track excess, corrected jet E_T, and track energy fraction ψ in jet half-cone. Each MC distribution is the sum of b,c,p jet distributions added in the ratios of Table C.2, normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

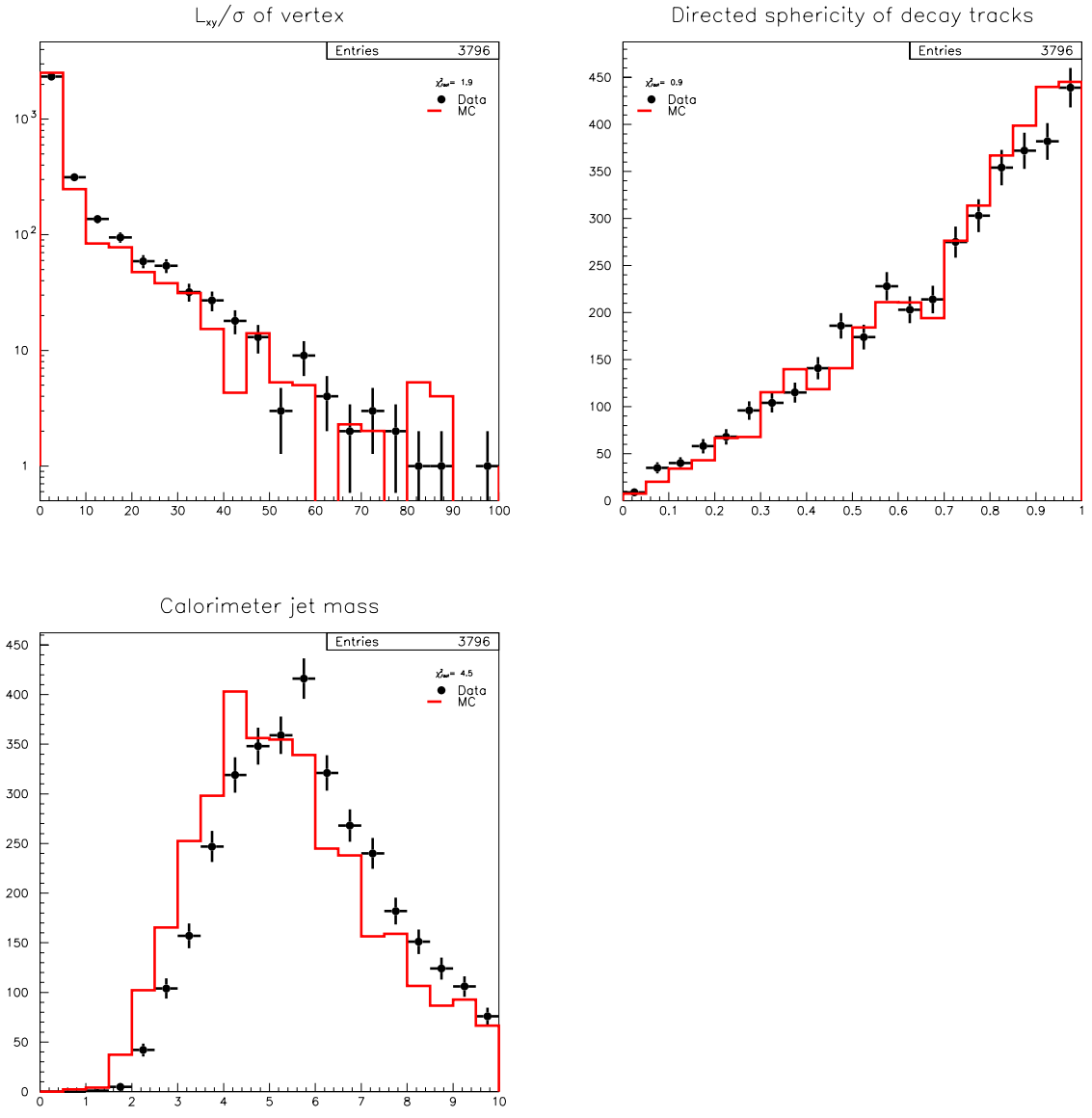


Figure C.17: Distributions of NN discrimination variables for taggable jets in the inclusive jets control samples: secondary vertex significance $L_{xy}/\sigma_{L_{xy}}$, directed sphericity of decay tracks, and calorimeter jet mass. Each MC distribution is the sum of b,c,p jet distributions added in the ratios of Table C.2, normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

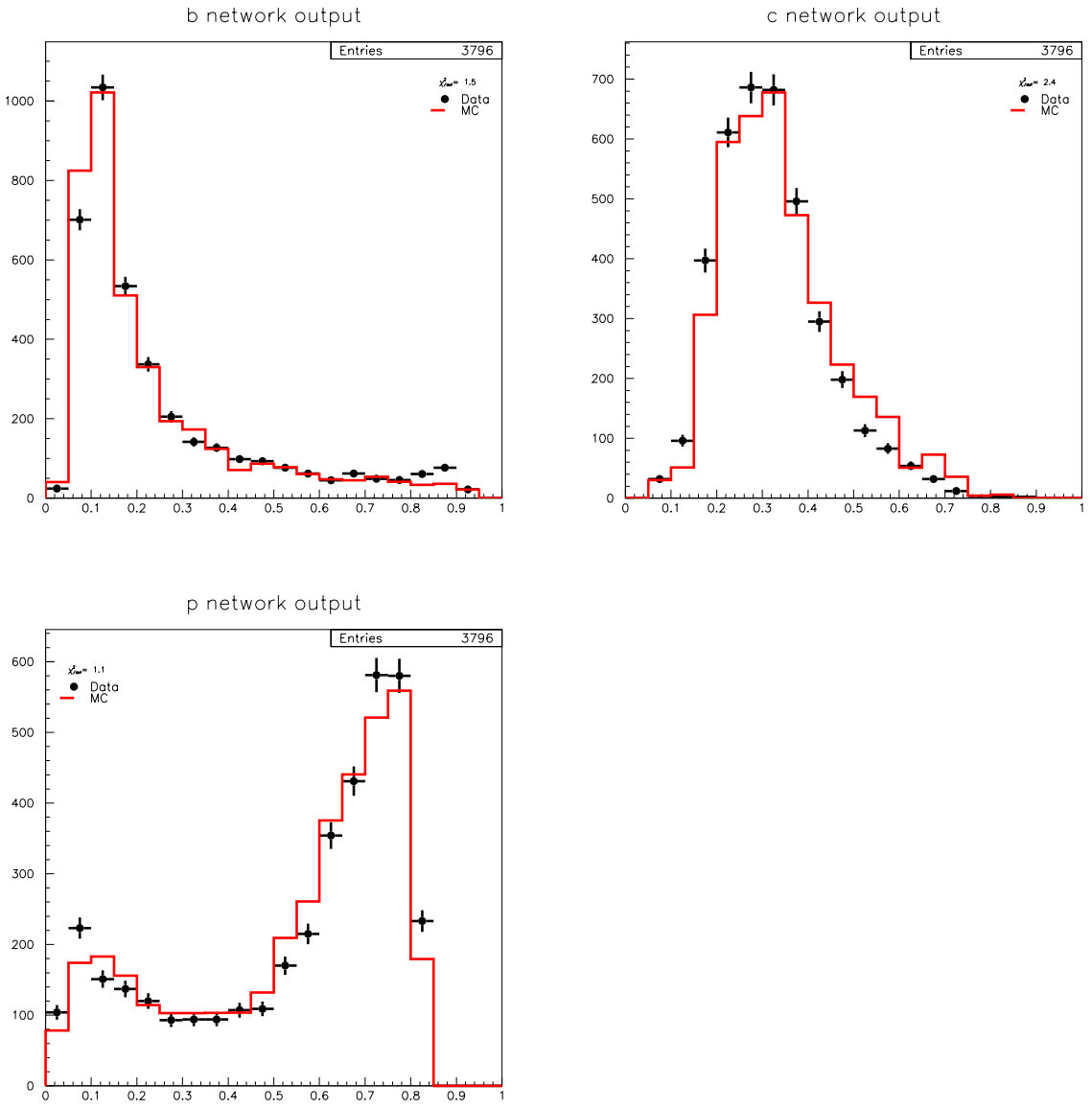


Figure C.18: Distributions of NN output variables for taggable jets in the inclusive jets control samples: b-, c-, and p-ness. Each MC distribution is the sum of b,c,p jet distributions added in the ratios of Table C.2, normalized to the data. Each plot includes a reduced χ^2 measure of the goodness-of-fit between the data and MC sum.

C.4 Remarks For All Control Samples

Data events tend to have a lower jet multiplicity than MC events in each control sample. We minimize the effect of this disagreement by using only dijet events in the scale factor measurement. The dijet opening angle distributions show that these dijets are mostly back-to-back in each control sample.

We see generally good agreement between the data and MC for all neural network variables. Note that because of the MC shortcomings mentioned in § 4.1.1 and our measurement of non-unity scale factors, we expect a small level of disagreement anyway.

Appendix D

Flavor-Tagging Statistics For Taggers in $t \to cg$ Search

This Appendix lists tagging statistics tables for each of the 10 NN taggers listed in Table 7.1. The tables appear in approximate order of increasingly strict decay track cuts for taggable jets. In each table, the dijet mass associated with the c-tag has a range of 140 GeV.

For most NN taggers, the observed numbers of tagged events is consistent (within error) with the predicted sum of backgrounds and SM-SM $t\bar{t}$ events. The largest exceptions are: 2.4-σ b-tag excess and 1.9-σ bc-tag excess with tagger "2a", and 1.4-σ c-tag excess with tagger "1a".

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.3155	0.56	0.0850	0.15	0.2036	0.36
	±0.37	±0.0190	±0.12	±0.0096	±0.04	±0.0209	±0.08
Single top (W^*)	0.53	0.3728	0.20	0.1262	0.07	0.1965	0.11
	±0.09	±0.0272	±0.04	±0.0170	±0.01	±0.0277	±0.02
Dibosons	6.27	0.0416	0.26	0.0115	0.07	0.0785	0.49
	±0.82	±0.0205	±0.13	±0.0115	±0.07	±0.0269	±0.18
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0151	0.19	0.0000	0.00	0.0796	0.98
	±2.58	±0.0068	±0.09	±0.0000	±0.00	±0.0161	±0.28
Wc	12.38	0.1123	1.39	0.0177	0.22	0.0874	1.08
	±3.11	±0.0247	±0.46	±0.0091	±0.13	±0.0211	±0.38
Wbb	6.99	0.3514	2.45	0.1039	0.73	0.1904	1.33
	±1.11	±0.0336	±0.45	±0.0277	±0.22	±0.0452	±0.38
Wcc	13.04	0.1610	2.10	0.0469	0.61	0.1688	2.20
	±2.55	±0.0337	±0.60	±0.0173	±0.26	±0.0343	±0.62
W + non-h.f.	132.32	0.0300	3.97	0.0049	0.65	0.0991	13.11
	±9.89	±0.0125	±1.67	±0.0049	±0.65	±0.0189	±2.69
Non-W e	12.03	0.1189	1.43	0.0274	0.33	0.0756	0.91
	±2.43		±0.50		±0.17		±0.35
Non-W μ	11.72	0.1519	1.78	0.0162	0.19	0.1058	1.24
	±2.36		±0.58		±0.11		±0.44
$\sum$ Backgrounds	212.86		14.33		3.01		21.80
			±2.77		±1.41		±3.31
SM-SM $t\bar{t}$	31.00	0.3694	11.45	0.1813	5.62	0.3281	10.17
	±7.07	±0.0146	±2.65	±0.0104	±1.32	±0.0190	±2.39
SM-FCNC $t\bar{t}$		0.2797		0.1707		0.4239	
		±0.0111		±0.0088		±0.0142	
$\sum$ Backgrounds	243.86		25.78		8.63		31.97
+ SM-SM $t\bar{t}$			±3.83		±1.93		±4.08
Data	244.00		38.00		11.00		42.00
Tag excess			12.22		2.37		10.03
			±6.36		±3.52		±6.97

Table D.1: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "1a", where a taggable jet requires ≥ 1 positive i.p. decay track.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2753	0.49	0.0325	0.06	0.0760	0.13
	±0.37	±0.0156	±0.10	±0.0054	±0.02	±0.0098	±0.03
Single top (W^*)	0.53	0.3231	0.17	0.0455	0.02	0.0886	0.05
	±0.09	±0.0246	±0.03	±0.0080	±0.01	±0.0140	±0.01
Dibosons	6.27	0.0228	0.14	0.0064	0.04	0.0179	0.11
	±0.82	±0.0134	±0.09	±0.0064	±0.04	±0.0126	±0.08
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0107	0.13	0.0000	0.00	0.0068	0.08
	±2.58	±0.0054	±0.07	±0.0000	±0.00	±0.0048	±0.06
Wc	12.38	0.0583	0.72	0.0040	0.05	0.0493	0.61
	±3.11	±0.0178	±0.29	±0.0041	±0.05	±0.0149	±0.24
Wbb	6.99	0.3096	2.16	0.0366	0.26	0.0819	0.57
	±1.11	±0.0335	±0.42	±0.0120	±0.09	±0.0214	±0.18
Wcc	13.04	0.1067	1.39	0.0168	0.22	0.0955	1.25
	±2.55	±0.0298	±0.47	±0.0090	±0.13	±0.0229	±0.39
W + non-h.f.	132.32	0.0126	1.67	0.0000	0.00	0.0118	1.56
	±9.89	±0.0091	±1.22	±0.0000	±0.00	±0.0068	±0.91
Non-W e	12.03	0.0889	1.07	0.0083	0.10	0.0349	0.42
	±2.43		±0.40		±0.08		±0.20
Non-W μ	11.72	0.1058	1.24	0.0085	0.10	0.0461	0.54
	±2.36		±0.44		±0.07		±0.24
$\sum$ Backgrounds	212.86		9.19		0.84		5.33
			±2.04		±0.87		±1.61
SM-SM $t\bar{t}$	31.00	0.3241	10.05	0.0618	1.92	0.1283	3.98
	±7.07	±0.0126	±2.33	±0.0048	±0.46	±0.0095	±0.95
SM-FCNC $t\bar{t}$		0.2307		0.0494		0.1585	
		±0.0093		±0.0037		±0.0084	
$\sum$ Backgrounds	243.86		19.24		2.76		9.31
+ SM-SM $t\bar{t}$			±3.10		±0.98		±1.87
Data	244.00		27.00		3.00		14.00
Tag excess			7.76		0.24		4.69
			±5.37		±1.93		±3.58

Table D.2: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "1b", where a taggable jet requires ≥ 1 positive i.p. decay track with track probability $< 10\%$.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2629	0.46	0.0360	0.06	0.0965	0.17
	±0.37	±0.0172	±0.10	±0.0060	±0.02	±0.0126	±0.04
Single top (W^*)	0.53	0.3103	0.17	0.0586	0.03	0.1016	0.05
	±0.09	±0.0245	±0.03	±0.0101	±0.01	±0.0166	±0.01
Dibosons	6.27	0.0197	0.12	0.0071	0.04	0.0410	0.26
	±0.82	±0.0116	±0.07	±0.0072	±0.05	±0.0202	±0.13
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0130	0.16	0.0000	0.00	0.0445	0.55
	±2.58	±0.0065	±0.09	±0.0000	±0.00	±0.0122	±0.19
Wc	12.38	0.0539	0.67	0.0067	0.08	0.0905	1.12
	±3.11	±0.0161	±0.26	±0.0050	±0.06	±0.0223	±0.39
Wbb	6.99	0.2944	2.06	0.0475	0.33	0.0729	0.51
	±1.11	±0.0333	±0.40	±0.0153	±0.12	±0.0232	±0.18
Wcc	13.04	0.0987	1.29	0.0214	0.28	0.1164	1.52
	±2.55	±0.0272	±0.44	±0.0105	±0.15	±0.0275	±0.47
W + non-h.f.	132.32	0.0056	0.74	0.0000	0.00	0.0434	5.74
	±9.89	±0.0056	±0.74	±0.0000	±0.00	±0.0124	±1.70
Non-W e	12.03	0.1081	1.30	0.0108	0.13	0.0515	0.62
	±2.43		±0.46		±0.09		±0.26
Non-W μ	11.72	0.1032	1.21	0.0085	0.10	0.0512	0.60
	±2.36		±0.43		±0.07		±0.26
$\sum$ Backgrounds	212.86		8.17		1.06		11.14
			±1.78		±0.95		±2.27
SM-SM $t\bar{t}$	31.00	0.3085	9.56	0.0880	2.73	0.1818	5.64
	±7.07	±0.0160	±2.24	±0.0071	±0.66	±0.0134	±1.35
SM-FCNC $t\bar{t}$		0.2058		0.0766		0.2553	
		±0.0106		±0.0061		±0.0132	
$\sum$ Backgrounds	243.86		17.73		3.79		16.78
+ SM-SM $t\bar{t}$			±2.86		±1.16		±2.64
Data	244.00		30.00		8.00		21.00
Tag excess			12.27		4.21		4.22
			±5.09		±2.26		±4.87

Table D.3: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "2a", where a taggable jet requires ≥ 2 positive i.p. decay tracks.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2614	0.46	0.0227	0.04	0.0520	0.09
	±0.37	±0.0165	±0.10	±0.0045	±0.01	±0.0082	±0.02
Single top (W^*)	0.53	0.2975	0.16	0.0436	0.02	0.0754	0.04
	±0.09	±0.0225	±0.03	±0.0087	±0.01	±0.0139	±0.01
Dibosons	6.27	0.0204	0.13	0.0066	0.04	0.0202	0.13
	±0.82	±0.0122	±0.08	±0.0066	±0.04	±0.0142	±0.09
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0135	0.17	0.0000	0.00	0.0101	0.12
	±2.58	±0.0068	±0.09	±0.0000	±0.00	±0.0059	±0.08
Wc	12.38	0.0337	0.42	0.0033	0.04	0.0548	0.68
	±3.11	±0.0129	±0.19	±0.0034	±0.04	±0.0165	±0.27
Wbb	6.99	0.2907	2.03	0.0350	0.24	0.0701	0.49
	±1.11	±0.0320	±0.39	±0.0125	±0.10	±0.0208	±0.16
Wcc	13.04	0.0730	0.95	0.0126	0.16	0.1140	1.49
	±2.55	±0.0250	±0.38	±0.0082	±0.11	±0.0277	±0.46
W + non-h.f.	132.32	0.0062	0.82	0.0000	0.00	0.0111	1.47
	±9.89	±0.0063	±0.83	±0.0000	±0.00	±0.0064	±0.86
Non-W e	12.03	0.0923	1.11	0.0025	0.03	0.0216	0.26
	±2.43		±0.41		±0.04		±0.14
Non-W μ	11.72	0.0896	1.05	0.0026	0.03	0.0324	0.38
	±2.36		±0.39		±0.04		±0.18
$\sum$ Backgrounds	212.86		7.29		0.62		5.15
			±1.62		±0.82		±1.69
SM-SM $t\bar{t}$	31.00	0.3085	9.56	0.0497	1.54	0.1093	3.39
	±7.07	±0.0150	±2.23	±0.0047	±0.38	±0.0100	±0.83
SM-FCNC $t\bar{t}$		0.1990		0.0416		0.1560	
		±0.0099		±0.0040		±0.0099	
$\sum$ Backgrounds	243.86		16.85		2.16		8.54
+ SM-SM $t\bar{t}$			±2.76		±0.90		±1.88
Data	244.00		23.00		3.00		12.00
Tag excess			6.15		0.84		3.46
			±4.94		±1.73		±3.48

Table D.4: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "2c", where a taggable jet requires ≥ 2 positive i.p. decay tracks, where $\geq$ has track probability $< 10\%$.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2666	0.47	0.0258	0.05	0.0643	0.11
	±0.37	±0.0171	±0.10	±0.0048	±0.01	±0.0095	±0.03
Single top (W^*)	0.53	0.3139	0.17	0.0456	0.02	0.0607	0.03
	±0.09	±0.0230	±0.03	±0.0089	±0.01	±0.0120	±0.01
Dibosons	6.27	0.0179	0.11	0.0063	0.04	0.0329	0.21
	±0.82	±0.0106	±0.07	±0.0063	±0.04	±0.0188	±0.12
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0137	0.17	0.0000	0.00	0.0195	0.24
	±2.58	±0.0069	±0.09	±0.0000	±0.00	±0.0080	±0.11
Wc	12.38	0.0509	0.63	0.0030	0.04	0.0665	0.82
	±3.11	±0.0157	±0.25	±0.0031	±0.04	±0.0203	±0.33
Wbb	6.99	0.3053	2.13	0.0380	0.27	0.0592	0.41
	±1.11	±0.0328	±0.41	±0.0128	±0.10	±0.0201	±0.15
Wcc	13.04	0.0976	1.27	0.0186	0.24	0.1153	1.50
	±2.55	±0.0292	±0.45	±0.0103	±0.14	±0.0295	±0.48
W + non-h.f.	132.32	0.0055	0.73	0.0000	0.00	0.0296	3.91
	±9.89	±0.0056	±0.74	±0.0000	±0.00	±0.0105	±1.42
Non-W e	12.03	0.0998	1.20	0.0133	0.16	0.0324	0.39
	±2.43		±0.43		±0.10		±0.19
Non-W μ	11.72	0.0922	1.08	0.0026	0.03	0.0461	0.54
	±2.36		±0.40		±0.04		±0.24
$\sum$ Backgrounds	212.86		7.97		0.85		8.18
			±1.76		±0.87		±2.04
SM-SM $t\bar{t}$	31.00	0.3191	9.89	0.0661	2.05	0.1293	4.01
	±7.07	±0.0154	±2.31	±0.0055	±0.50	±0.0107	±0.97
SM-FCNC $t\bar{t}$		0.2115		0.0649		0.2114	
		±0.0102		±0.0055		±0.0122	
$\sum$ Backgrounds	243.86		17.86		2.90		12.19
+ SM-SM $t\bar{t}$			±2.90		±1.00		±2.26
Data	244.00		25.00		5.00		10.00
Tag excess			7.14		2.10		-2.19
			±5.13		±1.98		±4.16

Table D.5: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing W + ≥ 3 jets cuts. The efficiencies were measured using NN tagger "2g", where a taggable jet requires ≥ 2 positive i.p. decay tracks with track probability $< 70\%$.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2694	0.48	0.0280	0.05	0.0610	0.11
	±0.37	±0.0173	±0.10	±0.0053	±0.01	±0.0094	±0.03
Single top (W^*)	0.53	0.3174	0.17	0.0482	0.03	0.0707	0.04
	±0.09	±0.0232	±0.03	±0.0096	±0.01	±0.0142	±0.01
Dibosons	6.27	0.0211	0.13	0.0078	0.05	0.0231	0.14
	±0.82	±0.0124	±0.08	±0.0079	±0.05	±0.0164	±0.10
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0138	0.17	0.0000	0.00	0.0151	0.18
	±2.58	±0.0069	±0.09	±0.0000	±0.00	±0.0068	±0.09
Wc	12.38	0.0391	0.48	0.0032	0.04	0.0547	0.68
	±3.11	±0.0143	±0.21	±0.0033	±0.04	±0.0174	±0.27
Wbb	6.99	0.3069	2.14	0.0376	0.26	0.0571	0.40
	±1.11	±0.0326	±0.41	±0.0135	±0.10	±0.0201	±0.15
Wcc	13.04	0.1085	1.42	0.0263	0.34	0.1108	1.44
	±2.55	±0.0323	±0.50	±0.0134	±0.19	±0.0293	±0.48
W + non-h.f.	132.32	0.0054	0.71	0.0000	0.00	0.0199	2.64
	±9.89	±0.0055	±0.73	±0.0000	±0.00	±0.0082	±1.10
Non-W e	12.03	0.0998	1.20	0.0133	0.16	0.0241	0.29
	±2.43		±0.43		±0.10		±0.15
Non-W μ	11.72	0.0896	1.05	0.0026	0.03	0.0435	0.51
	±2.36		±0.39		±0.04		±0.23
$\sum$ Backgrounds	212.86		7.96		0.96		6.44
			±1.82		±0.89		±1.81
SM-SM $t\bar{t}$	31.00	0.3256	10.09	0.0661	2.05	0.1337	4.15
	±7.07	±0.0155	±2.35	±0.0057	±0.50	±0.0117	±1.01
SM-FCNC $t\bar{t}$		0.2157		0.0586		0.1921	
		±0.0103		±0.0052		±0.0117	
$\sum$ Backgrounds	243.86		18.05		3.01		10.59
+ SM-SM $t\bar{t}$			±2.97		±1.02		±2.07
Data	244.00		25.00		5.00		9.00
Tag excess			6.95		1.99		-1.59
			±5.19		±2.01		±3.86

Table D.6: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing W + ≥ 3 jets cuts. The efficiencies were measured using NN tagger "2f", where a taggable jet requires ≥ 2 positive i.p. decay tracks with track probability $< 60\%$.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2523	0.45	0.0209	0.04	0.0477	0.08
	±0.37	±0.0167	±0.10	±0.0045	±0.01	±0.0084	±0.02
Single top (W^*)	0.53	0.3009	0.16	0.0513	0.03	0.0752	0.04
	±0.09	±0.0234	±0.03	±0.0113	±0.01	±0.0164	±0.01
Dibosons	6.27	0.0218	0.14	0.0075	0.05	0.0215	0.14
	±0.82	±0.0128	±0.08	±0.0077	±0.05	±0.0151	±0.10
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0257	0.32	0.0000	0.00	0.0121	0.15
	±2.58	±0.0109	±0.15	±0.0000	±0.00	±0.0060	±0.08
Wc	12.38	0.0690	0.85	0.0033	0.04	0.0169	0.21
	±3.11	±0.0193	±0.32	±0.0033	±0.04	±0.0099	±0.13
Wbb	6.99	0.2932	2.05	0.0234	0.16	0.0405	0.28
	±1.11	±0.0327	±0.40	±0.0103	±0.08	±0.0176	±0.13
Wcc	13.04	0.0948	1.24	0.0217	0.28	0.1200	1.56
	±2.55	±0.0271	±0.43	±0.0119	±0.17	±0.0306	±0.50
W + non-h.f.	132.32	0.0124	1.64	0.0000	0.00	0.0204	2.70
	±9.89	±0.0091	±1.20	±0.0000	±0.00	±0.0084	±1.12
Non-W e	12.03	0.0948	1.14	0.0025	0.03	0.0166	0.20
	±2.43		±0.41		±0.04		±0.12
Non-W μ	11.72	0.0811	0.95	0.0026	0.03	0.0324	0.38
	±2.36		±0.36		±0.04		±0.18
$\sum$ Backgrounds	212.86		8.93		0.66		5.74
			±1.94		±0.83		±1.85
SM-SM $t\bar{t}$	31.00	0.3043	9.44	0.0518	1.61	0.1160	3.60
	±7.07	±0.0155	±2.21	±0.0055	±0.40	±0.0127	±0.91
SM-FCNC $t\bar{t}$		0.2006		0.0433		0.1624	
		±0.0101		±0.0043		±0.0106	
$\sum$ Backgrounds	243.86		18.37		2.27		9.34
+ SM-SM $t\bar{t}$			±2.94		±0.92		±2.06
Data	244.00		22.00		3.00		8.00
Tag excess			3.63		0.73		-1.34
			±5.20		±1.77		±3.69

Table D.7: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "2b", where a taggable jet requires ≥ 2 positive i.p. decay tracks with track probability $< 50\%$.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2504	0.44	0.0156	0.03	0.0311	0.05
	±0.37	±0.0170	±0.10	±0.0038	±0.01	±0.0061	±0.02
Single top (W^*)	0.53	0.2978	0.16	0.0262	0.01	0.0577	0.03
	±0.09	±0.0235	±0.03	±0.0067	±0.00	±0.0126	±0.01
Dibosons	6.27	0.0222	0.14	0.0000	0.00	0.0000	0.00
	±0.82	±0.0132	±0.08	±0.0000	±0.00	±0.0000	±0.00
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0145	0.18	0.0000	0.00	0.0055	0.07
	±2.58	±0.0074	±0.10	±0.0000	±0.00	±0.0039	±0.05
Wc	12.38	0.0183	0.23	0.0029	0.04	0.0206	0.26
	±3.11	±0.0098	±0.13	±0.0030	±0.04	±0.0104	±0.14
Wbb	6.99	0.2912	2.03	0.0152	0.11	0.0192	0.13
	±1.11	±0.0344	±0.40	±0.0080	±0.06	±0.0100	±0.07
Wcc	13.04	0.0681	0.89	0.0097	0.13	0.0946	1.23
	±2.55	±0.0259	±0.38	±0.0076	±0.10	±0.0273	±0.43
W + non-h.f.	132.32	0.0079	1.05	0.0000	0.00	0.0090	1.19
	±9.89	±0.0081	±1.07	±0.0000	±0.00	±0.0053	±0.70
Non-W e	12.03	0.0923	1.11	0.0058	0.07	0.0083	0.10
	±2.43		±0.41		±0.06		±0.08
Non-W μ	11.72	0.0674	0.79	0.0000	0.00	0.0273	0.32
	±2.36		±0.31		±0.00		±0.16
$\sum$ Backgrounds	212.86		7.01		0.38		3.38
			±1.70		±0.74		±1.42
SM-SM $t\bar{t}$	31.00	0.3016	9.35	0.0360	1.11	0.0876	2.72
	±7.07	±0.0156	±2.19	±0.0039	±0.28	±0.0093	±0.68
SM-FCNC $t\bar{t}$		0.1923		0.0277		0.1159	
		±0.0105		±0.0033		±0.0086	
$\sum$ Backgrounds	243.86		16.36		1.49		6.10
+ SM-SM $t\bar{t}$			±2.77		±0.79		±1.57
Data	244.00		21.00		2.00		6.00
Tag excess			4.64		0.51		-0.10
			±4.90		±1.45		±2.93

Table D.8: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "2d", where a taggable jet requires ≥ 2 positive i.p. decay tracks with track probability $< 30\%$.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2438	0.43	0.0147	0.03	0.0322	0.06
	±0.37	±0.0172	±0.09	±0.0035	±0.01	±0.0061	±0.02
Single top (W^*)	0.53	0.2853	0.15	0.0333	0.02	0.0677	0.04
	±0.09	±0.0222	±0.03	±0.0073	±0.00	±0.0124	±0.01
Dibosons	6.27	0.0274	0.17	0.0000	0.00	0.0000	0.00
	±0.82	±0.0163	±0.10	±0.0000	±0.00	±0.0000	±0.00
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0145	0.18	0.0000	0.00	0.0028	0.03
	±2.58	±0.0079	±0.10	±0.0000	±0.00	±0.0028	±0.04
Wc	12.38	0.0197	0.24	0.0000	0.00	0.0171	0.21
	±3.11	±0.0112	±0.15	±0.0000	±0.00	±0.0086	±0.12
Wbb	6.99	0.2719	1.90	0.0185	0.13	0.0324	0.23
	±1.11	±0.0323	±0.38	±0.0086	±0.06	±0.0128	±0.10
Wcc	13.04	0.0813	1.06	0.0043	0.06	0.0638	0.83
	±2.55	±0.0299	±0.44	±0.0044	±0.06	±0.0220	±0.33
W + non-h.f.	132.32	0.0000	0.00	0.0000	0.00	0.0031	0.41
	±9.89	±0.0000	±0.00	±0.0000	±0.00	±0.0032	±0.42
Non-W e	12.03	0.0756	0.91	0.0000	0.00	0.0025	0.03
	±2.43		±0.35		±0.00		±0.04
Non-W μ	11.72	0.0572	0.67	0.0000	0.00	0.0162	0.19
	±2.36		±0.28		±0.00		±0.11
$\sum$ Backgrounds	212.86		5.72		0.23		2.03
			±1.36		±0.63		±1.03
SM-SM $t\bar{t}$	31.00	0.2951	9.15	0.0287	0.89	0.0744	2.31
	±7.07	±0.0148	±2.14	±0.0030	±0.22	±0.0071	±0.57
SM-FCNC $t\bar{t}$		0.1847		0.0177		0.0842	
		±0.0095		±0.0025		±0.0072	
$\sum$ Backgrounds	243.86		14.87		1.12		4.34
+ SM-SM $t\bar{t}$			±2.54		±0.67		±1.18
Data	244.00		20.00		1.00		5.00
Tag excess			5.13		-0.12		0.66
			±4.62		±1.25		±2.39

Table D.9: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "2e", where a taggable jet requires ≥ 2 positive i.p. decay tracks with track probability $< 20\%$.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2234	0.39	0.0124	0.02	0.0343	0.06
	±0.37	±0.0176	±0.09	±0.0033	±0.01	±0.0066	±0.02
Single top (W^*)	0.53	0.2425	0.13	0.0302	0.02	0.0689	0.04
	±0.09	±0.0216	±0.03	±0.0072	±0.00	±0.0133	±0.01
Dibosons	6.27	0.0194	0.12	0.0000	0.00	0.0105	0.07
	±0.82	±0.0142	±0.09	±0.0000	±0.00	±0.0106	±0.07
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0029	0.04	0.0000	0.00	0.0039	0.05
	±2.58	±0.0029	±0.04	±0.0000	±0.00	±0.0030	±0.04
Wc	12.38	0.0221	0.27	0.0000	0.00	0.0038	0.05
	±3.11	±0.0140	±0.19	±0.0000	±0.00	±0.0041	±0.05
Wbb	6.99	0.2671	1.87	0.0223	0.16	0.0290	0.20
	±1.11	±0.0336	±0.38	±0.0094	±0.07	±0.0116	±0.09
Wcc	13.04	0.0690	0.90	0.0051	0.07	0.0798	1.04
	±2.55	±0.0311	±0.44	±0.0051	±0.07	±0.0307	±0.45
W + non-h.f.	132.32	0.0000	0.00	0.0000	0.00	0.0015	0.19
	±9.89	±0.0000	±0.00	±0.0000	±0.00	±0.0019	±0.26
Non-W e	12.03	0.0623	0.75	0.0025	0.03	0.0058	0.07
	±2.43		±0.30		±0.04		±0.06
Non-W μ	11.72	0.0410	0.48	0.0000	0.00	0.0162	0.19
	±2.36		±0.22		±0.00		±0.11
$\sum$ Backgrounds	212.86		4.95		0.29		1.95
			±1.26		±0.53		±1.06
SM-SM $t\bar{t}$	31.00	0.2597	8.05	0.0223	0.69	0.0690	2.14
	±7.07	±0.0154	±1.90	±0.0029	±0.18	±0.0082	±0.55
SM-FCNC $t\bar{t}$		0.1574		0.0130		0.0691	
		±0.0096		±0.0020		±0.0060	
$\sum$ Backgrounds	243.86		13.00		0.98		4.09
+ SM-SM $t\bar{t}$			±2.28		±0.56		±1.19
Data	244.00		15.00		0.00		3.00
Tag excess			2.00		-0.98		-1.09
			±4.27		±1.14		±2.35

Table D.10: Event-tagging efficiencies and counts of background, $t\bar{t}$ signal, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "2h", where a taggable jet requires ≥ 2 positive i.p. decay tracks with track probability $< 10\%$.

Appendix E

Scale-Factor Transformation

In this appendix we derive the covariance matrix of the modified scale factors.

The variables we begin with are the scale factors $\left\{s_{ij}\right\}$ returned by the fit and the jet tagging efficiencies $\left\{\epsilon_{ij}^{MC}\right\}$ measured in each MC sample. (MC events were required to pass the $W + \geq 3$ jets criteria.) These 18 initial variables are represented by the vector:

$$\begin{aligned}\mathbf{v} = \left(\left\{s_{ij}\right\}, \left\{\epsilon_{ij}^{MC}\right\}\right) &= \left(s_{bt}, s_{bb}, s_{bc}, s_{ct}, s_{cb}, s_{cc}, s_{pt}, s_{pb}, s_{pc},\right. \\ &\left. \epsilon_{bb}^{MC}, \epsilon_{bc}^{MC}, \epsilon_{bp}^{MC}, \epsilon_{cb}^{MC}, \epsilon_{cc}^{MC}, \epsilon_{cp}^{MC}, \epsilon_{pb}^{MC}, \epsilon_{pc}^{MC}, \epsilon_{pp}^{MC}\right)\end{aligned}$$

The fit also returns the scale factor covariance matrix $\left\{\sigma_{s_{ij}s_{ik}}\right\}$. The tagging efficiency covariance matrix elements $\left\{\sigma^2_{\epsilon_{ij}^{MC}}\right\}$ were measured in the following way. Each MC sample was divided into N_{ss} subsamples of equal size. The covariance between each pair of efficiencies was calculated and by averaging their efficiency fluctuation product over the subsamples

$$\sigma_{\epsilon_{ij}^{MC}\epsilon_{kl}^{MC}} = \overline{\delta\epsilon_{ij}^{MC}\,\delta\epsilon_{kl}^{MC}} \simeq \frac{1}{N_{ss}} \sum_{\text{subsample } ss=1}^{N_{ss}} \left(\epsilon_{ij}^{ss} - \overline{\epsilon_{ij}^{MC}}\right)\left(\epsilon_{kl}^{ss} - \overline{\epsilon_{kl}^{MC}}\right)$$

where ϵ_{ij}^{ss} denotes the efficiency measured in a MC subsample, and we approximate the true mean efficiency $\overline{\epsilon_{ij}^{MC}}$ with the measured efficiency over the entire MC sample ϵ_{ij}^{MC}. We found the covariance terms to vary negligibly on N_{ss} around our choice $N_{ss} = 8$. [2]

[2]The 9 diagonal elements of the efficiency covariance matrix $\left\{\sigma^2_{\epsilon_{ij}^{MC}}\right\}$ are the variances on each of the tagging efficiencies. Since a tag is a "success" in a Bernoulli trial, these variances could be calculated analytically assuming binomial variance of the numbers of tags. But the off-diagonal covariance elements

Combining all these errors together, the 18×18 covariance matrix of our initial variables is:

$$\begin{aligned}
\mathbf{E} &= \left\{\sigma_{v_i v_j}\right\} \quad , \quad i,j = 1 \ldots 18 \\
&= \begin{pmatrix} \left\{\sigma_{s_{ij} s_{kl}}\right\} & 0 \\ 0 & \left\{\sigma_{\epsilon^{MC}_{mn} \epsilon^{MC}_{qr}}\right\} \end{pmatrix} \quad , \quad \begin{aligned} j,l &= t,b,c \text{ (scale factor type)} \\ i,k,m,q &= b,c,p \text{ (true jet flavor)} \\ n,r &= b,c,p,u \text{ (tag outcome)} \end{aligned}
\end{aligned} \tag{E.1}$$

where $\left\{\sigma_{s_{ij} s_{kl}}\right\}$ and $\left\{\sigma_{\epsilon_{mn}} \sigma_{\epsilon_{qr}}\right\}$ are each 9×9 covariance matrices, and the zero blocks come from the mutual independence of the scale factors and the tagging efficiencies.

Equations 7.1 define a transformation to a new set of variables, the 12 modified scale factors $\left\{s'_{ij}\right\}$, represented by the vector:

$$\mathbf{v}' = \{v'_k(\mathbf{v})\} = \left\{s'_{ij}\right\} = \left(s'_{bb}, s'_{bc}, s'_{bp}, s'_{bu}, s'_{cb}, s'_{cc}, s'_{cp}, s'_{cu}, s'_{pb}, s'_{pc}, s'_{pp}, s'_{pu}\right)$$

$$\text{where } k = 1 \ldots 12 \ , \ i = b,c,p \ , \ j = b,c,p,u$$

with a transformed 12×12 covariance matrix given by[68]:

$$\begin{aligned}
\mathbf{E}' &= \left\{\sigma_{v'_i v'_j}\right\} \quad , \quad i,j = 1 \ldots 12 \\
&= \left\{\sigma_{s'_{ij} s'_{kl}}\right\} \quad , \quad \begin{aligned} i,k &= b,c,p \text{ (true jet flavor)} \\ j,l &= b,c,p,u \text{ (tag outcome)} \end{aligned} \\
&= \tilde{\mathbf{T}}\, \mathbf{E}\, \mathbf{T}
\end{aligned} \tag{E.2}$$

where the elements of the 18×12 transformation matrix $\mathbf{T}$ are the function derivatives:

$$T_{ij} = \frac{\partial v'_j}{\partial v_i} \quad , \quad i = 1 \ldots 18 \ , \ j = 1 \ldots 12$$

$\left\{\sigma_{\epsilon^{MC}_{ij} \epsilon^{MC}_{kl}}\right\}$ must be *measured*, using multiple sets of trials (subsamples), since there is no model for the form of their covariance. We therefore used this subsample technique to measure all elements in the efficiency covariance matrix, including the diagonal elements – a consistent calculation technique for all elements was necessary to keep the pair-wise correlations between elements in the proper range

$$\left|\rho_{\epsilon^{MC}_{ij} \epsilon^{MC}_{kl}}\right| = \left|\frac{\sigma_{\epsilon^{MC}_{ij} \epsilon^{MC}_{kl}}}{\sigma_{\epsilon^{MC}_{ij}} \sigma_{\epsilon^{MC}_{kl}}}\right| \leq 1$$

We used the Maple V symbolic computation program to analytically calculate the derivatives of (7.1) for the transformation matrix. Having all necessary inputs to (E.2), the elements of $\mathbf{E}'$ are then calculated:

$$E'_{ij} \;=\; \sigma_{v'_i v'_j} \;=\; \sum_{k,l=1}^{18} \tilde{T}_{ik}\, E_{kl}\, T_{lj} \;=\; \sum_{k,l=1}^{18} \left(\frac{\partial v'_i}{\partial v_k}\right) \left(\frac{\partial v'_j}{\partial v_l}\right)\, E_{kl}$$

Note that we can calculate the covariance matrix of the data tagging efficiencies in a completely analogous way. In this case, we transform to the set of 12 variables

$$\mathbf{v}' = \left(\epsilon_{bb}^{data}, \epsilon_{bc}^{data}, \ldots, \epsilon_{pu}^{data}\right)$$

via Equations 4.5 (again written as functions of 9 scale factors and 9 MC efficiencies). Then the data tagging efficiency covariance matrix is again given by (E.2), where the transformation matrix $\mathbf{T}$ now contains the derivatives of this different transformation.

Appendix F

Variance of Tagged Data-Events From Modified Scale-Factors

In this appendix we derive the variance of the expected number of tagged events in the data (for a particular event sample) using the notation of Appendix E.

The expected number of tagged events in the data given by (7.2) is a function of the modified scale factors $\{s'_{ij}\}$ and the number of tagged MC events N_{tag}^{MC}. We keep N_{tag}^{MC} constant for the calculation here. The variance of the expected number of tagged events from the modified scale factors is then given by

$$\sigma^2_{N_{tag}^{data},s'} \;=\; \tilde{\mathbf{D}}\,\mathbf{E}'\,\mathbf{D} \tag{F.1}$$

where $\mathbf{E}'$ is the 12×12 covariance matrix of the modified scale factors derived in Appendix E and $\mathbf{D}$ is the 12-component vector of derivatives

$$D_k \;=\; \frac{\partial N_{tag}^{data}}{\partial v'_k} \;=\; \sum_{j=1}^{N_{tag}^{MC}} \frac{\partial\left(\prod_{i=1}^{N_{jet}^{j}} s'_{f_i t_i}\right)}{\partial v'_k} \;\;,\; k = 1 \ldots 12 \tag{F.2}$$

where the modified scale factors $\{s'_{ij}\}$ are mapped onto the variables $\{v'_k\}$ as in Appendix E. As the sum N_{tag}^{data} was accumulated for each tag type over the MC sample, so were these derivatives. The variance was then calculated by expanding (F.1) as

$$\sigma^2_{N_{tag}^{data},s'} \;=\; \sum_{i,j=1}^{12} \left(\frac{\partial N_{tag}^{data}}{\partial v'_i}\right)\left(\frac{\partial N_{tag}^{data}}{\partial v'_j}\right) E'_{ij} \quad .$$

Appendix G

Covariance of Event-Tagging Efficiencies of Different Background Samples

In this Appendix we derive the covariance matrix of the event tagging efficiencies expected in the data for the nine W/Z background samples (listed in the first nine rows of Table 7.4). This discussion is more general than the one-sample case described in Section 7.3.1 and Appendix F.

For a particular background sample, the event tagging efficiency expected in data is (7.2) normalized by the MC sample size:

$$\epsilon_{tag}^{data} = \frac{N_{tag}^{data}\left(\{s^l\}, N_{tag}^{MC}\right)}{N^{MC}} \tag{G.1}$$

where we have indicated the functional dependence of the numerator on the modified scale factors and the number of tagged MC events. The modified scale factors are sample dependent via the particular jet tagging efficiencies in (7.1). The denominator is known and constant for each sample.

There are nine such efficiencies (one for each background sample), and we wish to calculate their covariance matrix so that we can eventually calculate errors on functions of these efficiencies. We expect the efficiencies for different background samples to be correlated because the modified scale factors are correlated across different background samples through their common reliance on the original scale factors from the scale factor fit[3].

[3] Actually, 6 modified scale factors, $s_{bb}, s_{bc}, s_{cb}, s_{cc}, s_{pb}, s_{pc}$ are identical across samples; the other 6 vary because they also depend on jet tagging efficiencies.

G.1 Covariance of Modified Scale Factors Across Background Samples

In Appendix E we calculated the covariance of the 12 modified scale factors in a single sample. We now wish to calculate the covariance of the $12 \times 9 = 108$ modified scale factors in the nine background samples. This calculation is a straightforward generalization of the one-sample case.

Here we begin with 90 initial variables: the 9 scale factors from the scale factor fit and 81 tagging efficiencies (nine from each of the nine samples). The initial 90×90 covariance matrix is constructed analogously as (E.1). The scale factor covariance matrix from the scale factor fit occupies the upper-left 9×9 block, while the nine 9×9 covariance matrices of tagging efficiencies in each background sample fill down diagonally. This initial covariance matrix is again block-wise diagonal; the scale factors and MC tagging efficiencies (from different samples) are uncorrelated.

Our final set of 108 variables are the nine sets of 12 modified scale factors (one set per background sample). For each sample, the derivatives of this transformation are the same as those calculated in Appendix E. The only difference here is that derivatives from each sample fill successive 90×12 blocks of a larger 90×108 transformation matrix. The transformed 108×108 covariance matrix is again given by the matrix product (E.2).

G.2 Covariance of Event Tagging Efficiencies Across Background Samples

We now consider the covariance of the nine event tagging efficiencies (given by (G.1) separately for each background sample).

We begin with 117 variables: nine sets (one for each background sample) of 12

modified scale factors $\{s'\}$ and the number of tagged MC events:

$$\mathbf{v} = \left(\left\{ s'_{ij} \right\}_q , \left(N_{tag}^{MC} \right)_q \right) \quad , \quad \begin{aligned} i &= b, c, p \text{ (true jet flavor)} \\ j &= b, c, p, u \text{ (tag outcome)} \\ q &= 1 \ldots 9 \text{ (background sample)} \end{aligned}$$

The initial 117×117 covariance matrix is:

$$\mathbf{E} = \left\{ \sigma_{v_i v_j} \right\} \quad , \quad i, j = 1 \ldots 21$$

$$= \begin{pmatrix} \left\{ \sigma_{s'_{ij} s'_{kl}} \right\}_q & & 0 \\ & \ddots & \\ & & \left(\sigma^2_{N_{tag}^{MC}} \right)_q \\ 0 & & & \ddots \end{pmatrix} \quad , \quad \begin{aligned} i, k &= b, c, p \text{ (true jet flavor)} \\ j, l &= b, c, p, u \text{ (tag outcome)} \\ q &= 1 \ldots 9 \text{ (background sample)} \end{aligned}$$

where $\left\{ \left\{ \sigma_{s'_{ij} s'_{kl}} \right\}_q \right\}$ is the 108×108 covariance matrix of modified scale factors (described in § G.1), $\left\{ \left(\sigma^2_{N_{tag}^{MC}} \right)_q \right\}$ is the 9×9 diagonal covariance matrix of the tagged MC event counts[4], and the zero blocks reflect the mutual independence of the modified scale factors and MC event counts.

Equation (G.1) defines a transformation to a new set of variables, the event tagging efficiencies expected in the data for each background $\left(\epsilon_{tag}^{data} \right)_q$, represented by the vector:

$$\mathbf{v}' = \left\{ v'_q(\mathbf{v}) \right\} = \left\{ \left(\epsilon_{tag}^{data} \right)_q \right\} \ , \ q = 1 \ldots 9$$

with a transformed 9×9 covariance matrix given by:

$$\begin{aligned} \mathbf{E}' &= \left\{ \sigma_{v'_i v'_j} \right\} = \left\{ \sigma_{\left(\epsilon_{tag}^{data} \right)_i \left(\epsilon_{tag}^{data} \right)_j} \right\} \ , \ i, j = 1 \ldots 9 \\ &= \tilde{\mathbf{T}} \, \mathbf{E} \, \mathbf{T} \end{aligned}$$

[4]This matrix is diagonal because the numbers of tagged MC events in *different* samples are independent. The diagonal elements are simply the binomial variances on the number of tagged MC events for each sample:

$$\left(\sigma^2_{N_{tag}^{MC}} \right)_q = \left(N_{tag}^{MC} \right)_q \left(1 - \left(N_{tag}^{MC} \right)_q / \, N_q^{MC} \right) \ , \ q = 1 \ldots 9$$

where the elements of the 117×9 transformation matrix $\mathbf{T}$ are the function derivatives. For each background sample, the derivatives with respect to the modified scale factors were calculated as in (F.2). The only other non-zero derivatives are

$$\frac{\partial \left(\epsilon_{tag}^{data}\right)_j}{\partial \left(N_{tag}^{MC}\right)_j} \;=\; \frac{1}{N_j^{MC}} \frac{\partial \left(N_{tag}^{data}\right)_j}{\partial \left(N_{tag}^{MC}\right)_j} \;\simeq\; \frac{1}{N_j^{MC}} \frac{\left(N_{tag}^{data}\right)_j}{\left(N_{tag}^{MC}\right)_j} \;=\; \frac{\langle s' \rangle_j}{N_j^{MC}}$$

where $\langle s' \rangle_j$ is the average of the product of modified scale factors over tagged MC events in each background sample. The elements of $\mathbf{T}$ are summarized by

$$T_{ij} \;=\; \frac{\partial v_j'}{\partial v_i} \;=\; \begin{cases} & & & \text{Derivatives w/ respect to...} \\ \frac{1}{N_j^{MC}} \frac{\partial \left(N_{tag}^{data}\right)_j}{\partial v_i} & , & i = 12j - 11 \ldots 12j & \text{...modified scale factors} \\ & & & \quad \text{in sample } j \\ \frac{\langle s' \rangle_j}{N_j^{MC}} & , & i = j + 108 & \text{...number of tagged MC events} \\ & & & \quad \text{in sample } j \\ 0 & , & \text{otherwise} & \text{...quantities in other samples} \end{cases}$$

$$, \quad j = 1 \ldots 9 \quad \text{(background sample)}$$

The elements of the transformed covariance matrix $\mathbf{E}'$ are then calculated:

$$E'_{ij} \;=\; \sigma_{v_i' v_j'} \;=\; \sum_{k,l=1}^{117} \tilde{T}_{ik}\, E_{kl}\, T_{lj} \;=\; \sum_{k,l=1}^{117} \left(\frac{\partial v_i'}{\partial v_k}\right) \left(\frac{\partial v_j'}{\partial v_l}\right) E_{kl} \quad , \quad i,j = 1 \ldots 9$$

We separately calculated this transformed covariance matrix (of event tagging efficiencies expected in data for different background samples) for b-, bc-, and c-event tagging.

Appendix H

Variance of Sum of Tagged Background-Events

In this Appendix we calculate the error on the sum of tagged background events in the nine background samples.

For each event tag type, the sum of the expected numbers of tagged events has the following form:

$$S \;=\; S_1 + S_2$$

where S_1 is the sum of tagged events (expected in the data) of each of the nine MC background samples,

$$S_1 \;=\; \sum_{i=1}^{9} p_i \left(\epsilon_{tag}^{data}\right)_i$$

with the tagging efficiencies $\left(\epsilon_{tag}^{data}\right)_i$ given by G.1 and the pretag event counts p_i discussed in Sections 6.1 and 6.4, and S_2 is the contribution from the non-W backgrounds,

$$S_2 \;=\; \left(N_D^{tag}\right)_e + \left(N_D^{tag}\right)_\mu$$

where the tag counts N_D^{tag} are estimated separately for the electron and muon samples as discussed in Section 7.3.2.

We first consider the error on S_1. The variables going into S_1 are the pretag numbers p_i and the event tagging efficiencies $\left(\epsilon_{tag}^{data}\right)_i$. We assume that each pretag number is uncorrelated with any other pretag number or event tagging efficiency[5]. The covariances of event

[5]The Wc, W + non-h.f. (and perhaps other W/Z backgrounds) pretag numbers are probably correlated because of their method of calculation (described in Section 6.3 – reliance on common control samples and

tagging efficiencies in pairs of background samples $\sigma_{\left(\epsilon_{tag}^{data}\right)_i\left(\epsilon_{tag}^{data}\right)_j}$ are measured as described in Appendix G. This leads to a variance of the form

$$\begin{aligned}
\sigma_{S_1}^2 &= \sum_{i=1}^{9}\left(\frac{\partial S_1}{\partial p_i}\right)^2 \sigma_{p_i}^2 \quad + \quad \sum_{i,j=1}^{9}\frac{\partial S_1}{\partial\left(\epsilon_{tag}^{data}\right)_i}\frac{\partial S_1}{\partial\left(\epsilon_{tag}^{data}\right)_j}\sigma_{\left(\epsilon_{tag}^{data}\right)_i\left(\epsilon_{tag}^{data}\right)_j} \\
&= \sum_{i=1}^{9}\left(\epsilon_{tag}^{data}\right)_i^2 \sigma_{p_i}^2 \quad + \quad \sum_{i,j=1}^{9} p_i\, p_j\, \sigma_{\left(\epsilon_{tag}^{data}\right)_i\left(\epsilon_{tag}^{data}\right)_j}
\end{aligned}$$

Next we consider the error on S_2. The terms in S_2 are probably correlated since their only difference is the type of trigger lepton. We have insufficient statistics in the data to measure this correlation, so we assume full correlation. The error on S_2 is then the linear sum of errors on estimated tagged non-W events

$$\sigma_{S_2} = \sigma_{\left(N_D^{tag}\right)_e} + \sigma_{\left(N_D^{tag}\right)_\mu}$$

We now consider the final error on S. The tag counts of S_1 and S_2 are uncorrelated because the fluctuations of S_1 come from uncertainties in the pretag numbers and the event tagging efficiencies while the fluctuations of S_2 are from the statistics of the data. We therefore add these errors in quadrature to yield the final variance

$$\sigma_S^2 = \sigma_{S_1}^2 + \sigma_{S_2}^2$$

constrained fitting). We ignore these possible correlations because we assume their effect is small compared to other errors. The strongest correlation between the pretag numbers, between the single top event types, comes from the Run 1b integrated luminosity. We ignore this correlation because the relative luminosity error is small compared to their respective relative theory cross section errors. Finally, pretag numbers are clearly uncorrelated with event tagging efficiencies because their methods of measurement are completely independent.

Appendix I

Non-W-Background Tags Check

In this appendix we check the approximation for the non-W tagging efficiency made in Section 7.3.2. This non-W tagging efficiency, which may depend on the isolation and $\not\!\!E_{\rm T}$ of the event, was approximated with its average value over the non-W regions A,B,C (defined in Table 6.3):

$$\epsilon_{tag} \;=\; \frac{N_A^{tag} + N_B^{tag} + N_C^{tag}}{N_A + N_B + N_C}$$

We can check this approximation by comparing the actual number of tags in each region

$$N_i^{tag} \quad , \; i = A, B, C$$

with that expected using the average efficiency

$$N_i^{tag,expected} \;=\; N_i \cdot \epsilon_{tag} \quad , \; i = A, B, C.$$

For each predicted mean, we have only one trial comparison of the observed number of tags. If we had multiple trials per mean, we could check that the numbers of observed tags are distributed according to a Poisson probability *density* function $p(N^{tag,expected}, N^{tag})$ about the predicted mean $N^{tag,expected}$. Instead we consider for each trial the Poisson probability *distribution* function

$$P(N^{tag,expected}, N^{tag}) \;=\; \sum_{n=0}^{n<N^{tag}} p(N^{tag,expected}, n)$$

which gives the probability of observing fewer than N^{tag} tagged events in each trial. The distribution of P should be flat over trials for each predicted mean (to the extent that the

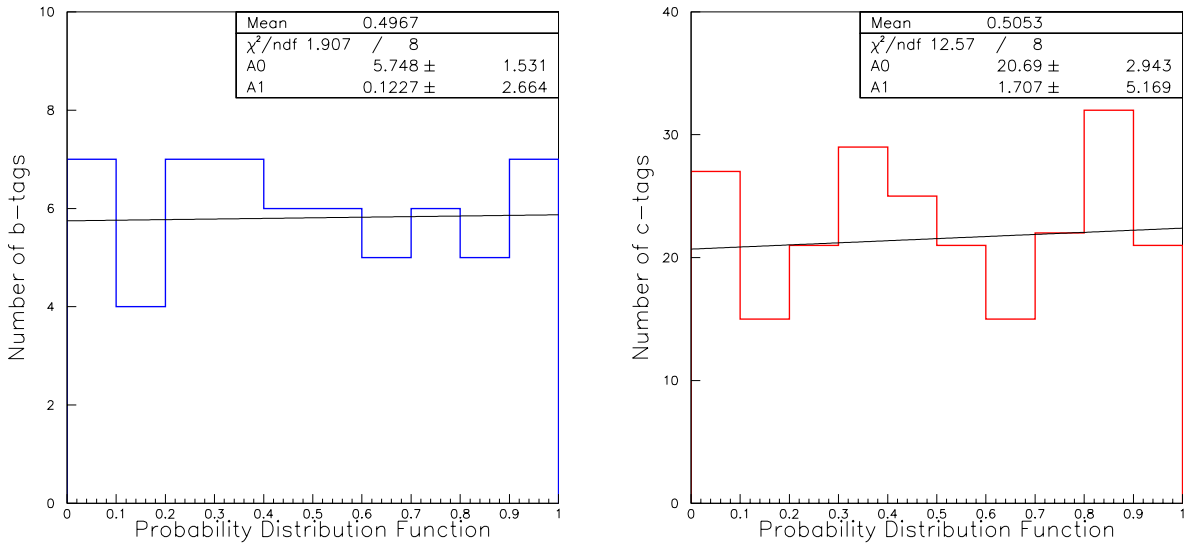


Figure I.1: The distribution of the Poisson probability distribution function for observed numbers of tagged non-W events, shown separately for b-tags (left) and c-tags (right).

predicted mean is a good approximation of the true Poisson mean)[6]. Although we still have only one trial per predicted mean, we still expect P to have a flat distribution across multiple means since P is begin sampled from multiple flat distributions.

We make the distribution of P separately for b- and c-tags by comparing the observed tags in the data with the following predicted means. We have a predicted mean for each of three non-W regions for both the electron and muon samples, using 10 different NN taggers. Additionally for c-tags, we have four different dijet mass window widths. We then have $3 \cdot 2 \cdot 10 = 60$ trials for b-tags, and $3 \cdot 2 \cdot 10 \cdot 4 = 240$ trials for c-tags. The distribution of P over these trials is shown separately for b- and c-tags in Figure I.1. Both of these distributions are consistent (within error) with a line with zero slope. Therefore we conclude that our method of using the average efficiency ϵ_{tag} yields reasonable Poisson means for tagged events.

[6] In N_{trials} trials, we expect to observe $N = N_{trials} P(N^{tag,expected}, N^{tag})$ trials with fewer than N^{tag} tagged events. Then $dN/dP = N_{trials}$ is constant $\Rightarrow$ P has a flat distribution.

Appendix J

Justification of Flavor-Tagger Choice

The method used to measure the branching fraction $B(t \to cg)$ and to set an upper limit is described in Chapter 8. Here we justify our choice of NN tagger "2e" from Table 7.1 with a c-tag dijet mass range of 140 GeV.

We considered separately each combination of tagger (1a,1b,...,2h) and dijet mass range (50,80,110,140 GeV). In addition, for each combination, we tried a branching fraction fit that used b-, c-, b,c-, or b,bc,c-tag information in the likelihood function (8.6), with the terms corresponding to the unused tags omitted. We refer to each such combination of tagger, dijet mass range, and tag information used as a "fit combination".

For each fit combination we performed null hypothesis pseudo-experiments (as in § 8.1.4) using tagging efficiencies and expected backgrounds from the appropriate tag table (Appendix D). Next we constructed a confidence belt (as in § 8.2.1) and used it to extract an expected upper limit (as in § 8.2.2) from the null hypothesis pseudo-experiments.

Table J.1 lists for each fit combination the expected upper limit, the fitted branching fraction in the data $B_{\rm fit}^{\rm data}$ and the corresponding upper limit $B_{95}^{\rm data}$. Included in the table are only those fit combinations whose pull distributions fit reasonably to Gaussians of unit width and zero mean, as in Figure 8.3 (right). We found that fits using only b- or c-tag information had significant biases, which explains their absence from the table. Adding a second type of tag information to the fit removed such biases.

The fit combinations in Table J.1 are listed in order of increasing expected upper

limit. *The combination on which we have reported in this analysis was selected because it yields the lowest expected upper limit (listed at the top of the table).* Several observations can be made about the other fit combinations.

First, for a given tagger and choice of used tags, the expected limit generally decreases as the dijet mass range is increased. Recall that the dijet mass cut was meant to select FCNC over SM top-quark decays (Figure 7.1). The trend note above shows that the c-tag efficiency decrease resulting from an additional mass cut is more detrimental (for limiting the branching fraction) than purifying the SM-FCNC signal.

Second, for a given tagger and choice of dijet mass range, the lowest expected limit is achieved by using all tagging information. Although there are significantly fewer bc-tags than either of the single tags, this correlated tag information is still useful in setting a lower limit.

Finally, the choice of tagger has the most effect on the expected limit. Taggers "2e" and "2h", with decay track probability cuts of 20% and 10%, respectively, give the lowest limits. The general rule throughout the table is that stricter taggable jet cuts lead to lower limits. Since taggers with strict taggable jet cuts discriminate well between b, c and prompt jets, we conclude that the suppression of non-h.f. backgrounds, using track i.p. cuts for decay tracks in taggable jets, is very important in setting the limit.

We also note that though we use the fit combination with the lowest expected limit to report the fitted branching fraction and its 95% CL upper limit, comparable values are obtained with many of the runners-up.

Tagger	Tags Used	M_{cj} width	$B_{95}^{\rm exp}$	$B_{\rm fit}^{\rm data}$	$B_{95}^{\rm data}$	Tagger	Tags Used	M_{cj} width	$B_{95}^{\rm exp}$	$B_{\rm fit}^{\rm data}$	$B_{95}^{\rm data}$
2e	b, c, bc	140	0.59	-0.41	0.32	2c	b, c, bc	80	0.66	-0.43	0.40
2e	b, c, bc	110	0.60	-0.36	0.35	2d	b, c, bc	80	0.66	-0.30	0.50
2e	b, c	140	0.61	-0.46	0.28	2d	b, c	110	0.66	-0.42	0.44
2h	b, c, bc	110	0.61	-0.07	0.56	2d	b, c, bc	50	0.67	-0.37	0.45
2h	b, c, bc	140	0.61	-0.04	0.58	2a	b, c, bc	80	0.67	-0.47	0.37
2h	b, c	110	0.62	-0.22	0.44	2c	b, c, bc	50	0.67	-0.34	0.45
2h	b, c	140	0.62	-0.20	0.47	2c	b, c	110	0.67	-0.42	0.44
2h	b, c, bc	80	0.62	-0.06	0.59	2b	b, c, bc	140	0.67	-0.44	0.47
2e	b, c, bc	80	0.62	-0.27	0.44	2d	b, c	80	0.67	-0.29	0.53
2e	b, c, bc	50	0.62	-0.24	0.45	2b	b, c, bc	110	0.67	-0.57	0.38
2h	b, c, bc	50	0.63	-0.02	0.62	2a	b, c, bc	50	0.68	-0.36	0.46
2c	b, c, bc	140	0.64	-0.33	0.45	1b	b, c, bc	50	0.68	-0.48	0.45
2d	b, c, bc	140	0.64	-0.43	0.37	2b	b, c, bc	80	0.70	-0.37	0.52
2d	b, c, bc	110	0.65	-0.42	0.42	2b	b, c	140	0.70	-0.42	0.51
2d	b, c	140	0.66	-0.42	0.41	1b	b, c	50	0.72	-0.50	0.51
2g	b, c, bc	140	0.66	-0.57	0.27						

Table J.1: Expected 95% CL upper limit $B_{95}^{\rm exp}$, fitted branching fraction in data $B_{\rm fit}^{\rm data}$ and corresponding upper limit $B_{95}^{\rm data}$, for different combinations of taggers, tag information used, and dijet mass range cuts, listed in order of increasing $B_{95}^{\rm exp}$. The approximate average error on $B_{\rm fit}^{\rm data}$ is 0.40 and on $B_{95}^{\rm exp}$ is 0.002.

Appendix K

Flavor-Tagging Statistics For $t \to cZ$ Search

Table K.1 shows the event-tagging statistics for NN tagger "2e" (with no dijet mass cuts for c-tags) used in the search for $t \to cZ$ described in § 8.3.

Event sample	Pretag	ϵ_{b-tag}	b-tags	ϵ_{bc-tag}	bc-tags	ϵ_{c-tag}	c-tags
Single top (Wg)	1.77	0.2438	0.43	0.0147	0.03	0.0435	0.08
	±0.37	±0.0172	±0.09	±0.0035	±0.01	±0.0075	±0.02
Single top (W^*)	0.53	0.2853	0.15	0.0333	0.02	0.0863	0.05
	±0.09	±0.0222	±0.03	±0.0073	±0.00	±0.0150	±0.01
Dibosons	6.27	0.0274	0.17	0.0000	0.00	0.0000	0.00
	±0.82	±0.0163	±0.10	±0.0000	±0.00	±0.0000	±0.00
$Z \to \tau\tau$	3.55	0.0000	0.00	0.0000	0.00	0.0000	0.00
	±0.53	±0.0000	±0.00	±0.0000	±0.00	±0.0000	±0.00
Z + jets	12.26	0.0145	0.18	0.0000	0.00	0.0028	0.03
	±2.58	±0.0079	±0.10	±0.0000	±0.00	±0.0028	±0.04
Wc	12.38	0.0197	0.24	0.0000	0.00	0.0444	0.55
	±3.11	±0.0112	±0.15	±0.0000	±0.00	±0.0147	±0.23
Wbb	6.99	0.2719	1.90	0.0185	0.13	0.0589	0.41
	±1.11	±0.0323	±0.38	±0.0086	±0.06	±0.0180	±0.14
Wcc	13.04	0.0813	1.06	0.0043	0.06	0.0962	1.26
	±2.55	±0.0299	±0.44	±0.0044	±0.06	±0.0275	±0.43
W + non-h.f.	132.32	0.0000	0.00	0.0000	0.00	0.0062	0.82
	±9.89	±0.0000	±0.00	±0.0000	±0.00	±0.0046	±0.61
Non-W e	12.03	0.0756	0.91	0.0000	0.00	0.0274	0.33
	±2.43		±0.35		±0.00		±0.17
Non-W μ	11.72	0.0572	0.67	0.0000	0.00	0.0324	0.38
	±2.36		±0.28		±0.00		±0.18
$\sum$ Backgrounds	212.86		5.72		0.23		3.90
	±11.60		±1.36		±0.63		±1.35
SM-SM	31.00	0.2951	9.15	0.0287	0.89	0.0849	2.63
$t\bar{t} \to WbWb$	±7.07	±0.0148	±2.14	±0.0030	±0.22	±0.0079	±0.65
SM-FCNC $t\bar{t}$		0.2125		0.0259		0.0999	
$t\bar{t} \to WbcZ$		±0.0129		±0.0038		±0.0083	
FCNC-FCNC $t\bar{t}$		0.2125		0.0259		0.0999	
$t\bar{t} \to cZcZ$		±0.0129		±0.0038		±0.0083	
$\sum$ Backgrounds	243.86		14.87		1.12		6.53
+ SM-SM $t\bar{t}$			±2.54		±0.67		±1.50
Data	244.00		20.00		1.00		7.00
Tag excess			5.13		-0.12		0.47
			±4.62		±1.25		±2.96

Table K.1: Event-tagging efficiencies and counts of background, $t\bar{t} \to WbWb, WbcZ, cZcZ$ signals, and data events passing $W + \geq 3$ jets cuts. The efficiencies were measured using NN tagger "2e", where a taggable jet requires ≥ 2 positive i.p. decay tracks with track probability $< 20\%$.

Bibliography

[1] J. Donoghue, E. Golowich, and B. Holstein, *Dynamics of the Standard Model*, Cambridge University Press (1992).

[2] W. Williams, *Nuclear and Particle Physics*, Clarendon Press (1991).

[3] G. Kane, *Modern Elementary Particle Physics*, Addison-Wesley Publishing Company (1993).

[4] V. Barger and R. Phillips, *Collider Physics*, Addison-Wesley Publishing Company (1987).

[5] The Particle Data Group, *Review of Particle Physics*, Eur. Phys. J. C **15**, (2000).

[6] P. Collins, A. Martin, and E. Squires, *Particle Physics and Cosmology*, John Wiley & Sons, Inc. (1989).

[7] J. Babcock, D. Sivers, and S. Wolfram, "Quantum-chromodynamic Estimates for Heavy-Particle Production," Phys. Rev. **D18**, 162 (1977).

[8] B. Combridge, J. Kripfganz, and J. Ranft, "Hadron Production at Large Transverse Momentum and QCD," Phys. Lett. **70B**, 234 (1977).

[9] R. Cutler and D. Sivers, "Quantum-chromodynamic Gluon Contributions to Large-P_T Reactions," Phys. Rev. **D17**, 196 (1978).

[10] F. Halzen and P. Hoyer, "Quantum-chromodynamic Gluon Contributions to Large-P_T Reactions," Phys. Lett. **154B**, 324 (1985).

[11] G. Eilam, J. Hewett, and A. Soni, "Rare Decays of the Top Quark in the Standard and Two-Higgs-Doublet Models," Phys. Rev. **D44**, 1473 (1991).

[12] G. Couture, "Top quark flavor changing neutral currents at CERN LEP-200: Signals and backgrounds," Phys. Rev. D **62**, 097503 (2000).

[13] F. Abe *et al.* [CDF Collaboration], "Search for flavor-changing neutral current decays of the top quark in p anti-p collisions at s**(1/2) = 1.8-TeV," Phys. Rev. Lett. **80**, 2525 (1998).

[14] V. Obraztsov, S. Slabospitsky and O. Yushchenko, "Search for anomalous top-quark interaction at LEP-2 collider," Phys. Lett. **B426**, 393 (1998).

[15] B. Denby, "Neural networks in high energy physics: A ten year perspective," Comput. Phys. Commun. **119**, 219 (1999).

[16] Fermilab Click N' Tour the Accelerators, http://www.fnal.gov/pub/tour.html, 1999.

[17] Y. Asano *et al.*, "Design Report For the Fermilab Collider Detector Facility (CDF)", 1981 (unpublished).

[18] F. Abe *et al.* [CDF Collaboration], "The CDF Detector: An Overview," Nucl. Inst. and Methods **A271**, 387 (1988).

[19] H. Wenzel, "Tracking in the SVX", CDF Note 1790, (1998, unpublished).

[20] D. Amidei *et al.*, "The Silicon Vertex Detector of the Collider Detector at Fermilab," Nucl. Inst. and Methods **A350**, 73 (1994).

[21] F. Abe *et al.*, "The CDF Vertex Time Projection Chamber System," Nucl. Inst. and Methods **A268**, 75 (1988).

[22] F. Bedeschi *et al.*, "Design and Construction of the CDF Central Tracking Chamber," Nucl. Inst. and Methods **A268**, 50 (1988).

[23] D. Bauer *et al.*, "The CDF Central Electromagnetic Calorimeter," Nucl. Inst. and Methods **A267**, 272 (1988).

[24] G. Apollinari *et al.*, "The CDF Central and Endwall Hadron Calorimeter," Nucl. Inst. and Methods **A267**, 301 (1988).

[25] G. Ascoli *et al.*, "CDF Central Muon Detector," Nucl. Inst. and Methods **A268**, 33 (1988).

[26] S. Belforte *et al.*, "The 1992 CDF Muon System Upgrade", CDF Note 2858, (1995, unpublished).

[27] S. Belforte *et al.*, "Sigma(BBC) Derivation and Results", CDF Note 2361, (1994, unpublished).

[28] G. Drake *et al.*, "CDF Front End Electronics: The RABBIT System," Nucl. Inst. and Methods **A269**, 68 (1988).

[29] E. Barsotti *et al.*, "FASTBUS Data Acquisition For CDF," Nucl. Inst. and Methods **A269**, 82 (1988).

[30] D. Amidei *et al.*, "A Two Level Fastbus Based Trigger System For CDF," Nucl. Inst. and Methods **A269**, 51 (1988).

[31] J. Carrol *et al.*, "The CDF Level 3 Trigger," Nucl. Inst. and Methods **A300**, 552 (1991).

[32] ALEPH Collaboration, "Evidence for charm production in W decays," EPS-HEP97 PA08, Jerusalem 19-26 August 1997.

[33] G. Unal, "Tracking Efficiency and SVX Scale Factor", CDF Note 3563, (1995, unpublished).

[34] D. Gerdes, "B Identification Using Jet Probability", CDF Note 2023, (1993, unpublished).

[35] H. Kolanoski, "Application of artificial neural networks in particle physics", Nucl. Inst. and Methods **A367**, 14-20 (1995).

[36] A. Caner, B. Denby, J. Wyss, "Search For Top in Multijet Events With a Neural Network: Part I", CDF Note 1914, (1993, unpublished).

[37] P. Berge *et al.*, "SVX B-tag Algorithm "Seed Vertexing" For Top Search", CDF Note 2716, (1994, unpublished).

[38] www.thep.lu.se/public_html/jetnet_30_manual/jetnet_30_manual.html

[39] R. Hughes *et al.*, "Calculation of the Scale Factor For the Tagging Efficiency in the b and c Content in Jets Analysis", CDF Note 3710, (1996, unpublished).

[40] T. Nakaya, "Primary Charm Quark Sample With D*", CDF Note 4443, (1998, unpublished).

[41] F. James and M. Roos, "'Minuit' A System For Function Minimization And Analysis Of The Parameter Errors And Correlations," Comput. Phys. Commun. **10**, 343 (1975).

[42] F. Abe *et al.* [CDF Collaboration], "Evidence for top quark production in anti-p p collisions at s**(1/2) = 1.8-TeV," Phys. Rev. **D50**, 2966 (1994).

[43] F. Abe *et al.* [CDF Collaboration], "Measurement of the Top Quark Mass," Phys. Rev. Lett. **80**, 2767 (1998).

[44] D. Cronin-Hennessy and A. Beretvas, "Luminosity at CDF", CDF Note 4721, (1998, unpublished).

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

[46] T. Sjöstrand, "PYTHIA 5.6 and JETSET 7.3: Physics and manual," CERN-TH-6488-92; T. Sjöstrand, Computer Physics Commun., **82**, 74 (1994).

[47] F.A. Berends, H. Kuijf, B. Tausk and W.T. Giele, "On the Production of a W and Jets at Hadron Colliders", Nucl. Phys. B, **357**, 32 (1991).

[48] J. Lewis and P. Avery, "CLEOMC: The CDF Interface to the CLEO Monte Carlo (QQ)", CDF Note 2724, (1997, unpublished).

[49] C. Loomis, "Using TAUOLA, A Decay Library for Polarized τ Leptons, at CDF", CDF Note 2796, (1994, unpublished).

[50] G. Apollinari *et al.*, "Observation of Semileptonic Decays of Heavy Quarks Produced in $p\bar{p}$ Interactions and the Existence of One Scalar Quark", CDF Note 4795, Appendix C, (1999, unpublished).

[51] G. Sganos, "Tutorial on CDFSIM", (1993, unpublished).

[52] M. Shapiro *et al.*, "An Updated User's Guide to QFL", (1996, unpublished).

[53] G. Marchesini *et al.*, Computer Phys. Commun. 67 (1992) 465.

[54] G. Unal, *$Wb\bar{b}$ and $Wc\bar{c}$ Backgrounds in Top SVX Channel*, CDF Note 3389, (1995, unpublished).

[55] R. Hughes *et al.*, "Sample Selection for Run 1B Top Search", CDF Note 2966, (1996, unpublished).

[56] R. Roser, private communication.

[57] R. Roser, "A Measurement of the Trigger Efficiencies for $t\bar{t}$ Events in Rub 1B", CDF Note 3442, (1996, unpublished).

[58] F. Abe *et al.* [CDF Collaboration], Phys. Rev. D. **43**, 2070 (1991).

[59] P. Giromini, F. Ptohos, and W. Yao, "Search for New Particles $X \to b\bar{b}$ Produced in Association with W Bosons at $\sqrt{s} = 1.8$ TeV", CDF Note 4112, (1997, unpublished).

[60] G. Apollinari *et al.*, "Measurement of the $t\bar{t}$ Production Cross Section Using SECVTX Tags", CDF Note 3855, (1997, unpublished).

[61] T. Stelzer, Z. Sullivan, and S. Willenbrock, Phys. Rev. D, **56**, 5919 (1997).

[62] M. Smith and S. Willenbrock, Phys. Rev. D, **54**, 6696 (1996).

[63] R. Bonciani *et al.*, Nuc. Phys. B, **529**, 424-450 (1998).

[64] R. Hughes *et al.*, "Non-W Backgrounds in W Plus Jets", CDF Note 2027, (1993, unpublished).

[65] P. Derwent, Ph.D. Thesis, University of Chicago (1990).

[66] R. Cousins and G. Feldman, Phys. Rev. D., "Unified approach to the classical statistical analysis of small signals", **57**, 3873 (1998).

[67] B. Roe, private communication.

[68] L. Lyons, *Statistics For Nuclear and Particle Physicists*, Cambridge University Press (1992).