Search for Vector-like Quark Production in the Lepton+jets and Dilepton+jets Final States Using 5.4 fb^{-1} of Run II Data

Seth Caughron

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Search for Vector-like Quark Production in the Lepton+jets and Dilepton+jets Final States Using 5.4 fb^{-1} of Run II Data

Seth Caughron

The Standard Model of particle physics provides an excellent description of particle interactions at energies up to $\sim$ 1 TeV, but it is expected to fail above that scale. Multiple models developed to describe phenomena above the TeV scale predict the existence of very massive, vector-like quarks. A search for single electroweak production of such particles in $p\bar{p}$ collisions at a center-of-mass energy of 1.96 TeV is performed in the W+jets and Z+jets channels. The data were collected by the DØ detector at the Fermilab Tevatron Collider and correspond to an integrated luminosity of 5.4 fb^{-1}. Events consistent with a heavy object decaying to a vector boson and a jet are selected. We observe no significant excess in comparison to the background prediction and set 95% confidence level upper limits on production cross sections for vector-like quarks decaying to W+jet and Z+jet. Assuming a vector-like quark – standard model quark coupling parameter $\tilde{\kappa}_{qQ}$ of unity, we exclude vector-like quarks with mass below 693 GeV for decays to W+jet and mass below 449 GeV for decays to Z+jet. These represent the most sensitive limits to date.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank first and foremost my advisor, Gustaaf Brooijmans, for his guidance, assistance, encouragement, support and patience during my six years as a graduate student. My understanding of how high-energy experimental particle physics works in practice is immeasurably improved thanks to his instruction, and as a result, I feel prepared for a career as a professional physicist. I would also like to thank John Parsons for his help both professional and personal during my graduate studies, and I would be remiss in not also extending thanks to the other Columbia professors in the ATLAS group, Michael Tuts and Emlyn Hughes.

I am grateful to my fellow Columbia University students for their day-to-day assistance, moral support and overall good company. Mark Cooke and Ning Zhou in particular were two people with whom I worked for nearly the entirety of my time on ATLAS and DØ, and I thank them for their help on matters big and small over the years. I would additionally like to thank Zachary Marshall, Kerstin Perez, David Lopez-Mateos, Heather Gray, Alex Penson, Evan Williams, Dustin Urbaniec and Evan Wulf for their more-occasional but always-willing help. I consider myself fortunate to have been part of such a talented and agreeable group of students.

Furthermore, I owe many thanks to the Columbia ATLAS group postdoctoral researchers who have always taken the time to help a lost graduate student figure out what he's doing. Katherine Copic, Valeria Perez-Reale, Francesco Spano, Andy Haas and Dominik Dannheim contributed greatly to my knowledge of the ATLAS experiment, as well as to my enjoyment in working on it. I would especially like to express my gratitude to Lidija Zivkovic and Thomas Gadfort for their tireless

assistance with the research contained in this dissertation.

Of course, this dissertation would not be possible if not for the work of the ATLAS and DØ Collaborations in general, and I would like to thank all of my ATLAS and DØ colleagues, particularly those with whom I have worked closely over the years, for their hard work and dedication.

I extend my sincerest thanks to all the close friends that I've met at Columbia, CERN and Fermilab. There are too many to list here, but they know who they are, and without them life would have been much less pleasant the past six years. Many people with whom I've never interacted professionally at all are owed many thanks for keeping me happy, motivated and sane.

Finally, I would like to thank my parents for getting me here. They impressed upon me the importance of education from a very young age, and their love and support over the years, even from across the ocean, has been invaluable.

For my family.

Foreword

I joined the Columbia ATLAS group in 2005. My first hardware work consisted of assisting with the testing of the Front End Boards for the readout of the ATLAS Liquid Argon Calorimeters at Nevis Labs. At the same time, I performed a sensitivity study for a search for Higgs boson decays to $b\bar{b}$ in the Minimal Supersymmetric Standard Model (MSSM) using the ATLAS detector at the CERN Large Hadron Collider.

After moving to CERN in the summer of 2006, I contributed to the simulated data flow and analysis projects called the Final Dress Rehearsals as part of my pre-commissioning work. Experience gained with the ATLAS trigger and data flow led me to join the ATLAS trigger rates group. My work there consisted of the development and maintenance of trigger-filtered "enhanced bias" datasets from simulated minimum bias events. I also participated (using these datasets) in the determination of predicted bandwidth usage for all trigger levels and slices in each new software release, work that helped drive the development of trigger menus for early data taking.

In February 2009, I switched to the DØ experiment in the interests of graduating with an analysis on real physics data, since an accident during LHC commissioning had delayed the first physics run to 2010. I first performed a study of the unclustered energy resolution in $Z \to ee$ events. The results of this study were used to update the missing energy significance algorithm in the DØ offline software. I then began the work summarized in this dissertation, a search for single production of vector-like quarks in the W+jets and Z+jets final states.

Chapter 1

Introduction

Particle physics is a field of physics that studies the fundamental constituents of matter and their interactions. The current knowledge of these constituents and their interactions is summarized in the Standard Model (SM), a theory dating to the 1970s which is one of the most successful and well-tested scientific theories in human history. However, the SM suffers from several shortcomings at higher energy scales, and it is widely acknowledged that previously undiscovered particles and/or interactions must exist to explain what the SM cannot. This dissertation describes a search for a particular type of hypothetical particle, the vector-like quark, that is allowed in several different extensions to the SM.

1.1 The Standard Model

The SM [1; 2; 3; 4] is above all a theory of interactions. It describes all known fundamental interactions, or forces, with the exception of gravity: the electromagnetic interaction, the weak interaction, and the strong interaction. There are two classes of particles in the SM: fermions, which make up the visible matter in the universe and interact with each other via the forces listed above, and bosons, the so-called "force carrier" particles which mediate the interactions. Each particle has a correspond-

ing anti-particle with the same mass and opposite electric charge. The two major theories which comprise the SM are quantum chromodynamics (QCD) and quantum electrodynamics (QED). The former describes the strong nuclear force, and the latter describes the electromagnetic force. The SM also provides an internally consistent description of the distinct electromagnetic and weak interactions as a result of the breaking of an electroweak symmetry through the so-called "Higgs mechanism."

At the mathematical level, the SM is a gauge theory with a symmetry group $SU(3) \otimes SU(2) \otimes U(1)$. Particles are represented by fields, and the fermions, which have no explicit mass terms in the SM Lagrangian, acquire masses via the introduction of a scalar field (the Higgs field). However the Higgs particle remains the only particle predicted by the SM which has not yet been observed in nature.

1.1.1 Fermions

Fermions are particles with half-integer spin obeying Fermi-Dirac statistics. They make up all the visible matter in the universe, and are divided into two sub-categories: leptons and quarks. Leptons are distinguished by their interactions via the electromagnetic and weak forces, and include electrons, muons and taus, along with their associated neutrinos. Quarks, while also coupling to electroweak force carriers, additionally interact through the strong force, and thus form baryonic matter, including protons, neutrons and integer-spin mesons.

Leptons

There are three generations of matter in the SM. The lepton content of each generation is a charged lepton and an associated neutral neutrino. Leptons are spin-1/2 particles, and participate in the electromagnetic and weak interactions, but not the strong interaction. The first generation consists of the electron (e^-) and electron neutrino (ν_e), the second generation consists of the muon (μ) and muon neutrino (ν_μ) and the third generation consists of the tau (τ) and tau neutrino (ν_τ). Each lepton has a lepton flavor number, L_e, L_μ or L_τ, which is $+1$ for each particle and -1

for each anti-particle. These lepton flavor numbers are individually conserved in the SM. Lepton properties are summarized in Table 1.1 [5]. Neutrinos are unique in the SM in that they are neutral and almost massless fermions, and in that right-handed neutrinos (and left-handed antineutrinos) are sterile, i.e. do not interact at all in the SM.

Lepton Content of the Standard Model

	Particle	Electric Charge	Mass (MeV/c^2)	Anti-particle
1$^{\text{st}}$ generation	e^-	-1	0.511	e^+
	ν_e	0	< 0.000002	$\bar{\nu_e}$
2$^{\text{nd}}$ generation	μ^-	-1	105.7	μ^+
	ν_μ	0	< 0.19	$\bar{\nu_\mu}$
3$^{\text{rd}}$ generation	τ^-	-1	1776.8	τ^+
	ν_τ	0	< 18.2	$\bar{\nu_\tau}$

Table 1.1: Leptons of the standard model, with charges, masses and anti-particles.

Quarks

Quarks are also spin-1/2 particles and are differentiated from leptons primarily by their participation in the strong interaction. They carry fractional electric charges of $+2/3$ or $-1/3$ of the electron charge, and are, like leptons, organized into three generations. The first generation consists of the up (u) and down (d) quarks, the second of the charm (c) and strange (s) quarks and the third of the top (t) and bottom (b) quarks. Strong interactions involving the second or third generation conserve quark flavor number, denoted by U (upness), D (down-ness), C (charmness), S (strangeness), T (topness) and B (bottomness), though the weak interaction does not conserve these numbers. Table 1.2 summarizes quark properties [5].

According to the SM, single, or free, quarks are not found in nature. Quarks are always confined in bound states called baryons (three quark bound states) and mesons (quark-antiquark pairs). They carry an additional quantum number, color, corresponding to their strong coupling and analagous to the charge of the electromagnetic interaction. Quarks can have 1 of 3 color charges – red (r), blue (b) or green (g) – and antiquarks carry anti-color, but their free states as baryons or mesons are color-neutral.

Quark Content of the Standard Model

	Particle	Electric Charge	Mass (MeV/c^2)	Anti-particle
1st generation	u	$+2/3$	$1.5 - 3.3$	$\bar{u}$
	d	$-1/3$	$3.5 - 6.0$	$\bar{d}$
2nd generation	c	$+2/3$	1270	$\bar{c}$
	s	$-1/3$	104	$\bar{s}$
3rd generation	t	$+2/3$	171200	$\bar{t}$
	b	$-1/3$	4200	$\bar{b}$

Table 1.2: Quarks of the standard model, with charges, masses and anti-particles.

1.1.2 Gauge bosons

Fermions interact through the exchange of gauge bosons, spin-1 particles which mediate the various forces described by the SM. There are four: The photon (γ) mediates the electromagnetic interaction, and corresponds to the generator of the $U(1)$ gauge group. The $W^{\pm}$ and Z bosons mediate the weak interaction, and correspond to the three generators of $SU(2)$. Finally, the gluon mediates the strong interaction, and the eight different color-anticolor states of the gluon generate $SU(3)$. These particles are summarized in Table 1.3 [5]. The photon and gluons are massless, whereas the W and Z bosons, like the fermions, acquire their masses via the Higgs mechanism.

1.1.3 Fundamental interactions

The electromagnetic interaction, through which particles that carry electric charge exchange virtual photons with each other under the symmetry group $U(1)_{em}$, is described in quantum mechanical terms by QED. The electromagnetic coupling strength is defined by the dimensionless quantity

Gauge Boson Content of the Standard Model

Force	Force carrier	Electric Charge	Mass (MeV/c^2)
Electromagnetic	γ	0	0
Charged weak	$W^{\pm}$	± 1	80.398 ± 0.025
Neutral weak	Z	0	91.1876 ± 0.0021
Strong	g	0	0

Table 1.3: Gauge bosons of the standard model, with charges and masses.

$$\alpha = \frac{e^2}{4\pi\epsilon_0 \hbar c} = \frac{1}{137.036...} \tag{1.1}$$

where e is the electron charge, ϵ_0 is the permittivity of free space and $\hbar = h/(2\pi)$ is the reduced Planck constant. α is known as the fine structure constant, and since $\alpha \ll 1$, electromagnetic interactions can be computed using perturbation theory by expanding matrix element calculations in orders of α. Indeed, predictions of QED have been verified in the laboratory to one part in 10^{12}, making it the most accurately tested theory in the history of science.

The weak interaction is understood in the SM as being distinct from the electromagnetic interaction only at low energy scales, a result of electroweak symmetry breaking via the Higgs mechanism. Above the electroweak symmetry breaking scale, the electromagnetic and weak forces are unified into one interaction under the gauge group $SU(2) \otimes U(1)$. The Higgs mechanism spontaneously breaks the symmetry of this gauge group, resulting in a massless photon and three massive gauge bosons, the $W^{\pm}$ and the Z. Due to the masses of the W and Z bosons, the weak interaction is short-range, as opposed to the long-range electromagnetic force.

QCD is the gauge field theory which describes the strong interaction between quarks through the exchange of gluons. As gluons carry both a color and an anti-color, there are eight color states of gluons which correspond to the generators of an

$SU(3)_{color}$ non-Abelian gauge symmetry group. Due to the fact that gluons themselves carry the charge of the force they mediate, they have a self-interaction which makes QCD non-Abelian (as opposed to QED). The coupling strength of QCD, α_s, changes (or "runs") as a function of the energy of the interaction, as do the electromagnetic and weak coupling strenghts. QCD is non-perturbative in energy regimes where α_s is large, i.e. at low energies.

1.2 Limitations of the Standard Model

The only particle predicted by the SM which has not yet been observed in the laboratory is the spin-0 Higgs boson. Experimental searches for the particle responsible for electroweak symmetry breaking have excluded a SM Higgs boson with masses in the regions $m_H < 114.4$ GeV/c^2 and $158 < m_H < 175$ GeV/c^2 at the 95% confidence level (C.L.) [5; 6]. Figure 1.1 shows the combined Tevatron and LEP exclusion plot as a function of Higgs mass.

However, even if one assumes that the Higgs will eventually be discovered with a mass consistent with the SM's predictions, the SM will still suffer from significant shortcomings. First of all, the SM is a theory of interactions. It describes the interactions between the fermions quite well, but has nothing to say about why, for example, the fermions have the masses they do. It requires 19 free parameters (including said fermion masses), which are not predicted by anything but rather are determined from empirical measurements. It does not include a theory of gravity or an explanation for dark matter. It contains only left-handed, massless neutrinos, a prediction which is contradicted by experimental results.

Even the confirmation of the existence of the Higgs boson would entail further problems with the SM. In general, the presence of a spin-0 particle in a quantum field theory with a mass much smaller than the cutoff scale of the theory requires an unnaturally delicate cancellation between the bare mass of the particle and its

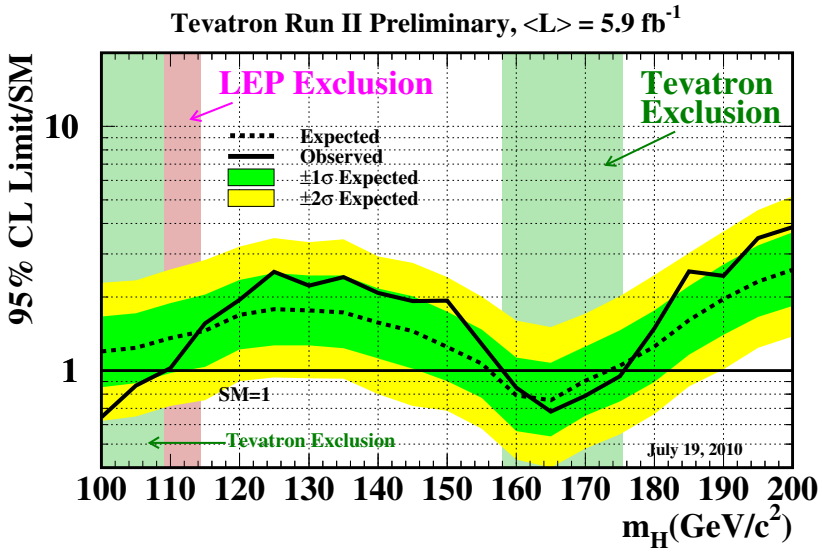

Figure 1.1: Upper limits on the ratios to the SM cross section as a function of Higgs mass. The horizontal line is the expected SM cross section; mass regions where the observed limit is below this line are excluded for an SM Higgs.

quadratic radiative corrections through loop diagrams. This unnaturalness is referred to as "the hierarchy problem." A Higgs particle with a mass less than 1 TeV and a cutoff scale of $\sim 10^{16}$ TeV, or even the Planck scale at $\sim 2 \times 10^{18}$ TeV, implies a fine-tuning on the order of 1 part in 10^{26}. The hierarchy problem can also be considered in terms of the extreme disparity between the strength of gravity and that of the weak force. The ratio of Fermi's constant G_F (associated with the electroweak symmetry breaking scale) and Newton's constant G_N (associated with the strength of gravity) is roughly 10^{32}, and it is a mystery why nature appears to have chosen such vastly disparate mass scales.

A wide range of theories suggests the existence of new particles or interactions close to the electroweak scale in order the stabilize the Higgs boson mass and/or explain phenomena that the SM does not. Several of these theories postulate the existence of vector-like quarks, and these are described in the next chapter.

Chapter 2

Vector-like Quarks

The property that distinguishes vector-like fermions from SM fermions is *non-chirality*, i.e. their left- and right-handed components transform in the same way under $SU(3)\otimes SU(2)\otimes U(1)$. Vector-like fermions do not exist in the SM: All SM fermions are chiral. A model which allows for the existence of vector-like quarks is described in Section 2.1, and a specific example of vector-like quark couplings to SM vector bosons is given in Section 2.2. References to other models are listed in Section 2.3, and the results of previous searches are given in Section 2.4.

2.1 Warped Extra Dimensions and the Randall-Sundrum Model

One method of addressing the hierarchy problem described in Section 1.2 is to assume that spacetime has more than four fundamental dimensions. If there are $4+n$ spacetime dimensions, where n is the number of spatial dimensions in addition to the three we observe, then the Planck scale, $M_{Pl} = 2 \times 10^{18}$ GeV, is only the 4-dimensional effective, or reduced, Plank scale. This reduced scale is determined by the fundamental scale M in $4+n$ dimensions and the geometry of the extra dimensions. For example, in the case of flat compactified extra dimensions with finite volume,

$$M_{Pl}^2 = M^{n+2} V_n, \tag{2.1}$$

where V_n is the volume of the n-dimensional compact space. If V_n is assumed to be very large, the hierarchy between the gravity scale and the weak scale can be eliminated. However, this setup introduces a new large hierarchy between the weak scale and the compactification scale μ_c ($\sim 1/V_n^{1/n}$).

The Randall-Sundrum (RS) model [7] attempts to solve this problem by proposing a scenario with extra dimensions that have a warped rather than a flat geometry. The non-factorizable metric describing a $4+1$-dimensional warped spacetime is

$$ds^2 = e^{-2kr_c\phi}\eta_{\mu\nu}dx^\mu dx^\nu + r_c^2 d\phi^2, \tag{2.2}$$

where k is a curvature scale on the order of the Planck scale, x^μ are the familiar 4-dimensional spacetime coordinates, and ϕ is the coordinate for the extra dimension, which has a finite interval whose size is set by r_c (the compactification radius of the 5^{th} dimension). This metric is a solution to Einstein's equations in a universe containing two 3-dimensional branes. SM fields are localized on one of the branes, while gravity originates on the other. Among the many theoretical and phenomenological implications of this setup, the reduced Planck scale reads as

$$M_{Pl}^2 = \frac{M^3}{k}[1 - e^{-2kr_c\pi}]. \tag{2.3}$$

Since the source of the hierarchy in this model is an exponential function of the compactification radius r_c, significant fine-tuning is not required to reproduce the hierarchy observed in nature. The introduced hierarchy between the fundamental 5-dimensional Planck scale and the compactification scale ($\mu_c \equiv 1/r_c$) is only of order 50.

2.2 Vector-like Quark Theory and Phenomenology

Fermions can be vector-like in five dimensions; therefore the existence of vector-like quarks is possible in models with extra dimensions such as the example described above. Of particular interest are those scenarios in which the modifications to SM quark couplings due to their interactions with vector-like quarks cancel out, allowing mixing between vector-like and light quarks to a degree unconstrained by precision electroweak measurements [8]. Otherwise, significant electroweak couplings between SM quarks and vector-like quarks are excluded by precision measurements of the couplings of SM quarks to each other, implying that the cross section for electroweak production of single vector-like quarks is heavily suppressed. A simple possibility [9] is to introduce two new vector-like quarks, Q_U and Q_D, with electric charges $+2/3$ and $-1/3$, respectively. Their charged- and neutral-current gauge interactions to the first generation SM quarks are

$$\frac{g}{\sqrt{2}}W^+_\mu(\kappa_{uD}\bar{u}_R\gamma^\mu Q_{D_R}+\kappa_{dU}\bar{d}_R\gamma^\mu Q_{U_R})+\frac{g}{2c_W}Z_\mu(\kappa_{uU}\bar{u}_R\gamma^\mu Q_{U_R}+\kappa_{dD}\bar{d}_R\gamma^\mu Q_{D_R})+\text{h.c.}, \tag{2.4}$$

where g is the weak coupling constant ($\sim 10^{-6}$), c_W is the cosine of the weak mixing angle ($\theta_W \sim 30^\circ$), u_R and d_R are the SM right-handed quark singlet fields and Q_D and Q_U are the vector-like quark fields. W and Z are the weak vector boson fields. The coupling strength is parametrized by the model-independent parameter κ_{qQ},

$$\kappa_{qQ} = \frac{v}{m_Q}\tilde{\kappa}_{qQ}, \tag{2.5}$$

where v is the vacuum expectation value of the SM Higgs field, and the dimensionless parameter $\tilde{\kappa}_{qQ}$ encodes the model-dependence of the coupling.

In models with warped extra dimensions, $\tilde{\kappa}_{qQ}$ can naturally be of order one. The possibility of large electroweak coupling between vector-like quarks and SM quarks makes searches for single production of vector-like quarks at the Tevatron attractive.

Figure 2.1 shows s- and t-channel production diagrams of single vector-like quarks. They are produced in association with a SM quark. Figure 2.2 shows charged- and neutral-current production cross sections for down- and up-type vector-like quarks, along with the much lower pair-production cross-section. Assuming a $\tilde{\kappa}_{qQ}$ of order one, even vector-like quarks with masses above 600 GeV/c^2 can be produced with cross sections greater than 100 fb.

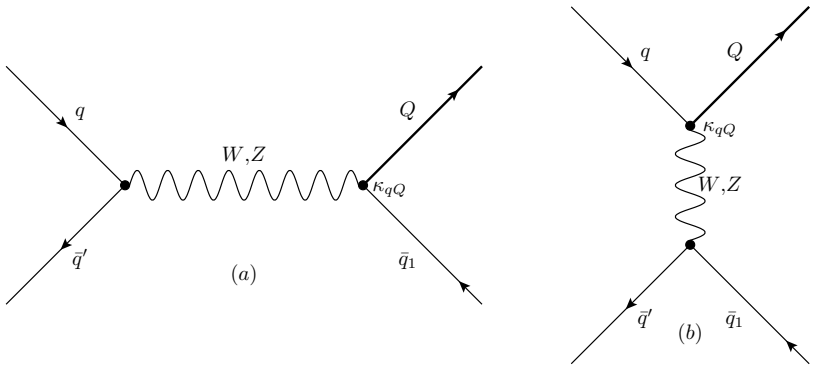

Figure 2.1: s-channel (a) and t-channel (b) Feynman diagrams for single electroweak production of vector-like quarks at the Tevatron.

Vector-like quarks in this model decay to a W or Z boson plus a SM quark q. Decays to the Higgs boson plus a SM quark are also allowed, but are not considered in this analysis. The implementation of the model which is used in this analysis sets $\tilde{\kappa}_{uU} = \tilde{\kappa}_{uD} = 1$ and $\tilde{\kappa}_{dU} = \tilde{\kappa}_{dD} = 0$. Allowing coupling only to the SM up quark implies that Q_D decays exclusively to Wq and Q_U decays exclusively to Zq. It also means that Q_D is produced only through charged-current interactions, and Q_U is produced only through neutral current interactions. This choice enables us to conduct two independent searches for vector-like quarks in two channels: $(W \to \ell\nu)qq$ for decays to Wq, and $(Z \to \ell\ell)qq$ for decays to Zq. There are two SM quarks in the final state: one from the associated production with the vector-like quark (see Figure 2.1), and one from the vector-like quark decay. Only decays of vector bosons to electrons or muons will be considered.

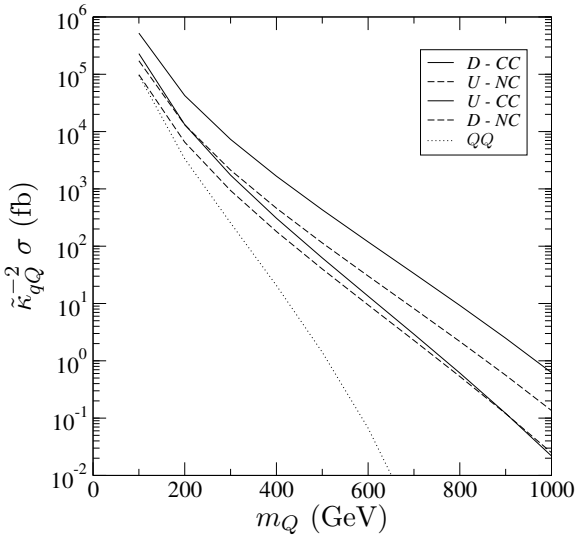

Figure 2.2: Vector-like quark production cross sections for charged current (solid lines) and neutral current (dashed lines) processes at the Tevatron as a function of vector-like quark mass m_Q. The dotted curve is for strong pair production.

The particular choice of coupling does not affect the final cross section limits shown in Section 9, as these are general for decays to Wq and Zq, respectively. However, it does affect the theoretical production cross sections of Q_U and Q_D by restricting Q_U to neutral-current production and Q_D to charged-current production. The excluded vector-like quark masses cited in Section 9 are therefore dependent on our choice of coupling.

2.3 Other Models

There are many other examples of theories of physics beyond the SM which predict the existence of vector-like quarks. Although a discussion of all of them is beyond the scope of this dissertation, notable examples include models with universal extra dimensions, which contain an entire vector-like fourth generation of fermions [10; 11], and little Higgs models [12; 13].

2.4 Previous Searches

Previous searches at the Tevatron for vector-like quarks, or for heavy quarks in general, have focused on pair production via the strong interaction. A Feynman diagram of this process is shown in Figure 2.3. The CDF collaboration has excluded a mass below 338 GeV/c^2 for a pair-produced heavy quark decaying to a W boson plus a SM quark [14]. Figure 2.4 shows the cross-section exclusion curve as a function of vector-like quark mass.

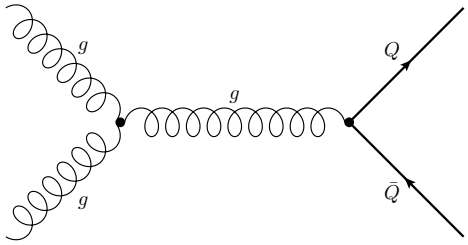

Figure 2.3: Strong pair production of vector-like quarks at the Tevatron.

The motivation for searching for pair-produced vector-like quarks, in spite of the reduced kinematic range accessible in the production of two heavy objects, is the assumption that vector-like quarks (and heavy quarks in general) cannot mix sizably with SM quarks. Such a search also depends only minimally on the strength of the coupling to weak bosons[1], making it essentially model-independent.

[1]The coupling does need to be large enough to ensure prompt vector-like quark decay.

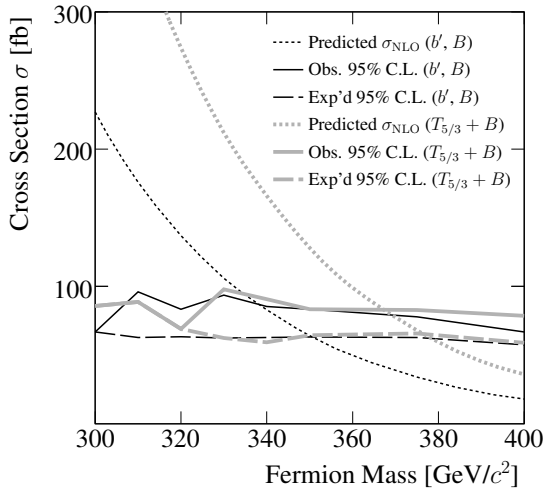

Figure 2.4: 1D cross section exclusion curve as a function of mass for a pair-produced heavy quark (curves labeled b') decaying to Wq (CDF Collaboration).

Chapter 3

Experimental Apparatus

The Fermilab Tevatron Collider [15; 16], located at the Fermi National Accelerator Laboratory [17], produces proton-antiproton ($p\bar{p}$) collisions at a center-of-mass energy ($\sqrt{s}$) of 1.96 TeV. This analysis uses data collected by the DØ experiment, one of two multipurpose detectors (CDF is the other) designed to study the outgoing particles produced by these collisions. The Fermilab Tevatron Collider and the DØ experiment are both described in detail below.

3.1 The Fermilab Tevatron

The Tevatron Accelerator Complex is a system of several different accelerators. The proton and antiproton beams that collide at $\sqrt{s} = 1.96$ TeV in the Tevatron are provided by the Proton Source (comprising the Cockcroft-Walton, Linac and Booster) and the Antiproton Source (comprising the Debuncher and Accumulator), respectively. Figure 3.1 shows an illustration of the full accelerator chain.

The process of creating two focused high-energy particle beams in the Tevatron starts with a small bottle of hydrogen. A Cockcroft-Walton accelerator negatively ionizes the hydrogen, producing H^- ions with an energy of 750 keV via electrostatic acceleration. The 750 keV ions are then sent to a linear accelerator (Linac) which

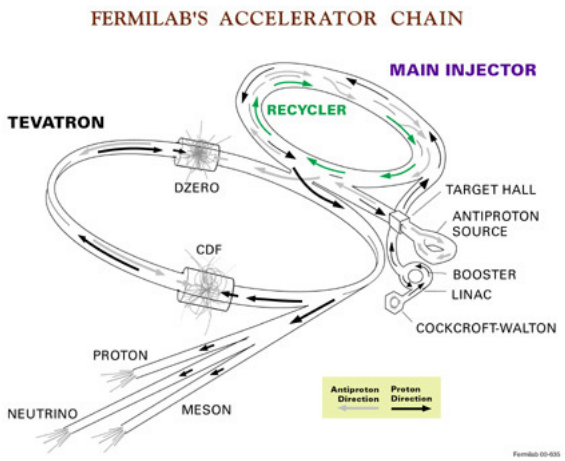

Figure 3.1: The Tevatron Accelerator Complex.

accelerates them to 400 MeV. The Linac consists of two stages, the first composed of five drift tubes, which accelerate the ions to 116 MeV, and the second composed of seven side-coupled cavities, which complete the acceleration to 400 MeV. While the two stages use different technologies, they both rely on oscillating electric fields in radio frequency (RF) cavities to apply a series of short pulses of acceleration to the hydrogen ions as they travel down the beamline. At the end of the Linac, the ions pass through a thin carbon foil which strips off the two electrons, leaving only 400 MeV protons.

The protons then enter the Booster, the first of two intermediate synchrotrons before injection into the Tevatron. The Booster accelerates the protons through roughly 20,000 revolutions around an an accelerator ring with a radius of 75 m, 13.25 times smaller than the Tevatron. The Booster's ring contains 18 RF cavities, which boost the proton's energy to about 8 GeV. The protons are then passed to the second intermediate synchotron, called the Main Injector (MI). With 18 RF cavities arranged in a circular ring of radius 525 m, the MI can acceletrate the protons from 8 GeV to either 120 GeV or 150 GeV. Protons with energy of 150 GeV are injected into the Tevatron for the final stage of acceleration and collisions.

The 120 GeV protons are sent from the MI to a nickel spallation target. The resulting collisions produce a stream of secondary particles, among them antiprotons. Magnets are used to separate out antiprotons with an energy near 8 GeV, which are then directed into the Debuncher (the rest of the particles continue to a beam dump). The purpose of the Debuncher, a roughly triangular ring with average radius 80.4 m, is to reduce the spread of the antiprotons in momentum space to 0.2% or less, which it achieves through techniques called RF bunch rotation and adiabatic debunching. The Debuncher passes these protons to the Accumulator ring, which sits inside the Debuncher ring with a smaller average radius of 75.5 m. The Accumulator stores and cools the 8 GeV antiprotons, receiving beam transfers from the Debuncher every few seconds until it reaches a point where it is essentially full, and adding more antiprotons becomes inefficient. At this time, the antiproton beam is transferred to the Recycler, an antiproton storage ring which is located in the same tunnel as the MI. Antiprotons stored in the Recycler are transferred to the MI at the beginning of a new set of collisions, or store, to be accelerated from 8 GeV to 150 GeV before injection into the Tevatron.

Both the proton and antiproton beams are injected into the Tevatron in 36 bunches, which contain over 10^{10} particles each. The beams travel in opposite directions around the Tevatron ring (radius 1000 m), and the particles are accelerated from 150 GeV to 980 GeV. 774 superconducting niobium-titanium dipole magnets, cooled with liquid helium, bend the protons and antiprotons around the circular ring, while 240 NbTi quadrupole magnets are used to keep the beams focused. The beams, now traveling at nearly the speed of light, cross paths at the two collision points at DØ and CDF with a filled bunch spacing of 396 nanoseconds.

The collision rate per unit area per unit time in the center of the detector is referred to as instantaneous luminosity, and it is generally measured in units of $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. It is given by the formula

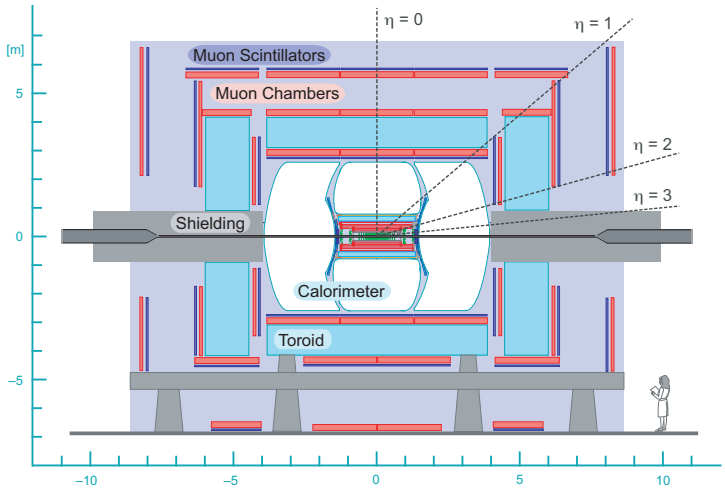

Figure 3.2: A side-view of the DØ detector. The inner tracking system (not labeled) can be seen at the center of the image, inside the calorimeter.

$$L = f\frac{n_1 n_2}{4\pi\sigma_x\sigma_y}, \tag{3.1}$$

where f is the collision rate, n_1 and n_2 are the number or particles per bunch in each beam, and $\sigma_x\sigma_y$ represents the overlapping area of the two beams in the plane transverse to the beam axis. The Tevatron currently holds the world record for instantaneous luminosity achieved at a hadron collider at $4.02 \times 10^{32}\text{cm}^{-2}\text{s}^{-1}$. Integrating the instantaneous luminosity over time gives the integrated luminosity, $\mathcal{L}$, a common measure of data set size which is usually expressed in inverse barns, where $1\text{b}^{-1} = 10^{-24}\text{cm}^{-2}$. The data set used in this analysis corresponds to an integrated luminosity of 5.4 fb^{-1}, recorded by the DØ detector from July 2002 to June 2009.

3.2 The DØ Detector

The DØ detector [18; 19] is one of two multipurpose detectors at the Fermilab Tevatron Collider. A cross-section displaying the main components of the detector is shown in Figure 3.2.

The central tracking system consists of a silicon microstrip tracker and a central fiber tracker, both of which are located within a 2 T superconducting solenoidal magnet. Hits in these two detectors are used to form charged particle tracks. Surrounding the two tracking subdetectors are liquid-argon and uranium calorimeters, both electromagnetic and hadronic, which are used to measure the energy of most particles coming from the collision point. The outer muon system consists of a layer of tracking detectors and scintillation trigger counters in front of 1.8 T toroidal magnets, with two similar layers behind the toroids.

3.2.1 Coordinate System

The DØ Detector utilizes a right-handed coordinate system with the z axis pointing in the direction of the proton beam and the y axis pointing upwards. The origin is located at the physical center of the detector. The azimuthal angle ϕ is defined in the xy plane measured from the x axis, and the polar angle θ is defined in the zy plane with respect to the z axis. Particle momenta are generally described in terms of their components perpendicular (transverse) and parallel (longitudinal) to the z axis. As θ is not a Lorentz-invariant quantity, it is convenient to define an alternate description of the longitudinal kinematics, the rapidity, y:

$$y = \frac{1}{2}\ln[\frac{E + p_L}{E - p_L}], \tag{3.2}$$

where E is the energy of the particle and p_L is its longitudinal momentum. In the limit where $|\vec{p}| \gg m$, a particle's energy and momentum are essentially equivalent, and rapidity reduces to another variable called pseudorapidity, η:

$$\eta = \frac{1}{2}\ln[\frac{|\vec{p}| + p_L}{|\vec{p}| - p_L}] = -\ln[\tan(\frac{\theta}{2})]. \tag{3.3}$$

Rapidity and pseudorapidity are useful in that the the difference in either quantity between two particles is independent of the boost along the beam axis. A related

quantity used in this analysis is ΔR, the separation between two objects in $\eta - \phi$ space:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{3.4}$$

In addition to the longitudinal measures defined above, this analysis also employs measures of energy and momentum in the plane transverse to the beam axis, denoted E_T and p_T:

$$E_T = E \sin\theta \tag{3.5}$$

$$p_T = \sqrt{p_x^2 + p_y^2}. \tag{3.6}$$

Coordinates measured using the detector center as $(0, 0, 0)$ are referred to as "detector coordinates." However, the interaction region has a significant spread along the z axis ($\sigma \approx 25$ cm) and the primary vertex from the $p\bar{p}$ collisions will usually not be located at $z = 0$ as defined by the detector itself. Coordinates defined with respect to the interaction point of a given event are referred to as "physics coordinates."

3.2.2 Central Tracking System

The DØ central tracking system, shown in Figure 3.3, contains a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both installed in a 2 T superconducting solenoid. It is designed for charged particle tracking and vertexing.

The SMT [20; 21] is the subdetector closest to the interaction point, surrounding the beam pipe over a length of 2.4 m. When charged particles pass through the doped silicon of the SMT, ionizing radiation produces free electrons and holes. An electric field acts on the electrons and holes, which travel to the electrodes arranged in long parallel strips over the length of the silicon "ladder," where they produce a measurable pulse.

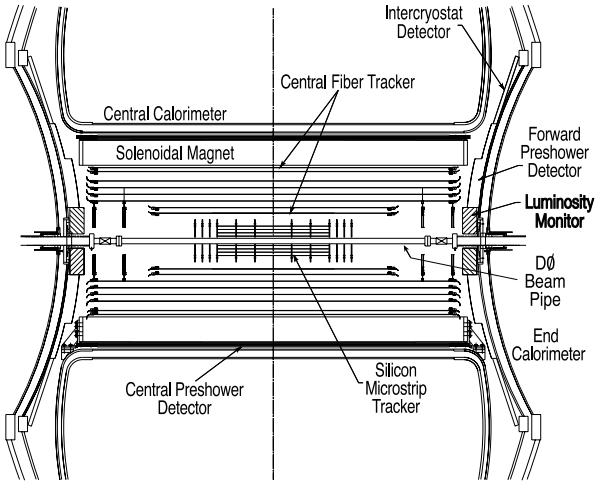

Figure 3.3: A schematic of the DØ inner tracker.

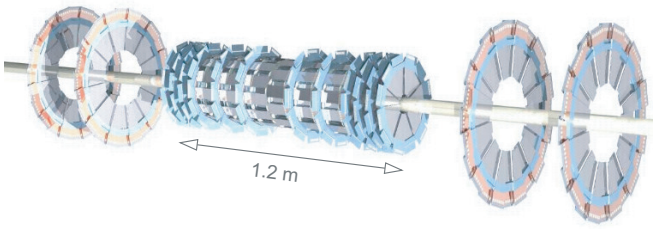

Figure 3.4: A drawing of the silicon microstrip tracker.

The SMT's design implements a combination of barrel and disk modules in order to provide detector surfaces perpendicular to the direction of particle tracks over the entire range in η which it covers, up to $|\eta| < 3$. The barrel detectors primarily measure the $r - \phi$ coordinate, while the disk detectors measure the $r - z$ coordinate in addition to $r - \phi$. The layout of barrel and disk modules in the SMT is shown in Figure 3.4.

The SMT has six barrel segments in the central region, each 12 cm long with four silicon readout layers. The first and third layers are single-sided and axial (microstrips are parallel to the z axis and thus provide the azimuthal position of tracks). The second and fourth layers are double-sided, with one axial side and one stereo side

oriented at an angle of 2° or 90° with respect to the beam line in order to provide the position of tracks in η space. Counting all layers, the barrels have an inner (outer) radius of 2.7 (10.5) cm. The barrels are interspersed with 12 F-disks, each made of 12 double-sided wedge detectors, both sides of which are offset by 15° from the radial direction. The F-disks are located at $|z| = 12.5$, 25.3, 38.2, 43.1 and 53.1 cm and have an inner (outer) radius of 2.6 (10.0) cm. In the forward regions, there are four H-disks consisting of two layers of 12 single-sided wedge detectors offset by 7.5° from the radial direction. The H-disks have an inner (outer) radius of 9.5 (26.0) cm and are located at $|z| = 100.4$ and 121.0 cm. The combination of their larger radius and forward placement provides tracking and vertexing coverage up to $|\eta| < 3$. The SMT provides individual track position resolution of about 10 μm and vertexing resolution of about 35 μm.

The CFT [22], which is used in combination with the SMT to reconstruct charged particle tracks, consists of scintillating fibers mounted on eight concentric support cylinders occupying the radial space from 20 to 52 cm from the center of the beam pipe. The six outer cylinders are 2.52 m long, while the two innermost cylinders are only 1.66 m long to accommodate the H-disks. This arrangement provides coverage of up to $|\eta| < 1.7$. Each cylinder contains two double-layers of fibers, one axial and one at a stereo angle of up to 3°. The double-layers are constructed such that one layer is offset by one half of the 835 nm fiber spacing with respect to the other for a gapless configuration. When charged particles pass through the fibers, they produce photons (scintillate) with a peak emission wavelength of 530 nm. Clear fiber waveguides conduct the light from one end of the scintillating fibers to visible light photon counters (VLPCs). The opposite ends of the fibers are coated in aluminum with a reflectivity of about 90%, making the photon collection efficiency very high. The VLPCs then convert the light signals into electric signals. The CFT provides spatial resolution on the order of 100 μm.

The two detectors of the central tracking system sit inside a 2 T superconducting

solenoidal magnet. The liquid helium-cooled solenoid is enclosed in a cryostat 2.73 m long and 1.4 m in diameter, and operates at a current of 4,749 A. The purpose of the solenoid is to provide a strong, uniform magnetic field in the central tracking volume, with field lines running parallel to the beam axis. The magnetic field bends charged particle tracks, allowing for measurement of particle momentum and charge.

3.2.3 Preshower Detectors

The preshower detectors, located between the solenoid and the calorimeters, aid in electron identification and background rejection. They can also be used to correct energy measurements in the calorimeters for energy losses in the upstream material and to help identify and reconstruct electrons or photons which begin to shower before reaching the calorimeters. The central preshower detector (CPS) covers the region $|\eta| < 1.3$ and the forward preshower detector (FPS) covers $1.5 < |\eta| < 2.5$.

Both the CPS and FPS are made from triangular strips of scintillator. The triangles are arranged in the pattern shown in Figure 3.5, such that there is no dead space between strips. Each triangular strip is wrapped in aluminized mylar for optical isolation and inward reflectivity and has a wavelength-shifting fiber at its center which collects the light emitted by the scintillating material and transports it to clear light-guide fibers. The light signal is then sent to VLPCs for conversion into an electronic signal.

The CPS, made of three concentric cylindrical layers of 1280 triangular scintillator strips each, sits in the 5 cm gap between the solenoid and the central calorimeter. The two FPS detectors (north and south) sit in front of the endcap calorimeters. The FPS detectors consist of two layers, separated in z, with a lead-stainless-steel absorber plate of thickness 2 X_0, where X_0 denotes a radiation length, between them. The 2 inner layers are called MIP layers and the 2 outer layers are called shower layers. Charged particles generally register as a minimum ionizing particle in the MIP layer, but both electrons and photons will usually shower in the absorber, producing a

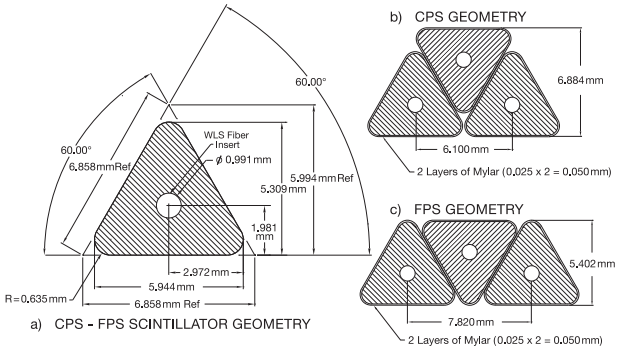

Figure 3.5: Cross sections of the preshower detectors.

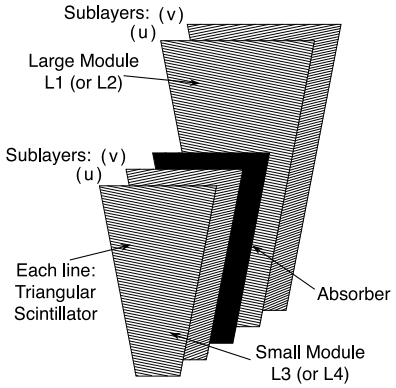

Figure 3.6: One complete ϕ-segment of the FPS.

cluster of energy in the shower layer. Both the MIP and shower layers are arranged in eight 22.5° wedges. Each MIP layer contains 206 scintillator strips and each shower layer contains 288. A drawing of a complete wedge with MIP layers, absorber and shower layers is shown in Figure 3.6.

3.2.4 Calorimeter

The DØ Calorimeter is a liquid argon / uranium sampling calorimeter [19]. It is designed to measure the energy of electrons, photons and jets; to assist in the identification of electrons, photons, jets and muons; and to measure the overall transverse energy balance of events. The calorimeter is made up of 3 sections, a Central

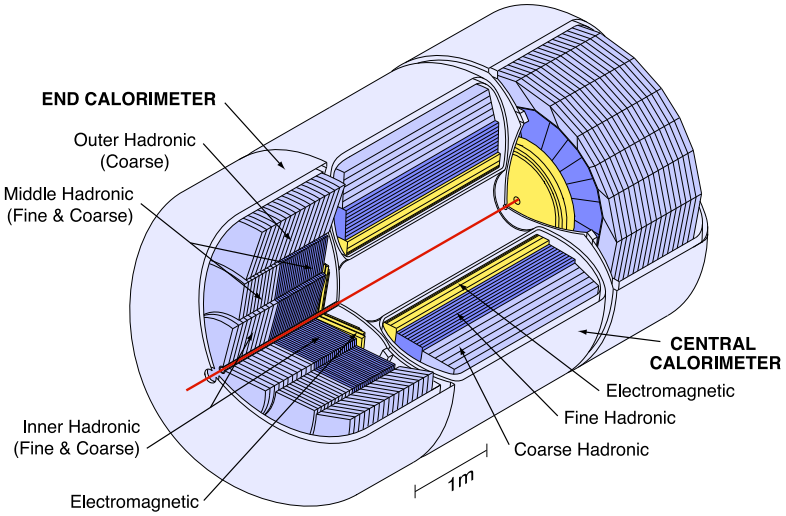

Figure 3.7: A drawing of the DØ calorimeter.

Calorimeter (CC) and two Endcap Calorimeters (EC), which are all contained in separate cryostats and cooled to 90 K. The CC covers $|\eta| < 1.1$ and the EC covers $1.3 < |\eta| < 4.0$. Each calorimeter has three main components: an electromagnetic (EM) layer with fine granularity and two hadronic (HAD) layers, one fine and one coarse. Figure 3.7 shows a cutaway of the calorimeter system.

The calorimeters measure incident particle energy by alternating absorber and active medium. Particles that enter the calorimeter, both charged and neutral, interact with the absorber and produce showers of secondary particles. These secondary particles ionize the active medium, and the resulting charge is collected by high voltage pads. The amount of charge collected is proportional to the energy deposited in the active medium by the incident particle.

The active medium in all of the calorimeters is liquid argon. The absorber material is depleted uranium in the EM layer, a uranium-niobium alloy in the fine hadronic layer, and copper (in the CC) or steel (in the EC) in the coarse hadronic layer. Electromagnetic showers, produced by electrons and photons, develop via the production of electron-positron pairs and Bremsstrahlung radiation, while hadronic showers, produced by hadrons, result from cascades of strong interactions. The different mecha-

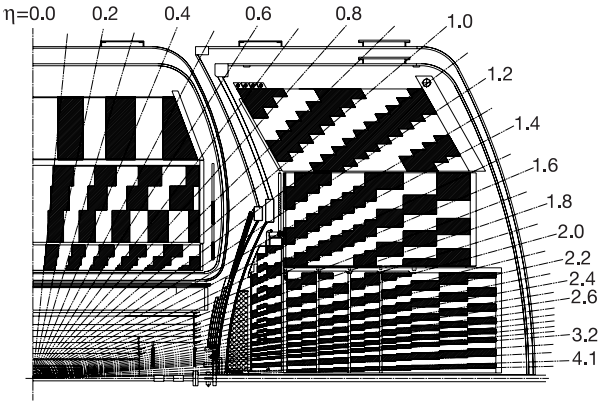

Figure 3.8: Layout of the calorimeter towers

nisms driving shower development can cause different detector responses for the two types of showers, an effect known as non-compensation. The different absorbing materials were chosen for the various layers of the calorimeter in order to achieve a ratio of detector response to EM and HAD showers, a quantity called e/π, of as close to 1 as possible.

Calorimeter cells are arranged in pseudo-projective towers. Starting from the innermost layer and working outwards, there are 4 layers for the EM modules, 3 layers for the fine HAD modules (4 in the EC) and 1 coarse HAD layer. Figure 3.8 displays the cell layout for one quadrant of the calorimeter system. Tables 3.1 and 3.2 list the radiation lengths and nuclear interaction lengths for the various layers.

Calo section	CC	EC
Number of layers	4	4
Layer radiation lengths (X_0)	1.4, 2.0, 6.8, 9.8	1.6, 2.6, 7.9, 9.3

Table 3.1: Radiation lengths of the EM layers.

The granularity of the calorimeter is such that each cell covers an area in $\eta-\phi$ space

Calo section	CC FH	CC CH	EC FH	EC CH
Number of layers	3	1	4	1
Layer interaction lengths (λ_A)	1.3, 1.0, 0.76	3.2	4×1.1	4.1

Table 3.2: Interaction lengths of the HAD layers.

of approximately 0.1×0.1. The third EM layer, however, is where the maximum of EM showers occurs, and so it has a finer segmentation of 0.05×0.05. The calorimeter delivers a relative energy resolution of $\sigma_E/E = 15\%/\sqrt{E} + 0.3\%$ for EM showers, and $\sigma_E/E = 45\%/\sqrt{E} + 4\%$ for HAD showers, where E is measured in GeV [23].

3.2.5 Intercryostat Detector

The use of separate cryostats for the three calorimeter sections means that the regions between them are not well-instrumented and also contain a large amount of dead material, resulting in poor coverage and energy resolution in the pseudorapidity range $0.8 < |\eta| < 1.4$. Additional sampling is therefore added in the form of the Intercryostat Detector (ICD). The ICD consists of 32 scintillating tiles with granularity $\Delta\eta \times \Delta\phi = 0.3 \times 0.4$, each of which is made up of twelve 0.1×0.1 sub-tiles. The ICD is attached to the exterior surface of the EC cryostat and provides coverage in the region $1.1 < |\eta| < 1.4$. Clear optical fibers carry the scintillator signal to photomultiplier tubes, which convert it to an electronic signal. The ICD signal electronics are designed to be compatible with the calorimeter readout system.

3.2.6 Muon Spectrometer

The muon system [19; 24] is the outermost subdetector of DØ , sitting outside the calorimeters. It provides detection and momentum measurement of muons independent of the central tracking system. Muons traverse the calorimeter as minimum

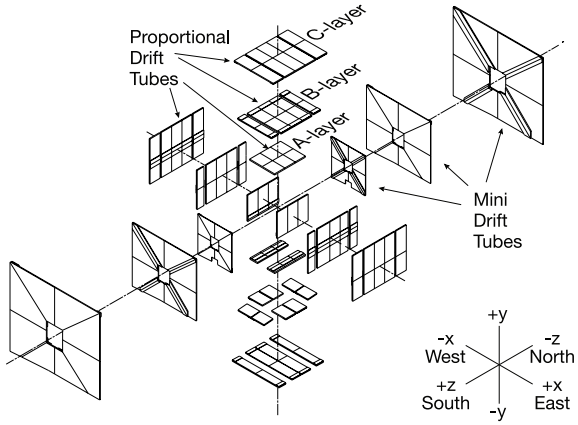

Figure 3.9: Blown-up view of the muon drift chambers

ionizing particles and generally lose a small fraction of their initial energy before reaching the muon system. Most other particles (electrons, photons and jets) are stopped in the calorimeter.

Three main components make up the muon system. Scintillation counters are used for triggering, drift tubes provide precise tracking in addition to triggering, and 1.8 T toroidal magnets bend muon tracks to allow momentum measurement. Figure 3.9 shows a blown-up view of the drift tube arrangement. The detector systems previously described in this chapter are all located inside the A-layer shown in Figure 3.9. The toroids of the muon system sit in between the A-layer and the B-layer.

The central muon tracking system, covering $|\eta| < 1.0$, utilizes Proportional Drift Tubes (PDTs) to reconstruct muon tracks. There are 6624 drift tubes of varying length in the central region, organized into 94 PDT chambers in 3 layers (A, B and C). The rectangular tubes have a 50 μm anode wire running down their center, along with cathode pads above and below the anode wire. They are filled with a gas mixture of 84% Argon, 8% CF_4 and 8% CH_4. When a muon passes through one of the drift tubes, it ionizes the gas. The anode wires are operated at 4.7 kV and the cathode pad electrodes are held at 2.3 kV, so the electrons in the ionized gas drift towards the anode and the positive gas ions drift towards the cathodes. Once

the ions reach their destinations, they create a measurable electronic signal. The muon's hit position along the length of the wire is determined using the electron drift time and the amount of charge deposition on the inner and outer cathode pads. A three-dimensional measurement of the muon's trajectory is achieved by using crossed $x \times y \times z$ tubes.

The forward muon tracking system, which extends coverage out to $|\eta| < 2.0$, uses Mini Drift Tubes (MDTs). The MDTs operate on the same principle described above. An MDT tube consists of eight 9.4×9.4 mm^2 cells, each with a 50 μm anode wire in the center. There are a total of 6080 tubes (or 48,640 anode wires) in the MDT system, again organized into three layers, each of which is further subdivided into eight octants (see Figure 3.9). The anodes are kept grounded, while the cathodes are operated at -3200 V. The MDT system is important for measuring muons which do not pass through all layers of the inner tracker.

Along with the drift tubes, the muon system employs scintillation counters in both the central and forward regions. Their layout is shown in Figure 3.10. In the central region, the cosmic cap and bottom counters are placed on the top, sides and bottom of the outermost layer of PDTs, the C-layer. There are a total of 372 scintillators, 0.6 m wide and 2.1 – 2.9 m long, in the cosmic cap and bottom. They provide a fast timing signal to associate a muon measured in the PDTs to an appropriate bunch crossing and to reject muons from cosmic rays. The A-layer of PDTs is also covered with scintillation counters, referred to as $A\phi$ counters, which also trigger on muons and reject out-of-time signals from cavern backgrounds. They additionally provide a time stamp for low-p_T muons that do not reach the cosmic cap or bottom counters. Their construction is similar to that of the cosmic cap and bottom, but requires finer segmentation as they sit closer to the detector center.

The forward muon scintillation counters are mounted on all three layers of the MDT system. Each layer is divided into octants containing about 96 counters each, for a total of about 4,600 counters in the forward system. Ranging in size from

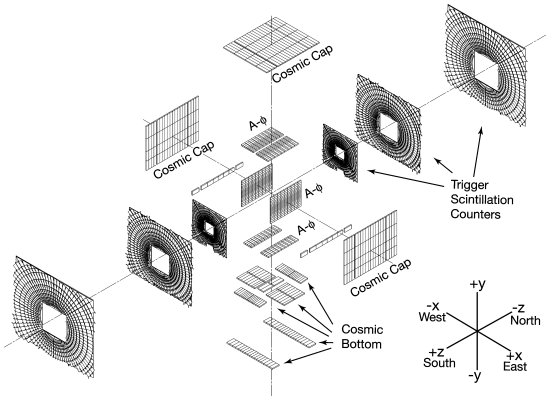

Figure 3.10: Blown-up view of the muon scintillation counters

17×24 cm^2 to 60×106 cm^2, they provide muon triggering and background rejection in the pseudorapidity range $1.0 < |\eta| < 2.0$.

The toroidal magnets of the muon system rest between layers A and B in both the central and forward regions. The central toroid is a square annulus 109 cm thick with an inner surface about 318 cm from the beam pipe. It is wound using twenty coils of ten turns each. The two end toroids are positioned at $454 \leq |z| \leq 610$ cm, and are wound with eight coils of eight turns each. The three toroids are operated in series at a current of 1500 A, producing a magnetic field of about 1.8 T. The polarity of the magnets is regularly reversed during data collection.

3.2.7 Luminosity Monitor

In order to provide an accurate measurement of the integrated luminosity recorded by the detector, luminosity monitors (LMs) are placed just in front of the endcap calorimeters at $|z| = 140$ cm. Figure 3.11 shows their placement along the beam pipe. They consist of arrays of 24 plastic scintillators each, which cover the pseudorapidity range $2.7 < |\eta| < 4.4$. The LMs are designed to measure the rate of inelastic $p\bar{p}$ collisions and therefore the instantaneous luminosity through the relation

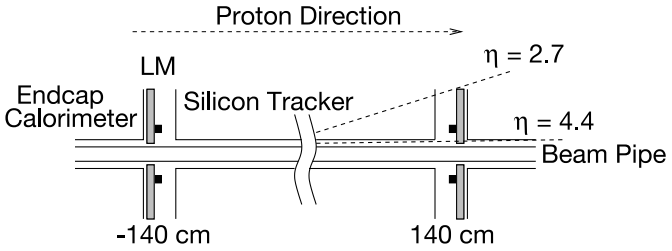

Figure 3.11: The DØ luminosity monitors.

$$L = f\frac{N_{LM}}{\sigma_{LM}}, \tag{3.7}$$

where f is the beam crossing frequency, N_{LM} is the average number of inelastic collisions per beam crossing, and σ_{LM} is the effective cross section for the LM, determined by the LM's acceptance and efficiency. The instantaneous luminosity L can then be integrated over all detector runs to determine the integrated luminosity $\mathcal{L}$. The z position of the primary vertex can also be measured using the difference between the time-of-flight measured for particles hitting the North and South monitors:

$$z_v = c\frac{t_- - t_+}{2}, \tag{3.8}$$

where c is the speed of light.

3.2.8 Trigger and Data Acquisition

The filled bunch crossing rate in the center of the detector is approximately 1.7 MHz. Since it is impossible to record detailed event data at such a rate, and since most bunch crossings do not result in uniquely interesting physics events, a 3-level trigger system is used to reduce the event rate to about 100 Hz. The Level 1 (L1) trigger is hardware-based, and it uses preliminary information from the tracking, calorimeter and muon systems to reduce the rate to about 2 kHz. The Level 2 (L2) trigger, consisting of processor cards and single-board computers (SBCs) arranged in VME crates, receives input from L1 and runs simple software algorithms to further reduce

the rate to 1 kHz. L2 has the capability of combining information from multiple subdetectors, including from the SMT. Finally, the Level 3 (L3) trigger runs more sophisticated algorithms on a computer farm, performing a partial reconstruction of the events and sending the accepted events for recording to tape at a rate of 100 Hz.

Trigger definitions are formed by requiring a certain set of criteria at each of the three levels. An example definition for a dielectron trigger is:

- L1: At least two EM objects with $E_T > 7$ GeV and $|\eta| < 3.2$, one of which has $E_T > 13$ GeV and satisfies isolation and EM fraction requirements.
- L2: At least one EM object with $E_T > 13$ GeV and a tighter isolation requirement.
- L3: At least two electrons with $E_T > 6$ GeV. One must have $E_T > 8$ GeV. One must be matched to a track with $p_T^{track} > 5$ GeV and pass loose shower shape requirements.

A trigger list is a combination of many trigger definitions, and different trigger lists and prescale sets are used depending on the instantaneous luminosity in the Tevatron in order to maintain an event rate that is as close as possible to 100 Hz. Every event written to tape has passed at least one trigger definition contained in the trigger list being used at the time the event was recorded.

The data acquisition system (L3DAQ) transports subdetector data from the VME readout crates to the L3 processing nodes, and finally on to tape storage. Coordination and control of triggering and data acquisition is handled by the main DØ run control program, COOR, which runs on the online host system. SBCs in the subdetector readout crates handle the routing of event fragments from the crates to the L3 farm nodes. The event builder (EVB) process constructs complete events from the various fragments, after which the events can be run through the L3 filtering process. Events which have passed the L3 trigger are assigned to various output streams based on which trigger components the event has satisfied. Each event is sent to only one

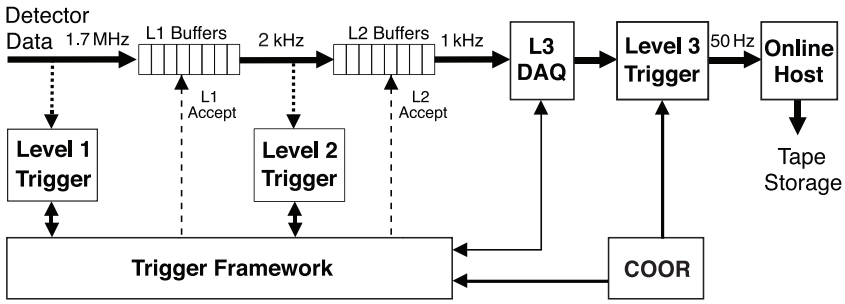

Figure 3.12: Overview of the DØ trigger and data acquisition systems

stream. These events are transmitted to the storage facility and written to tape. An overview of the data flow through the trigger and L3DAQ system is shown in Figure 3.12.

Chapter 4

Data Reconstruction and Object ID

This section describes the reconstruction of all objects used in the analysis. The DØ Offline Reconstruction Software Package (DØRECO) [25] is run on the triggered events and builds physics objects, stored according to the DØ Event Data Model [26].

4.1 Primary Vertex

Each bunch crossing at the center of the detector can contain multiple $p\bar{p}$ interactions, especially at high instantaneous luminosity. In order to determine the location of the hard scatter, or primary vertex (PV), which gave rise to the physics objects which passed the trigger in a given event, DØRECO uses an adaptive primary vertex algorithm [27].

First, tracks in the central tracking system with $p_T > 0.5$ GeV/c^2 are chosen with an additional requirement of 2 hits in the SMT if the track is within the SMT geometric acceptance. A clustering algorithm then identifies tracks from different interactions by clustering tracks that are within 2 cm of each other. A vertex fitting algorithm is then performed on the z-clusters, which determines the location and

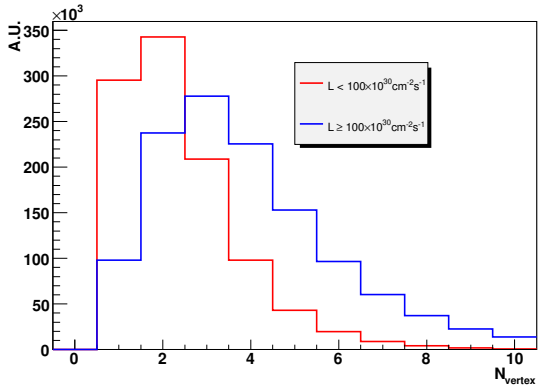

Figure 4.1: Number of primary vertices for low (in red) and high (in blue) luminosity ranges.

width of the beam, and tracks are preselected based on their distance of closest approach to the beam spot.

In the next step, the adaptive PV algorithm fits the preselected tracks to a common vertex. The track errors are then reweighted according to their individual χ^2 contributions to the vertex. Tracks with a χ^2 above a certain cutoff are dropped, and the fit is iteratively re-computed with the new weights and the new fitted vertex until convergence is reached, i.e. subsequent fits return the same vertex position. This process is performed for all clusters of tracks. The algorithm reconstructs primary vertices with an efficiency of close to 100% for $|z| < 80$ cm. In the final step, the reconstructed vertex with the lowest probability of coming from a minimum bias interaction [28] is chosen as the primary vertex. Figure 4.1 shows the number of primary vertices reconstructed for two different instantaneous luminosity regions.

4.2 Electrons

Electrons are reconstructed by combining calorimeter shower shape information with a track in the central tracking system. Various selection criteria are applied to clusters

of energy in the calorimeters to distinguish EM clusters from hadronic clusters due to jets. A track-match requirement is imposed in order to reject photons, which deposit energy in the calorimeters in a manner similar to electrons, but do not leave a track in the inner tracker, as they are not charged.

A Simple Cone Algorithm [29] builds calorimeter clusters from EM towers. EM towers are defined as the first five layers of calorimeter towers, i.e. the four EM layers and the first fine hadronic layer. The algorithm starts with a seed tower that has a minimum E_T of 0.5 GeV, and it then loops over all towers within a cone of radius $\Delta R < 0.4$ around the seed tower. It calculates the centroid of the cluster by weighting the towers by the energy deposited within them, and it then iteratively repeats the process with a cone drawn around this new centroid. A cluster is found once the position of the centroid is stable.

A set of electron identification criteria is defined to select high-quality EM clusters [30; 31; 32]. The clusters are required to deposit at least 90% of their energy in the EM calorimeter to reject jets. Additionally, they must have less than 15% of their energy in an annulus defined by the region between a cone of $R = 0.2$ and one of $R = 0.4$, as EM clusters tend to have a narrower transverse profile than hadronic clusters. To discriminate against photons that pass these selections, an EM cluster must be matched to a track with $p_T > 5$ GeV, and that track must point to a position along the z axis which is no more than 1 cm from the primary vertex chosen by the PV algorithm described in Section 4.1. The ratio of the calorimeter energy and the track momentum must also be less than 2.5.

Finally, multivariate discriminant variables formed from calorimeter shower shape measurements are used. The first is the"H-matrix" ($HMx7$) variable [30], built from the energy fractions deposited in all four EM layers, the total shower energy, the position of the primary vertex, and the shower width in the $r - \phi$ plane in the third EM layer. The second is the eight-variable EM likelihood ($LHood8$), constructed from a mix of tracker and calorimeter information (including $HMx7$). Figure 4.2

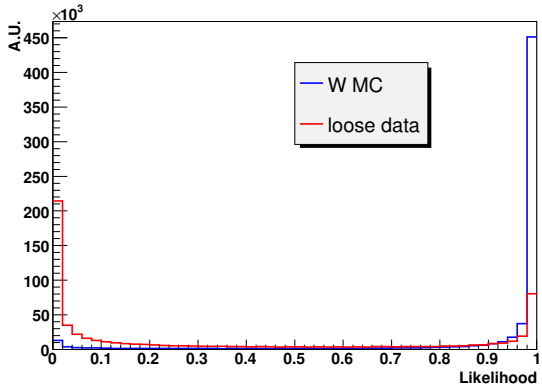

Figure 4.2: Electron $LHood8$ for W boson MC (in blue) and loosely selected data (in red), which is dominated by jets misidentified as electrons.

shows the $LHood8$ for electrons from W boson decays in Monte Carlo and for loosely selected data events.

Electrons passing all criteria except the $LHood8$ requirement are referred to as loose electrons. Those electrons which additionally satisfy $LHood8 > 0.85$ are referred to as tight. The tight selection has an efficiency of over 85% for signal electrons, with a fake rate from jets below 10%. Electrons in Monte Carlo (MC) simulation are corrected for differences between MC and data in reconstruction and identification efficiency.

4.3 Muons

Muons are identified by combining tracks in the muon spectrometer with tracks in the inner detectors. Specifically, the muon definition used in this analysis [33; 34] requires a track in the muon system that meets the following criteria:

- At least two A layer wire hits;
- At least one A layer scintillator hit;

- At least two B/C layer wire hits;
- at least one B/C layer scintillator hit.

This corresponds to the "mediumnseg3" requirement defined by the Muon ID Group, which has an overall efficiency of roughly 75% for isolated muons. The muon track must be matched to an inner detector track with a χ^2 per degree of freedom of less than 4. As with electrons, the inner detector track must be within 1 cm of the reconstructed primary vertex at its point of closest approach to the beam axis.

To reject cosmic rays, a timing requirement of $|t_{scint} - t_0| < 10$ ns is imposed for each layer of scintillation counters, where t_0 is the bunch crossing time and t_{scint} is the time recorded by the scintillator hits. Finally, muon candidates must be isolated from other tracks in the inner detector and from calorimeter activity in order to reject muons coming from heavy flavor decays, which usually also produce hadrons. Figure 4.3 shows calorimeter and tracker isolation for muons from W boson decays in MC and for loosely selected data. The tight isolation criteria are defined as follows: There must be less than 2.5 GeV of energy in the calorimeter in a cone of radius $R = 0.5$ around the muon's vector, extrapolated between the inner detector and the muon system. In addition, the scalar sum of the p_T of all tracks in the inner detector in a cone of the same size must also be less than 2.5 GeV. The loose isolation criterion uses an isolation variable defined as the calorimeter energy in a $R = 0.4$ cone around the muon plus the sum of the track momenta in a $R = 0.5$ cone, all divided by the muon's p_T. Muons in MC simulation are corrected for differences between MC and data in reconstruction and identification efficiency.

4.4 Jets

Quarks and gluons (collectively referred to as partons) produced in $p\bar{p}$ collisions do not travel through the detector as free particles, but quickly hadronize, producing

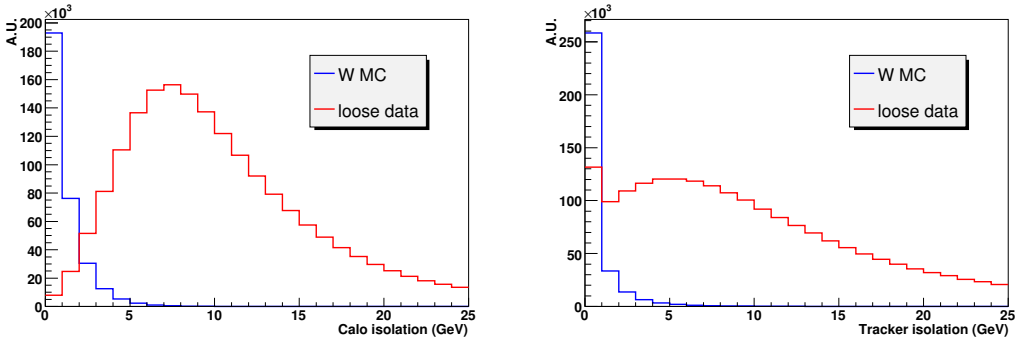

Figure 4.3: Muon calorimeter isolation (left) and tracker isolation (right) for W boson MC (in blue) and loosely selected data (in red), which is dominated by muons from heavy flavor decays.

energetic showers of mesons and baryons called jets. DØ software reconstructs jets from energy deposits in the calorimeters.

Before jet reconstruction, the T42 zero-suppression algorithm [35] is run to remove isolated energy deposits that are likely the result of calorimeter noise. This noise is caused by electronics noise from the calorimeter's 50,000 readout channels as well as decay of the uranium in the absorbing material. The T42 algorithm suppresses noise by requiring that calorimeter cells have an energy at least 4 standard deviations above the noise level ($E > 4\sigma_{noise}$). On average, only 0.003% of noisy cells, or about 1.5 cells per event, will pass this cut. Cells neighboring a 4σ cell are also accepted if they have energy greater than 2.5σ above the noise level.

The Run II midpoint cone algorithm [36] is run to reconstruct jets from the energy deposits which are not suppressed by T42. The midpoint cone algorithm works in a manner similar to that described for electrons in Section 4.2. However, in addition to using hadronic calorimeter towers with energy above a given threshold as seeds, the algorithm also seeds jets using the midpoint in $\eta - \phi$ space between two such towers. This jet-finding technique protects the algorithm, at first order, from so-called infrared sensitivity (see Figure 4.4), which causes simple cone algorithms to

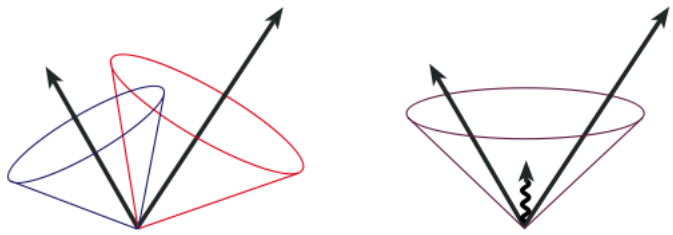

Figure 4.4: An illustration of infrared sensitivity. A simple cone algorithm that reconstructs two jets from two large energy deposits will merge them into one jet in the case of soft gluon radiation between the two deposits.

incorrectly merge jets in the case of soft gluon radiation between them. Jets cannot have more than 40% of their energy in the coarse hadronic layer of the calorimeter, as noise levels tend to be higher in that section. They also must have at least 5% of their energy in the EM layer, but not overlap with any reconstructed electrons in $\eta - \phi$ space.

DØRECO does not recover the full particle-level energies (i.e. the sum of the energies of all particles making up the jet), so a jet energy scale correction is applied to the reconstructed jets [37]. Jets in Monte Carlo (MC) simulation are further corrected for differences between MC and data in reconstruction and identification efficiency, energy resolution and calorimeter response. This is accomplished through a standard jet shifting, smearing and removal algorithm (JSSR) [38], which also removes jets with $p_T < 15$ GeV.

One final selection criteria can be applied to reconstructed jets. In running periods of higher luminosity, it can be advantageous to require that jets be *vertex-confirmed*, i.e. matched to the primary vertex of the event. A jet is considered vertex-confirmed if there are at least two tracks within the jet cone radius that point to the primary vertex. This selection rejects jets produced in other $p\bar{p}$ interactions from the same bunch crossing. Figure 4.5 shows the number of jets failing vertex confirmation (i.e. coming from other $p\bar{p}$ interactions) for high and low luminosity.

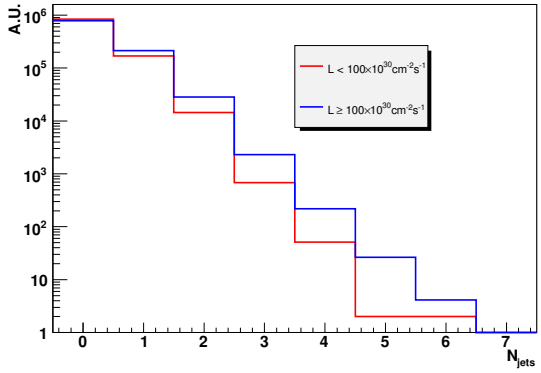

Figure 4.5: Number of non-vertex-confirmed jets for low (in red) and high (in blue) luminosity ranges.

4.5 Missing Transverse Energy

Neutrinos (and other particles that do not interact strongly or electromagnetically) do not interact with the detector and carry away momentum that can be partially inferred from momentum conservation in the transverse plane. The negative sum of the transverse momenta of all particles observed in the detector is associated with the sum of the transverse momenta of any neutrinos in the event. A similar inference cannot be made about the longitudinal momentum, as the longitudinal momentum of the partons involved in the hard scatter is not usually known.

In practice, missing transverse energy ($\not\!\!E_T$) is computed through a vector sum of the energies deposited in all the EM and fine HAD cells of the calorimeter (after application of the T42 algorithm) [39]. Cells in the coarse HAD calorimeter are only included in the calculation if they are part of a reconstructed jet passing all selection cuts. Any corrections applied to reconstructed electrons and jets are also propagated to the $\not\!\!E_T$. The lone exception is JSSR smearing, as applying this to the $\not\!\!E_T$ resulted in worse agreement between data and Monte Carlo. Since muons only deposit a small fraction of their energy in the calorimeters, the $\not\!\!E_T$ is also corrected for any muons passing the selection criteria. Figure 4.6 shows the observed linear

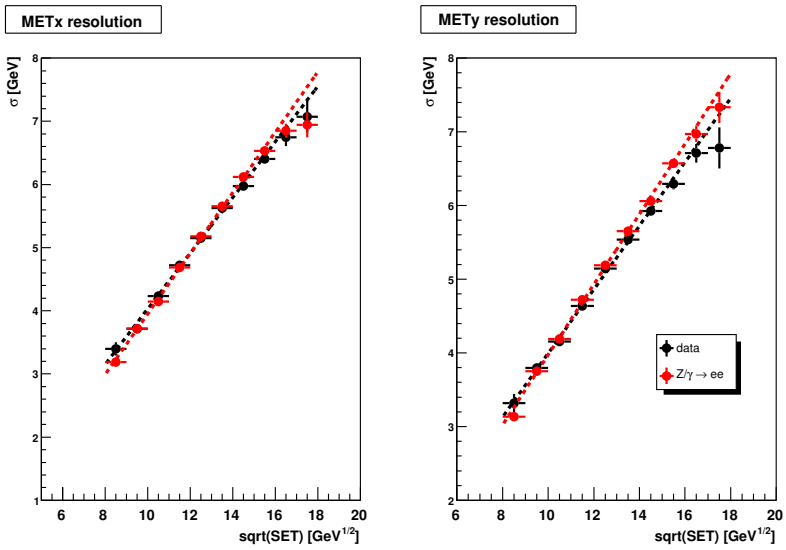

Figure 4.6: $\not\!\!E_T$ resolution in GeV as a function of the square root of the scalar sum of E_T.

relationship between the $\not\!\!E_T$ resolution along the x and y axes and the square root of the scalar sum of E_T from all objects in the event.

Chapter 5

Data and Monte Carlo Samples

5.1 Dataset

The data used in this analysis were recorded from July 2002 to June 2009. This comprises two separate run periods: Run IIa (or p17) data were recorded from July 2002 to February 2006 and correspond to about 1.1 fb^{-1}, and Run IIb (or p20) data were recorded from June 2006 to July 2009 and correspond to roughly 4.3 fb^{-1}. The period from February to June 2006 saw an upgrade of the detector, including installation of a new central tracking system and a significant upgrade to the muon system, therefore background modeling is done separately for the two run periods.

As our model for the signal contains two vector-like quarks which decay with 100% branching ratios to different decay products (see Section 2.2), the search is conducted in two final states: $(W \to \ell\nu)qq$, or "lepton+jets," and $(Z \to \ell\ell)qq$, or "dilepton+jets." The data in the lepton+jets final states are taken from the p17 and p20 EMinclusive and MUinclusive skims produced by the DØ Common Samples Group [40], and represent 5.4 fb^{-1} of Run II data. Data for the dilepton final states are obtained from the p17 and p20 2EMhighpt and 2MUhighpt skims, also representing 5.4 fb^{-1}. These skims are subsets of the complete data set recorded by DØ; Table 5.1 shows the total number of events in each skim. Events are placed into different skims

based on the reconstructed physics objects that they contain. Events in the skims used in the lepton+jets and dilepton+jets analysis meet the following criteria:

- EMinclusive: one EM object with $p_T > 20$ GeV **or**

  one EM object with $p_T > 8$ GeV and track $p_T > 5$ GeV **or**

  two EM objects each with $p_T > 7$ GeV **or**

  one EM object with $p_T > 12$ GeV and $Hmx7 < 75$ and a jet with $p_T > 8$ GeV.

- MUinclusive: one loose muon with $p_T > 8$ GeV **or**

  one loose muon with $p_T > 5$ GeV and two tracks with $p_T > 8, 5$ GeV **or**

  two loose muons with no p_T requirement.

- 2EMhighpt: two EM objects with $p_T > 12$ GeV **or**

  one EM object with $p_T > 15$ GeV and a track with $p_T > 7$ GeV.

- 2MUhighpt: two loose muons with $p_T > 10$ GeV **or**

  two loose muons, one with $p_T > 15$ GeV **or**

  one medium muon with $p_T > 15$ GeV and two tracks each with $p_T > 15$ GeV.

Skim name	Number of events (millions)
EMinclusive	980.3
MUinclusive	1525
2EMhighpt	129.4
2MUhighpt	69.36

Table 5.1: Initial number of events in each data skim.

Only good quality events are selected from these skims and used in the analysis, using criteria determined by the Data Quality (DQ) Group [41]. These criteria include complete detector read-out and the absence of known electronic noise patterns. All runs marked as *bad* by the SMT, CFT, Calorimeter, and Muon systems are removed. Luminosity blocks which are marked as *bad* by offline calorimeter selection are also removed from the analysis. Additionally, individual events with known noise patterns in the calorimeter are removed.

The dilepton analyses make no explicit trigger requirement, because the overall trigger efficiency for this channel is expected to be very high owing to the presence of two relatively high-p_T leptons in the final state. However, this is not the case in the lepton+jets analyses, so trigger requirements are imposed. The μ+jets analysis selects events using a logical OR of all triggers which require at least one muon, while the e+jets analysis uses a similar logical OR selecting events with have passed at least one electron+jets trigger. These choices were made in large part based on the success other DØ analyses have had in modeling the effects of using these trigger suites in Monte Carlo background predictions, as all signal and background MC must be corrected for trigger efficiencies. In all analyses the luminosity is determined by the fully unprescaled JT125 trigger. Table 5.2 shows the luminosity for the dilepton and lepton+jets analyses.

Run Period	Delivered [pb^{-1}]	Recorded [pb^{-1}]	Good DQ [pb^{-1}]
July 2002 - February 2006	1347.48	1244.66	1078.81
June 2006 - March 2008	1936.43	1787.37	1625.61
March 2008 - July 2009	2942.17	2823.22	2661.89
July 2002 - July 2009	6226.01	5855.19	5360.47

Table 5.2: Integrated luminosity delivered to and recorded by DØ and available for the analysis (good DQ) in units of pb^{-1}.

5.2 Monte Carlo Samples for Signal and Background

The vector-like quark signal Monte Carlo samples were generated using MADGRAPH [42] with CTEQ6L1 [43] parton distribution functions (PDF). The implementation of the model includes an up-type vector-like quark with charge 2/3 and a down-type vector-like quark with charge -1/3. Couplings to SM quarks are set by the model-independent parameter κ_{qQ}:

$$\kappa_{qQ} = \frac{v}{m_Q}\tilde{\kappa}_{qQ}, \tag{5.1}$$

where $\tilde{\kappa}_{qQ}$ encodes the model-dependence of the coupling and is set to 1. We produce samples with degenerate vector-like quarks, such that $Br[Q_D \to Wq] = Br[Q_U \to Zq] = 100\%$ [9], equivalent to $\tilde{\kappa}_{uU} = \tilde{\kappa}_{uD} = 1$ and $\tilde{\kappa}_{dU} = \tilde{\kappa}_{dD} = 0$. The final cross-section limits for $Q \to Wq$ and $Q \to Zq$ are independent of this choice.

Heavy quark widths are mass-dependent, and are numerically calculated using BRIDGE [44]. Parton shower evolution is performed with PYTHIA [45]. Signal samples are produced with vector-like quark masses from 280 GeV/c^2 up to 700 GeV/c^2.

Tables 5.3 and 5.4 show the cross-section, branching ratio, and number of generated events for each vector-like quark type and mass point. The branching ratios for $Q_D \to Wq \to \ell\nu q$ and $Q_U \to Zq \to \ell\ell q$ are due to the forcing of leptonic gauge boson decay in the second step, as the branching ratios for the first step are 100% in both cases.

The W/Z+jets and $t\bar{t}$ ($m_t = 172.5$ GeV/c^2) background samples are generated using ALPGEN [46] in tandem with the MLM jet-matching prescription [47]. ALPGEN is used as the generator for the primary backgrounds (W/Z+jets) due to its ability to model events with multiple hard jets via a $2 \to n$ hard scatter (rather than the $2 \to 2$ processes generated by, for example, PYTHIA). The matching algorithm ensures that hard jets (i.e. jets with large p_T) are generated by ALPGEN at the parton level and that PYTHIA fills in, via the parton shower, regions of phase space underpopulated by

$p\bar{p} \to Q_D q \to \ell\nu qq$ MADGRAPH Monte Carlo Sets			
Q_D Mass [GeV/c^2]	σ [pb]	BR	Number of Events
280	10.5	0.213	19745 (p17)
			19045 (p20)
320	5.52		18963 (p17)
			19213 (p20)
360	3.03		19155 (p17)
			18747 (p20)
400	1.73		12462 (p17)
			9969 (p20)
450	0.867		19801 (p17)
			9097 (p20)
500	0.450		19620 (p17)
			9767 (p20)
550	0.231		19671 (p17)
			8257 (p20)
600	0.124		19825 (p17)
			9100 (p20)
650	0.0637		18922 (p17)
			9923 (p20)
700	0.0348		14778 (p17)
			15342 (p20)

Table 5.3: The cross sections, branching fractions and initial numbers of events of the $Q_D q \to Wqq \to \ell\nu qq$ MADGRAPH Monte Carlo samples for Run IIa (p17) and Run IIb (p20). Branching fractions are the same for all listed samples.

$p\bar{p} \to Q_U q \to \ell\ell qq$ MADGRAPH Monte Carlo Sets			
Q_U Mass [GeV/c^2]	σ [pb]	BR	Number of Events
280	3.10	0.067	19360 (p17)
			11147 (p20)
320	1.60		19548 (p17)
			19213 (p20)
360	0.863		19545 (p17)
			18747 (p20)
400	0.480		18741 (p17)
			9380 (p20)
450	0.237		18695 (p17)
			9096 (p20)
500	0.119		19624 (p17)
			8870 (p20)
550	0.0608		19516 (p17)
			9123 (p20)
600	0.0315		19346 (p17)
			9668 (p20)
650	0.0160		19654 (p17)
			9846 (p20)
700	0.0085		15252 (p17)
			12805 (p20)

Table 5.4: The cross sections, branching fractions and initial numbers of events of the $Q_U q \to Zqq \to \ell\ell qq$ MADGRAPH Monte Carlo samples.

the fixed-order calculation. Both the W and Z + heavy flavor (i.e. $b\bar{b}$ and $c\bar{c}$) samples are generated separately from light flavor (i.e. $udsg$) samples. Because PYTHIA is allowed to generate heavy flavor from gluon splitting, such events are removed from the light parton samples to avoid double-counting. The procedure for removing these events is documented in Ref. [48]. The $t\bar{t}$ samples are scaled to the next-to-next-to-leading-order (NNLO) [49] cross section for $m_t = 172.5$ GeV/c^2. The individual W and Z+$n$lp samples are combined by their relative leading logarithm cross sections as determined by ALPGEN and then further scaled by the commonly used ALPGEN k-factors [50] shown in Table 5.5 to correct the cross sections to next-to-leading-order.

k-factors for W/Z+jets ALPGEN Samples

Process	Reco	k-factor
Z+jets	p17	1.30
	p20	1.30
$Z+c\bar{c}$+jets	p17	1.30×1.67
	p20	1.30×1.67
$Z+b\bar{b}$+jets	p17	1.30×1.52
	p20	1.30×1.52
W+jets	p17	1.30
	p20	1.30
$W+c\bar{c}$+jets	p17	1.30×1.47
	p20	1.30×1.47
$W+b\bar{b}$+jets	p17	1.30×1.47
	p20	1.30×1.47

Table 5.5: k-factors for ALPGEN W and Z+jets samples.

Single top ($m_t = 172.5$ GeV/c^2) background samples, used in the lepton + jets analysis, are generated using COMPHEP [51], with parton shower evolution performed by PYTHIA [52].

The Monte Carlo samples for diboson (i.e. WW, $W + Z$, and ZZ) production are generated using PYTHIA with the CTEQ6L1 PDF set. All samples are decayed inclusively. The samples are scaled to the NNLO cross sections [53].

For all processes, B-hadrons are decayed using EVTGEN [54] and tau leptons are decayed using TAUOLA [55].

All Monte Carlo samples are processed through a GEANT-based [56] simulation of the DØ detector, and are overlayed with data events from randomly chosen bunch crossings to simulate the effect of additional $p\bar{p}$ collisions and detector noise. The instantaneous luminosity profile of these events is chosen to match the profile of Run IIa and Run IIb triggered data.

Tables 5.6–5.22 contain listings of all Monte Carlo background samples used, including subprocesses, cross sections times branching ratios and numbers of events.

$t\bar{t}$ Monte Carlo Sets			
Event Type	Subprocess	$\sigma \cdot$ BR [pb]	Number of Events
$t\bar{t}$ ($m_t = 172.5\ GeV/c^2$)			
	$t\bar{t} + 0\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 0\text{lp}$	0.349	1516107 (p17)
	$t\bar{t} + 0\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 0\text{lp}$	0.352	749542 (p20)
	$t\bar{t} + 1\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 1\text{lp}$	0.147	963057 (p17)
	$t\bar{t} + 1\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 1\text{lp}$	0.142	452177 (p20)
	$t\bar{t} + 2\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 2\text{lp}$	0.112	701167 (p17)
	$t\bar{t} + 2\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 2\text{lp}$	0.068	281453 (p20)
	$t\bar{t} + 0\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 0\text{lp}$	1.398	771271 (p17)
	$t\bar{t} + 0\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 0\text{lp}$	1.400	777068 (p20)
	$t\bar{t} + 1\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 1\text{lp}$	0.589	492647 (p17)
	$t\bar{t} + 1\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 1\text{lp}$	0.577	457782 (p20)
	$t\bar{t} + 2\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 2\text{lp}$	0.422	288992 (p17)
	$t\bar{t} + 2\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 2\text{lp}$	0.267	321166 (p20)

Table 5.6: The cross sections times branching fractions and the initial numbers of events of the $t\bar{t}$ ALPGEN Monte Carlo samples used in dilepton analyses for Run IIa (p17) and Run IIb (p20). Branching fractions are the same for all listed samples.

$t\bar{t}$ Monte Carlo Sets			
Event Type	Subprocess	$\sigma \cdot$ BR [pb]	Number of Events
$t\bar{t}$ ($m_t = 170\ GeV/c^2$)			
	$t\bar{t} + 0\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 0\text{lp}$	0.375	301180 (p17)
	$t\bar{t} + 0\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 0\text{lp}$	0.379	1506566 (p20)
	$t\bar{t} + 1\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 1\text{lp}$	0.156	149465 (p17)
	$t\bar{t} + 1\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 1\text{lp}$	0.153	918361 (p20)
	$t\bar{t} + 2\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 2\text{lp}$	0.072	74942(p17)
	$t\bar{t} + 2\text{lp} \to 2\ell\ 2\nu\ 2\text{b} + 2\text{lp}$	0.0853	288840 (p20)
	$t\bar{t} + 0\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 0\text{lp}$	1.55	977226 (p17)
	$t\bar{t} + 0\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 0\text{lp}$	1.52	871904 (p20)
	$t\bar{t} + 1\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 1\text{lp}$	0.600	877783 (p17)
	$t\bar{t} + 1\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 1\text{lp}$	0.624	600399 (p20)
	$t\bar{t} + 2\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 2\text{lp}$	0.312	302011 (p17)
	$t\bar{t} + 2\text{lp} \to \ell\nu\ 2\text{b}\ \text{q}\bar{\text{q}}' + 2\text{lp}$	0.307	349273 (p20)

Table 5.7: The cross sections times branching fractions and the initial numbers of events of the $t\bar{t}$ ALPGEN Monte Carlo samples used in the ℓ+jets analyses.

s-channel Single Top Monte Carlo Sets			
Event Type	Subprocess	$\sigma \cdot$ BR [pb]	Number of Events
tb	$tb \to e\nu bb$	0.363	291094 (p17)
($m_t = 172.5\ GeV/c^2$)		0.363	248415 (p20)
	$tb \to \mu\nu bb$	0.363	289002 (p17)
		0.363	226343 (p20)
	$tb \to \tau\nu bb$	0.363	288909 (p17)
		0.363	249947 (p20)

Table 5.8: The cross sections times branching fractions and the initial numbers of events of the s-channel (tb) single top CompHEP Monte Carlo samples used in the ℓ+jets analyses.

t-channel Single Top Monte Carlo Sets			
Event Type	Subprocess	$\sigma \cdot$ BR [pb]	Number of Events
tqb	$tqb \to e\nu bqb$	0.758	290065 (p17)
($m_t = 172.5\ GeV/c^2$)		0.758	273518 (p20)
	$tqb \to \mu\nu bqb$	0.758	289311 (p17)
		0.758	274418 (p20)
	$tqb \to \tau\nu bqb$	0.758	290066 (p17)
		0.758	247397 (p20)

Table 5.9: The cross sections times branching fractions and the initial numbers of events of the t-channel (tqb) single top CompHEP Monte Carlo samples used in the ℓ+jets analyses.

Diboson Monte Carlo Sets			
Event Type	Subprocess	$\sigma \cdot$ BR [pb]	Number of Events
WW	Inclusive	11.6	2053448 (p17)
			593126 (p20)
WZ	Inclusive	3.25	412762 (p17)
			434206 (p20)
ZZ	Inclusive	1.33	296491 (p17)
			271675 (p20)

Table 5.10: The cross sections times branching fractions and the initial numbers of events of the diboson PYTHIA Monte Carlo samples.

W+jets Monte Carlo Sets		
Subprocess	$\sigma \cdot$ BR [pb]	Number of Events
$W \to \ell\nu$ + 0 lp	4520	12526514 (p17)
	4510	40274534 (p20)
$W \to \ell\nu$ + 1 lp	1280	9425157 (p17)
	1280	8365020 (p20)
$W \to \ell\nu$ + 2 lp	304	4285002 (p17)
	304	5936329 (p20)
$W \to \ell\nu$ + 3 lp	72.7	3535816 (p17)
	72.6	3754272 (p20)
$W \to \ell\nu$ + 4 lp	16.8	2493506 (p17)
	16.8	2602738 (p20)
$W \to \ell\nu$ + 5 lp	5.08	780938 (p17)
	5.15	2044335 (p20)

Table 5.11: The cross sections times branching fractions and the initial numbers of events of the W+jets ALPGEN Monte Carlo samples.

W+jets Monte Carlo Sets			
Event Type	Subprocess	$\sigma \cdot$ BR [pb]	Number of Events
$W + c\bar{c}$			
	$W \to \ell\nu + c\bar{c} + 0$ lp	23.3	1205996 (p17)
	$W \to \ell\nu + c\bar{c} + 0$ lp	24.5	934253 (p20)
	$W \to \ell\nu + c\bar{c} + 1$ lp	13.9	740093 (p17)
	$W \to \ell\nu + c\bar{c} + 1$ lp	13.5	738709 (p20)
	$W \to \ell\nu + c\bar{c} + 2$ lp	5.57	342472 (p17)
	$W \to \ell\nu + c\bar{c} + 2$ lp	5.50	554236 (p20)
	$W \to \ell\nu + c\bar{c} + 3$ lp	2.24	446288 (p17)
	$W \to \ell\nu + c\bar{c} + 3$ lp	2.53	469900 (p20)
$W + b\bar{b}$			
	$W \to \ell\nu + b\bar{b} + 0$ lp	9.30	1372108 (p17)
	$W \to \ell\nu + b\bar{b} + 0$ lp	9.37	1104413 (p20)
	$W \to \ell\nu + b\bar{b} + 1$ lp	4.14	666553 (p17)
	$W \to \ell\nu + b\bar{b} + 1$ lp	4.30	782487 (p20)
	$W \to \ell\nu + b\bar{b} + 2$ lp	1.58	248702 (p17)
	$W \to \ell\nu + b\bar{b} + 2$ lp	1.57	523717 (p20)
	$W \to \ell\nu + b\bar{b} + 3$ lp	0.763	276900 (p17)
	$W \to \ell\nu + b\bar{b} + 3$ lp	0.724	412747 (p20)

Table 5.12: The cross sections times branching fractions and the initial numbers of events of the $W + c\bar{c}$ and $W + b\bar{b}$ ALPGEN Monte Carlo samples.

Z+0lp Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to e^+e^-$ + 0lp	$15 \leq M_Z < 75$	337	532867 (p17)
		337	1580862 (p20)
	$75 \leq M_Z < 130$	132	2884817 (p17)
		132	1006195 (p20)
	$130 \leq M_Z < 250$	0.891	288923 (p17)
		0.891	352445 (p20)
	$250 \leq M_Z < 1960$	0.0692	199458 (p17)
		0.0668	723691 (p20)
$Z \to \mu^+\mu^-$ + 0lp	$15 \leq M_Z < 75$	336	533270 (p17)
		344	1556802 (p20)
	$75 \leq M_Z < 130$	132	2894120 (p17)
		134	1251862 (p20)
	$130 \leq M_Z < 250$	0.881	473275 (p17)
		0.887	351275 (p20)
	$250 \leq M_Z < 1960$	0.0677	197860 (p17)
		0.0694	549457 (p20)
$Z \to \tau^+\tau^-$ + 0lp	$15 \leq M_Z < 75$	337	534038 (p17)
		338	1113728 (p20)
	$75 \leq M_Z < 130$	133	2848048 (p17)
		131	968412 (p20)
	$130 \leq M_Z < 250$	0.885	281921 (p17)
		0.922	358601 (p20)
	$250 \leq M_Z < 1960$	0.0658	192995 (p17)
		0.0672	535656 (p20)

Table 5.13: The cross sections times branching fractions and the initial numbers of events of the Z+0lp ALPGEN Monte Carlo samples.

Z+1lp Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to e^+e^-$ + 1lp	$15 \leq M_Z < 75$	40.3	426271 (p17)
		40	956175 (p20)
	$75 \leq M_Z < 130$	40.8	1799040 (p17)
		40.1	566876 (p20)
	$130 \leq M_Z < 250$	0.373	181686 (p17)
		0.376	179033 (p20)
	$250 \leq M_Z < 1960$	0.0344	94645 (p17)
		0.0358	520452 (p20)
$Z \to \mu^+\mu^-$ + 1lp	$15 \leq M_Z < 75$	39.7	429859 (p17)
		40.1	570408 (p20)
	$75 \leq M_Z < 130$	40.6	1917640 (p17)
		41.4	604493 (p20)
	$130 \leq M_Z < 250$	0.347	369868 (p17)
		0.359	170242 (p20)
	$250 \leq M_Z < 1960$	0.0374	95823 (p17)
		0.035	440553 (p20)
$Z \to \tau^+\tau^-$ + 1lp	$15 \leq M_Z < 75$	39.9	428065 (p17)
		39.9	528075 (p20)
	$75 \leq M_Z < 130$	40.6	1881530 (p17)
		40.3	562760 (p20)
	$130 \leq M_Z < 250$	0.34	183371 (p17)
		0.375	171211 (p20)
	$250 \leq M_Z < 1960$	0.0353	90716 (p17)
		0.0363	461239 (p20)

Table 5.14: The cross sections times branching fractions and the initial numbers of events of the Z+1lp ALPGEN Monte Carlo samples.

Z+2lp Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to e^+e^-$ + 2lp	$15 \leq M_Z < 75$	10	162046 (p17)
		9.95	549058 (p20)
	$75 \leq M_Z < 130$	10	881188 (p17)
		9.98	267939 (p20)
	$130 \leq M_Z < 250$	0.0954	87783 (p17)
		0.102	160466 (p20)
	$250 \leq M_Z < 1960$	0.012	44507 (p17)
		0.0112	325986 (p20)
$Z \to \mu^+\mu^-$ + 2lp	$15 \leq M_Z < 75$	9.92	162803 (p17)
		9.87	275077 (p20)
	$75 \leq M_Z < 130$	9.81	956045 (p17)
		9.91	400558 (p20)
	$130 \leq M_Z < 250$	0.096	266045 (p17)
		0.0984	160267 (p20)
	$250 \leq M_Z < 1960$	0.011	43853 (p17)
		0.0105	326442 (p20)
$Z \to \tau^+\tau^-$ + 2lp	$15 \leq M_Z < 75$	9.94	163473 (p17)
		10	279564 (p20)
	$75 \leq M_Z < 130$	10	864222 (p17)
		9.99	274324 (p20)
	$130 \leq M_Z < 250$	0.0993	86752 (p17)
		0.0972	162320 (p20)
	$250 \leq M_Z < 1960$	0.0113	43695 (p17)
		0.0109	329708 (p20)

Table 5.15: The cross sections times branching fractions and the initial numbers of events of the Z+2lp ALPGEN Monte Carlo samples.

Z+3lp Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to e^+e^-$ + 3lp	$15 \leq M_Z < 75$	2.76	78677 (p17)
		2.77	536343 (p20)
	$75 \leq M_Z < 130$	3.15	835696 (p17)
		3.3	127038 (p20)
	$130 \leq M_Z < 250$	0.0319	82057 (p17)
		0.0331	300408 (p20)
	$250 \leq M_Z < 1960$	0.0039	39630 (p17)
		0.00404	301603 (p20)
$Z \to \mu^+\mu^-$ + 3lp	$15 \leq M_Z < 75$	2.81	78219 (p17)
		2.84	267966 (p20)
	$75 \leq M_Z < 130$	3.06	934623 (p17)
		3.25	146250 (p20)
	$130 \leq M_Z < 250$	0.0349	249232 (p17)
		0.0335	141929 (p20)
	$250 \leq M_Z < 1960$	0.00407	56575 (p17)
		0.00401	301395 (p20)
$Z \to \tau^+\tau^-$ + 3lp	$15 \leq M_Z < 75$	2.78	77433 (p17)
		2.77	278238 (p20)
	$75 \leq M_Z < 130$	3.18	828424 (p17)
		3.1	173657 (p20)
	$130 \leq M_Z < 250$	0.0322	83736 (p17)
		0.0362	157799 (p20)
	$250 \leq M_Z < 1960$	0.00387	41067 (p17)
		0.00406	146464 (p20)

Table 5.16: The cross sections times branching fractions and the initial numbers of events of the Z+3lp ALPGEN Monte Carlo samples.

$Z \to ee + c\bar{c}$ Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to e^+e^-$ + 0lp	$15 \leq M_Z < 75$	4.08	194772 (p17)
		4.09	182027 (p20)
	$75 \leq M_Z < 130$	0.9	202071 (p17)
		0.901	182485 (p20)
	$130 \leq M_Z < 250$	0.00743	95499 (p17)
		0.0075	181778 (p20)
	$250 \leq M_Z < 1960$	0.000661	192390 (p17)
		0.000645	230882 (p20)
$Z \to e^+e^-$ + 1lp	$15 \leq M_Z < 75$	1.04	87420 (p17)
		1.03	180347 (p20)
	$75 \leq M_Z < 130$	0.478	104998 (p17)
		0.506	89293 (p20)
	$130 \leq M_Z < 250$	0.00457	47903 (p17)
		0.00433	91938 (p20)
	$250 \leq M_Z < 1960$	0.000434	194033 (p17)
		0.000468	228935 (p20)
$Z \to e^+e^-$ + 2lp	$15 \leq M_Z < 75$	0.347	101246 (p17)
		0.382	172926 (p20)
	$75 \leq M_Z < 130$	0.289	48597 (p17)
		0.286	47357 (p20)
	$130 \leq M_Z < 250$	0.0028	48773 (p17)
		0.00267	93171 (p20)
	$250 \leq M_Z < 1960$	0.000288	193856 (p17)
		0.000306	231974 (p20)

Table 5.17: The cross sections times branching fractions and the initial numbers of events of the $Z \to ee + c\bar{c}$ ALPGEN Monte Carlo samples.

$Z \to ee + b\bar{b}$ Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to e^+e^-$ + 0lp	$15 \leq M_Z < 75$	0.546	201637 (p17)
		0.518	180549 (p20)
	$75 \leq M_Z < 130$	0.395	200830 (p17)
		0.401	200121 (p20)
	$130 \leq M_Z < 250$	0.00349	106152 (p17)
		0.0034	90970 (p20)
	$250 \leq M_Z < 1960$	0.000316	193233 (p17)
		0.000305	190731 (p20)
$Z \to e^+e^-$ + 1lp	$15 \leq M_Z < 75$	0.192	98780 (p17)
		0.207	90826 (p20)
	$75 \leq M_Z < 130$	0.193	100608 (p17)
		0.173	97474 (p20)
	$130 \leq M_Z < 250$	0.00171	11656 (p17)
		0.00167	47734 (p20)
	$250 \leq M_Z < 1960$	0.000184	201073 (p17)
		0.000201	183718 (p20)
$Z \to e^+e^-$ + 2lp	$15 \leq M_Z < 75$	0.0813	102688 (p17)
		0.0783	91895 (p20)
	$75 \leq M_Z < 130$	0.0973	49663 (p17)
		0.107	48269 (p20)
	$130 \leq M_Z < 250$	0.000969	48519 (p17)
		0.000938	48934 (p20)
	$250 \leq M_Z < 1960$	0.000113	193429 (p17)
		9.5e-05	181942 (p20)

Table 5.18: The cross sections times branching fractions and the initial numbers of events of the $Z \to ee + b\bar{b}$ ALPGEN Monte Carlo samples.

$Z \to \mu\mu + c\bar{c}$ Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to \mu^+\mu^-$ + 0lp	$15 \leq M_Z < 75$	4.14	196941 (p17)
		4.14	180580 (p20)
	$75 \leq M_Z < 130$	0.931	193789 (p17)
		0.932	193928 (p20)
	$130 \leq M_Z < 250$	0.00745	96075 (p17)
		0.00756	79493 (p20)
	$250 \leq M_Z < 1960$	0.000643	194548 (p17)
		0.000622	191017 (p20)
$Z \to \mu^+\mu^-$ + 1lp	$15 \leq M_Z < 75$	1.12	100243 (p17)
		0.953	93093 (p20)
	$75 \leq M_Z < 130$	0.504	101903 (p17)
		0.548	92744 (p20)
	$130 \leq M_Z < 250$	0.0046	47913 (p17)
		0.00439	45857 (p20)
	$250 \leq M_Z < 1960$	0.000445	192766 (p17)
		0.000439	170822 (p20)
$Z \to \mu^+\mu^-$ + 2lp	$15 \leq M_Z < 75$	0.374	97406 (p17)
		0.343	95436 (p20)
	$75 \leq M_Z < 130$	0.293	50563 (p17)
		0.281	51277 (p20)
	$130 \leq M_Z < 250$	0.00278	48783 (p17)
		0.00283	47946 (p20)
	$250 \leq M_Z < 1960$	0.000279	193214 (p17)
		0.000262	181158 (p20)

Table 5.19: The cross sections times branching fractions and the initial numbers of events of the $Z \to \mu\mu + c\bar{c}$ ALPGEN Monte Carlo samples.

$Z \to \mu\mu + b\bar{b}$ Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to \mu^+\mu^-$ + 0lp	$15 \leq M_Z < 75$	0.514	265270 (p17)
		0.509	188546 (p20)
	$75 \leq M_Z < 130$	0.404	193991 (p17)
		0.424	210139 (p20)
	$130 \leq M_Z < 250$	0.00341	105284 (p17)
		0.0034	90624 (p20)
	$250 \leq M_Z < 1960$	0.000336	193218 (p17)
		0.00034	178546 (p20)
$Z \to \mu^+\mu^-$ + 1lp	$15 \leq M_Z < 75$	0.188	103150 (p17)
		0.199	92018 (p20)
	$75 \leq M_Z < 130$	0.187	98877 (p17)
		0.195	101055 (p20)
	$130 \leq M_Z < 250$	0.00174	48137 (p17)
		0.00184	46222 (p20)
	$250 \leq M_Z < 1960$	0.000163	193985 (p17)
		0.000173	180821 (p20)
$Z \to \mu^+\mu^-$ + 2lp	$15 \leq M_Z < 75$	0.0782	102281 (p17)
		0.0784	89922 (p20)
	$75 \leq M_Z < 130$	0.105	50175 (p17)
		0.099	49600 (p20)
	$130 \leq M_Z < 250$	0.000966	47880 (p17)
		0.000884	45903 (p20)
	$250 \leq M_Z < 1960$	0.000104	198791 (p17)
		0.000106	187820 (p20)

Table 5.20: The cross sections times branching fractions and the initial numbers of events of the $Z \to \mu\mu + b\bar{b}$ ALPGEN Monte Carlo samples.

$Z \to \tau\tau + c\bar{c}$ Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to \tau^+\tau^-$ + 0lp	$15 \leq M_Z < 75$	4.13	201925 (p17)
		4.11	180024 (p20)
	$75 \leq M_Z < 130$	0.907	196058 (p17)
		0.898	260243 (p20)
	$130 \leq M_Z < 250$	0.00746	96318 (p17)
		0.00743	91957 (p20)
	$250 \leq M_Z < 1960$	0.0006	193769 (p17)
		0.000656	178927 (p20)
$Z \to \tau^+\tau^-$ + 1lp	$15 \leq M_Z < 75$	1.02	95266 (p17)
		1.05	181402 (p20)
	$75 \leq M_Z < 130$	0.505	97060 (p17)
		0.488	100802 (p20)
	$130 \leq M_Z < 250$	0.00452	48303 (p17)
		0.00441	48814 (p20)
	$250 \leq M_Z < 1960$	0.000459	194802 (p17)
		0.000466	181722 (p20)
$Z \to \tau^+\tau^-$ + 2lp	$15 \leq M_Z < 75$	0.352	98947 (p17)
		0.382	179172 (p20)
	$75 \leq M_Z < 130$	0.277	48058 (p17)
		0.298	50711 (p20)
	$130 \leq M_Z < 250$	0.00274	47683 (p17)
		0.0025	46629 (p20)
	$250 \leq M_Z < 1960$	0.000279	199836 (p17)
		0.000261	188560 (p20)

Table 5.21: The cross sections times branching fractions and the initial numbers of events of the $Z \to \tau\tau + c\bar{c}$ ALPGEN Monte Carlo samples.

$Z \to \tau\tau + b\bar{b}$ Monte Carlo Sets			
Subprocess	Mass [GeV]	$\sigma \cdot$ BR [pb]	Number of Events
$Z \to \tau^+\tau^-$ + 0lp	$15 \leq M_Z < 75$	0.532	201732 (p17)
		0.51	188371 (p20)
	$75 \leq M_Z < 130$	0.41	202015 (p17)
		0.424	196943 (p20)
	$130 \leq M_Z < 250$	0.00353	104451 (p17)
		0.0034	89812 (p20)
	$250 \leq M_Z < 1960$	0.000319	195386 (p17)
		0.000325	180448 (p20)
$Z \to \tau^+\tau^-$ + 1lp	$15 \leq M_Z < 75$	0.194	96058 (p17)
		0.189	95704 (p20)
	$75 \leq M_Z < 130$	0.191	101078 (p17)
		0.197	103105 (p20)
	$130 \leq M_Z < 250$	0.00172	48481 (p17)
		0.00171	46806 (p20)
	$250 \leq M_Z < 1960$	0.000184	194791 (p17)
		0.000158	182701 (p20)
$Z \to \tau^+\tau^-$ + 2lp	$15 \leq M_Z < 75$	0.0805	104491 (p17)
		0.0801	91011 (p20)
	$75 \leq M_Z < 130$	0.0946	50154 (p17)
		0.104	48476 (p20)
	$130 \leq M_Z < 250$	0.000955	48371 (p17)
		0.00103	45591 (p20)
	$250 \leq M_Z < 1960$	9.82e-05	370388 (p17)
		9.87e-05	182382 (p20)

Table 5.22: The cross sections times branching fractions and the initial numbers of events of the $Z \to \tau\tau + b\bar{b}$ ALPGEN Monte Carlo samples.

Chapter 6

Event Selection for the $Qq \to \ell\nu qq$ Analysis

6.1 Analysis Overview

The minimal distinguishing characteristic of down-type vector-like quark production and decay is a resonance in the lepton-neutrino system ($W \to \ell\nu$) in the presence of at least two jets, one from the associated production of a SM quark and one from the $Q \to Wq$ decay (see Figure 6.1). Preselection cuts designed to choose high-quality events of this kind are described in Section 6.2.

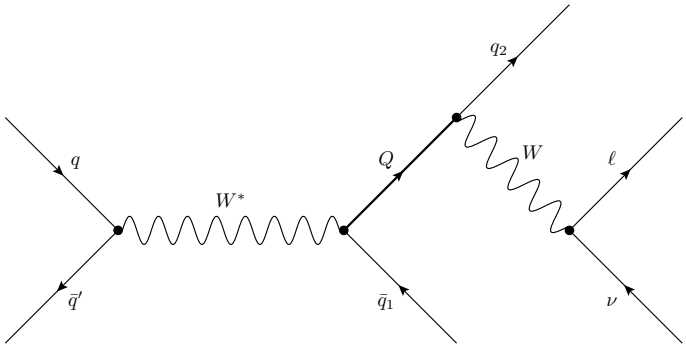

Figure 6.1: Vector-like quark production and decay to Wq.

SM W+jets production with leptonic W boson decay dominates the background, as the final state in these events contains the same objects as the signal. Also contributing to the background are $t\bar{t}$ and single top events with one $t \to (W \to \ell\nu)b$ decay. Z+jets events can pass the preselection if one decay lepton from the Z boson is not reconstructed (because, for example, it went into a poorly-instrumented region of the detector). There are several combinations of decay modes and instrumental factors which can cause SM diboson (WW, WZ and ZZ) events to mimic the final state of the signal. Finally, QCD multijet production can satisfy the preselection criteria if one jet is misidentified as a lepton. All backgrounds except QCD multijet production are modeled using Monte Carlo simulation; applied corrections for well-known mismodeling effects in the MC are detailed in Section 6.3. QCD multijet production is modeled using data as described in Section 6.4.

The preselection cuts described in this chapter are designed to produce a sample that is signal-poor and suitable for verifying the modeling of the background in the relevant kinematic region. Additional cuts to further isolate vector-like quark candidate events from SM background are detailed in Section 8.1.1, and the resulting distributions are shown in Section 8.1.3.

6.2 Preselection Cuts

Preselection cuts are applied in this analysis to select high quality $W \to e\nu$ and $W \to \mu\nu$ events both in the data and Monte Carlo.

- All events must satisfy the data quality criteria described in Section 5.1.
- There must be at least one hard scatter (primary) vertex in the event with at least three associated tracks. This vertex must lie within 60 cm of the detector center along the beam axis.
- Events must have at least two jets, each having $p_T > 20$ GeV and $|\eta^{\text{det}}| < 2.5$

where η^{det} is pseudorapidity defined using detector coordinates. For events recorded during Run IIb, the jets are required to be vertex-confirmed as described in Section 4.4.

- $\not\!\!E_T + 2M_T^W > 80$ GeV. A diagonal cut is applied in the $\not\!\!E_T - M_T^W$ plane, as QCD multijet background tends to have low missing E_T and low reconstructed W transverse mass, defined as $M_T^W \equiv \sqrt{2\not\!\!E_T p_{T_{lep}}(1 - cos[\Delta\phi(\not\!\!E_T, lep])}$. This is shown in Figure 6.2.

- In the e+jets channel, one tight electron (see Section 4.2) is selected with $p_T > 20$ GeV in the central calorimeter ($|\eta^{\text{det}}| < 1.1$). A veto is placed on events containing any additional electrons with $p_T > 15$ GeV.

- In the μ+jets channel, one "mediumnseg3" quality muon with tight isolation (see Section 4.3) is selected with $p_T > 20$ GeV and $|\eta^{\text{det}}| < 2$. A veto is placed on events with any additional mediumnseg3 muon with $p_T > 15$ GeV.

- Missing $E_T > 20$ GeV. The missing E_T ($\not\!\!E_T$) takes into account the charged lepton in the event.

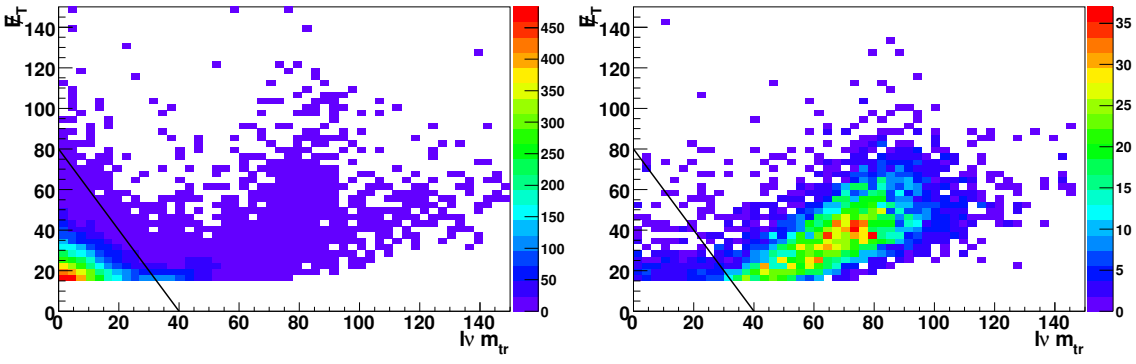

Figure 6.2: Missing E_T vs W transverse mass for QCD multijet events (left) and W boson MC events (right). The black line shows the applied cut.

6.3 Monte Carlo Corrections

Monte Carlo events which pass the event selection are reweighted for several known mismodeling effects. Individual events are assigned an overall multiplicative weight based on the reweightings described below, and this weight is carried through the analysis.

6.3.1 Luminosity and Primary Vertex Reweighting

All Monte Carlo events are overlaid with randomly triggered data events in order to simulate the effects of multiple $p\bar{p}$ interaction in addition to the hard scatter. However, the instantaneous luminosity profile of these events does not exactly match that of the data recorded by the physics triggers. In addition, the MC is generated with primary vertices which are more central along the z axis than those recorded in data. We re-weight the MC to the luminosity and primary vertex z-position distributions measured in data [57; 58].

6.3.2 W/Zp_T Reweighting

ALPGEN, which is used to generate most of the simulated backgrounds, is known to poorly model the transverse momentum distribution of the vector boson in V+jets events at low p_T (where V denotes a W or Z boson) due to the dominance of non-perturbative effects in this region. As the Z boson p_T profile is well-measured in data, the Monte Carlo is reweighted to better match the data in this distribution [59]. The W boson p_T distribution in W+jets events is treated in a similar manner [60]. Z boson p_T reweighting curves for 3 different dilepton mass bins are shown in Figure 6.3. The reweighting function is dependent on the number of jets in the event, and the relevant curves for this analysis (the 2-jet inclusive bin) are shown in blue.

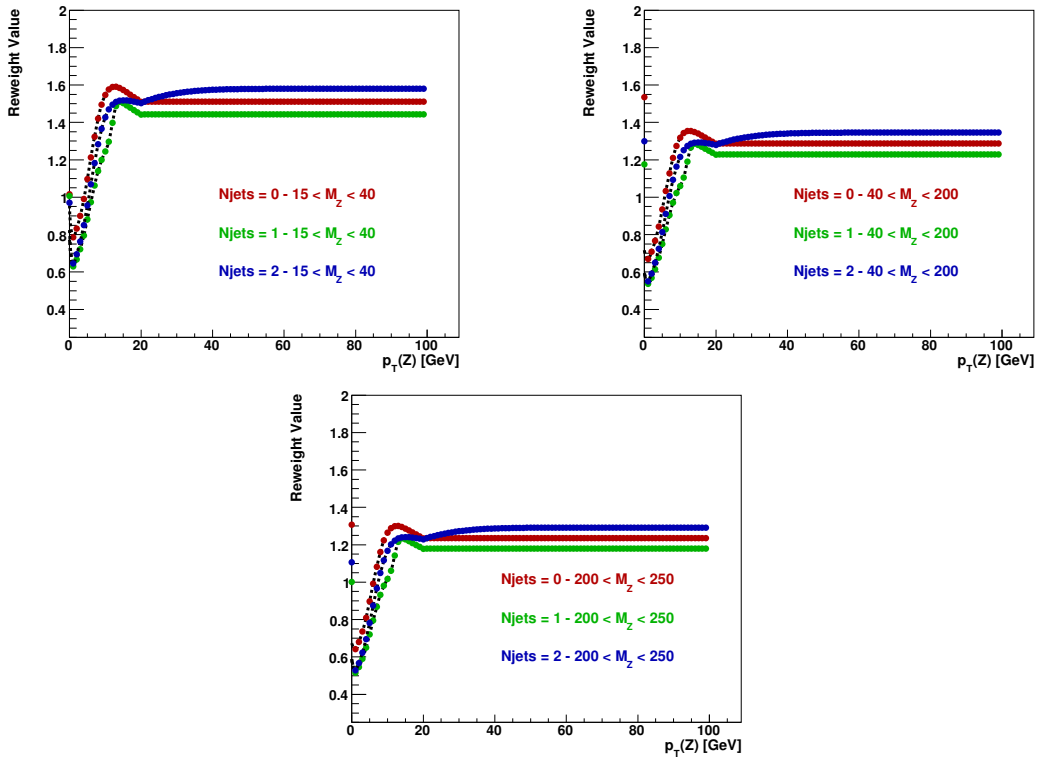

Figure 6.3: The reweighting function value as a function of the dilepton system p_T for events with no jets (red), one jet (green), and two or more jets (blue) in three dilepton mass bins ($15 < M_{\ell\ell} < 40$ GeV top left, $40 < M_{\ell\ell} < 200$ GeV top right, $200 < M_{\ell\ell} < 250$ GeV bottom left).

6.3.3 Jet η and ΔR Reweighting

The fixed-order calculation implemented by ALPGEN and its matching to the parton shower also produces known mismodeling effects in the angular distributions of jets in V+jets events, and the Monte Carlo is reweighted to better match the distributions measured in data and with other generators [61]. Specifically, the leading and sub-leading jet η distributions are rescaled by the functions shown in Figure 6.4, while the $\Delta R(j_1, j_2)$ (the ΔR between the two jets with the highest p_T in the event) is reweighted by the curve shown in Figure 6.5. A 15% systematic uncertainty is assigned to the V+jets samples to cover the shape and normalization discrepancies between data and Monte Carlo. This uncertainty includes effects from vector boson p_T modeling in addition to the jet modeling discussed in this section.

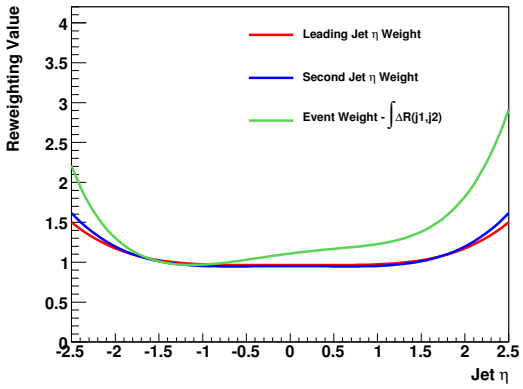

Figure 6.4: The reweighting curve applied to W and Z+ jets events as a function of the jet η value. The green curve shows the full weight integrated over the spatial separation between the two jets in the event. The red and blue curves show the contributions from the leading and sub-leading jets, respectively.

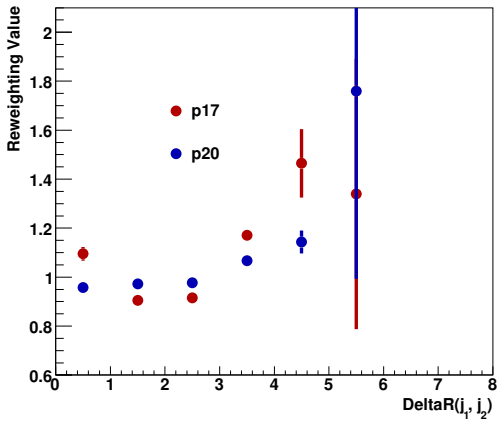

Figure 6.5: The reweighting function value as a function of $\Delta R(j_1, j_2)$ for p17 (red) and p20 (blue) reco versions.

6.4 QCD Multijet Modeling and Normalization

Multijet production is a background of particular concern for the single-lepton channels. Jets can fake electrons via production of narrow, EM-like showers, while non-isolated muons from jet punch-through or heavy flavor decays can mimic isolated muons from W boson decays.

The contribution of multijet events is estimated using a Matrix Method [62] approach in the electron channel (see Section 6.4.1). An anti-isolation method [63] is used for the muon channel (see Section 6.4.2).

In the following we use "QCD" as a shorthand for multijets.

6.4.1 Multijet background in the electron channel

We define three separate samples in data: *Loose*, where the selected electron satisfies the loose criteria, *tight*, where the selected electron satisfies tight criteria, and orthogonal, *orth*, where the selected electron satisfies loose but not tight (see Section 4.2). Each of these samples comprises *signal* events (with electrons from real W boson decays, estimated from all simulated MC events) and QCD events:

$$N_{loose} = N_{sig} + N_{QCD} \tag{6.1}$$

$$N_{tight} = \varepsilon_{sig} \times N_{sig} + \varepsilon_{QCD} \times N_{QCD} \tag{6.2}$$

$$N_{orth} = N_{loose} - N_{tight}, \tag{6.3}$$

where ε_{sig} (ε_{QCD}) is the efficiency for *signal* (QCD) events which have passed the loose selection to also pass the tight selection. These equations can be used to express the composition of the tight preselection data sample in terms of N_{sig} and N_{orth}:

$$N_{tight} = \frac{\varepsilon_{QCD}}{1-\varepsilon_{QCD}} \cdot N_{orth} + [1 - (\frac{\varepsilon_{QCD}}{1-\varepsilon_{QCD}}) \cdot (\frac{1-\varepsilon_{sig}}{\varepsilon_{sig}})] \cdot N^{sig}_{tight}, \tag{6.4}$$

where $N^{sig}_{tight} = \varepsilon_{sig} \times N_{sig}$. The coefficient for N_{orth} is applied to all orthogonal events, and determines the shape and normalization of the QCD multijet contribution to the tight preselection sample. The coefficient for N^{sig}_{tight} is applied to all MC passing the tight selection, to correct for leakage of real W bosons into the orthogonal data sample.

Derivation of ε_{sig}

ε_{sig} is determined from tag-and-probe studies on $Z \to ee$ events [30]. The efficiency is modeled one-dimensionally in bins of electron p_T, and the ratio between the efficiency for *tight* and *loose* electrons is fitted with the following function:

$$\varepsilon_{sig} = \begin{cases} a \cdot p_T^2 + b \cdot p_T + c & \text{if } p_T < 45 \text{ GeV} \\ 0.87 & \text{if } p_T \geq 45 \text{ GeV}, \end{cases} \tag{6.5}$$

where $a = -3.22 \cdot 10^{-5}$, $b = 4.04 \cdot 10^{-3}$ and $c = 0.76$, and shown in Figure 6.6.

Derivation of ε_{QCD}

The fake rate is measured in data events with $\not\!\!E_T < 10$ GeV. ε_{QCD} is derived from the ratio of tight to loose samples, where contributions from the *signals*, i.e. other

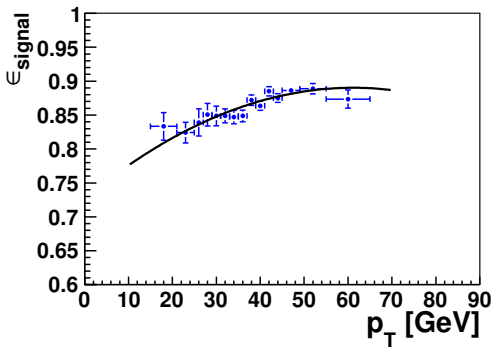

Figure 6.6: ε_{sig} as a function of electron p_T. Dots with error bars represent the measured relative efficiency for tight and loose electrons, and the black line is the fit function given by Eq. 6.5.

Standard Model backgrounds, are subtracted:

$$\varepsilon_{QCD}(p_T) = \frac{Data_{tight} - (Wjets + Zjets + top + Diboson)_{tight}}{Data_{loose} - (Wjets + Zjets + top + Diboson)_{loose}}. \quad (6.6)$$

We make the assumption that ε_{QCD} depends on the angle between the lepton and $\not\!\!E_T$ in transverse plane $\Delta\phi(EM,\not\!\!E_T)$ and the electron's $\eta_{\rm det}$ (the latter is used for Run IIa data only). The dependence of ε_{QCD} on the electron p_T is given by:

$$\varepsilon_{QCD} = e^{-par[0]\cdot p_T + par[1]} + par[2] + par[3] * p_T. \quad (6.7)$$

This is shown in Figure 6.7 for Run IIb.

6.4.2 Multijet background in the muon channel

Section 4.3 describes the calorimeter and tracker isolation requirements used to pre-select muons coming from real W bosons. To estimate the shape of the multijet background, we reverse our isolation criteria in data passing all other pre-selection requirements. To correct the anti-isolated data for real W boson events failing the isolation criteria, anti-isolated MC samples are subtracted in three distributions: $\not\!\!E_T$,

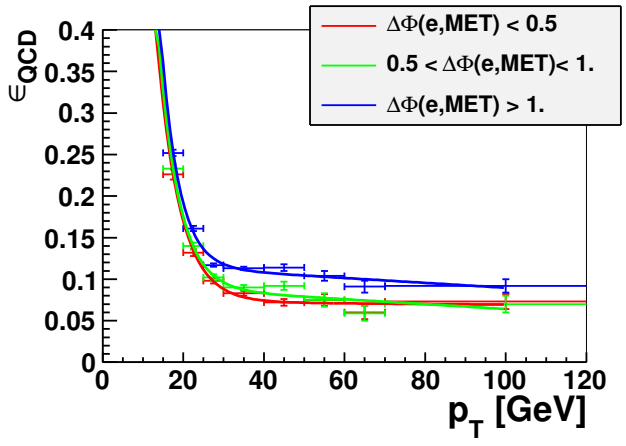

Figure 6.7: ε_{QCD} for the electron channel (p20).

p_T^{μ}, and M_T^W. Bin-by-bin weights are then determined in each of the three distributions:

$$qcd\_weight \times N_{anti-iso}^{DATA} = N_{anti-iso}^{DATA} - N_{anti-iso}^{MC}. \tag{6.8}$$

Reweighting the anti-isolated data events using the average of the $\not\!\!E_T$, p_T^{μ}, and M_T^W weights gives us shapes close to the original $N_{anti-iso}^{DATA} - N_{anti-iso}^{MC}$ shown in Figure 6.8.

To estimate the number of background QCD events, a MINUIT fit to the M_T^W distribution is performed using isolated MC, letting the V+Jets scale factor (s_{factor}) and QCD normalization float to get a best fit to isolated DATA. In order to remove a potential bias caused by the s_{factor}, the process of finding the $\not\!\!E_T$, p_T^{μ}, and M_T^W weights is repeated after applying the s_{factor} (bootstrapping), and a second MINUIT fit is performed. To remove any bias introduced by the choice of binning in the M_T^W distribution, 30 distributions with random binning are generated, and the fit is performed separately on each of them. The QCD normalization for all data-MC comparisons is taken from the average of the 30 fits, as is the final applied s_{factor}, which is 1.24 for p17 and 1.02 for p20.

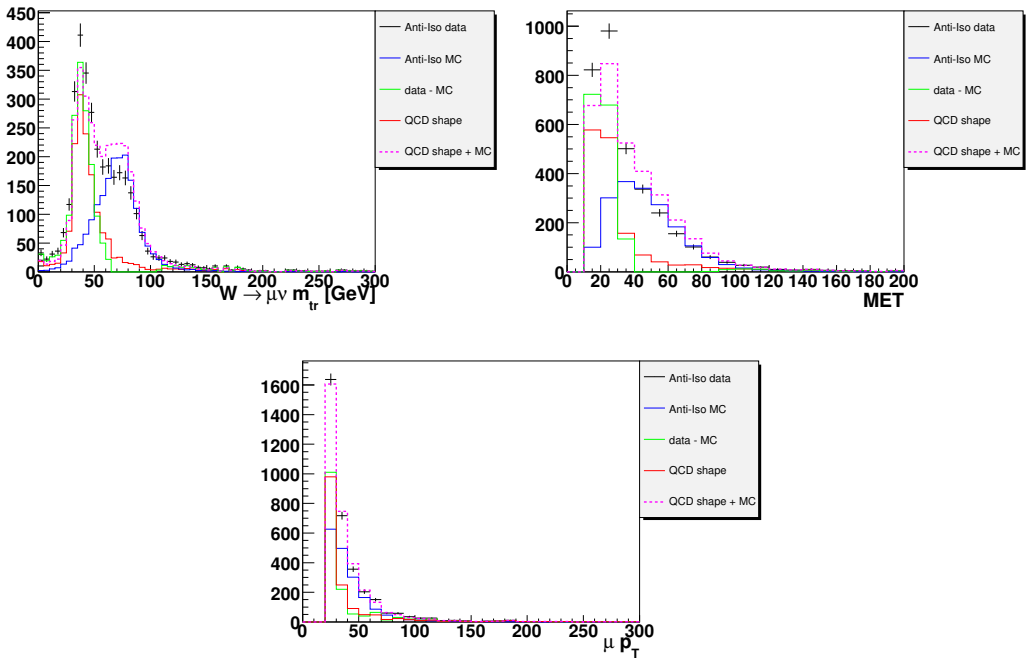

Figure 6.8: W transverse mass (top left), $\not\!\!E_T$(top right) and muon p_T (bottom) for the p20 muon channel. The black histogram is anti-isolated data, blue is anti-isolated MC, green is anti-iso (data - MC), red is reweighted anti-iso data (QCD), and purple is anti-isolated MC + QCD.

6.5 Comparison of the Data with the Background Expectation

Once all Monte Carlo corrections have been applied, and the QCD modeling and normalization has been performed, we can see how well the background expectation agrees with the observed data at preselection level. Expected and observed $\mu\nu jj$ and $e\nu jj$ event counts, along with theoretical signal yields, are shown for Run IIa and Run IIb in Table 6.1, and for total Run II combined in Table 6.2.

Figures 6.9-6.18 compare the combined p17 and p20 data in e+jets and μ+jets events with the background expectations.

Expected and Observed Yields for p17 and p20 e+jets and μ+jets events

Source	p17 e+jets	p17 μ+jets	p20 e+jets	p20 μ+jets
QCD Multijet	4002±24	553±16	8711±38	1366±27
Z+jets	1066±7	1292±6	2133±24	3126±23
W+jets	15844±47	11288±37	37455±111	28497±62
top	607±1	735±2	2049±4	1073±2
Diboson	374±1	422±2	1091±9	847±7
Background Sum	21894±53	14291±40	51440±120	34909±71
Data	21463	14252	51097	35285
$Qq \to \ell\nu qq$(M=280 GeV)	493±7	387±6	1584±30	1395±25
$Qq \to \ell\nu qq$(M=320 GeV)	273±4	210±3	887±16	760±14
$Qq \to \ell\nu qq$(M=360 GeV)	156±2	123±2	511±10	444±8
$Qq \to \ell\nu qq$(M=400 GeV)	91.9±2	66.2±1.3	293±7	273±6
$Qq \to \ell\nu qq$(M=450 GeV)	46.3±0.7	35.8±0.5	153±4	137±3
$Qq \to \ell\nu qq$(M=500 GeV)	24.8±0.4	18.5±0.3	79.0±1.9	73.1±1.7
$Qq \to \ell\nu qq$(M=550 GeV)	13.0±0.2	9.4±0.1	41.8±1.1	37.3±0.9
$Qq \to \ell\nu qq$(M=600 GeV)	7.1±0.1	5.2±0.1	22.5±0.6	20.2±0.5
$Qq \to \ell\nu qq$(M=650 GeV)	3.6±0.1	2.7±0.04	11.6±0.3	10.3±0.2
$Qq \to \ell\nu qq$(M=700 GeV)	2.0±0.03	1.4±0.03	6.6±0.1	5.3±0.1

Table 6.1: Estimated background yields and number of observed data events after the $\ell\nu jj$ preselection. Estimated signal yields for down-type vector-like quark production and decay using the same selection are also shown. The quoted errors are due to limited statistics.

Combined Run II Expected and Observed Yields for ℓ+jets events.

Source	$e\nu jj$ events	$\mu\nu jj$ events	$\ell\nu jj$ events
QCD Multijet	12713±45	1919±31	14632±55
Z+jets	3200±25	4417±23	7617±34
W+jets	53298±121	39785±72	93083±140
top	2657±4	1808±3	4465±5
Diboson	1465±9	1270±7	2735±12
Background Sum	73333±131	49199±82	122533±155
Data	72560	49537	122097
$Qq \to \ell\nu qq$(M=280 GeV)	2076±31	1782±26	3859±40
$Qq \to \ell\nu qq$(M=320 GeV)	1160±17	970±14	2131±22
$Qq \to \ell\nu qq$(M=360 GeV)	667±10	566±8	1233±13
$Qq \to \ell\nu qq$(M=400 GeV)	385±8	339±7	724±10
$Qq \to \ell\nu qq$(M=450 GeV)	199±4	173±3	373±5
$Qq \to \ell\nu qq$(M=500 GeV)	104±2	91.5±1.7	195±3
$Qq \to \ell\nu qq$(M=550 GeV)	54.8±1.1	46.8±0.9	102±1.4
$Qq \to \ell\nu qq$(M=600 GeV)	29.7±0.6	25.4±0.5	55.0±0.8
$Qq \to \ell\nu qq$(M=650 GeV)	15.3±0.3	12.9±0.2	28.2±0.4
$Qq \to \ell\nu qq$(M=700 GeV)	8.6±0.1	6.7±0.1	15.3±0.2

Table 6.2: Estimated background yields and number of observed data events after the $\ell\nu jj$ preselection. Estimated signal yields for down-type vector-like quark production and decay using the same selection are also shown. The quoted errors are due to limited statistics.

Preselection Plots For the e+jets Final State (1)

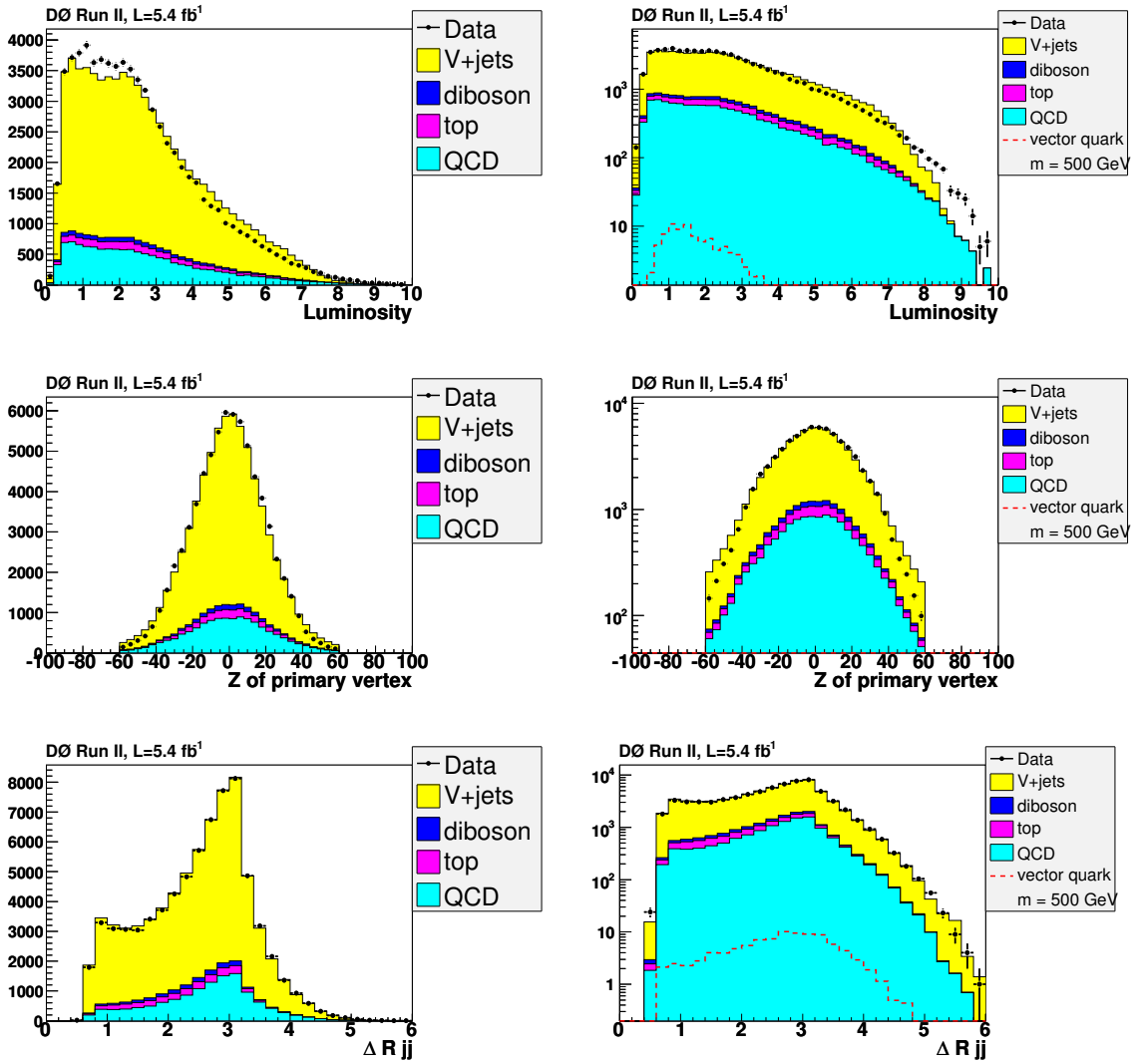

Figure 6.9: Comparison of the instantaneous luminosity (top row), z position of the primary vertex (middle row) and $\Delta R(j_1, j_2)$ (bottom row). V+jets refers to combined W+jets and Z+jets. Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots For the e+jets Final State (2)

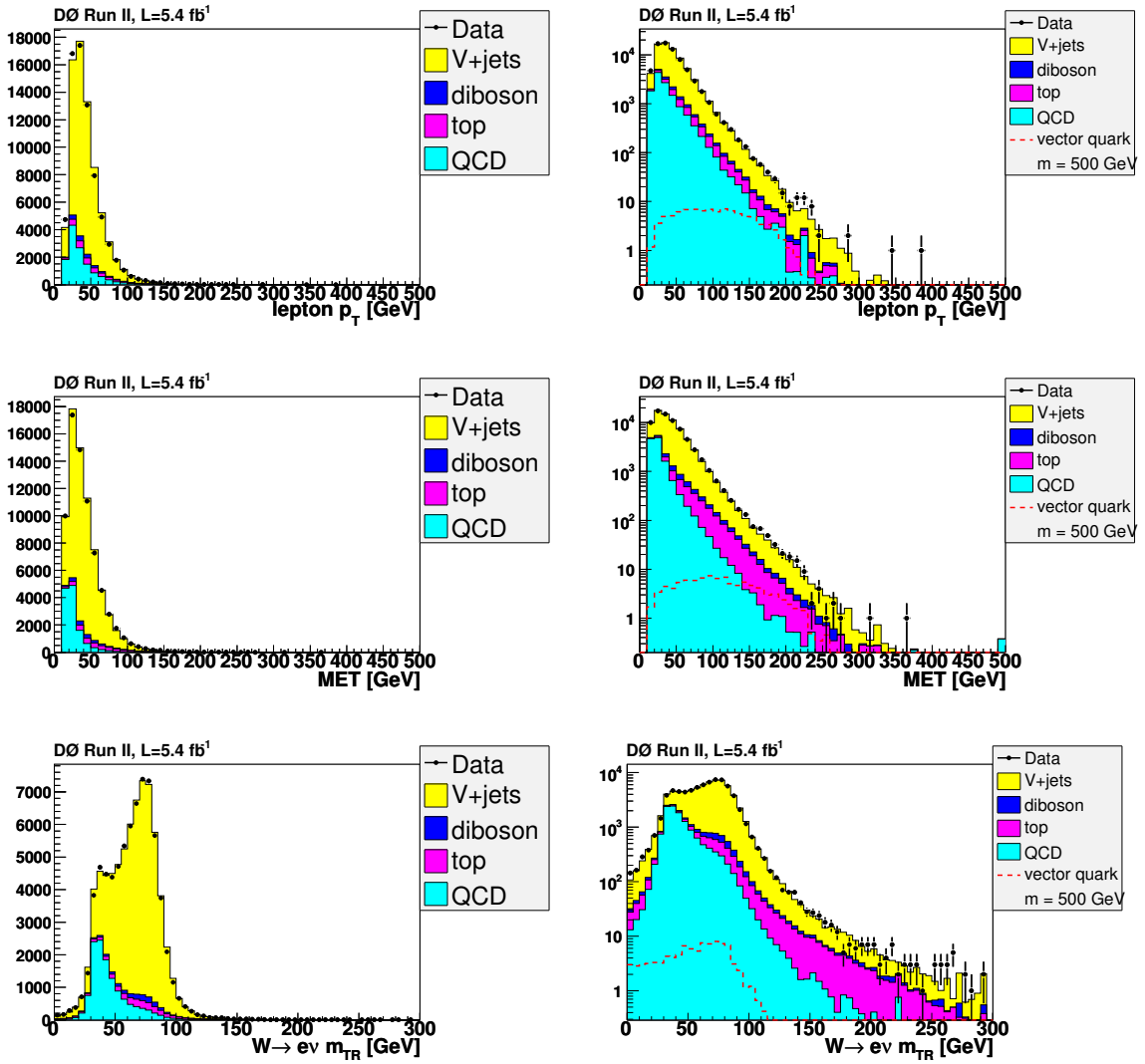

Figure 6.10: Comparison of the electron p_T (top row), missing E_T (middle row) and M_T^W (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots For the e+jets Final State (3)

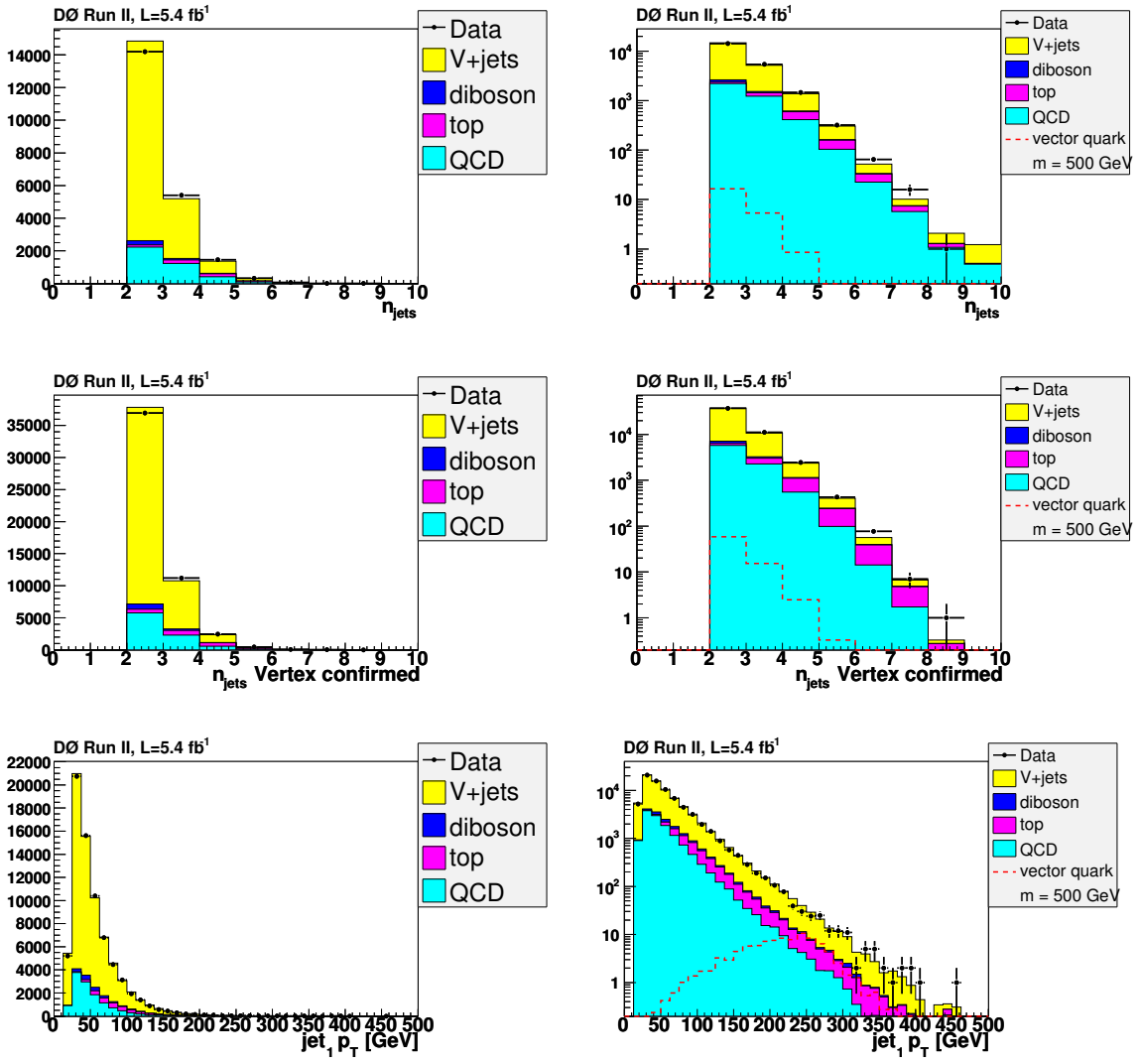

Figure 6.11: Comparison of the Run IIa jet multiplicity (top row), Run IIb vertex-confirmed jet multiplicity (middle row) and leading jet p_T (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots For the e+jets Final State (4)

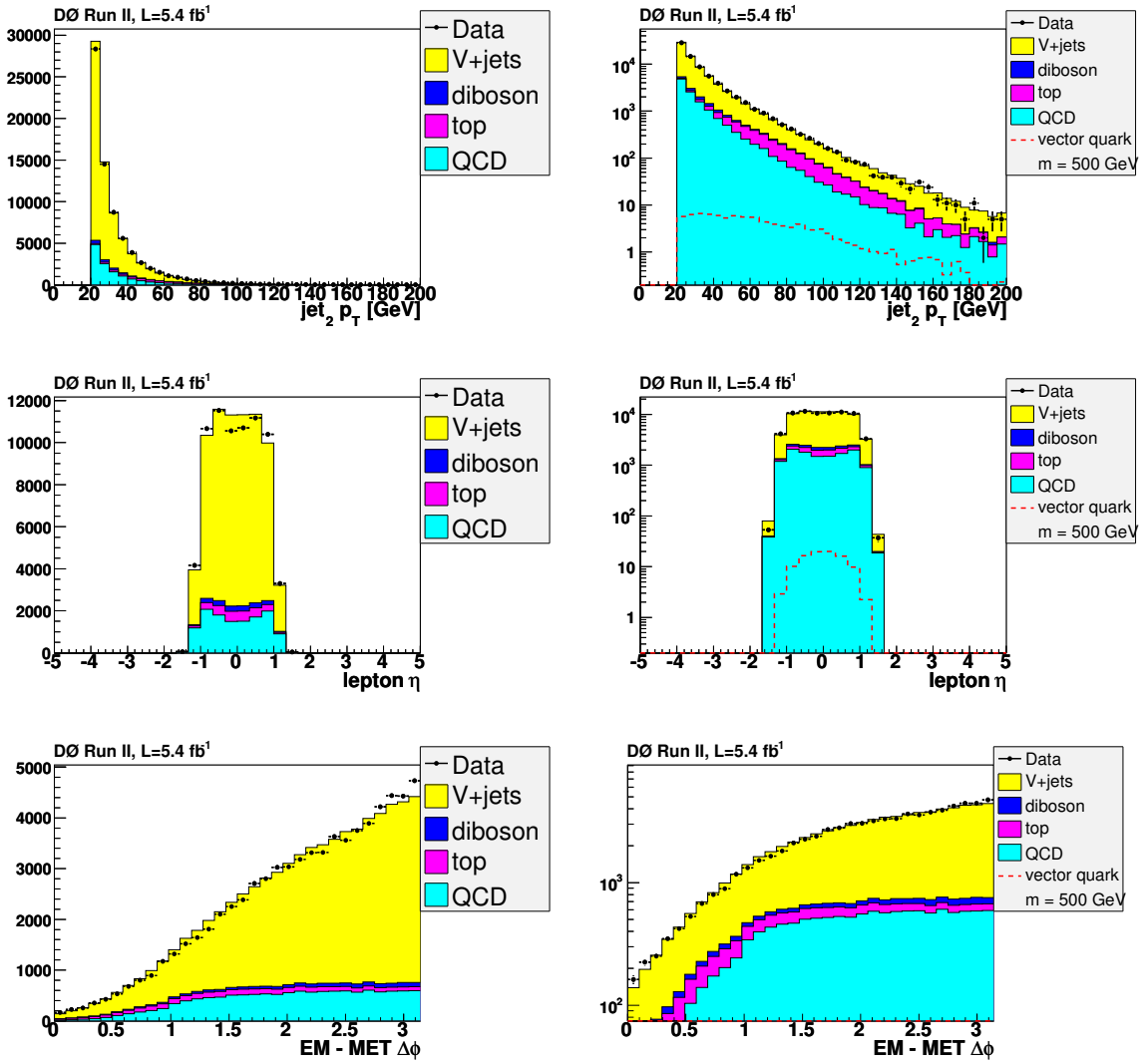

Figure 6.12: Comparison of the sub-leading jet p_T (top row), electron η (middle row) and $\Delta\phi(e, \not\!\!E_T)$ (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots For the e+jets Final State (5)

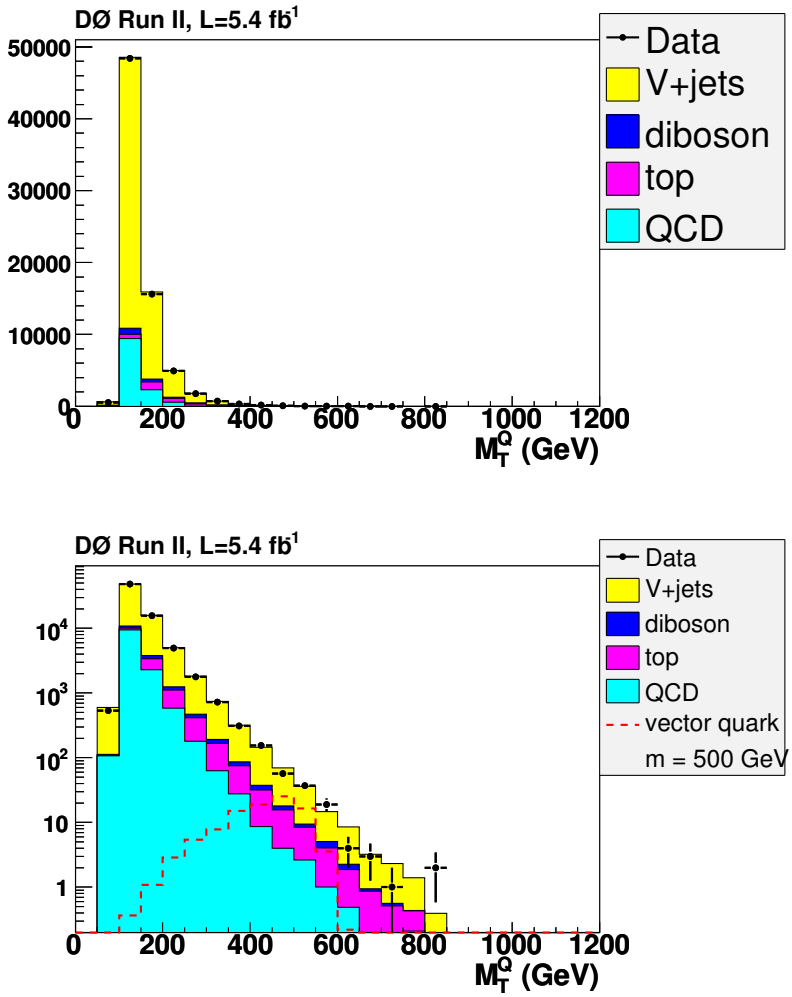

Figure 6.13: Comparison of the vector-like quark transverse mass, $(M_T^Q)^2 = (\sqrt{p_{T_W}^2 + M_W^2} + p_{T_j})^2 - (\vec{p}_{T_W} + \vec{p}_{T_j})^2$. Shown with linear scale at top, and log scale at bottom.

Preselection Plots For the μ+jets Final State (1)

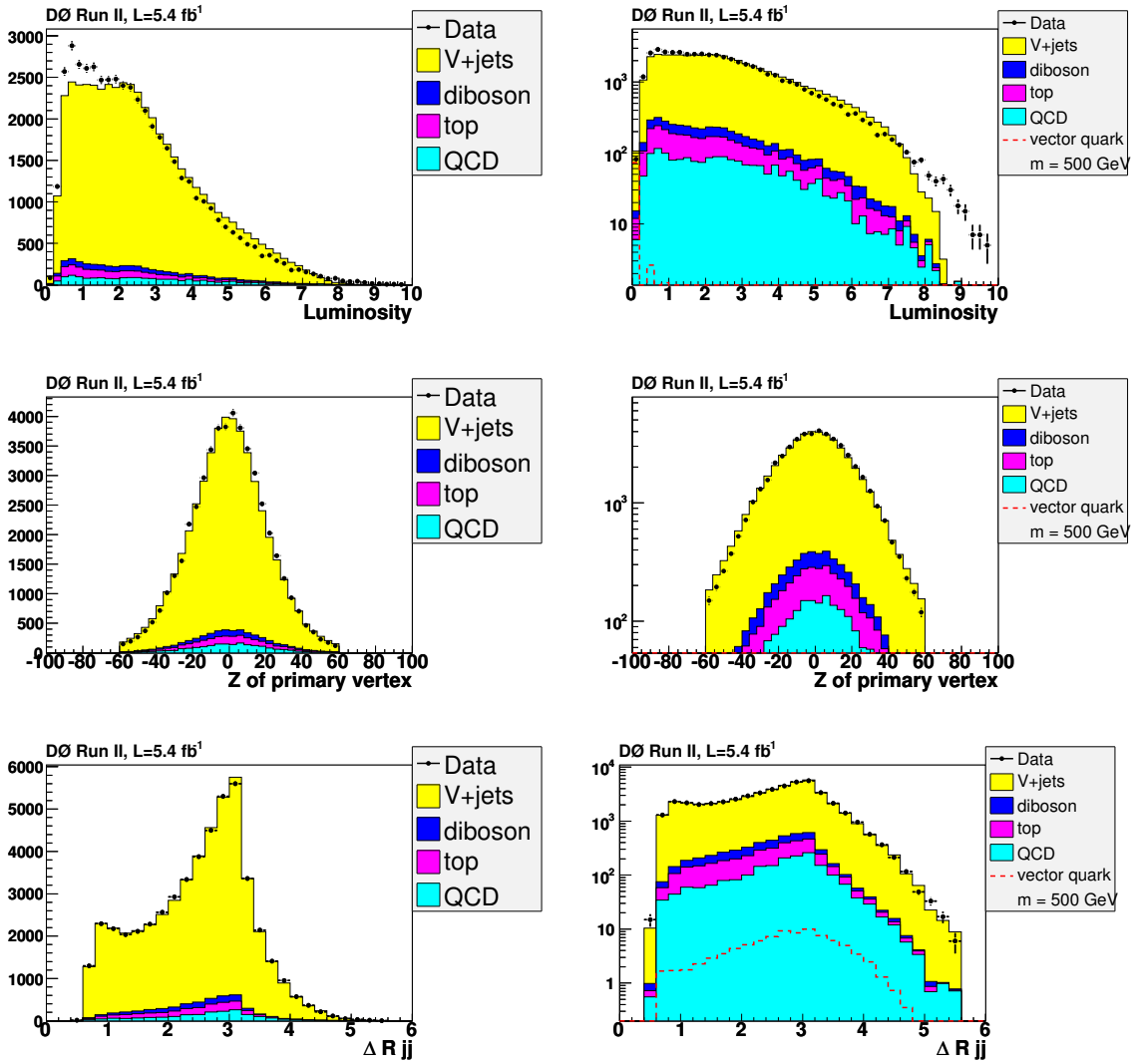

Figure 6.14: Comparison of the instantaneous luminosity (top row), z position of the primary vertex (middle row) and $\Delta R(j_1, j_2)$ (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots For the μ+jets Final State (2)

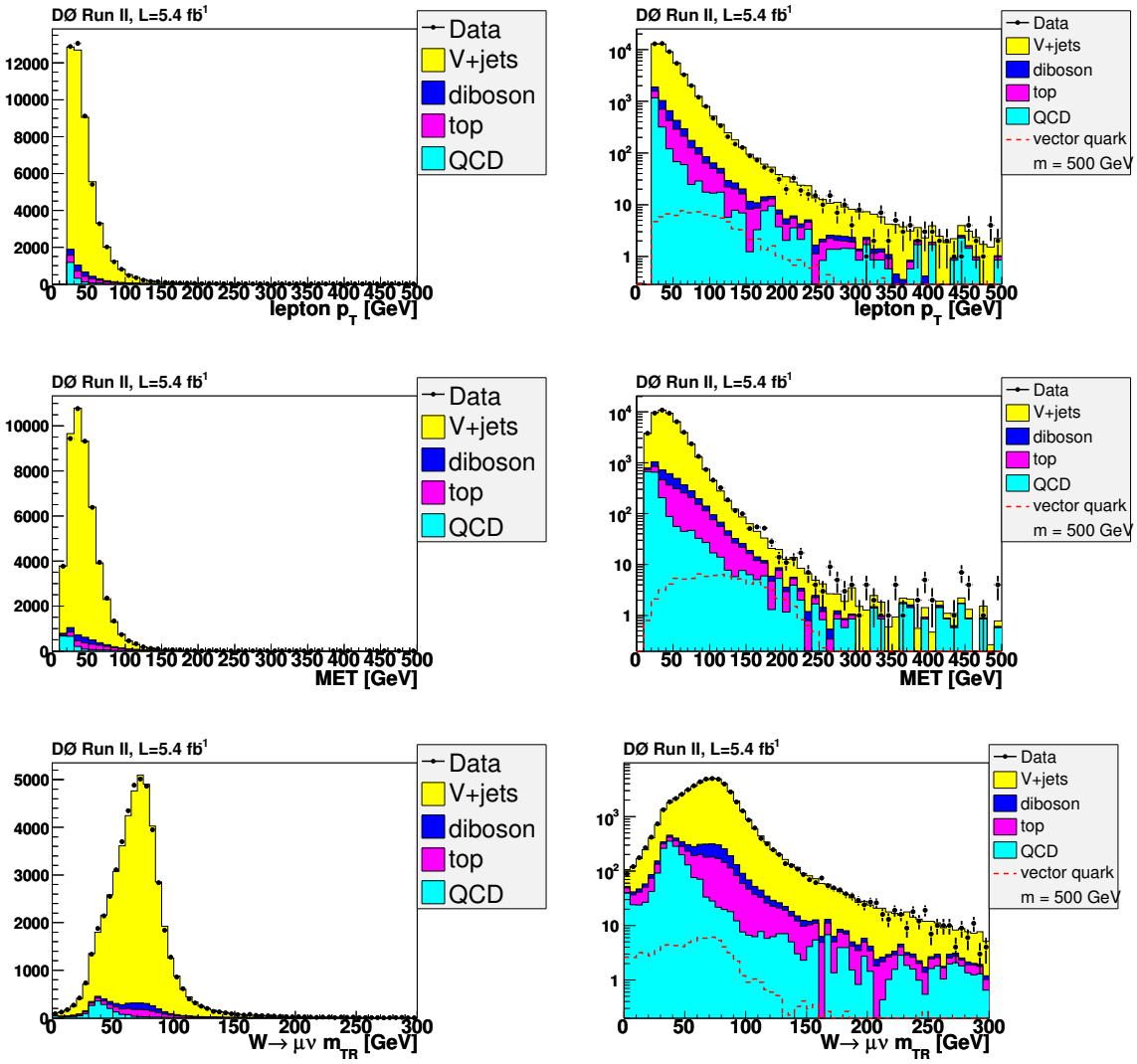

Figure 6.15: Comparison of the muon p_T (top row), missing E_T (middle row) and M_T^W (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots For the μ+jets Final State (3)

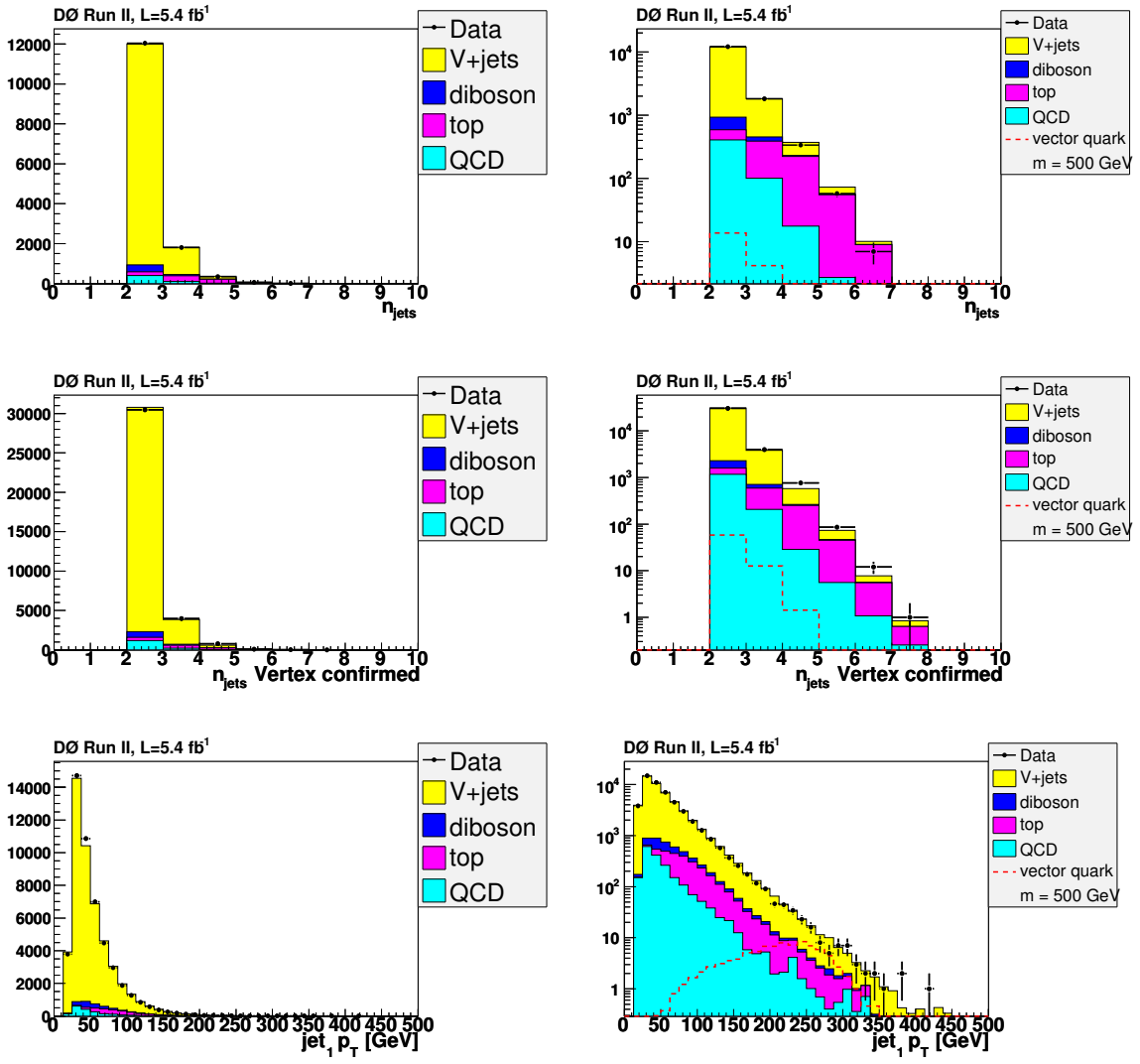

Figure 6.16: Comparison of the Run IIa jet multiplicity (top row), Run IIb vertex-confirmed jet multiplicity (middle row) and leading jet p_T (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots For the μ+jets Final State (4)

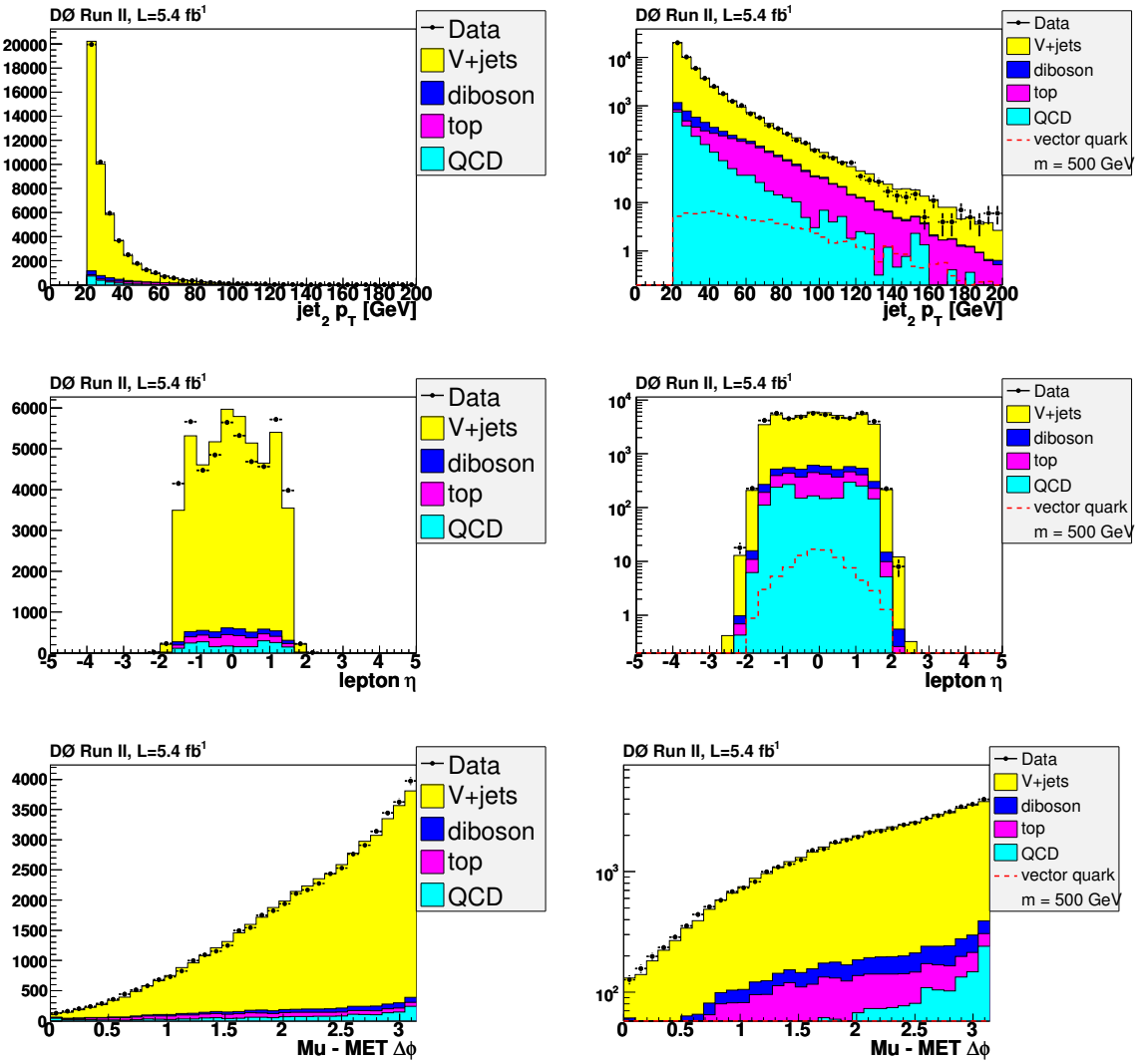

Figure 6.17: Comparison of the sub-leading jet p_T (top row), muon η (middle row) and $\Delta\phi(\mu, \not\!\!E_T)$ (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots For the μ+jets Final State (5)

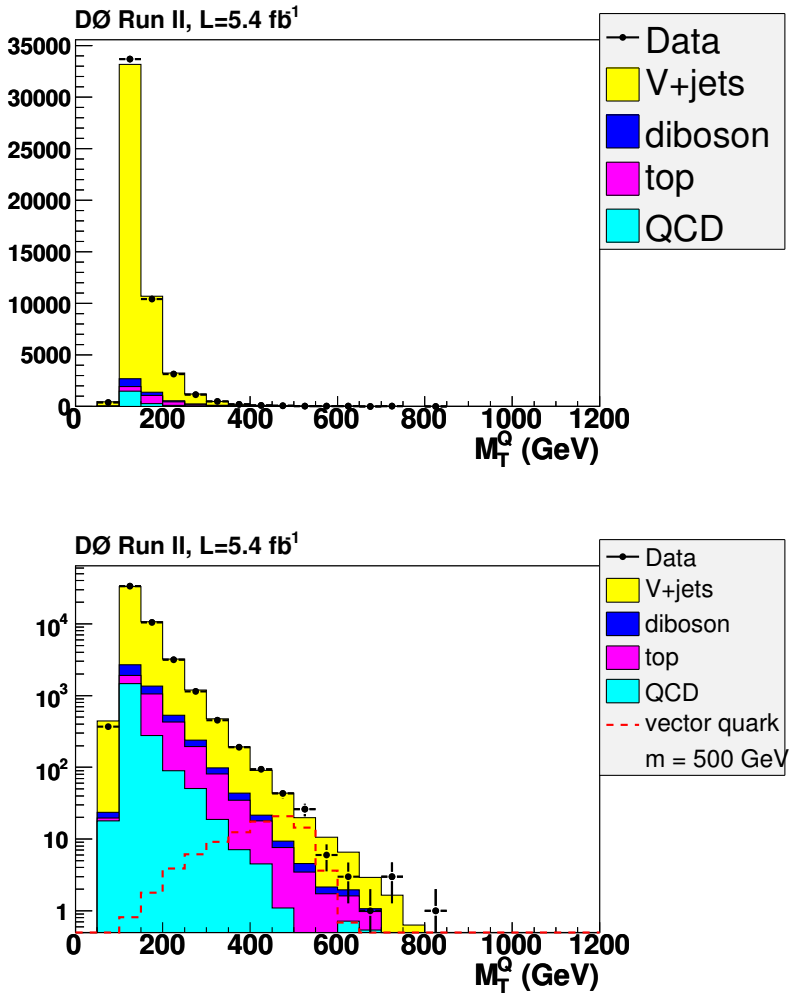

Figure 6.18: Comparison of the vector-like quark transverse mass, $(M_T^Q)^2 = (\sqrt{p_{T_W}^2 + M_W^2} + p_{T_j})^2 - (\vec{p}_{T_W} + \vec{p}_{T_j})^2$. Shown with linear scale at top, and log scale at bottom.

Chapter 7

Event Selection for the $Qq \to \ell\ell qq$ Analysis

7.1 Analysis Overview

The minimal distinguishing characteristic of up-type vector-like quark production and decay is a resonance in the dilepton system ($Z \to \ell\ell$) in the presence of at least two jets, one from the associated production of a SM quark and one from the $Q \to Zq$ decay (see Figure 7.1. Preselection cuts designed to choose high-quality events of this kind are described in Section 7.2.

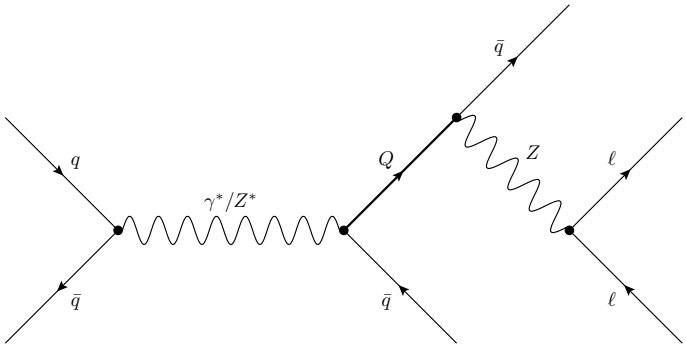

Figure 7.1: Vector-like quark production and decay to Zq.

As no explicit trigger selection is imposed for the dilepton channels, we must normalize the background prediction to account for the implicit trigger efficiency in the data. This procedure is described in Section 7.3. Other Monte Carlo corrections, similiar to those used in the ℓ+jets analyses, are described in Section 7.4. Finally, Section 7.5 describes the methods used to model and normalize the QCD multijet background in the dilepton analyses.

The preselection cuts listed below produce a sample that is very low in signal and is used for verifying the background modeling techniques. Additional cuts to further isolate up-type vector-like quark candidate events from SM background are detailed in Section 8.2.1, and the resulting distributions are shown in Section 8.2.2.

7.2 Preselection Cuts

Preselection cuts are applied in this analysis to select high quality $Z \to ee$ and $Z \to \mu\mu$ events both in the data and Monte Carlo.

- All events must satisfy the data quality criteria described in Section 5.1.
- There must be at least one hard scatter (primary) vertex in the event with at least three associated tracks. This vertex must lie within 60 cm of the detector center along the beam axis.
- Events must have at least two jets with $p_T > 20$ GeV. No vertex confirmation requirement is imposed.
- In the ee+jets channel, two top_loose electrons are selected with $p_T > 20$ GeV in either the CC ($|\eta^{\text{det}}| < 1.1$) or EC ($1.5 < |\eta^{\text{det}}| < 2.5$) region. A veto is placed on any additional electrons with $p_T > 15$ GeV.
- In the $\mu\mu$+jets channel, two muons are selected with $p_T > 20$ GeV and $|\eta^{\text{det}}| < 2$. The product of the sum of the muons' track and calorimeter scaled isolations is

required to be less than 0.03 (see Section 4.3).

- $\not\!\!E_T < 50$ GeV. As our final state contains no neutrinos, $\not\!\!E_T$ in the dilepton analyses is assumed to arise from poorly-measured jets. The $\not\!\!E_T$ is corrected for the two leptons in the event.

- The invariant mass of the two leptons is required to be consistent with a Z boson: 70 GeV $< M_{\ell\ell} <$ 110 GeV.

7.3 Corrections for Trigger Effects

The absence of an explicit trigger selection can manifest itself in two ways in our background prediction. One is in the lepton p_T distribution at low p_T, due to the turn-on effects of the implicit trigger requirement. However, this effect is negligible in this analysis, as our preselection cut of lepton $p_T > 20$ GeV removes the affected region.

The other effect is seen in the overall background normalization, due to the efficiency of the implicit trigger requirement. This efficiency can be measured in data, given an accurate prediction for the expected number of Z boson events [70]. The same efficiency is applied to all other Monte Carlo background and signal samples, since each has a similar or harder lepton p_T spectum. It is determined through the relation

$$f_{\mathcal{L}} = \frac{N_{Data}}{\mathcal{L} \times \sigma_Z^{NNLO}}, \tag{7.1}$$

where N_{Data} is the number of events seen in data, $\mathcal{L}$ is the integrated luminosity, and σ_Z^{NNLO} is the next-to-next-to-leading order calculation of the inclusive Z boson cross section. To determine the normalization in a region free from both non-Z boson events and any possible signal contamination, the following sample was used:

- Missing $E_T < 50$ GeV to remove $t\bar{t}$ and diboson backgrounds.

- $70 < M(Z \to \ell\ell) < 110$ GeV.
- $P_T(Z \to \ell\ell) < 100$ GeV.
- $\Delta R(\ell^+, \ell^-) > 2.0$.
- Veto on any additional leptons.

The derived global normalizations applied to p17 and p20 Monte Carlo backgrounds are shown in Table 7.1. More details on the normalization process can be found in Ref. [64]. The results are in broad agreement with previous measurements at DØ [65; 66; 67].

Z Normalization Factors

Reco Version	Final State	Normalization Factor
p17	$Z \to e^+e^-$	$0.99 \pm 4\%$
p17	$Z \to \mu^+\mu^-$	$0.91 \pm 4\%$
p20	$Z \to e^+e^-$	$0.87 \pm 4\%$
p20	$Z \to \mu^+\mu^-$	$0.73 \pm 4\%$

Table 7.1: Global normalization factors ($f_{\mathcal{L}}$) for p17 and p20 for both $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ samples. The 4% uncertainty arises from the theoretical uncertainty on the Z boson production cross section.

7.4 Monte Carlo Reweighting

Background Monte Carlo samples in the dilepton+jets analysis suffer from many of the same mismodeling issues as the ℓ+jets analysis. We therefore apply instantaneous luminosity and primary vertex reweighting, vector boson p_T reweighting, and jet η and ΔR reweighting, as detailed in Sections 6.3.1, 6.3.2, and 6.3.3.

Additionally, the Z+jets Monte Carlo is reweighted to address ALPGEN's imperfect modeling of jet multiplicity in the final state. Similar to the ΔR reweighting, Z+2 or more jet events are made to agree with the number of $Z \to \ell\ell$ data events with two or more jets [68]. The reweighting factor is determined in a region free from any possible signal contamination ($\Delta R(\ell_1, \ell_2) > 2.5$). Table 7.2 shows the average scaling factor applied to the Z+jets MC in each sub-sample.

As stated in Section 6.3.3, a 15% systematic uncertainty is associated with these reweightings which represents the maximum amount by which the $Z+ \geq 2$ jets Monte Carlo must be scaled.

Z+jets Scaling Factors for $Z+ \geq 1$ Jet Events

Reco Version	Final State	Dijet Scaling Factor
p17	$Z \to e^+e^-$	$1.02 \pm 15\%$
p17	$Z \to \mu^+\mu^-$	$1.02 \pm 15\%$
p20	$Z \to e^+e^-$	$0.98 \pm 15\%$
p20	$Z \to \mu^+\mu^-$	$1.00 \pm 15\%$

Table 7.2: Scaling factor for $Z \to \ell\ell$+jets Monte Carlo events in events with two or more jets.

7.5 Multijet Modeling and Normalization

The QCD multijet background is modeled using an *em-like* and a *muon-in-jet* multijet data sample. In the dielectron channel, objects that pass all electron ID criteria except the $Hmx7 > 50$ cut (see Section 4.2) are selected to model a jet faking an electron. In the dimuon channel, muons that pass all muon ID criteria except the joint isolation criterion (see Section 4.3) are selected to model a jet faking a muon. In both channels, events with two such objects are required, and they must also pass all other event selection criteria.

Normalization of this background is performed in the region relatively free from $Z \to \ell\ell$ events described below. A 100% systematic uncertainty is assigned to the modeling of this background due to the limited statistics of the reversed isolation / H-matrix sample. While this systematic uncertainty is large the net effect is negligible after the signal selection cuts are applied as seen in Section 8.2.2.

The following cuts are applied to select events for this sample.

- Require the dilepton mass be between 40 and 70 GeV to remove $Z \to \ell\ell$ events.
- Require missing $E_T < 25$ GeV to remove W+jets and $t\bar{t}$.
- Require each lepton $p_T < 40$ GeV to remove any remaining Z events.
- Require $\Delta R(l_1, l_2) > 2.0$ to remove any remaining possible signal.

Table 7.3 shows the amount by which the multijet events must be scaled to match the data in this region. These weights are used to normalize the multijet sample in the dilepton analysis. More details on the dilepton QCD method can be found in Ref. [64].

Normalization Factors

Reco Version	Final State	Scaling Factor
p17	$Z \to e^+e^-$	$1.34 \pm 100\%$
p17	$Z \to \mu^+\mu^-$	$1.54 \pm 100\%$
p20	$Z \to e^+e^-$	$2.90 \pm 100\%$
p20	$Z \to \mu^+\mu^-$	$3.20 \pm 100\%$

Table 7.3: QCD multijet scaling factors for both p17 and p20 reco version and both dielectron and dimuon samples.

7.6 Comparison of the Data with the Background Expectation

Once the Monte Carlo distributions are have been reweighted to correct known poor modelings we are ready to judge the overall level of agreement between data and the background estimation after preselection. Table 7.4 shows the expected and observed number of events for p17 and p20 dielectron and dimuon events passing the preselection cuts described in Section 7.2. The combined Run II data values are shown in Table 7.5.

The following Figures (7.2-7.11) compare the p17 and p20 data in dielectron and dimuon events with the background expectations.

Expected and Observed Yields for p17 and p20 dielectron and dimuon events

Source	p17 ee	p17 $\mu\mu$	p20 ee	p20 $\mu\mu$
QCD Multijet	20.4±5.0	53.4±7.8	52±73	159±22
Z+jets	1167±6	1554±7	3685±30	4064±26
W+jets	1.1±0.2	0.1±0.1	3.1±0.6	0.5±0.1
$t\bar{t}$	2.0±0.03	2.1±0.03	8.1±0.2	5.8±0.1
Diboson	19.5±0.3	23±0.3	66.9±1.1	58.0±0.9
Background Sum	1212±8	1633±10	3815±32	4287±35
Data	1281	1485	3775	4114
$Qq \to \ell\ell qq$(M=280 GeV)	31.6±0.6	23.2±0.5	117±3.1	82.1±2.4
$Qq \to \ell\ell qq$(M=320 GeV)	16.7±0.3	12.1±0.3	60.0±1.2	42.3±0.9
$Qq \to \ell\ell qq$(M=360 GeV)	9.6±0.2	6.5±0.1	34.0±0.7	24.2±0.6
$Qq \to \ell\ell qq$(M=400 GeV)	5.7±0.1	3.6±0.1	20.8±0.5	13.0±0.4
$Qq \to \ell\ell qq$(M=450 GeV)	2.9±0.05	1.7±0.04	10.1±0.3	6.2±0.2
$Qq \to \ell\ell qq$(M=500 GeV)	1.53±0.03	0.82±0.02	5.3±0.1	3.2±0.1
$Qq \to \ell\ell qq$(M=550 GeV)	0.79±0.01	0.40±0.01	2.9±0.1	1.4±0.04
$Qq \to \ell\ell qq$(M=600 GeV)	0.43±0.01	0.20±0.005	1.5±0.04	0.75±0.02
$Qq \to \ell\ell qq$(M=650 GeV)	0.22±0.004	0.09±0.002	0.80±0.02	0.32±0.01
$Qq \to \ell\ell qq$(M=700 GeV)	0.12±0.002	0.05±0.001	0.42±0.01	0.17±0.01

Table 7.4: Estimated background yields and number of observed data events passing preselection cuts. Also shown is the estimated signal yield for up-type vector-like quark production and decay using the same selection. The quoted errors are due to limited statistics.

Combined Run II Expected and Observed Yields for $\ell\ell$+jets events

Source	dielectron events	dimuon events	Combined Run II
QCD Multijet	72.2±13.5	212±24	285±27
Z+jets	4854±0.30	5618±27	10473±41
W+jets	4.1±0.7	0.5±0.2	4.7±0.7
$t\bar{t}$	10.1±0.2	7.9±0.1	18.0±0.2
Diboson	86.4±1.2	81.0±1.0	167±2
Background Sum	5027±33	5920±36	10947±49
Data	5056	5599	10655
$Qq \to \ell\ell qq$(M=280 GeV)	148±3	106±2	254±4
$Qq \to \ell\ell qq$(M=320 GeV)	76.7±1.2	54.8±0.9	132±2
$Qq \to \ell\ell qq$(M=360 GeV)	43.6±0.7	30.9±0.5	74.5±0.9
$Qq \to \ell\ell qq$(M=400 GeV)	26.5±0.5	16.7±0.4	43.2±0.7
$Qq \to \ell\ell qq$(M=450 GeV)	13.0±0.3	7.9±0.2	20.9±0.3
$Qq \to \ell\ell qq$(M=500 GeV)	6.8±0.1	4.1±0.1	10.9±0.2
$Qq \to \ell\ell qq$(M=550 GeV)	3.7±0.1	1.9±0.05	5.6±0.1
$Qq \to \ell\ell qq$(M=600 GeV)	1.9±0.04	0.96±0.02	2.9±0.05
$Qq \to \ell\ell qq$(M=650 GeV)	1.0±0.02	0.42±0.01	1.4±0.02
$Qq \to \ell\ell qq$(M=700 GeV)	0.54±0.01	0.22±0.01	0.76±0.01

Table 7.5: Estimated background yields and number of observed data events passing preselection cuts. Also shown is the estimated signal yield for up-type vector-like quark production and decay using the same selection. The quoted errors are due to limited statistics.

Preselection Plots for the ee+jets Final State (1)

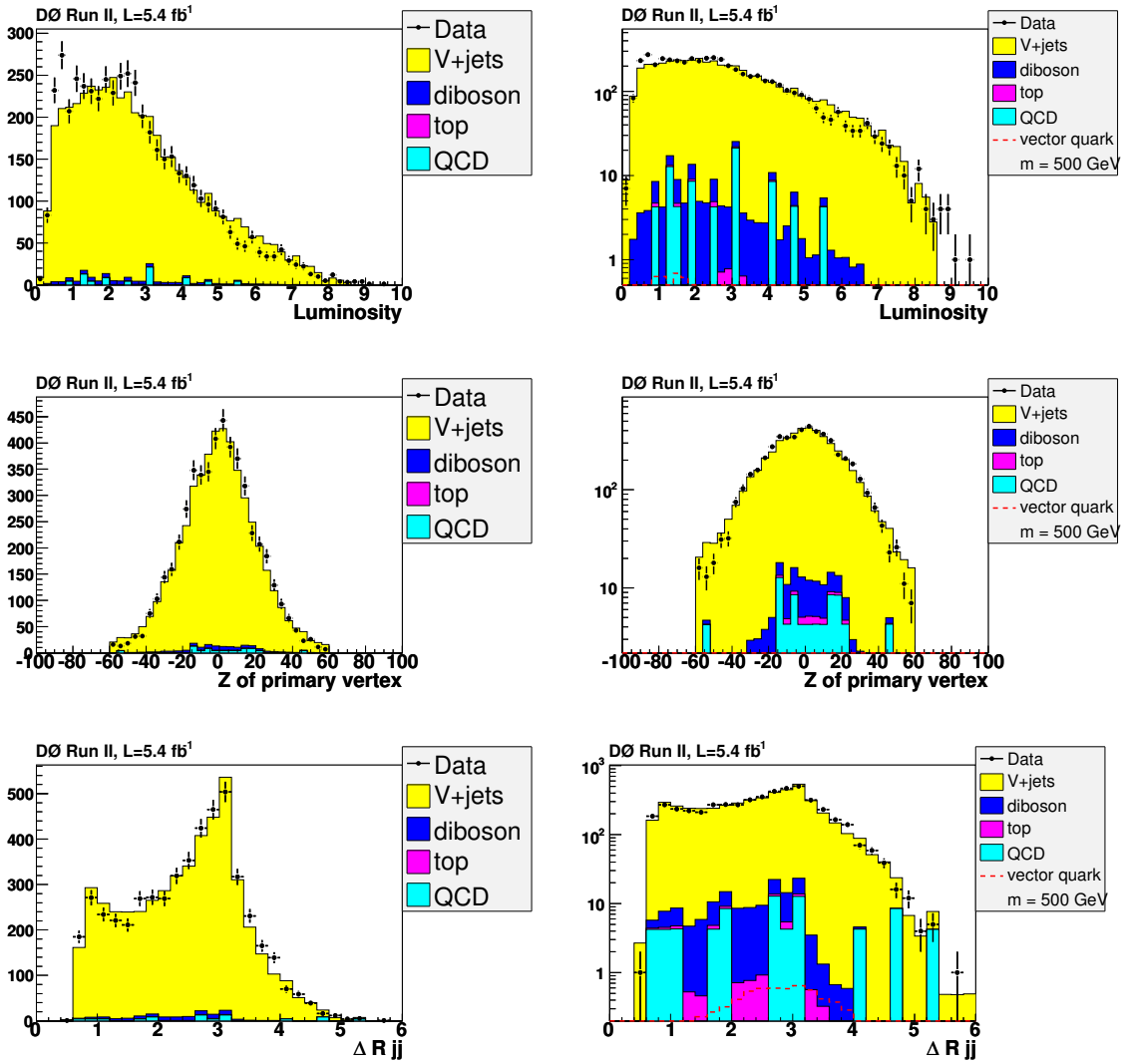

Figure 7.2: Comparison of the instantaneous luminosity (top row), z position of the primary vertex (middle row) and $\Delta R(j_1, j_2)$ (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots for the ee+jets Final State (2)

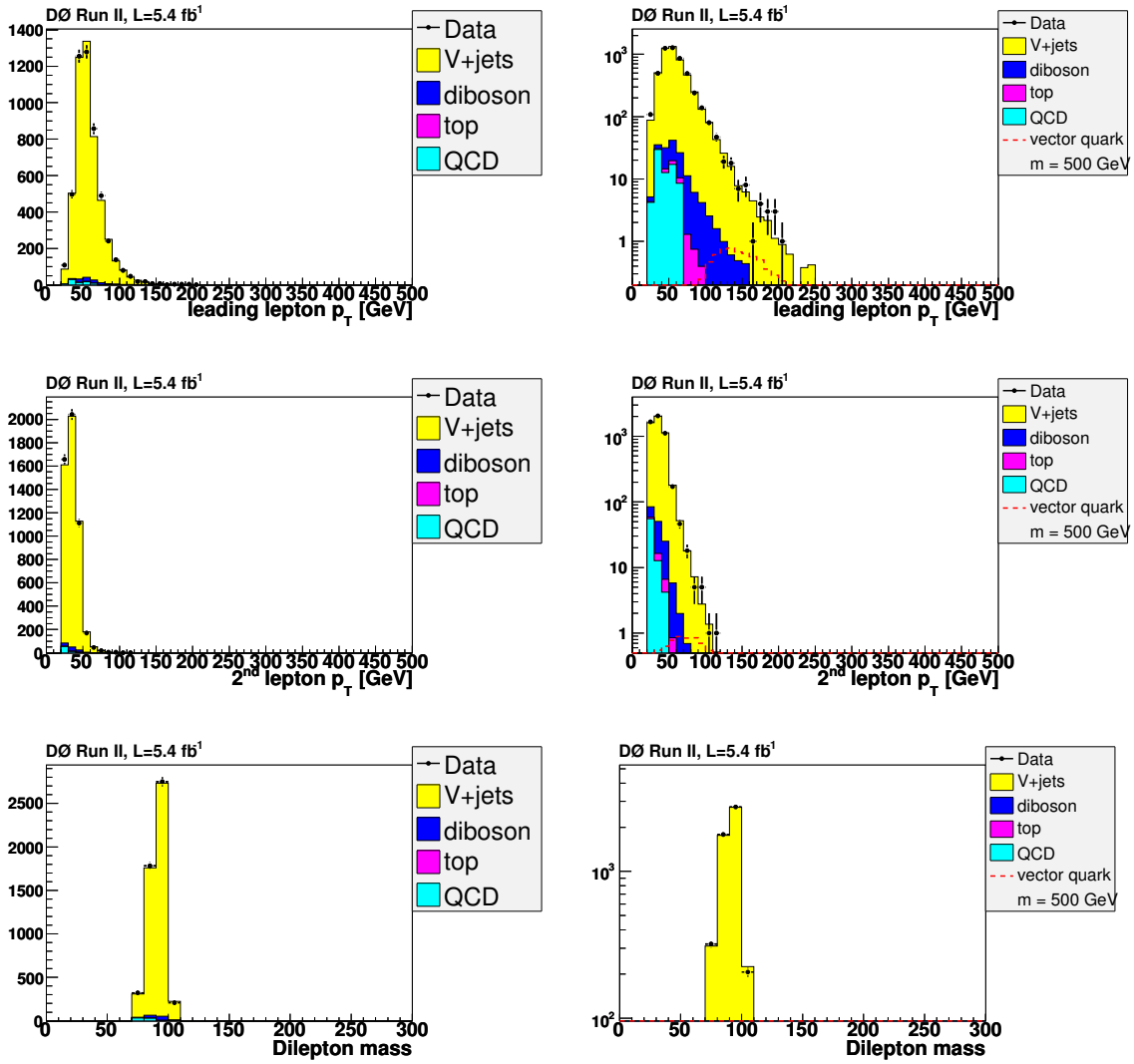

Figure 7.3: Comparison of the leading electron p_T (top row), sub-leading electron p_T (middle row) and dielectron mass (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots for the ee+jets Final State (3)

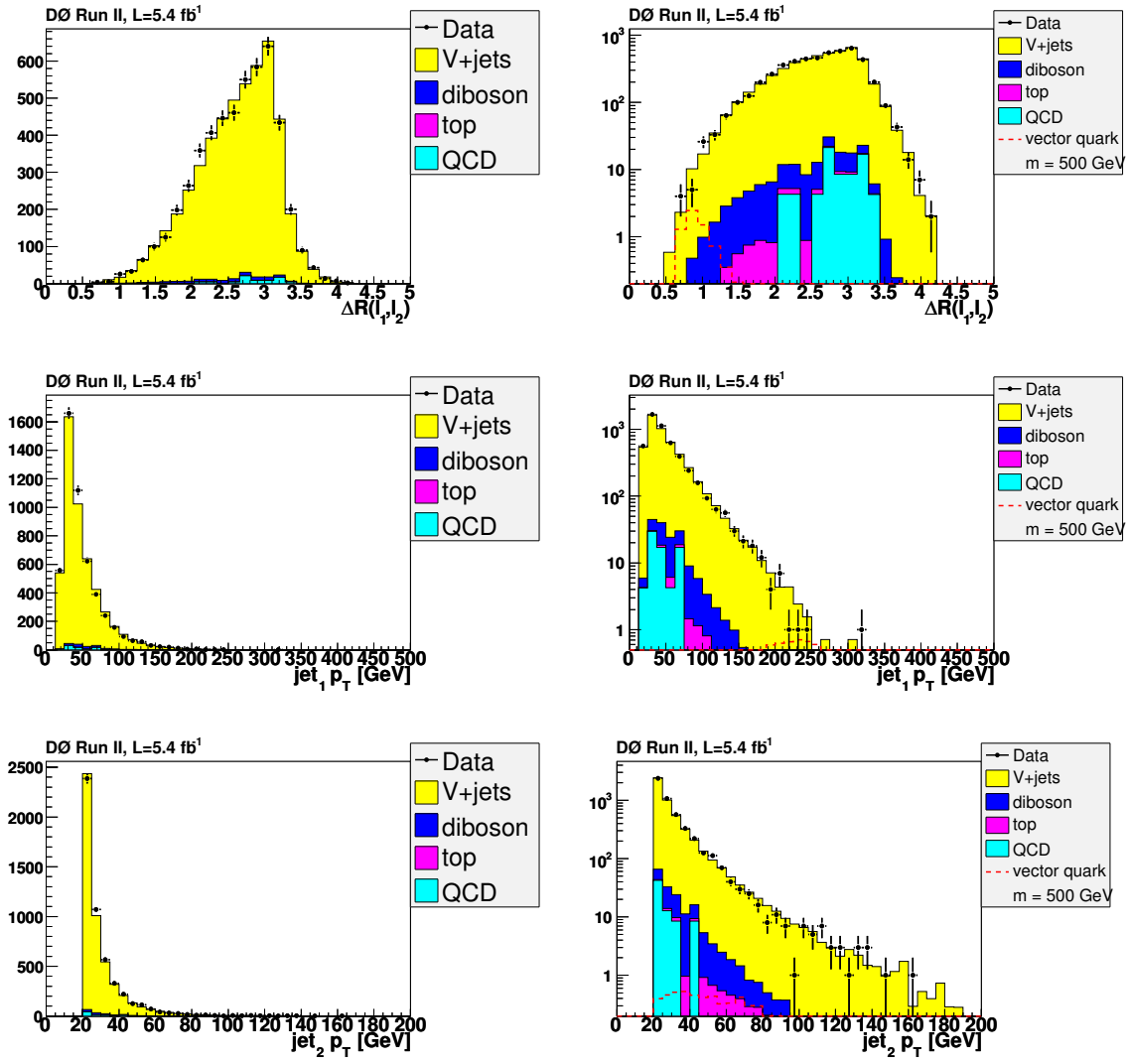

Figure 7.4: Comparison of the $\Delta R(e_1, e_2)$ (top row), leading jet p_T (middle row) and sub-leading jet p_T (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots for the ee+jets Final State (4)

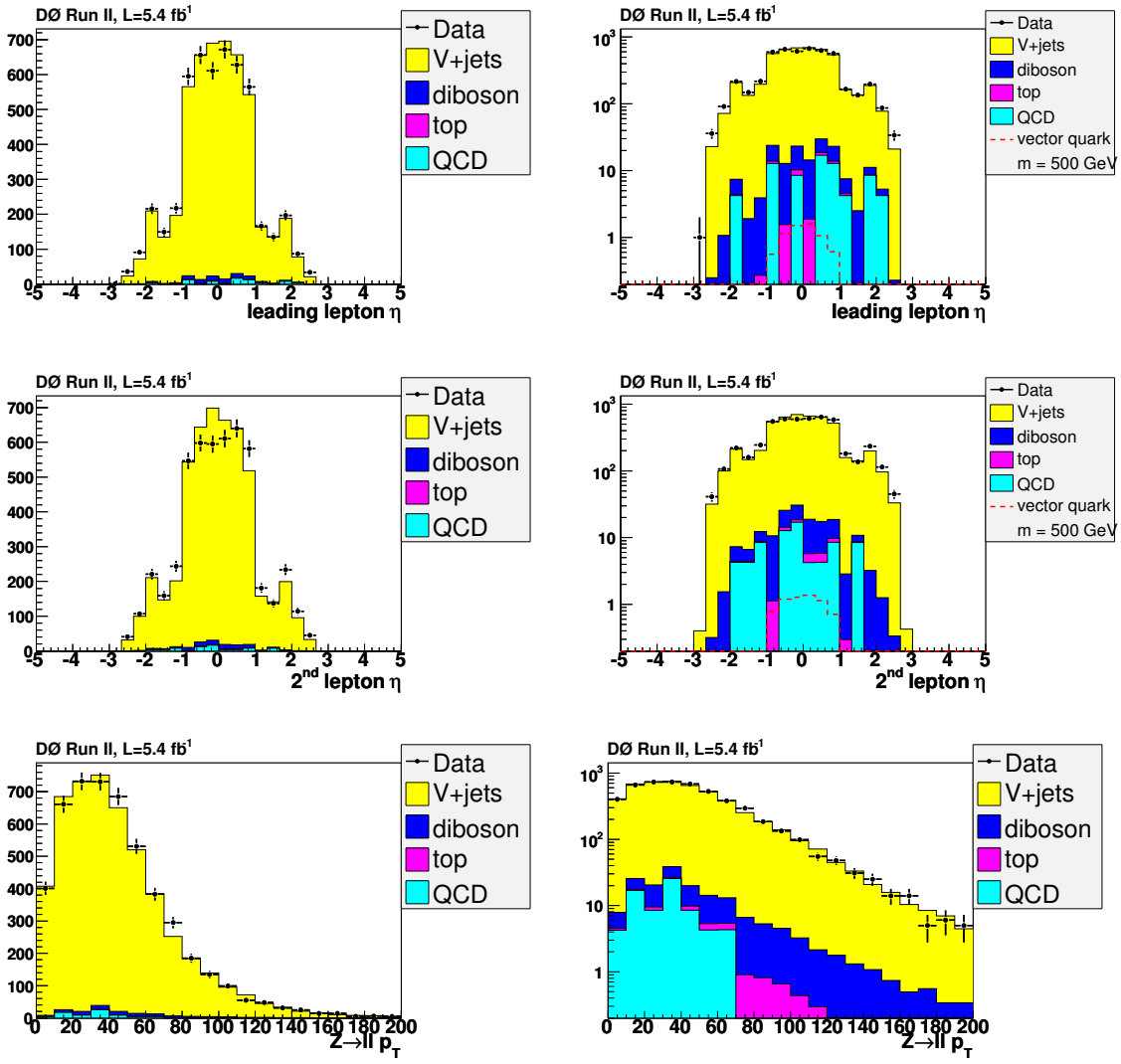

Figure 7.5: Comparison of the leading electron η (top row), sub-leading electron η (middle row) and p_T of the dilepton system (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots for the ee+jets Final State (5)

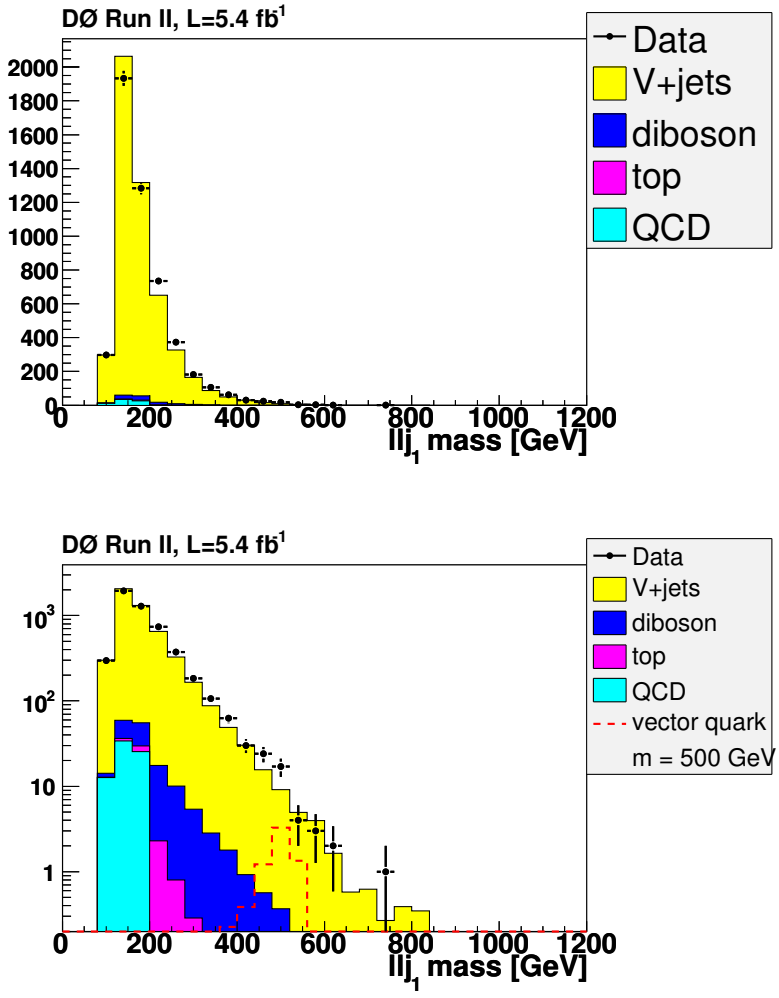

Figure 7.6: Comparison of the dielectron + leading jet invariant mass, M_{eej_1}. Shown with linear scale at top, and log scale at bottom.

Preselection Plots for the $\mu\mu$+jets Final State (1)

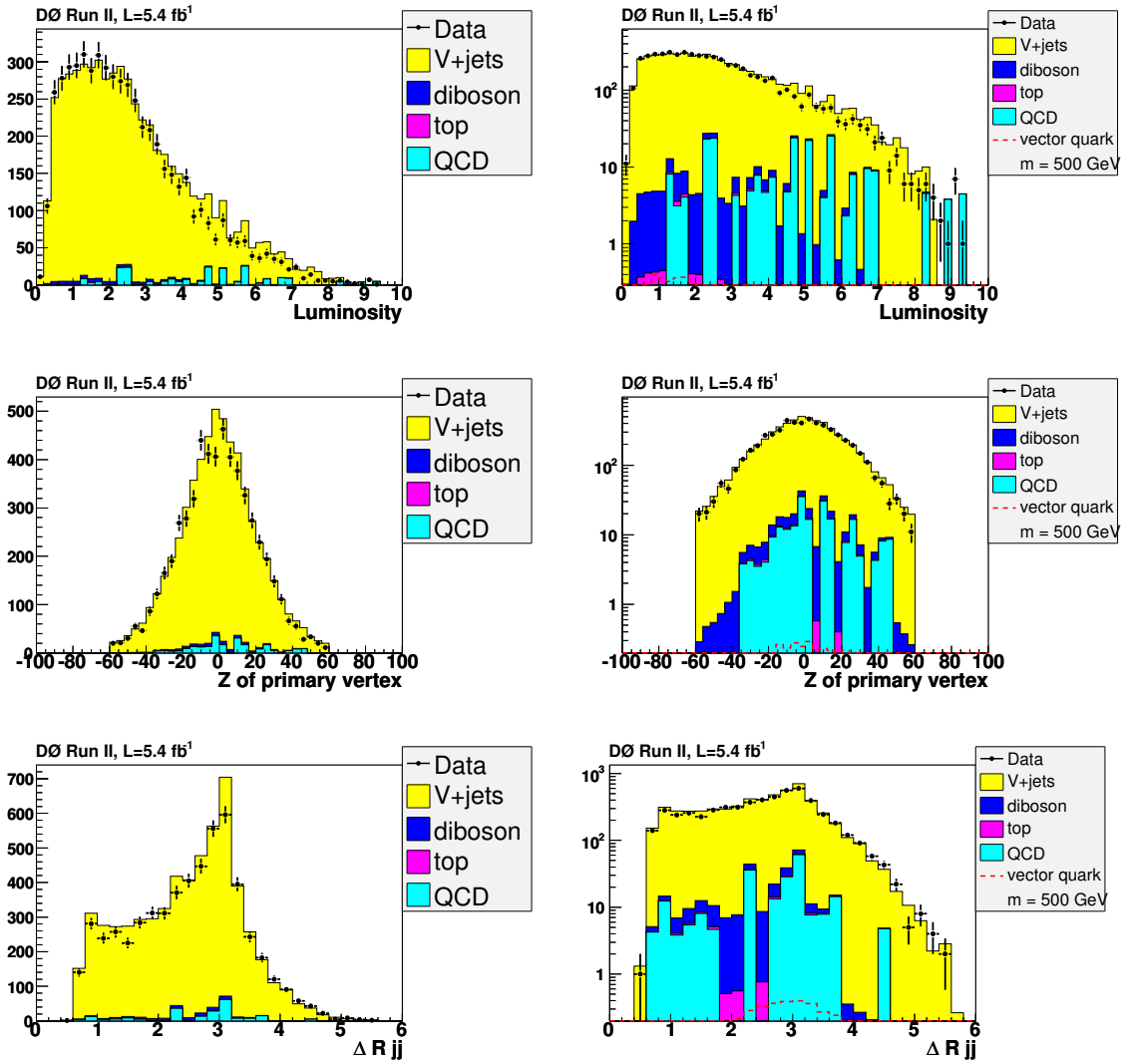

Figure 7.7: Comparison of the instantaneous luminosity (top row), z position of the primary vertex (middle row) and $\Delta R(j_1, j_2)$ (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots for the $\mu\mu$+jets Final State (2)

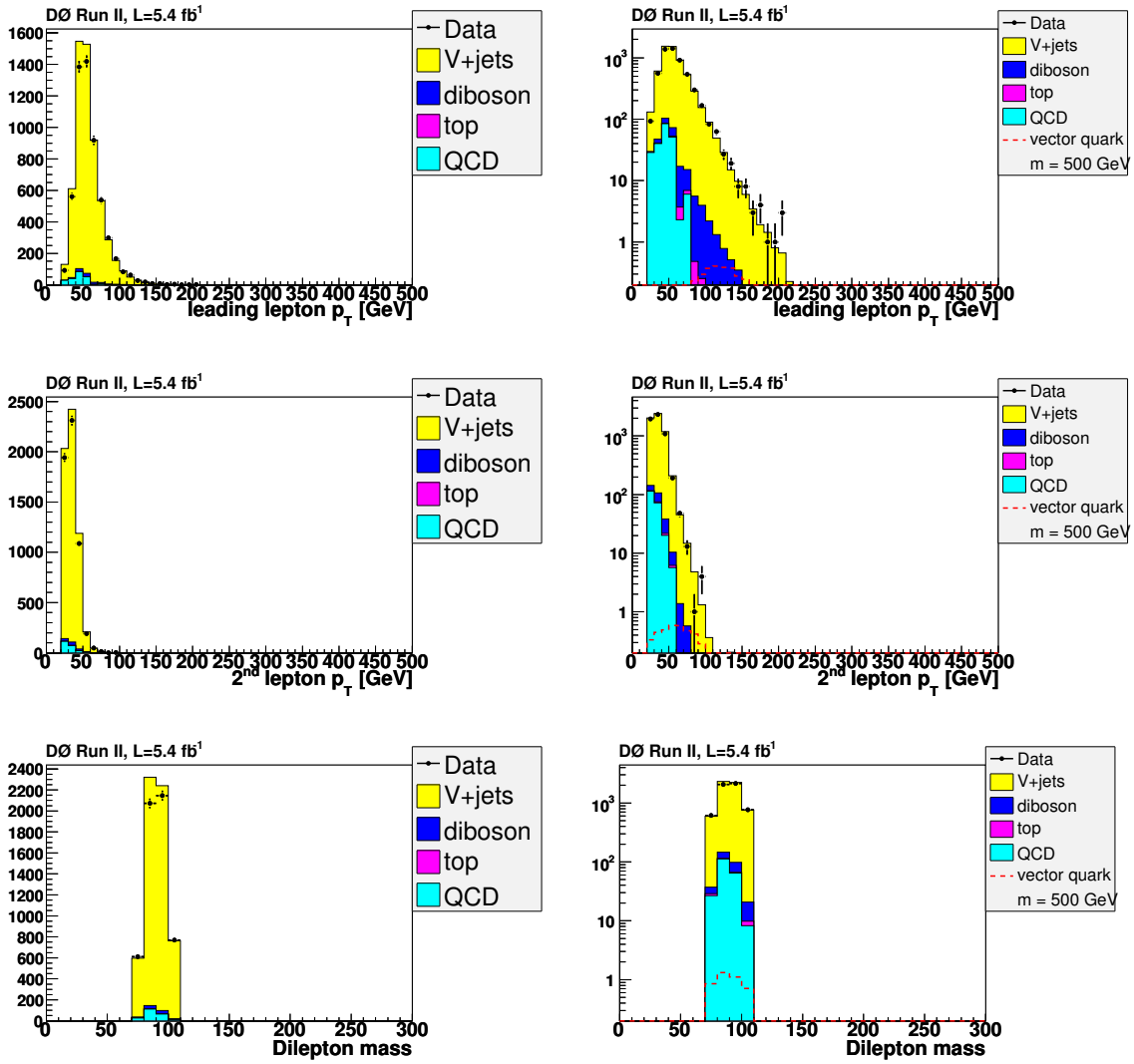

Figure 7.8: Comparison of the leading muon p_T (top row), sub-leading muon p_T (middle row) and dimuon mass (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots for the $\mu\mu$+jets Final State (3)

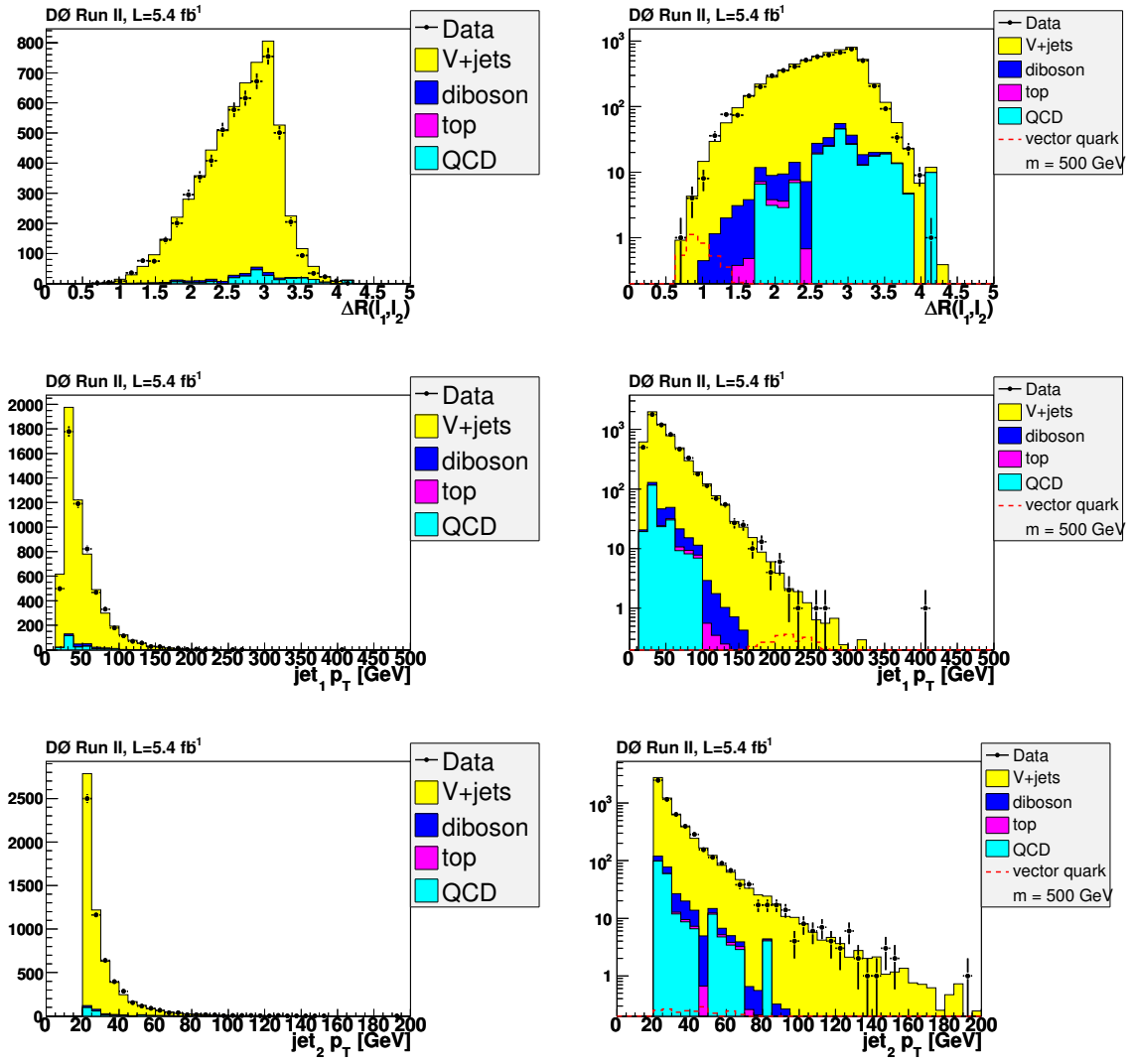

Figure 7.9: Comparison of the $\Delta R(\mu_1, \mu_2)$ (top row), leading jet p_T (middle row) and sub-leading jet p_T (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots for the $\mu\mu$+jets Final State (4)

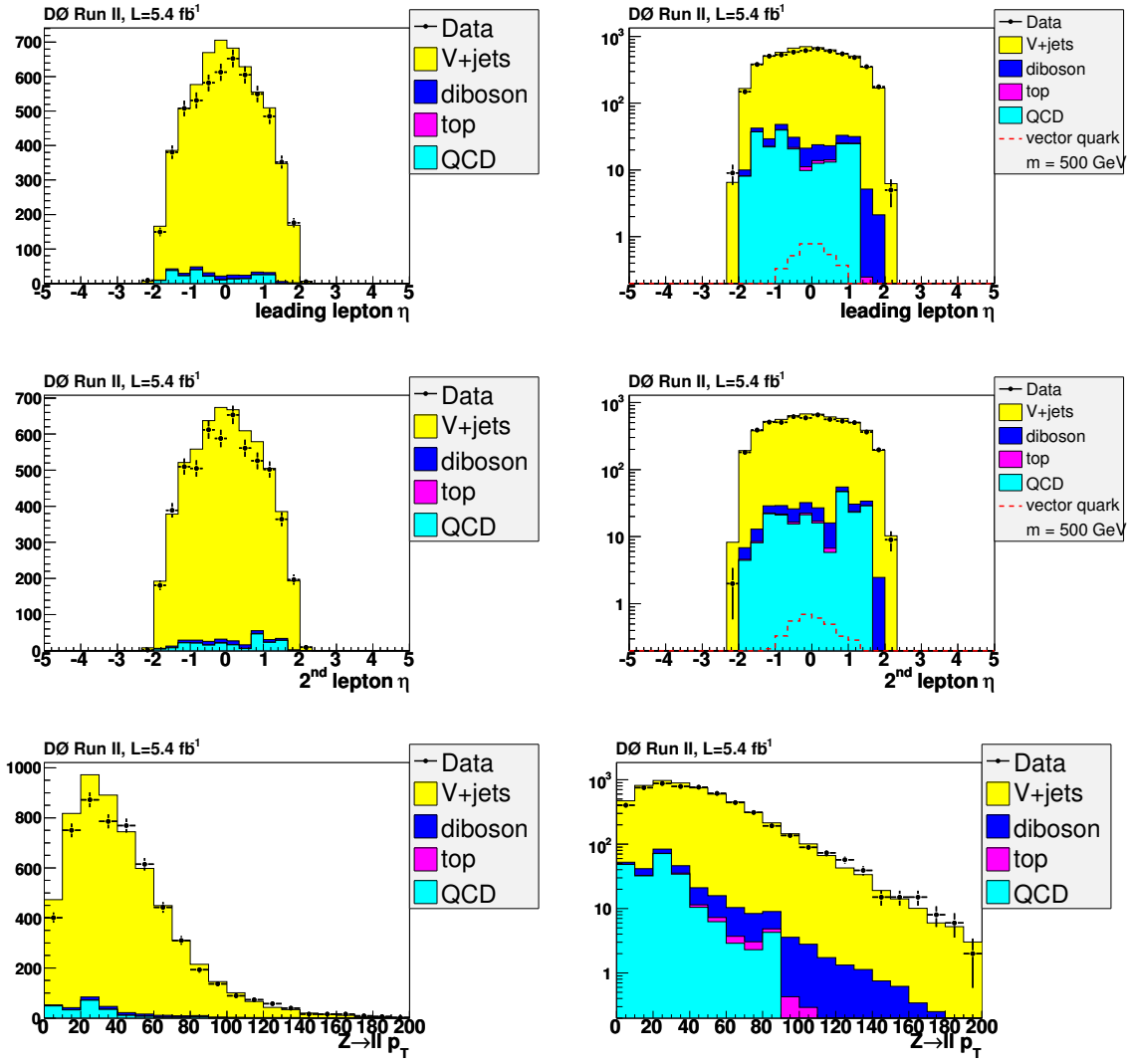

Figure 7.10: Comparison of the leading muon η (top row), sub-leading muon η (middle row) and p_T of the dilepton system (bottom row). Distributions are shown with linear scale at left, and log scale at right.

Preselection Plots for the $\mu\mu$+jets Final State (5)

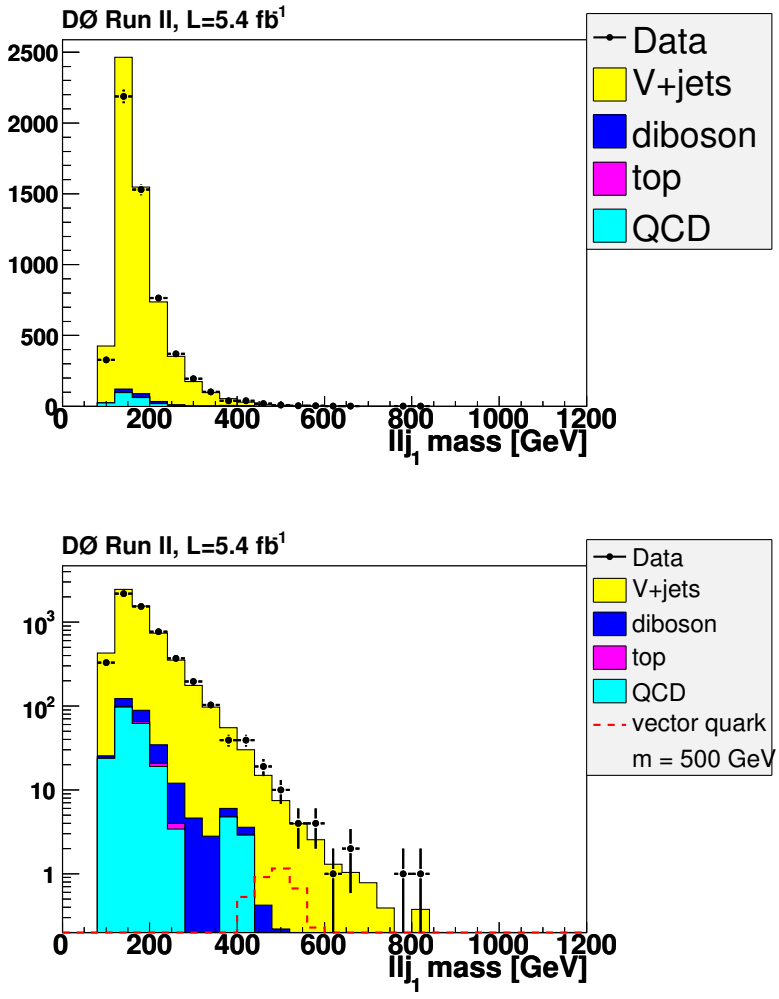

Figure 7.11: Comparison of the dimuon + leading jet invariant mass, $M_{\mu\mu j_1}$. Shown with linear scale at top, and log scale at bottom.

Chapter 8

Signal Enhancement

The signal-to-background ratio in the preselection sample is very low, as the purpose of the preselection stage is to verify that the background modeling reproduces the data. This chapter describes the signal enhancement procedure designed to isolate events consistent with a vector-like quark. Section 8.1 (8.2) describes this process for the single lepton (dilepton) analysis.

8.1 Signal Enhancement for the $Qq \to \ell\nu qq$ Analysis

This section describes the signal enhancement cuts applied to the data and the background prediction in order to reduce the large SM W+jets background in the $Qq \to \ell\nu qq$ searches. The cuts are designed to select events with kinematic features characteristic of a massive particle ($M > 300$ GeV/c^2)[1] decaying to a W boson and a quark and also to further reduce contamination from poorly-measured QCD events. These cuts are detailed in Section 8.1.1. An additional selection, targeted at a specific signature of down-type vector-like quark phenomenology, is described in Section 8.1.2.

[1]Vector-like quark masses below this are excluded by previous experiments.

The resulting event yields for data and background, as well as relevant distributions after final event selections, can be found in Section 8.1.3.

8.1.1 Final Selection Cuts

The following cuts were designed to isolate vector-like quark candidate events.

- Lepton $p_T > 50$ GeV.
- Leading jet $p_T > 100$ GeV. The leading jet in the event is assumed to originate from the heavy vector-like quark decay (see Figs. 6.11 and 6.16).
- $\not{E}_T > 40(50)$ GeV for the $\mu\nu jj$ ($e\nu jj$) channel. The electron analysis has a larger QCD multijet background at preselection, which is compensated by a harder $\not{E}_T$ cut (see Figs. 6.10 and 6.15).
- $\Delta\phi(\ell, \not{E}_T) < 2.0$. The W boson from the Q decay is significantly boosted, therefore the opening angle between the lepton and the neutrino is small (see Figs. 6.12 and 6.17).
- $M_T^W < 150$ GeV. The signal contains real W bosons and no other source of real $\not{E}_T$ (see Figs. 6.10 and 6.15).
- $Q_\ell \times \eta_{j_2} > 0$. A relationship between the lepton charge and the pseudorapidity of the sub-leading jet is exploited. This relationship is discussed in further detail in Section 8.1.2.

8.1.2 $W \to \ell\nu$ Charge and Jet η

As mentioned above, the leading jet in the event is assigned to the vector-like quark decay. The sub-leading jet is therefore assigned to the SM quark produced in association with the vector-like quark at the primary vertex. In signal events, this jet tends to be produced either forward or backward, rather than in the central region of

the detector. Furthermore, whether the jet is backwards or forwards in η is strongly correlated with the production of a down-type or anti-down-type heavy quark. Since this in turn determines the W boson charge from the heavy quark decay, we apply a cut $Q_\ell \times \eta_{j_2} > 0$, where Q_ℓ is the charge of the lepton in the event and η_{j_2} is the eta of the jet in the event with the second-highest p_T. This cut is approximately 85% efficient for the signal at all mass points (see Fig. 8.1), while reducing the background, which is symmetric in that distribution, by roughly a factor of two.

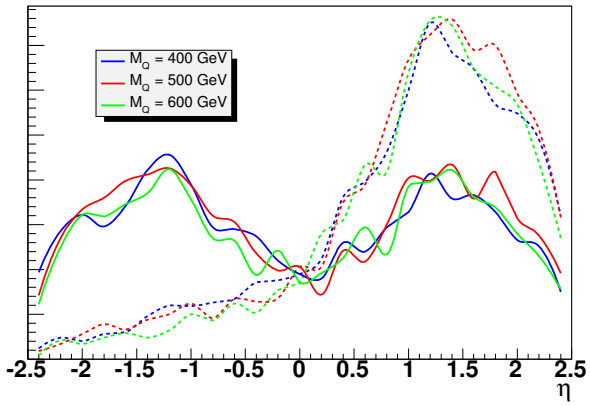

Figure 8.1: The distributions of sub-leading jet η for three down-type vector-like quark masses (solid lines). The result of multiplying by the charge of the lepton in the event is shown by the dashed curves.

8.1.3 Results of Final Selection

Expected and observed $\mu\nu jj$ and $e\nu jj$ event counts, along with theoretical signal yields, after final selection cuts are shown in Table 8.1. Cumulative efficiencies for each final selection cut relative to the preselection sample are shown in Tables 8.2 and 8.3. Figures 8.2-8.6 compare the combined Run II data in the two channels with the background expectations. Included is the vector-like quark transverse mass, or M_T^Q, defined as

$$(M_T^Q)^2 = (\sqrt{p_{T_W}^2 + M_W^2} + p_{T_j})^2 - (\vec{p}_{T_W} + \vec{p}_{T_j})^2, \tag{8.1}$$

where $M_W = 80.3$ GeV/c^2 is the W boson mass. M_T^Q is used as the final search variable in the $\ell\nu jj$ analysis.

Combined Run II Expected and Observed Yields for ℓ+jets Events After Final Selection

Source	$e\nu jj$ events	$\mu\nu jj$ events	$\ell\nu jj$ events
QCD Multijet	38.0±2.0	9.7±2.0	47.7±2.8
Z+jets	4.4±0.5	35.5±1.8	39.9±1.8
W+jets	511.7±10.1	390±6	901±12
top	117.2±1.0	75.6±0.6	193±1
Diboson	21.1±1.1	17.6±0.8	38.6±1.4
Background Sum	692±10.4	528±7	1220±12
Data	618	557	1175
$Qq \to \ell\nu qq$ (M=280 GeV)	380±13	369±12	748±18
$Qq \to \ell\nu qq$ (M=320 GeV)	346±9	313±8	659±12
$Qq \to \ell\nu qq$ (M=360 GeV)	264±6	202±5	467±8
$Qq \to \ell\nu qq$ (M=400 GeV)	182±5	158±5	340±7
$Qq \to \ell\nu qq$ (M=450 GeV)	111±3	89.1±2.3	200±4
$Qq \to \ell\nu qq$ (M=500 GeV)	63.4±1.5	53.0±1.3	116±2
$Qq \to \ell\nu qq$ (M=550 GeV)	35.5±0.9	27.6±0.7	63.2±1.1
$Qq \to \ell\nu qq$ (M=600 GeV)	20.1±0.5	15.6±0.4	35.7±0.6
$Qq \to \ell\nu qq$ (M=650 GeV)	10.8±0.2	8.2±0.2	19.0±0.3
$Qq \to \ell\nu qq$ (M=700 GeV)	6.1±0.1	4.4±0.1	10.5±0.1

Table 8.1: Estimated background yields and number of observed data events after final selection criteria are applied. Estimated signal yields for vector-like quark production and decay using the same selection are also shown. The quoted errors are due to limited statistics.

Final Selection Cut Efficiencies for e+jets Events

Source	p_{T_e}	p_{T_j}	$\not\!\!E_T$	$\Delta\phi(\ell, \not\!\!E_T)$	M_T^W	$Q_\ell \times \eta_{j_2}$
QCD Multijet	11.4%	2.0%	0.4%	0.4%	0.4%	0.2%
Z+jets	27.6%	4.6%	0.3%	0.2%	0.2%	0.1%
W+jets	16.4%	2.4%	1.1%	1.0%	1.0%	0.5%
top	38.4%	14.6%	8.6%	7.8%	7.7%	3.9%
Diboson	18.3%	3.2%	1.9%	1.8%	1.8%	0.8%
Background Sum	16.8%	2.9%	1.2%	1.1%	1.1%	0.6%
Data	16.0%	2.6%	1.2%	1.1%	1.1%	0.5%
$Qq \to \ell\nu qq$ (M=280 GeV)	56.9%	41.2%	23.8%	23.6%	23.6%	17.5%
$Qq \to \ell\nu qq$ (M=320 GeV)	63.8%	53.4%	36.1%	35.9%	35.9%	28.7%
$Qq \to \ell\nu qq$ (M=360 GeV)	73.3%	65.9%	48.5%	48.3%	48.3%	39.6%
$Qq \to \ell\nu qq$ (M=400 GeV)	76.4%	71.6%	57.1%	57.0%	57.0%	47.6%
$Qq \to \ell\nu qq$ (M=450 GeV)	81.9%	79.0%	65.8%	65.8%	65.8%	55.9%
$Qq \to \ell\nu qq$ (M=500 GeV)	84.8%	82.3%	70.9%	70.9%	70.9%	61.3%
$Qq \to \ell\nu qq$ (M=550 GeV)	88.7%	86.4%	76.5%	76.5%	76.4%	65.0%
$Qq \to \ell\nu qq$ (M=600 GeV)	88.8%	87.6%	79.3%	79.3%	79.3%	67.9%
$Qq \to \ell\nu qq$ (M=650 GeV)	91.4%	90.5%	82.9%	82.8%	82.7%	71.0%
$Qq \to \ell\nu qq$ (M=700 GeV)	91.6%	90.5%	84.1%	84.0%	83.9%	71.7%

Table 8.2: Cumulative efficiencies after each final selection cut for background, data and signal in the $e\nu jj$ channel.

Final Selection Cut Efficiencies for μ+jets Events

Source	p_{T_e}	p_{T_j}	$\not\!\!E_T$	$\Delta\phi(\ell, \not\!\!E_T)$	M_T^W	$Q_\ell \times \eta_{j_2}$
QCD Multijet	17.1%	3.0%	1.9%	1.1%	1.1%	0.5%
Z+jets	38.0%	4.5%	2.5%	1.6%	1.6%	0.8%
W+jets	29.4%	3.7%	2.4%	1.9%	1.9%	1.0%
top	40.1%	14.8%	10.9%	8.3%	8.2%	4.2%
Diboson	30.0%	4.7%	3.6%	3.1%	3.1%	1.4%
Background Sum	30.1%	4.2%	2.7%	2.1%	2.1%	1.1%
Data	29.2%	3.9%	2.8%	2.2%	2.2%	1.1%
$Qq \to \ell\nu qq$ (M=280 GeV)	55.1%	39.2%	29.5%	26.2%	26.1%	19.9%
$Qq \to \ell\nu qq$ (M=320 GeV)	60.5%	50.5%	41.1%	38.0%	37.7%	30.9%
$Qq \to \ell\nu qq$ (M=360 GeV)	68.7%	60.3%	51.4%	47.0%	46.6%	39.2%
$Qq \to \ell\nu qq$ (M=400 GeV)	75.1%	69.3%	61.5%	55.7%	55.0%	46.8%
$Qq \to \ell\nu qq$ (M=450 GeV)	78.1%	73.4%	66.6%	60.3%	59.7%	51.2%
$Qq \to \ell\nu qq$ (M=500 GeV)	81.5%	78.7%	73.6%	67.9%	66.9%	58.1%
$Qq \to \ell\nu qq$ (M=550 GeV)	84.5%	82.0%	77.5%	70.0%	68.7%	58.9%
$Qq \to \ell\nu qq$ (M=600 GeV)	86.3%	83.9%	79.2%	71.4%	70.0%	61.3%
$Qq \to \ell\nu qq$ (M=650 GeV)	88.2%	86.5%	82.7%	74.1%	72.9%	63.5%
$Qq \to \ell\nu qq$ (M=700 GeV)	88.9%	87.2%	83.7%	76.0%	74.4%	64.6%

Table 8.3: Cumulative efficiencies after each final selection cut for background, data and signal in the $\mu\nu jj$ channel.

Final Selection Plots for the e+jets Final State (1)

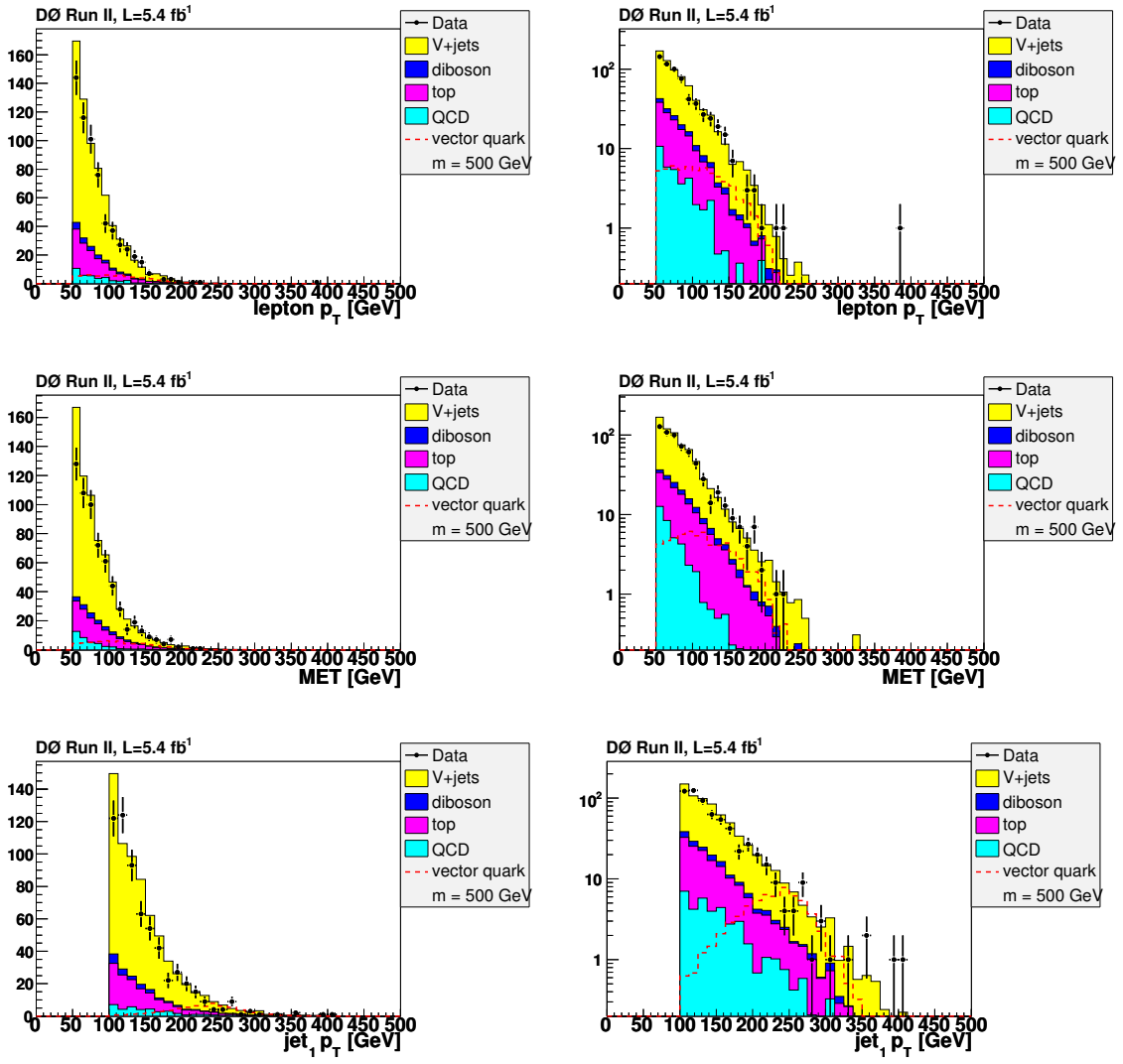

Figure 8.2: Comparison of the electron p_T (top row), missing E_T (middle row) and leading jet p_T (bottom row) after final selection cuts. Distributions are shown with linear scale at left, and log scale at right.

Final Selection Plots for the e+jets Final State (2)

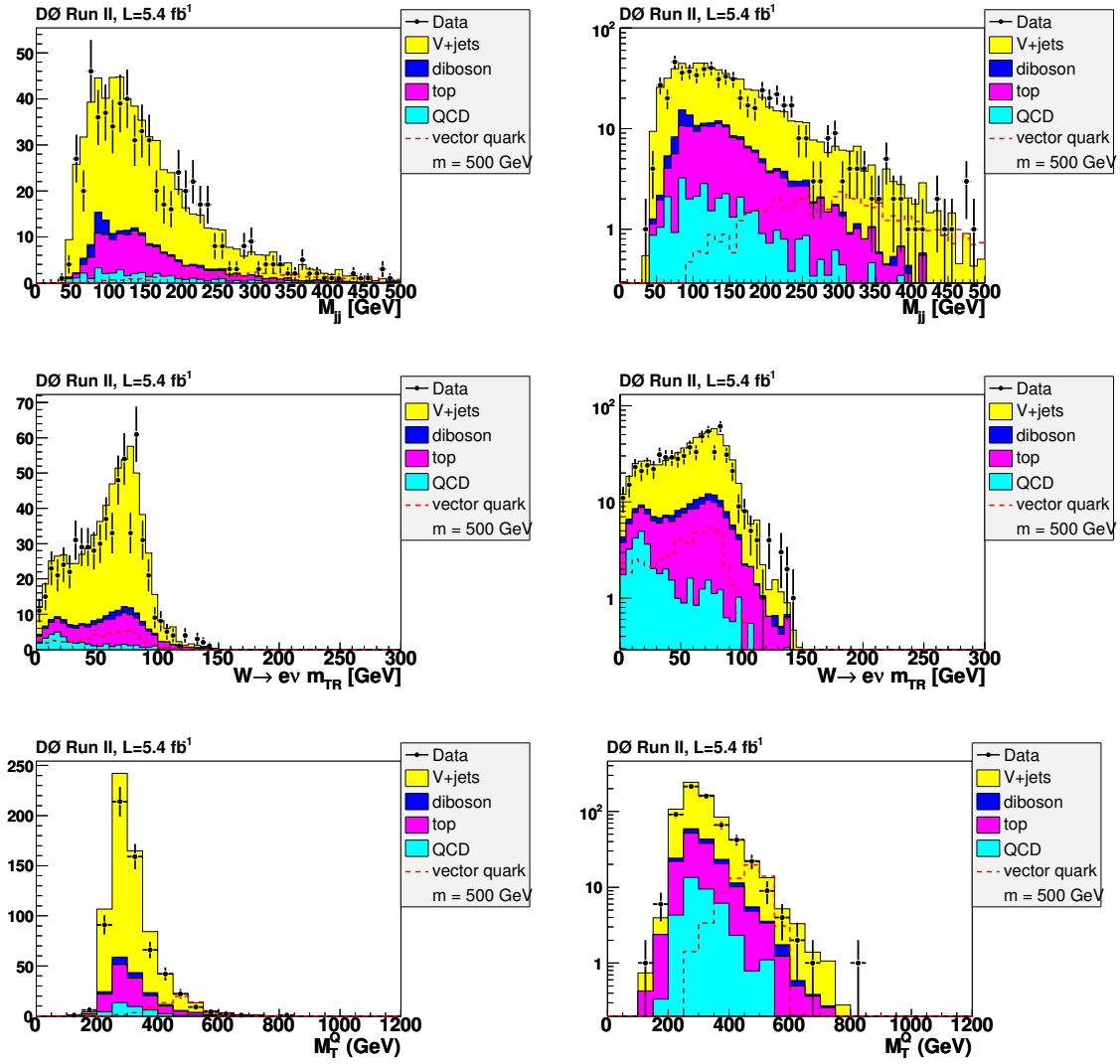

Figure 8.3: Comparison of the dijet mass (top row), W transverse mass (middle row) and vector-like quark transverse mass (bottom row) after final selection cuts. Distributions are shown with linear scale at left, and log scale at right.

Final Selection Plots for the μ+jets Final State (1)

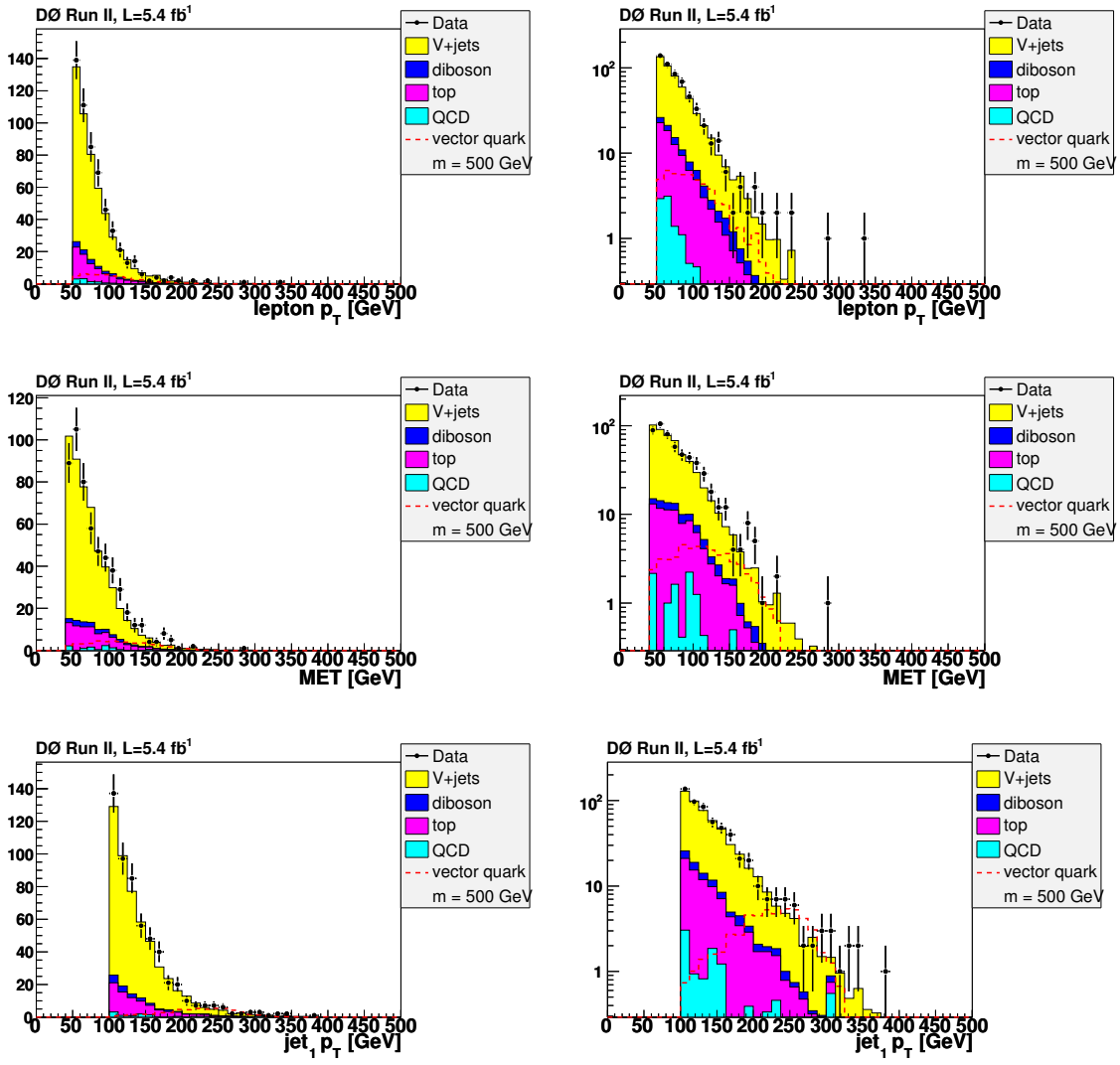

Figure 8.4: Comparison of the muon p_T (top row), missing E_T (middle row) and leading jet p_T (bottom row) after final selection cuts. Distributions are shown with linear scale at left, and log scale at right.

Final Selection Plots for the μ+jets Final State (2)

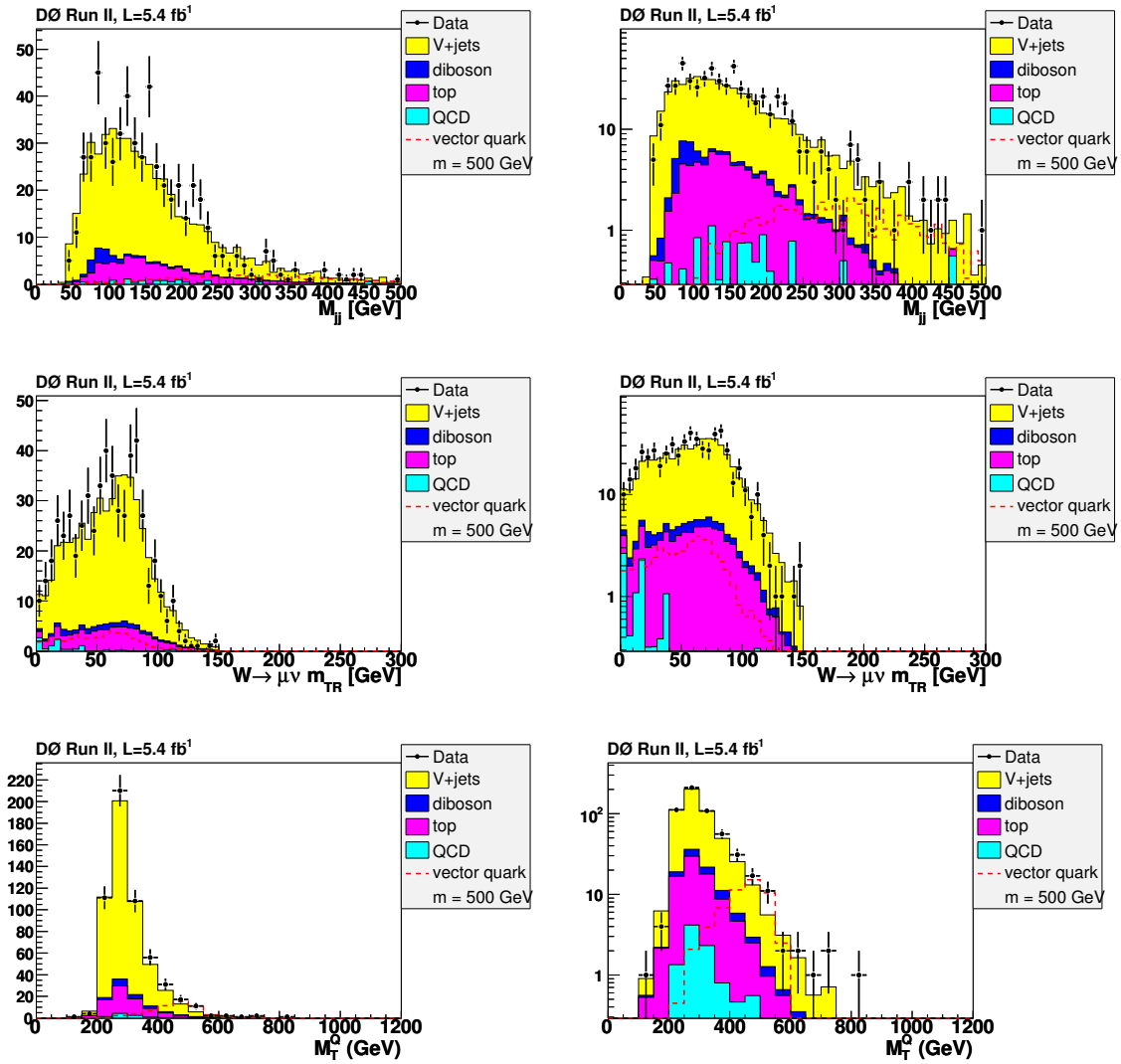

Figure 8.5: Comparison of the dijet mass (top row), W transverse mass (middle row) and vector-like quark transverse mass (bottom row) after final selection cuts. Distributions are shown with linear scale at left, and log scale at right.

Run II M_T^Q Mass Distribution

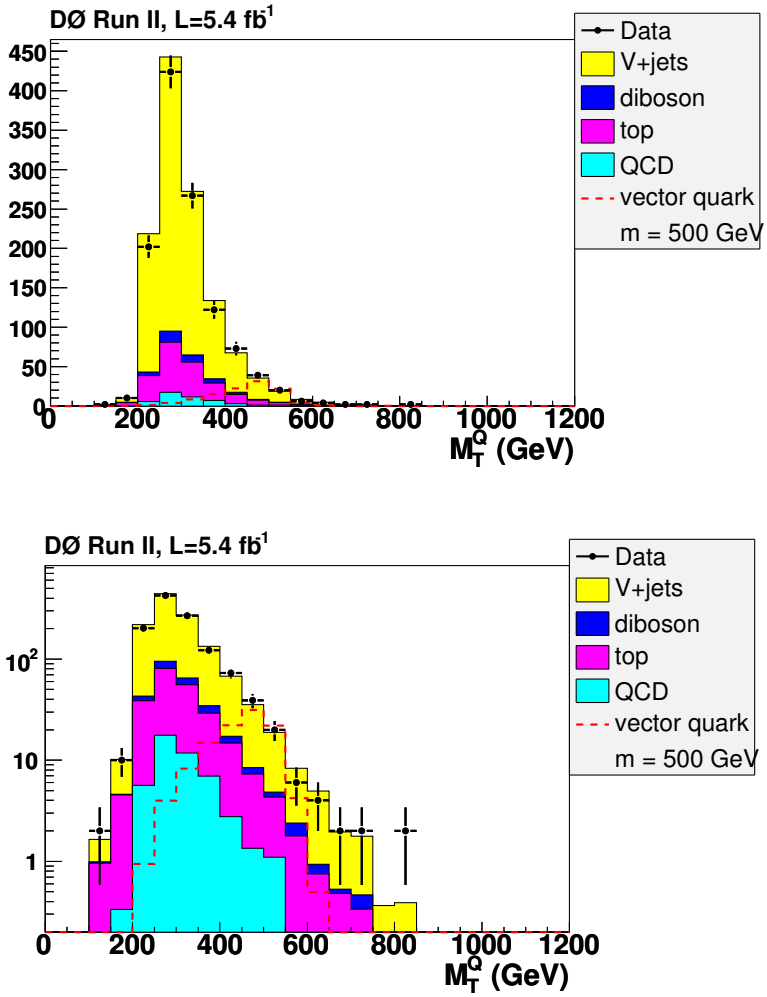

Figure 8.6: The reconstructed vector-like quark transverse mass (M_T^Q) for the Run II $e\nu jj$ and $\mu\nu jj$ combined samples after final selection criteria have been applied. The plot is shown in linear scale (top plot) and log scale (bottom plot).

8.2 Signal Enhancement for the $Qq \to \ell\ell qq$ Analysis

This section describes the signal enhancement cuts applied to the data and background prediction in order to reduce the large SM Z+jets background in the $Qq \to \ell\ell qq$ search. The cuts are designed to select events with kinematic features characteristic of a massive particle decaying to a Z boson and a jet, producing both decay objects at large transverse momentum. These cuts are detailed in Section 8.2.1. The resulting event yields for data and background, as well as relevant distributions after final event selections, can be found in Section 8.2.2.

8.2.1 Final Selection Cuts

The following cuts were designed to isolate vector-like quark candidate events.

- Leading jet $p_T > 100$ GeV. The leading jet in the event is assumed to originate from the Q decay (see Figs. 7.4 and 7.9).
- $Z \to \ell\ell$ $p_T > 100$ GeV. The reconstructed dilepton system is required to have large transverse momentum.
- $\Delta R(\ell, \ell) < 2.0$. The Z boson from the Q decay is significantly boosted, therefore the opening angle between two leptons is small (see Figs. 7.4 and 7.9).

8.2.2 Results of Final Selection

Expected and observed $\mu\mu jj$ and $eejj$ event counts, along with theoretical signal yields, after final selection cuts are shown in Table 8.4. Cumulative efficiencies for each final selection cut are shown in Tables 8.5 and 8.6. Figures 8.7-8.11 compare the combined Run II data in the two channels with the background expectations. Included is the full dilepton + leading jet invariant mass, $M_{\ell\ell j_1}$, used as the final search variable in the $\ell\ell jj$ analysis.

Combined Run II Expected and Observed Yields for $\ell\ell$+jets Events After Final Selection.

Source	$eejj$ events	$\mu\mu jj$ events	$\ell\ell jj$ events
QCD Multijet	0±0	0±0	0±0
Z+jets	144±4	118±3	262±5
W+jets	0.3±0.2	0.02±0.01	0.3±0.2
top	0.4±0.03	0.2±0.02	0.6±0.04
Diboson	4.9±0.3	3.4±0.2	8.3±0.3
Background Sum	150±4	122±3	271±5
Data	147	138	285
$Qq \to \ell\ell qq$(M=280 GeV)	59.5±2.0	37.0±1.4	96.5±2.5
$Qq \to \ell\ell qq$(M=320 GeV)	48.3±1.0	30.1±0.7	78.4±1.2
$Qq \to \ell\ell qq$(M=360 GeV)	33.8±0.6	21.6±0.5	55.3±0.8
$Qq \to \ell\ell qq$(M=400 GeV)	22.7±0.5	13.3±0.3	36.0±0.6
$Qq \to \ell\ell qq$(M=450 GeV)	12.0±0.3	6.6±0.2	18.6±0.3
$Qq \to \ell\ell qq$(M=500 GeV)	6.5±0.1	3.6±0.1	10.1±0.2
$Qq \to \ell\ell qq$(M=550 GeV)	3.6±0.1	1.7±0.04	5.3±0.1
$Qq \to \ell\ell qq$(M=600 GeV)	1.9±0.04	0.90±0.02	2.8±0.05
$Qq \to \ell\ell qq$(M=650 GeV)	1.0±0.02	0.40±0.01	1.4±0.02
$Qq \to \ell\ell qq$(M=700 GeV)	0.54±0.01	0.21±0.01	0.75±0.01

Table 8.4: Estimated background yields and number of observed data events after final selection criteria are applied. Estimated signal yields for up-type vector-like quark production and decay using the same selection are also shown. The quoted errors are due to limited statistics.

Final Selection Cut Efficiencies for ee+jets Events

Source	p_{T_j}	p_{T_Z}	$\Delta R(e_1, e_2)$
QCD Multijet	0.0%	0.0%	0.0%
Z+jets	6.6%	3.1%	3.0%
W+jets	15.7%	8.4%	6.2%
top	20.0%	3.9%	3.7%
Diboson	8.6%	5.9%	5.7%
Background Sum	6.6%	3.1%	3.0%
Data	6.1%	3.1%	2.9%
$Qq \to \ell\ell qq$ (M=280 GeV)	69.0%	41.0%	40.1%
$Qq \to \ell\ell qq$ (M=320 GeV)	79.8%	63.7%	63.0%
$Qq \to \ell\ell qq$ (M=360 GeV)	87.5%	78.2%	77.5%
$Qq \to \ell\ell qq$ (M=400 GeV)	91.3%	86.1%	85.7%
$Qq \to \ell\ell qq$ (M=450 GeV)	95.7%	92.3%	92.5%
$Qq \to \ell\ell qq$ (M=500 GeV)	97.2%	95.9%	95.8%
$Qq \to \ell\ell qq$ (M=550 GeV)	98.1%	97.4%	97.4%
$Qq \to \ell\ell qq$ (M=600 GeV)	99.0%	98.8%	98.7%
$Qq \to \ell\ell qq$ (M=650 GeV)	99.6%	99.5%	99.5%
$Qq \to \ell\ell qq$ (M=700 GeV)	99.6%	99.6%	99.6%

Table 8.5: Cumulative efficiencies after each final selection cut for background, data and signal in the $eejj$ channel.

Final Selection Cut Efficiencies for $\mu\mu$+jets Events

Source	p_{T_j}	p_{T_Z}	$\Delta R(\mu_1, \mu_2)$
QCD Multijet	0.0%	0.0%	0.0%
Z+jets	6.0%	2.2%	2.1%
W+jets	6.7%	4.3%	4.3%
top	19.2%	2.7%	2.6%
Diboson	7.4%	4.3%	4.2%
Background Sum	5.8%	2.2%	2.1%
Data	5.9%	2.5%	2.5%
$Qq \to \ell\ell qq$ (M=280 GeV)	66.5%	35.9%	35.1%
$Qq \to \ell\ell qq$ (M=320 GeV)	77.6%	56.1%	55.3%
$Qq \to \ell\ell qq$ (M=360 GeV)	85.0%	70.6%	70.2%
$Qq \to \ell\ell qq$ (M=400 GeV)	89.6%	80.6%	80.1%
$Qq \to \ell\ell qq$ (M=450 GeV)	91.3%	84.9%	84.4%
$Qq \to \ell\ell qq$ (M=500 GeV)	93.9%	89.3%	89.1%
$Qq \to \ell\ell qq$ (M=550 GeV)	95.6%	93.5%	93.4%
$Qq \to \ell\ell qq$ (M=600 GeV)	96.6%	94.0%	94.0%
$Qq \to \ell\ell qq$ (M=650 GeV)	97.4%	95.7%	95.7%
$Qq \to \ell\ell qq$ (M=650 GeV)	97.7%	96.2%	96.1%

Table 8.6: Cumulative efficiencies after each final selection cut for background, data and signal in the $\mu\mu jj$ channel.

Final Selection Plots for the ee+jets Final State (1)

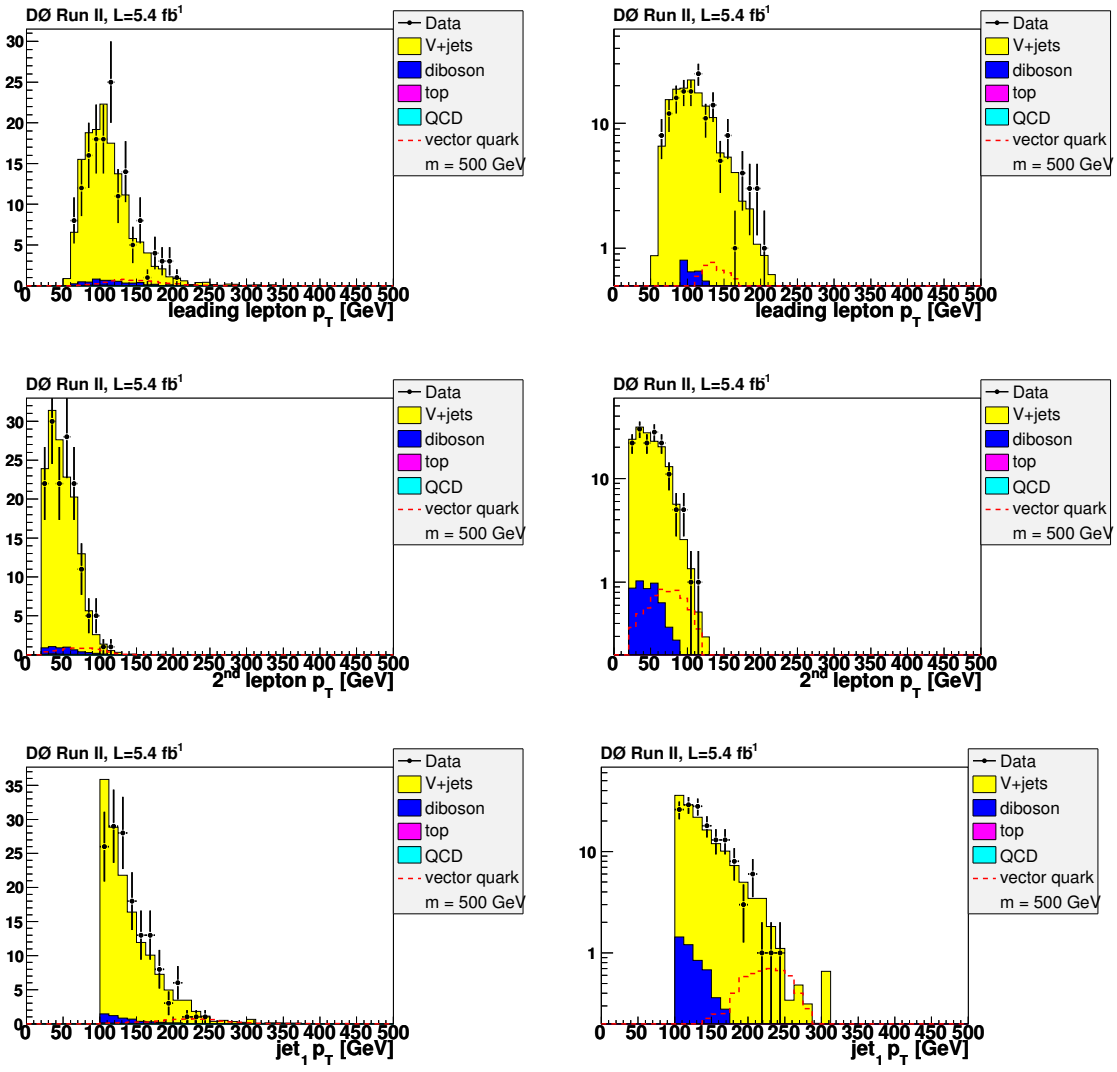

Figure 8.7: Comparison of the leading electron p_T (top row), sub-leading electron p_T (middle row) and leading jet p_T (bottom row) after final selection cuts. Distributions are shown with linear scale at left, and log scale at right.

Final Selection Plots for the ee+jets Final State (2)

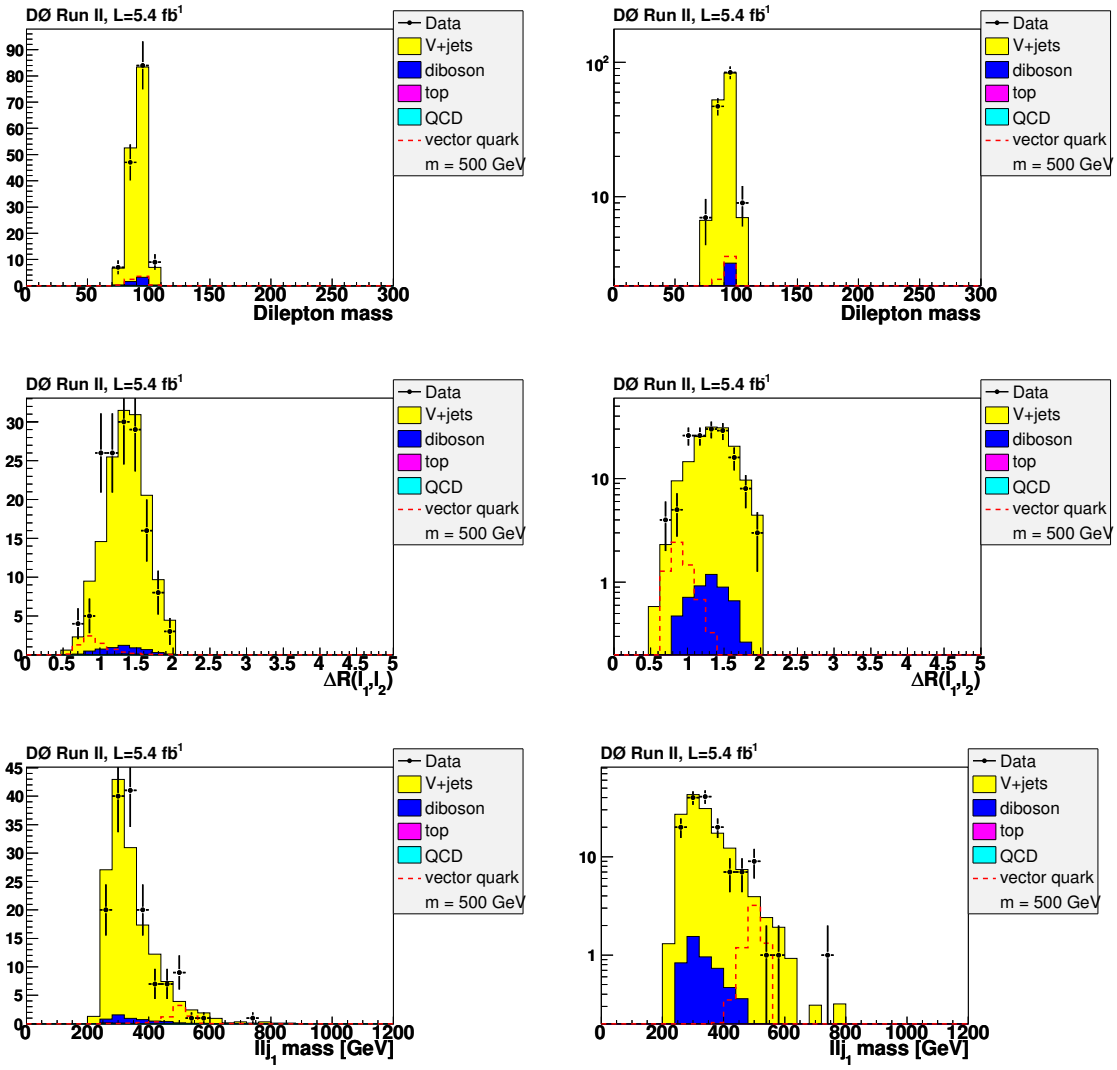

Figure 8.8: Comparison of the dielectron mass (top row), $\Delta R(e,e)$ (middle row) and M_{eej_1} (bottom row) after final selection cuts. Distributions are shown with linear scale at left, and log scale at right.

Final Selection Plots for the $\mu\mu$+jets Final State (1)

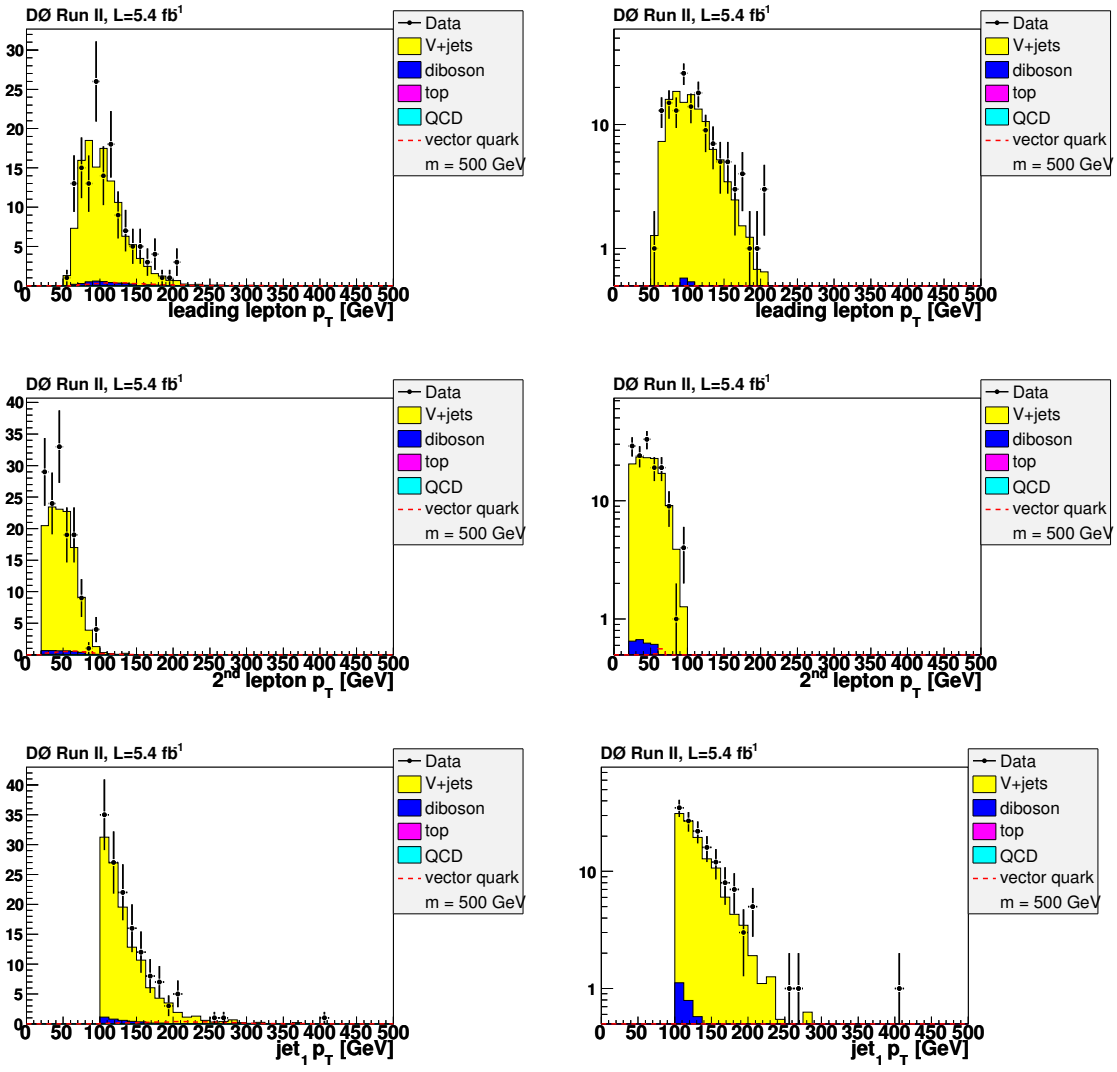

Figure 8.9: Comparison of the leading muon p_T (top row), sub-leading muon p_T (middle row) and leading jet p_T (bottom row) after final selection cuts. Distributions are shown with linear scale at left, and log scale at right.

Final Selection Plots for the $\mu\mu$+jets Final State (2)

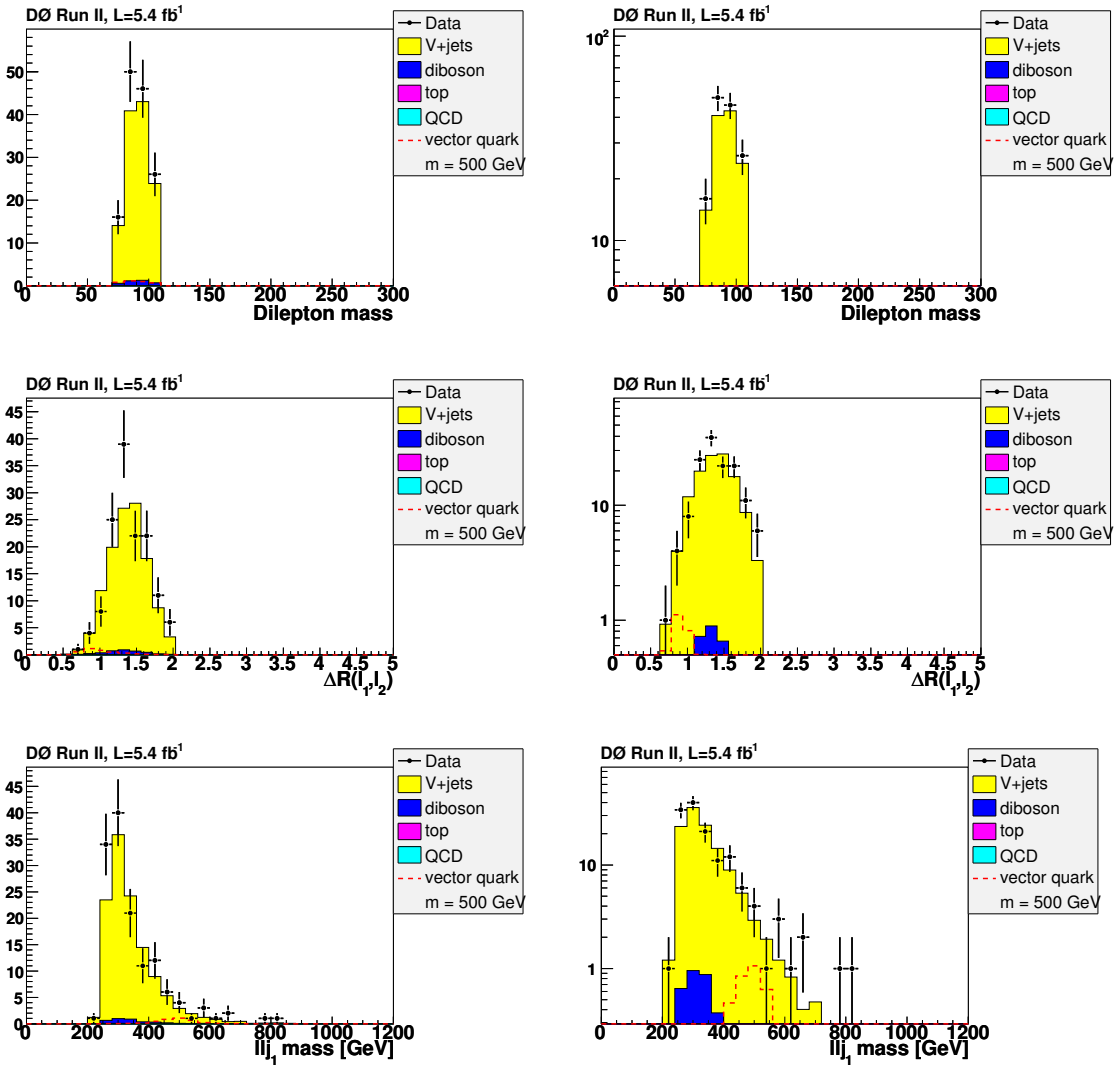

Figure 8.10: Comparison of the dimuon mass (top row), $\Delta R(\mu, \mu)$ (middle row) and $M_{\mu\mu j_1}$ (bottom row) after final selection cuts. Distributions are shown with linear scale at left, and log scale at right.

Run II $M(\ell\ell j_1)$ Mass Distribution

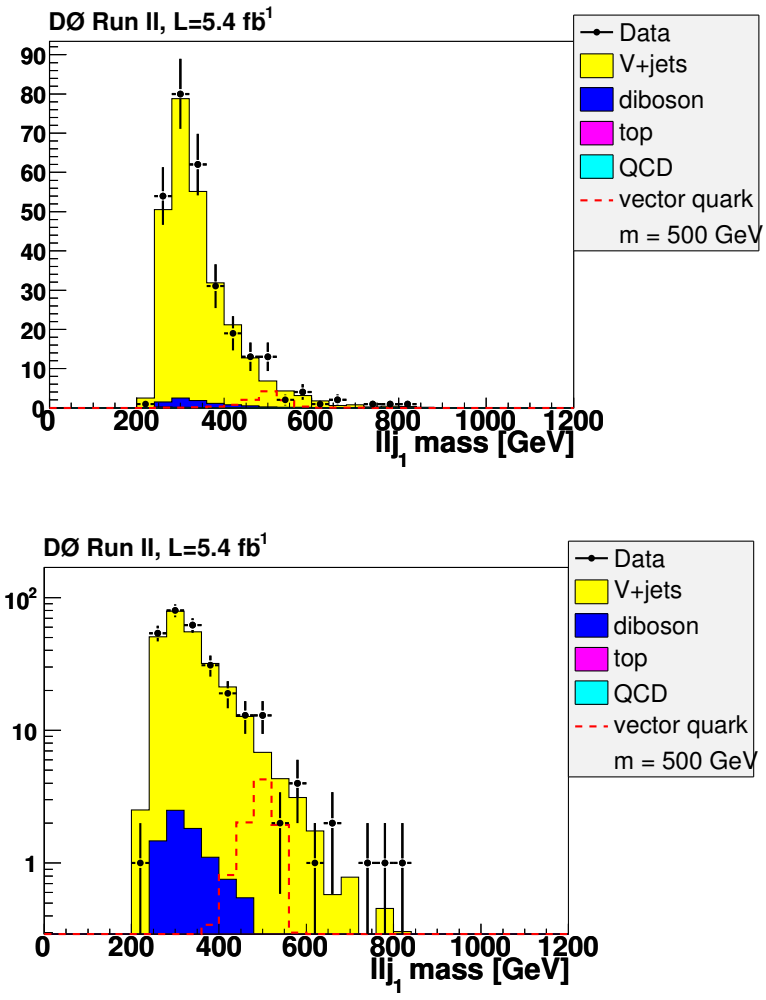

Figure 8.11: The reconstructed dilepton + leading jet mass ($M_{\ell\ell j_1}$) for the Run II $eejj$ and $\mu\mu jj$ combined samples after final selection criteria have been applied. The plot is shown in linear scale (top plot) and log scale (bottom plot).

Chapter 9

Results

No significant excess over the predicted background is observed in the data (see Table 9.1), therefore upper limits on vector-like quark production cross sections are calculated. Treatment of systematic uncertainties in the background and signal modeling is described in Section 9.1, and final results are shown in Section 9.3.

Source	$\ell\nu jj$ events	$\ell\ell jj$ events
QCD Multijet	47.7±4.7	< 0.1
Z+jets	39.9±7.4	262±45
W+jets	901±159	0.3±0.2
top	24±193	0.57±0.06
Diboson	38.6±3.8	8.3±0.7
Background Sum	1220±161	271±45
Data	1175	285

Table 9.1: Predicted number of background events with total uncertainties (including systematic uncertainties) and observed number of data events after final selection.

9.1 Systematic Uncertainties

Two types of systematic uncertainties are considered. Normalization uncertainties that affect the predicted number of events for each background and signal process, and shape uncertainties that affect the shapes of the final variable distributions: M_T^Q in the single lepton analysis, and $M_{\ell\ell j_1}$) in the dilepton analysis. All systematic uncertainties are treated as uncorrelated. The following systematic uncertainties are used in both the single lepton and dilepton analyses:

- V+jets modeling: (15%) determined conservatively by turning on/off the various V+jets reweightings described in Section 6.3.
- Production cross sections: on W/Z+jets (6%) [46], ttbar (10%) [49], diboson (6%) [53] production processes.

The following systematic uncertainties were applied to the single lepton analysis:

- Integrated Luminosity (6.1%) as suggested by the luminosity group [69].
- Muon ID efficiencies (3% per muon) as suggested in the muon ID certification notes [33] and [34].
- High-p_T muon modeling: We smear the muon p_T such that it reproduces the tail of the distribution in data, and propagate this smearing as a systematic uncertainty on the muon p_T determination.
- Electron ID efficiencies (3% per electron) as suggested in the electron ID certification notes [31] and [32].
- QCD multijet normalization: e+jets channel (6.5%), μ+jets channel (30%)[63].
- Trigger efficiency (1% in the electron channel, 4% in the muon channel). Signal trigger efficiency is over 99% in the electron channel. A conservative estimate is used for the muon channel.

- Jet Energy Resolution: $(1-5\%)$ A shape uncertainty determined by shifting the jet energy resolution in MC simulation by $\pm 1\sigma$, resulting in changes to the shapes of the predicted background and signal M_T^Q distributions.
- Jet Energy Scale: $(1-5\%)$ A shape uncertainty determined by shifting the jet energy scale in MC simulation by $\pm 1\sigma$.
- Jet ID efficiencies: $(1-5)$ A shape uncertainty determined by shifting the jet identification efficiency in MC simulation by $\pm 1\sigma$.
- Jet vertex confirmation: $(1-5)$ A shape uncertainty determined by shifting the jet vertex confirmation efficiency in MC simulation by $\pm 1\sigma$.

The following systematic uncertainties were applied to the dilepton analysis:

- Global MC background normalization (5%)[70].
- QCD normalization (100%) a conservative estimate based on the limited statistics of the reverse isolation sample.
- Jet Energy Resolution: (3%) Determined by shifting the jet energy resolution by $\pm 1\sigma$ in the signal Monte Carlo.
- Jet Energy Scale: (3%) Determined by shifting the MC jet energy scale by $\pm 1\sigma$ in the signal Monte Carlo.
- Jet ID efficiencies: (2%) Determined by shifting the MC jet ID efficiency by $\pm 1\sigma$ in the signal Monte Carlo.

9.2 Limit Setting Procedure

As no significant excess over the background expectation is observed in either the Wq or Zq channel, limits are set independently on production cross-sections for

vector-like quarks decaying to Wq and Zq final states. We employ a modified frequentist method, which uses a negative log likelihood ratio (NLLR) of the signal-plus-background (TEST) hypothesis to the background-only (NULL) hypothesis [71]. The NLLR can be expressed in terms of the expected background b, expected signal s and observed data d as:

$$\Gamma = 2 \sum_{i=1}^{N_{bins}} (s_i - d_i \ln(1 + \frac{s_i}{b_i})), \tag{9.1}$$

where the sum is over all bins of a given distribution. For the single lepton (Wq) channel, the vector-like quark transverse mass M_T^Q is used. For the dilepton (Zq) channel, the vector-like quark mass M_{llj_1} is used. A large number of pseudo-experiments is generated, smearing the expected number of background and signal events in each bin of the final variable by a Gaussian distribution whose width is determined by the total systematic and statistical uncertainty. The NLLR distributions generated by the pseudo-experiments are used to find CL_b, the probability for the NULL hypothesis to produce an outcome more background-like than the data, and CL_{s+b}, the probability for the TEST hypothesis to yield a more background-like outcome than the data. CL_s is defined as CL_{s+b}/CL_b, and the 95% confidence level cross section limit is calculated by finding the theoretical signal cross section for which $CL_s \leq 0.05$ [72].

9.3 Calculated Limits

Expected and observed upper cross-section limits versus vector-like quark mass are presented below. The expected limit is the limit that would have been obtained had the observed data exactly matched the background prediction. Figures 9.1 and 9.3 show the limits for vector-like quarks decaying to Zq and Wq final states, respectively. Figure 9.5 shows the combined limit, assuming degeneracy of up- and down-type vector-like quarks. The log-likelihood ratios of each limit calculation as a function of mass are shown in Figures 9.2, 9.4 and 9.6.

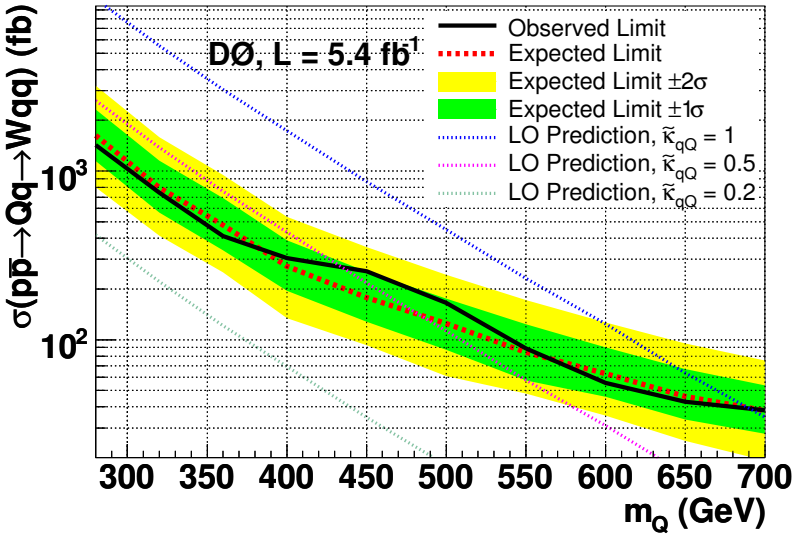

Figure 9.1: Exclusion curve from the $Qq \to \ell\nu qq$ analysis. Vector-like quark masses for which the predicted cross section is in excess of the observed cross section limit are excluded, implying a lower limit of 693 GeV/c^2 on the vector-like quark mass for a coupling $\tilde{\kappa}_{qQ} = 1$ in this channel.

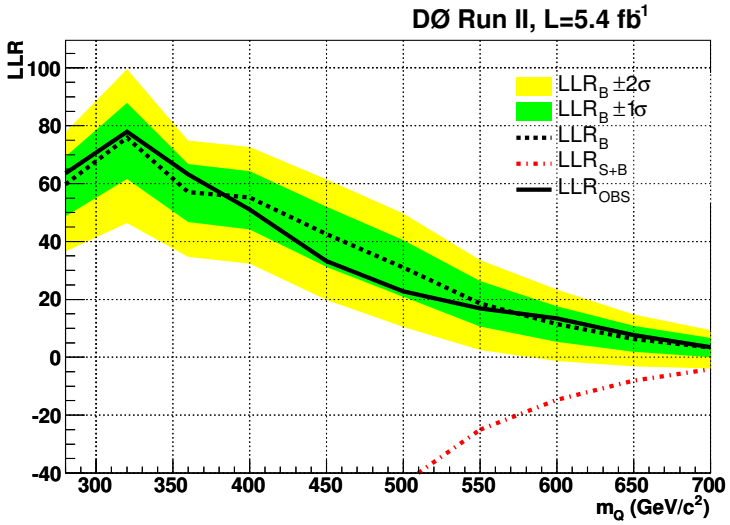

Figure 9.2: Log-likelihood ratio for the signal + background ($\mathrm{LLR_B}$) and background-only ($\mathrm{LLR_B}$) hypotheses, and for the observed data ($\mathrm{LLR_{OBS}}$), for the $Q \to Wq$ exclusion curve.

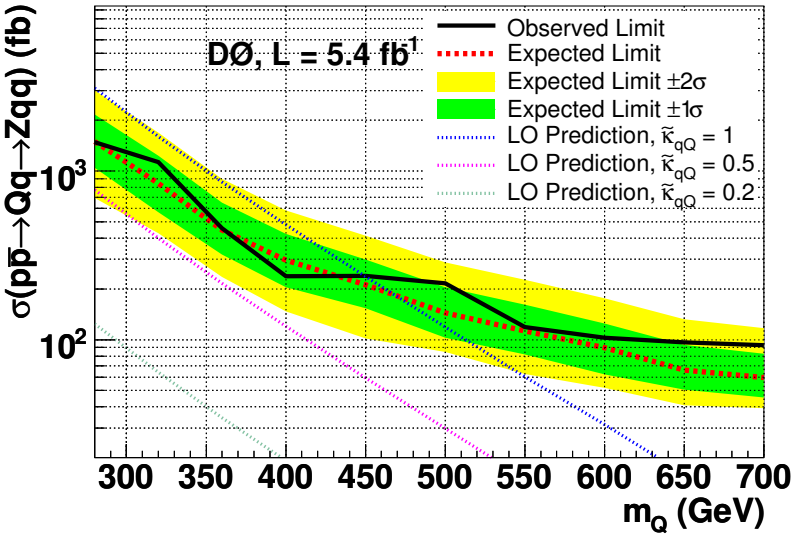

Figure 9.3: Exclusion curve from the $Qq \to \ell\ell qq$ analysis. For a coupling $\tilde{\kappa}_{qQ} = 1$, a vector-like quark mass below 449 GeV/c^2 is excluded in this channel.

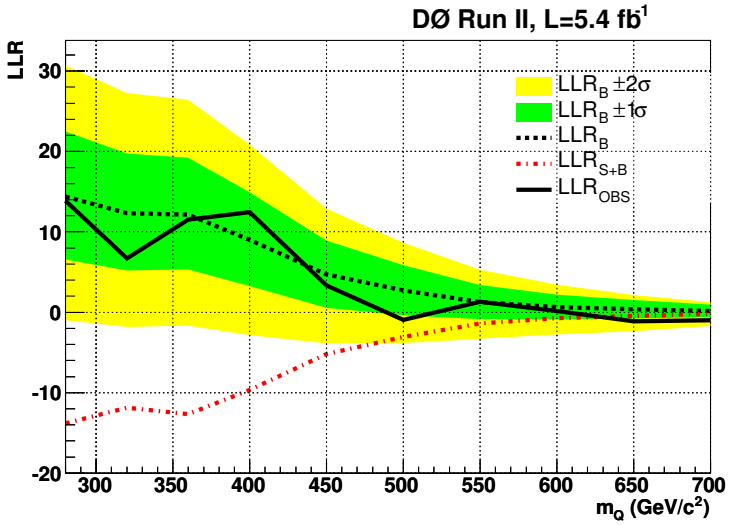

Figure 9.4: Log-likelihood ratios for the signal + background ($\mathrm{LLR_B}$) and background-only ($\mathrm{LLR_B}$) hypotheses, and for the observed data ($\mathrm{LLR_{OBS}}$), for the $Q \to Zq$ exclusion curve.

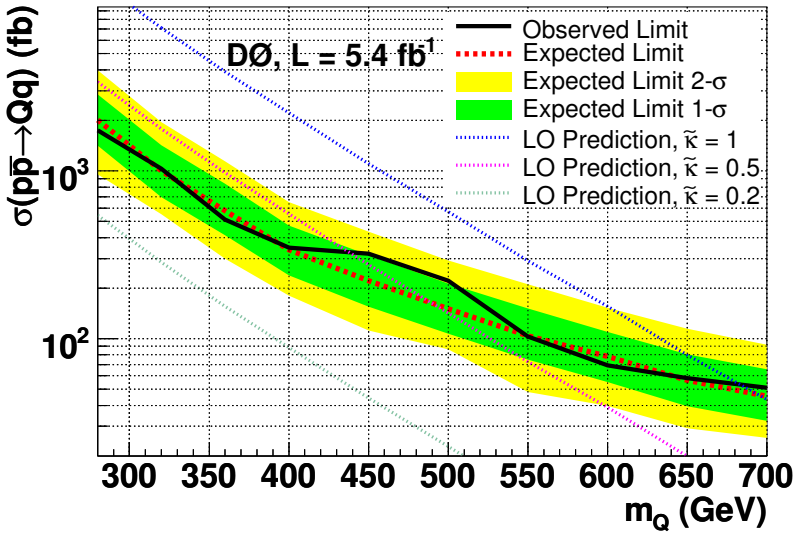

Figure 9.5: Combined exclusion curve. For degenerate Q_U and Q_D with a coupling $\tilde{\kappa}_{qQ} = 1$, masses below 685 GeV/c^2 are excluded.

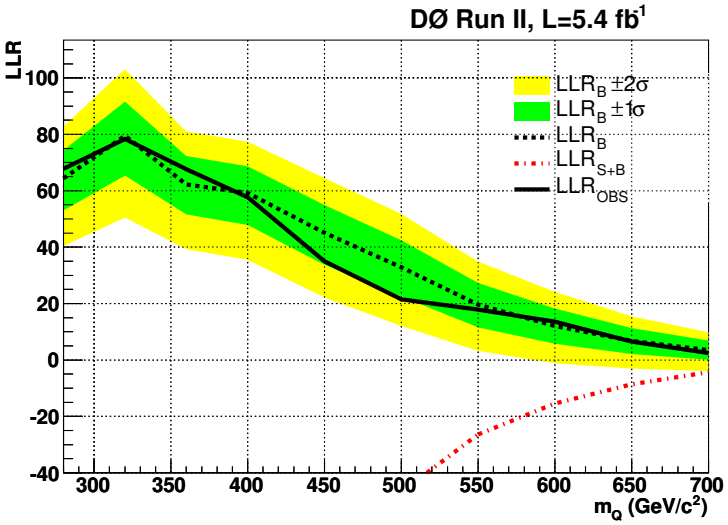

Figure 9.6: Log-likelihood ratio for the combined exclusion curve.

9.4 Conclusions

Using data corresponding to 5.4 fb^{-}1 from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, recorded by the DØ Detector at the Fermilab Tevatron, we have performed a search for theoretical vector-like quarks. With a combination of four channels ($e\nu jj$, $\mu\nu jj$, $eejj$ and $\mu\mu jj$), we have set the first limits on single vector-like quark production at the Tevatron. Assuming 100% branching ratios for $Q_D \to Wq$ and $Q_U \to Zq$, and a SM quark – vector-like quark mixing paramater $\tilde{\kappa}$ equal to one, we exclude a down-type vector-like quark of mass less than 693 GeV/c^2, and an up-type quark of mass less than 449 GeV/c^2 at 95% C.L. Under the same assumptions, we place a combined upper limit of 685 GeV/c^2 of vector-like quark mass.

Bibliography

[1] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19**, 1264 (1967).

[2] A. Salam, *Elementary Particle Theory*, Almquist and Wiksells, Stockholm (1968).

[3] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. D. **2**, 1285 (1970).

[4] R. K. Ellis *et al.*, *QCD and Collider Physics*, Cambridge (1996).

[5] C. Amsler *et al.*, *Review of particle physics 2008*, Phys. Lett. B. **667**, 1 (2008).

[6] CDF Collaboration, DØ Collaboration, Tevatron New Physics, Higgs Working Group, arXiv:hep-ex/1007.4587

[7] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370-3373 (1999).

[8] F. del Aguila, M. Perez-Victoria, and J. Santiago, J. High Energy Phys. **09**, 011 (2000).

[9] A. Atre, M. Carena, T. Han, and J. Santiago, Phys. Rev. D **79**, 054018 (2009).

[10] T. Applequist, H.C. Cheng, and B.A. Dobrescu, Phys. Rev. D **64**, 035002 (2002).

[11] K. Kong, S.C. Park, and T.G. Rizzo, arXiv:hep-ph/1004.4635.

[12] N. Arkani-Hamed, A.G. Cohen, E. Katz, and A.E. Nelson *et al.*, J. High Energy Phys. **07**, 034 (2002).

[13] T. Han, H. Logan, B. McElrath, and L. Wang, Phys. Rev. D **67**, 095004 (2003).

[14] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett.— **104**, 091801 (2010).

[15] TeVI Group, *Design report Tevatron 1 project*, FERMILAB-DESIGN-1984-01.

[16] P. C. Bhat and W. J. Spalding, *Fermilab collider Run II: accelerator status and upgrades*, AIP Conf. Proc. **753** 30 (2005).

[17] Frank T. Cole, (Ed.), Edwin L. Goldwasser, (Ed.), and Robert Rathbun Wilson, (Ed.), *National Accelerator Laboratory design report*, FERMILAB-DESIGN-1968-01.

[18] S. Abachi *et al.* (DØ Collaboration), *The DØ Detector*, Nucl. Inst. Methods A **338**, 185 (1994)

[19] V.M. Abazov *et al.* (DØ Collaboration), Nucl. Instrum. Methods Phys. Res. A **565**, 463 (2006).

[20] DØ Upgrade Collaboration, *DØ silicon tracker technical design report*, http://d0server1.fnal.gov/projects/silicon/www/silicon_old.html

[21] Michael S. Weber, *The DØ silicon strip tracker*, Nucl. Inst. Methods A **560** 14 (2006).

[22] DØ Upgrade Collaboration, *DØ Upgraded Central Fiber Tracker Technical Design Report*, http://d0server1.fnal.gov/projects/SciFi/cft_home.html

[23] Kotcher, J., *Design, Performance and Upgrade of the DØ Calorimeter*, DØ Note 2417 (1995).

[24] V. M. Abazov *et al.*, *The Muon System of the Run II DØ Detector*, Nucl. Inst. Methods A **552** 372-398 (2005).

[25] DØ Algorithms Group, *How to run RECO*, http://www-d0.fnal.gov/computing/algorithms/howto/howtoreco.html.

[26] DØ' Algorithms Group, *DØ Run II event data model*, http://www-d0.fnal.gov/software/data_model/data_model.html.

[27] A. Schwartzman and C. Tully, *Primary Vertex Reconstruction by Means of Adaptive Vertex Fitting*, DØ Note 4918 (2005).

[28] M. Narain *et al.*, *Probabilistic Primary Vertex Selection*, DØ Note 4042 (2002).

[29] F. Fleuret, *The DØ Electron/Photon Analysis Package emanalyze*, DØ Note 3888 (2001).

[30] J. Hays *et al.*, *Single Electron Efficiencies for p17 data and Monte Carlo*, DØ Note 5025 (2006).

[31] J. Hays *et al.*, *Electron Likelihood in p17*, DØ Note 5114 (2006).

[32] M. Aoki, *Electron Likelihood in p20*, DØ Note 5675 (2009).

[33] T. Gadfort *et al.*, *Muon Identification Certification for p17 Data*, DØ Note 5157 (2006).

[34] S. Choo *et al.*, *Muon ID Certification for p20 Data*, DØ Note 5824 (2008).

[35] J.-R. Vlimant *et al.*, *Technical Description of the T42 Algorithm for Calorimeter Noise Suppression*, DØ Note 4146 (2003).

[36] G. Blazey *et al.*, *Run II Jet Physics*, DØ Note 3750 (2000).

[37] *Preliminary p17 JES for Data and MC*, http://www-d0.fnal.gov/phys_id/jes/d0_private/certified/certified_jes.html.

[38] N. Makovec and J.-F. Grivaz, *Shifting, Smearing and Removing Simulated Jets*, DØ Note 4914 (2005).

[39] S. Trincaz-Duvoid and P. Verdier, *Missing E_T Reconstruction in p17*, DØ Note 4474 (2004).

[40] Common Samples Group
http://www-d0.fnal.gov/Run2Physics/cs/

[41] Data Quality Group
http://www-d0.fnal.gov/computing/data_quality/d0_private/forusers.html

[42] J. Alwall *et al.*, J. High Energy Phys. **09**, 028 (2007).

[43] J. Pumplin *et al.*, J. High Energy Phys. **07**, 012 (2002) and D. Stump *et al.*, J. High Energy Phys. **10**, 046 (2003).

[44] P. Meade and M. Reese, arXiv:hep-ph/0703031.

[45] T. Sjöstrand *et al.*, Comput. Phys. Commun. **135**, 238 (2001).

[46] M.L. Mangano *et al.*, J. High Energy Phys. **07**, 001 (2003).

[47] S. Höche *et al.*, arXiv:hep-ph/0602031.

[48] D. Gillberg, *Heavy Flavour Removal and Determination of Weighting Factors for* ALPGEN *W+jets Monte Carlo*, DØ Note 5129 (2006).

[49] S. Moch and P. Uwer, Phys. Rev. D **78**, 034003 (2008).

[50] J. Campbell and R.K. Ellis, Phys. Rev. D **65**, 113007 (2002).

[51] E. Boos *et al.*, Nucl. Instrum. Methods Phys. Res. A **534**, 250 (2004).

[52] A.S. Belyaev *et al.*, arXiv:hep-ph/0101232.

[53] J.M. Campbell and R.K. Ellis, Phys. Rev. D **60**, 113006 (1999).

[54] D.J. Lange, Nucl. Instrum. Meth. **A 462**, 152 (2001).

[55] S. Jadach *et al.*, Comput. Phys. Commun. **76**, 361 (1993).

[56] R. Brun and F. Carminati, CERN Program Library Long Writeup W5013, 1993 (unpublished).

[57] H. Schellman, *RunIIb longitudinal beam shape*, DØ Note 5540 (2007).

[58] H. Schellman, *The longitudinal shape of the luminous region at DØ*, DØ Note 5142 (2006).

[59] M. Shamim and T. Bolton, *Generator level reweighting of p_T of Z boson*, DØ Note 5565 (2008).

[60] G. Hesketh, *W p_T re-weighting for alpgen and pythia*, DØ Note 5786 (2008).

[61] V. Abazov *et al.* (DØ Collaboration), Phys. Lett. B. **669**, 278 (2008).

[62] E. Barberis *et al.*, *The Matrix Method and its Error Calculation*, DØ Note 4564 (2004).

[63] W. Fisher, J. Haley, J. Sekaric, *A Measurement of Diboson Production in Lepton Plus Jet Decays* DØ Note 5544, (2008).

[64] A. Askew *et al.*, *Search for $W' \to WZ$ production in the lepton+jets and dilepton+jets final states using 5.4 fb^{-1} of Run II Data*, DØ Note 6022 (2010). `In review`

[65] L. Han *et al.*, *Measurement of the Forward-Backward Charge Asymmetry in $p\bar{p} \to Z\gamma^* + X \to ee + X$ events produced at $\sqrt{s} = 1.96$ TeV*, DØ Note 5603 (2007).

[66] X. Bu *et al.*, *Search for the standard model Higgs boson in di-photon final states at DØ with 4.2 fb^{-1} data*, DØ Note 5846 (2007).

[67] A. Dominguez *et. al.*, *Search for the Higgs Boson in Dimuon plus Missing Transverse Energy Final States with 4.2 fb^{-1} of Data*, DØ Note 5932 (2007).

[68] V. Abazov *et al.* (DØ Collaboration), Phys. Lett. B. **658**, 112 (2008).

[69] T. Andeen *et al.*, FERMILAB-TM-2365 (2007).

[70] T. Nunnemann, *NNLO Cross-Sections for Drell-Yan, Z and W Production using Modern Parton Distribution Functions*, DØ Note 4476 (2004).

[71] T. Junk, Nucl. Instrum. Methods Phys. Res. A **434**, 435 (1999); A. Read, J. Phys. G **28** 2693 (2002).

[72] W. Fisher, FERMILAB-TM-2386-E.