Search for First Generation Leptoquarks in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV in the Dielectron + Dijet Channel using the DØ Detector at Fermilab

by

Shaohua Fu

Advisor: Philip Michael Tuts

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2004

ABSTRACT

Search for First Generation Leptoquarks in $p\bar{p}$ Collisions

at $\sqrt{s} = 1.96$ TeV in the Dielectron + Dijet Channel

using the DØ Detector at Fermilab

Shaohua Fu

We describe a search for first generation leptoquarks decaying into the $eejj$ final state in $p\bar{p}$ collisions at a center of mass energy of 1.96 TeV using the DØ detector at the Fermilab Tevatron. This search is based on data collected during 2002–2003 with an integrated luminosity of (130.4 ± 8.5) pb^{-1}. Leptoquarks are assumed to be produced in pairs and to decay into an electron and a quark with a branching ratio β. We observe no evidence for leptoquarks, and set an upper cross section limit of 0.086 pb at the 95% confidence level corresponding to a lower mass limit of 231 GeV/c^2 for scalar leptoquarks when $\beta = 1$.

Search for First Generation Leptoquarks in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV in the Dielectron + Dijet Channel using the DØ Detector at Fermilab

by

Shaohua Fu

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2004

ABSTRACT

Search for First Generation Leptoquarks in $p\bar{p}$ Collisions

at $\sqrt{s} = 1.96$ TeV in the Dielectron + Dijet Channel

using the DØ Detector at Fermilab

Shaohua Fu

We describe a search for first generation leptoquarks decaying into the $eejj$ final state in $p\bar{p}$ collisions at a center of mass energy of 1.96 TeV using the DØ detector at the Fermilab Tevatron. This search is based on data collected during 2002–2003 with an integrated luminosity of (130.4 ± 8.5) pb^{-1}. Leptoquarks are assumed to be produced in pairs and to decay into an electron and a quark with a branching ratio β. We observe no evidence for leptoquarks, and set an upper cross section limit of 0.086 pb at the 95% confidence level corresponding to a lower mass limit of 231 GeV/c^2 for scalar leptoquarks when $\beta = 1$.

Contents

List of Figures **vii**

List of Tables **xiv**

Acknowledgements **xviii**

Dedication **xx**

1 Introduction **1**

1.1 The Standard Model 3

1.1.1 Elementary Particles and Forces 3

1.1.2 Electroweak Interactions 5

1.1.3 Strong Interactions 9

1.2 Beyond the Standard Model 12

1.2.1 Grand Unified Theories 14

1.2.2 Strong Dynamics 15

1.2.3 Supersymmetry 15

1.2.4 String Theory 17

1.2.5 Summary 17

2 Leptoquark Phenomenology **18**

2.1 Overview 18

2.2 Leptoquarks in Theoretical Models 19

2.3 Leptoquark Pair Production 20

2.4 Signatures for this Search 23

2.5 Existing Limits 26

3 Experimental Apparatus **29**

3.1 The Tevatron Collider 29

3.2 The DØ Detector 33

3.3 The DØ Coordinate System 35

3.4 Central Tracking System 37

3.4.1 Silicon Microstrip Tracker 39

3.4.2 Central Fiber Tracker 41

3.4.3 Solenoid 43

3.5 Calorimeter System 43

3.5.1 Preshower Detectors 44

3.5.2 Inter-cryostat Detector 48
3.5.3 Liquid Argon Calorimeter 49
3.6 Muon System 59
3.7 Trigger System 60
3.7.1 Luminosity Monitor 60
3.7.2 Level 1 Trigger 62
3.7.3 Level 2 Trigger 63
3.7.4 Level 3 Trigger 64
3.7.5 Data Acquisition 64

4 Event Reconstruction **66**

4.1 Vertex Reconstruction 67
4.1.1 Primary Vertex 68
4.1.2 Secondary Vertex 69
4.2 Electron Reconstruction 70
4.2.1 EM Cluster Reconstruction 71
4.2.2 EM Identification 72
4.2.3 Track Matching 75
4.3 Jet Reconstruction 76
4.3.1 Cone Jet 76
4.3.2 Jet Identification 77

4.3.3 Jet Energy Correction 78
4.4 Missing E_T ($\not\!\!E_T$) Measurement 82
4.5 Muon Reconstruction 83
4.6 Monte Carlo Simulation 84
4.6.1 Event Generation 84
4.6.2 Detector Simulation 85

5 Data Analysis **87**

5.1 Data and Event Selection 87
5.1.1 Data Sets 88
5.1.2 Trigger Selection 88
5.1.3 Offline Selection 90
5.2 Background 94
5.2.1 QCD Background 95
5.2.2 Z/Drell-Yan Background 101
5.2.3 Top Background 106
5.3 Signal Monte Carlo 108
5.4 Observed Data vs. Expected Background 109
5.5 Optimization 110
5.6 Signal Efficiency 115
5.6.1 Acceptance 116

5.6.2 Efficiency 118

5.7 Systematic Uncertainties 133

5.7.1 Systematic Uncertainties on Signal Efficiency 133

5.7.2 Systematic Uncertainties on Background 138

6 Results **143**

6.1 Data and Background Comparison 143

6.2 Extracting a Limit 146

6.2.1 Bayesian Technique 146

6.2.2 Limit 152

6.3 Combined Results 154

6.3.1 Combining Run I Results 155

6.3.2 Combining $eejj$ and $e\nu jj$ Channels 162

6.4 Conclusion 166

Appendix **168**

A.1 More Discussion of the Systematic Uncertainties of the QCD Background 168

A.1.1 Systematic Uncertainty due to γ-jet Event Contamination 168

A.1.2 Systematic Uncertainty due to Trigger Selection 172

A.1.3 Systematic Uncertainty due to the Estimate Algorithm . . 175

A.2 Run II Prospect 178

Bibliography **179**

List of Figures

2.1 *Leading order Feynman diagrams for leptoquark pair production in hadron collisions for: (a) quark-antiquark annihilation and (b)–(e) gluon-gluon fusion.* 21

2.2 *Examples of next-to-leading order Feynman diagrams for leptoquark pair production: (a)–(c) gluon-quark subprocess, (d) a gluon radiated by an incoming quark in a process known as gluon bremsstrahlung.* 22

2.3 *The cross section for the production of scalar leptoquark pairs at the Tevatron energy* $\sqrt{s} = 1.96$ *TeV, as a function of the leptoquark mass* M_{LQ}*. The NLO result is compared with LO calculations. The variation of the NLO cross section with the value of the renormalization/factorization scale* μ *is indicated by the band.* . . 24

3.1 *Fermilab Tevatron Collider complex.* 31

3.2 *A side view of the upgraded DØ detector. From inside to outside are the inner tracking system, calorimeter, and muon system.* . . . 34

3.3 *The θ corresponding to some representative values of η.* 37

3.4 *A r–z view of the central tracking system. The trackers (not including the solenoid) cover a region of 52 cm in r and 252 cm in z.* 38

3.5 *The silicon microstrip tracker.* 40

3.6 *A side view* (left) *of the central fiber tracker with cross-sectional end view* (right) *of the fibers.* 42

3.7 *A cross-sectional end view* (left) *and side view* (right) *of the central preshower detector.* 45

3.8 *One quarter view of the forward preshower detector.* 47

3.9 *(a) Schematic of the ICD super-tile array hung on the inner surface of each end-cap calorimeter cryostat. (b) Schematic of the scintillator tiles within one ICD super-tile module.* 49

3.10 *The DØ calorimeter.* 52

3.11 *Schematic view of a calorimeter cell.* 53

3.12 *Side view of one quadrant of the calorimeter, with pseudorapidity lines plotted. The shading pattern indicates distinct calorimeter cells.* 54

3.13 *Readout chain of the calorimeter in Run II.* 57

3.14 *Design of the DØ trigger system.* 61

4.1 *Primary vertex and secondary vertex. Tracks from the secondary vertex have a large impact parameter with respect to the primary vertex.* 70

4.2 *EM ID variable distributions for electrons from* $Z \rightarrow e^+e^-$ *decays* (solid line) *compared with jets from multijet events which are reconstructed as (fake) EM objects* (dashed line)*: (a) EM-fraction, (b) isolation, (c) H-Matrix* (8×8) χ^2*.* 74

4.3 *Jet ID variable distributions for jets from dijet events: (a) EM-fraction, (b) Coarse-Hadronic-fraction, (c)* $n90$*, (d) Hot-fraction.* . 79

4.4 *Jet ID variable distributions for jets from dijet events: (a) nitm, (b)* $f90 + 0.5 \times CHF$*.* 80

5.1 *The di-EM invariant mass distribution from data* (triangles) *and QCD background* (dashed line) *for different event topologies: (a) both EM objects in the CC, (b) one EM object in the CC and the other in the EC, (c) both EM objects in the EC.* 93

5.2 *Fake rate as a function of EM object* E_T *for passing EM ID cuts in CC* (top left) *and EC* (top right)*, and for passing EM ID cuts with track match in CC* (bottom left) *and EC* (bottom right)*.* . . 99

5.3 *The di-EM invariant mass distribution from data* (triangles) *compared with background* (solid line) *of Z/Drell-Yan Monte Carlo plus QCD fake.* 104

5.4 *Number of 0.5 cone jets associated with a Z in data* (triangles) *and Monte Carlo* (shaded area). *QCD background has been subtracted from data. The uncertainties are also shown. The PYTHIA Monte Carlo fits an exponential slope of* -2.36 ± 0.10 (dashed line), *while the data fit an exponential slope of* -2.10 ± 0.10 (solid line). . . . 107

5.5 *The di-EM invariant mass distribution of the eejj events for data* (triangles) *compared with background* (solid line). 111

5.6 *The* S_T *distribution of the eejj events for data* (triangles), *Standard Model background* (solid line), *and leptoquark signal at a mass of 220 GeV* (dash-dot line). *The Z-veto cut has been applied.* 113

5.7 *The limit setting significance as a function of the* S_T *cut. The maximum limit setting significance corresponds to the optimized value of* S_T *cut* ($S_T > 375$ *GeV).* 116

5.8 *Trigger efficiency as a function of electron* E_T *for the CC* (left) *and EC* (right) *regions.* 120

5.9 *The invariant mass distributions for 2(tt)+(tp)* (left) *and 2(tt)+(tp)+(tf)* (right) *samples with the CC-CC topology. The points are data. The line is the expected QCD background.* 124

5.10 *The invariant mass distributions for 2(tt)+(tp)* (left) *and 2(tt)+(tp)+(tf)* (right) *samples with the EC-EC topology. The points are data. The line is the expected QCD background.* 124

5.11 *The invariant mass distributions for* $N_1 + 2N_2$ (left) *and* $N_0 + N_1 + N_2$ (right) *samples with the CC-CC topology. The points are data. The line is the expected QCD background.* 127

5.12 *The invariant mass distributions for* $N_1 + 2N_2$ (left) *and* $N_0 + N_1 + N_2$ (right) *samples with the EC-EC topology. The points are data. The line is the expected QCD background.* 127

5.13 *The jet ID efficiency with respect to jet* E_T (left) *and to jet* η_{det} (right)*, for jets in loose single-EM data sample.* 130

5.14 *The jet ID efficiency with respect to jet* E_T (left) *and to jet* η_{det} (right)*, for jets in QCD dijet data sample. The dashed line fits the efficiency with a slope of -0.00084* GeV^{-1}. 130

5.15 *The jet ID efficiency with respect to jet* E_T (left) *and to jet* η_{det} (right)*, for jets in QCD dijet Monte Carlo sample.* 131

6.1 *Comparison of eejj data to expected background for the leading EM object E_T* (left) *and the second EM object E_T* (right)*, where the Z-veto cut has been applied.* 145

6.2 *Comparison of eejj data to expected background for the leading jet E_T* (left) *and the second jet E_T* (right)*, where the Z-veto cut has been applied.* 145

6.3 *Comparison of eejj data to expected background for M_{ee}* (left) *and M_{ej}* (right)*, where the Z-veto cut has been applied. The hole in the M_{ee} distribution comes from the Z mass cut.* 147

6.4 *The S_T distribution of the eejj events for data* (triangles)*, Standard Model background* (solid line)*, and leptoquark signal at a mass of 220 GeV* (dash-dot line)*. The Z-veto cut has been applied.* 148

6.5 *The 95% confidence level upper limit on the cross section times decay fraction as a function of LQ mass* (open circles)*. The NLO theoretical cross sections are plotted for $\mu = 1, 0.5, 2 \times M_{LQ}$. A lower mass limit of 231 GeV for first generation scalar leptoquarks is achieved for $\beta = 1$.* 153

6.6 *The lower limit on the leptoquark mass at the 95% confidence level as a function of β. For $\beta = 0.5$, the mass limit on the first generation scalar leptoquark is 169 GeV.* 154

6.7 *The posterior probability $\rho(\sigma|k=0,I)$ for leptoquark mass = 220 GeV at Run II energy with respect to leptoquark cross section. The curved line shows a fit to the exponential function. The arrow indicates the cross section upper limit at 95% confidence level for this mass point.* 159

6.8 *The 95% confidence level upper limits on the cross section times decay fraction from DØ Run II only* (open circles)*, DØ Run II combined with DØ Run I* (triangles)*, and DØ Run II combined with DØ + CDF Run I* (inverted triangles) *leptoquark analyses. The NLO theoretical cross sections are plotted for $\mu = 1, 0.5, 2 \times M_{LQ}$.* 161

6.9 *The 95% confidence level lower limit on the mass of first generation scalar leptoquarks as a function of β.* 165

A.1 *Fake rate as a function of EM object E_T, for the loose single-EM samples passing different triggers.* 174

List of Tables

1.1 *Quarks and leptons of the Standard Model.* 4

1.2 *The fundamental forces.* 5

2.1 *DØ Run I mass limits for first generation scalar and vector leptoquarks, for three different values of the branching ratio β in all channels.* 28

5.1 *Number of events passing the selection criteria.* 94

5.2 *Z/Drell-Yan $\rightarrow$ ee Monte Carlo samples in different mass ranges.* 102

5.3 *Number of di-EM events from data and background (Z/Drell-Yan Monte Carlo and QCD fake). The error on the Z/Drell-Yan background is due to uncertainties on the EM ID efficiency and tracking efficiency. The error on the QCD background is mainly due to uncertainties on fake rates. The error coming from the luminosity is not included.* 104

5.4 *Number of events with $\geq n$ jets within Z mass range for data and QCD fake background. The error on number of events includes statistical uncertainties and systematic uncertainties (see Table 5.15).* 107

5.5 *Number of events with $\geq n$ jets within Z mass range for data and Z/Drell-Yan Monte Carlo samples. The error on number of events includes statistical uncertainties and systematic uncertainties (see Table 5.15). The errors of exponential slopes are fitting errors.* . . 108

5.6 *Number of events, PYTHIA (LO) cross section, and NLO cross section for scalar leptoquark Monte Carlo samples.* 109

5.7 *Number of eejj events from data and background (Z/Drell-Yan, QCD fake, and top), before applying Z-veto cut and after Z-veto cut.* 111

5.8 *Kinematic and geometric acceptance for leptoquark signals.* 117

5.9 *Efficiencies for EM object and jet.* 131

5.10 *Acceptance and overall efficiency of leptoquark signals after all the cuts (including Z-veto and S_T cuts).* 133

5.11 *Systematic uncertainties for leptoquark signals due to the choice of parton distribution function (PDF), and the jet energy scale (JES) uncertainties before and after the $S_T > 375$ GeV cut.* 137

5.12 *Signal efficiency uncertainties on leptoquark samples passing all the cuts, including the Z-veto and $S_T > 375$ GeV cuts.* 137

5.13 *Fake rate uncertainties.* 138

5.14 *QCD background uncertainties for di-EM sample, di-EM + dijet sample before and after the $S_T > 375$ GeV cut. "N/A"means the source of error is not appliable.* 139

5.15 *Background uncertainties due to jet energy scale and EM ID efficiency, for QCD background, Z/Drell-Yan PYTHIA Monte Carlo, and Z/Drell-Yan + 2jets ALPGEN Monte Carlo. "N/A" means the source of error is not appliable.* 140

5.16 *Summary of systematic uncertainties in background samples passing all selection cuts.* 141

6.1 *Number of eejj events from data and background (Z/Drell-Yan, QCD fake, and top), before applying the S_T cut and after the S_T cut.* 147

6.2 *Experimental upper limit of cross section × decay fraction for each leptoquark mass at the 95% confidence level. The decay fraction of a leptoquark pair decaying into eejj is β^2.* 152

6.3 *Results from the DØ Run II and DØ Run I experiments.* 155

6.4 *The 95% confidence level limits on the leptoquark cross section and the leptoquark mass as the combined results from DØ Run II and Run I leptoquark analyses.* 162

6.5 *The overall efficiency for leptoquark analyses in the $eejj$ and $e\nu jj$ channels as a function of the leptoquark mass.* 163

6.6 *The 95% confidence level lower limits on the first generation scalar leptoquark mass, for various values of β.* 166

A.1 *The cross section ratios and the fractions of direct photon events in loose single-EM sample, loose di-EM sample, and loose di-EM + dijet sample.* 170

Acknowledgements

In a collaboration such as the DØ experiment, there is a large number of people who contribute to a work such as this. I would like to thank all the people whose efforts have made this work possible. Even those whom I know contributed to this thesis are too numerous to name individually. My apology to those who are not named in these paragraphs.

First of all, I want to thank my advisor, Mike Tuts. Without his tireless guidance, this work could not have been finished. And I can not express enough my gratitude for his tremendous help and constant support throughout these many years.

I would like to thank all the members of the Columbia DØ group, who provided advice on my analysis. I thank Leslie Groer and Chris Hays who helped me to learn about the experiment when I first came to DØ. I would like to thank the members of the calorimeter group whom I worked with for the past four years. I especially want to thank Dean Schamberger for having taught me so much about the detector.

I would like to thank many people from the New Phenomena physics group, whose valuable suggestions directly helped to accomplish this work. I want to thank Sharon Hagopian who helped me to choose this analysis topic, and Greg

Landsberg who provided advice and analysis tools. I also would like to thank all the conveners and the editorial board members for reviewing the analysis, giving comments and criticism.

I am very grateful to Vishnu Zutshi for helping me through the difficulties of this analysis, for discussions and suggestions whenever I needed them. I am also indebted to Alex Melnitchouk and Junjie Zhu for sharing data samples and program codes.

Finally, I would like to thank all my Chinese friends here who organized the weekly CGS meeting, where I have gained knowledge and friendship. And I feel very furtunate to have my husband as a close collaborator during the past years.

To my parents and my husband.

Chapter 1

Introduction

Particle physics is the branch of the physical sciences that aims to understand the elementary constituents of the universe and the laws which govern their interactions.

The philosophy that the world was constructed out of small indivisible units of matter began with Democritus, a Greek philosopher in the 5th century BC. Later, people learned from chemistry that the world is made of molecules and the molecule is made of atoms. In the early 1900's, Rutherford found that the atom is composed of the nucleus and the electrons surrounding the nucleus. Then with the discovery of protons and neutrons, it became clear that the nucleus is made up of neutrons and protons. In the last four decades we have learned that neutrons and protons are not elementary particles, but are made up of *quarks*. Electrons

are considered to belong to the family of *leptons*. Quarks and leptons are spin-$\frac{1}{2}$ particles and are considered to be fundamental.

The elementary constituents that make up the universe interact with each other via forces. There are four kinds of forces: the gravitational, electromagnetic, weak, and strong forces. The gravitational force dominates the interactions at astronomical masses and distances. For the smallest objects like quarks and leptons, the gravitational force is so weak compared to the other forces that it can be neglected. The long range electromagnetic force is responsible for the force that exists between any charged objects, such as the attractive force between electrons and nuclei in atoms. The short range strong interaction acts between colored particles (like quarks and gluons) and binds quarks into protons. The very short range weak interaction is a force between leptons (such as electrons and neutrinos) and between quarks. The weak interaction is responsible for the radioactive β-decay of nuclei.

High energy physics gets its name from the high energies that are needed to probe the nature of the strong and electroweak forces and to create fundamental particles. The usual method to achieve these high energies is to accelerate particles, such as protons and electrons, and collide them at high energies. By measuring the particles produced in the debris of the collision we gain an understanding of the fundamental forces.

1.1 The Standard Model

In the 1960's, physicists developed a model to explain the fundamental building blocks of nature and the forces between them. This model, called the Standard Model (SM), is generally accepted as the description of the fundamental particles and their interactions.

1.1.1 Elementary Particles and Forces

In the framework of the Standard Model [1], the world is made up of elementary, structureless, point-like particles with spin-$\frac{1}{2}$. They are called fermions and obey Fermi-Dirac statistics. The fermions are further divided into families of quarks and leptons. As listed in Table 1.1, there are six quarks: up (u), down (d), charm (c), strange (s), top (t), and bottom (b); and six leptons: electron (e), electron neutrino (ν_e), muon (μ), muon neutrino (ν_μ), tau (τ), and tau neutrino (ν_τ). Quarks have fractional charge: $\frac{2}{3}e$ or $-\frac{1}{3}e$, where $-e$ is the charge of the electron. Leptons have integral charges: $-e$ or 0. The leptons and quarks are grouped into three generations. For example, u and d quarks, as well as e and ν_e belong to the first generation, and they make up the bulk of the "everyday" matter.

Every elementary particle has a corresponding antiparticle, which has the same mass and spin but opposite charge. The six quarks combine in particle-antiparticle pairs (mesons), or in particle or antiparticle triplets (baryons), to form all of the

Fermion	Quark		Lepton	
Charges	$\frac{2}{3}e$	$-\frac{1}{3}e$	0	$-e$
1st Generation	u	d	ν_e	e
2nd Generation	c	s	ν_μ	μ
3rd Generation	t	b	ν_τ	τ

Table 1.1: *Quarks and leptons of the Standard Model.*

known hadrons. For example, the familiar protons and neutrons, which comprise the atomic nucleus, are uud and udd bound states respectively.

The fermions interact with each other via the three forces contained in the Standard Model: electromagnetic, strong, and weak forces. The fourth physical force, gravity, is not contained in the framework of the Standard Model. The interactions are described in terms of the exchange of fundamental spin-1 bosons which obey Bose-Einstein statistics. These bosons, i.e. the quanta of the force fields, are called: the photon (γ), which mediates the electromagnetic force; the gluons (g), which mediate the strong force; and the $W^{\pm}$ and Z^0 bosons, which mediate the weak force.

The various particles in the Standard Model are sensitive to specific forces. The charged leptons undergo electromagnetic and weak interactions, while the neutrinos are only sensitive to the weak interactions. Quarks, in addition to weak and electromagnetic interactions, are subject to strong interactions. All of the forces, ordered by their relative strength, are summerized in Table 1.2 (including

gravity which is mediated by a spin-2 boson called the graviton).

Interaction	Strong	Electromagnetic	Weak	Gravitational
Relative Strength	1	1/137	10^{-5}	10^{-38}
Field Quanta	8 g	γ	$W^{\pm}$ / Z^0	G
Mass (GeV/c^2)	0	0	80.22 / 91.17	0
Range	10^{-15} m	∞	10^{-18} m	∞

Table 1.2: *The fundamental forces.*

Mathematically, the Standard Model is a gauge field theory based on the group $SU(3)_C \times SU(2)_L \times U(1)_Y$, where C refers to color, L refers to weak isospin, and Y is weak hypercharge, which will be discussed in more detail in the following.

1.1.2 Electroweak Interactions

The $U(1)$ symmetry corresponds to Quantum Electrodynamics (QED) [2], which was proposed by P. A. M. Dirac in 1927. In QED, a free electron is described by a four-component wave function which corresponds to two spin substates, $J_z = \pm\frac{1}{2}\hbar$, each with positive and negative energies. The negative energy state is interpreted as an antiparticle, the positron. The positron was discovered by Carl Anderson in 1932, validating the basic ideas of QED. One very important property of QED is *renormalizability*. Because a single electron can emit and reabsorb virtual photons, theoretical calculations of the electron "bare" mass, m_0, or the charge, e_0, lead to values for those quantities that are infinite. Divergent

terms of this type are present in all QED calculations. However, it is possible to absorb all the divergences into m_0 or e_0, and then redefine the mass and charge, replacing them by their physical values m, e. This process is called renormalization. The result of this renormalization process is that QED calculations, if expressed in terms of the physical quantities e and m, always give finite values for physical measurables like cross-sections, decay rates, etc. The generator of the $U(1)$ group [1] is the weak hypercharge Y, which is related to electric charge Q and the third component of isospin I_3 by $Q = I_3 + \frac{1}{2}Y$.

The generators of the $SU(2)$ group are the three components of the weak isospin. In the fundamental representation, the three generators are denoted as Pauli matrices:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{1.1}$$

with $i \equiv \sqrt{-1}$. Although the weak and electromagnetic forces appear to be quite different at low energy scales, i.e. $q^2 \ll M_W^2$, where q is the momentum transfer of the interaction and M_W is about 100 GeV/c^2, they are unified with the same intrinsic coupling strength at a higher energy scale, i.e. $q^2 \sim M_W^2$, based on the

[1] The generators of a group constitute the elements of the group and describe the symmetry. For example, taking the $U(1)$ group as rotations about the 3- (or z) axis, its generator is the third-component of the angular momentum, J_3. The element of $U(1)$ group can be written as $U(\theta) = e^{-i\theta J_3}$.

Glashow-Salam-Weinberg $SU(2) \times U(1)$ model [3] — the electroweak interactions model. The fundamental vector bosons are massless isotriplets $\mathbf{W}_\mu = W_\mu^i (i = 1, 2, 3)$ for $SU(2)$ and a massless isosinglet B_μ for $U(1)$. Spontaneous Symmetry Breaking (SSB) is the mechanism that gives rise to the gauge boson mass, without destroying renormalizability. This is achieved with the help of an isospin doublet of scalar particles called Higgs scalars, which generate mass as a result of self-interaction. Because of spontaneous symmetry breaking, three bosons (W_μ^+, W_μ^-, and Z_μ^0) acquire masses, and one (A_μ, the photon) remains massless. The Higgs particle is the last remaining unobserved particle in the Standard Model.

The interaction of the fermions with the fields $\mathbf{W}_\mu$ and B_μ is described by a Lagrangian density $\mathcal{L}$ as the product of the fermion currents with the fields:

$$\mathcal{L} = g\mathbf{J}_\mu \cdot \mathbf{W}_\mu + g' J_\mu^Y B_\mu \tag{1.2}$$

where $\mathbf{J}_\mu$ and J_μ^Y represent the isospin and hypercharge currents of the fermions (leptons and quarks) respectively, and g and g' are their couplings to $\mathbf{W}_\mu$ and B_μ. The relation between g, g', and the electromagnetic coupling e is:

$$e = g \sin\theta_W = g' \cos\theta_W \tag{1.3}$$

where θ_W is known as the weak mixing angle or Weinberg angle ($\sin^2\theta_W \approx 0.23$).

We know that the weak charged-current interaction is parity-violating, and connects, for example, the left-handed states of the electron and the neutrino. On the other hand, the electromagnetic interaction is parity-conserving and involves both left-handed and right-handed states of the electron. Hence, the lepton states are assigned to a left-handed doublet and a right-handed singlet:

$$\psi_L = \frac{(1+\gamma_5)}{2}\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \text{ with } T=\frac{1}{2}, Y=-1. \tag{1.4}$$

$$\psi_R = \frac{(1-\gamma_5)}{2}(e^-), \qquad \text{with } T=0, Y=-2. \tag{1.5}$$

For quarks,

$$\psi_L = \frac{(1+\gamma_5)}{2}\begin{pmatrix} u \\ d \end{pmatrix} \tag{1.6}$$

$$\psi_R = u_R \text{ or } d_R \tag{1.7}$$

where T and Y are the generators of the $SU(2)_L$ and $U(1)_Y$ groups of gauge transformations respectively, and γ_5 is a Dirac matrix:

$$\gamma_5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

In the limit of low energy scales, i.e. $q^2 \ll M_W^2$, the masses of the bosons can be written as

$$M_{W^\pm} = \left(\frac{\sqrt{2}g^2}{8G_F}\right)^{1/2} = \left(\frac{\sqrt{2}e^2}{8G_F \sin^2\theta_W}\right)^{1/2} \tag{1.8}$$

$$M_{Z^0} = \frac{M_W^2}{\rho \cos^2\theta_W} \tag{1.9}$$

$$M_\gamma = 0 \tag{1.10}$$

where G_F is the Fermi constant ($G_F/(\hbar c)^3 = 1.166 \times 10^{-5}$ GeV^{-2}), and ρ is a factor which is observed to be 1 by all experiments to date.

1.1.3 Strong Interactions

The strong interaction of quarks can be mathematically represented by a local gauge invariant $SU(3)_C$ color symmetry [4]. The field theory of quark-quark interactions is known as Quantum Chromodynamics (QCD). Quarks come in three "colors": red, green, and blue, denoted symbolically by R, G, and B, respectively. The antiquarks are assigned the complementary colors: $\overline{R}$, $\overline{G}$ and $\overline{B}$. All particles observed in nature are "colorless" or "white" (or, to be more precise, have wave functions that remain unchanged by rotations in R–G–B space). Baryons are bound states of three quarks with different colors. Mesons are bound states of a quark and an antiquark with color and complementary color respectively, and

hence "colorless" as well.

The generators of the $SU(3)_C$ group may be taken to be any $3 \times 3 - 1 = 8$ linearly independent traceless hermitian 3×3 matrices. The fundamental representation of $SU(3)_C$ is a triplet. The three color charges of a quark, R, G, and B, form the fundamental representation of the $SU(3)$ symmetry group. In this representation, the generators are traditionally denoted λ_i, $i = 1, 2, ..., 8$, known as the Gell-Mann matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \ \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \ \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \ \lambda_8 = \sqrt{\frac{1}{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Among them λ_3 and λ_8 are the diagonal matrices, with simultaneous eigenvectors

$$R = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \ , \ G = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \ , \ B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{1.11}$$

In the QCD theory, the strong color field is mediated by massless vector gluons, which come in eight different color combinations:

$$R\overline{G} \ , \ R\overline{B} \ , \ G\overline{R} \ , \ G\overline{B} \ , \ B\overline{R} \ , \ B\overline{G} \ , \ \sqrt{\frac{1}{2}}(R\overline{R}-G\overline{G}) \ , \ \sqrt{\frac{1}{6}}(R\overline{R}+G\overline{G}-2B\overline{B}) \tag{1.12}$$

The remaining combination, the $SU(3)$ color singlet,

$$\sqrt{\frac{1}{3}}(R\overline{R}+G\overline{G}+B\overline{B}) \tag{1.13}$$

does not carry color and cannot mediate the interaction between color charges.

The strong coupling parameter α_s is not a constant, instead, it changes with the energy scale. For violent collisions with very high q^2 (the momentum transferred), the strong coupling $\alpha_s < 1$. Thus perturbation theory may be used to calculate short-distance (i.e. high-energy) interactions. Hence, the theory is called "asymptotically free", which means that at very short distances, e.g. inside the radius of the proton, the quarks behave as if they were free particles. At low q^2

(or equivalently, larger distances), the coupling constant α_s becomes large and the theory is thus non-perturbative. However lattice gauge calculations can provide answers in that regime. This large-distance behavior is responsible for the confinement of quarks and gluons inside hadrons.

1.2 Beyond the Standard Model

Although the Standard Model of particle physics has been tested repeatedly and found to be in good agreement with the experimental evidence (aside from a few two or three standard deviation effects), there are reasons to believe that the Standard Model is not the ultimate theory of fundamental particles and their interactions. The Standard Model has an *ad hoc* choice of the gauge groups and particle multiplets. It contains no explanation for the apparent symmetry between the quark and lepton sectors (i.e. the similarity between the family and generational structure seen in Table 1.1). And there is no reason within the Standard Model framework for the number of generations.

One of the unsatisfactory features of the Standard Model is known as the *hierarchy* problem. There are huge differences in the energy scales of the various interactions: the QCD scale is of order 1 GeV/c^2 ($\sim M_{proton}$); the electroweak scale is of order 100 GeV/c^2 ($\sim M_{W^{\pm},Z^0}$); the scale of grand unification (GUT) is around

10^{16} GeV/c^2; and the Planck mass scale is about 10^{19} GeV/c^2 ($M_{Planck} = \sqrt{\hbar c/G}$).

Another unsatisfactory feature of the Standard Model is known as the *fine-tuning* problem. The Standard Model needs scalar Higgs bosons to induce spontaneous breaking of the gauge symmetry and to give mass to the other particles. However, the mass of the scalar Higgs boson diverges quadratically in perturbation theory. To lowest order in perturbation theory

$$m_H^2 = m_0^2 + \delta m_H^2 \approx m_0^2 - g^2\Lambda^2 \tag{1.14}$$

where m_H is the scalar Higgs boson mass, m_0 is the bare Higgs boson mass parameter, g is a dimensionless coupling constant, and Λ is the energy scale. The mass of the Higgs boson, m_H, is predicted to be comparable to the empirically measured electroweak scale ($\sim \mathcal{O}(10^2)$ GeV/c^2) [5]. So if $g^2 \approx 1$ and Λ is as large as M_{GUT} (10^{16} GeV/c^2) or M_{Planck} (10^{19} GeV/c^2), m_0^2 must be precisely adjusted so that the two terms in Equation 1.14, each of which is in the order of 10^{30} (GeV/c^2)^2, cancel to leave $m_H \sim \mathcal{O}(10^2)$ GeV/c^2. While this is not impossible, the requirement of such a precise adjustment of the input parameters to 23 decimal places (or fine-tuning) is unnatural.

Finally there is the problem that gravity is not included in the Standard Model. As stated before, the gravitational interaction is sufficiently weak that it plays no

role in fundamental particle interactions. Nonetheless, a theory which does not include one of the four forces of physics is probably not a fundamental theory. While there is no empirical evidence for it, most physicists believe that the four forces are just different aspects of a single, unified force that would be apparent at sufficiently high energy scale (e.g., M_{GUT} or M_{Planck}, the scale available at the moment of the big bang). Within the Standard Model, this unification is not possible. Thus, many physicists believe that a new theory is required which goes beyond the Standard Model. Today, there are many theories going beyond the Standard Model. A few popular models are introduced in the following sections. To date, no conclusive experimental evidence for physics beyond the Standard Model has been observed.

1.2.1 Grand Unified Theories

Grand Unified Theories (GUTs) [6] propose a single interaction to describe electromagnetic, weak, and strong interactions with a unique intrinsic coupling at the unification energy $M_{GUT} \approx 10^{16}$ GeV/c^2, and appeal to further symmetry-breaking processes to account for their different apparent strengths in the energy domain well below the unification scale.

The simplest of the GUT theories is based on the $SU(5)$ symmetry group which incorporates the $SU(3)$, $SU(2)$, and $U(1)$ symmetry groups within it. However,

the $SU(5)$ group has been ruled out because it predicts that protons will decay with a half-life which has been excluded by experiment. Other popular examples of GUTs include the $SO(10)$ and E_6 groups.

1.2.2 Strong Dynamics

In the Standard Model the electroweak symmetry breaking is introduced with the help of Higgs bosons. An alternative view puts the origin of the symmetry breaking in a different sector of the theory, one with new fundamental fermions that have new gauge interactions. In this approach elementary scalar bosons are completely absent. Scalar and pseudoscalar bosons that are built of the new fundamental fermions can exist (in analogy with the way mesons such as pions are built of quarks). Such theories include, e.g. *Technicolor* [7], *Extended Technicolor* [8], *Walking Technicolor* [9], *Topcolor-assisted Technicolor* [10].

1.2.3 Supersymmetry

Supersymmetry Theories (SUSY) [11] postulate a relation between the bosons and the fermions in the Standard Model. Each Standard Model particle has a supersymmetric partner (a superpartner, called a "sparticle") with spin differing by $\frac{1}{2}\hbar$, but with the other quantum numbers remaining the same. Thus the Standard Model bosons have superpartner fermions and the Standard Model fermions ac-

quire superpartner bosons. The superpartners cancel the quadratic divergences in the masses of the Standard Model particles thus yielding finite calculation results without fine-tuning.

If the SUSY particles have the same masses as the Standard Model particles, we should have observed them in experiment. The fact that the SUSY particles must be much more massive than the corresponding Standard Model particles suggests that SUSY is a broken symmetry. There exist several ways of supersymmetry breaking which result in different SUSY models:

- In Minimal Supersymmetric Standard Model (MSSM) the supersymmetry breaking is simply introduced explicitly.
- Gauge Mediated Supersymmetry Breaking (GMSB) theories introduce a new gauge field whose interactions cause SUSY-breaking.
- In Supergravity (SUGRA) theories, SUSY-breaking is propagated through the gravitational interaction.

In addition, there are R-parity Violating (RPV) SUSY models in which the R-parity is not conserved. R is a new quantum number in SUSY models defined as: $R = (-1)^{(3B+L+2S)}$, where B is the baryon number, L the lepton number, and S the particle spin.

1.2.4 String Theory

String Theory [12] can incorporate gravity in a natural fashion. String theory describes all matter particles as one dimensional loops, rather than the zero dimensional points of field theory. It predicts the "graviton", the mediator of the gravitational force. It also predicts additional six or seven dimensions beyond the four observed space-time dimensions. Such extra dimensions could exist if they are curled in on themselves, instead of extending linearly in the manner of the familiar dimensions. Matter and the three non-gravitational forces are confined to our four-dimensional subspace whereas gravity may propagate throughout higher-dimensional volume. Based on this, the superstring theory is a supersymmetric version of string theory.

1.2.5 Summary

The unsatisfactory features of the Standard Model motivate new theories as well as experimental searches for new physics beyond the Standard Model. This dissertation describes a search for a new particle, the leptoquark, which will be described in detail in the next chapter.

Chapter 2

Leptoquark Phenomenology

We search for leptoquarks as a means to search for physics beyond the Standard Model. This chapter discribes the phenomenology of leptoquarks: the theories, the leptoquark production and decay, and the results of previous searches for leptoquarks.

2.1 Overview

The observed symmetry in the generational structure of quark and lepton families is a mysterious occurrence within the Standard Model, and hints that the two kinds of fermions (quarks and leptons) may be related at a more fundamental level. Almost all models beyond the Standard Model which deal with the connection of leptons and quarks predict the existance of some type of *leptoquarks*.

Leptoquarks [13]–[19] are particles that provide the link between leptons and quarks. They carry both baryon number and lepton number, couple to both leptons and quarks, and are color triplets under $SU(3)_C$. Leptoquarks carry fractional electric charge and integral spin (they are bosons). In models where baryon and lepton numbers are separately conserved, leptoquarks can be light ($<$ 1 TeV/c^2) [1] and still avoid conflict with rapid proton decay. Depending on the structure of each specific model, leptoquarks can have spin 0 or 1; electric charge $-\frac{4}{3}$, $-\frac{1}{3}$, $\frac{2}{3}$ or $\frac{5}{3}$; isospin 0, $\frac{1}{2}$ or 1; baryon number $\pm\frac{1}{3}$; and lepton number ± 1. Leptoquarks with spin $= 0$ are called scalar leptoquarks, and those with spin $= 1$ are called vector leptoquarks. Several leptoquark models are described below.

2.2 Leptoquarks in Theoretical Models

In Grand Unified Theories (GUTs) [13], a leptoquark is introduced to transform a quark into a lepton. As a result, the strong similarity between the weak lepton and quark doublet patterns occurs as a natural consequence of lepton-quark unification.

Leptoquarks can be found in the $SU(4)$ Pati-Salam unification model [14], where quarks are assumed to carry four colors: three of these are the conventional color quantum numbers, and the fourth represents the lepton number L.

[1] 1 TeV = 10^{12} electron volt, 1 GeV = 10^9 electron volt.

The unification of baryonic and leptonic matter arises by extending the gauge symmetry $SU(3)$ of the three colors to $SU(4)$ for the four colors. As the result, leptoquarks are found as exotic gauge mesons carrying both baryonic as well as leptonic quantum numbers.

In the Technicolor model [15], the problems of the Standard Model, such as the arbitary choice of the scalar sector (Higgs), the unknown decay modes of the scalars, and the ambiguity of the elementary scalar solution of the electroweak interaction are investigated through the technicolor approach. Leptoquarks are color-triplet *technipions* with baryon number $\frac{1}{3}$ and lepton number -1, which can decay into quarks and leptons.

In addition, there are other types of leptoquarks that might exist [16]. For example, leptoquarks might arise as scalar Higgs [17], as composite particles [18], or as R-parity violating supersymmetric particles [19].

2.3 Leptoquark Pair Production

At hadron colliders, such as the Fermilab Tevatron, leptoquarks (LQ) may be produced in pairs ($p\overline{p} \rightarrow LQ\overline{LQ}$) via strong interactions. Figure 2.1 shows the Feynman diagrams for the leading order (LO) processes [20]: quark-antiquark annihilation in the first diagram, and gluon-gluon fusion in the other four diagrams. Some of the next-to-leading order (NLO) processes [21] of leptoquark pair pro-

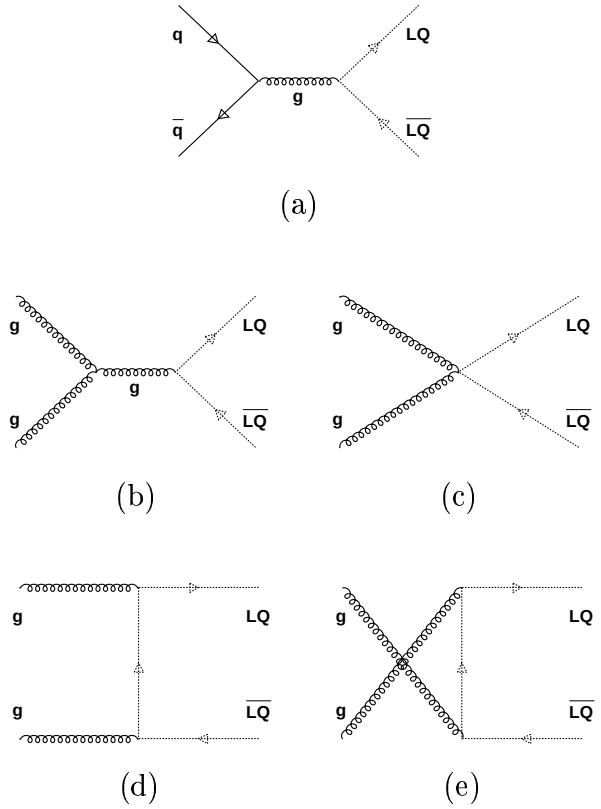


Figure 2.1: *Leading order Feynman diagrams for leptoquark pair production in hadron collisions for: (a) quark-antiquark annihilation and (b)–(e) gluon-gluon fusion.*

duction are represented by the Feynman diagrams shown in Figure 2.2, including the gluon-quark interactions and the bremsstrahlung of gluons.

Given the Feynman diagrams, we can calculate the production rate of leptoquarks. The production rate in terms of the effective cross sectional area per collision is known as the "cross section" σ, in units of *picobarns* (1 pb = 10^{-36} cm^2). The number of collisions per cross sectional area is defined as the "luminos-

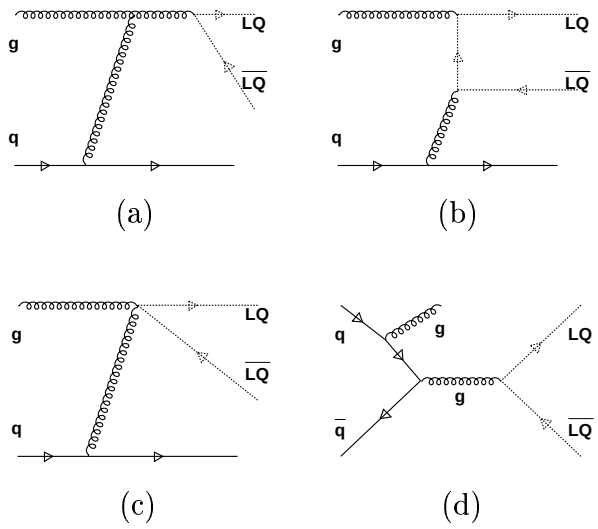


Figure 2.2: *Examples of next-to-leading order Feynman diagrams for leptoquark pair production: (a)–(c) gluon-quark subprocess, (d) a gluon radiated by an incoming quark in a process known as gluon bremsstrahlung.*

ity" L, in unit of pb^{-1}. Thus the number of leptoquarks produced at the collider is the product $L \times \sigma$.

By demanding the Lagrangian which describes the interactions of leptoquarks with the known particles to respect the $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry of the Standard Model, the couplings of the leptoquarks to the gauge bosons are widely induced. These leptoquarks are color triplets. For the scalar leptoquarks, their couplings to the gauge bosons are completely determined [22]. Only the strengths of the leptoquark-fermion couplings (i.e. leptoquark–quark–lepton couplings) remain as free parameters. The production cross section for scalar leptoquarks is independent of the LQ–q–l coupling [17]. In the case of vector leptoquarks,

the production cross sections are complicated by the anomalous leptoquark-gluon couplings and thus are model dependent. For example, there is a Minimal Vector coupling [22] and a Yang-Mills coupling [23].

Figure 2.3 shows the theoretical cross section for scalar leptoquark pair production at the Tevatron as a function of the leptoquark mass. On the plot, the leading order (LO) cross sections are compared with the ones including the next-to-leading order (NLO) processes [21]. In the cross section calculations using QCD theory, we choose the renormalization/factorization scale (μ) to be: $\mu = M_{LQ}$ for the central values of the LO and the NLO cross sections; and $\mu = 2M_{LQ}$ ($\mu = \frac{1}{2}M_{LQ}$) for the lower (higher) bound values of the NLO cross sections which are taken as the uncertainties of the theoretical calculations [2], where M_{LQ} is the leptoquark mass.

2.4 Signatures for this Search

If leptoquarks are produced in the $p\bar{p}$ collisions, they will decay before traveling a significant distance as long as their coupling strengths are not significantly weaker than the electroweak coupling strength (a reasonable assumption,

[2]The renormalization/factorization scale μ is an unknown value in the QCD perturbation theory. A conventional choice in the calculations for the leptoquark cross section is to take $\mu = M_{LQ}$ to calculate the central value based on the assumption that the most possible momentum transfer in leptoquark production is equal to the leptoquark mass; and to take the variation of μ between $\mu = 2M_{LQ}$ and $\mu = \frac{1}{2}M_{LQ}$ to obtain the theoretical uncertainties by convention.

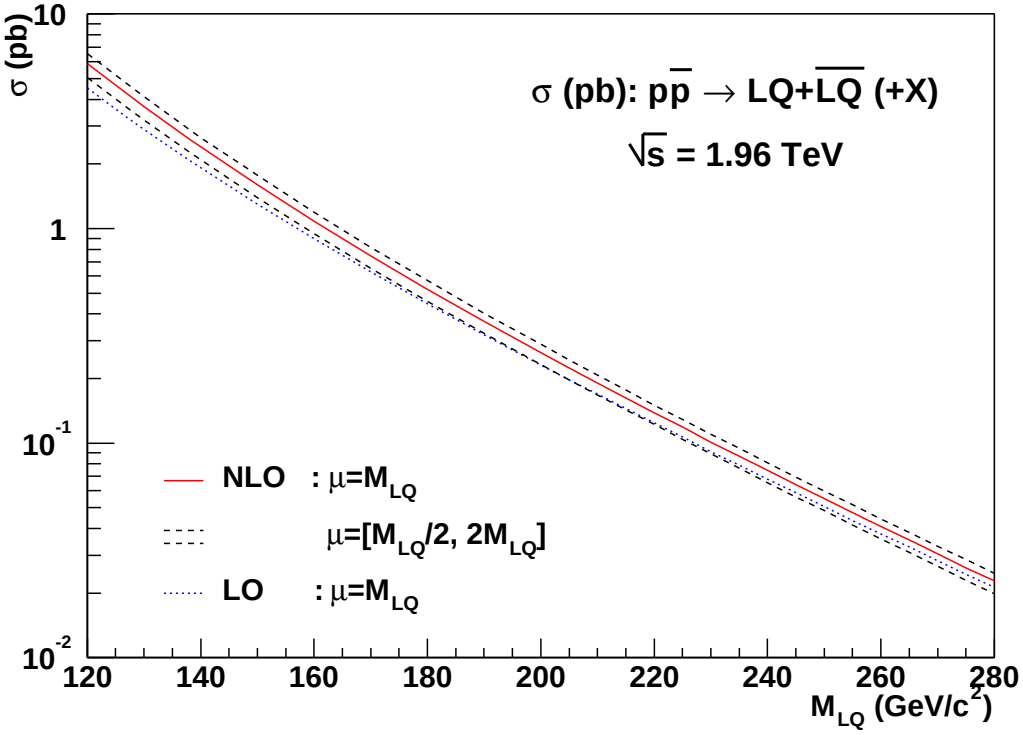


Figure 2.3: *The cross section for the production of scalar leptoquark pairs at the Tevatron energy $\sqrt{s} = 1.96$ TeV, as a function of the leptoquark mass M_{LQ}. The NLO result is compared with LO calculations. The variation of the NLO cross section with the value of the renormalization/factorization scale μ is indicated by the band.*

particularly if the leptoquarks are gauge bosons in a Unified Theory). The leptoquark search in lepton–quark collisions usually assumes the LQ–q–l coupling $\lambda_{LQ} \sim \mathcal{O}(10^{-1})$ [16].

Leptoquarks with universal couplings to all leptons and quarks will give rise to flavor-changing neutral currents (FCNC's) and are severely constrained [24]. We assume that there are three distinct generations of leptoquarks and each leptoquark couples only to the corresponding generation of quarks and leptons. Under these conditions, experimental constraints can be satisfied without requiring very large leptoquark masses, thereby making leptoquarks accessible at current energies. Thus a leptoquark decays to a lepton (a charged lepton l or a neutrino ν) and a quark (q) of the same generation.

The decays of leptoquark pairs would therefore be expected to yield one of three possible final states: $l^+l^-q\overline{q}$, $l^\pm\nu q\overline{q}$, and $\nu\overline{\nu}q\overline{q}$. If the *branching ratio* (i.e. the probability of a particle to decay into certain channel) of a leptoquark decaying into a charged lepton and a quark is defined as β, then the decay fractions of a leptoquark pair in the $l^+l^-q\overline{q}$, $l^\pm\nu q\overline{q}$, and $\nu\overline{\nu}q\overline{q}$ final states are β^2, $2\beta(1-\beta)$, and $(1-\beta)^2$ respectively.

In this work, we describe a search for the first generation scalar leptoquarks in the $e^+e^-q\overline{q}$ final state, where q can be u or d quark. This search is most sensitive to the case of the branching ratio $\beta = 1$. Electrons (positrons) can be identified

in the detector by a charged particle track and an energy deposition pattern in the calorimeter that is consistent with an electron pattern. The u and d quarks will hadronize by forming jets which can be identified by their energy deposition. Therefore the signatures for this search are 2 e + 2 jets in the detector.

2.5 Existing Limits

Since leptoquarks were first predicted, much experimental work has been done to constrain the mass of leptoquarks and the coupling constants.

Experiments at the CERN Large Electron-Positron (LEP) collider searched for leptoquarks which would be pair produced in e^+e^- collisions via a virtual γ or Z^0 in the s-channel. Based on the data of an integrated luminosity of 56 pb^{-1} collected at a center of mass energy $\sqrt{s} \simeq 183$ GeV, the OPAL collaboration has searched for vector and scalar leptoquarks with specific weak isospins and decay models [25]. For first generation scalar leptoquarks with weak isospin of 1 and the decay $LQ \to eq$ branching ratio of $\beta = 1$, OPAL has set a mass limit of 87.8 GeV/c^2. For other values of weak isospin, the mass limit ranges from 81.8 GeV/c^2 to 87.0 GeV/c^2.

The production of leptoquarks at HERA, the ep collider in DESY, Germany, is dependent on the unknown leptoquark coupling. In ep collisions, a leptoquark would be produced as an s-channel resonance via electron-quark fusion. The H1

and ZEUS groups at HERA searched for first generation leptoquarks based on data samples collected at $\sqrt{s}$ of 300–320 GeV [26]. In February 1997, the H1 and ZEUS experiments reported an excess of events at high Q^2 (momentum transfer) [27]. A possible interpretation of these events is the production of first generation leptoquarks at a mass near 200 GeV/c^2 [28]. Later on, the existance of leptoquarks of such mass was excluded by the experiments at Fermilab [29]. The observation of the excess events have not been confirmed with more recent data at HERA [30] and their statistical significance has been much reduced.

At the Fermilab Tevatron, the pair production of scalar leptoquarks is nearly independent of the leptoquark coupling constant. The DØ and CDF experiments have searched for scalar and vector leptoquarks of all three generations. Table 2.1 lists the previous mass limits for first generation leptoquarks set by the DØ experiment with Run I data (1992–1996) of an integrated luminosity about 120 pb^{-1} collected in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.8 TeV [31][32]. The CDF experiment achieved similar results in the leptoquark search during Run I [33]. For example, a mass limit of 213 GeV/c^2 was set by CDF on the first generation scalar leptoquarks at $\beta = 1$.

In this work we shall present a search for first generation scalar leptoquarks in 2 e + 2 jets final state based on the Tevatron Run II data, with a higher center-of-mass energy and an increased integrated luminosity compared to Run I. In the

$\beta =$ BR($LQ \to eq$)	Scalar LQ (GeV/c^2)	Vector LQ (GeV/c^2)	
		Minimal Coupling	Yang-Mills Coupling
1.0	225	292	345
0.5	204	282	337
0.0	98	238	298

Table 2.1: *DØ Run I mass limits for first generation scalar and vector leptoquarks, for three different values of the branching ratio β in all channels.*

next chapter, the experimental apparatus will be described.

Chapter 3

Experimental Apparatus

The data are produced by using two major instruments. First the Tevatron collider creates the $p\bar{p}$ collision at the center-of-mass energy of 1.96 TeV. Then the DØ detector measures the particles produced in the collisions.

3.1 The Tevatron Collider

The Fermi National Accelerator Laboratory (Fermilab), located in Batavia, IL, is home to the Tevatron proton-antiproton Collider, which is the highest energy particle collider in the world today. Two collider detectors, called CDF and DØ, are located along the Tevatron ring.

During the previous run of 1992–1996, referred to as Run I, the Tevatron operated at a center-of-mass energy of 1.8 TeV and delivered an integrated luminosity

of ~ 130 pb^{-1}. Both the CDF and DØ experiments produced significant physics results during Run I, including the discovery of the top quark.

The Tevatron has been upgraded since that time for Run II. The major upgrades include the construction of the Main Injector and the Antiproton Recycler, within a common tunnel. A schematic view of the upgraded Tevatron Collider complex is shown in Figure 3.1, and a detailed description can be found in [34]. The net result of these upgrades is an increase in the center-of-mass energy to 1.96 TeV and an ever increasing peak luminosity (for the work described in this thesis, the peak luminosity was 4.25×10^{31} cm^{-2}s^{-1}; during the time this thesis was written, the highest luminosity was 6.75×10^{31} cm^{-2}s^{-1}).

The starting point for the chain of accelerators is a Cockcroft-Walton accelerator. Electrons are added to hydrogen atoms to make negative hydrogen ions and the beam of H^- ions is accelerated to an energy of 750 keV. The H^- ions are then injected into a 150 m long linear accelerator called the Linac, where they are accelerated to 400 MeV. After leaving the Linac, the beam passes through a carbon foil which strips off the two electrons in the H^-, thus protons are produced. Then the protons are injected into the Booster, a 150 m diameter synchrotron, and are accelerated to an energy of 8 GeV.

Protons coming from the Booster are injected into the Main Injector. The Main Injector is a rapid cycling synchrotron with a 3 km circumference. It accelerates

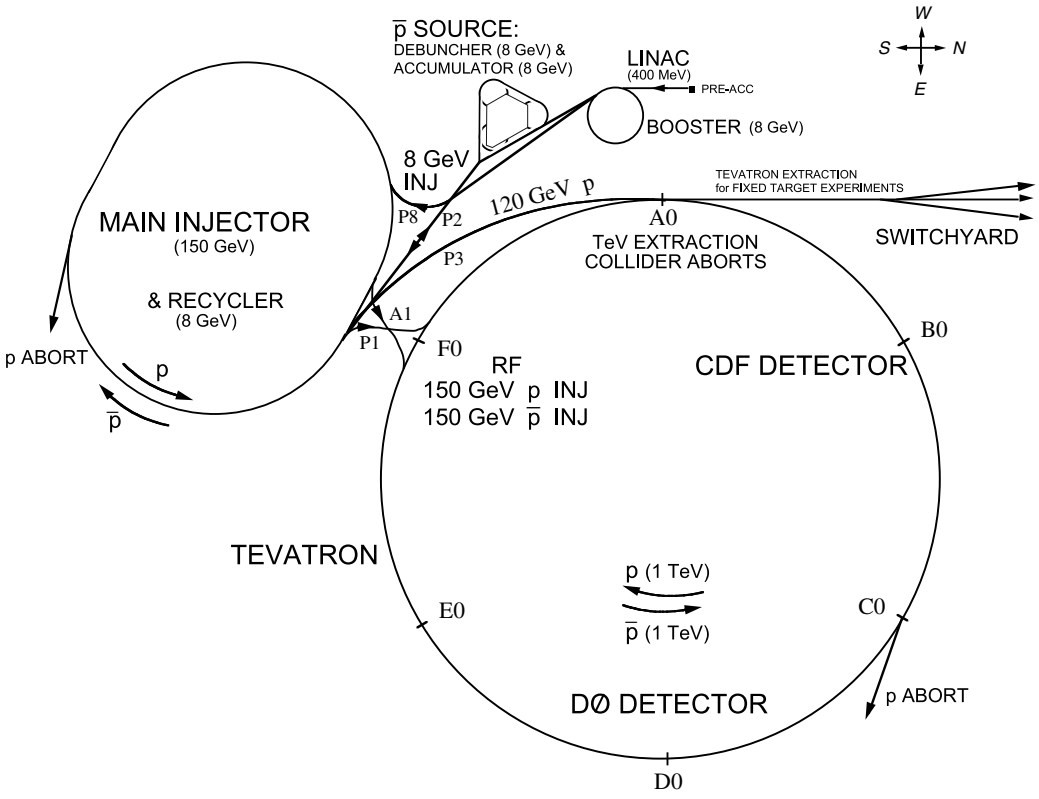


Figure 3.1: *Fermilab Tevatron Collider complex.*

the proton beams to 120 GeV, then the proton beams can be either extracted to the antiproton production target or to an external physics target. For injection of either protons or antiprotons into the Tevatron, the Main Injector accelerates the energy to 150 GeV.

Antiprotons are produced by hitting a nickel target with proton beams extracted from the Main Injector. The debris from those collisions contains antiprotons. About 10^7 antiportons are produced for every 10^{12} protons striking the

target. Immediately following the target, there is a lithium "lens", which produces an azimuthal magnetic field to focus the antiprotons. Following the lens, there is a bending magnet that selects antiprotons with energies of about 8 GeV, and transports them to the Debuncher.

The Debuncher is an 8 GeV storage ring in which the bunches of antiprotons are reduced in momentum spread and increased in longitudinal distance spread. From the Debuncher antiprotons are transferred to the Accumulator, another 8 GeV storage ring, where they are cooled (i.e. reduced in momentum spread and transverse oscillations) and accumulated before being transferred to the Recycler ring.

The 8 GeV Recycler ring is located in the same tunnel as the Main Injector, directly above the Main Injector beamline. The Recycler provides the following functionality to increase the number of antiprotons, and thus increase the luminosity:

(1) A high reliability storage ring for antiprotons. It can store 2.5×10^{12} antiprotons.

(2) A post-Accumulator ring. By emptying the antiprotons from the Accumulator into the Recycler periodically, the Accumulator can operate in its optimum intensity regime.

(3) A receptacle for antiprotons left over (about 80%) at the end of Tevatron stores. By re-use these antiprotons, the luminosity is more than doubled. This function is not commissioned yet.

The Tevatron is the final acceleration stage. It is a synchrotron with superconducting magnets, housed in a 6 km long tunnel. It accelerates the 150 GeV protons and antiprotons injected from the Main Injector to an energy of 980 GeV. The Tevatron is filled with 36 bunches of protons and 36 bunches of antiprotons, circulating in opposite directions in the same beampipe. The beams are made to collide at the desired interaction points B0 and D0, where the CDF and DØ detectors are located respectively. The time interval between collisions is 396 ns.

By June 2003 an integrated luminosity of about 220 pb^{-1} had been delivered. The current goal of the Tevatron is to deliver an integrated luminosity between 4.4 fb^{-1} and 8.6 fb^{-1} by 2009.

3.2 The DØ Detector

The DØ detector is a general-purpose detector aimed at studying high-transverse-momentum physics at the Tevatron. It weights 5500 tons and measures 13 m (height) $\times$ 11 m (width) $\times$ 17 m (length). The design was first proposed in 1983 and the detector was completed in 1992. A full description of Run I DØ detector

can be found in [35].

The DØ detector has undergone an upgrade for Run II [36], which is designed to enhance its capabilities from Run I and to accommodate the decrease in bunch spacing from 3.56 μs in Run I to 396 ns in Run II. Figure 3.2 shows schematic side view of the Run II DØ detector.

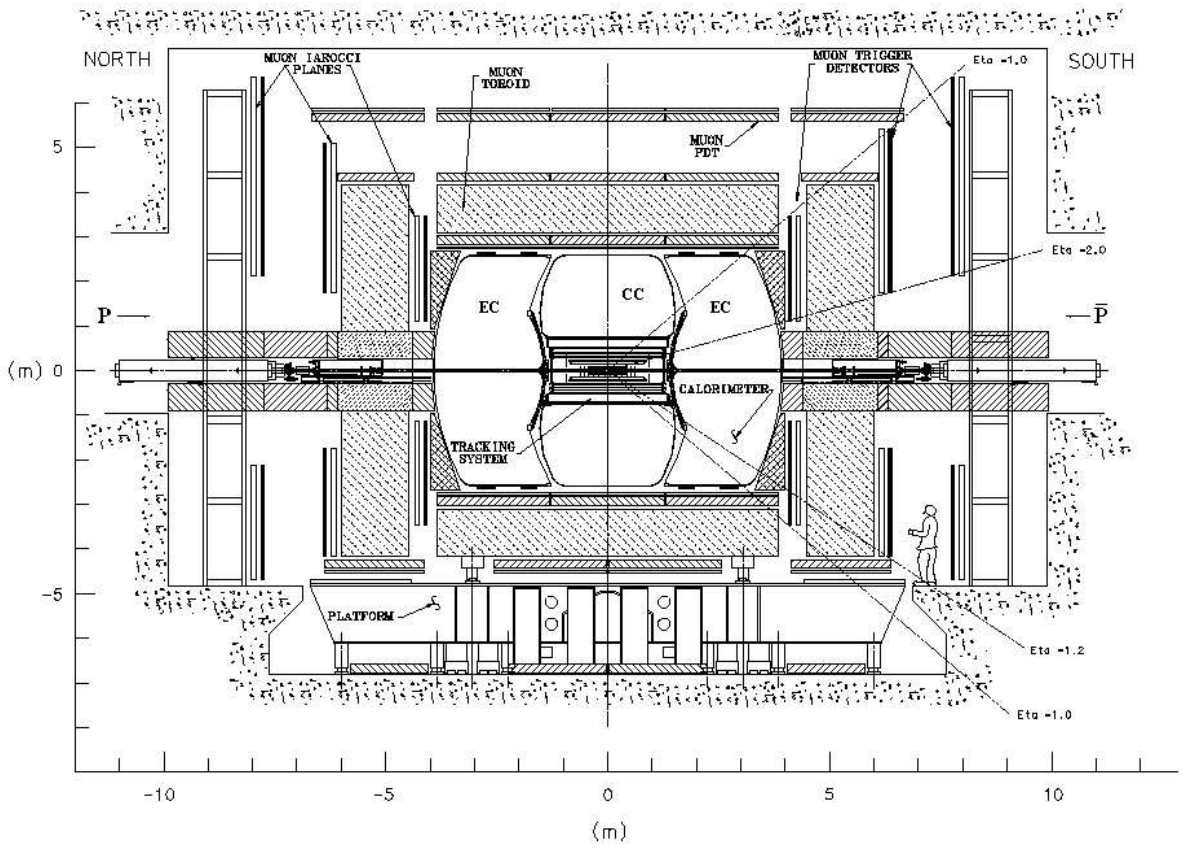


Figure 3.2: *A side view of the upgraded DØ detector. From inside to outside are the inner tracking system, calorimeter, and muon system.*

The upgraded DØ detector consists of three primary detector systems as we move from inside to outside: inner trackers, calorimeter, and muon systems. Here we give an overview of the upgraded Run II detector. The inner tracking system has been completely replaced, and sits inside a 2 Tesla magnetic field provided by a super-conducting solenoid. The calorimeter itself remains unchanged, although the readout electronics has been completely replaced. Preshower detectors have been added between the solenoid and the calorimeter to provide electron identification and to compensate for energy loss in the solenoid. New muon trigger detectors have been added to provide a fast muon trigger. A new 3-level trigger system and data acquisition system are used to handle the high event rate. In the following sections, each sub-detector system will be discussed in more detail.

3.3 The DØ Coordinate System

First, we define the coordinate system used in the experiment. DØ uses a Cartesian coordinate system (x, y, z), in which the positive z-axis is along the proton beam direction and the positive y-axis points upward. The x-axis is then fixed by the right hand rule.

Based on the Cartesian coordinates, the cylindrical coordinates (r, ϕ, z) and the spherical angles (ϕ, θ) are also used. The r coordinate corresponds to the perpendicular distance from the z-axis. The azimuthal angle ϕ and the polar

angle θ are defined so that $\phi = 0$ is along the $+x$ direction and $\theta = 0$ along the $+z$ direction.

Instead of the angle θ, it is more convenient to use the pseudorapidity η, defined as

$$\eta = -\ln(\tan\frac{\theta}{2}) \tag{3.1}$$

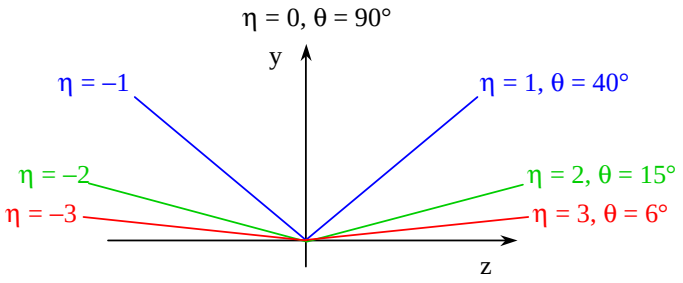
It approximates the true rapidity y

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z} \tag{3.2}$$

in the limit that $m \ll E$, where m is the particle's rest mass and E is energy. The rapidity is Lorentz invariant. So we use η and ϕ to describe the spatial angles, instead of θ and ϕ. For example, Figure 3.3 shows the values of θ corresponding to some representative values of η.

In addition, a detector pseudorapidity η_{det} is computed with respect to an interaction point of $z = 0$. The real $p\bar{p}$ interaction point which we use to calculate η is usually not at $z = 0$ and it varies from event to event, so η_{det} is slightly different from η. We usually use η_{det} to locate a particle in different detector regions.

The commonly used kinematic variables are the transverse energy $E_T = E\sin\theta$, and the transverse momentum $p_T = p\sin\theta$. Such a choice of variables is motivated



η	0	1	1.5	2	2.5	3	4
θ	90°	40°	25°	15°	9°	6°	2°

Figure 3.3: *The θ corresponding to some representative values of η.*

by the fact that besides the parton scattering, a significant fraction of energy escapes the detector as the nucleon remnants (*spectators*) carry it away down the uninstrumented beampipe. This is a consequence of the parton structure of a nucleon, i.e., the partons carry a fraction of the total nucleon energy. So the total energy balance can not be used. However, the transverse energy balance can be used since it is known to be zero before the collision, and its undetectable fraction (due to the limitation of the detector coverage) is negligible.

3.4 Central Tracking System

The goals of the new tracking system include: detection of charged particles over a large range in η ($\approx \pm 3$); charged particle momentum measurement in the

magnetic field of the solenoid; secondary vertex measurement for identification of b-jets.

A r–z view of the tracking system is shown in Figure 3.4. It consists of the Silicon Microstrip Tracker in the center and the Central Fiber Tracker surrounding it. A 2 Tesla magnetic field is provided by a superconducting solenoid surrounding the Fiber Tracker. The trackers (not including the solenoid) cover a region of 52 cm in r and 252 cm in z.

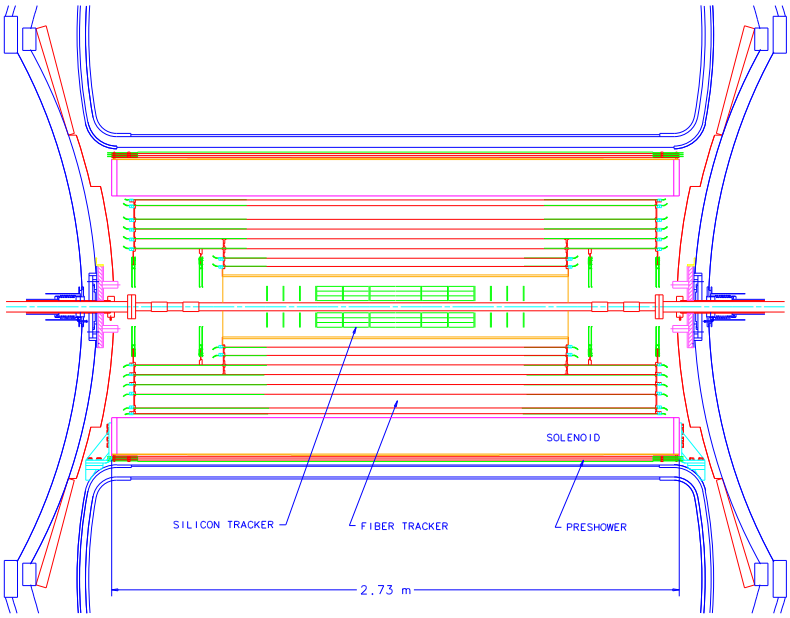


Figure 3.4: *A r–z view of the central tracking system. The trackers (not including the solenoid) cover a region of 52 cm in r and 252 cm in z.*

3.4.1 Silicon Microstrip Tracker

The Silicon Microstrip Tracker (SMT) [37] is the high resolution ($\leq 10\mu$m) part of the tracking system, and it is the first set of detectors encountered by the particles emerging from the collision. The silicon tracker is designed as a hybrid system consisting of *barrel* detectors measuring primarily the r–ϕ coordinate and *disk* detectors measuring the r–z coordinate as well as r–ϕ. This design was motivated by the fact that the interaction point is distributed over the z coordinate with a $\sigma_z \approx 25$ cm. Under these conditions it is difficult to design a detector such that the tracks are generally perpendicular to the detector surfaces at all η, so the hybrid system provides a solution.

The silicon tracker consists of: six barrel modules where silicon sensors are parallel to the beamline; twelve small disks ("F-disks") and four large disks ("H-disks") where silicon sensors are normal to the beamline. Figure 3.5 shows the barrels and disks of the silicon detector. Each barrel is 12 cm in length and contains four detector layers, starting at a radius of 3 cm and extending out to a radius of 10 cm. Four of the F-disks are sandwiched in the four 8 mm gaps between barrel segments, and the remaining eight F-disks are symmetrically located on each side of the central barrels. The four H-disks are located at $|z| = 110$ cm and 120 cm.

The *ladders* and *wedges* are the basic mechanical units in the barrels and in

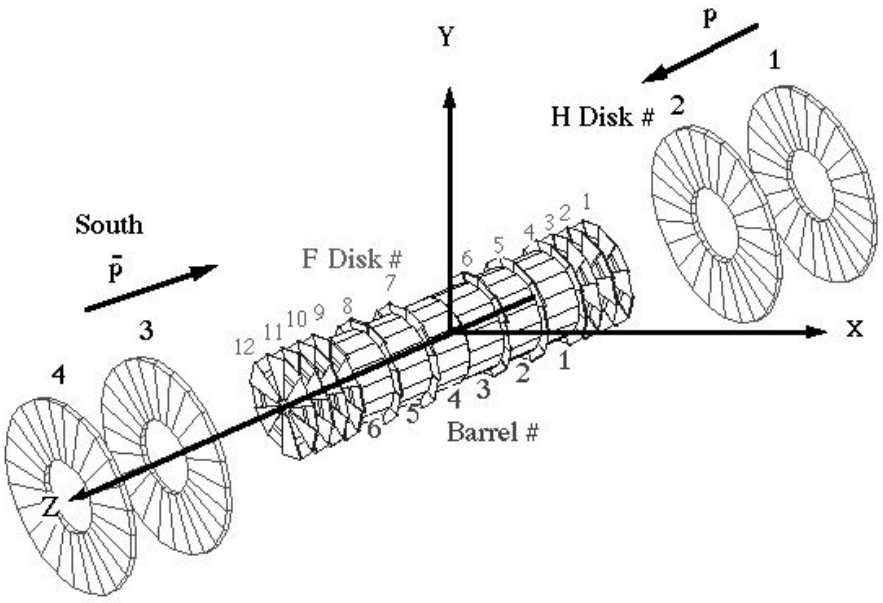


Figure 3.5: *The silicon microstrip tracker.*

the disks respectively. Conceptually they have the same structure, which consists of:

- Silicon microstrip sensors. Each sensor consists of thin silicon wafers implanted with very narrow and closely spaced strips (50–80 μm strip pitch).
- SVX II readout chips mounted on a high density interconnect (HDI) circuit. SVX II is a 128-channel chip which collects the signal from the silicon strips and performs pipelining and digitization.
- Supporting components and cables connecting to the port cards which control the SVX II chips.

The units are mounted on beryllium bulkheads, which serve as mechanical supports and provide cooling at the readout end by means of an integrated coolant channel. The total number of channels is around 793k.

3.4.2 Central Fiber Tracker

The Central Fiber Tracker (CFT) [38] serves two main purposes. First, together with the silicon tracker, it enables track reconstruction and momentum measurement for all charged particles within the range $|\eta| < 2.0$. Second, the fiber tracker provides fast "Level 1" track triggering within the range $|\eta| < 1.6$.

The fiber tracker consists of a total of 76,800 scintillating fibers mounted on eight concentric cylinders, as shown in Figure 3.6. The cylinders range in radius from 20 to 50 cm and are 2.5 m in length, except for the inner two which are 1.7 m in length. The scintillating fibers are double clad, 835 μm in diameter, and are arranged in a "doublet" layer structure where the centers of fibers in one layer are offset by half a fiber width relative to the centers of fibers in the adjacent layer.

There are eight doublet *axial* layers of scintillating fibers, x layers, which are aligned parallel to the beam direction. There are eight doublet *stereo* layers, u and v layers, which are oriented at a $+3°$ and $-3°$ angle with respect to the beam direction. From the innermost one outward, the order of layers is xu-xv-xu-xv-xu-xv-xu-xv.

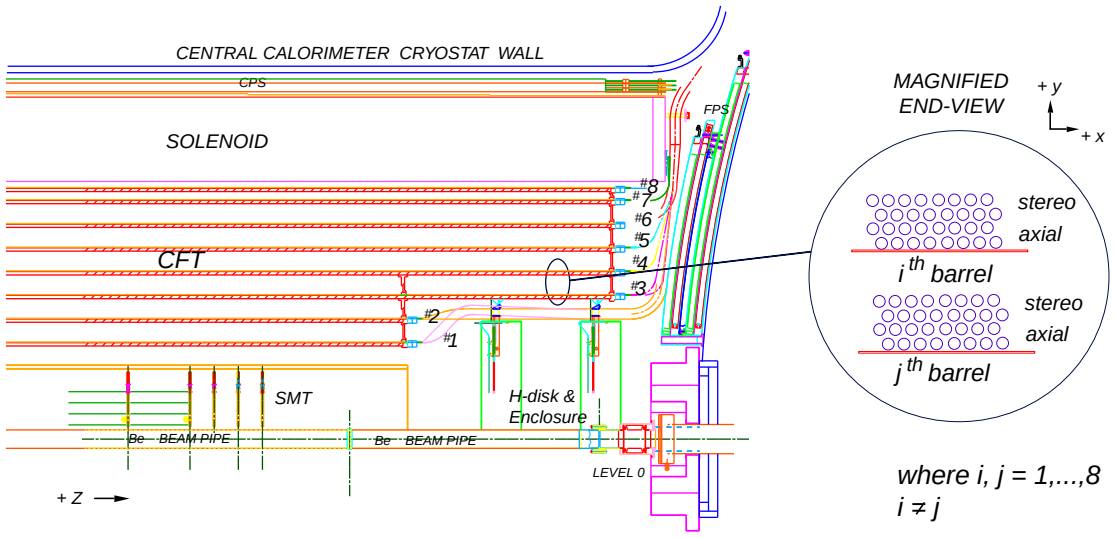


Figure 3.6: *A side view* (left) *of the central fiber tracker with cross-sectional end view* (right) *of the fibers.*

When a charged particle travels through a fiber, the scintillator emits light, which is internally reflected down the length of the fiber. A connection is made at the edge of the tracker to a clear fiber which transports the light to a solid-state light detector, called a Visible Light Photon Counter (VLPC). The VLPC is a silicon avalanche photodetector, which operates at a temperature of 6–15 Kelvin, with a quantum efficiency of over 80% and a gain of 20,000–50,000.

The readout electronics of the fiber tracker is similar to that of the silicon tracker, with the addition of a Level 1 trigger. The fiber tracker is divided into 80 equal azimuthal sectors for parallel processing. The signals (hits) from all eight

axial layers are input to the Field Programmable Gate Arrays (FPGAs) to look for coincidences as Level 1 tracks. These Level 1 tracks can then be combined with information from other sub-detectors to form electron triggers and muon triggers.

3.4.3 Solenoid

A Superconducting Solenoid magnet [39] which provides a magnetic field of 2 Tesla along the z-axis was added for Run II, surrounding the central trackers. This upgrade to the DØ detector enables the measurement of the momentum and the charge of a particle, by measuring the curvature of a track in the r–ϕ plane.

The solenoid is a 2.7 m long, two-layer coil with a mean radius of 60 cm. The coil is inside a cryostat at 4.7 Kelvin. The thickness of the coil plus its cryostat has been minimized (about 17 cm, or 0.9 radiation lengths) to preserve the largest possible tracking volume as well as to minimize the particle energy loss. Because of the energy loss in the solenoid, layers of lead were added before the preshower detectors to make the radiation lengths uniform in all directions.

3.5 Calorimeter System

The calorimeter system measures the energy and position of incident particles, and distinguishes particle types by their energy deposition patterns. Usually we

refer the DØ calorimeter as the liquid argon calorimeter, which consists of three liquid argon cryostats – a Central Calorimeter (CC) and two End-cap Calorimeters (EC). It is shown in Figure 3.2, and will be discussed in Section 3.5.3.

First we will describe the new Preshower detectors which are attached to the outside of the solenoid as well as on the inner surfaces of the EC calorimeter, and the Inter-cryostat detectors which are also placed on the inner surfaces of the EC calorimeter at a larger radius.

3.5.1 Preshower Detectors

The Preshower detectors are designed to enhance the electron and photon identification, and to sample energy after the solenoid thus to provide an additional calorimetric measurement. The system includes the Central Preshower detector and the Forward Preshower detectors. Their fine spatial resolution (1–2 mm) in both the r–ϕ and r–z planes, together with additional energy measurement, helps to distinguish electron, photon, and pion showers.

Central Preshower Detector

The Central Preshower (CPS) detector [40] (Figure 3.7) is located in the 51 mm gap between the solenoid and the central calorimeter cryostat at a radius of 72 cm, and covers the pseudorapidity region $|\eta| < 1.2$. It consists of three layers of

scintillating strips with wavelength-shifting (WLS) fiber readout. The innermost layer is an axial layer (x layer), and the outer two layers (u and v layers) are stereo layers with stereo angles of $\pm 23°$. A lead absorber before the preshower is tapered along the beam direction to make the thickness of the solenoid plus the lead equal to two radiation lenghths ($2X_0$) for all particle trajectories.

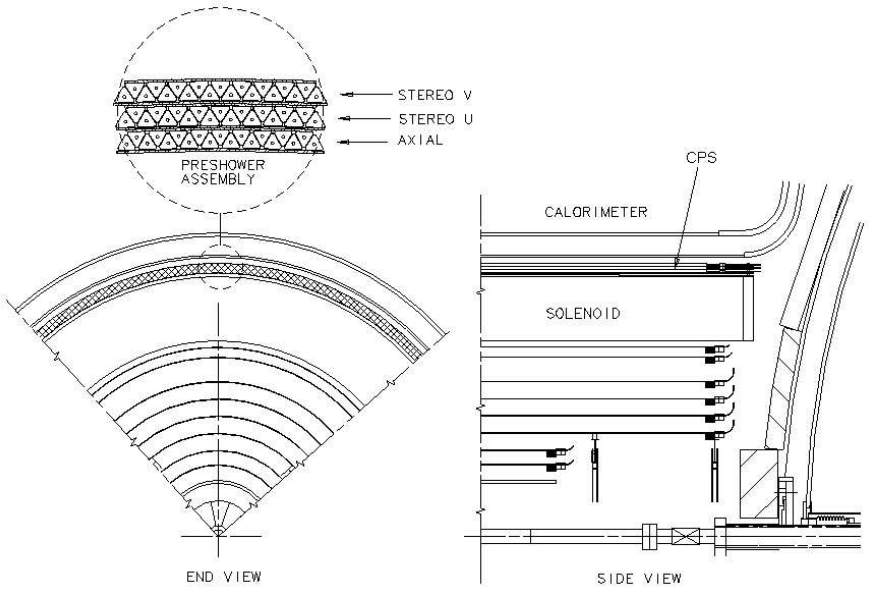


Figure 3.7: *A cross-sectional end view* (left) *and side view* (right) *of the central preshower detector.*

Each CPS layer consists of eight azimuthal octants which are 273 cm long. The scintillating strips have a triangular cross section with a 7 mm base and a 1 mm diameter hole in the center for housing the WLS fibers. The WLS fibers are split at $z = 0$ and are connected to the clear fibers at both ends. Clear fibers of

8–11 m in length transport the scintillation light to the VLPC, then the signals from the VLPC are sent to the SVX II for amplification and digitization. There are a total of 7680 readout channels in the CPS. The axial layer of the CPS is treated as the 9th layer of the fiber tracker for the purposes of triggering.

Forward Preshower Detectors

The Forward Preshower (FPS) detectors [41] (Figure 3.8) consist of two domes mounted on the inner face of each end-cap calorimeter cryostat and cover the pseudorapidity range $1.5 < |\eta| < 2.5$. Each detector is composed of two active scintillator planes with a two radiation length ($2X_0$) lead absorber sandwiched in between. Each scintillator plane consists of one u and one v layer, thus one FPS detector has four scintillator layers.

The outer scintillator plane covers the full available range $1.5 < |\eta| < 2.5$. Since electromagnetic particles traversing the solenoid ($1.5 < |\eta| < 1.65$) are likely to develop showers in the material upstream of the FPS, the inner scintillator plane is not needed in this region. Therefore, the inner scintillator plane and the lead absorber of $2X_0$ in thickness cover the range $1.65 < |\eta| < 2.5$.

Each of the four scintillator layers is made of eight modules, which are azimuthal 45° wedges. Each wedge consists of a 22.5° active scintillator volume in the center, and the wedge positions in successive layers are staggered by 22.5° to

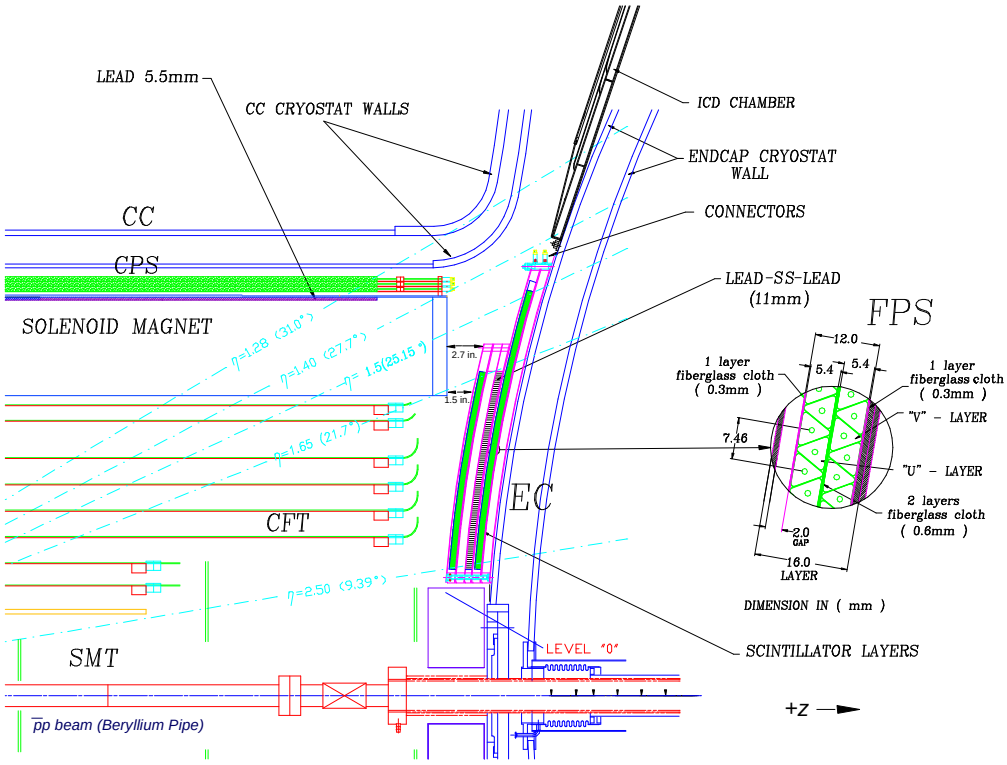


Figure 3.8: *One quarter view of the forward preshower detector.*

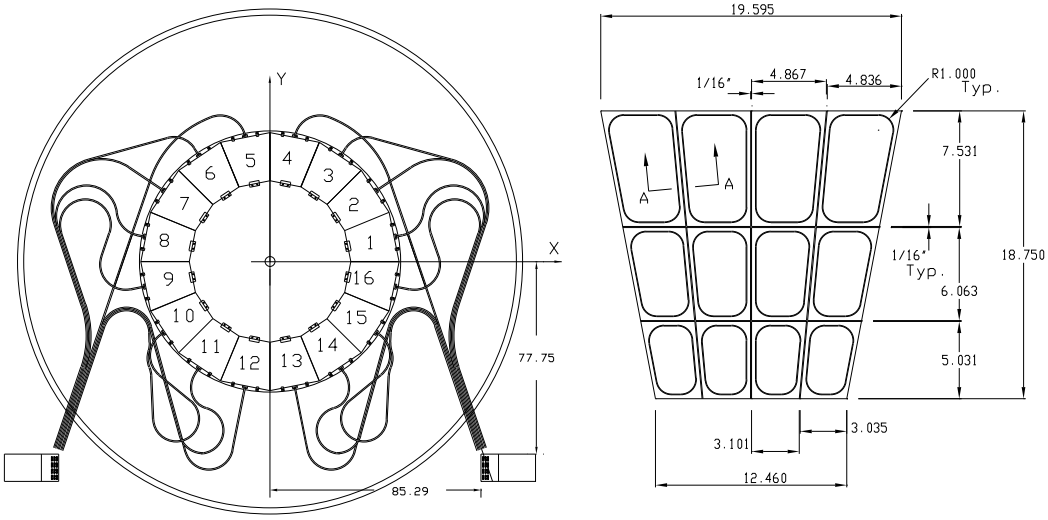
cover the full azimuthal angle. The FPS has a total of about 14,000 channels. As in the central preshower, the FPS signals from the WLS fibers are transferred to clear fibers routed at the periphery of the FPS and down to the VLPC on the platform below the detector.

3.5.2 Inter-cryostat Detector

In the intermediate region between the central and end-cap calorimeters ($1.1 < |\eta| < 1.4$), the rapidly changing material profile and extra uninstrumented material lead to reduced sampling of showers and hence a degradation of energy measurements in this region. The Inter-cryostat Detector (ICD) [42] restores energy resolution by providing additional sampling in this region, which improves the jet energy and missing energy measurements.

The ICD detectors are placed on the inner faces of each end-cap calorimeter cryostat to cover the range $1.1 < |\eta| < 1.4$ (the position of the ICD is shown in Figure 3.8). The ICD detectors consist of a single layer array of scintillating tiles, 16 super-tiles per ICD detector, and 12 tiles per super-tile, as shown in Figure 3.9.

The tile segmentation of 0.1×0.1 in $\Delta\eta \times \Delta\phi$ matches that of the calorimeter projective towers. The light is picked up by wavelength shifting fibers in the tiles, and transported 5–6 m in clear fiber waveguides to photomultiplier tubes located in a region of reduced magnetic field. The ICD detectors have 378 channels, and the readout electronics is compatible with the liquid argon calorimeter electronics system.



(a) (b)

Figure 3.9: *(a) Schematic of the ICD super-tile array hung on the inner surface of each end-cap calorimeter cryostat. (b) Schematic of the scintillator tiles within one ICD super-tile module.*

3.5.3 Liquid Argon Calorimeter

The DØ calorimeter [43] is a sampling calorimeter, and is unchanged for Run II. It uses mainly uranium as the dense absorber, and liquid argon as the active medium which is sensitive to particles passing through it. When a particle enters the absorber medium, it interacts with the material and produces a cluster of low-energy secondary particles, which is called a *shower*. As shower particles pass through the active medium, their energy is reduced by ionization, which in turn is measured in the detector. Since a majority of the incident energy is absorbed by the absorber medium, only a fraction of energy is detected by the active medium.

This is known as the sampling fraction, and thus the calorimeter is known as a sampling calorimeter.

When a high energy electron passes through a material with a high atomic number (the absorber), it loses energy primarily through Bremsstrahlung, in which a charged particle interacts with the Coulomb field around a nucleus and emits a photon. A high energy photon, on the other hand, interacts predominately via pair production, in which a photon converts into an electron-positron pair in the vicinity of a nucleus. The particles emitted in these interactions can themselves undergo Bremsstrahlung or pair production. Thus, an energetic electron or photon passing through a dense material results in a shower of secondary electrons, positrons, and photons. This process is called an electromagnetic shower. The shower will continue to develop until all secondary particles have sufficiently low energies that other energy loss mechanisms (mostly ionization) become important. The rate at which an incident electromagnetic particle loses energy is a property of the material, and is usually specified as the radiation length X_0:

$$\frac{dE}{E} = -\frac{dx}{X_0} \tag{3.3}$$

The radiation length for uranium is about 3.2 mm.

Hadronic particles also generate showers, but the showers are produced by

different mechanisms and have different patterns from the electromagnetic showers. Hadrons lose energy primarily through inelastic collisions with nuclei. These collisions produce secondary hadrons [1], which can in turn undergo inelastic collisions. This process is called a hadronic shower, and it continues until all particles have either been stopped by ionization losses or absorbed by nuclear processes. Hadronic showers are much more extended in space than electromagnetic showers of similar energy. The rate of the energy loss of a hadronic particle is described by the nuclear interaction length λ_A for the material, which is the analog to the radiation length for electromagnetic showers. For uranium, $\lambda_A \approx 10.5$ cm.

Calorimeter Structure

As shown in Figure 3.10, the calorimeter consists of one Central Calorimeter (CC) covering the region $|\eta| < 1.2$, and two End-cap Calorimeters (EC) extending the coverage to $|\eta| \approx 4$. The use of liquid argon requires a containment vessel (cryostat), in which the argon is maintained at a temperature of 78 Kelvin.

The calorimeter is highly modular and finely segmented in the transverse and longitudinal shower directions. Each module consists of a row of interleaved absorber plates and signal readout boards, as shown in Figure 3.11. The 2.3 mm gap separating adjacent absorber plates and signal boards is filled with liquid argon

[1] Among the secondary particles of a hadron, about 1/3 of them are neutral pions. The neutral pions produce photons and electrons which interact electromagnetically.

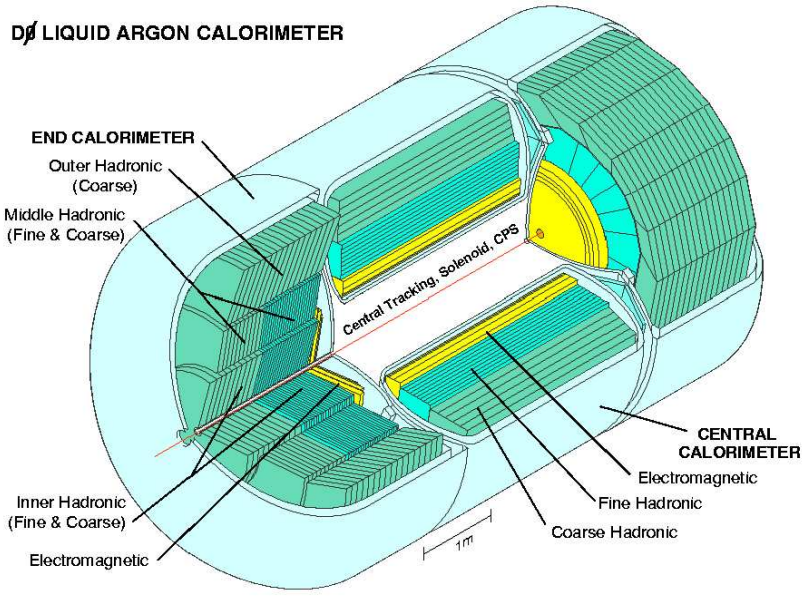


Figure 3.10: *The DØ calorimeter.*

as the active medium. The signal boards are constructed by laminating a copper pad with two separate 0.5 mm thick G10 sheets at each end. The outer surfaces of the boards are coated with a highly resistive epoxy. Different absorber plate materials are used in different locations. The electromagnetic modules use 3 mm thick (CC) and 4 mm thick (EC) uranium plates. The fine hadronic modules have 6 mm thick uranium-niobium (2%) alloy plates. The coarse hadronic modules contain thick (46.5 mm) plates of either copper (CC) or stainless steel (EC).

An electric field is established by grounding the absorber plate while applying

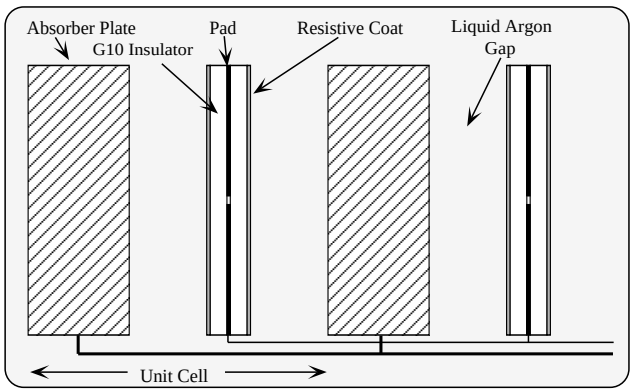


Figure 3.11: *Schematic view of a calorimeter cell.*

a positive potential of 2.0 kV to the resistive surface of the signal board. The electron drift time in the liquid argon gap is about 430 ns, which sets the time scale for signal charge collection. Signals from several signal boards at approximately the same η and ϕ are ganged together in depth to form a readout cell. The readout cells are designed to be arranged into *pseudo-projective* towers, i.e., the centers of cells lie on rays projecting from the center of the detector, but the cell boundaries are aligned perpendicular to the absorber plates (Figure 3.12).

Central Calorimeter

The central calorimeter (CC) is composed of three concentric cylindrical shells.

- The electromagnetic (EM) modules used in the inner ring are segmented into four layers, known as the EM1, EM2, EM3, and EM4 layers, with depths of

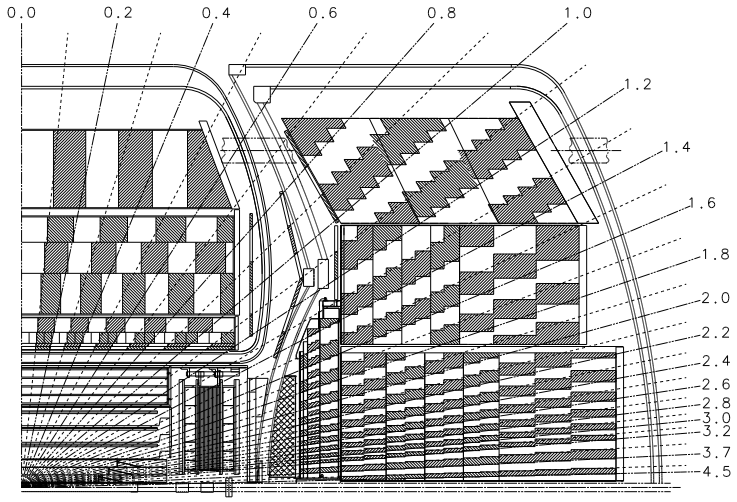


Figure 3.12: *Side view of one quadrant of the calorimeter, with pseudorapidity lines plotted. The shading pattern indicates distinct calorimeter cells.*

2.0, 2.0, 6.8, and 9.8 X_0 (radiation length) respectively. A full EM module comprises 20.5 X_0 and 0.76 λ_A (nuclear absorption length).

- The fine hadronic (FH) modules used in the surrounding ring have three layers, known as the FH1, FH2, and FH3 layers, with depths of 1.3, 1.0, and 0.9 λ_A respectively.

- The coarse hadronic (CH) modules used in the outer ring contain one layer of 3.2 λ_A.

The transverse segmentation of all the layers is 0.1×0.1 in $\Delta\eta \times \Delta\phi$, except for the EM3 layer which has finer segmentation with a cell size of 0.05×0.05 in $\Delta\eta \times \Delta\phi$ [2]. The total number of channels in the CC is about 18k.

End-cap Calorimeters

The two mirror-image end-cap calorimeters (EC) contain four module types: an EM module and three kinds of hadronic modules at increasing radii (Figure 3.10).

- The EM module contains four layers (EM1, EM2, EM3, and EM4) with depths of 0.3, 2.6, 7.9 and 9.3 X_0, comprising about 0.75 λ_A.
- The inner hadronic module consists of four fine hadronic layers of 1.1 λ_A each, and one coarse hadronic layer of 4.1 λ_A.
- The middle hadronic modules have four fine hadronic layers of 0.9 λ_A each, and a single coarse hadronic layer of 4.4 λ_A.
- The outer hadronic modules consist of three coarse hadronic layers. In the range $0.7 < |\eta| < 1.1$, they provide the CH modules for the towers whose EM and FH modules are in the CC.

[2] In Run I, the EM3 layer spanned the region of the maximum EM shower energy deposits thus had finer segmentation. In Run II, there is an additional $2X_0$ in front of the calorimeter, which makes the maximum EM shower happen at the front part of the EM3 layer.

The cell segmentation is the same as in the CC for $|\eta| < 2.6$. For $2.6 < |\eta| < 3.2$, the cell size in the EM3 layer is increased to 0.1×0.1. For $|\eta| > 3.2$, the cell size in all layers is increased to 0.2×0.2 and continues to increase up to 0.4×0.4 at $|\eta| \approx 4$, since the physical size of showers measured in η becomes larger at high η regions. There are about 36k channels in the two EC detectors.

Calorimeter Electronics

The calorimeter readout electronics has been upgraded [44] to accommodate the reduced Run II bunch spacing time, and to maintain the Run I noise performance. It was necessary to optimize the noise contributions (including electronic, uranium and pile-up), and to add an analog storage device in which to store the signals until a Level 1 trigger decision can be made (in about 4 μs).

A schematic diagram of the Run II calorimeter readout chain is shown in Figure 3.13. The calorimeter has a total of about 55,000 channels. Signals from the calorimeter cells are brought to the feed-through ports (a sealed interface between the inside and outside of the cryostat) and carried out via 30 Ω co-axial cables to the preamplifiers (preamps) located on top of the cryostats.

The charge-integrating preamp is a hybrid circuit on ceramic, with a preamplifier circuit and output driver. Two input low-noise jFETs in parallel are used for better noise performance. To make the preamp output wave-form similar for

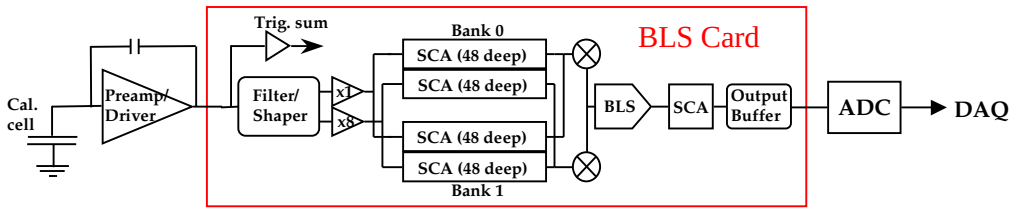


Figure 3.13: *Readout chain of the calorimeter in Run II.*

all channels, fourteen (an additional one for ICD) different species of preamps are used to compensate for various capacitances of calorimeter cells.

The preamps drive signals through 115 Ω twist and flat cables of about 25 m in length to the Baseline Subtractor (BLS) boards which sit on the platform underneath the detector. The BLS board contains signal shapers, trigger pick-off and summation circuits, and the Switched Capacitor Array (SCA) chips. The SCA is an analog signal storage device, which contains a 12-channel by 48-depth capacitor array to pipeline the signals. The SCA can store analog voltages from 0–5 V with a precision greater than 12 digital bits. In order to achieve a 15-bit dynamic range, the shaped signals are passed through two different gain paths: $\times 1$ for large signals and $\times 8$ for small signals.

If the Level 1 trigger system accepts the event, a *baseline subtraction* is performed and the result is sent to a Level 2 SCA, where it awaits a Level 2 trigger decision. After the Level 2 trigger acceptance is received, the signal is read out and

sent to the Analog to Digital Converters (ADCs), which digitize and zero-suppress the signals and then send them to the Data Acquisition System (DAQ).

The shaped signals are sampled every 132 ns, and pipelined in Level 1 SCAs for storage. The Level 1 trigger decision time is about 4 μs, so the 48-depth SCA provides a deadtimeless operation for the Level 1 trigger. The timing is tuned such that the shaped output can be sampled at its peak at about 320 ns. Since the preamp output is an integral of the detector signal characterized by a rise time of 430 ns and a long recovery time of 15 μs, it can pile up on earlier signals over successive bunch crossings. In order to remove slowly varying offsets and the pile-up of events from neighboring bunch crossings, the pulse height sampled 396 ns earlier is used as the baseline, and subtracted from the current pulse height. This process is called "baseline subtraction".

A fast trigger pick-off signal for each channel is obtained from the preamp output, and the calorimeter Level 1 trigger uses the energy summed in trigger towers of size 0.2×0.2 in $\Delta\eta \times \Delta\phi$ by the "trigger summers". The calorimeter electronics is calibrated using the pulser calibration system, which can supply a precise charge pulse of known value to selected preamps. Coordination between the calorimeter readout and the trigger framework is done by the timing and control system through the Field Programmable Gate Array (FPGA) chips.

3.6 Muon System

Muons are minimum ionizing particles which only deposit a small amount of energy in the calorimeter, so the identification and measurement of muons relies on the muon system [45] which is the outermost detector system. As shown in Figure 3.2, the DØ muon system is divided into two components: the Wide Angle Muon System (WAMUS) in the central region covering $|\eta| < 1$, and the Forward Angle Muon System (FAMUS) in the forward regions covering $1 < |\eta| < 2$.

Both the central and forward systems consist of three detection layers, labeled A, B, and C from inside to outside. A 1.8 Tesla toroidal iron magnet is sandwiched between the A-layer and the B-layer, to bend the trajectory of the muon and thus obtain a measurement of its momentum and charge. The central WAMUS uses proportional drift tubes (PDT) in three layers, and scintillators in two layers (no middle-layer scintillator). The forward FAMUS uses three planes of mini-drift tubes (MDT), and three layers of scintillators. The PDTs and MDTs are used for muon tracking and momentum measurement, while the scintillators are used for fast muon triggering and cosmic muon rejection. Shielding has been added around the beam pipe from the calorimeter to the accelerator tunnel to reduce background rates.

3.7 Trigger System

At a hadron collider, elastic or inelastic hadron collisions happen at every bunch crossing. The total collision rate by far exceeds the rate at which event processing and recording can be performed. So it is important to determine the quality and physical importance of an event right after a bunch crossing, and store the physically interesting events to disks. This selection is done by the means of triggers.

The DØ trigger framework consists of three levels of triggers: Level 1 (L1), Level 2 (L2), and Level 3 (L3), with increasingly sophisticated event selection and decreasing output rate. The expected trigger accept rate limits in Run II are 2.5 kHz, 1 kHz, and 50 Hz at L1, L2, and L3 respectively. A block diagram of the trigger system is shown in Figure 3.14. The luminosity monitor is also known as the Level 0 trigger, which detects inelastic collisions and measures luminosity.

3.7.1 Luminosity Monitor

The DØ luminosity monitor [46] consists of two arrays of plastic scintillation counters located on the inside faces of the EC cryostats and arranged symmetrically around the beam pipe. The detector is placed at $|z| \approx 135$ cm and covers the region $2.7 < |\eta| < 4.4$. The scintillation light is read out using fine-mesh photomultipliers which can operate in large magnetic fields.

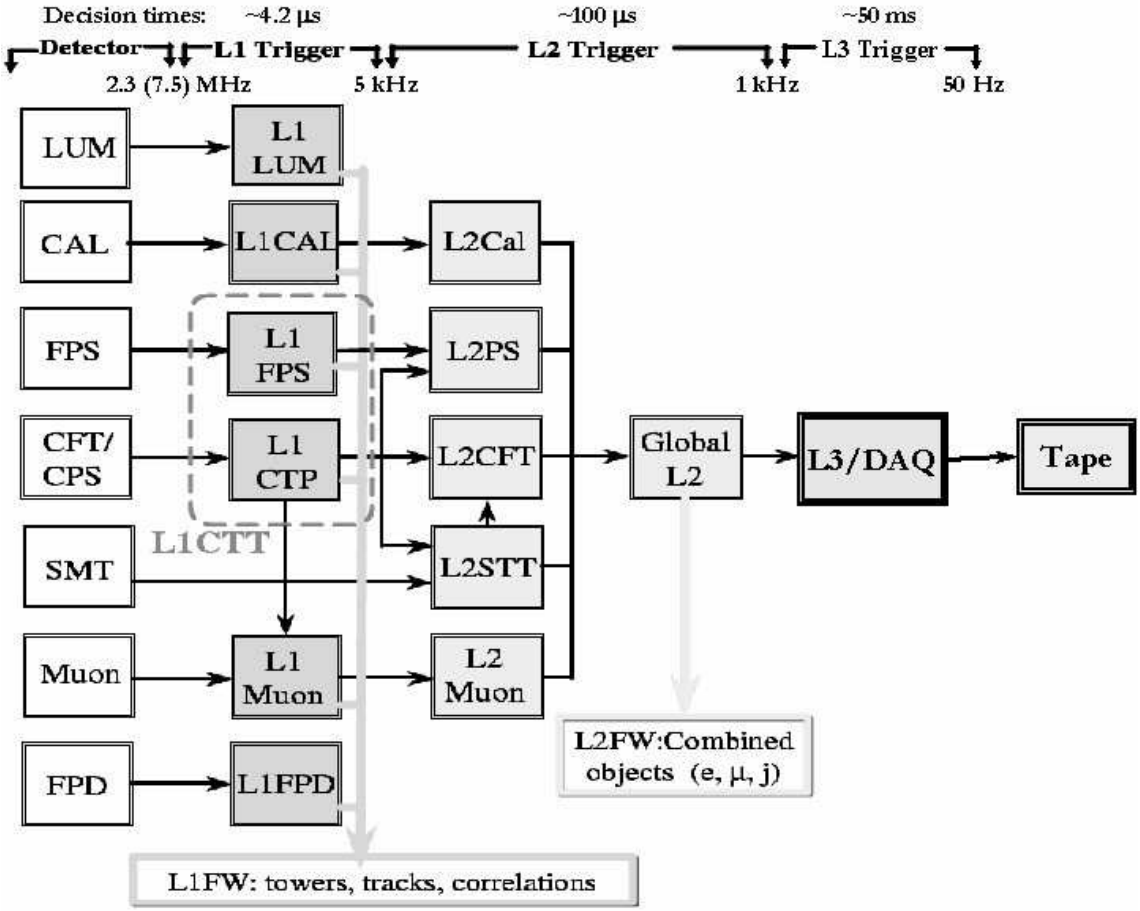


Figure 3.14: *Design of the DØ trigger system.*

When an interaction occurs, the remnants of the incident proton and antiproton give a pair of time correlated hits in the scintillators. The scintillation counters detect such non-diffractive inelastic collisions with high efficiency. The rate of these collisions is used to determine the luminosity. The difference in hit times provides a fast measurement of the vertex z position. In addition, the luminosity monitor is a tool that provides diagnostic information regarding the accelerator performance and also can be used to help identify the number of interactions per beam crossing.

3.7.2 Level 1 Trigger

The Level 1 trigger is a hardware trigger system. It consists of the L1 trigger framework [47] and the L1 trigger subsystems, which include the central fiber tracker, the central and forward preshower, the calorimeter, and the muon detectors. The calorimeter, fiber tracker, and preshower detectors provide electron triggering for $|\eta| < 2.5$. The fiber tracker and the muon systems provide muon triggering in the region $|\eta| < 2.0$.

The L1 trigger subsystem processes information from specific subdetectors and reports its findings (trigger terms) to the L1 framework. The L1 framework supports 128 criteria (L1 triggers) and each criteria is a combination of trigger terms. If any one of the 128 criteria is satisfied, the L1 framework issues an

accept and the event signals are digitized and moved from the pipeline to a series of buffers to await a L2 trigger decision.

3.7.3 Level 2 Trigger

The Level 2 trigger [48] consists of a hardware framework and a separate set of processors. The L2 trigger is aimed at reducing the 2.5 kHz L1 accept event rate by roughly a factor of two to three within 100 μs. It has two stages, the preprocessor stage and the global processor stage. The preprocessor stage refines L1 trigger information from individual subdetectors (silicon detector, fiber tracker, preshower, calorimeter, and muon detector) for use in the second stage. The global processor stage combines different trigger objects found by the preprocessors, and makes the L2 trigger decision.

The algorithm in the global processor stage is implemented by a fast digital beta processor on a VME-based CPU card. This provides flexibility to combine the trigger information in ways not possible at L1 across different subsystems. There are 128 L2 triggers, and the L2 decision is reported to the L2 trigger framework to be used to guide L3 triggering.

3.7.4 Level 3 Trigger

The Level 3 trigger is a software based system which is often described as an event filter [49]. The L3 trigger uses a farm of standard, high performance commercial processors to run event filter algorithms. It reduces the input rate of 1 kHz to an output rate of 50 Hz.

The L3 trigger combines and partially reconstructs data for each event, to find physics objects (jets, electrons, muons, tracks, scalar E_T, missing E_T, etc.) in a fast and coarse process. The partially reconstructed event is analyzed by a L3 filtering process, which selects events based on a set of criteria on physics objects. If any of the L3 criteria is passed, the event will be recorded via the data acquisition system.

3.7.5 Data Acquisition

The DØ Data Acquisition (DAQ) system adopts an ethernet-based design [50]. The data flow in parallel out of about 70 VME readout crates, each corresponding to a section of a subdetector system or the trigger framework. Each crate is read out by a Single Board Computer (SBC). Data size in each crate is 1–10 kB, and the total event size is about 250 kB.

The data are moved out of SBCs over the Ethernet via a series of Ethernet connections, which transfer them over to the main switch via a 1 GB/s optical

fiber. The L3 farm nodes receive data fragments through the main switch and do L3 filtering. Events passing the L3 trigger will be stored on a tightly coupled disk and robotic tape system, and made available for analysis.

Chapter 4

Event Reconstruction

The events recorded by the data acquisition system are in raw data form. The raw data contain information such as hits in the central tracking system, digitized counts in the calorimeter cells, etc. The process to convert the raw data into physics objects (electrons, photons, jets, etc.) is called event *reconstruction* [51]. At DØ, the software package RECO is used for event reconstruction, which is performed in four steps:

- **Hit Finding.** The digitized signals from the tracking detectors are converted into spatial locations of hits. The signals from calorimeter cells are converted to energies.

- **Tracking and Clustering.** The tracking hits are combined to form tracks. The calorimeter energy deposits are grouped to form clusters.

- **Vertexing.** The location of the $p\bar{p}$ interaction is calculated for determination of various kinematic quantities.

- **Particle Identification.** The tracks and clusters are combined to form candidates of electrons, photons, jets, muons, etc.

4.1 Vertex Reconstruction

The primary and secondary vertices are reconstructed from tracks using the impact parameter technique [52]. The track finding algorithm, called the road-following algorithm [53], is based on the TRF++ package [54] and uses paths during tracking finding. The algorithm takes the hit information from the different layers of the central tracking system and uses a Kalman filter approach to build tracks. First, track segments are produced in each layer, and the segments in the first few layers are used to build a "seed" track. The seed track is extrapolated to the remaining layers, where the propagation takes into account the effects of magnetic fields, multiple scattering and energy loss in materials [55]. A fit of a track and a segment at a new surface is performed to determine whether to add or reject the segment depending on the χ^2. Finally filters are applied to reject tracks based on the overall χ^2 of their fit.

4.1.1 Primary Vertex

The primary vertex refers to the hard-scattering vertex that triggered the event (i.e., the interaction point). If additional interactions occurr, there can be more than one primary vertex in an event. The position of the primary vertex is used to precisely calculate many physical quantities, such as the transverse momentum of tracks and the transverse energy of jets. The finding of the primary vertex starts from all selected tracks, and the procedure is:

1. Fit a vertex from a set of tracks (total number N_{trk}) and compute $\chi^2(N_{trk})$.
2. Each track is excluded separately and a new $\chi^2(N_{trk}-1)$ is computed.
3. Select the track with maximum difference $\Delta_{max} = \chi^2(N_{trk}) - \chi^2(N_{trk}-1)$.
4. Exclude the track from the set if $\Delta_{max} > \Delta_{threshold}$.
5. Repeat the procedure while there are tracks whose $\Delta_{max} > \Delta_{threshold}$.

This algorithm can also handle events due to multiple interactions by storing the rejected tracks into a separate set and reusing this set for vertex finding. A minimum of two tracks is required to fit a vertex candidate.

4.1.2 Secondary Vertex

The secondary vertices are the displaced vertices that arise from the decay of long-lived particles, e.g., B and D mesons whose decay length is a few mm. Their reconstruction is important for b-tagging. The tracks from the secondary vertex usually have a large impact parameter with respect to the primary vertex (the distance of closest approach to the primary vertex), as shown in Figure 4.1. Because the beam spot is very narrow (about 30 μm in diameter), these tracks will also have a large r–ϕ parameter with respect to the beam line. There can be more than one secondary vertex in an event. The secondary vertex search uses displaced tracks with respect to the primary vertex, and the procedure is as follows:

1. Make a seed vertex from two tracks not belonging to the primary vertex or a previously found secondary vertex.

2. Fit the vertex from the set of N_{trk} ($N_{trk} = 2$ for the first time) tracks and compute $\chi^2(N_{trk})$.

3. Compute the $\chi^2(N_{trk}+1)$ obtained by adding each of the other tracks one by one.

4. Select the track giving the smallest contribution $\Delta_{min} = \chi^2(N_{trk}+1) - \chi^2(N_{trk})$.

5. Add the good track to the set if $\Delta_{min} < \Delta_{threshold}$.

6. Go back to step 2 and repeat the procedure until there is no more good tracks.

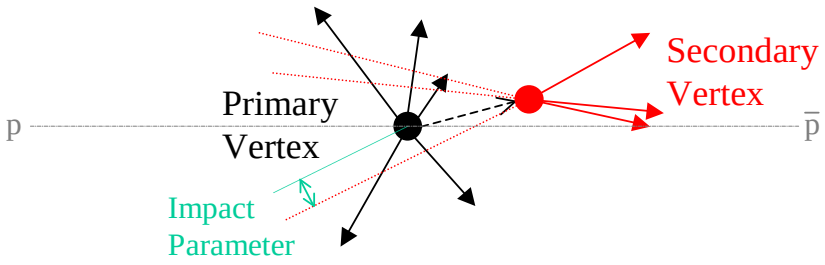


Figure 4.1: *Primary vertex and secondary vertex. Tracks from the secondary vertex have a large impact parameter with respect to the primary vertex.*

4.2 Electron Reconstruction

Electrons and photons have very similar shower shapes in the calorimeter, and they deposit the majority of their energy in the electromagnetic (EM) layers of the calorimeter. Electrons are expected to have a track in the central tracking system while photons do not. We call both electrons and photons EM objects.

4.2.1 EM Cluster Reconstruction

A cluster is a list of adjacent cells with significant energy deposits. The standard EM cluster algorithm at DØ is the simple cone algorithm [56] based on towers [1]. The centroid of the cluster is taken as the energy weighted mean value of coordinates of the cells in the EM3 layer of the calorimeter.

An initial EM cluster has to pass crude selection criteria as follows:

- $E_T > 1.5$ GeV.
- The fraction of the total energy deposited in the calorimeter EM layers to the total energy deposited in all layers > 0.9.
- The isolation < 0.2. The variable *isolation* is defined as:

$$\text{isolation} = \frac{E_{total}(0.4) - E_{EM}(0.2)}{E_{EM}(0.2)} \tag{4.1}$$

where $E_{total}(0.4)$ is the total energy (EM and hadronic) in the towers within a cone of radius 0.4 in the η–ϕ space around the centroid of the cluster, and $E_{EM}(0.2)$ is the energy deposited in the EM layers of the calorimeter within a core cone of radius 0.2. Small values of isolation correspond to the situation that most of the energy is deposited in a narrow region of the

[1] Calorimeter cells with the same η and ϕ are grouped together to form towers. A calorimeter read-out tower is of size 0.1×0.1 in $\Delta\eta \times \Delta\phi$, and a trigger tower is of size 0.2×0.2 in $\Delta\eta \times \Delta\phi$.

EM layers in the calorimeter. Such a pattern of energy deposition is the characteristic of an electron or a photon.

4.2.2 EM Identification

The above criteria for initial cluster finding are very loose. At the final stage of reconstruction, a number of variables that describe EM cluster properties are computed and stored. Among those, the ones important for EM identification are EM-fraction, isolation, and H-matrix χ^2. The EM-fraction is the ratio of the cluster energy deposited in the EM layers to the total energy deposited in all EM and hadronic layers of the calorimeter. The isolation has the same definition as in Equation 4.1.

The H-matrix is a multi-variable tool to analyze the detailed shape of a cluster, and therefore provide a more reliable discrimination between EM clusters and jets. Eight variables are used to characterize the difference in the shower profiles:

- Fractions of the energy deposited in each of the four EM layers of the calorimeter: EM1, EM2, EM3, EM4.
- Cluster width in r–ϕ plane, and cluster width in z (CC) or r (EC).
- Logarithm of the cluster energy.
- Reconstructed interaction point position in z.

For a sample of N Monte Carlo electrons, a covariance matrix M (8×8) is defined as:

$$M_{ij} = \frac{1}{N} \sum_{n=1}^{N} (x_i^n - \overline{x_i})(x_j^n - \overline{x_j}) \tag{4.2}$$

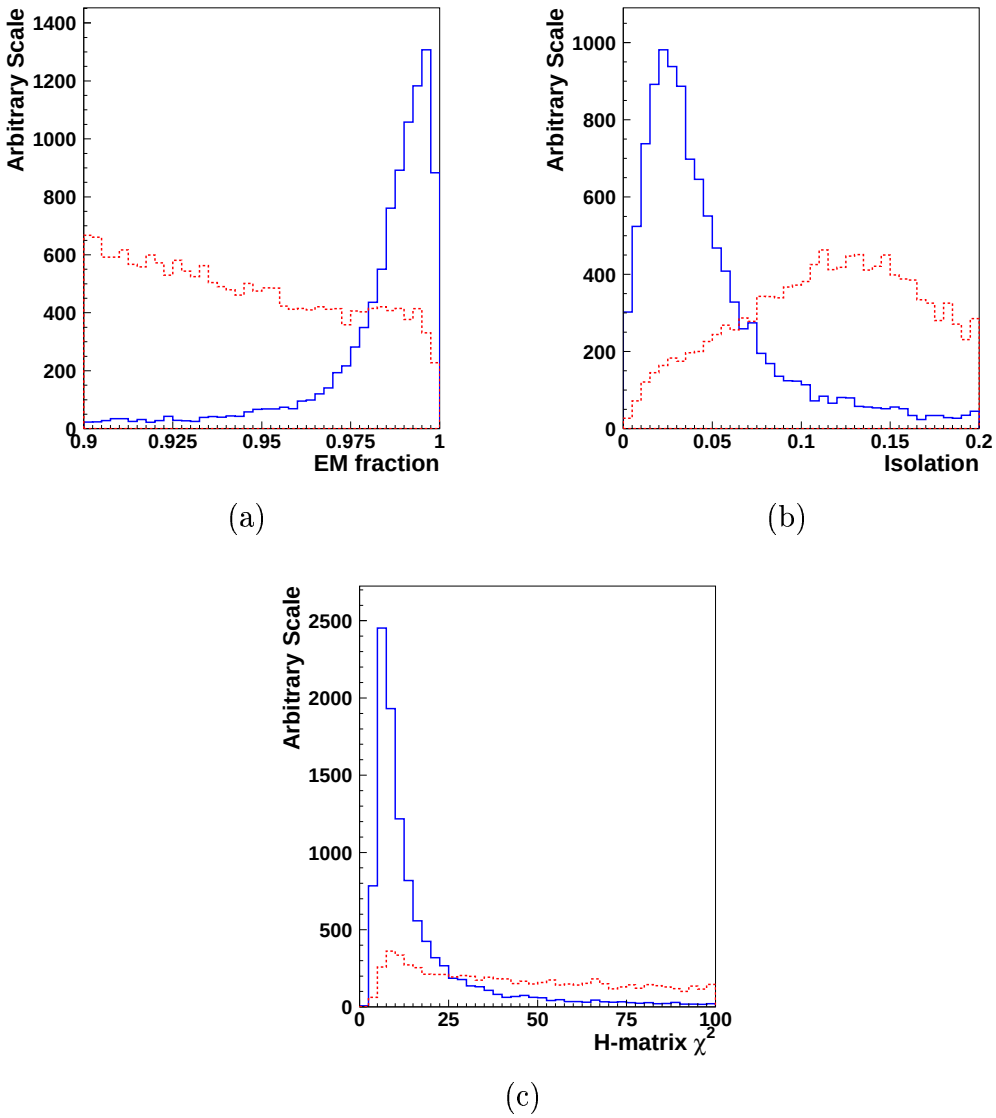
where x_i^n is the ith variable of the nth electron, and $\overline{x_i}$ is the mean value of the ith variable from the N measurements. The H-matrix is defined as the inverse of the covariance matrix: $H = M^{-1}$. For a subsequent measurement y, we define a χ^2 to quantify how consistent the shower shape is with that of electrons or photons:

$$\chi^2 = \sum_{i,j=1}^{8} (y_i - \overline{x_i}) H_{ij} (y_j - \overline{x_j}) \tag{4.3}$$

The standard EM identification [57] at DØ requires:

- EM-fraction > 0.9
- isolation < 0.15
- H-matrix (8×8) $\chi^2 < 20$

Figure 4.2 shows the distributions of above variables for electrons from $Z \to e^+e^-$ decays and jets from multijet events which are reconstructed as (fake) EM objects.



(a) (b)

(c)

Figure 4.2: *EM ID variable distributions for electrons from* $Z \to e^+e^-$ *decays* (solid line) *compared with jets from multijet events which are reconstructed as (fake) EM objects* (dashed line)*: (a) EM-fraction, (b) isolation, (c) H-Matrix* (8×8) χ^2.

4.2.3 Track Matching

An electron is expected to have an associated track in the central tracking system. To measure the spatial distance between a cluster and a track, the track is extrapolated to the EM3 layer of the calorimeter. We define a track matching significance, χ^2, for electrons in CC and EC regions as:

$$\chi^2(CC) = (\frac{\Delta\phi}{\sigma_\phi})^2 + (\frac{\Delta z}{\sigma_z})^2 + (\frac{E/p-1}{\sigma_{E/p}})^2 \quad (4.4)$$

$$\chi^2(EC) = (\frac{\Delta\phi}{\sigma_\phi})^2 + (\frac{\Delta r}{\sigma_r})^2 + (\frac{E/p-1}{\sigma_{E/p}})^2 \quad (4.5)$$

where $\Delta\phi$, Δz, and Δr are the spatial differences between the cluster and the track; E is the cluster energy and p is the track momentum; σ_ϕ, σ_z, σ_r, and $\sigma_{E/p}$ are the corresponding resolutions.

Then we calculate the probability of getting a χ^2 value for a specific track, according to the standard χ^2-distribution obtained by using electrons from $Z \to e^+e^-$ events. We take the track with a χ^2 probability > 0.01 as a good match, which is optimized to keep high efficiency and low background.

4.3 Jet Reconstruction

QCD theory implies that color particles are confined by the strong force, thus a parton (quark or gluon) only exists in hadrons and cannot exist alone. When an energetic parton is produced from a $p\bar{p}$ collision and moving away from other partons, the potential of the strong force field between the partons grows as the distance increases until the energy is large enough to create a parton-antiparton pair out of the vacuum. Such processes take place as more particles are generated out of the vacuum, until the energy is too low to permit the creation of parton pairs. This *hadronization* process produces a group of hadrons moving in about the same direction of the original parton. These particles deposit a cluster of energy in the calorimeter, which is called a *jet*. Pions make up the majority of particles in a jet, since they are the lightest hadrons.

4.3.1 Cone Jet

The standard jet reconstruction at DØ uses a cone jet algorithm [58], which proceeds as follows:

- **Preclustering.** Beginning with the highest E_T tower, preclusters are formed of adjacent towers within $\Delta\eta < 0.3$ and $\Delta\phi < 0.3$. Only towers with $E_T > 1$ GeV are included in preclusters.

- **Cone Clustering.** The E_T weighted centroid of a precluster is used as the initial cone center of a jet. From it, all towers within a radius of $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ are included in the jet, and the E_T weighted centroid is recalculated. This search uses a jet cone size of $R = 0.5$. This process is repeated until the jet is stable. A jet must have $E_T > 8$ GeV.

- **Merging and Splitting.** If two jets share some towers, the fraction of E_T that is shared between them is examined. If it is more than 50% of the E_T for the lower E_T jet, the two jets are merged and the center is recalculated. Otherwise, they are split, and each shared tower is assigned to the nearest jet.

4.3.2 Jet Identification

There are a number of quantities which are important for jet identification (jet ID) [59], especially to distinguish the real jets from fake ones due to hot cells or calorimeter noise. These quantities are:

- EM-fraction, which is the fraction of total jet energy deposited in the EM layers of the calorimeter. Jet ID requires $0.05 <$ EM-fraction < 0.95.

- Coarse-Hadronic-fraction (CHF), which is the fraction of the total jet energy deposited in the coarse hadronic layers of the calorimeter. Jet ID requires

CHF < 0.4.

- $n90$, which is the number of towers whose summed $E_T \geq 90\%$ of the total jet E_T. The tower E_T is summed up starting from the highest E_T tower and in the order of decreasing E_T. Since jet shower is wide in the calorimeter, a good jet should have $n90 > 1$.

- Hot-fraction, which is the ratio of the hottest cell E_T and the next-to-hottest cell E_T. Hot-fraction < 10 is required for a good jet.

- $f90$, which is defined as $f90 = n90/nitm$, where $nitm$ is the total number of towers in a jet. The DØ certified jet ID version 2.1 [60] applies a requirement of $f90 < 0.8 - 0.5\times$CHF or CHF < 0.1, which greatly reduces the number of fake jets due to calorimeter noise.

The distributions for the above jet ID variables are shown in Figure 4.3 and Figure 4.4, for jets from dijet events.

4.3.3 Jet Energy Correction

The energy of a reconstructed jet is the sum of energies in all the calorimeter towers within the jet cone. So it may lose some energy due to out-of-cone showering. The reconstructed jet energy also depends on the response of the calorimeter, and an offset energy which is due to underlying events, pile-up, and

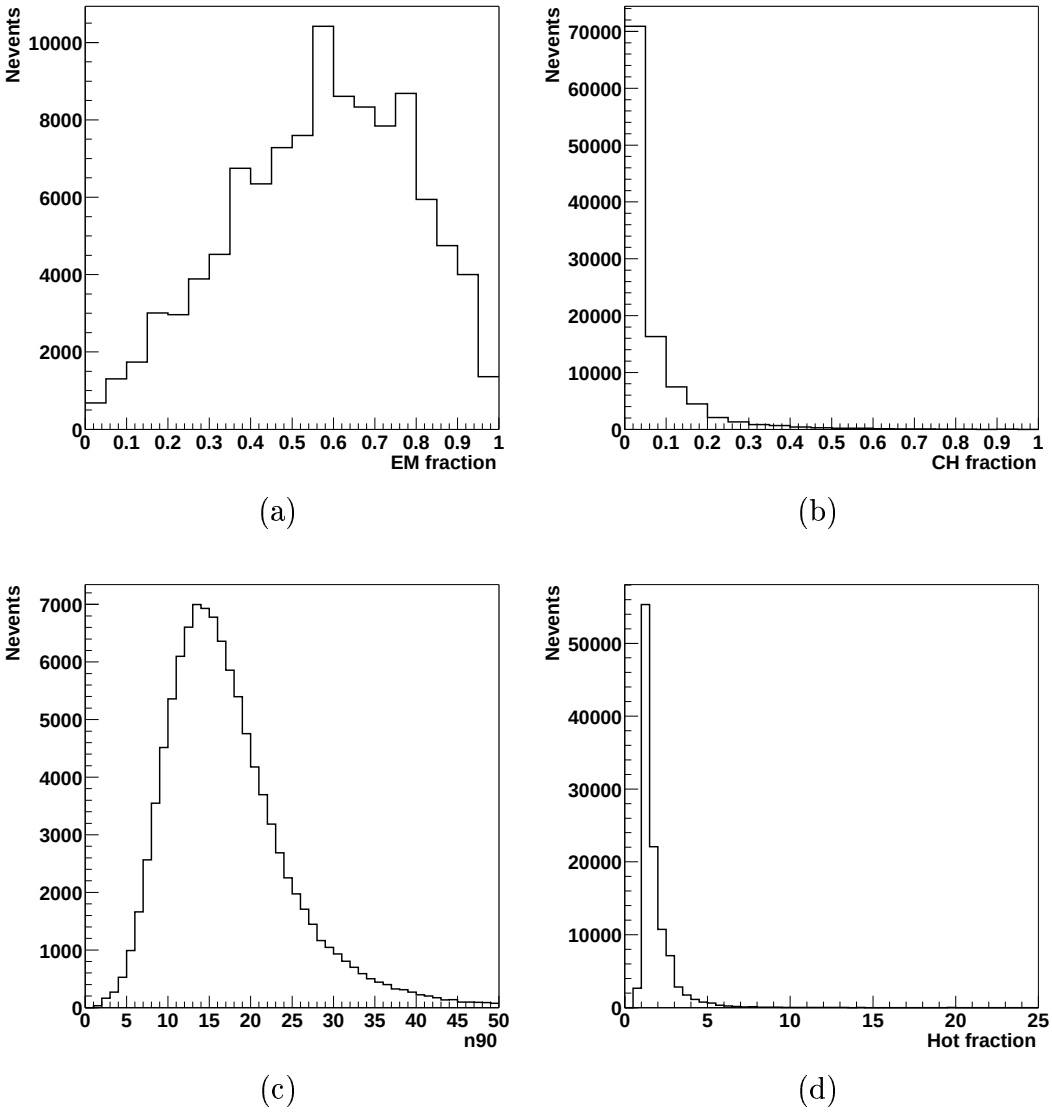
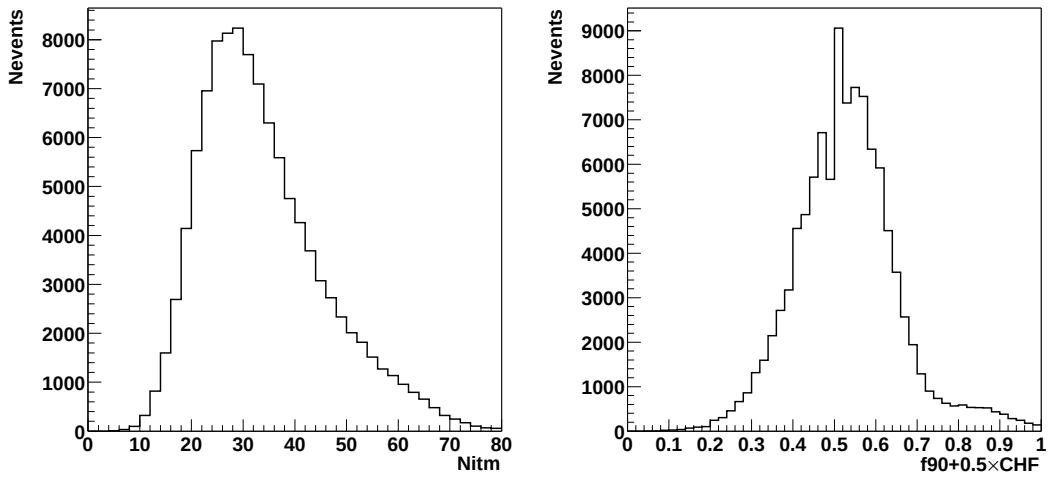


Figure 4.3: *Jet ID variable distributions for jets from dijet events: (a) EM-fraction, (b) Coarse-Hadronic-fraction, (c) n*90*, (d) Hot-fraction.*



(a) (b)

Figure 4.4: *Jet ID variable distributions for jets from dijet events: (a) nitm, (b)* $f90 + 0.5 \times CHF$.

uranium noise. The jet energy scale (JES) corrects the jet energy measured by the calorimeter (E_{jet}^{calor}), on average, back to the energy of the final state particle level jet ($E_{jet}^{particle}$) as follows [61]:

$$E_{jet}^{particle} = \frac{E_{jet}^{calor} - E_{offset}}{R_{jet} \cdot F_S} \tag{4.6}$$

where E_{offset} is the offset energy; R_{jet} is the calorimeter energy response to jets; F_S is the fraction of the jet energy showered inside the algorithm cone.

The offset term E_{offset} comes from: underlying events, electronic pile-up from the previous $p\bar{p}$ crossing, multiple $p\bar{p}$ interactions during the same beam crossing,

and noise from the natural radioactivity of the uranium absorbers in the calorimeter. It is measured from the transverse energy density (calculated from calorimeter cells) in minimum bias events [2] as a function of detector pseudorapidity η_{det}.

The jet energy response R_{jet} depends on jet cone size, jet energy and η_{det}. It is measured using conservation of p_T in photon + jet events. The photon energy is well measured, since the EM energy scale has been determined from $Z \rightarrow e^+e^-$ events using the known Z mass. By requiring the photon and the jet to be back to back, any p_T imbalance can be attributed to the mismeasurement of the jet energy.

Due to the fixed cone radius in the cone jet algorithm, the energy of particles showering outside the cone is not included in the energy measurement. The shower profile of the jet in photon + jet events is measured as the average energy density as a function of the distance from the jet center. The showering correction F_S is calculated as the ratio of the energy contained within the algorithm cone to the energy contained in a much larger cone (size 1.0 in the CC and 1.5 in the EC) where the energy density decreases to zero at the edge of the large cone.

The last correction applied to the jet energy adjusts the particle level jet energy $E_{jet}^{particle}$ to the energy of the original parton which initiates the jet, on average. The Monte Carlo jet events are used to compare the parton with the particle jet

[2] A *minimum bias* event refers to an event where a hard interaction has occurred. A *zero bias* event is an event where a particle beam crossing has occurred.

to obtain this correction as a parameterized function of jet E_T and η_{det}. Finally, it should be noted that the jet energy scale is different for the light quarks (u, d, s and c) than for b quarks [3], because the b quark decays semi-leptonically about 1/3 of the time (into an electron or muon with a neutrino and a quark).

4.4 Missing E_T ($\not\!\!E_T$) Measurement

The missing transverse energy, $\not\!\!E_T$, shows the momentum imbalance of an event. It can be the signature of particles escaping from the detector without interacting in it, such as neutrinos. By definition, $\not\!\!E_T$ is the opposite vector to the vector sum of transverse energies of all measured particles. In measurement, the transverse energies of all the calorimeter cells are added vectorially, except "hot cells" and cells with a high level of noise (the fraction of such cells is $< 1\%$ of the total number) [62]. Just as jet energies are modified by the jet energy scale correction, the $\not\!\!E_T$ is also corrected for the corresponding jets. Since muons are minimum ionizing particles, their presence may cause transverse energy imbalance. To correct this, the transverse momentum of each muon in an event is subtracted from $\not\!\!E_T$.

[3] The t quark is not considered here because it decays rapidly and leads to several jets.

4.5 Muon Reconstruction

Reconstruction of muons [63] starts with the conversion of hits from the muon detector into three dimensional position information. After the individual hits are found, track segments in each layer are formed by fitting groups of hits to a straight line. The tracking is done separately for segments before and after the toroid magnet. The segments are then matched and the momentum is determined from the measurement of the bend of the track when passing through the magnetized iron. A veto on cosmic muons is performed by requiring the time information recorded at each layer of the muon detector to be consistent with a muon coming from the interaction region.

The muon tracks are then extended to the point of the closest approach to the beam and their parameters are compared with those of central tracks by performing a global fit. By matching tracks in the central tracking system to those in the muon system, the momentum is corrected for the loss of energy in the calorimeter. Finally, the results of object reconstruction in the muon system are combined with the information provided by the central tracking system and the calorimeter to construct a good muon object.

4.6 Monte Carlo Simulation

The Monte Carlo simulation is used to simulate the collision of particle beams and to model the observation of the scattered collision products within an appropriately designed detector. We use Monte Carlo simulated samples to model the backgrounds and to study the signal. The simulation proceeds in two steps: event generation and detector response simulation.

4.6.1 Event Generation

Event generators are programs that generate the physics events of a $p\bar{p}$ interaction, and the output is usually a list of vertices and particles that are produced at those vertices. The PYTHIA [64] event generator is widely used at DØ for various physics processes.

The PYTHIA program is intended to generate complete events, in as much detail as experimentally observable ones, within the bounds of our current understanding of the underlying physics. The quantum mechanical variability between events in nature is simulated by Monte Carlo methods, to properly obtain both the average behavior and the fluctuations. The underlying physics, such as the strong interaction, is not understood well enough to give an exact description, therefore the program contains a combination of analytical results and various models. PYTHIA provides a simulation of several physics aspects of the interactions, in-

cluding hard and soft subprocesses, parton distributions, initial and final parton showers, beam remnants and underlying events [4], fragmentation and decays, etc. DØ uses PYTHIA for a wide range of Monte Carlo simulations of Standard Model processes as well as processes beyond the Standard Model, because PYTHIA is a good representation of the fragmentation model and the handling of color flows.

Other event generators used by DØ include the ALPGEN [65] generator. ALPGEN simulates the leading order matrix elements for $2 \to n$ (where n can be greater than 4) multiparton processes, while PYTHIA mainly simulates the $2 \to 2$ and $2 \to 1 \to 2$ processes. So ALPGEN is especially well suited to simulating boson(s)+jets (e.g., Z+jets, W+jets) production.

4.6.2 Detector Simulation

The next step in a simulation is to compute the detector response to the generated events. DØ has a fast simulation program and a full simulation program.

The fast simulation program [66] uses a simplified and parameterized description of the detector response. It takes objects (electrons, jets, etc.) from the Monte Carlo generators as input, smears their energy and spatial angles (η, ϕ),

[4]The beam remnants are left behind by each incoming beam paritcle, without taking part in the initial state radiation or hard scattering process. There are two kinds of underlying events. In hadron-hadron collisions, the composite nature of the two incoming beam particles implies the additional possibility that several parton pairs undergo separate hard or semi-hard scattering, which is called "multiple interactions". In high-luminosity colliders, it is also possible to have several collisions between beam particles in one and the same beam crossing, which is called "pile-up events".

and directly produces smeared 4-vectors for the objects. This program is fast [5], but a number of parameters in the program need to be tuned in order to give an accurate simulation. Also, it cannot produce the object identification quantities, such as the EM-fraction or H-matrix for EM objects, because of the lack of detector information.

The full simulation consists of two programs: D0GSTAR [67] and D0SIM [68]. D0GSTAR is based on the GEANT [69] program from CERN. This is a program that can describe the true geometry of a detector by building it up from a library of known shapes. It also has extensive knowledge of the interactions of particles with materials. D0GSTAR simulates the behavior of particles passing through a GEANT model of the DØ detector, and the output consists of hits in detector cells. The program D0SIM uses D0GSTAR output as input and does the digitization for each detector, pile-up (overlapping minimum bias events), and the addition of noise. The output events of the full simulation are in the same format as the raw data, and are passed to the reconstruction program RECO.

[5]The fast simulation is about 2000 times faster than the full simulation. For example, the full simulation takes about 1–4 minutes to simulate one event on a 2GHz CPU, while the fast simulation can produce 10k events in about ten minutes.

Chapter 5

Data Analysis

This chapter describes the search for first generation leptoquarks in the two electron and two jet final state. In this chapter we discuss the event selection, background, signal, and analysis optimization.

5.1 Data and Event Selection

The data were taken through the three level trigger and data acquisition system, and then processed by the event reconstruction software RECO. After reconstruction, the data are in the form of physics objects and their physical quantities, and the analysis is performed on that data.

5.1.1 Data Sets

We used the data collected by the DØ detector from September 2002 to June 2003 during the Tevatron Run II with $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. The data were reconstructed with the DØ reconstruction program RECO version p13.05.00 or p13.06.01 [1] [70]. We used the sample containing two EM objects selected by the DØ New Phenomena Physics group [71].

Two known bad runs with "hot" calorimeter towers are removed from the sample. Due to electronics component failure, a readout cell or tower may give fake signals in every event taken during that time, and we call it a hot cell or hot tower.

5.1.2 Trigger Selection

As described in Section 3.7, a trigger is an event filter with a set of criteria. There are EM triggers, jet triggers, muon triggers, etc. Due to the limitations of event recording rate, some triggers with relatively loose criteria are *prescaled*, i.e., only a fraction of events passing such trigger are recorded. Triggers are used to choose the events of interest, which in our case are those with two energetic electrons. The selected events were required to pass at least one of the following

[1] The majority of data samples have been reconstructed with RECO p13.06.01, which is a newer version than p13.05.00. There are some bug fixes in this newer version, but there are no observable differences for EM objects or jets between these two versions.

un-prescaled EM triggers:

- *EM_HI_SH*

  Level 1: at least one EM trigger tower (of size 0.2×0.2 in $\Delta\eta \times \Delta\phi$) having $E_T > 10$ GeV;

  Level 2: one EM candidate with $E_T > 12$ GeV;

  Level 3: one electron with $|\eta| < 3$ and $E_T > 20$ GeV, meeting loose criteria (see below) including a transverse shower shape requirement.

- *EM_MX_SH*

  Level 1: at least one EM trigger tower having $E_T > 15$ GeV;

  Level 3: one electron with $|\eta| < 3$ and $E_T > 20$ GeV, meeting loose criteria including a transverse shower shape requirement.

- *2EM_HI*

  Level 1: two calorimeter EM trigger towers with $E_T > 10$ GeV;

  Level 3: one electron with $|\eta| < 3$ and $E_T > 20$ GeV, meeting loose criteria.

- *2EM_HI_SH*

  Level 1: two calorimeter EM trigger towers with $E_T > 10$ GeV;

  Level 3: one electron with $E_T > 15$ GeV, meeting loose criteria including a transverse shower shape requirement.

The above mentioned loose criteria at Level 3 is a loose EM fraction cut. The trigger region covers up to $|\eta| < 2.4$.

The integrated luminosity for this data sample is estimated to be 130.4 ± 8.5 pb^{-1}, using tools provided by DØ Luminosity ID group [72]. Among the above data, the sample reconstructed by RECO p13.05.00 has an integrated luminosity of 26.1 pb^{-1}, and the sample of p13.06.01 has an integrated luminosity of 104.3 pb^{-1}. The systematic uncertainty of the integrated luminosity is 6.5%, including the uncertainties from the cross section of inelastic collisions, the efficiency and acceptance of the luminosity detector [73].

5.1.3 Offline Selection

From the data samples, we select events with two electrons and two jets. This is called the offline selection, where the name offline means that it is performed after the data are recorded. We use kinematic and geometric cuts to remove background events while maintaining a majority of signal events. And we use physics object quality cuts to remove fake physics objects in an event, e.g., the electron EM-fraction cut helps to distinguish real electrons from hadronic jets. The following cuts are imposed on the data sample:

(a) That there be two EM objects, each with $E_T > 25$ GeV in the calorimeter fiducial region. The fiducial region is $|\eta_{det}| < 1.1$ in the central calorime-

ter(CC), and $1.5 < |\eta_{det}| < 2.4$ in the endcap calorimeters (EC). This requirement eliminates the poorly instrumented regions of calorimeter ($1.1 < |\eta_{det}| < 1.5$) and the high $|\eta_{det}|$ regions which have no trigger coverage. We further require at least one of the EM objects be in the CC. The events that have both EM objects in the EC are rejected because the QCD background is much higher for that topology, as illustrated in Figure 5.1. For a leptoquark signal (mass from 120 to 280 GeV), only less than 2% events have EC-EC di-EM topology. An EM object is defined by the EM ID certification cuts version 4.0 [57]. It requires

- EM-fraction > 0.9
- isolation < 0.15
- H-matrix (8×8) $\chi^2 < 20$

(b) That at least one of the EM objects has a matched track in the central tracking detectors. The track matching requirement is defined in the EM ID certification as the probability of a global track χ^2 (including both spatial and E_T/p_T information) $> 10^{-2}$.

(c) That the event has at least two jets of cone size 0.5 with $E_T > 20$ GeV within region $|\eta_{det}| < 2.5$. The energy of the jet is corrected using the Jet Energy Scale certified version 4.1 [74]. Jets are defined by the Jet ID certification

cuts version 2.1 [60]. The requirements are

- $0.05 <$ EM-fraction < 0.95
- Coarse-Hadronic-fraction (CHF) < 0.4
- $n90 > 1$
- Hot-fraction < 10
- $f90 < 0.8 - 0.5 \times$CHF or CHF < 0.1

The $f90$ cut is largely motivated by the need to remove "fake" jets generated by calorimeter noise.

(d) Jets are separated from EM objects by a distance of jet cone size, i.e.,

$$\Delta R_{ej} \equiv \sqrt{(\Delta\eta_{ej})^2 + (\Delta\phi_{ej})^2} > 0.5$$

where $\Delta\eta_{ej}$ and $\Delta\phi_{ej}$ are distances between an EM object and a jet in η and ϕ, respectively.

(e) The invariant mass of the two EM objects is required to be outside of the Z-mass window, i.e., we require $M_{ee} < 80$ GeV or $M_{ee} > 102$ GeV. We call this the Z-veto cut. This cut removes the majority ($> 85\%$) of the $Z \to ee$ background.

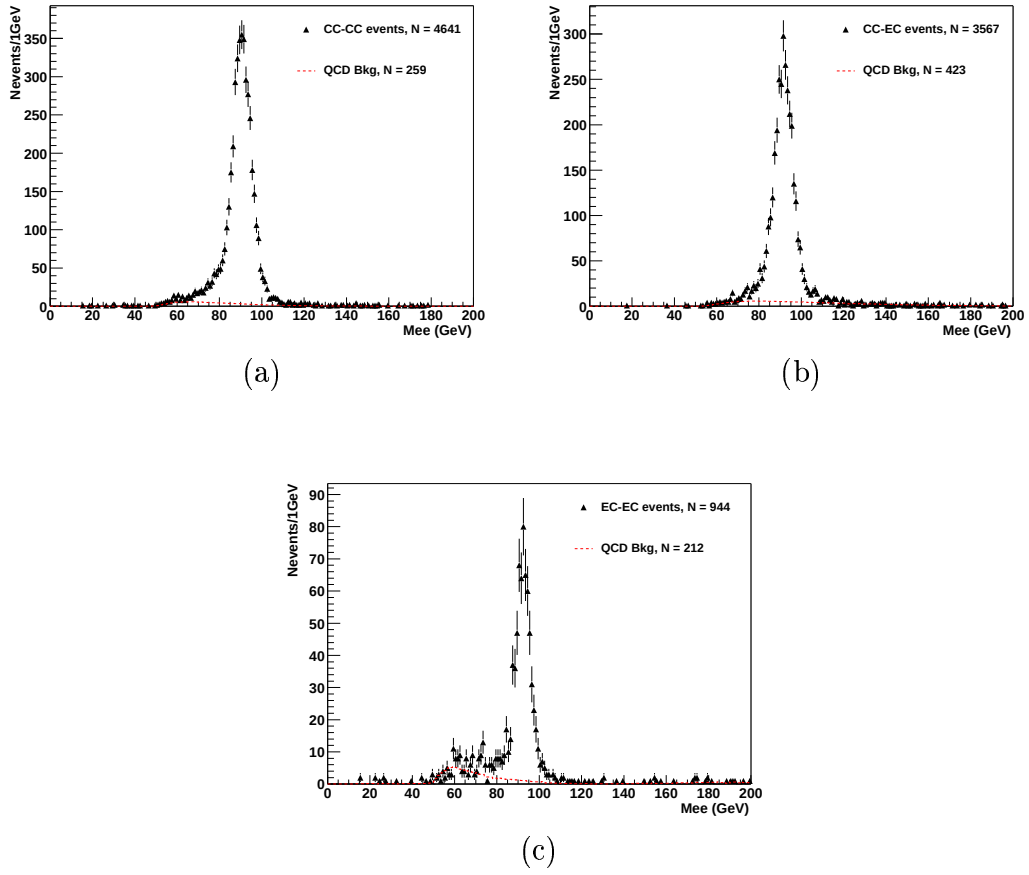


Figure 5.1: *The di-EM invariant mass distribution from data* (triangles) *and QCD background* (dashed line) *for different event topologies: (a) both EM objects in the CC, (b) one EM object in the CC and the other in the EC, (c) both EM objects in the EC.*

The number of events that pass these selection cuts is listed in Table 5.1. We call an EM object with a track match an electron (e), and an EM object without a track match a photon (γ). Since we require at least one of the EM objects with a track match, the events in our sample are either in the ee channel or in the $e\gamma$ channel.

Selection Criteria	Number of Events
di-EM sample	7439987
Pre-selection and trigger selection (Pre-selection requires loose di-EM with $E_T > 25$ GeV)	46521
Two EM objects (Certified EM ID, at least one EM has track match)	8208
At least two 0.5 cone jets ($E_T > 20$ GeV, certified Jet ID, $\Delta R_{ej} > 0.5$)	173
Z veto ($M_{ee} < 80$ GeV or $M_{ee} > 102$ GeV)	36

Table 5.1: *Number of events passing the selection criteria.*

5.2 Background

The major Standard Model backgrounds in the dielectron + dijet decay channel of the leptoquark pair are from 1) QCD multijet production where two of the jets are misidentified as two EM objects, 2) Z/Drell-Yan production, and 3) top pair production.

5.2.1 QCD Background

The QCD background is also known as the mismeasurement background. For QCD events with four or more jets, if two of the jets are misidentified as EM objects, the events may pass the offline cuts. Although the probability of two jets being misidentified as two EM objects is very small (in the order of 10^{-7}), this background cannot be ignored due to the large cross section of QCD production (which is 10^7–10^8 times larger than that of leptoquark production).

The traditional method [31] used to obtain the QCD background is first to estimate the probabilities of a jet being misidentified as an EM object (an electron or a photon), which are also called the "fake" rates (denoted as f_e and f_γ). Then the QCD background to the $eejj$ signal is estimated by applying the fake rates to the weighted number of 4-jet events in the QCD sample. The weight assigned to each event is the number of jet permutations which can be used for faking a pair of EM objects. The background in the ee and $e\gamma$ channels is calculated by multiplying the weighted number of events by f_e^2 and $2f_ef_\gamma$, respectively.

We use a similar method, but with the starting point of *very loose* EM objects instead of jets. These very loose EM objects are those which can be reconstructed by the EMRECO package, but are not required to pass any offline EM ID cuts. Because the QCD cross section greatly exceeds the cross section for genuine electron production, such loose EM objects are, in fact, mostly jets ($> 99.8\%$). The

criteria to select events containing these very loose EM objects are described in the following section. We define the fake rates based on a loose single-EM sample, and we estimate the QCD background by using a loose di-EM sample.

By applying this method on the very loose EM objects which are mostly jets, we actually work on a subset of the QCD jet sample. Thus this method gives the equivalent QCD background estimation as that using the traditional method. Moreover, we avoid the difficulties inherent in using the QCD jet sample, such as an event selection bias caused by the high jet trigger threshold, and a systematic uncertainty due to the jet energy scale uncertainties.

Misidentification Rate

We select the loose single-EM data sample for the misidentification rate calculation using the following criteria:

- The event should pass the 2EM_HI trigger.
- There is only one EM object. This requirement removes $Z \to ee$ events.
- The EM object should have $E_T > 25$ GeV and be in the detector fiducial region.
- The missing transverse energy $\not\!\!E_T < 15$ GeV. This requirement removes $W \to e\nu_e$ events.

Events selected in this way are mostly QCD multijet (dominated by dijet) events, where one of the jets is misidentified as an EM object. A small number of the events ($< 2\%$) are from the direct photon (γ+jets) process. The direct photon contamination is estimated using the cross sections for γ+jets and multijet processes as well as the probability of a jet being reconstructed as an EM object (see Appendix A.1.1).

We define the fake rate, i.e., the probability of a loose EM object passing the electron identification cuts, as

$$f_{em \to e_{hmx}} = N_{e_{hmx}}/N_{em} \tag{5.1}$$

$$f_{em \to e_{trk}} = N_{e_{trk}}/N_{em} \tag{5.2}$$

where N_{em} is the number of loose EM objects in the sample. $N_{e_{hmx}}$ is the number of fake electrons passing EM ID cuts, where we use the subscript hmx since the H-matrix (8×8) χ^2 cut is the decisive EM ID cut here. $N_{e_{trk}}$ is the number of fake electrons passing EM ID cuts and with a track match.

We choose the 2EM_HI trigger for the fake rate calculation since more than 96% of the events in our analysis data sample passed this trigger. The other triggers used to select the data sample have a shower shape requirement on the Level 3 electron, so they tend to have higher fake rates for the triggering EM

object. We account for this difference in the systematic uncertainty of the QCD background, which is estimated to be 1.5% (see Appendix A.1.2).

We expect the fake rates to be different between the CC and EC regions. The H-matrix is calculated using different cluster width parameters in the CC (in z) and the EC (in r) as defined in Section 4.2.2, so $f_{em\to e_{hmx}}$ is expected to be slightly different between the CC and EC. For $f_{em\to e_{trk}}$, since track matching in the CC contains an E/p requirement while the tracks in EC region do not have E/p information available, we expect a much lower fake rate in CC region. The fake rates as a function of the EM object E_T (in GeV) are shown in Figure 5.2, for CC and EC regions. We use straight lines to fit the fake rates and obtain the following functions

$$f^{CC}_{em\to e_{hmx}}(E_T) = 0.228 + 0.00063 \cdot E_T \quad (5.3)$$

$$f^{EC}_{em\to e_{hmx}}(E_T) = 0.276 + 0.00040 \cdot E_T \quad (5.4)$$

$$f^{CC}_{em\to e_{trk}}(E_T) = 0.00561 \quad (5.5)$$

$$f^{EC}_{em\to e_{trk}}(E_T) = 0.0176 + 0.00002 \cdot E_T \quad (5.6)$$

where $f^{CC}_{em\to e_{trk}}$ is a flat line independent of E_T. The other fake rates increase very slowly with E_T, which is similar to Run I results [75].

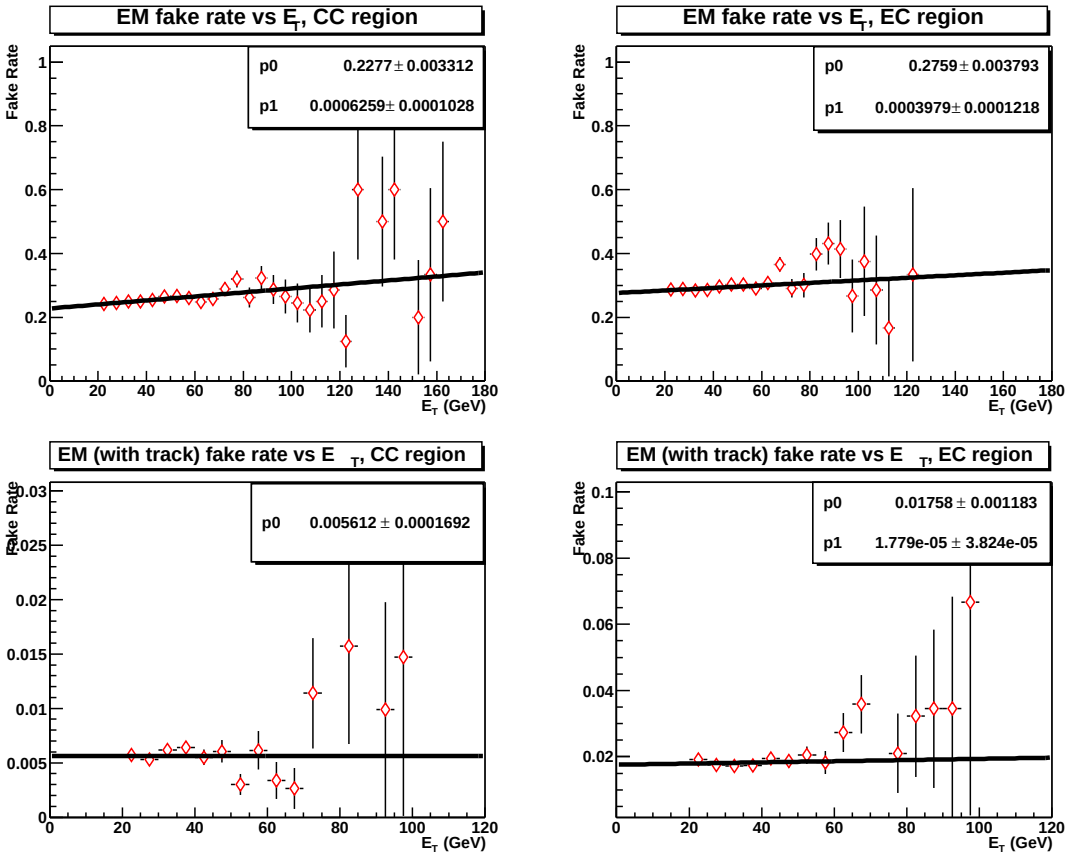


Figure 5.2: *Fake rate as a function of EM object E_T for passing EM ID cuts in CC* (top left) *and EC* (top right), *and for passing EM ID cuts with track match in CC* (bottom left) *and EC* (bottom right).

Misidentification Background

To estimate the QCD misidentification background to $eejj$ sample, we use the loose di-EM sample. We start from the sample containing two EM objects selected by the DØ New Phenomena Physics group and use the same triggers as we used to select the analysis data sample, so the QCD background sample has the same integrated luminosity as the data sample. For events passing our kinematic and geometric cuts, we loop over all possible permutations of the event that would give us 2 EM objects and 2 good jets, not requiring EM ID cuts on EM objects. The contribution of a QCD event to the background is then given by multiplying the number of combinations with the fake probability (p) of the event.

$$p = f^{(1)}_{em\to e_{hmx}} f^{(2)}_{em\to e_{trk}} + f^{(1)}_{em\to e_{trk}} f^{(2)}_{em\to e_{hmx}} - f^{(1)}_{em\to e_{trk}} f^{(2)}_{em\to e_{trk}} \tag{5.7}$$

where $f^{(1)}_{em\to e_{hmx}(e_{trk})}$ and $f^{(2)}_{em\to e_{hmx}(e_{trk})}$ are the fake rates for the first and the second EM object respectively.

This QCD background sample is dominated by multijet events, but it contains real electrons from Z(+jets) events. The contamination of Z(+jets) events is found to be 15% near the Z mass (80 GeV$< M_{ee} <$102 GeV), and negligible in other mass ranges. To obtain the di-EM invariant mass distribution (M_{ee}), we can remove the Z(+jets) events by using a side-band technique to re-estimate the

number of events within the Z mass range. For our analysis in $eejj$ channel, the Z-veto cut ensures that we do not have Z(+jets) events in the QCD background.

In the QCD background sample, a small fraction of the events ($< 4\%$) are from the direct photon (γ+jets) process, which is a part of a general multijet background. The method we used above to estimate the fake rates also accounts for direct photon background. If we estimate the QCD background for γ+jets and multijet events separately, the difference between it and our method is less than 1%. A detailed discussion on this is in Appendix A.1.1.

5.2.2 Z/Drell-Yan Background

The major physics background to $eejj$ signal is the Z/Drell-Yan process

$$q\bar{q} \rightarrow Z/\gamma^*(\text{+jets}) \rightarrow ee(\text{+jets})$$

Monte Carlo samples of the Z/Drell-Yan (DY) process are generated with PYTHIA, simulated through full detector simulation and reconstructed by RECO version p13.05.00. More than 200k events are used to ensure enough statistics after requiring two jets. The samples are generated in different mass ranges and joined together after being normalized to the same integrated luminosity as the data. The number of events and PYTHIA cross section of every Monte Carlo sample

are listed in Table 5.2. The cross sections from the PYTHIA generator are leading order (LO) ones. To include the next to leading order (NLO) and the next to next to leading order (NNLO), the LO cross sections need to be scaled by a K-factor of 1.3 [75][76].

Z/DY mass range (GeV)	Number of events	PYTHIA σ (pb)
20–60	25000	152
60–130	177750	183
130–250	10000	1.37
250–500	11000	0.115
500–up	5000	0.00466

Table 5.2: *Z/Drell-Yan $\rightarrow ee$ Monte Carlo samples in different mass ranges.*

To illustrate that we understand this source of background, we compare the Z/Drell-Yan Monte Carlo events with data. We select a dielectron data sample using the same trigger criteria and requiring two EM objects just as for the $eejj$ data sample, but we do not require any jet or Z-veto at this stage. We compare the data with Z/Drell-Yan Monte Carlo plus QCD background. The Z/Drell-Yan Monte Carlo samples are normalized to have the same integrated luminosity as the data, and are required to pass the same di-EM requirements. Since the Monte Carlo samples do not have a trigger requirement, and the EM ID efficiency and jet ID efficiency are not exactly the same for electrons in the Monte Carlo sample and in the data sample (see Section 5.6.2), we correct the Monte Carlo sample for the residual difference in efficiencies by multiplying the ratio of data efficiency

to Monte Carlo efficiency. We do not require track matching in the Z/Drell-Yan Monte Carlo sample, instead we apply the data track efficiency on the Monte Carlo sample to correct this. We know the energy resolution of electrons in the Monte Carlo does not match that in the data, so the energy of Monte Carlo electrons has been smeared by $4.7\% \times E$. The QCD background is obtained using the method described in Section 5.2.1, without jet requirement.

Figure 5.3 shows that the di-EM data (8208 events) are consistent with a Z/Drell-Yan plus QCD background (total 8618 ± 800 events). The numbers of events with errors are listed in Table 5.3. The systematic uncertainties are calculated in Section 5.7.2.

To evaluate the Z/Drell-Yan background contribution to the $eejj$ signal, we make the further requirement that there be two jets in the event. The energy of the Monte Carlo jet is corrected using the Jet Energy Scale (JES) certified version 4.1 [74]. As expected the PYTHIA Z/Drell-Yan Monte Carlo sample does not reproduce the jet multiplicity observed in data, since we know that RECO p13.05.00 and JES v4.1 need improvement for Monte Carlo jets [2]. We correct the PYTHIA Z/Drell-Yan Monte Carlo sample with a scaling factor of 1.3 per jet. This scaling factor is determined in the following way: We plot the spectrum of the

[2]The jet energy scale corrects the E_T of a jet according to its spatial position and measured energy. The jet multiplicity is counted for jets above certain E_T threshold (20 GeV), so it is sensitive to the jet energy scale.

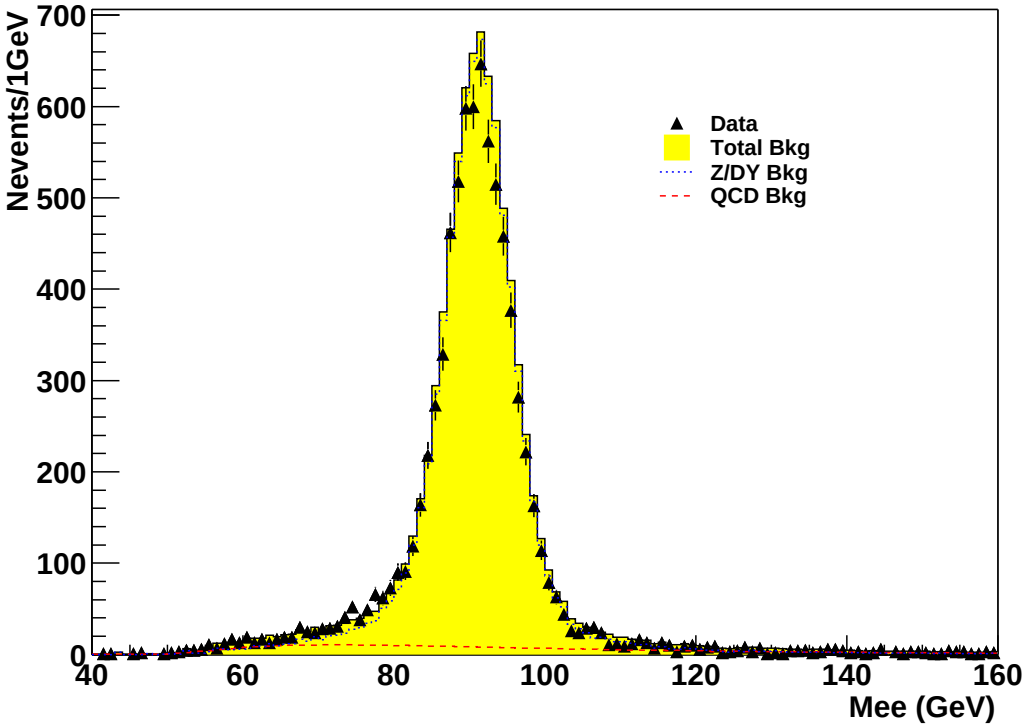


Figure 5.3: *The di-EM invariant mass distribution from data* (triangles) *compared with background* (solid line) *of Z/Drell-Yan Monte Carlo plus QCD fake.*

Data	8208
Expected events	8618 ± 800
Drell-Yan/Z	7936 ± 794
QCD fake	682 ± 96

Table 5.3: *Number of di-EM events from data and background (Z/Drell-Yan Monte Carlo and QCD fake). The error on the Z/Drell-Yan background is due to uncertainties on the EM ID efficiency and tracking efficiency. The error on the QCD background is mainly due to uncertainties on fake rates. The error coming from the luminosity is not included.*

number of jets (i.e., the first bin is the number of events containing 0 or more jets, the second bin is the number of events containing 1 or more jets, etc.) and fit the spectrum to an exponential function. The events which we are interested in are Z(+jets) events, i.e., 80 GeV $< M_{ee} <$ 102 GeV. Figure 5.4 shows the spectrum of the number of jets for data (QCD background subtracted) and for PYTHIA Z/Drell-Yan Monte Carlo. The number of events and the fitted exponential slopes are listed in Tables 5.4 and 5.5. The difference in the fitted exponential slopes between data and Monte Carlo gives the scaling factor of 1.3 per jet to correct the Monte Carlo sample.

As a cross check on the jet multiplicity, we generated a Z/Drell-Yan+2jets Monte Carlo sample using the ALPGEN generator and reconstructed it using RECO version p14.03.00 (note this is a newer version of the reconstruction program used earlier). We select events with two EM objects and at least two jets. We then correct the sample for the difference in the efficiencies between data and Monte Carlo. The energy of a jet in the Monte Carlo sample is corrected using the Jet Energy Scale version v4.2. The spectrum of the number of jets using the ALPGEN Z/Drell-Yan+2jets Monte Carlo sample is also plotted in Figure 5.4. The plot and the numbers in Table 5.5 show that the ALPGEN Z/Drell-Yan+2jets Monte Carlo has a jet multiplicity that better matches the data than the PYTHIA

Z/Drell-Yan Monte Carlo for Z + at least 2, 3, 4 jets [3]. In the following sections, we continue to use the PYTHIA Z/Drell-Yan Monte Carlo sample because it has more events at higher mass range (the PYTHIA sample is $> 10^3$ larger than the ALPGEN sample when $M_{ee} > 250$ GeV).

5.2.3 Top Background

The process of top pair production with both Ws (from top decay) decaying into electrons ($q\bar{q} \to t\bar{t} \to ee$) is another physics background to $eejj$ signal. This background becomes significant after the Z-veto cut and the optimization cut (Section 5.5). We use Monte Carlo samples of the process $t\bar{t} \to ee$ generated by PYTHIA and reconstructed by RECO version p13.05.00. The top mass used in Monte Carlo production is 175 GeV and about 22k events are used. The next to next to leading order (NNLO) cross section for top pair production at $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV is calculated to be about 7.5 pb, thus the cross section times branching ratio (1/81) is about 0.093 pb [77]. Similar to the Z/Drell-Yan background estimate, we require the top Monte Carlo events to have two EM objects and two jets, then we correct the sample for difference in efficiencies between data and Monte Carlo, and normalize the sample to the same integrated luminosity as data.

[3] Since the ALPGEN Monte Carlo sample is generated for Z/DY+2jets process, the numbers of events containing 0 jet and 1 jet are not available.

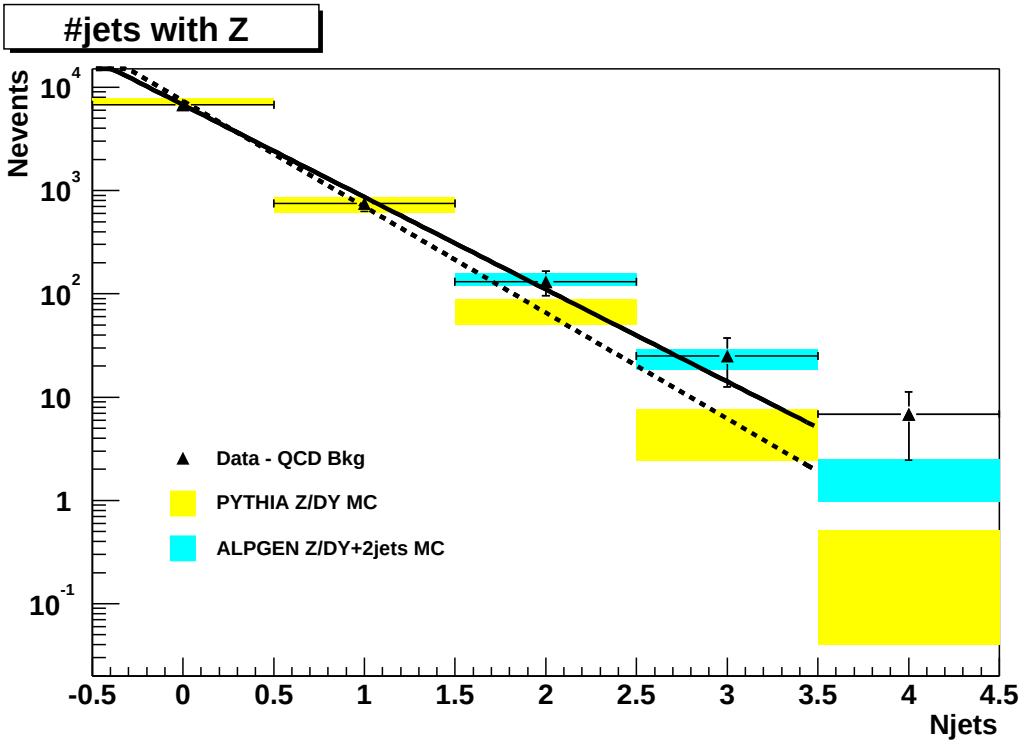


Figure 5.4: *Number of 0.5 cone jets associated with a Z in data* (triangles) *and Monte Carlo* (shaded area). *QCD background has been subtracted from data. The uncertainties are also shown. The PYTHIA Monte Carlo fits an exponential slope of* -2.36 ± 0.10 (dashed line), *while the data fit an exponential slope of* -2.10 ± 0.10 (solid line).

	Data	QCD fake	Data − QCD
≥0 jet	6944 ± 83	183 ± 26	6761 ± 87
≥1 jet	785 ± 121	33.4 ± 6.4	751 ± 121
≥2 jets	137 ± 35	5.8 ± 1.7	131 ± 35
≥3 jets	26 ± 13	0.87 ± 0.37	25 ± 13
≥4 jets	7 ± 4.4	0.15 ± 0.08	6.8 ± 4.4

Table 5.4: *Number of events with* $\geq n$ *jets within Z mass range for data and QCD fake background. The error on number of events includes statistical uncertainties and systematic uncertainties (see Table 5.15).*

	Data − QCD	MC Z/DY PYTHIA p13.05.00	MC Z/DY+2jets ALPGEN p14.03.00
≥0 jet	6761 ± 87	7048 ± 628	N/A
≥1 jet	751 ± 121	716 ± 126	N/A
≥2 jets	131 ± 35	68 ± 19	135 ± 19
≥3 jets	25 ± 13	4.8 ± 2.5	22.8 ± 5.0
≥4 jets	6.8 ± 4.4	0.27 ± 0.23	1.68 ± 0.73
Exponential slope from fitting	−2.10 ± 0.10	−2.36 ± 0.10	−2.09 ± 0.16

Table 5.5: *Number of events with ≥ n jets within Z mass range for data and Z/Drell-Yan Monte Carlo samples. The error on number of events includes statistical uncertainties and systematic uncertainties (see Table 5.15). The errors of exponential slopes are fitting errors.*

5.3 Signal Monte Carlo

Monte Carlo samples for a scalar leptoquark pair decaying into an $eejj$ final state were generated using PYTHIA, simulated with the full detector simulation and reconstructed by RECO version p13.05.00. Over 5,000 events per mass point were produced for LQ masses from 120 to 280 GeV in steps of 20 GeV. The cross sections from PYTHIA are leading order (LO). The next to leading order (NLO) theoretical cross sections of scalar leptoquark pair production are calculated using a program provided by the authors of [21].

Table 5.6 lists the number of events for each leptoquark Monte Carlo sample, the PYTHIA cross sections, and the NLO cross sections. The NLO cross sections are calculated for different renormalization/factorization scale, μ. By convention,

the value at $\mu = M_{LQ}$ is taken as the center value of NLO cross section, while the values at $\mu = 2M_{LQ}$ and $\mu = M_{LQ}/2$ are taken as the lower and higher bound of the theoretical cross section.

LQ mass (GeV)	Number of events	PYTHIA LO σ (pb)	NLO σ (pb)		
			$\mu = 2M_{LQ}$	$\mu = M_{LQ}$	$\mu = M_{LQ}/2$
120	5500	4.21	5.04	5.86	6.53
140	5000	1.79	2.09	2.40	2.65
160	5000	0.843	0.947	1.08	1.19
180	6000	0.423	0.458	0.522	0.572
200	6000	0.216	0.232	0.264	0.289
220	6500	0.116	0.122	0.138	0.150
240	8000	0.0637	0.0655	0.0746	0.0811
260	7700	0.0356	0.0358	0.0408	0.0443
280	7600	0.0200	0.0199	0.0228	0.0248

Table 5.6: *Number of events, PYTHIA (LO) cross section, and NLO cross section for scalar leptoquark Monte Carlo samples.*

5.4 Observed Data vs. Expected Background

After combining the backgrounds, we compare the observed data with the expected background when requiring two EM objects and two jets. Figure 5.5 shows the dielectron invariant mass (M_{ee}) distribution for the data and the combined background, where the Z-veto cut is not applied yet. The number of events in the data and the expected background before and after applying the Z-veto cut are listed in Table 5.7. The number of events expected from a leptoquark signal

at a mass of 220 GeV is obtained by using the signal cross section and efficiency (see Section 5.6.2) and is also listed in the table. The uncertainties are estimated in Section 5.7.1. We choose a leptoquark signal of mass 220 GeV for comparison since this is about the mass value most sensitive to our analysis (see Section 6.2.2).

From the above comparison, the observed data are consistent with the Standard Model background expectations. We do not see an excess of events, which would be expected from leptoquark decay. In the next section, we will optimize the event selection cuts, then compare the data with the background after optimization.

5.5 Optimization

To search for the phenomena of new physics, we focus on the regions where such production occurs. We make selection cuts to maximize the statistical significance of the excesses we would observe in the presence of the new physical processes. In other words, we look for a set of cuts determined by the characteristics of a specific new process to distinguish it from the Standard Model backgrounds. This is called the optimization.

For the leptoquark search, we know that a leptoquark pair decays into four energetic objects – two electrons and two jets in our case. Since the Standard

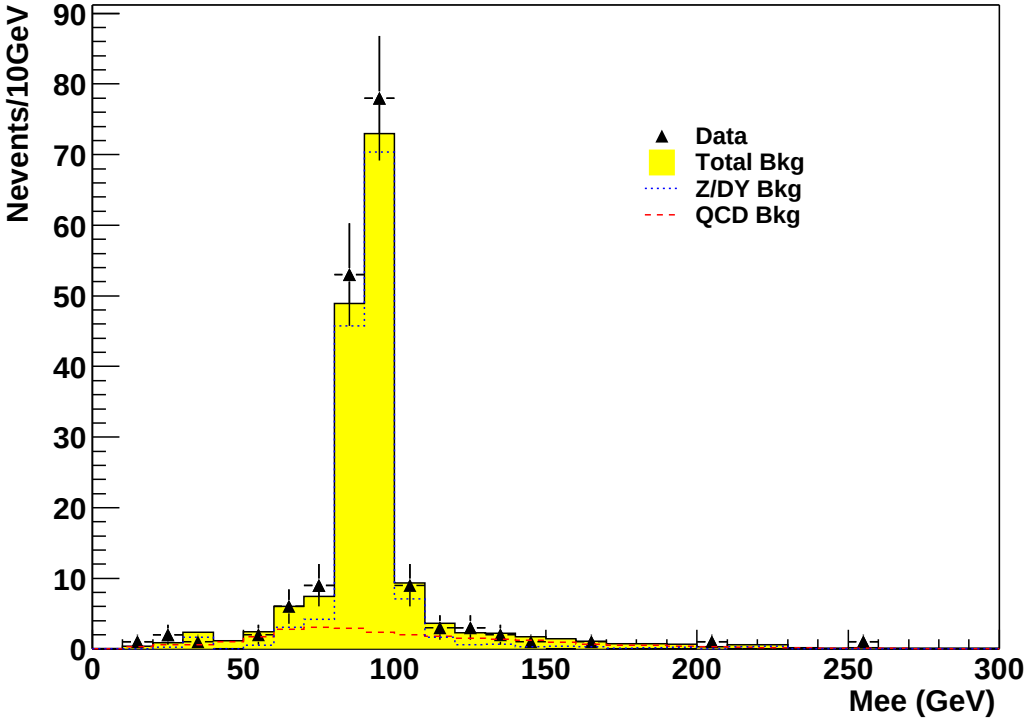


Figure 5.5: *The di-EM invariant mass distribution of the eejj events for data* (triangles) *compared with background* (solid line).

	Before Z-veto cut	After Z-veto cut
Data	173	36
Total background	164 ± 44	43.7 ± 11.2
Z/Drell-Yan	133 ± 38	19.5 ± 5.4
QCD fake	27.8 ± 8.0	22.0 ± 6.3
Top	2.67 ± 0.29	2.18 ± 0.24
LQ (m=220 GeV/c^2)	5.71 ± 0.39	5.39 ± 0.37

Table 5.7: *Number of eejj events from data and background (Z/Drell-Yan, QCD fake, and top), before applying Z-veto cut and after Z-veto cut.*

Model processes rarely have energetic many-object final states, it is reasonable to use the sum of energies of all four objects as a cut. Thus we define such a kinematic quantity, S_T, as the scalar sum of the transverse energies of the two electrons and two leading jets in an event:

$$S_T = E_T^{e1} + E_T^{e2} + E_T^{j1} + E_T^{j2}$$

S_T can serve as a powerful variable on which to cut in order to suppress the background while maintaining high efficiency for the leptoquark signal, especially for high mass leptoquarks. As shown in Figure 5.6, S_T distributions for the Standard Model background is well separated from the leptoquark signal.

We want to choose an optimized S_T cut in order to maximize a "significance" measure, such as $s/\sqrt{b}$, where s is the number of signal events and b is the number of background events. With large data samples, maximizing the quantity $s/\sqrt{b}$ corresponds to minimizing the probability that the expected background b might fluctuate up to or above the expected number of events $s + b$. When s and b are small numbers, such as in our case, we need to use a more accurate measure of significance.

As we do not observe any leptoquark signal, we can use a *limit setting significance* [78] to optimize the limit setting. Eventually what we want to do is

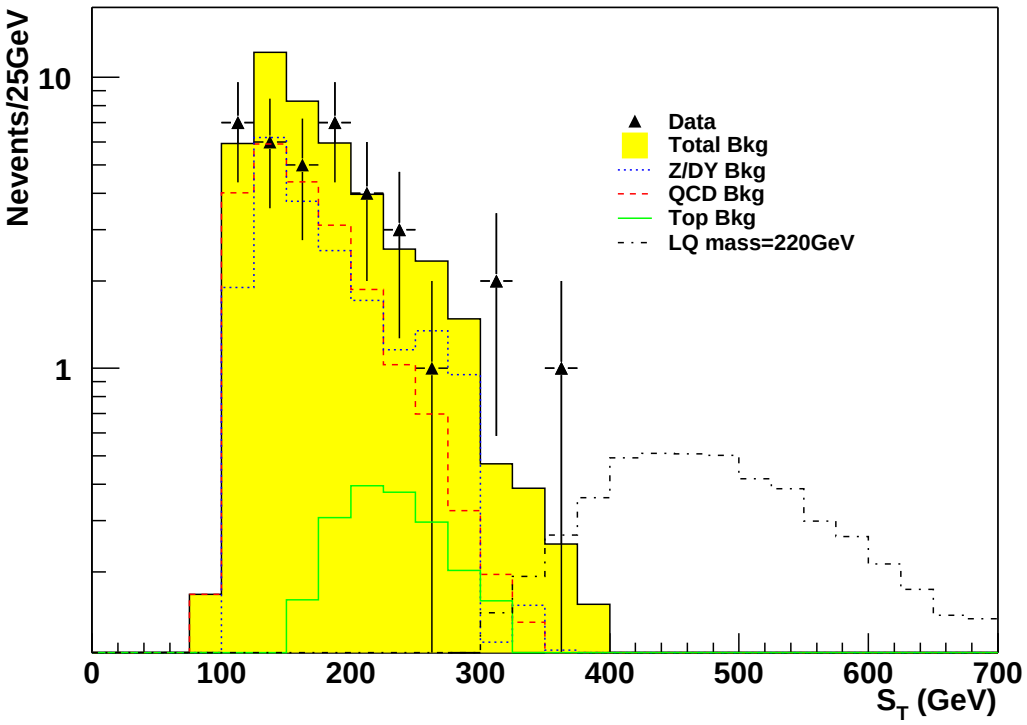


Figure 5.6: *The S_T distribution of the eejj events for data* (triangles)*, Standard Model background* (solid line)*, and leptoquark signal at a mass of 220 GeV* (dash-dot line)*. The Z-veto cut has been applied.*

to choose a set of cuts that maximizes our expected lower limit on the leptoquark mass. Since we are interested in maximizing a lower mass limit at the 95% confidence level, we can write the "average limit" as

$$\hat{M}^{95\%}(C) = \sum_{k=0}^{\infty} P(k,b) M^{95\%}(k,b,s(M)) \tag{5.8}$$

where $P(k,b)$ denotes the probability that k data events pass the set of cuts C,

and $M^{95\%}(k, b, s(M))$ is the mass limit we will set if we do in fact see k data events passing these cuts. We write $\hat{M}^{95\%}(C)$ to emphasize the dependence of $\hat{M}^{95\%}$ upon the choice of cuts C.

If we assume Poisson statistics, $P(k, b)$ is simply given by

$$P(k, b) = \frac{e^{-b} b^k}{k!} \tag{5.9}$$

The second piece of the summand in Equation 5.8, $M^{95\%}(k, b, s(M))$, depends on the assumed number of data events k, the number of expected background events b, and the number of expected signal events $s(M)$ passing our set of cuts C. The parameters b and $s(M)$ are functions of our cuts, and $s(M)$ is explicitly written as a function of the leptoquark mass M.

The 95% confidence level upper limit on the cross section may be found by employing Helene's formula [79], which states

$$0.95 = 1 - \frac{\sum_{i=0}^{k} (e^{-N} N^i)/i!}{\sum_{i=0}^{k} (e^{-b} b^i)/i!} \tag{5.10}$$

where k is the assumed number of events seen in the data, b is the expected number of background events, and N is the value of $s(M) + b$ (expected signal + expected background) with the property that the probability for b to fluctuate up to or above N is 5%. Given k and b, we may solve Equation 5.10 for N numerically.

And then we numerically solve

$$s(M) = N - b \tag{5.11}$$

for M to obtain the desired $M^{95\%}(k, b, s(M))$ referred to above.

The code to calculate our limit setting significance

$$\text{Limit Setting Significance} = \hat{M}^{95\%}(C) = \sum_{k=0}^{\infty} \frac{e^{-b}b^k}{k!} M^{95\%}(k, b, s(M)) \tag{5.12}$$

can be found at [80]. The limit setting significance in our case is the statistically expected leptoquark mass limit [4].

For every S_T cut, we obtain a limit setting significance, as shown in Figure 5.7. The optimized S_T cut is chosen as 375 GeV corresponding to the maximum limit setting significance. So from now on, we add a cut of $S_T > 375$ GeV.

5.6 Signal Efficiency

The efficiency for leptoquark signal sample is estimated in two steps. First, we use the leptoquark Monte Carlo samples to estimate the acceptance for kine-

[4] We set the leptoquark mass limits in the next chapter, but not using this method. By definition, $\hat{M}^{95\%}(C)$ given by this method is the average of the mass limits that would be obtained if the experiment were performed many times and the data contained no signal. Besides, the uncertainties of $s(M)$ and b are not taken into account here for simplicity, but it does not affect the significance comparison to select the best cut value.

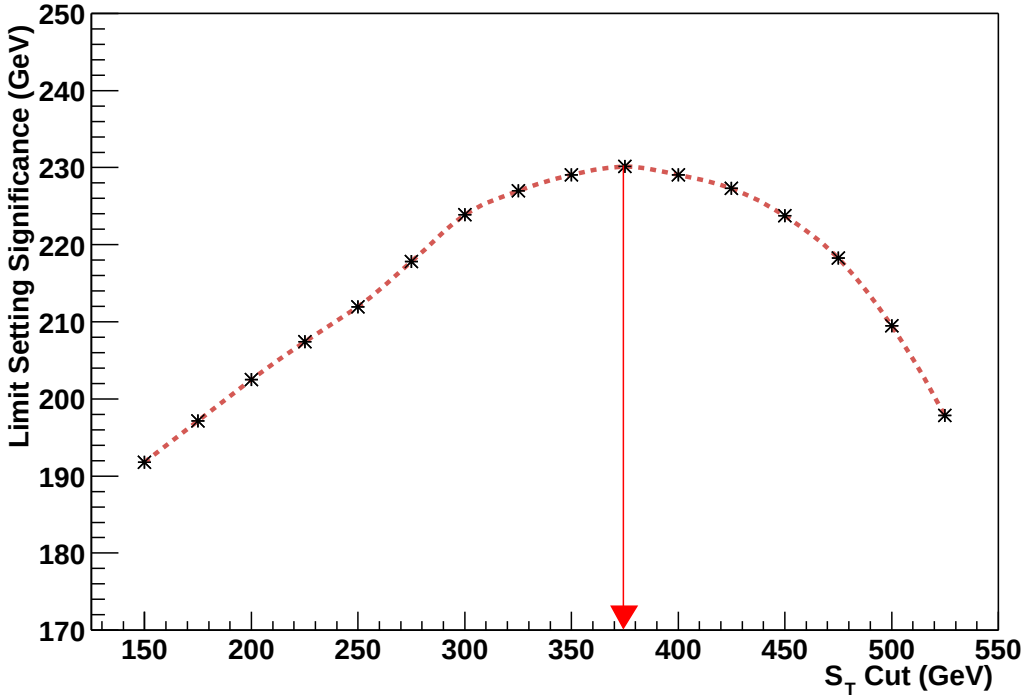


Figure 5.7: *The limit setting significance as a function of the S_T cut. The maximum limit setting significance corresponds to the optimized value of S_T cut (S_T > 375 GeV).*

matic and geometric cuts. Then we combine the trigger efficiency and object identification efficiencies measured from data to obtain the overall efficiency.

5.6.1 Acceptance

The kinematic and geometric acceptance for the leptoquark signal at each mass point is calculated with Monte Carlo samples which have gone through the full detector simulation and reconstruction. We do not require object identification cuts

on EM objects or jets, so that we do not mix the object identification efficiencies in the acceptance. We require the event to have two EM objects with $E_T > 25$ GeV, each of which can match an electron with similar E_T at the generator level within a window of size 0.2×0.2 in $\Delta\eta \times \Delta\phi$. The two EM objects should be in the CC or EC fiducial regions, and we reject EC-EC di-EM events as we did in the data selection. We further require the event to have at least two 0.5 cone jets with $E_T > 20$ GeV, within $|\eta_{det}| < 2.5$, and separated from EM objects by $\Delta R_{ej} > 0.5$. We then apply the Z-veto cut ($M_{ee} < 80$ GeV or $M_{ee} > 102$ GeV) and the S_T cut ($S_T > 375$ GeV) sequentially. The acceptance for each leptoquark Monte Carlo sample is listed in Table 5.8.

LQ mass (GeV)	Acceptance (%)			
	2 EM objects	2EM + 2jets	After Z-veto	After S_T cut
120	50.1	39.3	34.1	6.0
140	54.4	45.1	40.6	12.5
160	58.9	49.9	45.7	20.8
180	60.4	52.4	48.4	30.5
200	62.0	54.3	50.6	39.4
220	63.4	55.5	52.4	45.5
240	63.6	56.3	53.6	49.5
260	64.4	57.3	55.0	52.8
280	64.1	57.4	55.0	53.6

Table 5.8: *Kinematic and geometric acceptance for leptoquark signals.*

5.6.2 Efficiency

In this section, we first estimate the efficiency for objects, including the EM trigger efficiency, the electron identification efficiency, the electron track matching efficiency and the jet identification efficiency. Then we use these to obtain the overall efficiency for leptoquark signals.

Trigger Efficiency

The trigger efficiency for EM objects is calculated from $Z \to ee$ candidates using a "tag-and-probe" method. The idea of the tag-and-probe method is to tag one of the two EM objects from Z decay using tight EM ID requirements (including track match), and use the other EM object as a probe to calculate the trigger efficiency.

We estimate the trigger efficiency for an EM object passing the EM_HI_SH or EM_MX_SH trigger using the tag-and-probe method [81]. The Z candidate events are required to have a di-EM invariant mass between 50 GeV and 120 GeV. The highest E_T electron in the event is used as the tag electron, which is required to pass the EM ID cuts, to have a matched track and matched trigger objects at all trigger levels. The matched trigger objects are required to pass the trigger requirements at each level. The matching requirements are:

- Level 1: $\Delta\phi < 0.4$

- Level 2: $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$
- Level 3: $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$

At Level 1, we use $\Delta\phi$ instead of ΔR because only the ϕ information is available from the data. The second electron in the event (the probe electron) is required to have matching objects at all trigger levels using the same matching requirements, with the trigger objects passing the trigger's requirements. By calculating the fraction of probe electrons passing the trigger criteria, we extract the trigger efficiency. Figure 5.8 shows the trigger efficiencies as a function of the electron E_T for EM_HI_SH or EM_MX_SH triggers. In our di-EM data sample, the E_T weighted efficiencies (i.e., using the E_T spectrum as weights to average the efficiency) for CC and EC electrons are

$$\epsilon_{trig}^{CC} = 0.978 \pm 0.001$$

$$\epsilon_{trig}^{EC} = 0.955 \pm 0.001$$

The error of trigger efficiency is the binomial error[5].

The trigger efficiency for a di-EM event to pass the EM_HI_SH or EM_MX_SH

[5]The binomial error of $\epsilon = a/b$ is defined as $\delta_\epsilon \equiv \sqrt{a(b-a)/b^3} = \sqrt{\epsilon(1-\epsilon)/b}$.

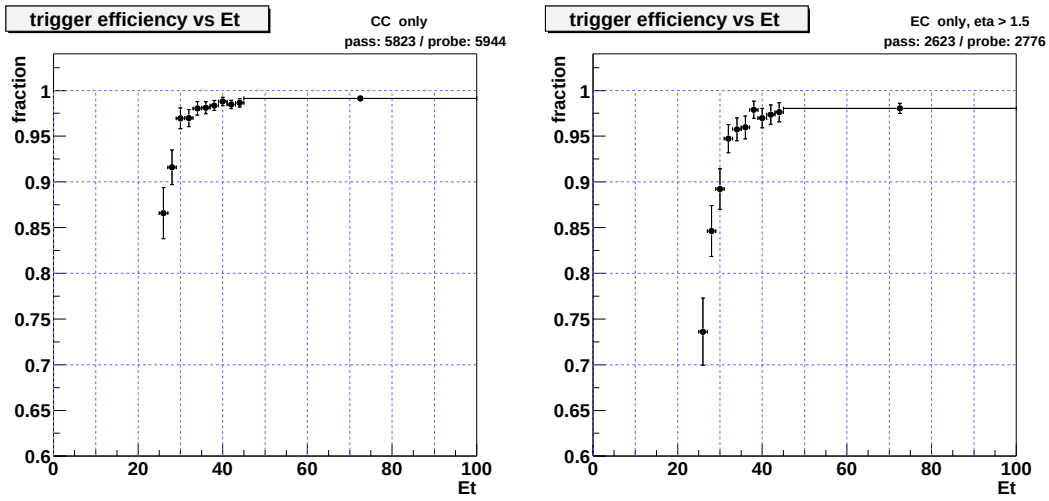


Figure 5.8: *Trigger efficiency as a function of electron E_T for the CC* (left) *and EC* (right) *regions.*

trigger is then computed for CC-CC and CC-EC topologies as

$$\epsilon_{trig}^{CC-CC} = 1 - (1 - \epsilon_{trig}^{CC})^2 \tag{5.13}$$

$$\epsilon_{trig}^{CC-EC} = 1 - (1 - \epsilon_{trig}^{CC})(1 - \epsilon_{trig}^{EC}) \tag{5.14}$$

For the di-EM data sample, about 60% consists of the CC-CC topology and 40% of the CC-EC topology. The trigger efficiency for the di-EM data sample is 99.9%. For the leptoquark signal sample, the two electrons have much higher E_T, and the trigger efficiency is close to 1.

The above method of trigger efficiency estimation only applies to single-EM triggers. In our di-EM data sample, about 99.4% of the events passed the EM_HI_SH

or EM_MX_SH triggers. For the other di-EM triggers used in the data sample selection (2EM_HI and 2EM_HI_SH), the trigger efficiency for a di-EM event can be estimated in the following way.

We choose unbiased di-EM events using the same criteria for the two EM objects, but require the events to pass the muon triggers. Such a data sample has no bias for the EM triggers, since muon triggers are based on the muon detector and are uncorrelated to EM triggers which are based on the calorimeter detector. The fraction of such events passing the EM triggers is the trigger efficiency. Using this method, we estimated the trigger efficiency for the 2EM_HI or 2EM_HI_SH triggers to be 92% ± 3% [6]. The binomial error is large due to the lack of events (about 100 events in total). Including this small fraction of events passing the di-EM triggers, the trigger efficiency for data sample is 99.9% ± 0.02%.

EM ID Efficiency

$Z \rightarrow ee$ candidate events are used to estimate the EM ID efficiency, i.e., the efficiency of an EM object to pass the EM ID cuts. The events are required to have two EM objects where one EM object ("tag") is required to satisfy all EM ID cuts and have a matched track, while the other EM object ("probe") only needs to have $E_T > 25$ GeV as well as the fiducial requirement. The efficiency is then

[6] We also estimated the trigger efficiency for the EM_HI_SH or EM_MX_SH triggers to be 98% ± 2% by this method. This result is consistent with that from the other method.

given by the number of probe electrons passing the EM ID cuts divided by the total number of probe electrons [81]:

$$\epsilon_{EMID} = \frac{2(tt) + (tp)}{2(tt) + (tp) + (tf)} \tag{5.15}$$

where

- tt = number of events where the probe electron passes the EM ID cuts and has a track match
- tp = number of events where the probe electron passes the EM ID cuts but fails the track match requirement
- tf = number of events where the probe electron fails the EM ID cuts

The invariant mass distributions are made for the combined $2(tt)+(tp)$ sample and the combined $2(tt) + (tp) + (tf)$ sample. The distributions are fit to a QCD background shape estimated from data plus a Z/Drell-Yan shape estimated from the Monte Carlo sample. The QCD background invariant mass distribution is estimated using events in which there are two EM objects that both have an H-matrix (8×8) $\chi^2 > 35$ (i.e., "reversing" the EM ID cut). The normalization of the QCD background and Z/Drell-Yan Monte Carlo is obtained by fitting them to the data in the two sideband regions [50, 75] and [105, 130] GeV on both sides of the Z peak.

After the estimated QCD background is subtracted, the EM ID efficiency is the ratio of the number of events from the $2(tt)+(tp)$ sample to the $2(tt)+(tp)+(tf)$ sample for the invariant mass in the Z mass range [80, 100] GeV. The invariant mass distributions of these samples restricted to for the CC-CC topology are shown in Figure 5.9 and are used to estimate the efficiency for CC electrons. The distributions restricted to the EC-EC topology are shown in Figure 5.10 and are used to estimate the efficiency for EC electrons. The EM ID efficiencies are

$$\epsilon_{EMID}^{CC} = 0.850 \pm 0.020$$

$$\epsilon_{EMID}^{EC} = 0.918 \pm 0.043$$

The error on EM ID efficiency includes both statistical and systematic errors (added in quadrature). The statistical error (binomial error) is 0.5% for CC and 0.9% for EC. The systematic error is estimated by varing the number of QCD background events subtracted within the fitting error. The systematic error is 2.3% for CC and 4.6% for EC.

Using the same method, the EM ID efficiencies for Monte Carlo electrons are obtained from the Z/Drell-Yan Monte Carlo sample (with no QCD background to subtract) to be $\epsilon_{EMID}^{CC}(MC) = 0.965 \pm 0.020$ in the CC and $\epsilon_{EMID}^{EC}(MC) = 0.968 \pm 0.037$ in the EC.

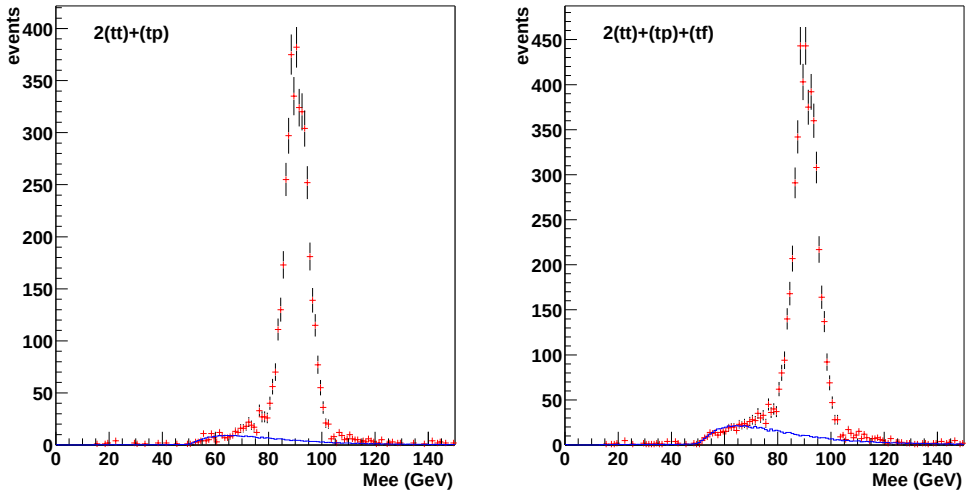


Figure 5.9: *The invariant mass distributions for 2(tt)+(tp)* (left) *and 2(tt)+(tp)+(tf)* (right) *samples with the CC-CC topology. The points are data. The line is the expected QCD background.*

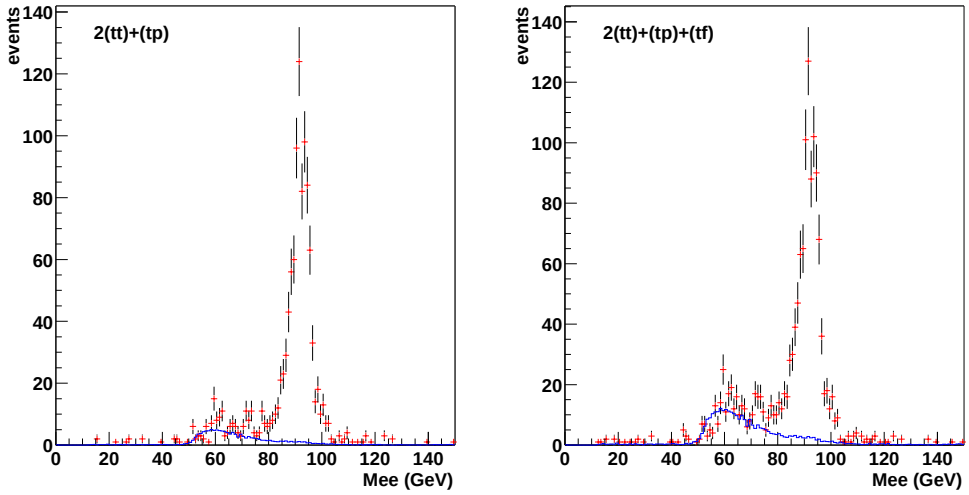


Figure 5.10: *The invariant mass distributions for 2(tt)+(tp)* (left) *and 2(tt)+(tp)+(tf)* (right) *samples with the EC-EC topology. The points are data. The line is the expected QCD background.*

EM Tracking Efficiency

The EM track matching efficiency is estimated using Z candidate events with both EM objects passing the EM ID cuts. Such events are divided into samples in which both EM objects have a track match (N_2), only one has a track match (N_1), or neither object has a matched track (N_0). The efficiency is thus given by:

$$\epsilon_{trk} = \frac{N_1 + 2N_2}{2(N_0 + N_1 + N_2)} \tag{5.16}$$

The invariant mass distributions of the combined $N_1 + 2N_2$ sample and the combined $N_0 + N_1 + N_2$ sample are shown in Figure 5.11 for the CC-CC topology and in Figure 5.12 for the EC-EC topology. The QCD background estimation is similar to that in the EM ID efficiency estimation. The number of events in each combined sample is counted between 80 and 100 GeV after subtracting the QCD background. The track matching efficiencies for CC and EC electrons are

$$\epsilon_{trk}^{CC} = 0.769 \pm 0.027$$

$$\epsilon_{trk}^{EC} = 0.558 \pm 0.042$$

The tracking efficiency is significantly lower in EC than in CC because the central fiber tracker only covers the range up to $|\eta| = 2$ (the silicon microstrip tracker

covers $|\eta| < 3$). The quoted error includes statistical and systematic errors. The statistical error is 0.6% for CC and 1.1% for EC. The systematic error is estimated by varying the number of subtracted background events within its fitting error, and is found to be 3.4% for CC and 7.4% for EC. By using the same method, the tracking efficiencies for Monte Carlo electrons in the Z/Drell-Yan Monte Carlo sample are $\epsilon_{trk}^{CC}(MC) = 0.825 \pm 0.020$ in the CC and $\epsilon_{trk}^{EC}(MC) = 0.716 \pm 0.031$ in the EC.

Jet ID Efficiency

The jet ID efficiency is estimated using the QCD dijet events. For jets in the data sample, we use two methods to estimate the jet ID efficiency, taking into account the dependence of the efficiency on jet detector η and jet E_T. The difference of the average jet ID efficiencies between the two methods is taken as the systematic error.

In the first method, we use the loose single-EM sample. As we discussed in Section 5.2.1, the loose single-EM sample we used to estimate the fake rate is dominated by QCD dijet events and γ+jet events. We use jets in such an EM+jet sample to calculate the jet ID efficiency.

The events are required to have only one EM object (no EM ID requirement)

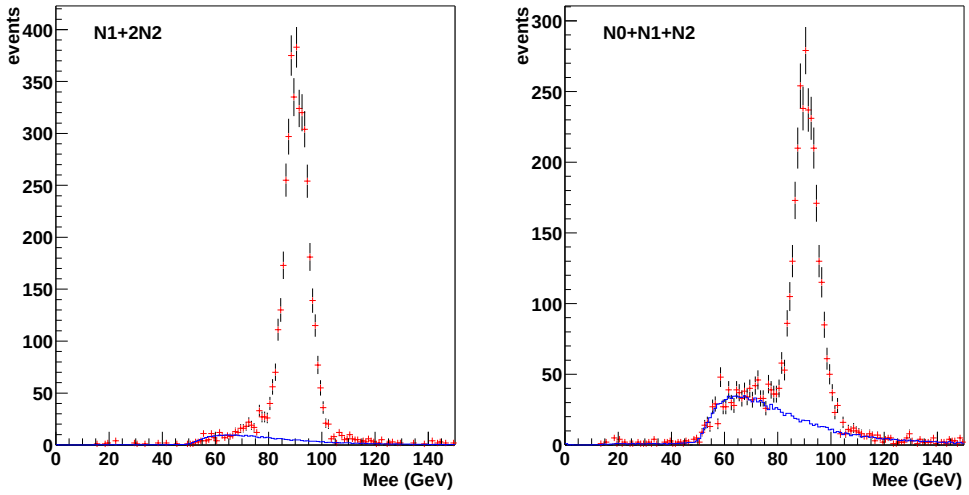


Figure 5.11: *The invariant mass distributions for* N_1+2N_2 (left) *and* $N_0+N_1+N_2$ (right) *samples with the CC-CC topology. The points are data. The line is the expected QCD background.*

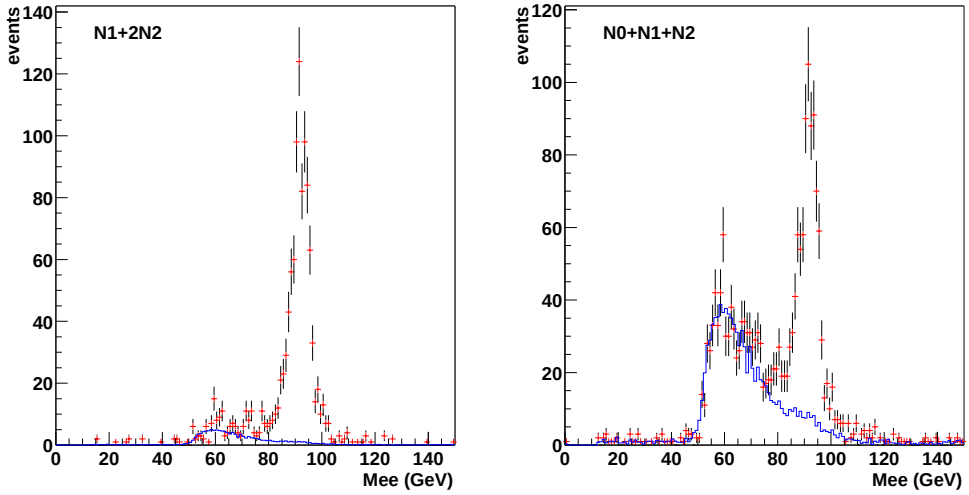


Figure 5.12: *The invariant mass distributions for* N_1+2N_2 (left) *and* $N_0+N_1+N_2$ (right) *samples with the EC-EC topology. The points are data. The line is the expected QCD background.*

and one jet, with the EM object and the jet back-to-back, i.e., $|\Delta\phi_{ej} - \pi| < 0.1$ [7]. The number of jets passing the jet ID cuts divided by the total number of jets is taken as the jet ID efficiency. Figure 5.13 shows the jet ID efficiency with respect to jet E_T and to jet detector η (η_{det}). The jet ID efficiency is clearly dependent on η_{det}, and it is lower in the ICD region $0.8 < |\eta_{det}| < 1.4$ which is the gap between the CC and EC calorimeters. The η_{det} weighted average value of efficiency is 0.952.

The other method we have used is to choose QCD dijet events from the data by requiring two jets in the event, where one of them (the "tag" jet) passes jet ID requirements and the two jets are back-to-back ($|\Delta\phi_{jj} - \pi| < 0.1$). We use the other jet in the event as a "probe" to estimate the jet ID efficiency. As shown in Figure 5.14, such a sample has more events at high E_T than the first method, and the jet ID efficiency is found to decrease as the jet E_T increases. We find an average efficiency of 0.969, which is close to that found from using the loose single-EM sample, and the difference between the two (0.017) is taken as the systematic error. To take into account the effect of decreasing efficiency at high E_T, we fit the plot of efficiency vs E_T for $E_T > 70$ GeV with a straight line to obtain the slope $s = -0.00084$ GeV^{-1}, and we take the efficiency for a jet with $E_T > 70$ GeV

[7] We only impose this requirement on $\Delta\phi$ but not on $\Delta\eta$ because the QCD dijet event can be boosted in the z direction, however the total momentum in x–y plane is close to 0. $|\Delta\phi - \pi| < 0.1$ is a tight cut on $\Delta\phi$, since the half-width of the $\Delta\phi$ distribution is about 0.3.

as

$$\epsilon_{jetID}(E_T) = \epsilon_{jetID}(\eta_{det}) + s \cdot (E_T - 70)$$

where $\epsilon_{jetID}(\eta_{det})$ is the jet ID efficiency at the corresponding η_{det}. We choose E_T of 70 GeV as the starting point to correct the efficiency because the plot is flat for efficiencies below 70 GeV.

When we estimate the overall efficiency for leptoquark signals, we use the jet ID efficiency as a function of the jet η_{det} and E_T. For uncertainty estimation purposes, the average jet ID efficiency weighted by η_{det} is

$$\bar{\epsilon}_{jetID} = 0.960 \pm 0.017$$

The error includes a statistical error of 0.07% and a systematic error of 1.8%. The systematic error is taken from the efficiency difference between the loose single-EM sample and the QCD dijet sample.

Finally, we estimate the jet ID efficiency for Monte Carlo jets in the QCD dijet Monte Carlo sample, using the same method. The Monte Carlo jet ID efficiency with respect to jet E_T and η_{det} is shown in Figure 5.15. The η_{det} weighted efficiency is $\bar{\epsilon}_{jetID}(MC) = 0.966 \pm 0.016$. The error includes a statistical error of 0.14%, and a systematic error of 1.6% which is estimated as the difference of efficiencies between the QCD dijet Monte Carlo sample and a γ+jet Monte Carlo sample.

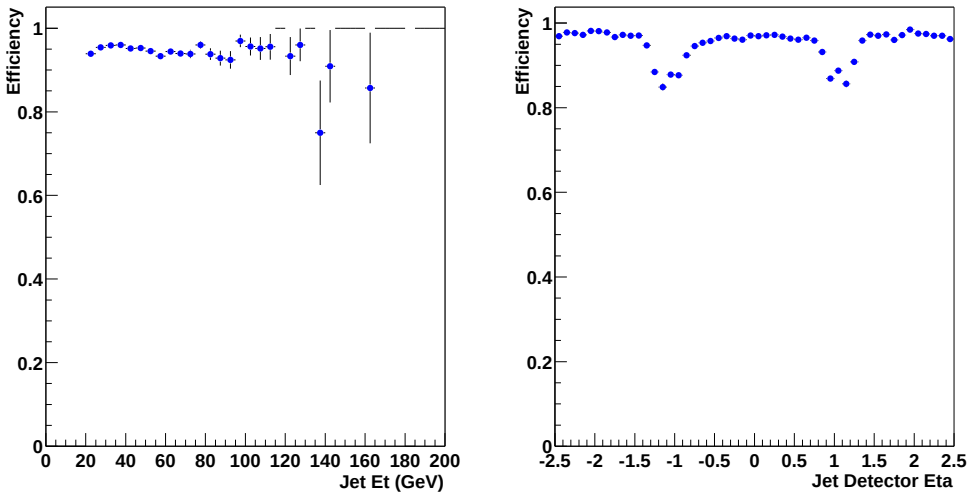


Figure 5.13: *The jet ID efficiency with respect to jet* E_T (left) *and to jet* η_{det} (right), *for jets in loose single-EM data sample.*

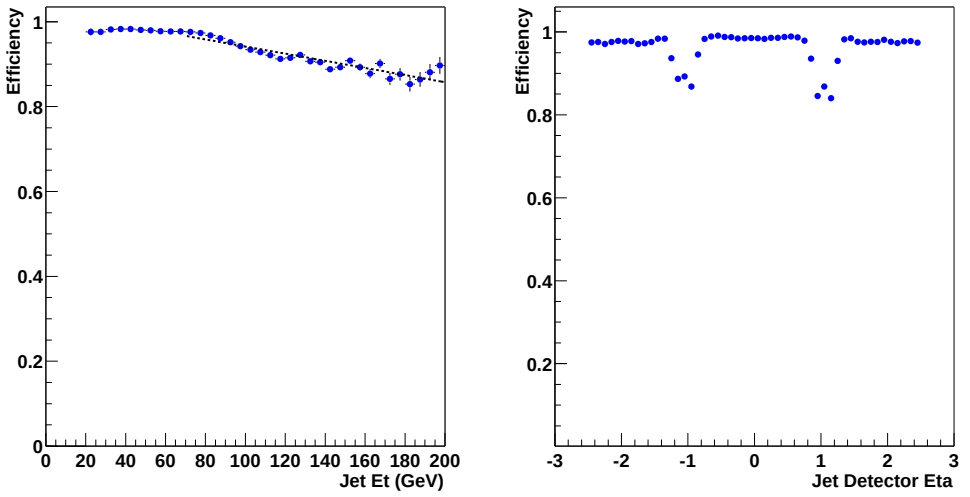


Figure 5.14: *The jet ID efficiency with respect to jet* E_T (left) *and to jet* η_{det} (right), *for jets in QCD dijet data sample. The dashed line fits the efficiency with a slope of -0.00084* GeV^{-1}.

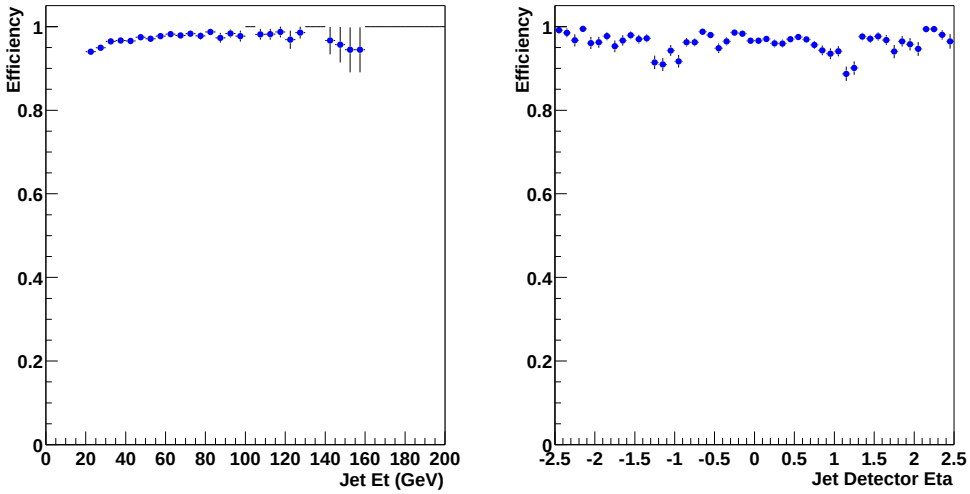


Figure 5.15: *The jet ID efficiency with respect to jet* E_T (left) *and to jet* η_{det} (right), *for jets in QCD dijet Monte Carlo sample.*

Overall Efficiency

The efficiencies per object, including the EM trigger efficiency, the EM ID efficiency, the EM tracking efficiency and the jet ID efficiency are summarized in Table 5.9. We list the efficiencies for EM objects in the CC and the EC separately, and the average value of jet ID efficiency.

	CC	EC
EM trigger	0.978 ± 0.001	0.955 ± 0.001
EM ID	0.850 ± 0.020	0.918 ± 0.043
EM track	0.769 ± 0.027	0.558 ± 0.042
Jet ID	0.960 ± 0.017	

Table 5.9: *Efficiencies for EM object and jet.*

A signal event consisting of a leptoquark pair decaying into $eejj$ has either the dielectron CC-CC topology or the dielectron CC-EC topology. The event selection efficiencies for different topologies are

$$\begin{aligned}\epsilon^{CC-CC} &= [1-(1-\epsilon^{CC}_{trig})^2]\cdot\epsilon^{CC}_{EMID}\cdot\epsilon^{CC}_{EMID} \\ &\quad\cdot[1-(1-\epsilon^{CC}_{trk})^2]\cdot\epsilon_{jetID}(j1)\cdot\epsilon_{jetID}(j2) \qquad (5.17)\\ \epsilon^{CC-EC} &= [1-(1-\epsilon^{CC}_{trig})(1-\epsilon^{EC}_{trig})]\cdot\epsilon^{CC}_{EMID}\cdot\epsilon^{EC}_{EMID} \\ &\quad\cdot[1-(1-\epsilon^{CC}_{trk})(1-\epsilon^{EC}_{trk})]\cdot\epsilon_{jetID}(j1)\cdot\epsilon_{jetID}(j2) \qquad (5.18)\end{aligned}$$

where $\epsilon_{jetID}(j1)$ and $\epsilon_{jetID}(j2)$ are the jet ID efficiencies for the first and second jets (dependent on the jet η_{det} and E_T).

The overall efficiency for a signal sample is estimated as the sum of the above event efficiency (using either ϵ^{CC-CC} for CC-CC event or ϵ^{CC-EC} for CC-EC event) for all the events which passed the kinematic and geometric requirements, then divided by the total number of events. Ideally, if the event selection efficiency is the same for every event, then the overall efficiency is just the product of the acceptance and the selection efficiency. Table 5.10 lists the acceptance and overall efficiency for leptoquark signals of different masses after all the cuts, including the Z-veto cut and the S_T cut. The errors quoted are calculated in the next section.

LQ mass (GeV)	Acceptance (%)	Overall efficiency (%)
120	6.0	3.4 ± 0.5
140	12.5	7.2 ± 1.1
160	20.8	11.9 ± 1.6
180	30.5	17.4 ± 2.1
200	39.4	22.5 ± 2.1
220	45.5	26.0 ± 2.1
240	49.5	28.3 ± 2.1
260	52.8	30.2 ± 2.1
280	53.6	30.7 ± 2.1

Table 5.10: *Acceptance and overall efficiency of leptoquark signals after all the cuts (including Z-veto and S_T cuts).*

5.7 Systematic Uncertainties

This section discusses the systematic uncertainties for the signal and each of the backgrounds. For example, when we estimate the efficiency of the signal or the number of events from the backgrounds, we apply the jet energy scale corrections, then the systematic error is induced from the uncertainties of the jet energy scale. In the following, we will estimate the uncertainties from every systematic source, and obtain the combined errors for the signal and the backgrounds.

5.7.1 Systematic Uncertainties on Signal Efficiency

The systematic errors on the overall efficiency for leptoquarks include uncertainties in the: trigger efficiency; particle identification efficiencies; jet energy scale; and effect of parton fragmentation in the signal modeling. The statistical

uncertainty on the signal acceptance is less than 0.8%, which is the binomial error, given that the number of events in every signal Monte Carlo sample is greater than 5000.

The EM ID efficiency error of a di-EM event can be obtained for the CC-CC and CC-EC topologies separately, using the EM ID efficiency uncertainties per EM object (listed in Table 5.9). Then the efficiency error of a signal sample can be estimated by combining the CC-CC and CC-EC event error according to the fraction of events for each topology. For all leptoquark signal samples from a mass of 120 GeV to 280 GeV, that fraction is estimated to be about the same, namely 80% CC-CC events and 20% CC-EC events. The error due to the EM ID efficiency uncertainty for the signal sample is 4.8%.

The error due to uncertainties in the EM track matching efficiency is also calculated for the CC-CC and CC-EC topologies, using the efficiency formulas (Equations 5.17 and 5.18). Taking the CC-CC and CC-EC event fractions, we estimate the error due to the EM track matching efficiency to be 1.8% for our signal sample.

The signal efficiency error due to uncertainties in the EM trigger efficiency is estimated to be negligible (0.02%). The error on signal efficiency due to the uncertainty in jet ID efficiency is simply twice the jet ID efficiency error per jet, which is 3.6%. The above errors are due to the uncertainties in object identification

efficiencies.

The uncertainty due to the choice of structure functions is obtained by comparing the acceptance for the signal samples generated with PYTHIA using different parton distribution functions (PDF). We use the parton distribution functions of CTEQ4L, CTEQ3L, and GRV98LO [82] for comparison, since they are all the leading order ones which are suitable for the leading order generator PYTHIA. The uncertainty on each leptoquark sample is listed in Table 5.11. We use the average error value of 2.1%.

The uncertainty due to the jet energy scale is estimated by varying the jet energy scale by plus or minus one standard deviation. The standard deviation of jet energy scale for a single jet is given by the Jet Energy Scale software [74], and it includes both systematic and statistical errors. The standard deviation for the jet energy scale is 6%–10% per jet for $|\eta| < 1.2$, and 10%–18% per jet for $|\eta| > 1.2$ (the variation depends on the jet position and E_T).

We vary the jet energy scale by plus or minus one standard deviation, then re-calculate the signal acceptance. The difference in acceptance is taken as the error due to the jet energy scale uncertainty. This error is sensitive to the jet E_T spectrum, so it is sample dependent. For leptoquark signal samples, we estimate the error of jet energy scale on the final samples which pass all kinematic and geometric cuts including the Z-veto and S_T cuts. The errors for leptoquark samples

of different masses are listed in Table 5.11. For comparison, the jet energy scale errors before the $S_T > 375$ GeV cut are also listed in the table.

The error due to the jet energy scale uncertainty increases after the S_T cut, especially for low mass leptoquark samples, because the S_T cut effectively chooses events with higher E_T jets. For a low mass leptoquark sample, e.g., leptoquark mass = 120 GeV, the jet E_T distribution peaks at about 60 GeV. A high S_T cut of 375 GeV mainly selects events with jet E_T greater than 90 GeV, so a small variation in jet energy scale results in a large difference in the number of selected events. However, for high mass leptoquarks, e.g. mass = 280 GeV, the jet E_T spectrum peaks around 140 GeV. The S_T cut can keep the majority of such leptoquark events, so the variation in jet energy scale has a much smaller effect on acceptance.

The errors on the signal overall efficiencies from various error sources are summerized in Table 5.12. The total error for each signal sample is obtained by adding the uncertainties of all sources in quadrature. The luminosity uncertainty is not included into the total signal error in the table, because in the next chapter we use the signal efficiency uncertainty and the luminosity uncertainty separately.

LQ mass (GeV)	PDF error	JES error before S_T cut	JES error after S_T cut
120	2.2%	3.3%	14.5%
140	1.8%	2.9%	14.0%
160	2.4%	2.0%	12.1%
180	1.7%	1.8%	10.1%
200	1.6%	1.4%	6.8%
220	2.5%	1.2%	4.5%
240	2.1%	1.2%	3.3%
260	2.4%	1.0%	2.5%
280	2.3%	1.0%	1.9%

Table 5.11: *Systematic uncertainties for leptoquark signals due to the choice of parton distribution function (PDF), and the jet energy scale (JES) uncertainties before and after the $S_T > 375$ GeV cut.*

Source of Systematics	Efficiency Uncertainty		
	$M_{LQ} = 120$	$M_{LQ} = 220$	$M_{LQ} = 280$
EM ID	4.8%	4.8%	4.8%
EM track	1.8%	1.8%	1.8%
Jet ID	3.6%	3.6%	3.6%
Jet energy scale	14.5%	4.5%	1.9%
Parton distribution function	2.1%	2.1%	2.1%
Monte Carlo statistics	0.8%	0.8%	0.8%
Total	16.0%	8.0%	6.9%

Table 5.12: *Signal efficiency uncertainties on leptoquark samples passing all the cuts, including the Z-veto and $S_T > 375$ GeV cuts.*

5.7.2 Systematic Uncertainties on Background

Systematic Uncertainties on the QCD Background

The systematic error on the QCD background is dominated by the uncertainties in the fake rate and the jet energy scale. The statistical error on QCD background sample is 4.2%.

The fake rate errors are listed in Table 5.13, including the error from fitting the fake rate vs E_T plot and the systematic error. The systematic error is estimated from the difference in the fake rates calculated by varying: the $\not\!\!E_T$ cut (from 15 GeV to 10 GeV); the plot bin size (from 5 GeV to 2 GeV and 10 GeV) before fitting; the loose single-EM sample used (for different data collection periods).

	Fitting error	Systematic error	Total error
$f^{CC}_{em\to e_{hmx}}$	1.5%	4.3%	4.6%
$f^{EC}_{em\to e_{hmx}}$	1.4%	4.5%	4.7%
$f^{CC}_{em\to e_{trk}}$	3.0%	7.0%	7.6%
$f^{EC}_{em\to e_{trk}}$	6.7%	7.9%	10.4%

Table 5.13: *Fake rate uncertainties.*

The error on the QCD background due to fake rate uncertainties is 11%. The error due to jet energy scale is 25% for $eejj$ sample before the $S_T > 375$ GeV cut, and 26% after the S_T cut. Other errors due to direct photon contamination, trigger

selection, and the background estimation algorithm are calculated in Appendix A.1.1 to A.1.3. The errors on the QCD background are summarized in Table 5.14.

Source of systematics	Uncertainty di-EM sample	Uncertainty di-EM + dijet before S_T cut	Uncertainty di-EM + dijet after S_T cut
Fake rates	11%	11%	11%
Jet energy scale	N/A	25%	26%
γ+jet contamination	1.0%	1.0%	1.0%
Trigger selection	3.0%	1.5%	1.5%
Algorithm	8.0%	8.0%	8.0%
Statistics	0.2%	1.1%	4.2%
Total	14%	28%	29%

Table 5.14: *QCD background uncertainties for di-EM sample, di-EM + dijet sample before and after the $S_T > 375$ GeV cut. "N/A"means the source of error is not appliable.*

Systematic Uncertainties on the Z/Drell-Yan Background

The systematic error on the Z/Drell-Yan background includes the uncertainties in the jet energy scale and the particle identification efficiencies (EM ID efficiency, EM tracking efficiency, and jet ID efficiency).

The error on the Z/Drell-Yan background due to the jet energy scale uncertainty is estimated for the $eejj$ sample to be 26%, which is the same before or after the S_T cut.

When we normalize the Z/Drell-Yan background, we correct the difference in

particle ID efficiencies between data and Monte Carlo by scaling the Monte Carlo sample. That scale factor is the ratio of the particle ID efficiency for data events to that for Monte Carlo events, and it propagates the uncertainties of both the data efficiency and the Monte Carlo efficiency. From the uncertainties in particle ID efficiencies and by taking into account the fraction of events in the Z/Drell-Yan sample (about 60% for CC-CC and 40% for CC-EC), we can estimate the error on the Z/Drell-Yan background due to particle ID efficiency to be 10%.

In Section 5.2.2, we fit the jet multiplicity spectrum to obtain the scale factor of 1.3 per jet for the Z/Drell-Yan PYTHIA Monte Carlo sample. We estimate the number of events from the QCD background and the Z/Drell-Yan Monte Carlo for Z + (at least) n jets events. The errors due to the jet energy scale (JES) and EM ID efficiency are estimated and listed in Table 5.15. These errors explain the uncertainties on the number of events in Table 5.4 and 5.5.

		QCD background	Z/DY MC (PYTHIA)	Z/DY+2jets MC (ALPGEN)
JES error	$Z+ \geq$1jet	13%	15%	N/A
	$Z+ \geq$2jets	25%	26%	14%
	$Z+ \geq$3jets	40%	50%	21%
	$Z+ \geq$4jets	70%	80%	38%
EM ID error		N/A	10%	10%

Table 5.15: *Background uncertainties due to jet energy scale and EM ID efficiency, for QCD background, Z/Drell-Yan PYTHIA Monte Carlo, and Z/Drell-Yan + 2jets ALPGEN Monte Carlo. "N/A" means the source of error is not appliable.*

Systematic Uncertainties on the Top Background

The systematic error on the top background includes the uncertainties in jet energy scale and particle identification efficiencies. These errors are estimated similarly to the Z/Drell-Yan background errors. The error due to the jet energy scale is 4.4% for the $eejj$ sample before the S_T cut, and 26% after the S_T cut. The top background sample has about 80% CC-CC events and 20% CC-EC events. The error due to particle ID efficiency is estimated to be 10%.

Summary

The uncertainties in each background sample after all selection cuts, including the S_T cut, are summarized in Table 5.16. The statistical errors are not listed, since they are negligible in comparison to the systematic errors.

Source of systematics	Background uncertainty		
	QCD	Z/DY	Top
Fake rate etc.	14%	N/A	N/A
Particle ID	N/A	10%	10%
Jet energy scale	26%	26%	26%
Total	29%	28%	28%

Table 5.16: *Summary of systematic uncertainties in background samples passing all selection cuts.*

When we combine the backgrounds, i.e., summing up the number of events,

the absolute errors [8] in the number of events are combined in the following way: the errors from the same source (e.g., particle ID efficiency, or jet energy scale) are first added directly, then the errors from different sources are added in quadratures assuming that different error sources are uncorrelated to each other. It is a reasonable assumption that the fake rate, particle ID, and jet energy scale are uncorrelated error sources. If there is some correlation, we obtain a more conservative combined error when we add the errors as uncorrelated.

[8] If we express a value as $N \pm \delta N$, then the δN is called the abolute error. $\delta N/N$ is called the relative error, usually in percentage.

Chapter 6

Results

Having selected the data, predicted the background, and optimized the selection cut for the leptoquark signal, we are ready to compare the data to the background and determine the sensitivity of our search for leptoquarks. We first compare the distributions of a set of variables for the data and the background. Based on the consistency of the data with the background, we further extract the limits on the existence of leptoquarks.

6.1 Data and Background Comparison

We start by comparing the total number of observed events (before applying an S_T cut) to our expected background. By requiring events to have two EM objects and two jets and applying the Z-veto cut (i.e., $M_{ee} < 80$ GeV or $M_{ee} > 102$ GeV),

the number of events observed in the data (36) is consistent with the expected number of background events (43.7 ± 11.2).

In addition, a comparison of various data distributions to background distributions shows no evidence for leptoquarks. We have compared the data to the background for the following distributions: two EM objects E_T (Figure 6.1); two jets E_T (Figure 6.2); M_{ee} and M_{ej} (Figure 6.3). M_{ej} is the invariant mass of a pair of an EM object and a jet. There are two possible combinations of associating the EM objects (e_1, e_2) with the jets (j_1, j_2) into two ej pairs (e_1j_1–e_2j_2, or e_1j_2–e_2j_1) in an $eejj$ event, and thus each event can have two possible M_{ej} pairs. We choose the pair that has the closest M_{ej} values.

In Section 5.5, we determined that the optimized cut for S_T was $S_T > 375$ GeV, where S_T is the scalar sum of transverse energies of two electrons and two jets. Figure 6.4 shows the S_T distribution of data compared with background, with the Z-veto cut applied.

In the comparison plots (Figure 6.1–6.4), the triangle-points represent the data, with the error bars representing the statistical errors. The shaded histogram represents the total background, which is the sum of the Z/Drell-Yan background (dotted line), the QCD fake background (dashed line), and the top background (light solid line). The dashed-dot line represents the expected signal with the leptoquark mass of 220 GeV/c^2.

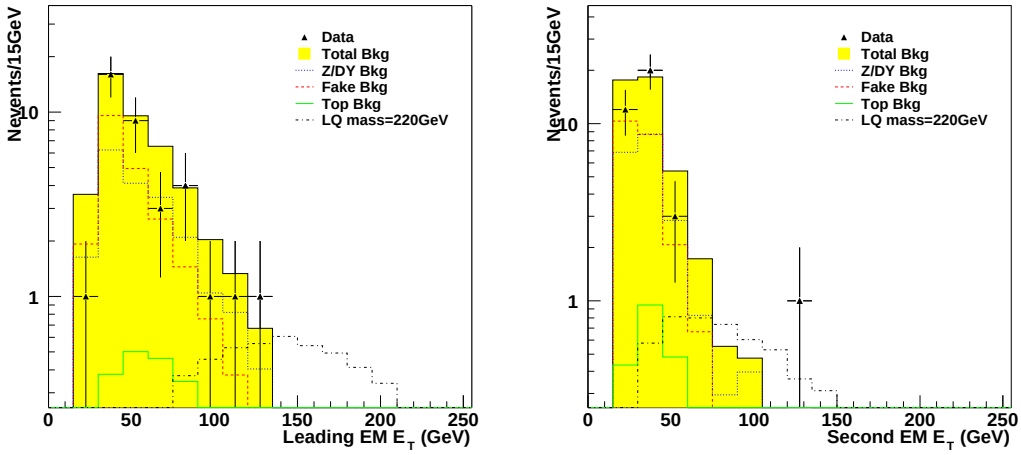


Figure 6.1: *Comparison of eejj data to expected background for the leading EM object E_T* (left) *and the second EM object E_T* (right)*, where the Z-veto cut has been applied.*

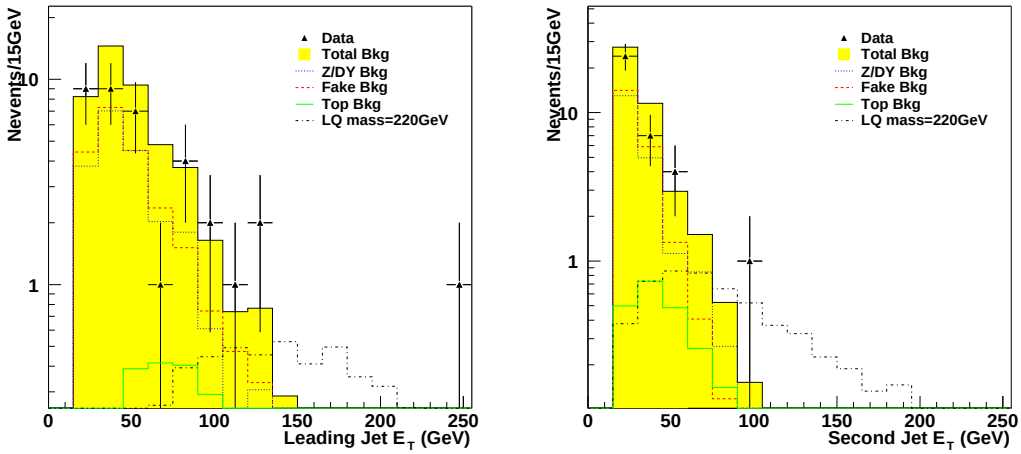


Figure 6.2: *Comparison of eejj data to expected background for the leading jet E_T* (left) *and the second jet E_T* (right)*, where the Z-veto cut has been applied.*

After applying the S_T cut, the data are still consistent with the expected background, and there is no evidence of a leptoquark signal. Table 6.1 lists the number of events in data and the number expected from background before and after applying the $S_T > 375$ GeV cut. After the S_T GeV cut, we observe zero events, and the number of expected background events is 0.44 ± 0.12.

Given the good agreement of the data with the Standard Model background expectation, we conclude that there is no evidence for leptoquark production.

6.2 Extracting a Limit

In the absence of leptoquark signal, we can set an upper limit on the product of the cross section times decay fraction as a function of the leptoquark mass.

6.2.1 Bayesian Technique

We use the Bayesian approach [83] for determining the limit. For a discrete variable A, we use the symbol $P(A|B)$ to represent the probability of proposition A, given that proposition B is true. When dealing with a continuous parameter x, we define a differential probability density $\rho(x|B)$, such that, given B, the probability of the continuous variable being between x and $x + dx$ is given by $P(x|B) = \rho(x|B)\, dx$.

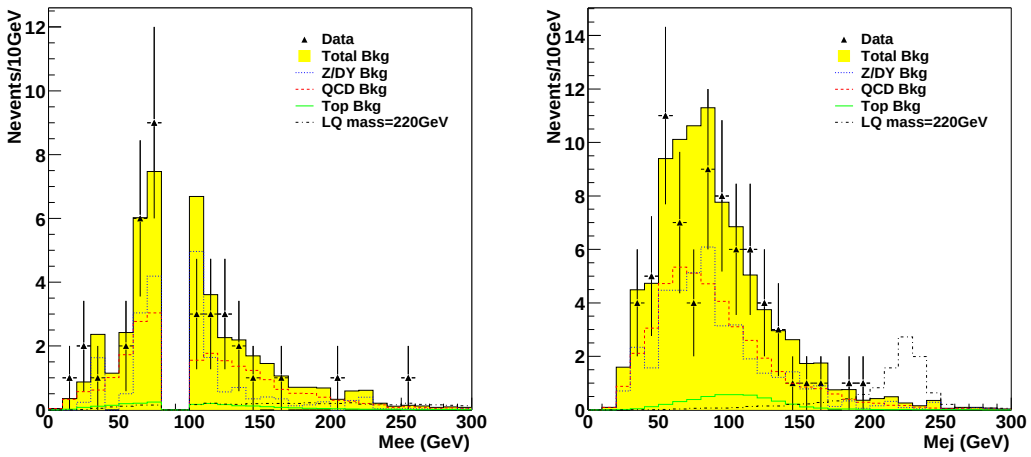


Figure 6.3: *Comparison of eejj data to expected background for* M_{ee} (left) *and* M_{ej} (right), *where the Z-veto cut has been applied. The hole in the* M_{ee} *distribution comes from the Z mass cut.*

	Before S_T cut	After S_T cut
Data	36	0
Total background	43.7 ± 11.2	0.44 ± 0.12
Z/Drell-Yan	19.5 ± 5.4	0.18 ± 0.05
QCD fake	22.0 ± 6.3	0.13 ± 0.04
Top	2.18 ± 0.24	0.13 ± 0.04
LQ (m=220 GeV/c^2)	5.39 ± 0.37	4.68 ± 0.38

Table 6.1: *Number of eejj events from data and background (Z/Drell-Yan, QCD fake, and top), before applying the* S_T *cut and after the* S_T *cut.*

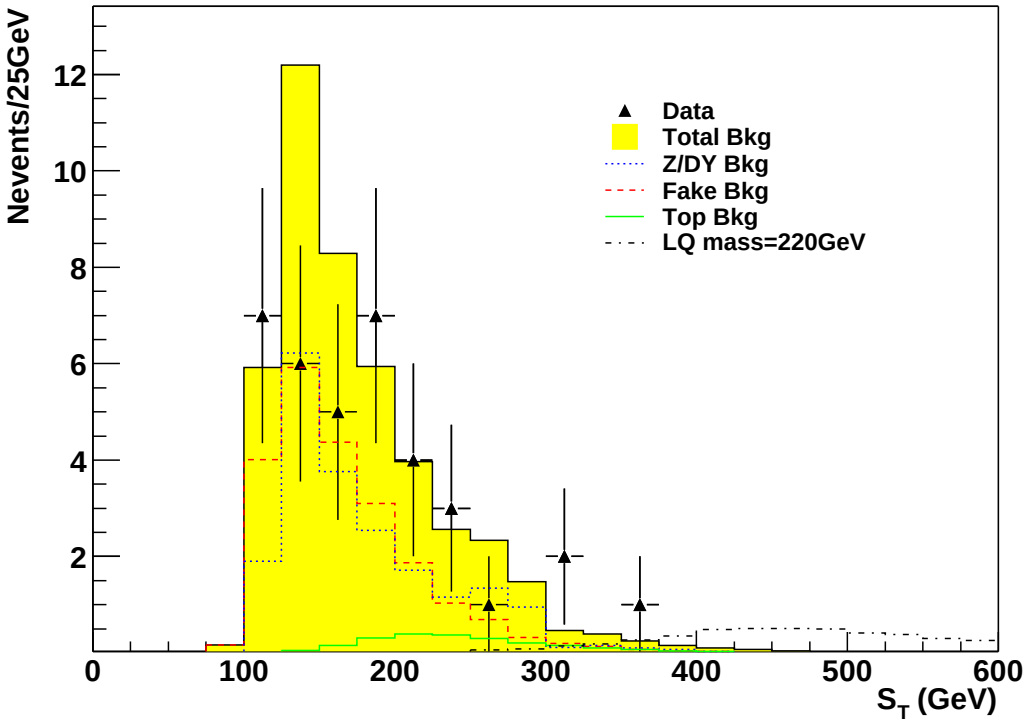


Figure 6.4: *The S_T distribution of the eejj events for data* (triangles), *Standard Model background* (solid line), *and leptoquark signal at a mass of 220 GeV* (dash-dot line). *The Z-veto cut has been applied.*

In a particular counting experiment, if k represents the number of observed events and μ represents the expected number of events, the likelihood function of the probability $P(k|\mu, I)$ can be described using the Poisson distribution function:

$$P(k|\mu, I) = \frac{e^{-\mu}\mu^k}{k!} \tag{6.1}$$

where I indicates all the information used to calculate μ, as well as the assumption that the Poisson distribution is the correct function to describe the probability. We call I the *prior* condition. The expected number of events μ is related to the signal cross section σ, the signal overall efficiency ϵ, the integrated luminosity L, and the expected background b, as:

$$\mu = b + L\epsilon\sigma \tag{6.2}$$

With the above model, the likelihood function (i.e., the probability of observing k events) is

$$P(k|\sigma, L, \epsilon, b, I) = \frac{e^{-(b+L\epsilon\sigma)}(b + L\epsilon\sigma)^k}{k!} \tag{6.3}$$

Now, given the k observed events, the probability that the leptoquark cross

section is σ is given by Bayes' Theorem as

$$\begin{aligned}\rho(\sigma, L, \epsilon, b|k, I) &\propto P(k|\sigma, L, \epsilon, b, I)\rho(\sigma, L, \epsilon, b|I) \\ &\propto \frac{e^{-(b+L\epsilon\sigma)}(b + L\epsilon\sigma)^k}{k!}\rho(\sigma|I)\rho(L|I)\rho(\epsilon|I)\rho(b|I) \qquad (6.4)\end{aligned}$$

where the constant of proportionality is determined by the condition

$$\int_0^\infty d\sigma \int_0^\infty dL \int_0^1 d\epsilon \int_0^\infty db\ \rho(\sigma, L, \epsilon, b|k, I) = 1 \qquad (6.5)$$

Since the most basic assumption about the signal is that it cannot be negative, but otherwise can be anything, a natural choice for the signal prior is $\rho(\sigma|I) = \theta(\sigma)$, where $\theta(x)$ is θ-function, defined as 0 for $x < 0$ and 1 for $x \geq 0$. We can also use a flat prior probability density of finite range as

$$\rho(\sigma|I) = \begin{cases} 1/\sigma_{max} & \text{if } 0 \leq \sigma \leq \sigma_{max} \\ 0 & \text{otherwise} \end{cases} \qquad (6.6)$$

where σ_{max} is the cross section value chosen to be sufficiently large that the likelihood function for $\sigma > \sigma_{max}$ is negligible. The other prior probability distributions

are assumed to be Gaussian functions

$$\rho(L|I)\rho(\epsilon|I)\rho(b|I) = \frac{1}{\delta_L\sqrt{2\pi}}e^{-\frac{(L-\bar{L})^2}{2\delta_L^2}} \cdot \frac{1}{\delta_\epsilon\sqrt{2\pi}}e^{-\frac{(\epsilon-\bar{\epsilon})^2}{2\delta_\epsilon^2}} \cdot \frac{1}{\delta_b\sqrt{2\pi}}e^{-\frac{(b-\bar{b})^2}{2\delta_b^2}} \tag{6.7}$$

where δ_L, δ_ϵ and δ_b are the errors for L, ϵ and b respectively; $\bar{L}$, $\bar{\epsilon}$ and $\bar{b}$ are the mean values for L, ϵ and b respectively.

Because the interest is in σ, we integrate over the other parameters (nuisance parameters) L, ϵ, and b in Equation 6.4, and get

$$\rho(\sigma|k,I) = \int_0^\infty dL \int_0^1 d\epsilon \int_0^\infty db\ \rho(\sigma, L, \epsilon, b|k, I) \tag{6.8}$$

After normalization, $\rho(\sigma|k,I)$ is called the *posterior* probability distribution. $\rho(\sigma|k,I)$ is the probability distribution for the leptoquark production cross section given the data observed. The 95% confidence level (CL) cross section upper limit can be determined by solving the equation

$$0.95 = \int_0^{\sigma^{95\%}} d\sigma\rho(\sigma|k,I) \tag{6.9}$$

where $\rho(\sigma|k,I)$ is normalized and $\sigma^{95\%}$ is what we call the cross section upper limit at the 95% confidence level. The program used to perform the calculation can be found in [84].

6.2.2 Limit

Using a 95% confidence level upper limit for the leptoquark cross section as a function of leptoquark mass (see Table 6.2), we arrive at the open circles shown in Figure 6.5. Comparing these experimental upper limits with the next-to-leading order (NLO) theoretical calculations for the scalar leptoquark pair production cross section (in Table 5.6), we are able to place a lower mass limit.

LQ mass (GeV/c^2)	$\sigma \times \beta^2$ upper limit (pb)
120	0.737
140	0.346
160	0.205
180	0.138
200	0.105
220	0.090
240	0.082
260	0.078
280	0.077

Table 6.2: *Experimental upper limit of cross section $\times$ decay fraction for each leptoquark mass at the 95% confidence level. The decay fraction of a leptoquark pair decaying into $eejj$ is β^2.*

Using the lower bound for the theoretical cross section band we find a lower limit on the leptoquark mass of 231 GeV/c^2, assuming a leptoquark decay branching ratio $\beta = 1$ (Figure 6.5). By varying β (the branching ratio of a leptoquark decaying into an electron and a quark), which corresponds to a decay fraction of β^2 for a leptoquark pair decaying into the $eejj$ channel, we can set the corre-

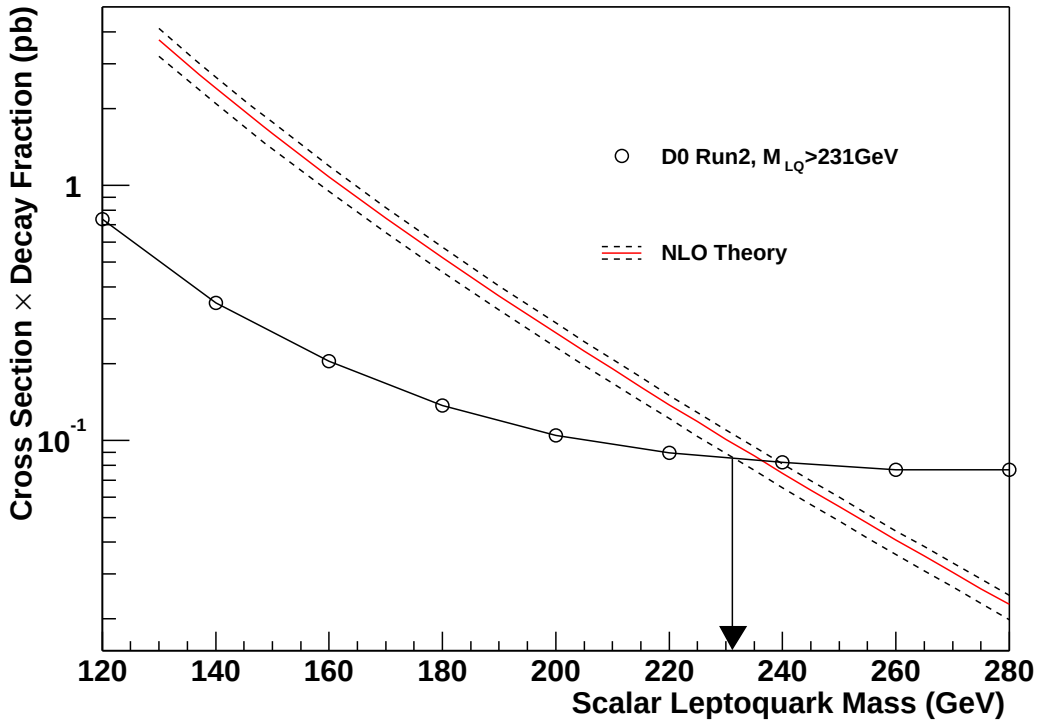


Figure 6.5: *The 95% confidence level upper limit on the cross section times decay fraction as a function of LQ mass* (open circles). *The NLO theoretical cross sections are plotted for* $\mu = 1, 0.5, 2 \times M_{LQ}$. *A lower mass limit of 231 GeV for first generation scalar leptoquarks is achieved for* $\beta = 1$.

sponding leptoquark mass limit. Figure 6.6 shows the leptoquark mass limit for different β. For example, at $\beta = 0.5$, the lower limit on the leptoquark mass is 169 GeV/c^2.

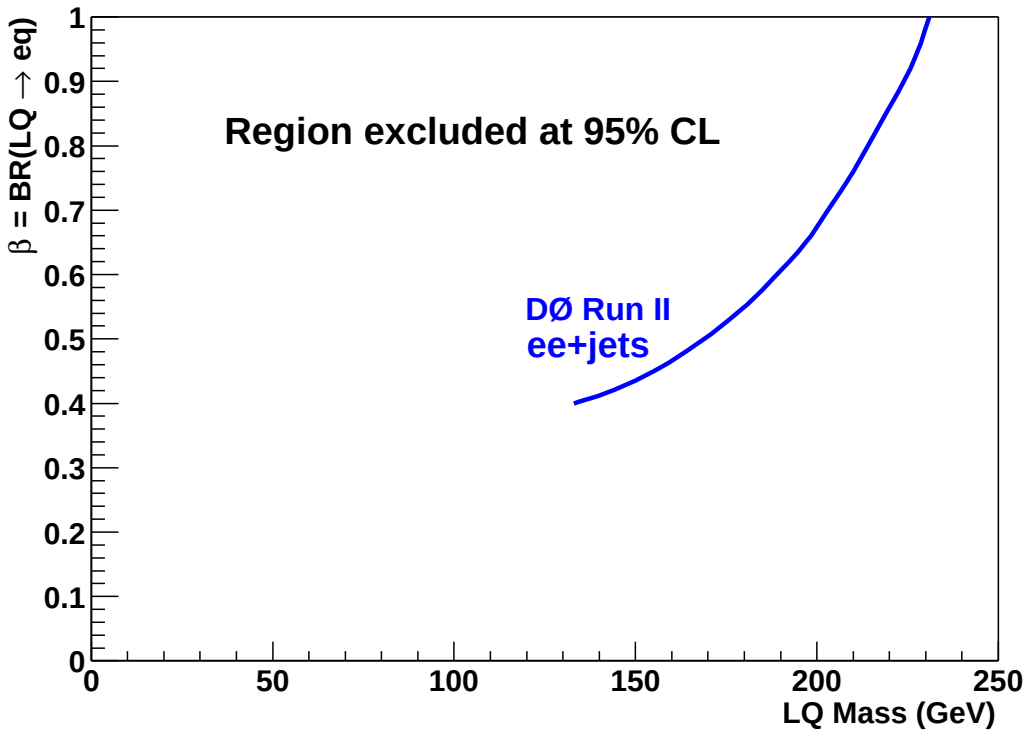


Figure 6.6: *The lower limit on the leptoquark mass at the 95% confidence level as a function of β. For $\beta = 0.5$, the mass limit on the first generation scalar leptoquark is 169 GeV.*

6.3 Combined Results

To obtain the most stringent limits in leptoquark production, we can combine our result with earlier upper limit results from Run I experiment. Moreover, if we combine our analysis in the $eejj$ channel with the one of leptoquark search in the $e\nu jj$ channel, we are able to obtain better limits for the case of leptoquarks decaying into both two channels. In the following, we will describe these two kinds

of combinations.

6.3.1 Combining Run I Results

In Run I, the mass limit of the first generation scalar leptoquark was measured to be 225 GeV by DØ [31], and 242 GeV for the combined result from the CDF and DØ experiments [85]. It is natural for us to combine the DØ Run II limits with the Run I results, using the Bayesian approach. The results from the DØ Run II and Run I experiments are summarized in Table 6.3.

	DØ Run II	DØ Run I
Number of candidates	0	0
Background	0.44 ± 0.12	0.44 ± 0.06
Efficiency ($M_{LQ} > 220$ GeV)	26%–30% $\pm$ 2.1%	36%–39% $\pm$ 5%
Integrated luminosity	130.4 ± 8.5 pb^{-1}	123 ± 6.5 pb^{-1}

Table 6.3: *Results from the DØ Run II and DØ Run I experiments.*

In the Bayesian approach, we will define probability functions for each experiment in such a way that they take into account correlated and uncorrelated uncertainties. When we combine Run II and Run I results, we assume the systematic errors in the integrated luminosities, efficiencies, and backgrounds are completely uncorrelated between Run II and Run I. This is a more conservative assumption than that a fraction of the errors are correlated. We assume Gaussian errors in all the parameters.

As we described in Section 6.2.1, the probability of observing k events is given by Equation 6.3. Since neither experiment observed any candidate events ($k = 0$), the likelihood function of each measurement is give by

$$P(k = 0|\sigma, L, \epsilon, b, I) = e^{-(b+L\epsilon\sigma)} \tag{6.10}$$

We apply Bayes' theorem and obtain the posterior probability for the cross section σ, given the observation of zero events in the data:

$$\rho(\sigma|k = 0, I) = \frac{\int_0^\infty dL \int_0^1 d\epsilon \int_0^\infty db P(k = 0|\sigma, L, \epsilon, b, I)\rho(L|I)\rho(\epsilon|I)\rho(b|I)\rho(\sigma|I)}{\int_0^\infty d\sigma \int_0^\infty dL \int_0^1 d\epsilon \int_0^\infty db \rho(L|I)\rho(\epsilon|I)\rho(b|I)\rho(\sigma|I)} \tag{6.11}$$

where $\rho(L|I)$, $\rho(\epsilon|I)$, $\rho(b|I)$ are prior probability densities for integrated luminosity, efficiency and background, and are Gaussian by assumption. $\rho(\sigma|I)$ is the prior for the signal cross section, which we choose as a θ-function, defined as $\theta(\sigma) = 0$ for $\sigma < 0$ and $\theta(\sigma) = 1$ for $\sigma \geq 0$.

An efficient way to calculate the integral in Equation 6.11 is to use Monte Carlo integration by generating random values of L, ϵ and b according to their Gaussian priors. The value of the integral is simply the average value of $P(k = 0|\sigma, L, \epsilon, b, I)$ obtained in the series of the Monte Carlo trials, since by definition probability density functions are normalized to unity.

We then vary the input value of σ for the Monte Carlo trials to obtain $\rho(\sigma|k =$

$0, I)$ in the entire range: $0 \leq \sigma < \infty$. The upper limit at 95% confidence level on signal cross section, $\sigma^{95\%}$, can then be obtained by solving the following integral equation

$$\int_0^{\sigma^{95\%}} \rho(\sigma|k=0,I)d\sigma = 0.95 \int_0^{\infty} \rho(\sigma|k=0,I)d\sigma \tag{6.12}$$

(Here we have to normalize the posterior probability to unity since $\rho(\sigma|I) = \theta(\sigma)$ is not properly normalized.)

From Equation 6.10 it is natural to expect that $\rho(\sigma|k=0,I)$ can be parameterized as $Ae^{-a\sigma}$. In this case Equation 6.12 transforms into

$$e^{-a\sigma^{95\%}} = 1 - 0.95 = 0.05$$

which can be easily solved as

$$\sigma^{95\%} = \frac{1}{a} \ln 20 = \frac{3.00}{a} \tag{6.13}$$

In general, the efficiency ϵ depends on the leptoquark mass. For Run II leptoquark samples, we know the efficiency at various leptoquark masses. For Run I results, we will use the fact that for the leptoquark masses above 200 GeV the efficiency changes very slowly, and we simply take the average value.

When combining results from the Run II and Run I experiments, since we

still have zero candidates observed, the only required modification to Equation 6.10 is the background plus signal expectation, which changes from $b + L\epsilon\sigma$ to $b_1 + L_1\epsilon_1\sigma + b_2 + L_2\epsilon_2\sigma$, where indices 1 and 2 correspond to Run I and Run II experiments respectively. Under the assumption of completely uncorrelated errors, it leads to a combined posterior probability of

$$\rho_{combined}(\sigma|k=0,I) = \rho_{RunI}(\sigma|k=0,I)\rho_{RunII}(\sigma|k=0,I) \tag{6.14}$$

The posterior probability for the Run II DØ experiment is calculated for each leptoquark mass point (Figure 6.7 shows one mass point as an example). The plot proves that the exponential function $Ae^{-a\sigma}$ is a good approximation to the posterior probability $\rho(\sigma|k=0,I)$. We use the exponential approximation for the posterior probability for Run I DØ experiment and the Run I CDF and DØ combined posterior probability.

The Run I results of 95% confidence level upper limits on the leptoquark cross sections are [85]:

$$\sigma^{95\%}_{RunI D\emptyset} = 0.0663 \text{ pb} \tag{6.15}$$

$$\sigma^{95\%}_{RunI D\emptyset+CDF} = 0.0383 \text{ pb} \tag{6.16}$$

Using the exponential approximation and the relation in Equation 6.13, we obtain

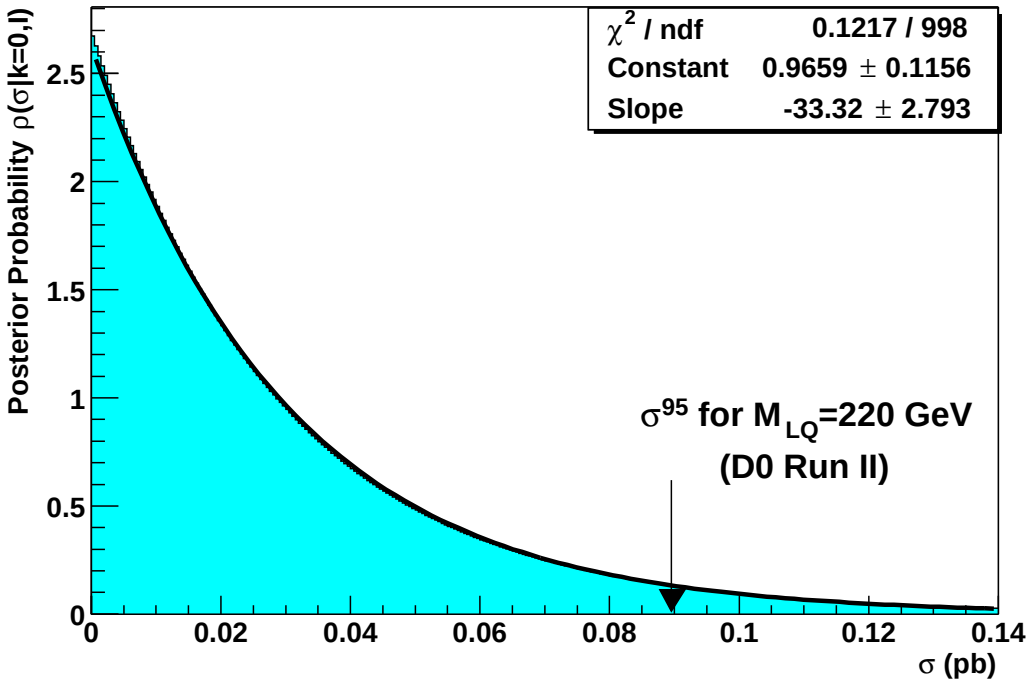


Figure 6.7: *The posterior probability $\rho(\sigma|k=0,I)$ for leptoquark mass = 220 GeV at Run II energy with respect to leptoquark cross section. The curved line shows a fit to the exponential function. The arrow indicates the cross section upper limit at 95% confidence level for this mass point.*

the Run I posterior probabilities as

$$\rho'_{RunI D\emptyset}(\sigma'|k=0,I) = A'e^{-45.2\sigma'} \quad (6.17)$$

$$\rho'_{RunI D\emptyset+CDF}(\sigma'|k=0,I) = A'e^{-78.1\sigma'} \quad (6.18)$$

The above σ' is the leptoquark cross section at the Run I energy. Considering the difference in leptoquark production cross sections due to the energy difference

between Run I and Run II, we need to normalize the Run I cross section limit to the Run II cross section in order to correctly combine them. We take the mass limit obtained in Run I and find the equivalent Run II cross section limit which would give the same mass limit. Such an equivalent Run II cross section limit is larger than the Run I cross section limit. We then re-scale the x-axis (σ') of the posterior probability plot from Run I by that ratio, and we get the normalized posterior probabilities:

$$\rho_{RunI D\emptyset}(\sigma|k=0,I) = Ae^{-28.9\sigma} \quad (6.19)$$

$$\rho_{RunI D\emptyset+CDF}(\sigma|k=0,I) = Ae^{-48.4\sigma} \quad (6.20)$$

We use the combined posterior probability, and solve the similar integral equation to obtain $\sigma^{95\%}$ at each leptoquark mass point, as shown in Figure 6.8. Comparing with the NLO theoretical cross sections, we obtain the 95% confidence level mass limits for the combined results, which are listed in Table 6.4. The combined mass limit from DØ Run II and DØ Run I experiments is 253 GeV/c^2, which is the most stringent limit on the mass of first generation scalar leptoquark to date.

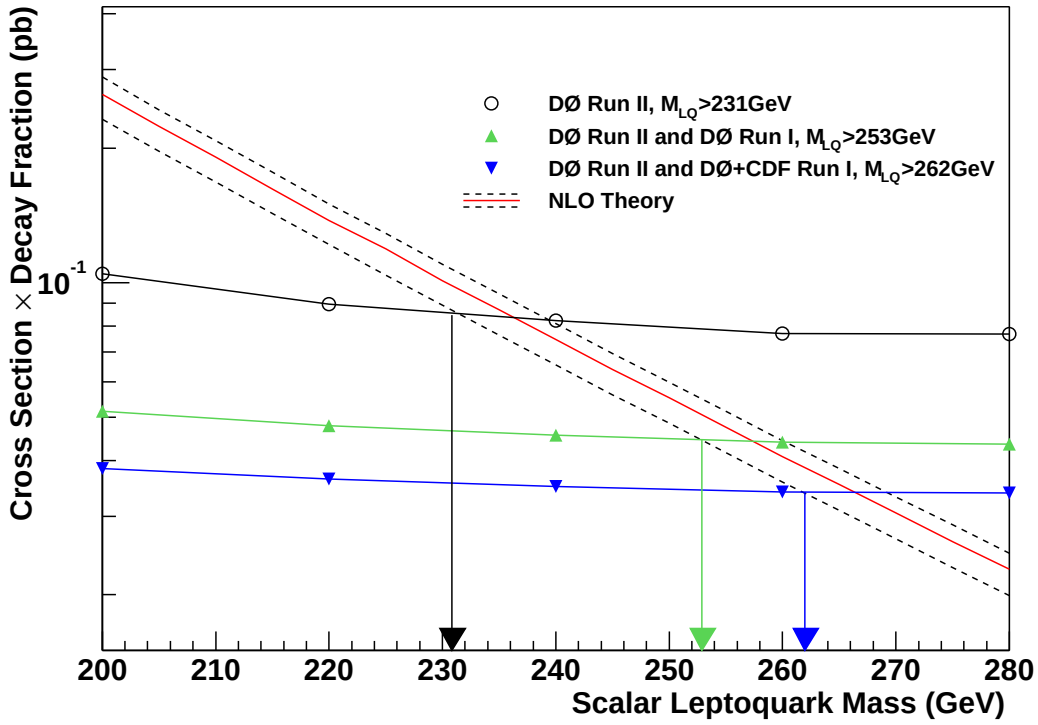


Figure 6.8: *The 95% confidence level upper limits on the cross section times decay fraction from DØ Run II only* (open circles), *DØ Run II combined with DØ Run I* (triangles), *and DØ Run II combined with DØ + CDF Run I* (inverted triangles) *leptoquark analyses. The NLO theoretical cross sections are plotted for* $\mu = 1, 0.5, 2 \times M_{LQ}$.

	Upper σ limit (pb)	Lower mass limit (GeV/c^2)
DØ Run II	0.086	231
DØ Run II combined with DØ Run I	0.045	253
DØ Run II combined with DØ + CDF Run I	0.034	262

Table 6.4: *The 95% confidence level limits on the leptoquark cross section and the leptoquark mass as the combined results from DØ Run II and Run I leptoquark analyses.*

6.3.2 Combining $eejj$ and $e\nu jj$ Channels

The search for leptoquarks in the $eejj$ channel alone gives the best result when the branching ratio for a leptoquark decaying into an electron and a quark is $\beta = 1$. When $\beta < 1$, combining the results from both the $eejj$ and the $e\nu jj$ channels will prove more sensitive in searching for leptoquarks.

The search for first generation leptoquarks in the $e\nu jj$ channel has been performed at DØ with Run II data [86]. The $e\nu jj$ analysis is based on 121 $\pm$ 12.1 pb^{-1} of data and observes 3 events with the predicted background of 4.18 $\pm$ 0.99 events. Table 6.5 summarizes the efficiencies used in the $eejj$ and $e\nu jj$ analyses. We again use the Bayesian technique to combine the results, with correlated errors taken into account [87].

In order to combine the two channels for various values of β (the branching ratio of a leptoquark decaying into an electron and a quark), we note that the

LQ mass (GeV)	Efficiency $eejj$ (%)	Efficiency $e\nu jj$ (%)
120	3.4 ± 0.5	3.4 ± 0.5
140	7.2 ± 1.1	5.8 ± 0.7
160	11.9 ± 1.6	8.9 ± 1.1
180	17.4 ± 2.1	13.2 ± 1.6
200	22.5 ± 2.1	15.9 ± 1.9
220	26.0 ± 2.1	18.3 ± 2.2
240	28.3 ± 2.1	19.5 ± 2.4
260	30.2 ± 2.1	20.9 ± 2.5

Table 6.5: *The overall efficiency for leptoquark analyses in the $eejj$ and $e\nu jj$ channels as a function of the leptoquark mass.*

decay fraction for the leptoquark pair decaying into the $eejj$ and $e\nu jj$ channels is β^2 and $2\beta(1-\beta)$, respectively. We multiply the luminosity in each channel by the corresponding decay fraction, which would result in the limits expressed in terms of pure cross section (instead of cross section times decay fraction).

The systematic error on the signal efficiency in each analysis can be divided into correlated and uncorrelated errors. The correlated error comes from the luminosity uncertainty, parton distribution function dependence, electron and jet efficiency uncertainties, and the jet energy scale. The uncorrelated error comes from Monte Carlo statistics and the fact that there is one more EM object in the $eejj$ final state. The correlated error dominates the overall uncertainty. The uncorrelated error is very little and amounts to 2–3%.

In order to treat the correlated systematic errors correctly, we remove them from one of the channels, e.g., the $e\nu jj$ channel, leaving just the uncorrelated er-

rors. We then produce the posterior probability for the unmodified channel (which is the same as in the $eejj$ analysis), as well as the similar posterior probability for the other channel with reduced errors (which is slightly different from that in the $e\nu jj$ analysis) [1].

To set the combined limits, we multiply these two posterior probabilities and solve the integral equation, to find the cross section value that corresponds to 95% of the area below this combined posterior probability curve. This value represents the combined cross section limit at the 95% confidence level. We repeat this procedure for every leptoquark mass and the limit is then plotted as a function of the leptoquark mass. The intersection of the limit curve with the lower band of the NLO leptoquark pair production cross section gives the combined lower limit on the leptoquark mass.

The leptoquark mass limits from the individual analysis, as well as the combined limit as a function of β are listed in Table 6.6 and also shown in Figure 6.9.

[1] Alternatively, we can keep the full errors in the $e\nu jj$ analysis and use only the uncorrelated errors in the $eejj$ channel, and we obtain the same combined results.

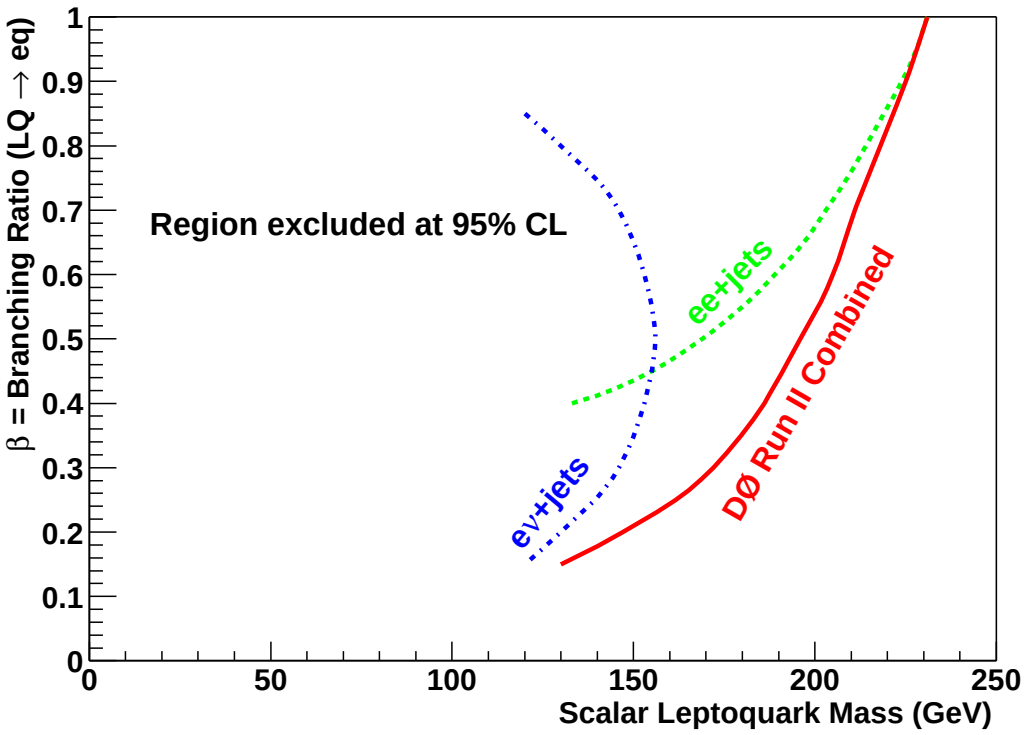


Figure 6.9: *The 95% confidence level lower limit on the mass of first generation scalar leptoquarks as a function of* β.

$\beta =$	95% CL mass limit (GeV)		
BR($LQ \to eq$)	$eejj$	$e\nu jj$	Combined
0.00	0	0	0
0.10	<120	<120	<120
0.15	<120	120	130
0.20	<120	130	147
0.30	<120	146	172
0.40	133	153	186
0.50	169	156	196
0.60	189	153	205
0.70	203	146	211
0.80	214	130	218
0.90	224	<120	225
1.00	231	<120	231

Table 6.6: *The 95% confidence level lower limits on the first generation scalar leptoquark mass, for various values of β.*

6.4 Conclusion

A search for first generation leptoquarks decaying into the $eejj$ channel has been performed using 130.4 $\pm$ 8.5 pb^{-1} of data collected with the DØ detector between September 2002 and June 2003. The observed data are consistent with the Standard Model backgrounds. Optimization of the kinematic cut has been performed. We find no evidence of the leptoquark signal, and set an upper limit on the leptoquark production cross section of 0.086 pb at 95% confidence level and a lower limit on the leptoquark mass of 231 GeV/c^2 for the first generation scalar leptoquarks when $\beta = 1$. When combining with the DØ Run I result, we set the world's most stringent mass limit (253 GeV/c^2, when $\beta = 1$) for the first

generation scalar leptoquarks to date.

Appendix

A.1 More Discussion of the Systematic Uncertainties of the QCD Background

A.1.1 Systematic Uncertainty due to γ-jet Event Contamination

In the QCD background estimation, we use the loose single-EM sample to obtain fake rates, and then estimate the fake background using the loose di-EM sample. These samples are dominated by QCD dijet and multijet events. There are a small number of events from the direct photon (γ+jets) process, which is also a background to our analysis. We cannot separate direct photon events from

multijet events, but we can estimate its contamination in the sample. In the loose single-EM sample, the direct photon events contamination is

$$\frac{N_{\gamma j}}{N_{2j}} \sim \frac{\sigma_{\gamma j}}{\sigma_{2j} \cdot 2 f_{j \to em}}$$

where $N_{\gamma j}$ and N_{2j} are the number of γ+jet events and dijet events respectively. $\sigma_{\gamma j}$ and σ_{2j} are the cross sections of γ+jet and dijet processes (with $E_T > 25$ GeV for the leading two objects). $f_{j \to em}$ is the probability that a jet is reconstructed as a loose EM object, which is estimated to be about 2% in CC and 5% in EC [1]. The fraction of direct photon events in the loose single-EM sample, $frac^{\gamma}_{1em} = N_{\gamma j}/(N_{2j} + N_{\gamma j})$, is estimated to be $< 2\%$.

Similarly, in the loose di-EM sample, the direct photon contamination is

$$\frac{N_{\gamma j}}{N_{2j}} \sim \frac{\sigma_{\gamma j} \cdot f_{j \to em}}{\sigma_{2j} \cdot f^2_{j \to em}}$$

And the fraction of direct photon events in the loose di-EM sample, $frac^{\gamma}_{2em}$, is $< 4\%$. In our QCD background sample to the $eejj$ signal, i.e., the loose di-EM + dijet sample, the ratio of the number of γ+3jet events to the number of 4-jet

[1] We list the fake probabilities for jets and loose EM objects in Appendix A.1.3, and we estimate $f_{j \to em} \sim f_{j \to e_{hmx}} / f_{em \to e_{hmx}}$.

events is

$$\frac{N_{\gamma+3j}}{N_{4j}} \sim \frac{\sigma_{\gamma+3j} \cdot 3 f_{j\to em}}{\sigma_{4j} \cdot 6 f_{j\to em}^2}$$

where $\sigma_{\gamma+3j}$ and σ_{4j} are the cross sections of γ+3jet and 4-jet processes (with $E_T > 25$ GeV for the leading two objects and $E_T > 20$ GeV for the next two). The fraction of direct photon events in the loose di-EM + dijet sample, $frac^{\gamma}_{2em2j}$, is estimated to be $< 3\%$.

The cross section ratios and the direct photon contamination fractions are listed in Table A.1.

$\sigma_{2j}/\sigma_{\gamma j}$ (two objects $E_T > 25$ GeV)		2140
$frac^{\gamma}_{1em}$ (single-EM)	CC	1.2%
	EC	0.6%
$frac^{\gamma}_{2em}$ (di-EM)	CC-CC	3.5%
	CC-EC	0.6%
$\sigma_{4j}/\sigma_{\gamma+3j}$ (two objects $E_T > 25$ GeV, next two $E_T > 20$ GeV)		1590
$frac^{\gamma}_{2em2j}$ (di-EM+dijet)	CC-CC	2.4%
	CC-EC	0.4%

Table A.1: *The cross section ratios and the fractions of direct photon events in loose single-EM sample, loose di-EM sample, and loose di-EM + dijet sample.*

By definition, the fake rates estimated using the loose single-EM sample are

$$f_{em\to e_{hmx}} = N_{e_{hmx}}/N_{1em} \quad \text{(A.1)}$$

$$f_{em\to e_{trk}} = N_{e_{trk}}/N_{1em} \tag{A.2}$$

If we remove the direct photon events, the fake rates without direct photons are

$$F_{em\to e_{hmx}} = \frac{N_{e_{hmx}} - \epsilon N_{1em} frac^{\gamma}_{1em}}{N_{1em} - N_{1em} frac^{\gamma}_{1em}} \tag{A.3}$$

$$F_{em\to e_{trk}} = \frac{N_{e_{trk}}}{N_{1em} - N_{1em} frac^{\gamma}_{1em}} \tag{A.4}$$

where ϵ is the efficiency of the direct photon passing the EM ID cuts. We can easily obtain

$$F_{em\to e_{hmx}} = f_{em\to e_{hmx}} \frac{1 - \epsilon frac^{\gamma}_{1em}/f_{em\to e_{hmx}}}{1 - frac^{\gamma}_{1em}} \tag{A.5}$$

$$F_{em\to e_{trk}} = f_{em\to e_{trk}} \frac{1}{1 - frac^{\gamma}_{1em}} \tag{A.6}$$

When we estimate the QCD fake background, the fake probability for a loose di-EM event is expressed in Equation 5.7. To simplify the derivation, we take the fake rates to be independent of the E_T or η of the EM object, and we neglect the last term of Equation 5.7 (which is the higher order term since $f_{em\to e_{trk}} \ll f_{em\to e_{hmx}}$), leading to a fake probability of

$$p = 2 f_{em\to e_{hmx}} f_{em\to e_{trk}} \tag{A.7}$$

If we try to estimate the fake background for direct photon events and multijet events separately, then the fake probability of an event is

$$P = (1 - frac^{\gamma}_{2em2j}) \cdot 2F_{em\to e_{hmx}} F_{em\to e_{trk}} + frac^{\gamma}_{2em2j} \cdot \epsilon F_{em\to e_{trk}} \quad \text{(A.8)}$$

Comparing the above P with the p we used, the ratio is

$$\frac{P}{p} = \frac{(1 - frac^{\gamma}_{2em2j})(1 - \epsilon frac^{\gamma}_{1em}/f_{em\to e_{hmx}})}{(1 - frac^{\gamma}_{1em})^2} + \frac{\epsilon frac^{\gamma}_{2em2j}}{2(1 - frac^{\gamma}_{1em}) f_{em\to e_{hmx}}} \quad \text{(A.9)}$$

To estimate the difference, we take $frac^{\gamma}_{1em} = 0.02$, $frac^{\gamma}_{2em2j} = 0.03$, $\epsilon = 0.9$, $f_{em\to e_{hmx}} = 0.3$, and we get $P/p = 0.995$.

Our method of QCD background estimation does not separate direct photon events from multijet background events, but because the direct photon contamination is small, the systematic uncertainty due to this is less than 1%.

A.1.2 Systematic Uncertainty due to Trigger Selection

We have chosen the 2EM_HI trigger for the fake rate calculation because it has the smallest bias in extracting fake rates. The other triggers we used to select the data sample (EM_HI_SH, EM_MX_SH, and 2EM_HI_SH) have a shower shape requirement on the Level 3 electron. If we use any of these triggers to choose the loose single-EM sample, the triggering EM object is more likely to pass the

EM ID cuts and thus such a sample tends to have higher fake rates. This trigger bias effect on fake rates is shown in Figure A.1, especially for EM objects with $E_T < 60$ GeV.

We select the loose di-EM sample for the QCD background estimate, using the same triggers that we used for data analysis. About 95% of the events in the loose di-EM sample pass the 2EM_HI trigger, on which we can apply the fake rates calculated using the same trigger. For the remaining 5% of the events, since they passed the triggers that included a shower shape requirement, the triggering EM object should have higher fake rate.

To estimate the systematic uncertainty caused by the trigger bias, we can estimate the QCD background in the following way and check the difference. For events that pass the 2EM_HI trigger, we apply the fake rates for that trigger. If the event did not pass the 2EM_HI trigger but passed one of the other triggers, we apply the higher fake rate on the leading EM object (assuming it is the triggering EM object), and apply the normal fake rate on the other EM object. We have estimated the difference on the QCD fake background to be 3.0% for the di-EM sample, and 1.5% for the di-EM + dijet sample.

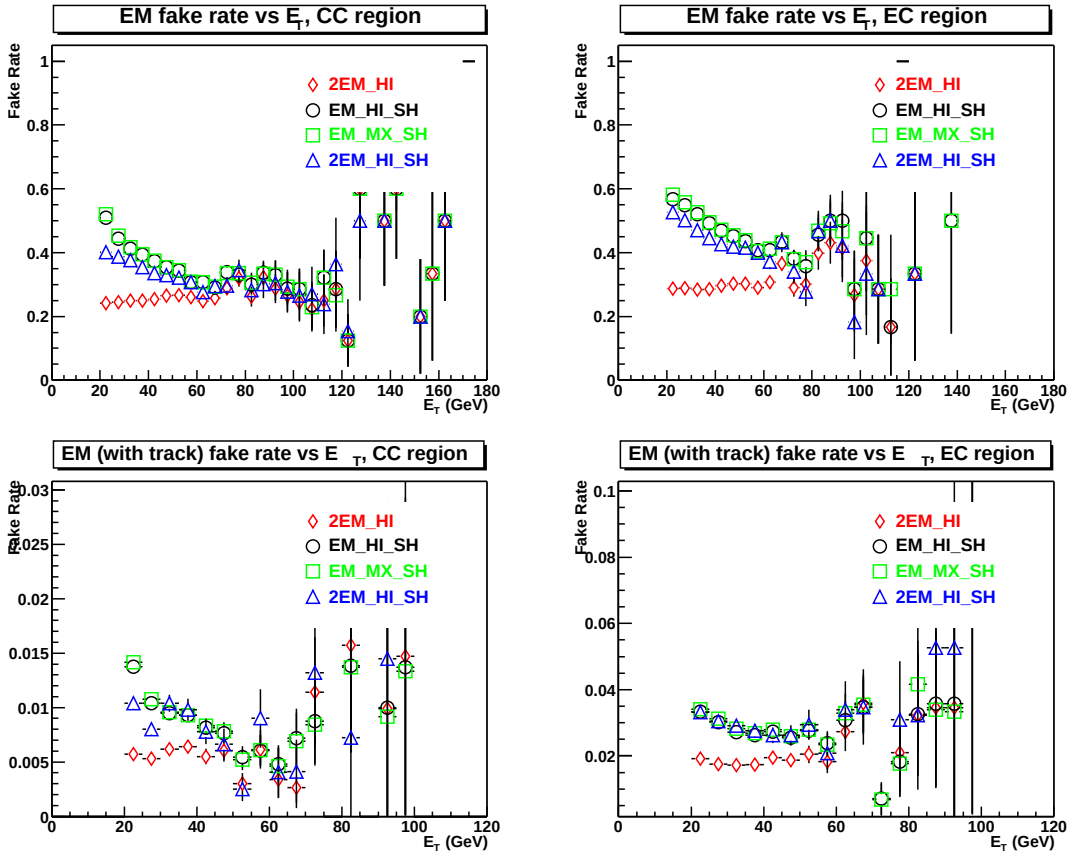


Figure A.1: *Fake rate as a function of EM object E_T, for the loose single-EM samples passing different triggers.*

A.1.3 Systematic Uncertainty due to the Estimate Algorithm

Our method of estimating the QCD background using the loose di-EM sample only accounts for the loose two EM objects faking two electrons passing the EM ID cuts. Since every event of the QCD background sample has four objects – two EM objects and two jets, there is certain probability that one or two of the jets can also fake electrons passing EM ID cuts. In this section, we estimate the systematic uncertainty caused by not including such a probability.

In an event with two loose EM objects and two jets, the two jets may not fake electrons since they may not pass the kinematic and geometric cuts for electrons (electrons have a higher E_T cut and a tighter η cut than jets). To simplify the calculation, here we take the same approximation as in Appendix A.1.1 Equation A.7 (i.e., assuming the fake rates are independent of E_T or η, and neglecting the higher order term of Equation 5.7).

The fake rates we need are

$$\begin{aligned} f^{CC}_{em\to e_{hmx}} &= 0.25 \\ f^{CC}_{em\to e_{trk}} &= 0.006 \\ f^{CC}_{j\to e_{hmx}} &= 0.005 = 0.02 \times f^{CC}_{em\to e_{hmx}} \end{aligned}$$

$$
\begin{aligned}
f^{CC}_{j\to e_{trk}} &= 0.00012 = 0.02 \times f^{CC}_{em\to e_{trk}} \\
f^{EC}_{em\to e_{hmx}} &= 0.3 \\
f^{EC}_{em\to e_{trk}} &= 0.02 \\
f^{EC}_{j\to e_{hmx}} &= 0.015 = 0.05 \times f^{EC}_{em\to e_{hmx}} \\
f^{EC}_{j\to e_{trk}} &= 0.001 = 0.05 \times f^{EC}_{em\to e_{trk}}
\end{aligned}
$$

The above fake rate values are approximate values, and the probabilities for jets faking electrons (i.e., jet fake rates, $f_{j\to e_{hmx}(e_{trk})}$) are from [88]. The specific numbers we will use are the ratios of jet fake rates to loose EM object fake rates. For simplicity, we take the average value of CC and EC,

$$f_{j\to e_{hmx}} = 0.03 \times f_{em\to e_{hmx}} \tag{A.10}$$

$$f_{j\to e_{trk}} = 0.03 \times f_{em\to e_{trk}} \tag{A.11}$$

In our simplified method, the probability for a loose di-EM + dijet event faking an $eejj$ event is

$$p = 2 f_{em\to e_{hmx}} f_{em\to e_{trk}}$$

If we consider the case that all four objects can fake electrons, then the fake

probability of the event is

$$\begin{aligned} p'_{2j} &= 2f_{em\to e_{hmx}} f_{em\to e_{trk}} \\ &+ 4f_{em\to e_{hmx}} f_{j\to e_{trk}} \\ &+ 4f_{em\to e_{trk}} f_{j\to e_{hmx}} \\ &+ 2f_{j\to e_{hmx}} f_{j\to e_{trk}} \qquad \text{(A.12)} \\ p'_{2j} &= 1.1209\ p \qquad \text{(A.13)} \end{aligned}$$

The difference is about 12%. Similarly, if only one of the jets can pass the electron's kinematic and geometric cuts in order to fake an electron, the fake probability is

$$\begin{aligned} p'_{1j} &= 2f_{em\to e_{hmx}} f_{em\to e_{trk}} \\ &+ 2f_{em\to e_{hmx}} f_{j\to e_{trk}} \\ &+ 2f_{em\to e_{trk}} f_{j\to e_{hmx}} \qquad \text{(A.14)} \\ p'_{1j} &= 1.06\ p \qquad \text{(A.15)} \end{aligned}$$

In the QCD background sample, we find about 80% of the events have at least one jet passing the electron's kinematic and geometric cuts, and about 40% of the events have two jets passing electron's kinematic and geometric cuts. Combining these effects, the difference in event fake probability is 8%.

In conclusion, our method to estimate the QCD background neglects the probability that the jets in the loose di-EM sample may also fake the electrons (with much smaller fake rates). So the QCD background is under-estimated, and the systematic uncertainty due to this is estimated to be 8% [2].

A.2 Run II Prospect

The goal of the Tevatron Run II is to deliver an integrated luminosity between 4.4 fb^{-1} and 8.6 fb^{-1} by 2009, which is 30–60 times greater than the luminosity used in this analysis. With a 50 times more integrated luminosity, given the same efficiencies and optimization as in this analysis, DØ will be able to extend the mass limit for scalar leptoquarks to $\sim$350 GeV/c^2 if there is no excess of events observed in the data. If leptoquarks exist with a mass less than that, we expect to see a 3σ evidence for scalar leptoquarks with mass up to $\sim$320 GeV/c^2, or a 5σ discovery for leptoquarks with mass up to $\sim$300 GeV/c^2.

[2]More accurately, this error is asymmetric, i.e. $^{+8}_{-0}$%. For simplicity, we just take the error as $\pm$8%, which is more conservative than the asymmetric one.

Bibliography

[1] F. Halzen and A. D. Martin, *Quarks and Leptons,* John Wiley & Sons (1984); D. H. Perkins, *Introduction to High Energy Physics,* Addison-Wesley (1987); C. Quigg, *Gauge Theories of the Strong, Weak, and Electromagnetic Interactions,* Benjamin-Cummings Publishing Co. (1983).

[2] D. Griffiths, *Introduction to Elementary Particle Physics,* pp. 213–256, John Wiley & Sons (1987).

[3] S. Glashow, Nucl. Phys. **22**, 579 (1967); S. Weinberg, Phys. Lett. **12**, 132 (1967); A. Salam, *Elementary Particle Physics,* N. Svartholm ed., 367 (1968).

[4] H. Fritzsch and M. Gell-Mann, *Proceedings of the Sixteenth international Conference on High Energy Physics,* Vol. **2**, p.135, Chicago (1972); S. Weinberg, Phys. Rev. Lett. **31**, 494 (1973); D.J. Gross and F. Wilczek, Phys. Rev. D **8**, 3633 (1973); H. Fritzsch, M. Gell-Mann and H. Leutwyler, Phys.Lett. B **47**, 365 (1973).

[5] T. Hambye and K. Riesselmann, Phys. Rev. D **55**, 7255 (1997).

[6] G. Ross, *Grand Unified Theories,* Benjamin-Cummings Publishing Co. (1995); P. D. B. Collins, A. D. Martin, and E. J. Squires, *Particle Physics and Cosmology,* John Wiley & Sons (1989).

[7] S. Weinberg, Phys. Rev. D **19**, 1277 (1979); L. Susskind, Phys. Rev. D **20**, 2619 (1979).

[8] S. Dimopoulos and L. Susskind, Nucl. Phys. B **155**, 237 (1979); E. Eichten and K. Lane, Phys. Lett. B **90**, 125 (1980).

[9] B. Holdom, Phys. Lett. B **150**, 301 (1985); M. Bando, T. Morozumi, H. So, and K. Yamawaki, Phys. Rev. Lett. **59**, 389 (1987); T. Appelquist and L.C.R. Wijewardhana, Phys. Rev. D **36**, 568 (1987).

[10] C. T. Hill, Phys. Lett. B **345**, 483 (1995); K. Lane and E. Eichten, Phys. Lett. B **352**, 383 (1995); K. Lane, Phys. Lett. B **433**, 96 (1998).

[11] J. Wess and J. Bagger, *Supersymmetry and Supergravity,* Princeton University Press (1983); Y.A. Gol'fand and E.P. Likhtam, JETP Lett. **13**, 323 (1971); D.V. Volkov and V.P. Akulov, Phys. Lett. B **46**, 109 (1973); J. Wess and B. Zumino, Nucl. Phys. B **70**, 39 (1974).

[12] G. Veneziano, Nuovo Cim. A **57**, 190 (1968); J. Scherk and J. H. Schwarz, Nucl. Phys. B **81**, 118 (1974); Phys. Lett. B **57**, 463 (1975); F. Gliozzi, J. Scherk and D. Olive, Phys. Lett. B **65**, 282 (1976).

[13] H. Georgi and S. Glashow, Phys. Rev. Lett. **32**, 438 (1974).

[14] J. C. Pati and A. Salam, Phys. Rev. D **10**, 275 (1974).

[15] E. Eichten *et al.*, Phys. Rev. D **34**, 1547 (1986); E. Eichten and K. Lane, hep-ph/9609297, hep-ph/9609298.

[16] E. Baver and M. Leurer, Phys. Rev. D **51**, 260 (1995); W. Buchmüller, R. Rückl, and D. Wyler, Phys. Lett. B **191**, 442 (1987).

[17] P. H. Frampton, Mod. Phys. Lett. A **7**, 559 (1992).

[18] J. L. Hewett and T. G. Rizzo, Phys. Rep. **183**, 193 (1989); E. Accomando *et al.*, Phys. Rep. **299**, 1 (1998).

[19] D. Choudhury and S. Raychaudhuri, Phys. Lett. B **401**, 54 (1997); G. Altarelli *et al.*, Nucl. Phys. B **506**, 3 (1997).

[20] S. Dawson, E. Eichten, and C. Quigg, Phys. Rev. D **31**, 1581 (1985); P. R. Harrison and C. H. Llewellyn Smith, Nucl. Phys. B **213**, 223 (1983) and Errata B **223**, 542 (1983).

[21] M. Krämer, T. Plehn, M. Spira, and P.M. Zerwas, *Pair Production of Scalar Leptoquarks at the Fermilab Tevatron,* Phys. Rev. Lett. **79**, 341 (1997).

[22] J. Blümlein, E. Boos, and A. Kryukov, Z. Phys. C **76**, 137 (1997).

[23] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[24] H.-U. Bengtsson, W.-S. Hou, A. Soni, and D. H. Stork, Phys. Rev. Lett. **55**, 2762 (1985).

[25] OPAL Collaboration, *Search for Pair-Produced Leptoquarks in e^+e^- interactions at $\sqrt{s} \simeq 183$ GeV,* Eur. Phys. J. C **13**, 15 (2000).

[26] D. Acosta and S. Blessing, *Leptoquark Searches at HERA and the Tevatron,* Ann. Rev. Nucl. Part. Sci. **49**, 389 (1999).

[27] C. Adloff *et al.* (H1 Collaboration), Z. Phys. C **74**, 191 (1997); J. Breitweg *et al.* (ZEUS Collaboration), Z. Phys. C **74**, 207 (1997).

[28] J. L. Hewett and T. Rizzo, Phys. Rev. D **56**, 5709 (1997).

[29] DØ Collaboration, Phys. Rev. Lett. **79**, 4321 (1997).

[30] Roberto Sacchi, J. Phys. G Nucl. Partic. **28 (5)**, 1141 (2002).

[31] DØ Collaboration, *Search for First-Generation Scalar and Vector Leptoquarks,* Phys. Rev. D **64**, 092004 (2001).

[32] DØ Collaboration, Phys. Rev. Lett. **81**, 38 (1998); DØ Collaboration, Phys. Rev. Lett. **84**, 2088 (2000); DØ Collaboration, Phys. Rev. Lett. **88**, 191801 (2002).

[33] CDF Collaboration, Phys. Rev. Lett. **79**, 4327 (1997); CDF Collaboration, Phys. Rev. Lett. **85**, 2056 (2000).

[34] J. Thompson, *Introduction to Colliding Beams at Fermilab,* Fermilab-TM-1909 (1994);

Fermilab Beam Division, *Run II Handbook,* http://www-bd.fnal.gov/runII .

[35] DØ Collaboration, *The DØ Detector,* Nucl. Instrum. and Methods. A **338**, 185 (1994).

[36] DØ Collaboration, *DØ Upgrade: the Detector and its Physics,* Fermilab Pub-96/357-E (1996).

[37] DØ Upgrade Collaboration, *DØ Silicon Tracker Technical Design Report,* DØ Note 2169 (1994).

[38] DØ Upgrade Collaboration, *DØ Upgrade Central Fiber Tracker Technical Design Report,* DØ Note 4164 (1999).

[39] J. Brzezniak *et al.*, *Conceptual Design of a 2 Tesla Superconducting Solenoid for the Fermilab DØ Detector Upgrade,* Fermilab TM-1886/DØ Note 2167 (1994).

[40] M. Adams *et al.*, *Design Report of the Central Preshower Detector for the DØ Upgrade,* DØ Note 3014 (1996).

[41] A. Gordeev *et al.*, *Technical Design Report of the Forward Preshower Detector for the DØ Upgrade,* DØ Note 3445 (1998).

[42] L. Sawyer *et al.*, *Technical design report for the upgrade of the ICD for DØ Run II,* DØ Note 2686 (1997).

[43] R. D. Schamberger, *The DØ Calorimeter Performance and Calibration,* Proc. fifth Calorimetry in High Energy Physics, H. Gordon and D. Rueger, Upton, NY: World Scientific, 1994, pp. 39–50.

[44] A. Kotwal *et al.*, *Calorimeter Electronics Upgrade for Run II: Technical Design Report,* DØ Note (1998);
http://www-d0.fnal.gov/hardware/cal/intro/tdr/tdrformat.htm .

[45] R. Jayanti *et al.*, *The DØ Muon System Upgrade,* DØ Note 2780 (1996).

[46] A. Lo, C. Miao, and R. Partridge, *Luminosity Monitor Technical Design Report,* DØ Note 3320 (1997).

[47] M. Abolins, D. Edmunds, S. Gross, and P. Laurens, *DØ Run II Level 1 Trigger Framework Technical Design Report,* (1998), http://www.pa.msu.edu/hep/d0/ftp/l1/framework/l1fw_tdr_05june98.txt .

[48] D. Edmunds, S. Gross, P. Laurens, J. Linnemann, and R. Moore, *Technical Design Report for the Level 2 Global Processor,* DØ Note 3402 (1998).

[49] A. Boehnlein, G. Brooijmans, D. Claes, and M. Souza, *Description of DØ L3 Trigger Software Components,* DØ Note 3630 (1999).

[50] A. Haas *et al.*, *Ethernet-based Data Acquisition for the DØ Experiment at Fermilab,* SCI 2002 conference proceedings; http://d0.phys.washington.edu/~haas/d0/l3/sci2002.pdf .

[51] http://www-d0.fnal.gov/computing/algorithms/ .

[52] M. Narain, *Vertex Reconstruction using the Impact Parameters Technique,* DØ Note 3560 (1999).

[53] D. Adams, *Finding Tracks,* DØ Note 2958 (1996).

[54] The TRF++ homepage, http://www-d0.fnal.gov/~dladams/trf/ .

[55] G. Hesketh, *Central Track Extrapolation Through the DØ Detector,* DØ Note 4079 (2003).

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

[57] http://www-d0.fnal.gov/phys_id/emid/d0_private/certification /main_v4_0.html .

[58] L. Duflot, *Simple cone preclustering for cone jets,* DØ Note 3749 (2000).

[59] JetMet group, *Jets in Run II,* DØ Note 3985 (2002).

[60] http://www-d0.fnal.gov/~d0upgrad/d0_private/software/jetid/certification /v2_0/jetid_p13.html .

[61] S. Fatakia *et al.*, *Jet Energy Scale and Resolution for p13 Data and Monte Carlo,* DØ Note 4115 (2003).

[62] http://www-d0.fnal.gov/~d0upgrad/d0_private/software/jetid/jetid.html .

[63] http://www-d0.fnal.gov/computing/algorithms/muon/muon_algo.html .

[64] T. Sjőstrand, L. Lőnnblad, S. Mrenna, and P. Skands, *PYTHIA 6.2: Physics and Manual,* hep-ph/0108264 (2002).

[65] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A.D. Polosa, *ALPGEN: a generator for hard multiparton processes in hadronic collisions,* hep-ph/0206293 (2002).

[66] http://www-d0.fnal.gov/computing/MonteCarlo/pmcs/pmcs_doc /pmcs.html .

[67] http://www-d0.fnal.gov/computing/MonteCarlo/simulation/d0gstar.html .

[68] http://www-d0.fnal.gov/computing/MonteCarlo/simulation/d0sim.html .

[69] F. Carminati *et al.*, *GEANT Users Guide,* CERN Program Library, 1991.

[70] http://www-d0.fnal.gov/computing/algorithms/status/p13.html .

[71] http://www-d0.fnal.gov/Run2Physics/np/ .

[72] http://www-d0.fnal.gov/phys_id/luminosity/ .

[73] T. Edwards *et al.*, *The Updated DØ Luminosity Determination,* DØ Note 4328 (2004).

[74] http://www-d0.fnal.gov/phys_id/jes/d0_private/certified/certified.html .

[75] A. Gupta, N. K. Mondal, and A. V. Kotwal, *Search for Quark and Lepton Compositeness using the Drell-Yan Cross Section at High Dielectron Invariant Mass,* DØ Note 3285 (1998).

[76] R. Hamberg, W. L. Van Neerven, and T. Matsuura, *A complete calculation of the order α_s^2 correction to the Drell-Yan K-factor,* Nucl. Phys. B **359**, 343 (1991).

[77] http://www-d0.fnal.gov/Run2Physics/top/d0_private/wg/mc /p13_p13.05.html .

[78] B. Knuteson, *Algorithms for Computing Significance,* DØ Note 3345 (1997).

[79] O. Helene, Nucl. Instr. and Meth. **212**, 319 (1983).

[80] http://mit.fnal.gov/~knuteson/limitSensitivity.cpp .

[81] J. Zhu, *EM Certification Tools (Version 2.0),* DØ Note 4171 (2003).

[82] http://consult.cern.ch/writeup/pdflib/ .

[83] I. Bertram, G. Landsberg, J. Linnemann, R. Partridge, M. Paterno, and H.B. Prosper, *A Recipe for the Construction of Confidence Limits,* DØ Note 3476 (1999).

[84] http://www-d0.fnal.gov/~hobbs/limit_calc.html .

[85] C. Grosso-Pilcher, G. Landsberg, and M. Paterno, *Combined Limits on First Generation Leptoquarks from the CDF and DØ Experiments,* hep-ex/9810015 (1998).

[86] M.-C. Cousinou, A. Cothenet, and G. Landsberg, *Search for First Generation Leptoquarks in the $e\nu jj$ channels,* DØ Note 4226 (2003).

[87] M.-C. Cousinou, A. Cothenet, S. Fu, G. Landsberg, and V. Zutshi, *Combined Limits on First Generation Leptoquarks in the eejj and evjj channels,* DØ Note 4228 (2003).

[88] http://www-d0.fnal.gov/phys_id/emid/d0_private/minutes /20030312tuan.ppt .