A Measurement of the Production Cross Section of $t\bar{t}$ Pairs in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using Secondary Vertex b-Tagging

by

Henri Bachacou

Grad. (Royal Institute of Technology, Sweden) 1999
Grad. (Ecole Centrale des Arts et Manufactures, France) 1999

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in Charge:
Professor James Siegrist
Professor Mark Strovink
Professor Eugene Haller

Fall 2004

The dissertation of Henri Bachacou is approved:

Chair Date

Date

Date

University of California, Berkeley
Fall 2004

A Measurement of the Production Cross Section of $t\bar{t}$ Pairs in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using Secondary Vertex b-Tagging

Abstract

A Measurement of the Production Cross Section of $t\bar{t}$ Pairs in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using Secondary Vertex b-Tagging

by

Henri Bachacou

Doctor of Philosophy in Physics

University of California, Berkeley

Professor James Siegrist, Chair

A measurement of the $t\bar{t}$ pair production cross section is presented using 162 pb^{-1} of data collected by the CDF experiment during Run II at the Tevatron. $t\bar{t}$ events in the lepton+jets channel are isolated by identifying electrons and muons, reconstructing jets and transverse missing energy, and identifying b jets with a secondary vertex tagging algorithm. The efficiency of the algorithm is measured in a control sample using a novel technique that is less dependent on the simulation. For a top quark mass of 175 GeV/c^2, a cross section of $\sigma_{t\bar{t}} = 5.6^{+1.2}_{-1.1}(\text{stat.})^{+0.9}_{-0.6}(\text{syst.})$pb is measured.

Chair Date

To my parents.

Contents

1 Introduction **1**

2 Theoretical Overview **5**
- 2.1 The Standard Model of Particle Physics 6
  - 2.1.1 The Strong Force: Quantum Chromodynamics 7
  - 2.1.2 The Electromagnetic and Weak Forces: the Electroweak Theory 8
  - 2.1.3 The Higgs Mechanism 9
  - 2.1.4 The Parameters of the SM 10
- 2.2 Physics Beyond the Standard Model 11
- 2.3 Top Quark Properties 12
  - 2.3.1 Standard Model Top Quark 12
  - 2.3.2 The Top Quark Beyond the Standard Model 14
- 2.4 $t\bar{t}$ Production at Hadron Colliders 15
- 2.5 Analysis Overview and Motivation 17
  - 2.5.1 Method 17
  - 2.5.2 Motivation 18

3 Experimental Apparatus **20**
- 3.1 The Tevatron and the Fermilab Accelerator System 21
  - 3.1.1 Proton Injector 21
  - 3.1.2 Antiproton Production and Recycling 22
  - 3.1.3 The Tevatron storage ring 22
- 3.2 The Collider Detector at Fermilab (CDF) 23
  - 3.2.1 The CDF coordinate system 23
  - 3.2.2 The Solenoid 24
  - 3.2.3 The Silicon Vertex Detector 25
  - 3.2.4 The Central Outer Tracker 30
  - 3.2.5 The Calorimetry 33
  - 3.2.6 The Muon System 35
  - 3.2.7 The Cerenkov Luminosity Counter 37
  - 3.2.8 The Trigger System 41
  - 3.2.9 Detector Simulation 44

4 Event Reconstruction and Selection 45
4.1 Tracking Reconstruction 46
  4.1.1 COT Tracking 46
  4.1.2 Silicon Detector Tracking 47
4.2 Electron and Photon Identification 48
4.3 Muon Identification 49
4.4 Jet Reconstruction 51
4.5 Missing Transverse Energy Measurement 52
4.6 Total Transverse Energy Measurement (H_T) 52
4.7 Data Samples 53
  4.7.1 W Sample 53
  4.7.2 Low p_T Electron Sample 55
  4.7.3 Jet Samples 55

5 Secondary Vertex b-Tagging: The SecVtx Algorithm 56
5.1 Introduction 57
5.2 Kinematics of b Jets in $t\bar{t}$ Events 58
5.3 Event-by-Event Primary Vertex Finding 59
5.4 The SecVtx Algorithm 61
5.5 Track Selection 63

6 Performance of the SecVtx b-Tagging Algorithm 68
6.1 Data and Monte Carlo Samples 69
6.2 Method 70
6.3 Measurement of The Electron-Jet Heavy Flavor Content 74
  6.3.1 F_b using $D^0 \rightarrow K\pi$ 74
  6.3.2 F_b from Cascade Muons 75
  6.3.3 Measurement of F_c/F_b 76
  6.3.4 Determination of F_{HF} 77
6.4 Result 78
6.5 Mistag Asymmetry 78
  6.5.1 Conversion Jets 81
  6.5.2 Generic Jets 83
6.6 Jet E_T Dependence 83
6.7 Semileptonic vs Hadronic decay 87
6.8 Measurement of the Mistag Rate 88

7 Understanding the Heavy Flavor Content of the W + Jets Sample 91
7.1 Heavy Flavor Production in Monte Carlo 92
7.2 Measurement of Heavy Flavor Fraction in Simulated W + Jets Events 94
7.3 Calibration of Heavy Flavor Fraction Using Jet Data 94

8 Backgrounds 100
8.1 QCD Background . . . 101
8.1.1 Pretag QCD Background . . . 101
8.1.2 Tagged QCD Background . . . 102
8.2 Mistags (W + Light Jets) . . . 106
8.3 W + Heavy Flavor Backgrounds . . . 107
8.4 Other Backgrounds . . . 109
8.5 Background Summary . . . 110

9 Event Selection Optimization and Acceptance 112
9.1 Event Selection Optimization with the H_T Variable . . . 113
9.2 Acceptance . . . 114
9.3 Remarks on the Systematics Associated with the H_T Cut . . . 121

10 Measurement of the $\sigma_{t\bar{t}}$ Production Cross Section 123
10.1 The Z^0 + Jets Samples: Validation of the Background Estimate . . . 124
10.2 The W + Jets Sample: $t\bar{t}$ Production Cross Section . . . 124

11 Conclusion and Outlook 132

A Derivation of the b-Tagging Efficiency Estimate 139

B List of Event Candidates 143
B.0.1 3-Jet Events . . . 144
B.0.2 $\geq$ 4-Jet Events . . . 145

C Particle Level Study of an Alternative Jet Reconstruction Algorithm Using Tracking Information 146
C.1 Introduction . . . 146
C.2 Event Generation . . . 147
C.3 Detector Simulation . . . 148
C.3.1 Projection of Particles into the Towers . . . 148
C.3.2 Efficiency and Resolution . . . 148
C.3.3 Jet Reconstruction . . . 150
C.3.4 Comparison with the Run I CDF Fast Simulation (QFL) . . . 151
C.4 Analysis . . . 151
C.4.1 Event Selection . . . 151
C.4.2 Tower Classification . . . 153
C.5 Idealized Case . . . 155
C.6 Resolution Improvement . . . 161
C.7 Conclusion . . . 163

D Monitoring of the Silicon Vertex Detector with the SVXMon Software 165
D.1 Introduction 165
D.2 Data Validation 166
D.3 Error Reporting 167
  D.3.1 SVX Error Logger GUI 167
D.4 SVXMon Plots 171
  D.4.1 Strip Plots 171
  D.4.2 Occupancy, Mean Charge, *etc.* vs DAQ Parameters 176
  D.4.3 Chip Plots 177
  D.4.4 Tracking Plots 180
  D.4.5 Pipeline Plots 181
  D.4.6 History Plots 185
D.5 Online Quality Control 187
D.6 Offline Histogram Analysis: SvxMonRunCompare 194
D.7 SVXMon Messages 205

List of Figures

2.1 *Higgs potential, projected onto the real axis (the potential is independent of the complex phase of the field).* 11
2.2 *Dotted contour: direct top and* $W^{\pm}$ *mass measurements from the Tevatron and LEP II. Solid contour: predicted top and* $W^{\pm}$ *masses from other electroweak measurements from LEP I, SLD, and neutral current experiments. The straight lines show the SM expected* $W^{\pm}$*-top mass dependence for three different Higgs masses[7].* 13
2.3 *Parton distribution functions of quarks and gluons in the proton at a momentum transfer* $\mu^2 = 10\text{GeV}^2$ *[4].* 17
2.4 *Feynman diagram of the leading order quark-antiquark annihilation top quark pair production process.* 17
2.5 *Feynman diagrams of the leading order gluon fusion top quark pair production processes.* 18

3.1 *The chain of accelerators at Fermilab.* 21
3.2 *Section of the CDF upgrade.* 24
3.3 *Section of the CDF tracking system (outward: SVX, ISL, and COT). L00 is not shown.* 25
3.4 *3D view of the SVX.* 26
3.5 *End view of the L00, SVX, and ISL.* 26
3.6 *View of a Layer 0 SVX module (*ϕ *side).* 27
3.7 *View of a Layer 0 SVX module (z side).* 27
3.8 *End view of L00.* 28
3.9 *Picture and diagram of the SVX3D chip.* 29
3.10 *DAQ of the silicon vertex detector.* 29
3.11 *COT cell layout in superlayer 2.* 31
3.12 *Equipotential lines in a COT cell of superlayer 1.* 32
3.13 *COT quadrant.* 32
3.14 *Wedge of the Central Electromagnetic calorimeter (CEM).* 35
3.15 *View of the Plug calorimeter (PEM and PHM).* 36
3.16 *Segmentation of the plug calorimeter.* 36
3.17 *Coverage (in the* $\eta \times \phi$ *plane) of the upgraded CDF muon system.* . . . 38
3.18 *Number of nuclear absorption lengths in front of the muon systems as a function of pseudorapidity.* 38

3.19 *Section of a CMU module in the $x \times y$ plane.* 39
3.20 *The CDF Cerenkov Luminosity Counter.* 40
3.21 *Time distribution in the East and West modules. Actual $p\bar{p}$ collisions deposit a coincident signal in the two modules.* 40
3.22 *Block diagram of the data flow at CDF.* 42
3.23 *Block diagram of the Level 1 and Level 2 triggers.* 43

5.1 *Pseudorapidity of the b quarks from top decays in $t\bar{t}$ lepton+jets Monte Carlo events.* 58
5.2 *Multiplicity of prompt tracks, tracks from B hadron, and tracks from sequential C hadron (from left to right).* 59
5.3 *Transverse momentum of prompt tracks, tracks from B hadron, and tracks from sequential C hadron (from left to right).* 60
5.4 *Impact parameter (in cm) of prompt tracks, tracks from B hadron, and tracks from sequential C hadron (from left to right).* 60
5.5 *Impact parameter significance (d_0/σ_{d_0}) of prompt tracks, tracks from B hadron, and tracks from sequential C hadron (from left to right).* . . 61
5.6 *Schematic view of a positive (left) and negative (right) SecVtx tag.* . . 62
5.7 *Transverse decay length of secondary vertices reconstructed by SecVtx for b jets in $t\bar{t}$ Monte Carlo events. Blue: all vertices. Red: vertices with $L_{2D}/\sigma_{L_{2D}} > 3$ (positive tags).* 63
5.8 *Vertex display of a W + 3 jets data event with two SecVtx positive tags. Only tracks within a jet cone and passing the selection are shown. Prompt tracks are shown in black; displaced tracks are shown in solid blue, or red if they are part of the final secondary vertex fit.* 64
5.9 *Distribution of the number of missing hits (upper left), number of good hits, number of hits (axial only), and track χ^2 (clockwise) for all tracks in the control sample satisfying the baseline selection [40].* 66

6.1 *Data/Monte Carlo comparison of some kinematic quantities in events where the electron-jet is positively tagged (identified conversions are vetoed). From top-left, clockwise: electron E_T, electron-jet E_T, away-jet E_T, electron p_T.* 71
6.2 *Data/Monte Carlo comparison of some quantities of tagged electron-jets (identified conversions are vetoed). From top-left, clockwise: number of selected tracks, number of tracks in vertex, and vertex mass of positively tagged electron-jets; pseudo-$c\tau$ of (positively or negatively) tagged electron-jets.* 72
6.3 *Schematic composition of the electron-jet with (bottom) and without the requirement that the away-jet be positively tagged. Among light electron-jets (conversion electrons and fake electrons), the fraction of conversion electrons is independent of the away-jet heavy flavor property.* 72
6.4 *Radius of identified conversions in data.* 73
6.5 *Reconstructed $K\pi$ mass in data.* 75

6.6 *Reconstructed $K\pi$ mass in Monte Carlo. No requirement on the away-jet has been applied, in order to increase statistics.* 76
6.7 *A fit of flavor templates to the SecVtx mass distribution. The round markers are the data and the line shows the sum of the fitted templates.* 77
6.8 *Efficiency to tag a b jet as a function of jet E_T in the low p_T electron sample for data and Monte Carlo (top), and data/MC scale factor (bottom). Errors include statistical uncertainties, and systematic uncertainties due to F_{HF}.* 80
6.9 *Pseudo-$c\tau$ of positively tagged electron-jets in which the electron is identified as conversion.* 82
6.10 *Top: positive excess tag rate in Jet 50 data and* PYTHIA *MC vs jet E_T. Bottom: ratio of data over MC for silicon-fiducial jets, and taggable jets.* 84
6.11 *Number of COT tracks in electron-jet for data (solid line) and Monte Carlo.* 85
6.12 *Number of default silicon tracks in electron-jet for data (solid line) and Monte Carlo.* 85
6.13 *Number of selected silicon tracks in electron-jet for data (solid line) and Monte Carlo. Jets in the Monte Carlo contain in average 12% more selected tracks than in the data.* 86
6.14 *Scale factor relative to the number of selected tracks vs electron-jet E_T.* 86
6.15 *SecVtx negative tag rate as a function of jet E_T and track multiplicity in the inclusive jet data.* 89

7.1 *Observed W + Jets cross section compared with the* ALPGEN *W + Jets prediction as a function of number of jets.* 93
7.2 *Pseudo-$c\tau$ distribution for jet data, including fitted contributions for the different components of heavy flavor and secondary interactions in light flavor jets.* 95
7.3 *Distribution of closest jets in $\Delta\phi$ for tagged 3-jet data events with fake tag prediction (top) and for double-tagged events in which the tagged jets are also the two closest jets (bottom).* 97
7.4 *Positive tag excess rate in data and* ALPGEN *Monte Carlo as a function of $\Delta\phi$.* 98

8.1 *Tagging rate (per taggable jet) at high and low $\not\!\!E_T$ vs electron isolation, for events with 1 jet only.* 104
8.2 *Electron isolation before and after tagging in the high and low $\not\!\!E_T$ regions.* 105
8.3 *Muon isolation before and after tagging in the high and low $\not\!\!E_T$ regions.* 106
8.4 *Dependence of the isolation cut for the electron and muon channels on the tagged QCD background estimate.* 108
8.5 *Tagging rate as a function of H_T for electron events with 2 or more jets in regions A, B, C, and D.* 109
8.6 *Tagging rate as a function of H_T, for muon events with 2 or more jets in regions A, B, C, and D.* 110

8.7 *Comparison of observed and predicted negative SecVtx tags versus jet E_T in the lepton + jets sample.* 111
8.8 *Number of events passing selection criteria with at least one tagged jet and the background prediction.* 111

9.1 *Distribution of the H_T variable for $t\bar{t}$ Monte Carlo, and for various backgrounds normalized to an integrated luminosity of* $107\,\mathrm{pb}^{-1}$. 114
9.2 *Estimate of S/B, statistical and total significance for* $107\,\mathrm{pb}^{-1}$ *integrated luminosity, as a function of H_T cut.* 115
9.3 *Isolation of the electrons and muons in* PYTHIA *$t\bar{t}$ Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after tight lepton selection. The event selection requires* $I > 0.1$. 117
9.4 *Transverse missing energy in* PYTHIA *$t\bar{t}$ Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after tight lepton selection and* $I > 0.1$. *The event selection requires* $\not{E}_T > 20$ *GeV.* 118
9.5 *Transverse energy of the first four jets (in decreasing order of transverse energy) satisfying* $|\eta| < 2$ *in* PYTHIA *$t\bar{t}$ Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after tight lepton selection,* $I > 0.1$, *and* $\not{E}_T > 20$ *GeV, with and without the* $H_T > 200$ *GeV requirement. The event selection requires the third jet to have* $E_T > 15$ *GeV.* . . . 118
9.6 *Number of jets with* $E_T > 15$ *GeV and* $|\eta < 2|$ *in* PYTHIA *$t\bar{t}$ Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after tight lepton selection,* $I > 0.1$, *and* $\not{E}_T > 20$ *GeV. The event selection requires at least three jets.* 119
9.7 *SecVtx b-tagging efficiency for b jets with* $E_T > 15$ *GeV and* $|\eta < 2|$ *in* PYTHIA *$t\bar{t}$ Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after full event selection (no requirement on H_T) as a function of the jet transverse energy. The jets are required to match a b parton.* 120

10.1 *Comparison of the observed and predicted number of events in the b-tagged* Z^0 *+ Jets sample.* 125
10.2 *Background and $t\bar{t}$ signal expectation as a function of jet multiplicity.* 127
10.3 *Comparison of the $t\bar{t}$ production cross section measured in Run I at* $\sqrt{s} = 1.8$ *TeV [53] (combining measurements using a soft lepton tagger and a secondary vertex tagger in the Lepton+Jets channel), the Run II measurement at* $\sqrt{s} = 1.96$ *TeV, and the Standard Model prediction [9]* ($m_t = 175\,\mathrm{GeV}/c^2$). 128
10.4 *H_T distribution of the tagged candidates with one and two jets in 162* pb^{-1} *of data, compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross section of 6.7 pb).* 129
10.5 *H_T distribution of the 57 tagged candidates in 162* pb^{-1} *of data (three or more jets), compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross section of 6.7 pb).* 129

10.6 *Transverse energy of the tagged jets in the 48 tagged candidates in 162* pb^{-1} *of data (three or more jets,* $H_T > 200$ *GeV), compared to the expected background and* $t\bar{t}$ *signal (normalized to the theoretical cross section of 6.7 pb).* 130
10.7 *Transverse decay length of the secondary vertices in the 48 tagged candidates in 162* pb^{-1} *of data (three or more jets,* $H_T > 200$ *GeV), compared to the expected background and* $t\bar{t}$ *signal (normalized to the theoretical cross section of 6.7 pb).* 130
10.8 *Reconstructed mass of the secondary vertices in the 48 tagged candidates in 162* pb^{-1} *of data (three or more jets,* $H_T > 200$ *GeV), compared to the expected background and* $t\bar{t}$ *signal (normalized to the theoretical cross section of 6.7 pb).* 131
10.9 *Pseudo-*$c\tau$ *of the secondary vertices in the 48 tagged candidates in 162* pb^{-1} *of data (three or more jets,* $H_T > 200$ *GeV), compared to the expected background and* $t\bar{t}$ *signal (normalized to the theoretical cross section of 6.7 pb).* 131

11.1 *Comparison of the result with some other current CDF Run II* $t\bar{t}$ *production cross section measurements. The systematic uncertainties (blue) are added in quadrature with the statistical ones (red).* 133

C.1 *CES photon efficiency (Run I minimum bias event data).* 150
C.2 *Dijet mass distribution in* $Z^0 \to b\bar{b}$ *events. ISR and FSR on. QFL (dotted line) is compared to the toy detector with CDF segmentation (solid line :* $\frac{100\%}{\sqrt{E}}$ *HAD resolution; dashed line :* $\frac{50\%}{\sqrt{E}} + 3\%$ *HAD resolution).* 152
C.3 *Fraction of EM and HAD energy from particles with no associated track in Track and Mixed towers. The mean values determine the parameters used in the T&C definition.* $Z^0 \to b\bar{b}$ *sample with CDF segmentation. ISR and FSR on. Real efficiencies.* 154
C.4 *Ratio of jet* p_T *to corresponding quark* p_T *(top) and dijet mass (bottom).* $W \to q\bar{q}$ *sample with perfect resolution and efficiencies. CDF segmentation. ISR and FSR off. T&C (dashed line) is compared to the calorimeter only (solid line).* 157
C.5 *Same as figure C.4 on the* $Z^0 \to b\bar{b}$ *sample.* 158
C.6 *Ratio of tower T&C energy to calorimeter only energy (with perfect calorimeter resolution), assuming perfect efficiencies (solid line) and real efficiencies (dashed line).* $Z^0 \to b\bar{b}$ *sample, ISR and FSR on. a)Track towers, no neutral. b)Mixed towers, no neutral. c)Track towers with neutral(s). d)Mixed towers with neutral(s).* 158
C.7 *Number of towers of each type. Solid line : CDF segmentation. Dashed line : EM and HAD finer segmentation (0.05×0.05).* $Z^0 \to b\bar{b}$ *sample, ISR and FSR on. Perfect efficiencies are assumed.* 159

C.8 *Fraction of the jet energy for each type of tower. Solid line : CDF segmentation. Dashed line : EM and HAD finer segmentation (0.05×0.05).* $Z^0 \to b\bar{b}$ *sample, ISR and FSR on. Perfect efficiencies are assumed.* . 160
C.9 $W \to q\bar{q}$ *with real resolution and efficiencies. CDF segmentation. ISR and FSR on.* 161
C.10 $Z^0 \to b\bar{b}$ *with real resolution and efficiencies. CDF segmentation. ISR and FSR on.* 162
C.11 *same as figure C.10 except HAD resolution* $= \frac{50\%}{\sqrt{E}} + 3\%$. 162
C.12 $Z^0 \to b\bar{b}$ *with real resolution and efficiencies. T&C with 3 different segmentations : CDF seg., finer EM seg., finer EM and HAD seg. (0.05×0.05). ISR and FSR on.* 163

D.1 *SVX Error Logger GUI* 168
D.2 *Example plot of occupancy vs strip number for the phi side of the SVX half-ladder in the west barrel, west half-barrel, layer 2, wedge 4.* . . . 173
D.3 *Example pulse height distribution for the L00 westmost sensor in wedge 0. The SVXMon command "histo strips2d adc {west west 0 0 phi} {nevents} 0" has been used to book this plot.* 175
D.4 *Example plot of RMS charge vs the strip and pipeline cell numbers for the L00 westmost sensor in wedge 0. This plot has been booked with SVXMon command "histo strips2d cellid {west west 0 0 phi} {stdev} 0".* 175
D.5 *Example plot of number of channels read out vs the SRC bunch crossing number for the L00 westmost sensor in wedge 0. Together with the plot on the next figure, it has been booked with the command "histo daqparam bunchx {west west 0 0 phi} {nhits mean} 0". Since L00 is used in "read all" mode, this plot essentially displays the distribution of bunch crossing numbers (up to a constant multiplication factor).* . 177
D.6 *Example plot of mean read out charge vs the SRC bunch crossing number for the L00 westmost sensor in wedge 0. Since the mean charge is relatively flat across all bunch crossings present in the run, we can deduce that the front-end preamp reset signal has been safely confined to beam gaps.* 178
D.7 *Example plot of the bunch crossing number vs the event number.* . . 179
D.8 *Example chip occupancy plots for layers L00, L0, L1 and L2.* 180
D.9 *SVT Tracking Occupancy.* 182
D.10 *Example of track quality plot (here for SVT algorithm). When the SVT algorithm is chosen, this canvas shows the average* χ^2 *of the tracks. For the other algorithms, it shows the average residual of hits on tracks.* 182
D.11 *Example plot of chip cell id distribution. This chip's pipeline appears to be in sync with the rest of the detector.* 184
D.12 *Example SVX pipeline status map.* 185

D.13 *Distribution of the number of chips whose pipeline has lost synchronization. One entry is added to this histogram every time SVXMon plots are updated. The two solid bins are the ones most recently updated.* 186
D.14 *Example history plots of median chip occupancy by layer and side.* . . 186
D.15 *Example SVX chip status map.* 187
D.16 *Example occupancy monitoring plots for the west side of the west barrel in layers L3 to ISL1. The chip numbers on the horizontal axis increase with the phi wedge number. Within each wedge they increase in readout order. The monitored quantity is drawn in green color if everything is normal, or in red if there is a problem. The acceptable range limits are shown in blue.* 194
D.17 *Single chip occupancy, mean charge, charge standard deviation, and stdevVar, for reference runs (black line) and the run being compared (red line). The data for this chip looks like its ladder was unbiased, and an alarm should go off.* 201
D.18 *Occupancy significance of all chips in the run being compared. The spike at 0 comes from chips that never send any data (due to truncated readout chains). The rest of the distribution is a Gaussian centered at 0, with an rms close to 1: the run shows no deviation from the reference runs.* 202
D.19 *Chip Mean Charge for a given run (red line) and a set of reference runs (black line).* 204
D.20 *Chip Occupancy vs Run Number.* 205

List of Tables

1.1 *The elementary particles of the Standard Model and their masses [4].* 3

2.1 *Quantum numbers associated with the fermions in the electroweak theory: Q is the electric charge; T_3 is the third component of the weak isospin; Y_W is the weak hypercharge ($Q = T_3 + Y_W/2$).* 9
2.2 *$t\bar{t}$ decay modes and branching ratios (the particle charge has been omitted for simplicity).* 14

3.1 *Properties of the SVX modules.* 30
3.2 *Properties of L00 and ISL.* 31
3.3 *Some properties of the CDF calorimeters. The last row shows the single particle energy resolution.* 35

4.1 *Electron selection cuts. All energies (momenta) are in GeV (GeV/c).* 50
4.2 *Muon selection cuts. All energies (momenta) are in GeV (GeV/c).* . . 51

5.1 *Bins used to define the track classes (280 classes total). For 3-hit tracks, the hit pattern (i.e. what layers the hits are on) is also considered.* . 66
5.2 *In addition to the SVX requirements specified above, the associated COT track is required to have at least two axial and two stereo superlayers with at least 6 hits each, and a total of at least 19 axial hits (16 stereo hits). "gd" = number of good hits; "m" = number of missed hit.* 67

6.1 *Numbers of* $\mathrm{e}-\mu$ *pairs and* F_{b} *for the muon method. Errors on* F_b *are statistical (data and MC, resp.), and systematic (reconstruction efficiency.)* 76
6.2 *Sources of uncertainty on F_b, for each method, and on F_{HF} (relative errors, in %).* 78
6.3 *Tag counts and conversion yields found in the inclusive electron data and Monte Carlo.* 79
6.4 *Efficiency to tag a b-jet in the low p_T electron sample for data and Monte Carlo, and data/MC ratio (tagging scale factor). Errors on efficiencies are statistical only. The error on the scale factor includes both statistical and systematic uncertainties.* 79
6.5 *Sources of uncertainty on the SecVtx scale factor (relative errors, in %).* 79

6.6 *Efficiency in data and Monte Carlo for various values of the conversion jet negative tag scale factor ($f_C = 1$ corresponds to default).* 81
6.7 *Efficiency in data and Monte Carlo for various values of the generic jet negative tag scale factor ($f = 1$ corresponds to default).* 83
6.8 *Differences in predicted and observed negative tagging rates for various samples. The first four rows with labels of the form Sample1-Sample2 compare observed tag rates in Sample 2 to the rates predicted by a matrix made from Sample 1. The last three rows compare the observed tag rates for trigger jets, non-trigger jets, and jets in the ΣE_T sample with predictions from the standard mistag matrix derived from all four jet samples.* 90
6.9 *Systematic uncertainties assigned to the negative tag rate matrix.* . . 90

7.1 *Summary of systematic uncertainties in the heavy flavor fraction determination.* 94
7.2 *Fitted contributions from b, c and secondary interactions / long-lived light flavor particles in data events. The errors on the b and c fraction include the template systematic uncertainty. The ratio $\Delta N/N$ estimates the excess of positive over negative tags in data events, due to secondary interactions and long-lived light flavor particles.* 96
7.3 *Ratio of W + Heavy Flavor production to total W + Jets production, for different jet multiplicities. The heavy flavor ratios include the correction factor 1.5 ± 0.4 as measured from the data, and the SecVtx event tagging efficiencies include the scale factor described in Chap. 6.* 99

8.1 *Number of electron events in data, and corrections, for pretag events.* 103
8.2 *Number of muons events in data, and corrections, for pretag events.* . 103
8.3 *Pretag QCD fraction and tagged QCD background estimates (using the tagging rate in region B).* 103
8.4 *Number of electron events in data, and corrections, for tagged events. Regions A and C are defined with $I > 0.1$.* 107
8.5 *Number of muon events in data, and corrections, for tagged events. Regions A and C are defined with $I > 0.1$.* 107
8.6 *QCD background estimates (using the tagged event method).* 108
8.7 *Tagged QCD background estimates (combined methods, electrons and muons).* 109

9.1 *Number of events in the $t\bar{t}$ MC sample passing the CEM electron selection cuts.* 116
9.2 *Number of events in the $t\bar{t}$ MC sample passing the CMUP muon selection cuts.* 116
9.3 *Number of events in the $t\bar{t}$ MC sample passing the CMX muon selection cuts.* 117

9.4 *Summary table of the $t\bar{t}$ acceptance, for a top quark mass of* $175\ \mathrm{GeV}/c^2$. *Only systematic uncertainties are shown (the statistical uncertainties from the Monte Carlo sample are negligible).* 121
9.5 *Relative systematic uncertainties (%) on the signal acceptance relative to CEM electrons, CMUP muons, and CMX muons.* 121
9.6 *Relative systematic uncertainties on the signal acceptance which are common to all lepton types.* 122

10.1 *The predicted number of* Z^0 + *Jets events and the observed number, along with the* Z^0 + *Jets contribution in the* W + *Jets sample and the estimate of the resulting extra* b*-tags in that sample. (The prediction of extra* b*-tagged events is included in the predicted background summary for the* W + *Jets sample.)* 125
10.2 *Background summary for the single-tag selection. The total backgrounds are given before and after the correction for* $t\bar{t}$ *events in the pretag W+jets sample.* 126
10.3 *Measured cross section for different top quark mass assumptions.* . . . 127
10.4 *Relative uncertainties contributing to the* $t\bar{t}$ *production cross section error.* 127

B.1 *Positively-tagged jets in* W + *3 jet events.* 144
B.2 *Tagged jets in* $W + \geq$ *4 jet events.* 145

C.1 *Number of towers and fraction of energy for each type of tower. CDF segmentation,* $Z^0 \rightarrow b\bar{b}$ *sample, perfect resolution and efficiencies, ISR and FSR on.* 159
C.2 *Same as table C.1, except segmentation : 0.05×0.05 for both EM and HAD calorimeters.* 160
C.3 *Gaussian fit* σ *improvement with T&C reconstruction compared to Calorimeter-only reconstruction (after offset correction).* 164

D.1 *Description of the run block variables.* 195
D.2 *Description of the chip array block variables.* 197

Acknowledgements

When James Siegrist invited me to visit Berkeley six years ago for an internship to complete my Master's thesis, I did not expect to settle in this country for so many years. I want to thank him for giving me such a wonderful opportunity, and for his constant support and guidance throughout the years as my adviser.

I spent my first six months at Berkeley working on the ATLAS experiment with Ian Hinchliffe. This was my first research experience in particle physics, and a very rewarding one. Ian's passion for physics is legendary, and I tried to learn as much as I could from him. Ian also advised me to apply to graduate schools in the US. I would probably not be here if he had not done so.

Laurent Vacavant arrived at Berkeley soon after I did. His patience for answering my most inane questions and for bailing me out at some critical times has been invaluable. I would also like to thank him for giving me his priceless three-cylinder car.

Weiming Yao was my mentor and like a second adviser as I worked with the CDF group at LBL. His wisdom, judgment, and immense knowledge of experimental particle physics were inspiring and guided me toward the completion of my PhD. I am grateful that he took the time to try to transfer his experience to me.

I spent a lot of time with Jason Nielsen working on the analysis and commuting between Chicago and Fermilab. He deserves special thanks for helping me in difficult times, and for always doing so with a smile on his face. I owe him a lot. His performance of *Sex Bomb* will not be forgotten either.

The CDF group at LBL was a constant source of support. I want to thank them all here. In particular, Lina Galtieri, for leading the group so enthusiastically. Marjorie Shapiro, for always being available and giving me sound advice when I needed it the most. Igor Volobouev, for showing me the ropes of the silicon detector and of object-oriented programming.

I also want to thank Mark Strovink and Eugene Haller for reading my thesis with scrutiny and for their insightful comments.

I met so many people during my stay in California and in Illinois that, I am sure, I will forget to thank some of them. My Berkeley friends: Arthur, Jean-David, Tiffany, Alika. Vicky, for telling me that I am not that old after all. All my friends at LBL: Nicolas, Frederic, Klas, Jeroen, Gwenael, Pierre-Antoine, Aldo, Sven. Mona, for long discussions about how hard it is to be a graduate student. Anne Takizawa, for making me feel at home the first day I entered grad school. Special thanks to Charles for the relaxing breaks during the writing period. I apologize to Aathavan and my numerous roommates for neglecting my chores during the same period. Bill and Myriam became a second family, after so much time spent on the water and at the Berkeley Yacht Club. Estrellita, my little star. Sorcerer and its crew: Greg, Melanie, Ivan, Matt. Jennifer, who always chooses to live in wonderful places and whose hospitality is unique.

My friends from Chicago, and Stéphane Tourneur, for introducing me to them:

Julia, Marie-Jeanne, Mike, and so many others. The crew of Jeannine, for their exoticism and for giving me an unforgettable ride to Mackinac. Ju Mei Wei and Dino Kostas for restoring my health: no doubt that I would not have completed this thesis without their help. Marta for going through my injury with me and showing me the best of Chicago. Jill and James: you should get married. Alex and Cristian shared two winters in Aurora with me, and their house was always open for me. My fellow graduate students: Hung-Chung, Johannes, Amanda, John, Erik, Adam, and Greg. Amy, for the LaTex templates, the good tips about writing the thesis, and her good mood. Chris, for so many rides to Fermilab.

In the "old Europe": Pailin, Anmai, Yann, Laurent, Christophe, Fred, and Nicolas, for keeping in touch with me in spite of the distance. Julien, for making sure that I don't completely lose my French identity.

Finally, I want to thank my family for supporting me even when I made the difficult decision to live so far away from them.

Chapter 1

Introduction

In the 4^{th} century B.C., Democritus was the first one to reject the idea that matter is a continuum and suggested instead that matter is made of elementary — indivisible — particles, that he named "atoms". He went one step further by postulating that particles are of four kinds — air, water, stone, and fire — and that any material is made of a combination of them, in some specific proportions. Two millennia later, modern physics was indeed able to describe matter as made of a handful of particles, although not of the kind Democritus expected. Whether those particles are truly "elementary", or can be broken into more fundamental constituents, is an open question, but the corpuscular nature of matter has been established, first with the development of chemistry in the early 19^{th} century, then with the creation of modern particle physics at the turn of the 20^{th} century. In particular, J. J. Thompson showed in 1897 that the electron is a particle, not a wave, by measuring the mass and charge of an individual electron. In 1911, Rutherford conducted an experiment that marks the very beginning of experimental particle physics as we know it today: Rutherford studied how α particles produced by a radioactive source interact with a thin foil of gold. Although very rudimentary, this was the first scattering experiment ever made, and similar techniques are still used nowadays.

During the 20^{th} century, the development of the theory of relativity and of quantum mechanics, in parallel with the improvement of experimental techniques, has completely revolutionized our understanding of the world. In the 1970's, the Standard Model (Chap. 2) emerged as a theory able to describe the interactions of all known particles. For decades, the Standard Model predictions were tested at an extreme level of precision, and no significant discrepancy between the theory and the laboratory experimental measurements was found. In the recent years, however, several observations cannot be explained by the Standard Model and call for extensions of the theory (Sec. 2.2). One of the main successes of the Standard Model was to predict the existence of the $W^{\pm}$ and Z^0 gauge bosons, discovered at the SPS accelerator at CERN, in 1983, confirming the electroweak model invented by Sheldon Glashow, Abdus Salam, and Steven Weinberg [5] a decade earlier. The spectrum of particles in the Standard Model is shown in Tab. 1.1. The Standard Model describes the interactions of quarks and leptons under the influence of three forces: the electromagnetic force, mediated by photons (γ) and the weak force, mediated by the heavy $W^{\pm}$, Z^0 bosons, are unified within the electroweak model; the strong force, mediated by gluons (g), affects only the quarks and is described by Quantum Chromo-Dynamics (QCD). Quarks and leptons (spin 1/2 fermions) come in three generations, which differ only by the mass of the particles. To each fermion also corresponds an antiparticle (not shown in the table), which has the same properties, but comes with opposite charges. The Higgs boson, yet to be observed, is the result of the Higgs mechanism responsible for breaking the electroweak symmetry and giving mass to the $W^{\pm}$ and Z^0 bosons and to the fermions. The energy at which the electroweak symmetry breaks is called the electroweak scale. It should be noted that the fourth known force, gravity, is not described by the Standard Model. Indeed, combining gravity with quantum theories has proven to be a difficult task; theories such as superstring theories try to address

Generation	1^{st}	2^{nd}	3^{rd}
Leptons	e	μ	τ
(spin 1/2)	0.511 MeV	106 MeV	1.78 GeV
	ν_e	ν_μ	ν_τ
	< 3 eV	< 0.19 MeV	< 18.2 MeV
Quarks	u	c	t
(spin 1/2)	1.5-4 MeV	1.15-1.35 GeV	178 GeV
	d	s	b
	4-8 MeV	80-130 MeV	4.5 GeV

Gauge bosons (spin 1)	γ	$W^\pm$	Z^0	g
	0	80.4 GeV	91.2 GeV	0
Higgs boson (spin 0)	H^0			
	> 114.4 GeV			

Table 1.1: *The elementary particles of the Standard Model and their masses [4].*

this issue. Except for the Higgs boson, every particle of the Standard Model has been directly observed. The top quark was discovered in 1995 [1] after a long search at the Tevatron collider by the Collider Detector at Fermilab (CDF) and the D0 experiments, completing the 3^{rd} generation of quarks. The Tevatron and CDF, which are used for the analysis presented in this thesis, are described in Chap. 3. Although very little doubt remained about its existence, the tau-neutrino ν_τ was directly observed by the DONUT experiment [2] as late as July 2000, completing the 3^{rd} generation of leptons.

Because of its surprisingly large mass ($178.0 \pm 4.3\,\mathrm{GeV}/c^2$ [3]), the top quark can only be produced through collisions of very energetic particles, which requires extremely powerful accelerators. The Tevatron at Fermilab (Illinois, USA) is currently the most powerful accelerator in the world, colliding protons and antiprotons head-on at a center-of-mass energy of 1.96 TeV. Until the Large Hadron Collider (LHC) turns on at CERN (Switzerland) in 2007, the Tevatron is the only place where top quarks can be produced and observed. Besides its mass, which is known at the 3% level, the top quark properties are only poorly measured, if at all, and require further studies. In particular, the top quark production mechanisms and decay branching ratios are well predicted by the SM, but have been only poorly measured during the Tevatron Run I, due to the lack of a sufficiently large data sample. The Tevatron Run II, which officially started in February 2002 and will run until 2009 should accumulate a significantly larger data set and allow more precise measurements. The large mass of the top quark, of the order of the electroweak scale, is responsible for its particular properties; it is also a hint that it could play a special role in the breaking of the electroweak symmetry. For these reasons, precision measurements of the top quark are important to indirectly test the Standard Model, but could also happen to directly reveal new forms of phenomena beyond the Standard Model.

This thesis presents a measurement of the $t\bar{t}$ production cross section at CDF using the 162 pb^{-1} of data taken at the Tevatron between February 2002 and September 2003. It is organized as follows. Chapter 2 is a brief introduction to quantum field theory, the Standard Model, and top quark physics, as well as an overview of the analysis, and the motivation for it. Chapter 3 describes the experimental setup: the chain of accelerators, the Tevatron collider, and the Collider Detector at Fermilab (CDF). Chapter 4 describes the event reconstruction, the data samples, and the event selection. Chapter 5 goes into the details of the secondary vertex b-tagging algorithm. Chapter 6 describes the measurement of the b-tagging algorithm efficiency and fake rate. Chapter 7 describes the estimation of the heavy flavor composition of the W + Jets sample, necessary to understand the background due to W + Heavy Flavor production. Chapter 8 describes the estimate of the various backgrounds. Chapter 10 shows Z^0 + Jets and W + Jets data samples used for this analysis and gives the result of the measurement. Chapter 11 discusses the result and concludes on the prospects of top quark physics. A derivation of the method used to measure the b-tagging algorithm efficiency is given in Appendix A. A list of the candidate events together with some of their characteristics is given in Appendix B. Appendix C presents a study of a jet algorithm that combines both tracking and calorimeter information in an attempt to improve the jet energy measurement resolution. Appendix D presents the SvxMon software dedicated to monitoring the silicon vertex detector.

Chapter 2

Theoretical Overview

2.1 The Standard Model of Particle Physics

The Standard Model of particle physics (SM) was developed in the 1970's and has been astonishingly successful in describing all known particles and their interactions over a large range of energies: the SM has been tested extensively up to the electroweak scale ($\approx$ 200 GeV) and some of its predictions have been tested with incredible precision. However, some theoretical caveats and recent observations of neutrino oscillations, dark matter, and dark energy imply that the SM is an incomplete theory.

The SM is a quantum field theory that describes interactions of spin 1/2 fermions (see Tab. 1.1): the quarks and the leptons come in three generations, which differ only by their masses. To every fermion corresponds an antiparticle (not shown in the table), which has the same properties (especially, same mass), but opposite charges. For a given force, the "charge" of a particle is the quantum number that defines the coupling of this particle to the force. The interactions between fermions are mediated by spin-1 gauge bosons, and are dictated by local gauge invariance. The strong force, which affects only quarks, follows the SU(3) gauge symmetry and is mediated by gluons. The electromagnetic and weak forces follow the SU(2)$\otimes$U(1) gauge symmetry, and unify at the electroweak scale; they are mediated by the photon, and the massive $W^{\pm}$ and Z^0 bosons. The electroweak symmetry is broken through the Higgs mechanism in order to give mass to the $W^{\pm}$ and Z^0 bosons and to the fermions.

The hierarchy in the masses of fermions (Table 1.1) is yet to be explained: the second generation is heavier than the first, and the third is heavier than the second by several orders of magnitude. The electrically charged leptons (charge -1) are the electron (0.511 MeV), the muon (106 MeV) and the tau (1.78 GeV); the neutral leptons are the neutrinos (ν_e, ν_μ, ν_τ). In the SM, neutrinos are massless. There is recent evidence of neutrino flavor-to-flavor oscillations, implying that neutrinos do have a mass, which can be accommodated in the framework of the SM but is a clear sign of physics beyond the SM. The current upper limit on the electron neutrino mass is 3 eV [4]. Limits on the mass of the other two generations of neutrinos from direct measurement are much larger. However, neutrino oscillations allow measurements of *differences* of mass between generations; those differences are small, so that neutrinos are not expected to have a mass larger than $\approx$ 3 eV.

The quarks follow the same mass hierarchy as the charged leptons: the up and down quarks (1.5-4 MeV and 4-8 MeV, resp.) which are the building blocks of ordinary matter form the first generation; the charm (1.15-1.35 GeV) and strange (80-130 MeV) quarks form the second generation; the top ($\approx$175 GeV) and bottom ($\approx$4.5 GeV) quarks form the third.

2.1.1 The Strong Force: Quantum Chromodynamics

Quantum Chromodynamics (QCD) describes the strong interaction between quarks and gluons and follows an SU(3) gauge symmetry. The charge associated with the strong force is named "color". Each quark comes in three colors (red, green or blue). Eight massless gluons mediate the strong force between quarks. The corresponding QCD Lagrangian is:

$$\mathcal{L}_{QCD} = -\frac{1}{4} F^{(a)}_{\mu\nu} F^{(a)\mu\nu} + i \sum_q \bar{\psi}^i_q \gamma^\mu (D_\mu)_{ij} \psi^j_q - \sum_q m_q \bar{\psi}^i_q \psi_{qi} \qquad (2.1)$$

with:

$$F^{(a)}_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_S f_{abc} A^b_\mu A^c_\nu$$

$$(D_\mu)_{ij} = \delta_{ij}\partial_\mu + i g_S \sum_a \frac{\lambda^a_{ij}}{2} A^a_\mu$$

where g_S is the strong coupling constant (usually expressed in the form $\alpha_S = \frac{g_s^2}{4\pi}$), f_{abc} are the structure constants and λ^a are the generators of the SU(3) Lie algebra; ψ^i_q is the Dirac spinor for a quark of color i and flavor q; A^a_μ is the Yang-Mills gluon field corresponding to a $\mu - \bar{a}$ color exchange; γ^μ are the Dirac matrices; m_q is the quark mass. It should be noted that quarks are allowed to self-interact, as a result of the non-commutative nature of the SU(3) group. A consequence of this is that the renormalized effective QCD coupling decreases as the interaction energy scale increases:

$$\alpha_S(q^2) = \frac{4\pi}{(11 - \frac{2}{3}N_f(q^2)) \ln\left(\frac{-q^2}{\Lambda^2_{QCD}}\right)} \qquad (2.2)$$

where Λ_{QCD} is the QCD energy scale, and $N_f(q^2)$ is the number of quark flavors that can kinematically be pair produced at a given energy, *i.e.* the number of quark flavors with $m < \sqrt{-q^2}/2$. The coupling is small enough at high energy to allow perturbative calculations: this behavior is referred to as "asymptotic freedom". At low energy, however, the coupling becomes strong and diverges at $\Lambda_{QCD} \approx 0.2$ GeV, making perturbative calculations impossible. Techniques such as lattice QCD are being developed in order to cope with this non-perturbative behavior. This strong coupling at low energy is responsible for the phenomenon of quark confinement: in nature, quarks are not observed in an isolated state. Instead, they are confined in color-singlet bound states of combinations of two or three quarks and antiquarks, the hadrons. When a quark or a gluon is produced with a large momentum as the result of an interaction, quark antiquark pairs are produced and recombine themselves into hadrons (hadronization). Experimentally, such a process leads to the observation of a large number of collimated particles, called a jet. An exception to this is the top quark:

because of its exceptionally large mass, the top quark decays before hadronization takes place (see Sec. 2.3 for details).

2.1.2 The Electromagnetic and Weak Forces: the Electroweak Theory

The electromagnetic and weak interactions are unified within the electroweak theory developed by Glashow, Weinberg and Salam in the 1960's [5]. The electroweak theory obeys a SU(2)⊗U(1) gauge symmetry and describes the interactions of leptons and quarks mediated by four vector bosons: B^0 corresponds to the U(1) generator; W^1, W^2 and W^3 correspond to the SU(2) generators. Two new quantum numbers are introduced: the hypercharge relates to interactions with B^0, while the weak isospin couples fermions to the W^i fields. Under the SU(2) group, left-handed[1] states are doublets, while right-handed states are singlets. Table 2.1 summarizes the spectrum of fermionic fields in the electroweak theory together with their quantum numbers. In the SM, neutrinos are all left-handed and massless. Given recent evidences that neutrinos do have a mass, right-handed neutrinos should exist as well, but do not couple at all, and have never been observed.

As is, the theory is made of massless particles only; the Higgs mechanism (next section) gives mass to the fermions, and mixes the B^0 and W^3 fields into a massless state A (the photon), and a massive state Z^0 (the Z^0 boson):

$$Z^0 = -B^0 \sin\theta_W + W^3 \cos\theta_W \tag{2.3}$$

$$A = B^0 \cos\theta_W + W^3 \sin\theta_W \tag{2.4}$$

where θ_W is a free parameter of the theory, the weak angle. The $W^\pm$ bosons are a combination of the W^1 and W^2 fields: $W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp W^2)$. The hypercharge Y_W and the third component of the weak isospin T_3 are related to the electric charge by the relation $Q = T_3 + Y_W/2$. The photon couples to fermions in proportion of their electric charge. The Z^0 boson coupling (neutral current) is relative to both the electric charge and the weak isospin third component. Finally, the $W^\pm$ bosons (charged current) couple only to the left-handed component of the fermions; the coupling is the same for every fermion. Because the Z^0 and $W^\pm$ bosons are massive, the weak force is short-ranged ($\approx 10^{-15}$ m). In the limit of low energy, the weak force is identical to the point-like Fermi coupling: $\frac{g^2}{q^2 - m_W^2} \to g^2/m_W^2 = \frac{8}{\sqrt{2}} G_F$ at low momentum transfer q^2, where g is the weak coupling constant and G_F is the Fermi constant.

The quark QCD eigenstates are not exactly electroweak eigenstates, so that the three generations of quarks mix. This mixing is parametrized by the Cabibbo-

---

[1] A particle is said to be left-handed if its spin is opposite to its momentum.

	1^{st}	2^{nd}	3^{rd}	Q	T_3	Y_W
Leptons	$\binom{\nu_e}{e}_L$	$\binom{\nu_\mu}{\mu}_L$	$\binom{\nu_\tau}{\tau}_L$	$\binom{0}{-1}$	$\binom{1/2}{-1/2}$	-1
	e_R	μ_R	τ_R	-1	0	-2
Quarks	$\binom{u}{d'}_L$	$\binom{c}{s'}_L$	$\binom{t}{b'}_L$	$\binom{2/3}{-1/3}$	$\binom{1/2}{-1/2}$	1/3
	u_R	c_R	t_R	2/3	0	4/3
	d_R	s_R	b_R	-1/3	0	-2/3

Table 2.1: *Quantum numbers associated with the fermions in the electroweak theory: Q is the electric charge; T_3 is the third component of the weak isospin; Y_W is the weak hypercharge ($Q = T_3 + Y_W/2$).*

Kobayashi-Maskawa (CKM) 3x3 unitary matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{2.5}$$

where (q') are weak eigenstates and (q) are mass eigenstates. The non-diagonal elements of the matrix are small, but different from 0, allowing for the weak interaction to couple quarks of different generations.

2.1.3 The Higgs Mechanism

If the electroweak symmetry were exact, the $W^{\pm}$ and Z^0 bosons and the fermions would be massless, in obvious contradiction with observations. A solution to this problem is to introduce in the theory a new scalar field with non-vanishing vacuum expectation value, that spontaneously breaks the electroweak Lagrangian, giving mass to both the $W^{\pm}$ and Z^0 bosons, and the left-handed fermions (leaving the neutrinos massless): this is known as the Higgs mechanism [6].

The Higgs field of the SM is a weak isospin doublet of complex scalar fields $\Phi = \binom{\Phi^+}{\Phi^0}$ in a potential of the form:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2 \tag{2.6}$$

where λ is positive and μ^2 is negative, so that the potential reaches a degenerate minimum in $|\Phi| = \sqrt{-\mu^2/2\lambda}$ (see Fig. 2.1). By "choosing" one particular minimum and acquiring a non-vanishing vacuum expectation value, the Higgs field spontaneously breaks the vacuum symmetry, while the Lagrangian symmetry is preserved. Without loss of generality, this particular minimum can be assumed to be $\Phi_0 = \frac{1}{\sqrt{2}}\binom{0}{v}$, where $v = \sqrt{-\mu^2/2\lambda}$ is the vacuum expectation value of the Higgs field. After spontaneous symmetry breaking, the photon and gluons remain massless, but the $W^{\pm}$ and Z^0 bosons acquire the following masses:

- $m_{Z^0} = \frac{vg}{2\cos\theta_W}$
- $m_{W^\pm} = \frac{vg}{2}$

The weak angle and the masses of the $W^\pm$ (80.4 GeV/c^2) and Z^0 (91.2 GeV/c^2) bosons have been measured directly, leading to a Higgs vacuum expectation value $v \approx 246$ GeV. A relic of the Higgs field remains in the form of the Higgs boson, with a mass $\sqrt{-2\mu^2}$. The Higgs boson has not been directly observed, but is the subject of an intense search; the LEP II experiments have managed to rule out the possibility of a SM Higgs boson lighter than 114.4 GeV (at the 95% confidence level) [8]. Finally, the mechanism naturally gives mass to the fermions by the addition of terms of the form: $\Delta\mathcal{L}_f = -k_f \bar{f}_L \phi f_R$, where k_f is the Yukawa coupling of the fermion f, and f_L (f_R) is the left-handed (resp. right-handed) component of the field f. After symmetry breaking, the electroweak Lagrangian for the fermions is:

$$\begin{aligned}\mathcal{L}_F = &\sum_i \bar{\psi}_i(i\partial - m_i - \frac{gm_iH}{2M_W})\psi_i \\ &- \frac{g}{2\sqrt{2}}\sum_i \bar{\psi}_i\gamma^\mu(1-\gamma^5)(T^+W_\mu^+ + T^-W_\mu^-)\psi_i \\ &- e\sum_i q_i\bar{\psi}_i\gamma^\mu\psi_i A_\mu \\ &- \frac{g}{2\cos\theta_W}\sum_i \bar{\psi}_i\gamma^\mu(g_V^i - g_A^i\gamma^5)\psi_i Z_\mu\end{aligned} \tag{2.7}$$

where ψ_i is the fermion Dirac spinor, m_i is the fermion mass, H is the Higgs boson field, M_W is the W boson mass, g is the weak coupling constant, e is the electron charge magnitude, and θ_W is the weak angle defined earlier. Of particular importance is the coupling of the Higgs boson with fermions, which is proportional to the fermion mass. As a result, the top quark couples to the Higgs sector much more than any other fermion. The second term corresponds to the charged current interaction. T^+ and T^- are the weak isospin raising and lowering operator, respectively; the $(1-\gamma^5)$ projects the field onto its left-handed component. The third term is the electromagnetic interaction, coupling electrically charged fermions (charge q_i) with photons (field A_μ). The last term corresponds to the neutral current interaction (field Z_μ), which can be divided into two components: the axial coupling ($g_A^i = T_i^3$) affects both right- and left-handed particles, while the vector coupling ($g_V^i = T_i^3 - 2q_i sin^2\theta_W$) affects only left-handed particles.

2.1.4 The Parameters of the SM

The SM has 19 free parameters. One particular parametrization is the following:

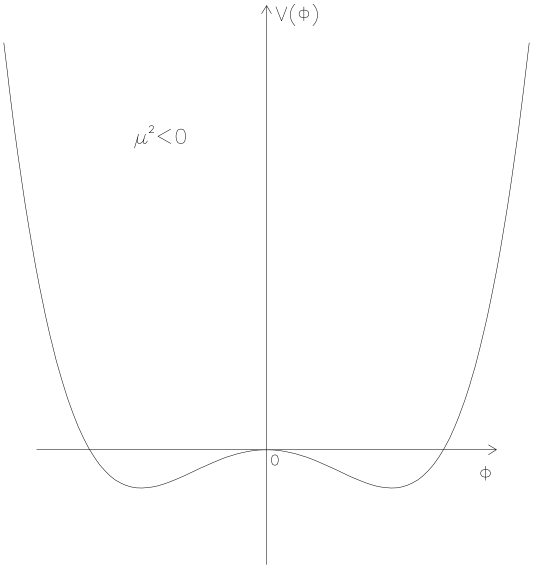


Figure 2.1: *Higgs potential, projected onto the real axis (the potential is independent of the complex phase of the field).*

- The three coupling constants of the electromagnetic, strong, and weak forces: the electromagnetic coupling $\alpha = \frac{e^2}{4\pi}$, the strong coupling $\alpha_S = \frac{g_s^2}{4\pi}$, and the Fermi constant $G_F = -\frac{\lambda}{\sqrt{2}\mu^2}$
- Three lepton and six quark masses
- The mass of the Z^0 boson
- The four parameters of the CKM matrix
- The mass of the Higgs boson
- The coefficient of a possible CP-violating interaction among gluons in QCD

Neutrino masses and mixing require additional parameters.

2.2 Physics Beyond the Standard Model

In spite of all its successes, the SM is incomplete, and many mysteries remain. The mass of neutrinos calls for an extension of the SM. Another caveat of the SM demands to be solved: the Higgs mass is subject to divergent quadratic radiative corrections that need to be somehow controlled in order for the Higgs mass to remain at the electroweak scale. If the SM is to remain valid up to the Plank scale without extension, such a cancellation requires some fine-tuned cancellation: this is referred to as the hierarchy problem. Several ways of solving the hierarchy problem have been explored.

For example, new strong dynamics could appear around 1 TeV (technicolor theories). Another possibility is that the radiative corrections are canceled by a new spectrum of particles at the electroweak scale: supersymmetric (SUSY) theories propose that to every SM particle corresponds a supersymmetric partner with different spin, so that radiative correction contributions to the Higgs mass from a particle is canceled by the contribution from its supersymmetric partner. To SM fermions correspond bosonic superpartners (squarks $\tilde{q}$ and sleptons $\tilde{l}$), and to SM gauge bosons correspond fermionic superpartners (gluinos $\tilde{g}$ and gauginos $\tilde{\chi}^0$, $\tilde{\chi}^{\pm}$). SUSY requires additional Higgs fields in order to provide mass to both up and down families. In the minimal supersymmetric extension of the SM (MSSM), there are five Higgs bosons: h, H, A, and $H^{\pm}$. Furthermore, the Standard Model is unable to describe gravity, and cannot explain the existence of dark matter and dark energy without some extension.

2.3 Top Quark Properties

2.3.1 Standard Model Top Quark

Since its discovery in 1995, the top quark has been the subject of intensive studies by the CDF and D0 experiments. However, many of its properties, although well predicted by the SM, have not been measured — if at all — with a good level of precision. Physics beyond the SM is likely to affect the $t\bar{t}$ signal, justifying the need for top precision measurements.

Mass

The top quark surprisingly large mass was measured at the 3% level. The current top mass measurement is $178.0 \pm 4.3\,\mathrm{GeV/c^2}$ [3]. Most of the properties of the top quark are a direct result of its large mass. Because of its large mass, the top quark causes large radiative corrections that can be observed in electroweak precision measurements. Figure 2.2 shows a good agreement between the SM prediction of the $W^{\pm}$ and top masses, based on LEP I, SLD and neutral current experiments precision measurements, and the direct measurements from LEP II and Tevatron Run I. Since the Higgs boson also contributes to the radiative corrections, a precise knowledge of the $W^{\pm}$ and top masses provides information about the Higgs mass. Figure 2.2 shows that, currently, the electroweak data favor a light Higgs ($m_H^0 \approx 110\,\mathrm{GeV/c^2}$).

Decay Width

The SM predicts the top width Γ_t to be of the order of 1.5 GeV; this is of course very dependent on the mass. This large width ($\Gamma_t \gg \Lambda_{QCD}$) causes the top quark to decay before hadronizing, thus allowing its observation in a free state. In particular, this feature enables precision mass measurement, which is impossible for the other quarks due to non-perturbative effects in the hadronic bound state. The top width

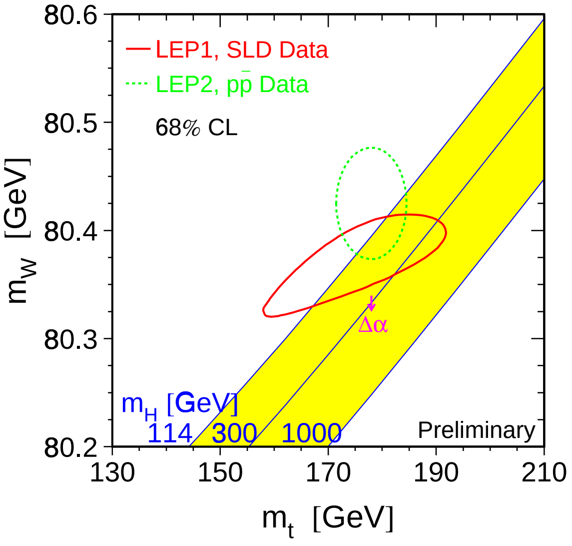


Figure 2.2: *Dotted contour: direct top and* $W^{\pm}$ *mass measurements from the Tevatron and LEP II. Solid contour: predicted top and* $W^{\pm}$ *masses from other electroweak measurements from LEP I, SLD, and neutral current experiments. The straight lines show the SM expected* $W^{\pm}$*-top mass dependence for three different Higgs masses[7].*

has not been measured yet; this measurement is challenging, and requires a large statistical sample that the Tevatron will not be able to produce. The LHC, with a much higher center-of-mass energy than the Tevatron (14 TeV instead of 1.96 TeV) will collect a much larger sample of top events, allowing such a measurement.

Decay Branching Ratios

The SM predicts that the top quark decays almost exclusively to a bottom quark and a W boson ($t \to W^+ b$ and $\bar{t} \to W^- \bar{b}$). Although highly suppressed, decays to charm and down quarks are also allowed. Assuming unitarity of the CKM matrix, $|V_{tb}|$ is predicted to be between 0.9990 and 0.9992 (90% confidence level). The direct measurement, however, is not very accurate: with the assumption of the SM three generations, a 95% confidence level limit of $|V_{tb}| > 0.76$ is found.

Assuming an exclusive decay to Wb, $t\bar{t}$ events produce two W's and two b quarks. The decay channels can then be categorized as a function of the W decays. W bosons decay to either of the three generations of leptons $W^+ \to e^+ \nu_e$, $W^+ \to \mu^+ \nu_\mu$, $W^+ \to \tau^+ \nu_\tau$, or to the two lightest generations of quarks $W^+ \to u\bar{d}$, $W^+ \to c\bar{s}$ ($W^+ \to t\bar{b}$ is not kinematically allowed since $m_t + m_b > m_W$). Because quarks come in three colors, each hadronic branching ratio is three times as large as each

	$W \to e\nu_e$	$W \to \mu\nu_\mu$	$W \to \tau\nu_\tau$	$W \to qq'$
$W \to qq'$	e+jets (12/81)	μ+jets (12/81)	τ+jets (12/81)	all hadronic (36/81)
$W \to \tau\nu_\tau$	$e\tau$ (2/81)	$\mu\tau$ (2/81)	$\tau\tau$ (1/81)	-
$W \to \mu\nu_\mu$	$e\mu$ (2/81)	$\mu\mu$ (1/81)	-	-
$W \to e\nu_e$	ee (1/81)	-	-	-

Table 2.2: $t\bar{t}$ *decay modes and branching ratios (the particle charge has been omitted for simplicity).*

leptonic one[2]: BR$(W^+ \to q\bar{q}')= 3/9$ and BR$(W^+ \to l^+\nu_l) = 1/9$. It is customary to combine the hadronic decays since they are not easily distinguishable experimentally. Table 2.2 shows the $t\bar{t}$ decay modes and their branching ratios. Experimentally, one distinguishes the following channels:

- "Hadronic" channel (44%): both W's decay hadronically.
- Tau channels (21%): since taus are difficult to identify, these channels are challenging and are the subject of specific analyses, while the electron and muon channels are very similar and usually treated in parallel.
- "Di-lepton" channel (5%): both W's decay leptonically to either an electron or a muon.
- "Lepton+jets" channel (30%): one W decays leptonically to an electron or a muon, the other one hadronically. This is the decay channel used for the analysis presented in this document. The final state for this channel consists of a single electron or muon, a neutrino, and four jets, including two b jets produced by the bottom quarks from the top decays.

2.3.2 The Top Quark Beyond the Standard Model

The top quark mass is of the order of the electroweak scale, and is much larger than any other fermion in the SM. This is an indication that the top quark may play a special role in the SM and in the mechanism of electroweak symmetry breaking. To begin with, the top mass implies that its Yukawa coupling is of the order of 1, thus the top quark couples strongly to the Higgs sector. More generally, many theories have been developed, in which the top quark behaves differently from the other quarks. These models may in general result in significant modifications to the top quark production cross section. A non-exhaustive review of such exotic top models follows.

---

[2] Neglecting the small effect of the difference of masses between fermions.

Exotic decays

As discussed in the previous section, the top decay branching ratios have not been measured with good precision so far. Hence it is not excluded that the top quark decays to some exotic (non-SM) particle. In particular, many models, such as supersymmetry, predict the existence of a charged Higgs boson H^+, which could be lighter than the top quark. The current experimental limit for a charged Higgs in the framework of the Minimal Supersymmetric Standard Model (MSSM) is $m_{H^+} >$ 79.3 GeV at the 95% confidence level [4]. This would allow the decay $t \rightarrow H^+ b$, where H^+ preferentially decays to $\tau\nu_\tau$. Compared to the SM $t \rightarrow W^+ b$, this would enhance the presence of τ's in the final state.

Another possibility is that the top squark ($\tilde{t}$, the supersymmetric partner of the top quark) contributes to $t\bar{t}$ production: if the top quark is heavier than the top squark, the decay $t \rightarrow \tilde{t}\chi^0$, where χ^0 is a neutralino, is allowed. The top squark subsequently decays either semileptonically ($\tilde{t} \rightarrow bl\tilde{\nu}$) or hadronically, together with flavor changing ($\tilde{t} \rightarrow c\chi^0$), hence modifying the lepton content of the final state. Alternatively, if the top squark is heavier than the top quark, it may be pair produced and decay into a top quark and a neutralino. In many theories, the lightest neutralino is stable, hence would escape the detector undetected. Thus such exotic decays would result in a larger unbalance in the event momentum. Such a contribution may significantly increase the apparent $t\bar{t}$ cross section.

Extended interactions

Several models propose that the third generation of quarks (and leptons in some cases) could be subject to a new interaction. For example, the strong interaction can be extended so that it treats differently the third generation at high energy, but still treats all generations equally at low energy:

$$SU(3)_{3^{\rm rd}} \otimes SU(3)_{1^{\rm st},2^{\rm nd}} \rightarrow SU(3)_{\rm QCD}$$

This implies that the symmetry of $SU(3)_{3^{rd}}$ and $SU(3)_{1^{st},2^{nd}}$ must be broken below a certain energy, producing a color octet of heavy gauge bosons that couple preferentially to the third generation. Phenomenologically, such gauge bosons could be observed as new resonances in the $t\bar{t}$ and $b\bar{b}$ invariant mass spectra.

2.4 $t\bar{t}$ Production at Hadron Colliders

The cross section (σ) is a Lorentz-invariant measure of the probability of interactions in a two-particle initial state [10]. It relates the interaction rate dN/dt in a volume $d^3\vec{r}$ to the density and velocity of the initial state particles:

$$\frac{dN}{dt} = \rho_1(\vec{r},t)\rho_2(\vec{r},t)\sqrt{|\vec{u_1} - \vec{u_2}|^2 - |\vec{u_1} \times \vec{u_2}|^2}\,\sigma d^3\vec{r} \tag{2.8}$$

where ρ_1 and ρ_2 are the density of the two particle species and $\vec{u_1}$ and $\vec{u_1}$ are their velocity. σ has the dimension of surface area and is usually expressed in barns (1 barn = 10^{-24} cm). Integrating over the interaction region and averaging over time, one defines the instantaneous luminosity:

$$L = \int \rho_1(\vec{r},t)\rho_2(\vec{r},t)\sqrt{|\vec{u_1}-\vec{u_2}|^2 - |\vec{u_1}\times\vec{u_2}|^2}d^3\vec{r} \tag{2.9}$$

The instantaneous luminosity is representative of the intensity of colliding beam machines and is usually expressed in $\text{cm}^{-2}\text{s}^{-1}$. The average event rate is then given by $\frac{dN}{dt} = L \cdot \sigma$. Integrating over time, one can define the (time) integrated luminosity $\mathcal{L} = \int Ldt$. The average number of interactions over a certain period of time is:

$$N = \mathcal{L} \cdot \sigma \tag{2.10}$$

Hadron colliders collide protons into protons (LHC), or protons into antiprotons (Tevatron). Because protons and antiprotons are not elementary particles, but are made of quarks and gluons, the initial state of the actual hard scatter interaction is complicated, and is subject to non-perturbative effects. Thanks to asymptotic freedom, however, if the momentum of the initial particles is high enough ($\gg \Lambda_{QCD} \approx 200$ MeV), it is possible to make the approximation that the interaction occurs between two and only two elementary particles (quark or gluon), one in each incoming particle, and to neglect interactions with the other particles in the proton or antiproton; the nearly free particles are called partons. Unfortunately, the initial momentum of the two partons is not accessible on a event by event basis. The parton carries a fraction (x) of the proton momentum according to a statistical distribution that depends on its type (gluon, or quark flavor) and on the energy scale: such a distribution is called Parton Distribution Function (PDF). Fig 2.3 shows the PDF's for the proton at a scattering energy scale $\mu^2 = 10\,\text{GeV}^2$. The valence quarks (u_v and d_v) are the most likely to carry a large fraction of the momentum, while gluons tend to carry a smaller fraction. To theoretically evaluate a cross section, one must sum over all possible interactions, weighting for their probability according to the PDF's. For example for a proton-antiproton collision:

$$\sigma = \sum_{i,j} dx_i dx_j F_p^i(x_i,\mu^2) F_{\bar{p}}^j(x_j,\mu^2)\sigma(p_i,p_j \to \text{Final State}) \tag{2.11}$$

where x_i (resp. x_j) is the fraction of momentum carried by a parton of type i in the proton (resp. of type j in the antiproton), μ the energy scale of the interaction, and F_p^i (resp. $F_{\bar{p}}^j$) the PDF of a parton of type i in the proton (resp. the PDF of a parton of type j in the antiproton).

At the Tevatron Run II, protons and antiprotons collide at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV. Top-antitop quark pairs can be produced through several mechanisms. Figure 2.4 shows the Feynman diagram of the quark-antiquark annihilation process at tree level; quark-antiquark annihilation is expected to contribute 85% of

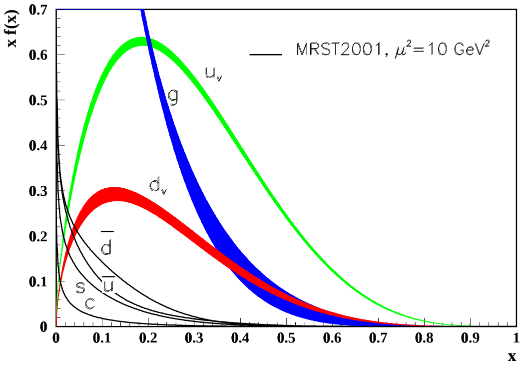


Figure 2.3: *Parton distribution functions of quarks and gluons in the proton at a momentum transfer* $\mu^2 = 10\text{GeV}^2$ *[4].*

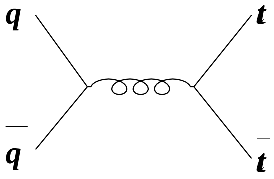


Figure 2.4: *Feynman diagram of the leading order quark-antiquark annihilation top quark pair production process.*

the $t\bar{t}$ production. Gluon fusion processes (Figure 2.5 shows the tree level gluon fusion processes) are responsible for the remaining 15%. The theoretical SM prediction for $t\bar{t}$ pair production is $6.7^{+0.7}_{-0.9}$ pb for a top quark mass of 175 GeV/c^2 [9].

It is interesting to notice that at the LHC, gluon fusion is the dominant $t\bar{t}$ production process (about 90%), for two reasons. Firstly, because of the much larger center-of-mass energy (14 TeV), $t\bar{t}$ pairs can be produced with a much smaller x, for which the gluon PDF is very large, thus the gluon fusion process is enhanced compared to the Tevatron. Secondly, in proton-proton collisions, the probability for colliding a quark and its antiquark is small (because the antiquark cannot be a valence quark), thus the quark-antiquark annihilation process is suppressed.

2.5 Analysis Overview and Motivation

2.5.1 Method

This analysis measures the inclusive production cross section of $t\bar{t}$ production, assuming an SM top quark (in particular, SM branching ratios) with a mass $m_t \approx$ 175 GeV. Because this measurement is sensitive to the top mass itself, the result will be given as a function of m_t. The method used to measure the cross section is based

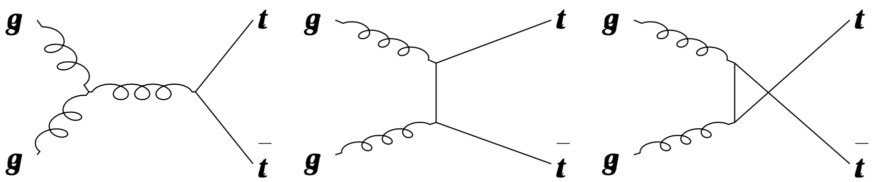


Figure 2.5: *Feynman diagrams of the leading order gluon fusion top quark pair production processes.*

on a counting analysis. From Equ. 2.10, the cross section is given by:

$$\sigma_{t\bar{t}} = \frac{N_{\text{produced}\,\mathrm{t\bar{t}}}}{\mathcal{L}} \tag{2.12}$$

where $N_{\text{produced}\,\mathrm{t\bar{t}}}$ is the total number of $t\bar{t}$ events produced (independently of the decay mode) over a period corresponding to an integrated luminosity $\mathcal{L}$. The integrated luminosity measurement is described in 3.2.7. $\mathrm{N}_{\text{produced}\,\mathrm{t\bar{t}}}$ is evaluated as follows:

- A sample of $N_{\text{observed events}}$ events, rich in $t\bar{t}$ lepton+jets events, is isolated.
- Only a small fraction of the $t\bar{t}$ events produced are actually observed, either because the event selection is not appropriate for their decay mode, or because of limitations in the instrumentation and reconstruction. The acceptance is the fraction of $t\bar{t}$ events that are actually observed and the cross section can be expressed:

  $$\sigma_{t\bar{t}} = \frac{N_{\text{observed}\,\mathrm{t\bar{t}}}}{\mathrm{A}_{t\bar{t}} \cdot \mathcal{L}} \tag{2.13}$$

- Even though the sample is expected to contain mostly $t\bar{t}$ events, a non-negligible fraction of events are due to various background processes. Each background contribution is estimated separately, and the number of background events $N_{\text{background events}}$ that are expected to contribute to the sample is subtracted from the number of observed events $N_{\text{observed events}}$:

$$\sigma_{t\bar{t}} = \frac{N_{\text{observed events}} - N_{\text{background events}}}{\mathrm{A}_{t\bar{t}} \cdot \mathcal{L}} \tag{2.14}$$

Each step of the analysis will be described in the subsequent chapters.

2.5.2 Motivation

Measuring the $t\bar{t}$ production cross section is the most basic test of the QCD properties of the top quark. In the SM, QCD treats all quarks identically (except for their mass), while several theories beyond the SM treat the top quark differently, which could result in a deviation from the SM. Furthermore, non-SM processes, such as SUSY stop-antistop production, have a signature very similar to $t\bar{t}$ events: even though the final state is different, this could contribute to, and enhance, the measured cross section.

Besides the absolute measurement, deviations from the SM could appear in the comparison of the same measurement in different decay channels. Indeed, exotic top decays could enhance the measured cross section in one channel, and suppress it in another. For example, a top quark decaying to a charged Higgs would enhance the di-lepton cross section measurement compared to the lepton+jets one, since a charged Higgs decays predominantly to a τ, which in turns decays leptonically with a large branching ratio.

On a more technical note, measuring the cross section paves the way to other analyses by clearly defining the top sample and understanding its composition. Measuring the $t\bar{t}$ production cross section is a prerequisite to top mass measurements. Besides, Higgs and new physics processes may contaminate the $t\bar{t}$ sample: this could appear in anomalies in the heavy flavor content or in the kinematic properties of the $t\bar{t}$ candidates.

Chapter 3

Experimental Apparatus

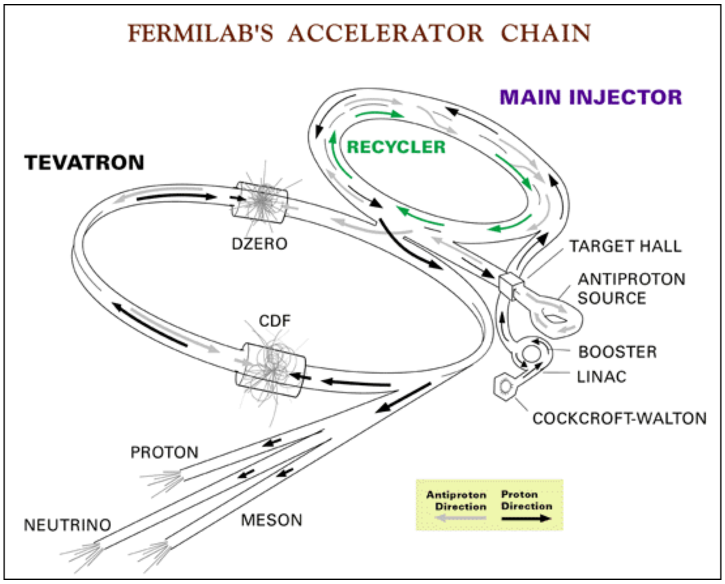


Figure 3.1: *The chain of accelerators at Fermilab.*

The analysis presented in this thesis uses data collected between February 2002 and September 2003 from proton-antiproton collisions produced by the Tevatron at a center-of-mass energy of 1.96 TeV, and observed by the Collider Detector at Fermilab (CDF). This chapter describes the collider apparatus and the detector.

3.1 The Tevatron and the Fermilab Accelerator System

The Tevatron is a proton-antiproton storage ring complex located at the Fermi National Laboratory (Fermilab), in Batavia, Illinois. With a center-of-mass energy of $\sqrt{s} = 1.96$ TeV, it is currently the most powerful collider, and the only apparatus capable of producing top quarks. The 2 km diameter storage ring is the last step of a complex chain of accelerators that produce and accelerate the proton and antiproton beams (see Fig. 3.1). Each step of this process is described in the subsequent sections. Major upgrades have been made to the Fermilab accelerator system in order to increase the luminosity for the Tevatron Run II. The center-of-mass energy was also increased from 1.8 to 1.96 TeV. Two technical challenges limit the increased luminosity: the antiproton production, and the control of beams of very high intensity. To cope with these challenges, the Main Injector (MI) and the Recycler were built.

3.1.1 Proton Injector

The proton beam is initiated with di-hydrogen molecules (H_2) that are split by an intense local electrostatic field; the resulting H^- anions are accelerated by a 750 kV Cockroft-Walton machine. The continuous beam of H^- ions is segmented into bunches, and transfered into a 150 m long linear accelerator (LINAC) and brought to an energy of 400 MeV. Upon exiting the LINAC, the anions pass through a thin foil of graphite that tears apart the pair of electrons of each ion, leaving only the proton and thus converting the anion beam into a proton beam. The beam is transfered into a 475 m long circular synchrotron (the Booster) that accelerates the protons to 8 GeV, which are in turn transfered into the MI. The MI is used both for accelerating and injecting the proton beam into the Tevatron, and for producing an intense proton beam used to produce antiprotons (see next section). Before injection into the Tevatron, protons are accelerated up to 150 GeV within a few seconds.

3.1.2 Antiproton Production and Recycling

Antiprotons ($\bar{p}$) are produced by colliding a 120 GeV proton beam from the MI onto a Nickel target. Roughly, one antiproton is produced for every million proton collisions. Hence the interaction products need to be efficiently filtered in order to reject non-$\bar{p}$ particles. In a first step, charged particles are focused by a "lithium lens", a tubular piece of lithium traversed by a 650 kA current. Then a pulsed magnet vetoes positively charged particles and selects negative particles with a momentum of the order of 8 GeV. After the non-$\bar{p}$ particles decay in flight, only $\bar{p}$ remain. The resulting $\bar{p}$ beam is de-bunched and focused further through betatron stochastic cooling and momentum cooling. Antiprotons are stored into the accumulator, a set of concentric synchrotrons 80 m long in circumference. Accumulation rates of the order of 10^{11} antiprotons per hour have been reached. Once enough antiprotons are accumulated (typically $150 \cdot 10^{10}$), they are transfered into the MI in the direction opposite to the protons, accelerated up to 150 GeV together with the proton beam, and transfered into the Tevatron.

Because $\bar{p}$ production is one of the limiting factors to increasing the luminosity, the recycler was designed to recoup the antiprotons that remain in the Tevatron at the end of a store, and re-inject them in the Tevatron for the next store, together with a new stack of antiprotons from the accumulator. In the current operations, however, the recycler is used as an extension of the accumulator: the antiprotons are transfered from the accumulator into the recycler, where the beam is cooled down before injection into the Tevatron.

3.1.3 The Tevatron storage ring

Unlike the other accelerators in the chain, the Tevatron is equipped with superconducting magnets. The beams in the Tevatron have a 36x36 bunch structure, with a 396 ns time spacing between bunches. It should be noted that some bunches are kept empty (the so-called "abort gap") in order to be able to dump the beam. At the beginning of a store, the Tevatron beams contain about 10^{13} protons and 10^{12} antiprotons.

The luminosity is given by the following formula:

$$L = \frac{N_p N_{\bar{p}} B f}{4\pi\sigma_x\sigma_y}$$

where N_p ($N_{\bar{p}}$) is the number of protons (antiprotons) per bunch, B is the number of bunches, f is the revolution frequency, and σ_x and σ_y are the beam dimensions in the plane transverse to the beam at the interaction point. However, luminosity is not determined from this formula, but from the measured rate of some reference physical processes. The measurement of the luminosity delivered by the Tevatron to the CDF experiment is described in Sec. 3.2.7.

Peak luminosities above $10^{32}/\mathrm{cm}^2/\mathrm{s}$ have been reached. The luminosity decays exponentially during a store, due to beam losses and collisions, with a semi-life-time of the order of 6 hours, so that stores can last 12 hours or more before the Tevatron is emptied and filled again with the next store.

Beams are brought into collision in the center of the CDF and D0 detectors, where the resulting events are recorded.

3.2 The Collider Detector at Fermilab (CDF)

CDF is a multipurpose high-energy particle detector with approximate cylindrical and forward-backward mirror symmetry with respect to the beam axis. The CDF collaboration gathers more than 700 physicists and 49 institutions from from 12 different countries. In normal operations, several dozens of physicists and technicians take shifts to run, monitor, and maintain the very complex detector. It is composed of several coaxial sub-detectors which serve different purposes, and which will be described in the following sections: the Silicon Vertex Detector and the Central Outer Tracker, which lie inside a 1.41 T superconducting solenoid magnet, form the tracking system, and precisely measure trajectories and momenta of charged particles as they go through the detector; the calorimeters lie outside the solenoid and measure particle energies; the muon system lies outside the calorimeter and detects muons that escape the calorimeters; the Cerenkov Luminosity Counter (CLC) measures the rates of interactions near the beam used for luminosity measurement. A Time-Of-Flight detector (TOF) located between the COT and the solenoid, and sub-detectors used to study diffractive physics, are not used in this thesis; their description has been omitted. Figure 3.2 shows a section of the CDF detector; Figure 3.3 shows details of the tracking system.

Many components of the detector have been upgraded for the Tevatron Run II: the Technical Design Report [11] contains a lot of useful information about these upgrades.

3.2.1 The CDF coordinate system

The CDF experiment uses a standard right-handed coordinate system:

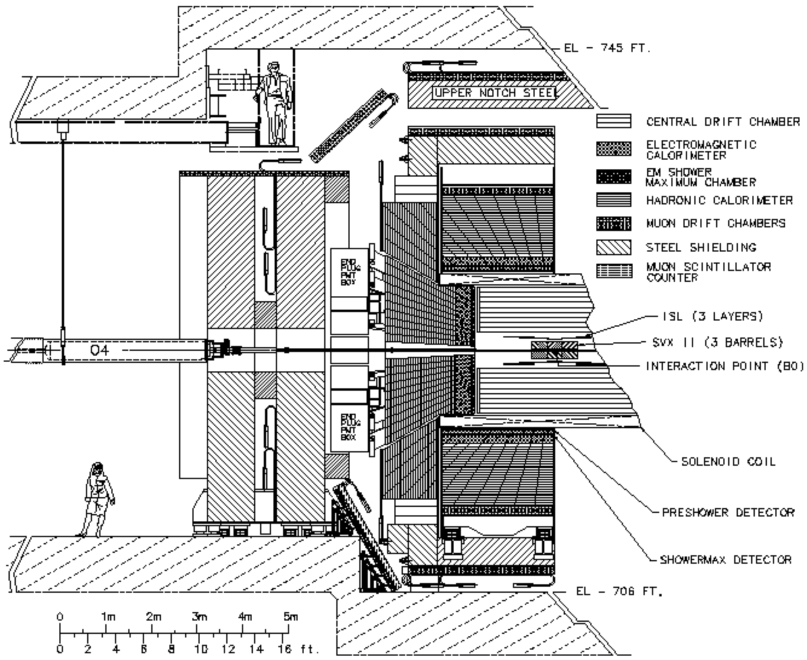


Figure 3.2: *Section of the CDF upgrade.*

- the z axis is defined by the proton beam (toward East),
- the y axis is vertically upward,
- the x axis is horizontal, pointing North (outward w.r.t the center of the Tevatron).

Spherical coordinates are also commonly used:

- the polar angle θ is defined w.r.t to the proton beam,
- the azimuthal angle ϕ is defined w.r.t to the x axis.

A commonly used variable related to the polar angle is the pseudorapidity $\eta = -\ln\tan\frac{\theta}{2}$. In the limit of massless particles, the pseudorapidity is equivalent to the rapidity $y = \ln\frac{E+p_z}{E-p_z}$ (where E is the energy and p_z is the z component of the momentum of the particle). The variable r refers to the distance to the beam. The (x,y) plane is named the transverse plane. Projected quantities onto the transverse plane are referred to as "transverse", and noted with a "T" subscript (such as transverse momentum, p_T).

3.2.2 The Solenoid

A superconducting solenoid coil creates a nearly uniform magnetic field in the tracking region ($|z| < 1.5$ m, $r < 1.4$ m). The coil itself is 4.8 m long and ≈ 25 cm

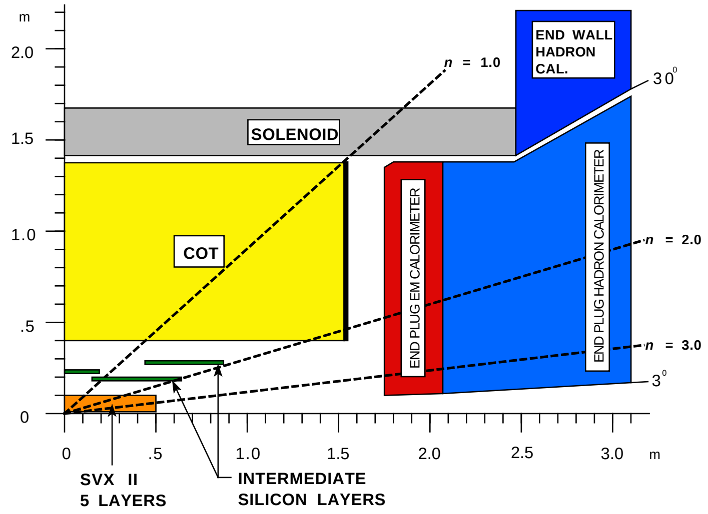


Figure 3.3: *Section of the CDF tracking system (outward: SVX, ISL, and COT). L00 is not shown.*

thick, with an inner radius r = 1.4 m. A current of 4650 A goes through the 1164 turns of aluminum-stabilized Nb Ti to produce a magnetic field of 1.41 T oriented in the $-z$ direction. The coil amounts to 0.85 radiation length at normal incidence.

3.2.3 The Silicon Vertex Detector

The role of a vertex detector is to provide precise tracking information near the interaction point in order to identify displaced tracks, and reconstruct secondary vertices, produced by long life-time particle decays (see Chap. 5). Hence it is built as close to the interaction point as possible, and must be resistant to high doses of radiation. Another requirement for this type of detector is to be able to function in a high track density environment. Semi-conductor detectors satisfy such requirements.

The CDF II vertex detector [12] is divided into three co-axial silicon strip devices. The L00 is a single-sided layer mounted directly on the beam pipe, in order to be as close as possible to the interaction point ($r \approx 1.3$ cm). The SVX II is composed of five 90 cm long double-sided layers ($2.55 < r < 10.6$ cm). The ISL (Intermediate Silicon Layer) is made of one double-sided layer in the central region ($r \approx 22$ cm) and two in the forward region ($r \approx 20$ and 28 cm), and extends the tracking capability to $|\eta| = 2$, while helping with the track reconstruction in the central region by filling the gap between the COT and the SVX (see Fig. 3.3).

The SVX is composed of three 27 cm long identical barrels with a ≈ 1.5 cm gap between adjacent barrels, hence covering the $|z| < 45$ cm region (see Fig. 3.4). In each barrel, the five double-sided layers are arranged with a 12-fold ϕ symmetry into 30° "wedges" (see Fig. 3.5). Layers are numbered from 0 to 4 (starting with the innermost one). Double-sided sensors provide ϕ measurement with axial strips (*i.e.*

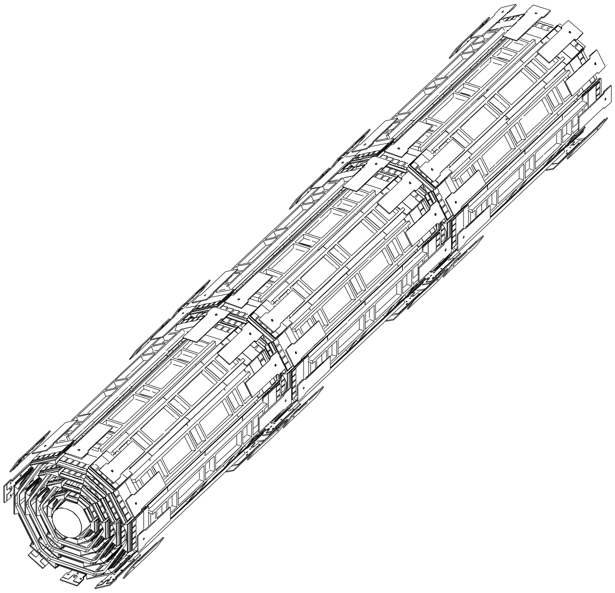


Figure 3.4: *3D view of the SVX.*

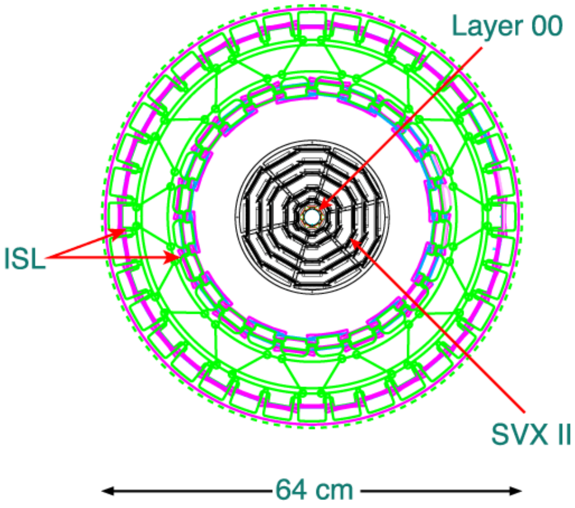


Figure 3.5: *End view of the L00, SVX, and ISL.*

parallel to the z axis, the "ϕ side"), and z measurement with stereo angle strips (the "z side").

90° Hamamatsu sensors on layer 0 (L0), L1, and L4; 1.2° small stereo angle Micron sensors on L2 and L3. 90° sensors use 4″ silicon wafers with double metal technology. They provide a better z-resolution than the small stereo angle sensors, but suffer from larger hit ambiguities for tracking reconstruction. As is visible in Fig. 3.4, power supply and readout devices are connected to the sensors near both edges of each barrel, on each side of the sensors (*i.e.* overlapping with the sensors in the z dimension). Sensor strips are connected to beryllium oxide ceramic electrical

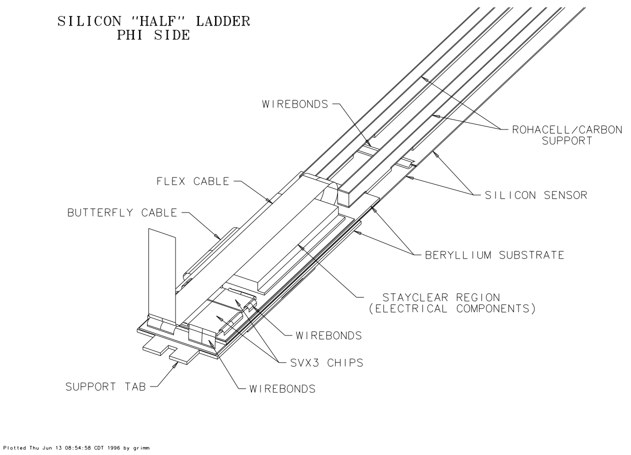


Figure 3.6: *View of a Layer 0 SVX module (ϕ side).*

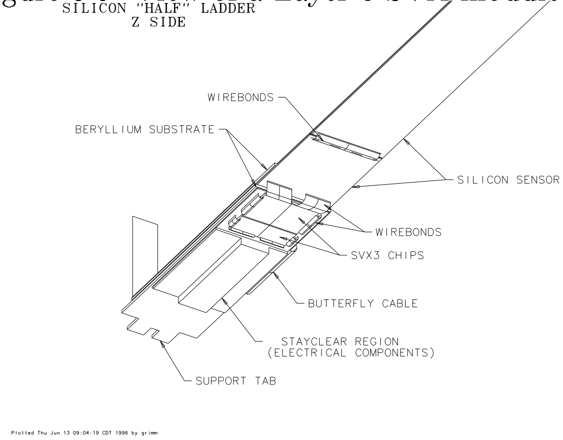


Figure 3.7: *View of a Layer 0 SVX module (z side).*

hybrids via wire-bonds. A hybrid on the ϕ side is connected to the one on the z side via a 100 μm diameter wire (the "jumper") at the edge of the sensor. Figures 3.6 and 3.7 show a view of the ϕ and z sides, respectively, of a L0 module. The hybrids support SVX3D chips, described in the next section.

The L00 is also composed of 12 wedges, where alternate wedges have different width and are located at a different radius, hence presenting a 6-fold ϕ symmetry, as shown in Fig. 3.8. L00 is split in two identical barrels. Sensors are single-sided, with axial-only 25 μm wide strips. Wide (narrow) wedges contain 512 (256) strips, with an overlap of about 60 strips. Only one out of two strips is actually read out. The sensors used for L00 are radiation-hard high-resistivity wafers. L00 was not used for the analysis presented in this document because an unexpected source of noise in the analog part of the readout system requires additional offline treatment, that was not available at the time this analysis was made.

The ISL has a more complex geometry, shown in figures 3.5 and 3.3. The central barrel consists of a single layer of 28 wedges. The two forward barrels are composed

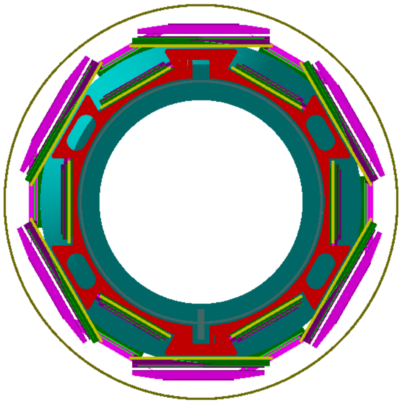


Figure 3.8: *End view of L00.*

of two layers covering roughly the same pseudorapidity range, the outer one (ISL1) being located at larger $|z|$ than the inner one (ISL0). ISL0 (ISL1) is made of 24 (36) wedges.

Some characteristics of the sensors are summarized in Tables 3.1 and 3.2.

The SVX3D Chip and the DAQ system

SVX3D chips are built with 0.8 μm bulk CMOS Honeywell radiation-hard technology, designed to resist up to 4 MRad. Hence radiation damage on the chips is not expected to be a problem during Run II.

Figure 3.9 shows a picture and a diagram of the SVX3D chip. Each SVX3D chip reads out 128 strips. The Front End is a set of 128 integrators and 47-event deep pipelines. Forty-six events can actually be stored in the pipeline, waiting for the Level 1 (L1) trigger decision (see Sec. 3.2.8 for details about the trigger system);

if an event is accepted by the L1 trigger, it is taken out of the loop and waits for the Level 2 (L2) trigger decision. Four events can be waiting for L2 decision at once. The Back End is equipped with 128 8-bit Wilkinson ADC's which allow continuous analog acquisition during digitization and readout. An important feature of the SVX3D chip is its ability to subtract the common mode noise dynamically, on an event-by-event basis (Dynamic Pedestal Subtraction, DPS). The DPS significantly improves the signal over background ratio, and reduces the readout occupancy. The chip is designed to run at a L1 trigger rate of 50 kHz without dead-time. The data acquisition system of the SVX is schematically shown in Fig. 3.10. Chips within a given module form a "daisy chain" and are initialized and read out sequentially. The five layers within a wedge (totaling 44 chips) are connected to a port card (PC) via five copper/Kapton laminate cables (High Density Interconnect, HDI). Each PC is equipped with five (one per layer) Dense Optical Interface Modules (DOIM's) that convert the output of the chips into optical signal, which is sent through optical fibers to Fiber Interface Boards (FIB's) located next to the CDF detector. FIB's gather

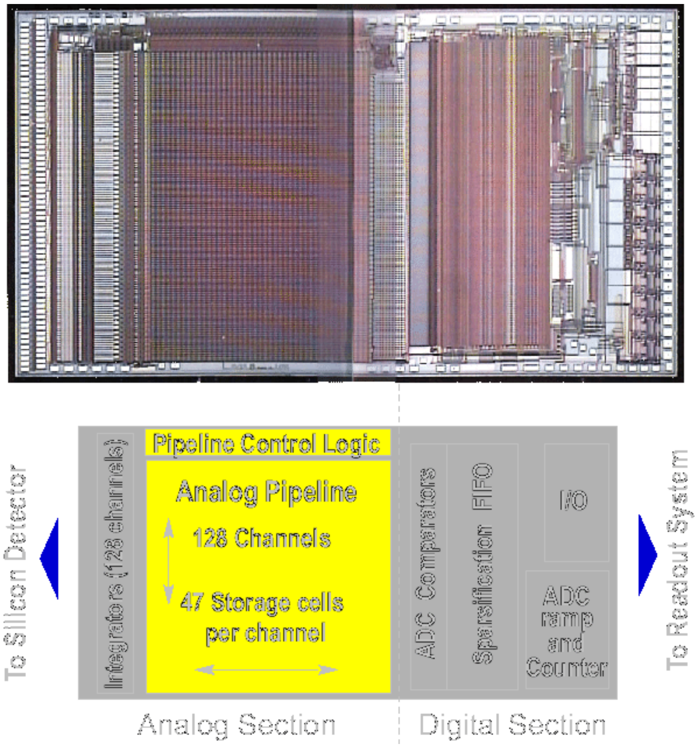


Figure 3.9: *Picture and diagram of the SVX3D chip.*

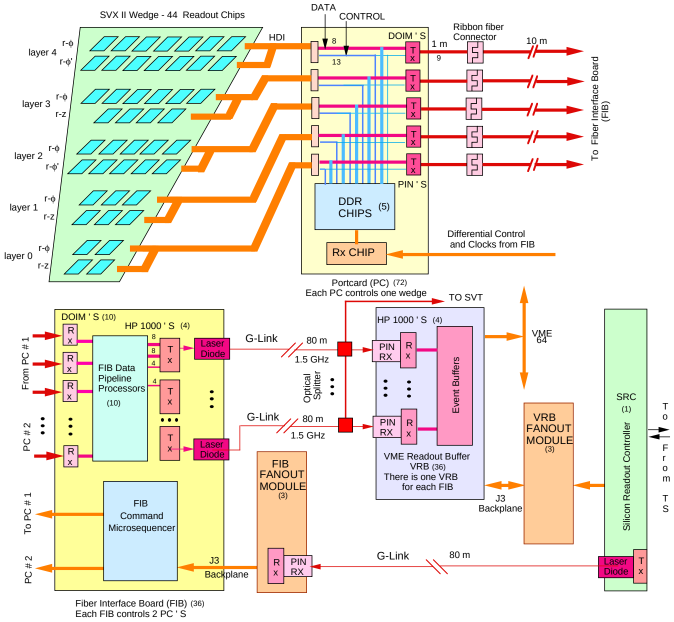


Figure 3.10: *DAQ of the silicon vertex detector.*

	L0	L1	L2	L3	L4
Sensor	Ham 90	Ham 90	Micron 1.2	Ham 90	Micron 1.2
Stereo angle	90^o	90^o	-1.2^o	90^o	$+1.2^o$
Radius (cm)	2.55/3.00	4.12/4.57	6.52/7.02	8.22/8.72	10.1/10.6
# modules	72	72	72	72	72
Thickness (μm)	300	300	300	300	300
# ϕ strips	256	384	640	768	896
# z strips	512	576	640	512	896
# ϕ chips	2	3	5	6	7
# z chips	2	3	5	4	7
ϕ pitch (μm)	60	62	60	60	65
z pitch (μm)	141 (x2)	125.5 (x1.5)	60	141	65
Lifetime (fb^{-1})	4.3	8.5	10.7	23	14

Table 3.1: *Properties of the SVX modules.*

the data from the DOIM's into an 80 m long high speed optical link (G-Link) which carries the data to VME readout buffer cards (VRB's) which hold the data until a L2 decision is made. The silicon readout controller (SRC) supervises the entire SVX, and in particular communicates with the trigger system. All 405,504 channels of the SVX can be read out within 10 μs, which allows the L2 silicon vertex trigger (SVT) to use silicon tracking information for triggering on tracks with large impact parameter. L00 (13824 channels) and ISL (303104 channels) use the same DAQ system as the SVX, although it had to be adapted to their specific geometry.

For a particle emitted at $z = 0$, the entire silicon detector represents several percents of a radiation length at normal incidence, and up to 40% of a radiation length at $\eta = 2$ [13]. Because of the complex geometry of the detector, the number of radiation lengths varies widely depending on the particle direction.

3.2.4 The Central Outer Tracker

The Central Outer Tracker (COT) [11] is a large open-cell drift chamber built to replace the one used in Run I (CTC). The initially planned 132 ns bunch spacing for Run II requires a shorter drift time: the COT was designed in order to achieve a maximal drift time of 100 ns, corresponding to a maximal drift distance of 0.88 cm. The COT extends to $|z| < 150$ cm and radii between 40 and 137 cm from the beam, covering the pseudorapidity range $|\eta| < 1$.

Figure 3.11 shows the layout of open cells. Each open cell contains 12 sense wires and is tilted by 35° w.r.t the radial direction. Figure 3.12 shows equipotential lines in a cell. Cells are gathered into eight so-called "superlayers", as shown in Fig. 3.13. There are four radial and four stereo superlayers; wires in radial layers are parallel to the z axis, while in stereo layers they are tilted by $+3°$ and $-3°$ (alternatively) w.r.t

	L00	ISL
Sensor	Ham/SGS	Micron 1.2
Stereo angle	-	$+1.2^{o}$
Radius (cm)	1.35/1.62	20.0/22.0/28.0
# modules	48	296
Thickness (μm)	300	300
# ϕ strips	128/256	1024
# z strips	-	768
# ϕ chips	3/2	4
# z chips	-	3
ϕ pitch (μm)	25 (x2)	55 (x2)
z pitch (μm)	-	73 (x2)
Lifetime (fb^{-1})	7.4	-

Table 3.2: *Properties of L00 and ISL.*

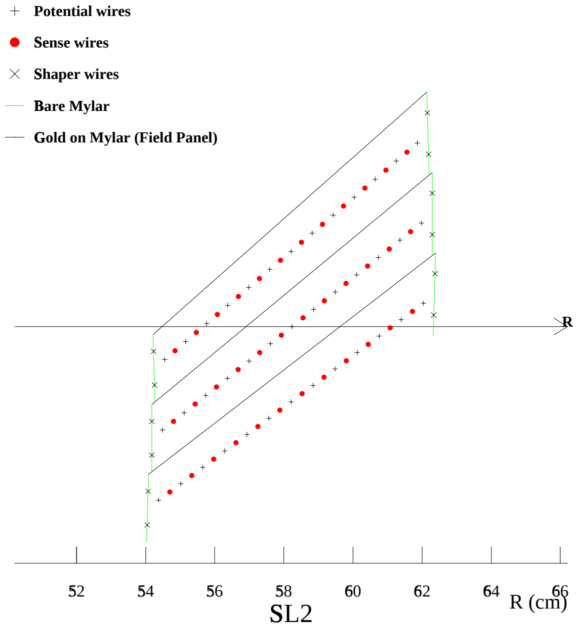


Figure 3.11: *COT cell layout in superlayer 2.*

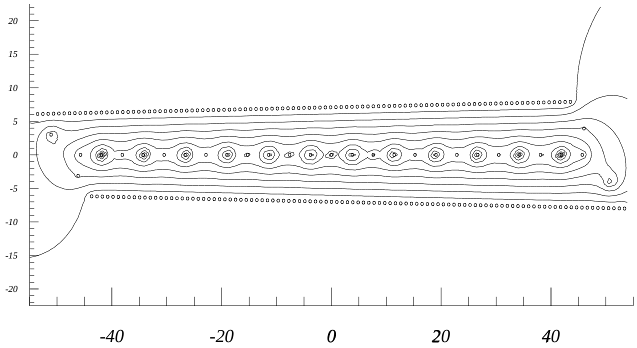


Figure 3.12: *Equipotential lines in a COT cell of superlayer 1.*

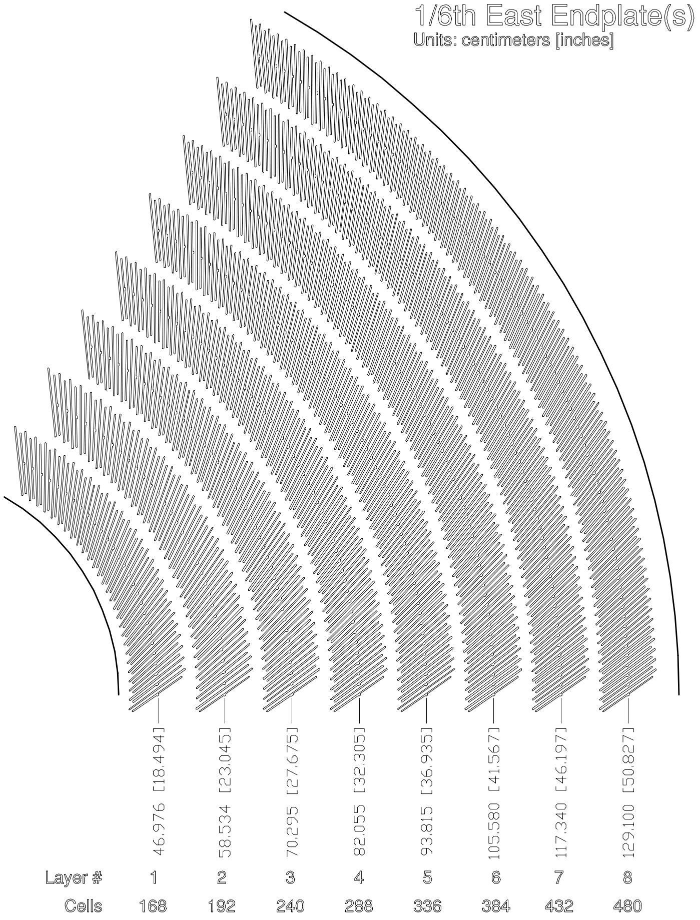


Figure 3.13: *COT quadrant.*

the z axis, in order to provide measurement relative to this axis. The number of cells varies from 168 for the innermost superlayer to 480 for the outermost one. The COT amounts to 1.69% of a radiation length at normal incidence.

The gas mixture chosen is 50:35:15 Argon-Ethane-CF_4. Indeed, adding 15% of CF_4 reduces the aging of the wires. However, in the course of Run II data taking, some premature aging of the innermost superlayers, visible in a large gain loss, was discovered. A significant increase of the gas flux and the addition of a very small fraction of oxygen to the mixture solved the problem .

The performance of the tracking system is described in detail in Sec. 4.1.

3.2.5 The Calorimetry

The CDF calorimetry consists of both electromagnetic calorimeters (large number of radiation lengths X_0, and small number of interaction lengths λ, for photon and electron identification and energy measurement), and hadronic calorimeters (large number of interaction lengths for hadron energy measurement). The central region of the detector is covered by the central electromagnetic (CEM) [14] and central hadronic (CHA) [15] calorimeters, in the pseudorapidity ranges $|\eta| < 1.1$ and $|\eta| < 0.9$, respectively. In the forward region, the plug electromagnetic (PEM) and hadronic (PHA) calorimeters [16] cover the regions $1.1 < |\eta| < 3.6$ and $1.3 < |\eta| < 3.6$, respectively. The WHA [15] is a hadronic calorimeter that fills the gap between the CHA and the PHA in the pseudorapidity range $0.7 < |\eta| < 1.3$.

This calorimetry system provides a nearly 4π solid angle coverage around the interaction point, which is particularly important in order to identify the presence of high-momentum neutrinos by looking for an imbalance in the event transverse energy (see Sec. 4.5).

The CEM calorimeter is a sample device made of 31.5 mm thick layers of polystyrene scintillator, alternated with 1/8 in thick (3.18 mm) layers of aluminum-clad lead. In order to maintain a constant number of radiation lengths as a function of θ ($19X_0$), some lead layers are replaced by acrylic (Plexiglas), so that the actual number of absorber layers varies from 30 near the center to 20 at $\eta \approx 1.1$. The CEM is divided into four arches (North-West, South-West, North-East and South-East), made of identical 15° modules, each of them being segmented into 10 projective towers. Thus each tower covers a solid angle of 0.1 by 15° in the $\eta \times \phi$ space. The blue light emitted by the scintillators is collected on each side of the tower by acrylic wavelength shifters that convert it into green light and guide the light toward two photomultipliers (Hamamatsu R580) outside the CHA (see Fig. 3.14). The two most forward towers of one of the CEM and CHA modules are not instrumented (the so-called "chimney"), in order to provide access for cryogenics to the solenoid. Based on test beam data, the CEM energy resolution for an electron going through the center of a tower is found to be $\frac{13.7\%}{\sqrt{E}} \oplus 2\%$[1].

---

[1] The symbol $\oplus$ means that the errors are to be added in quadrature.

The Central Electron Strip Chambers (CES) are used to determine shower position and transverse shower development. They are located between the 8^{th} lead layer and the 9^{th} scintillator layer (counting outward), which is the expected position of shower maximum ($\approx 6X_0$, including tracking and solenoid material). In each CEM module, a CES module is a multi-wire proportional chamber with 64 anode wires parallel to the beam axis, spaced 0.73 cm apart and split at $|z| = 121$ cm, and 128 cathode strips, with a pitch of 1.67 cm for $|z| < 121$ cm, and 2.01 cm for $|z| < 121$ cm. The spatial resolution achieved is ≈ 2 mm.

The CEM is also equipped with a pre-shower detector (CPR), useful in discriminating between hadrons and photons/electrons. The CPR is a set of multi-wire proportional chambers with wires parallel to the beam providing transverse measurements and strip cathodes providing z information, with a resolution of the order of a few millimeters.

The CHA is a sampling hadronic calorimeter surrounding the CEM, following the same segmentation (0.1 by 15° in $\eta \times \phi$). The WHA extends the CHA coverage and uses the same technology as the CHA. Altogether, a wedge contains 12 towers, 6 of which are fully in the CHA, 3 in the WHA, and 3 are shared between the two. The number of interaction lengths (λ) is constant through the entire range of pseudorapidity and is equal to 4.5. The CHA is made of 32 layers of 2.5 cm thick steel absorber, and 1.0 cm thick scintillator. The WHA is made of 15 layers of 5.0 cm thick steel absorber, and 1.0 cm thick scintillator. 2 PMT's per tower are linked to the scintillators by a wavelength shifter and a light guide. The CHA and WHA single pion energy resolutions are $\frac{50\%}{\sqrt{E}} \oplus 3\%$ and $\frac{75\%}{\sqrt{E}} \oplus 4\%$, respectively.

The plug calorimeter (Fig. 3.15) was designed and built to replace the CDF Run I forward calorimeters, and to cope with the Run II requirements (higher luminosity and 132 ns bunch spacing). It consists of an electromagnetic (PEM) and a hadronic (PHA) calorimeter with the same projection segmentation. Figure 3.16 shows the segmentation pattern of a 15° module: towers cover an azimuthal angle of 7.5° down to $\eta = 2.11$, and 15° further; similarly, the segmentation in η becomes coarser as one moves closer to the beam. Figure 3.16 also shows how towers are combined for the purpose of being used by the trigger system. The PEM is made of 22 layers of 4.5 mm lead and 4 mm thick scintillator tiles. Each scintillator tile is read by a single PMT. In front of the 22 sampling layers is a 1 cm thick scintillator tile read out by a multi-anode photomultiplier (MAPMT) which is used as a pre-shower detector. The PEM, including the pre-shower, amounts to a total of 21 radiation lengths. The PEM energy resolution is $\frac{16\%}{\sqrt{E}} \oplus 1\%$. The plug maximum shower detector (PES) is made of two sets of scintillating strips that provide precise 2D shower position measurement (resolution ≈ 1 mm). The PHA is made of 23 layers of 2 in (5.08 cm) thick steel absorber, and 6 mm thick scintillator, corresponding to 7 interaction lengths. Its resolution is $\frac{80\%}{\sqrt{E}} \oplus 5\%$.

Several regions of the detector have low calorimetry response (called "cracks"): at the junction between East and West module near $\eta = 0$, the "chimney", near the azimuthal boundaries between wedges, and at the junction between central and

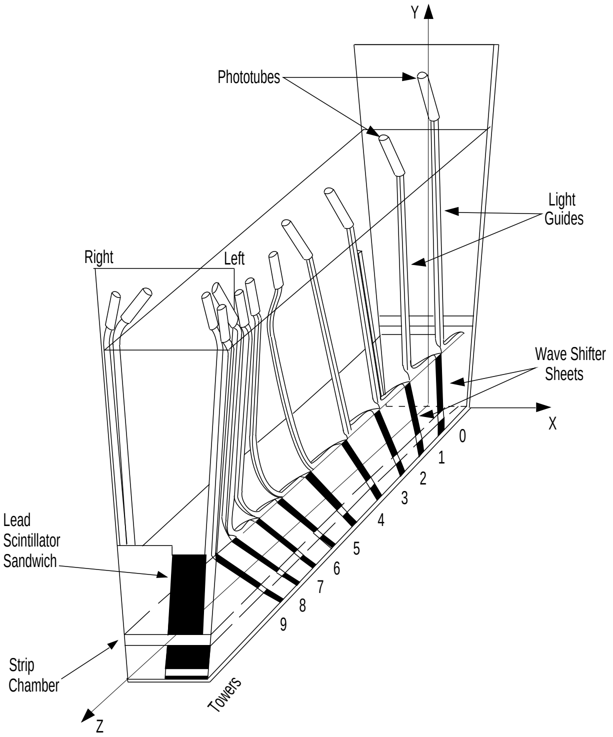


Figure 3.14: *Wedge of the Central Electromagnetic calorimeter (CEM).*

	CEM	CHA	WHA	PEM	PHA
X_0	19	-	-	21	-
λ	1	4.5	4.5	1	7
# layers	21	32	15	22	23
Absorber	3.18 mm lead	2.5 cm steel	5 cm steel	4.5 mm lead	5.08 cm steel
Scint.	5 mm SCSN-38	1 cm PMMA	1 cm PMMA	4 mm PMMA	6 mm PMMA
σ_E	$\frac{13.7\%}{\sqrt{E}} \oplus 2\%$	$\frac{50\%}{\sqrt{E}} \oplus 3\%$	$\frac{75\%}{\sqrt{E}} \oplus 4\%$	$\frac{16\%}{\sqrt{E}} \oplus 1\%$	$\frac{80\%}{\sqrt{E}} \oplus 5\%$

Table 3.3: *Some properties of the CDF calorimeters. The last row shows the single particle energy resolution.*

forward calorimeters at $\eta \approx 1.3$.

Table 3.3 summarizes some properties of the various calorimeters.

3.2.6 The Muon System

The muon system has been significantly upgraded for Run II, in particular to complete the coverage in the central region. Three muon detectors are used for the analysis described in this document: the Central Muon Detector (CMU) [17] and the Central Muon Upgrade (CMP) [18], covering the pseudorapidity range $|\eta| < 0.6$, and

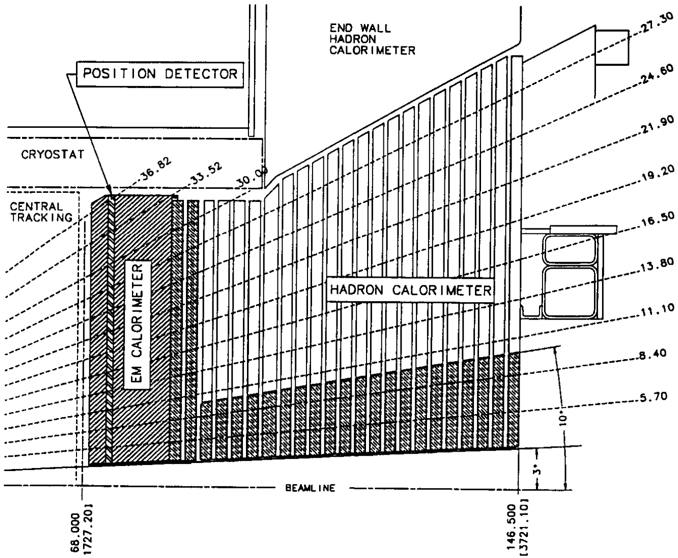


Figure 3.15: *View of the Plug calorimeter (PEM and PHM).*

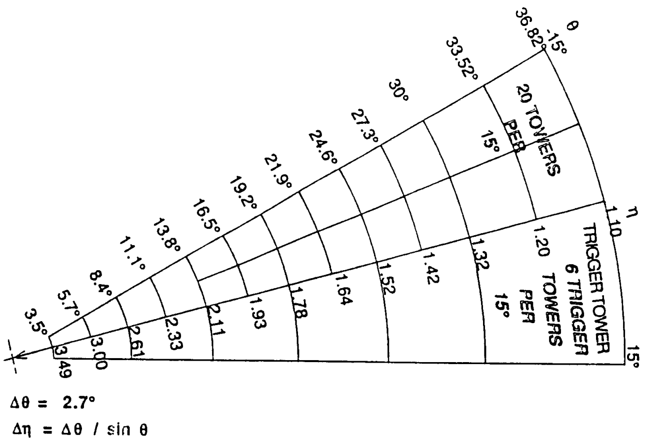


Figure 3.16: *Segmentation of the plug calorimeter.*

the CMX, covering $0.6 < |\eta| < 1.0$, now have a full 2π azimuthal coverage. Fig. 3.17 shows the coverage of each subdetector in the $\eta \times \phi$ coordinates. It should be noted that the CMU and the CMP coverages do not exactly overlap. The CMU is located outside the Central Hadronic Calorimeter ($\approx 5\lambda$) at a radius of 3.47 m from the beam. It can be reached by muons with $p_T > 1.4$ GeV. The CMP lies behind a 60-cm-thick steel shielding (providing an extra 2.4λ shielding at normal incidence) in order to reduce the rate of hadrons escaping the hadronic calorimeter and faking a muon, and can be reached by muons with $p_T > 2.2$ GeV. The CMX is an arch-shaped detector built around the plug calorimeter. The CMU and CMX (resp. CMP) can be reached by muons of $p_T > 1.4$ GeV (resp. $p_T > 2.2$ GeV). Figure 3.18 shows the number of nuclear absorption lengths a particle coming from the center of the detector goes through before reaching the muon detectors, as a function of pseudorapidity.

The muon system relies on proportional wire chambers to provide coarse tracking information, and scintillation counters for triggering. The three detectors are designed with the same four-layer configuration of drift chambers. Wires in the first and third layer are slightly offset in ϕ w.r.t. the second and fourth layers, in order to remove the ϕ ambiguity in the track reconstruction.

Figure 3.19 shows the four-layer configuration of a CMU module. Each cell is $6.35 \times 2.68 \times 226\,\text{cm}^3$ with a 50 μm stainless steel sense wire in the center parallel to the beam axis. The muon z location is measured using charge division at each end of the wire. The maximum drift time is 800 ns.

CMP cells are $2.5 \times 15 \times 640\,\text{cm}^3$, providing ϕ information only, with a maximum drift time of 1.4 μs. Additionally, a layer of scintillator counters covers the outside surface of the drift chambers. Each counter covers two chambers in width, and half the chamber length, and is read out by a single photo-tube located at the center of the detector.

The CMX cells differ from the CMP cells only by their length (180 cm). The polar angle of the tracks is measured thanks to the stereo angle between adjacent cells, which partially overlap. Because of the conical arrangement of the cells, the overlap is more important near the inner edge, where a track can typically cross six cells. The CMP is equipped with two layers of scintillators, one on each face. Scintillator counters on the inside and outside layers are half-cell staggered, in order to effectively double the granularity of the system. Each counter is read out by a single photo-tube.

In the forward region ($1.0 < |\eta| < 2.0$), the IMU detector replaces the Run I detector; since the IMU commissioning was still in progress during the time data was taken, it is not used for this analysis.

3.2.7 The Cerenkov Luminosity Counter

The Cerenkov Luminosity Counter (CLC) [19] was designed for the Tevatron Run II in order to achieve a precision measurement of the instantaneous luminosity up to $\approx 2 \cdot 10^{32}/\text{cm}^2/\text{s}$) and to cope with the 132 ns bunch-spacing that was originally

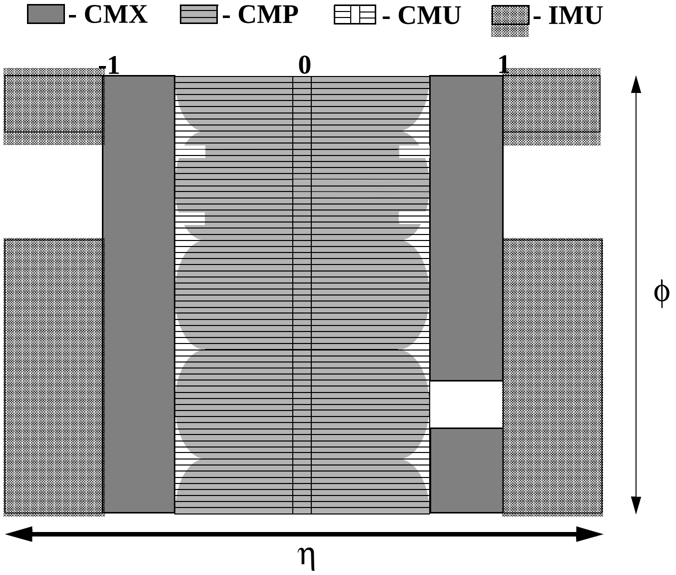


Figure 3.17: *Coverage (in the $\eta \times \phi$ plane) of the upgraded CDF muon system.*

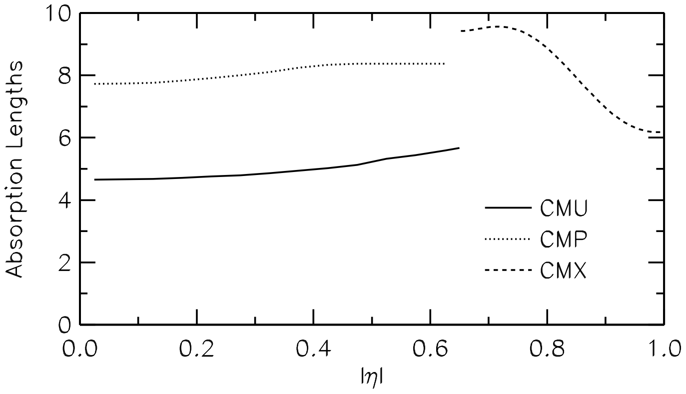


Figure 3.18: *Number of nuclear absorption lengths in front of the muon systems as a function of pseudorapidity.*

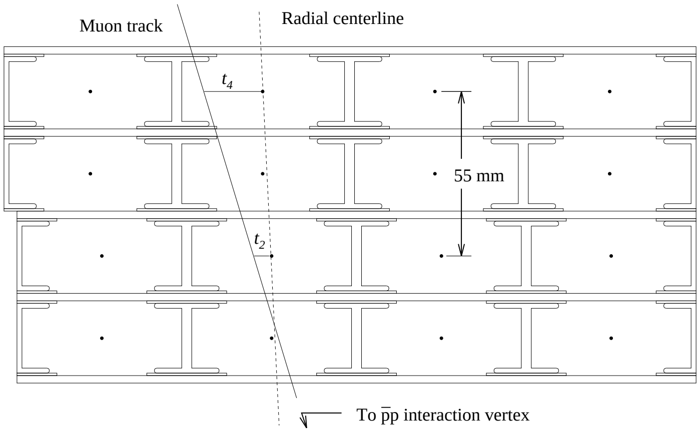


Figure 3.19: *Section of a CMU module in the $x \times y$ plane.*

envisioned. Since luminosity measurement is critical to the cross section measurement presented in this thesis, it explained here in some detail. The detector is made of two identical modules located near the beam pipe and inside the plug calorimeter, on each side of the interaction point. Figure 3.20 shows two views of such a module. Each module is composed of 48 conical gaseous Cerenkov counters pointing toward the interaction point and covering the pseudorapidity range $3.7 < |\eta| < 4.7$. The outer cones are 180 cm long, while the inner ones are 110 cm long; their diameter varies from 2 to 6 cm. At the widest end of each cone (the furthest away from the interaction point), a conical mirror collects the Cerenkov light into photomultiplier tubes (Hamamatsu R5800Q), capable of operating at gains up to $2 \cdot 10^6$ (see detailed view in Fig. 3.20). The modules are filled with iso-butane at atmospheric pressure; it is however possible to increase the pressure up to 2 atm, in order to increase the yield of Cerenkov light. Iso-butane was chosen because of its large refractive index at atmospheric pressure (1.0043) and its good transparency to ultra-violet light.

Because of the narrow shape and the orientation of the cones, particles produced by $p\bar{p}$ interactions close to the center of the detector are likely to go through a large portion of the CLC, producing an important light yield (several hundred photo-electrons), while particles from the beam halo or from secondary interactions traverse the detector at large angle, and have lower energy, hence producing a much smaller light signal. Thus the background is easily rejected by requiring a certain minimal light yield threshold in each channel; the number of particles is measured from the total light yield in the module. Thanks to the CLC excellent time resolution (less 100 ps), it is also possible to select hits from prompt particles by requiring time coincidence between hits in the two different modules. The time distribution of hits in each module is shown in Fig. 3.21. One can clearly see the contributions from the proton beam halo, the antiproton beam halo, and actual $p\bar{p}$ collisions.

The instantaneous luminosity can be expressed as a function of the measured number of hits per bunch-crossing as follows:

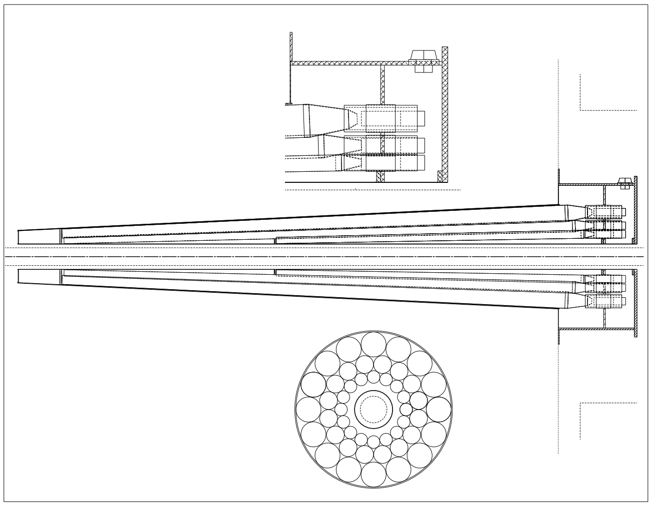


Figure 3.20: *The CDF Cerenkov Luminosity Counter.*

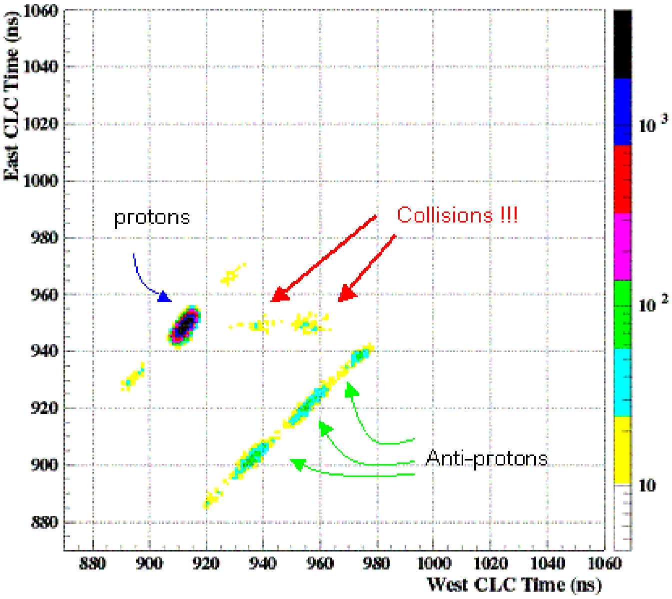


Figure 3.21: *Time distribution in the East and West modules. Actual $p\bar{p}$ collisions deposit a coincident signal in the two modules.*

$$L = \frac{f_{\mathcal{BC}}}{\sigma_{\text{inelastic}} \cdot \epsilon} \cdot \frac{< N_H >}{< N_H^1 >}$$

where:

- $f_{\mathcal{BC}}$ is the bunch-crossing frequency.
- $\sigma_{\text{inelastic}}$ is the inelastic scattering cross section.
- ϵ is the acceptance of the CLC for inelastic scattering events.
- $< N_H >$ is the average number of hits (particles) per bunch-crossing.
- $< N_H^1 >$ is the average number of hits (particles) for a single $p\bar{p}$ inelastic interaction.

The total inelastic scattering cross section was measured in several experiments using a luminosity-independent method combining Regge theory [20] to obtain the total cross section and the optical theorem to compute the inelastic cross section from the total cross section. CDF Run I and E811 measurements were combined, giving $\sigma_{\text{inelastic}} = 60.4 \pm 2.3$ mb at 1.8 TeV, which can be extrapolated to 61.7 ± 2.4 mb at 1.96 TeV. This is the value used in this thesis.

A precision of 5.9% on the luminosity is achieved with the CLC, with the main systematic uncertainties due to the detector stability (1.8%), the CLC acceptance (4%), and the inelastic cross section (3.8%).

3.2.8 The Trigger System

The CDF II trigger system is composed of two hardware levels — Level 1 (L1) and Level 2 (L2) — and a software level — Level 3 (L3) —, to select interesting events and reduce the very high rate of collisions to an acceptable one. For Run II, the trigger system was designed to cope with a bunch crossing spacing of 132 ns (7.6 MHz); currently, the Tevatron actually operates with a 396 ns bunch spacing, which corresponds to an interaction rate of 1.7 MHz, since not all 36 bunches of each beam actually contain particles (see Sec. 3.1).

Figure 3.22 shows a block diagram of the data flow at CDF II. In a first stage, the Level 1 gathers information from the calorimeters, the COT and the muon system and makes a decision within 5.5 μs. In particular, a route-based hardware tracking algorithm (the "extra-fast tracker", XFT) provides tracking information from the COT to the Level 1. In order to avoid dead time while the Level 1 decision is being made, a 42 deep pipeline stores subsequent events until the Level 1 is available to process them. In a second stage, the Level 2 further selects events that pass the Level 1 selection. The Level 2 is equipped with a 4 event deep buffer, and can cope with input rates up to 50 kHz; currently, rates of the order of 25 kHz are common. The Level 2 takes $\approx 20\mu$s to make a decision. The Level 2 is able to reconstruct

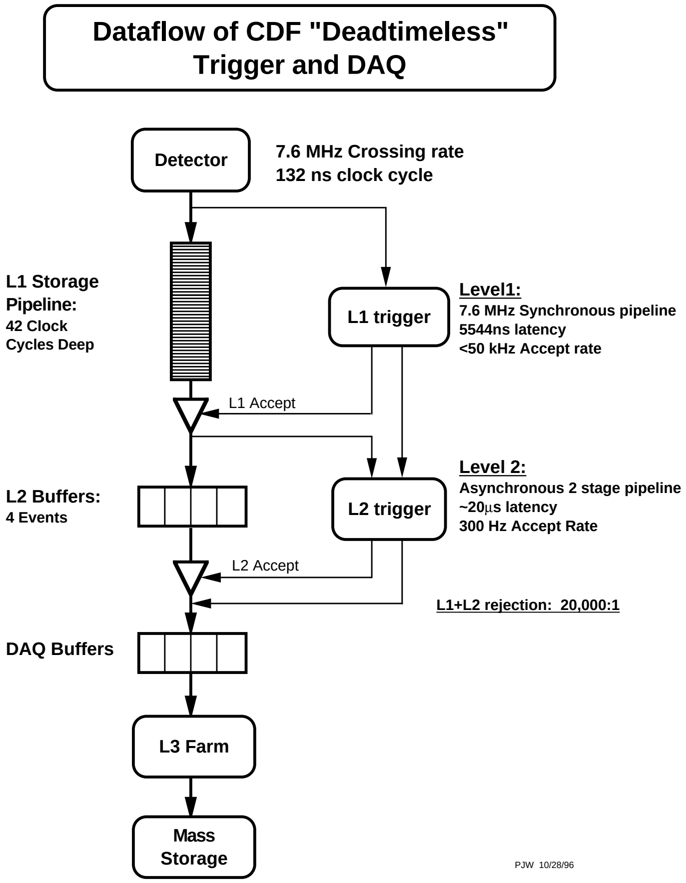


Figure 3.22: *Block diagram of the data flow at CDF.*

calorimeter clusters, and to use the maximum shower detector information. A novelty in hadronic physics, it is also able to use the silicon vertex detector: the silicon vertex trigger (SVT) uses XFT tracks as an input and tries to reconstruct tracks based on silicon hits in the neighborhood of an XFT tracks; this technique significantly reduces the number of candidate hits, hence allows very fast reconstruction, while achieving a resolution comparable with the full tracking reconstruction. The SVT is able to identify tracks that are significantly displaced from the beam location, hence selecting heavy flavor enriched events. Figure 3.23 shows what information is available to Level 1 and Level 2. Finally, a cluster of $\approx$ 200 Linux PCs reconstructs the events that pass Level 2 using a simplified version of the software used for offline analyses. Level 3 reduces the rate of events from $\approx$ 300 Hz to $\approx$ 50 Hz. Events that pass Level 3 are monitored in real time in the control room (see Appendix D for details on the silicon vertex detector online monitoring), and stored on tape at the Feynman Computer Center.

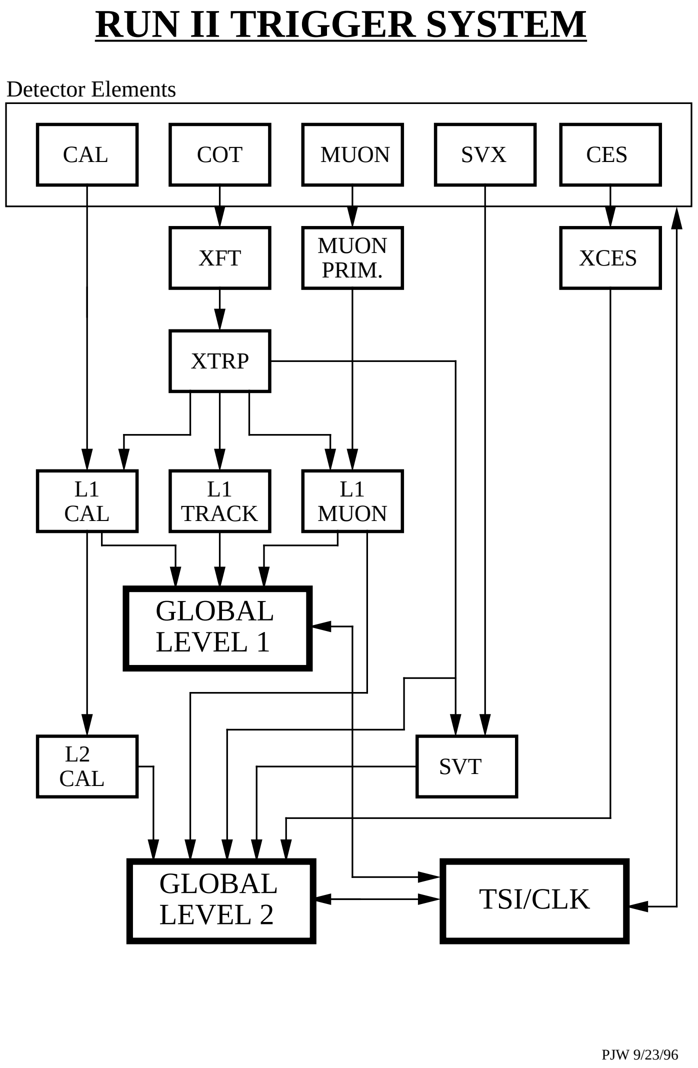


Figure 3.23: *Block diagram of the Level 1 and Level 2 triggers.*

3.2.9 Detector Simulation

Physical processes of the $p\bar{p}$ interactions are simulated by Monte Carlo (MC) generators: the HERWIG [23, 24], PYTHIA [25] and ALPGEN [26] generators are used in conjunction with the CTEQ5L PDF's [21] throughout this document. The QQ v9.1 [22] decay tables are interfaced to provide proper modeling of b and c hadron decays. For each event, these generators produce a list of particles (and their four-vector) that define the final state of the process[2]. In order to simulate the detector response, a complex simulation of the interaction of the particles present in the final state with the detector is needed. The GEANT 3 [27] package is used together with a detailed description of the CDF sub-detectors and the passive material (such as readout systems, cables, *etc.*). Specific models are used for some of the sub-detectors. The calorimeter response is simulated with a parametrized shower simulation package GFLASH [28] tuned for single particle response and shower shape based on data. The simulation of the COT drift time uses the GARFIELD package [29, 30]. For the silicon, a parametrized charge deposition model [31] is used in place of GEANT 3. The simulation includes the modeling of noise and dead regions in the silicon detectors: the hardware and calibration silicon detector databases were used to account for bad silicon strips, noise, and defective modules in order to reproduce the actual detector configuration in the Monte Carlo. The product of the simulation is stored in the same format as the real data, allowing the reconstruction and selection of simulated events to be performed in exactly the same way as real events.

---

[2] To be accurate, ALPGEN only produces a list of partons, while HERWIG and PYTHIA also simulate the hadronization of gluons and quarks. ALPGEN needs to be used in combination with another generator in order to produce the actual final state. More details are given in Chap. 7.

Chapter 4

Event Reconstruction and Selection

This chapter describes the algorithms and techniques used to reconstruct charged particle trajectories, to identify electrons, muons, photons and jets, and to measure kinematic quantities. The event selection specific to the $t\bar{t}$ lepton+jets channel and the data samples used in this document are described in the last section.

4.1 Tracking Reconstruction

4.1.1 COT Tracking

The COT drift chamber is able to efficiently reconstruct the trajectory of charged particles ("tracks") and measure their momenta up to $|\eta| = 1$. The tracking pattern recognition algorithm works in several successive steps [34]:

- **Segment finding:** in each of the eight superlayers, sets of three hits consistent with a straight line trajectory are used as a segment-seed, and fit to a straight line by the method of least squares. Other hits in the superlayer lying within a road of 20 ns ($\approx$ 1 mm, for a drift time of 55 μm/ns) are added to the straight line fit in an iterative procedure.

- **Axial tracking reconstruction:** in the first pass of the tracking reconstruction, tracks are reconstructed in 2D (transverse plane), using axial superlayers only. Two algorithms are run in parallel for this purpose, in order to increase the reconstruction efficiency:

  - The "segment-linking" algorithm matches segments from different superlayers, then fits all the hits in the segments by minimizing a reduced χ^2.
  - The "histogram-linking" algorithm starts with a simple circle fit defined by one segment and the beam position. It then looks for hits within a 1 cm road about the circle and fills a 200 μm-binned histogram with the radius (*i.e.* the distance to the center of the track circle) of each hit. If the most populated bin contains at least 10 hits, a track is made out of those hits. Finally, the algorithm attempts to find additional hits within 750 μm of the track, and re-fits the track.

  Track duplicates (reconstructed by both algorithms) are removed in the end.

- **Stereo tracking reconstruction:** information from the stereo angle superlayers is added to the axial tracks in this last step. Again, two algorithms are combined, but this time run in series:

  - For every axial track, the stereo segment-linking algorithm attempts to add a stereo segment from the outer stereo superlayer, re-fits the track to get a first estimate of the track z and $\cot\theta$, and then looks for additional stereo segments in the inner layers.

  - Once the stereo segment-linking is completed, the stereo tracks are used to identify the z coordinate of vertices in the event, using a histogram method. Axial tracks that could not be extended to a 3D track with the stereo segment-linking algorithm go through a stereo hit-linking algorithm, which scans the list of z vertices and attempts to add stereo hits for a given z vertex.

- **Re-fit:** finally, tracks are re-fit taking into account corrections due to the actual magnetic field map, and drift model.

The efficiency for isolated tracks of large transverse momentum ($p_T > 1.5$ GeV) was found to be larger than 99%. This efficiency drops to $\approx 95\%$ for $p_T \approx 500$ MeV [35].

4.1.2 Silicon Detector Tracking

Hit Clustering Algorithm

A charged particle going through a silicon sensor often deposits charge in several contiguous strips. In the SVX and the ISL, only strips above a certain threshold and their immediate neighbors are read out. The purpose of the hit clustering algorithm is to identify sets of strips that belong to the same track, gather them into a strip cluster (or "hit"), and estimate the exact location of the track impact. In a first step, the algorithm identifies sets of contiguous above-threshold strips ("super-clusters"). It then splits super-clusters into clusters at local minima, and identifies local maxima (peaks). If any local maximum exists, the strip at a local minimum is assigned to the closest peak; if the two peaks are equally distant, the local minimum strip charge is split between the two clusters. Finally, the hit position is estimated by the charge centroid of the cluster (weighting the central location of each strip by its charge).

Outside-In Algorithm

COT tracks are used as seeds to look for hits on the silicon detector. The hits are attached to the track in an iterative procedure, starting with the outermost layer, and moving inward (referred to as "outside-in" (OI) algorithm [36]). In a first pass, only axial hits are attached to the track. In a second pass, the algorithm attempts to add stereo (small angle and 90°) hits. Only tracks with at least three axial hits are kept. The algorithm looks for hits lying in a "road" four standard deviations wide around the seed track. At each iteration (*i.e.* in each layer), several hits may be found within the road: a new track candidate is considered for each hit found, so that a "tree" of candidate tracks is produced from a single COT seed track. In order to speed up the algorithm, only the tracks with the most and second most number of silicon hits are considered for the following iteration. The seed track is also kept for the following iteration, in case none of the hits found belong to the track. At each iteration, the track is re-fit in order to improve the pointing resolution and to take into account the amount of scattering material encountered. At the end of the process, only one track is kept, based on the number of silicon hits attached to it, and on the χ^2 of the fit.

4.2 Electron and Photon Identification

Electron[1] and photon identification relies on the combination of the tracking and calorimeter information. Electrons and photons leave a characteristic signature in the calorimeters, since an electromagnetic (EM) shower almost entirely contained in the EM calorimeter. Electrons are distinguished from photons in part by the slight difference of shape of the EM shower, but mostly by requiring that its trajectory be reconstructed and point to the calorimeter cluster produced by the EM shower. Photons, being neutral, do not leave any trace in the tracking system.

Charged hadrons can mimic an electron signature if they shower early in the solenoid or in the EM calorimeter. Also, an electron can be misidentified as a photon if the electron track is not reconstructed. Alternatively, a photon can be misidentified as an electron if the photon converts to an electron-positron pair as it goes through material, or if a track is wrongly associated with the EM cluster. Photon conversions are identified by looking for pairs of COT tracks satisfying the following cuts:

- The two tracks have opposite sign,
- $|\Delta(xy)| < 2\,\mathrm{mm}$,
- $|\Delta(\cot\theta)| < 0.04$,

where $\Delta(xy)$ is the distance between the tracks in the transverse plane at the point where they are tangent in that plane, and $\Delta(\cot\theta)$ is the difference between the polar angle cotangent of the two tracks. If a candidate electron combined with a partner track is consistent with a photon conversion, it is rejected. However, if a third track can be combined with the electron partner track to form a conversion as well ("trident"), the conversion is likely to be due to a high-energy bremsstrahlung photon emitted by the initial electron as it goes through matter: in this case, the candidate electron is real, and is not rejected.

Since photon identification is not used for the analysis presented here, it is not discussed any further. Electron identification in the central calorimeter (CEM electron) for Run II is almost identical to the one used in Run I [37]. An electron candidate is considered if a tower in the CEM has $E_T > 2$ GeV and a track points to this tower. Adjacent towers in the same wedge (*i.e.* towers with same ϕ) are added to the cluster. The energy attributed to the electron is the total energy of the cluster. The momentum of the electron is the momentum of the track with highest p_T associated with the cluster; the direction of this track defines the direction of the electron, and is used to compute the electron $E_T = E \cdot \sin\theta$. The selection cuts shown in Table 4.1 using the variables described below efficiently select electrons and reject the background:

---

[1] Electron refers to both the electron and its antiparticle, the positron.

- The track is required to satisfy some quality requirements: it should come from the luminous region ($|z_0| < 60$ cm), and have a large number of hits attached to it (at least three axial and three stereo superlayers must provide segments with at least seven hits to the track reconstruction). The track must point to an instrumented region of the detector; in particular it should be away from the tower edges, and from the "chimney".

- E is the total energy of the electron calorimeter cluster. p is the electron momentum measured from the track. p_T is the projection of the electron momentum onto the transverse $x \times y$ plane.

- E_{HAD} (E_{EM}) is the energy of the electron in the hadronic (electromagnetic) calorimeter.

- L_{shr} is a variable relative to the EM shower shape. The energy deposited in the two towers adjacent to the tower at the center of the cluster is compared to the one expected from test beam data:

  $L_{shr} = 0.14 \cdot \frac{\Sigma_i (E_i^{obs} - E_i^{exp})}{\sqrt{(0.14 \cdot \sqrt{E_{EM}})^2 + \Sigma_i (\delta E_i^{exp})^2}}$

  L_{shr} is required to be less than 0.2.

- The shower maximum detector (CES) is used to reject possible hadron contamination. The track is required to match a CES cluster in both axial ($|\Delta z| < 3$ cm) and azimuthal ($-1.5 \leq Q \times \Delta x \leq 3$ cm, where Q is the charge of the electron) directions. In the azimuthal direction, the shower asymmetry caused by the electron bremsstrahlung is taken into account. The shape of the CES cluster is required to be similar to the one evaluated from test beam data based on a χ^2 test ($\chi^2_{CES} \leq 10$).

The electron transverse energy is reconstructed from the electromagnetic cluster with a precision $\sigma/E_T = 13.5\%/\sqrt{E_T(\mathrm{GeV})} \oplus 2\%$ [38].

4.3 Muon Identification

Unlike electrons, muons do not initiate an EM shower in the calorimeters due to their larger mass (105 MeV compared to 0.511 MeV). Unlike hadrons, muons do not interact strongly, hence do not shower in the hadronic calorimeter either. As a result, muons with a transverse momentum of a few GeV or more deposit only a small fraction of their energy in the calorimeters due to ionization, and escape the detector. Muons are thus identified by matching hits in the muon chambers with a reconstructed track, and requiring that little energy be deposited in the calorimeter on the trajectory of the particle. In each muon system (CMU, CMP, CMX), the 4-layer scintillator structure allows to reconstruct a track segment (a so-called "stub"). A muon is reconstructed if such a stub is found in one of the muon systems, and if a

Variable	Value
Fiducialty	Requirement to be in the active region of the CES/CEM, and away from the missing tower (the "chimney").
$\|z_0\|$	$\leq$ 60 cm
# COT axial seg.	$\geq$ 3 with $\geq$7 hits each
# COT stereo seg.	$\geq$ 3 with $\geq$7 hits each
E/p	$\leq$ 2.0 (if $p_T \leq 50$ GeV/c)
E_{HAD}/E_{EM}	$\leq$ 0.055 + 0.00045*E
L_{shr}	$\leq$ 0.2
$\|\Delta z\|$	$\leq$ 3 cm
$Q \times \Delta x$	$\geq$ -1.5 and $\leq$ 3.0 cm
χ^2_{CES}	$\leq$ 10

Table 4.1: *Electron selection cuts. All energies (momenta) are in GeV (GeV/c).*

track points to this stub. Muons can be mimicked by hadrons that shower unusually late — or not at all — in the calorimeter and manage to escape the detector ("punch-through"). Another source of background is due to muons from cosmic rays, which are vetoed by using timing information of the muon chambers and of the COT. Table 4.2 shows the selection cuts that are used in order to efficiently select muons and reject backgrounds. The variables are defined as follows:

- The requirements on the COT track quality are the same as for the electron.
- Additionally, in order to reject cosmic ray background, the track is required to have a low impact parameter d_0 (the distance between the track and the beam at the point of closest approach). If hits from the Silicon vertex detector are attached to the track, the requirement on the impact parameter is more stringent, since the resolution is greatly improved.
- For CMX muons (higher pseudorapidity), the track is required to go through all COT superlayers, *i.e.* exit the COT volume at a radius of 140 cm.
- The energies E_{EM} and E_{HAD} deposited in the EM and hadronic calorimeters, respectively, along the trajectory of the muon are required to be small.
- The track is required to match the muon stub in the axial direction: Δx is the distance in the $x \times y$ plane between the muon chambers' hits and the track extrapolated to the muon chambers' radius. The cut value varies depending on the type of muon chambers involved. There are two types of muons: CMX muons (with a stub in the CMX system), and CMUP muons (with a stub in both the CMU and CMP systems).

Variable	Value
$\|z_0\|$	$\leq$ 60 cm
$\|d_0\|$ (no Si hits)	$\leq$ 0.2 cm
$\|d_0\|$ (Si hits)	$\leq$ 0.02 cm
# COT axial segments	$\geq$ 3 with $\geq$7 hits each
# COT stereo segments	$\geq$ 3 with $\geq$7 hits each
COT exit radius (CMX only)	$\geq$ 140 cm
E_{EM}	$\leq$ max(2, 2 + 0.0115*(p-100))
E_{HAD}	$\leq$ max(6, 6 + 0.0280*(p-100))
$\|\Delta x\|_{CMU}$	$\leq$ 3.0 cm
$\|\Delta x\|_{CMP}$	$\leq$ 5.0 cm
$\|\Delta x\|_{CMX}$	$\leq$ 6.0 cm

Table 4.2: *Muon selection cuts. All energies (momenta) are in GeV (GeV/c).*

The muon transverse momentum is measured by the COT with the following precision:

$$\frac{\sigma_{p_T}}{p_T} = 0.0011 \times p_T[\mathrm{GeV}/c]$$

At high momentum, multiple scattering is negligible and the uncertainty is dominated by the drift chamber intrinsic resolution.

4.4 Jet Reconstruction

Jets are the experimental signature of the production of high-momentum gluons and quarks, which hadronize into a large number of collimated particles that deposit a cluster of energy in EM and HAD calorimeter towers. Jets are reconstructed using a cone algorithm similar to the one used in Run I. As a first step, every tower with a transverse energy $E_T > 1$ GeV ("seed tower") initiates the algorithm. A tower transverse energy is defined as $E_T = E \cdot \sin\theta$, where E is the energy in the tower, and θ is the polar angle of the center of the tower[2]. Towers with $E_T > 100$ MeV in a cone of semi-angle $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ around the seed tower are gathered into a cluster. The centroid of these towers (in the $\eta \times \phi$ space), weighted by the energy of each tower, is computed and the list of towers in the cluster is re-evaluated by now centering the cone on the energy-weighted tower centroid. In an iterative procedure, the centroid and the list of towers in the cluster are computed until the list of towers remains unchanged from one iteration to the next. The last step of the algorithm consists in dealing with overlapping jets: if the transverse energy common to two jets

---

[2] In order to compute tower transverse energies, the z coordinate of the interaction point from which the jet is produced needs to be known; several methods, described later, can be used to evaluate the z coordinate of the interaction.

amounts to more than 75% of the jet with smaller tranverse energy, the two jets are merged; otherwise towers in the overlapping region are assigned to the nearest jet (in the $\eta \times \phi$ space). A cone of semi-angle $\Delta R = 0.4$ is used everywhere in the present document. The approximate jet energy resolution is $(0.1 \cdot E_T[\mathrm{GeV/c}] + 1.0)$ GeV [39].

Several corrections need to be applied to the raw energy measurement in order to flatten the response of the detector and to obtain the same response in the simulation as in the data: η-dependence, energy scale calibration, and multiple interactions are taken into account. No attempt is made to measure the actual momentum of the primary parton at the origin of the jet, which is not needed for this particular analysis (although necessary for top mass measurement, for example). Uncertainties relative to the jet energy measurement are discusses in Chap. 9.

Jet energy resolution is paramount for analyses such as the top mass measurement or Higgs searches; techniques of reconstruction leading to a better resolution than the algorithm described above are being investigated. Appendix C presents a study of a jet algorithm that combines both tracking and calorimeter information in order to take advantage of the good tracking resolution.

4.5 Missing Transverse Energy Measurement

Even though the momentum of the initial parton is inaccessible (see Sec. 2.4), its momentum in the transverse plane is known to be null (in the detector frame). Thus, due to conservation of momentum, the final state is expected to have a null total transverse momentum as well. Because neutrinos escape the detector unnoticed, they create an apparent imbalance in the event transverse momentum. This imbalance is measured by combining the information from every calorimeter tower, and defining the event missing transverse energy $\not{E}_T$ (which is, actually, a momentum):

$$\vec{\not{E}_T} = - \sum_{all\,towers\,i} \vec{E_T^i} \tag{4.1}$$

where $\vec{E_T^i}$ is a 2D vector (in the transverse plane) pointing from the interaction point to the center of tower i, with norm the transverse energy of this tower.

Muons deposit only a small fraction of their energy in the calorimeter: this is taken into account by adding their momentum measured from the COT to the transverse energy sum, and subtracting the energy deposited in the calorimeter along the path of the muons. Finally, in order to improve the resolution on the measurement, the corrections mentioned in the previous section are applied to every jet with raw $E_T \geq$ 8 GeV and $|\eta| < 2.5$.

4.6 Total Transverse Energy Measurement (H_T)

The event global variable H_T is defined as the scalar sum of the transverse energy of all the kinematic objects in the event: electrons, muons, jets, and $\not{E}_T$. For muons, the transverse momentum measured by the COT is used. All jets with $E_T > 8$ GeV and $|\eta| < 2.5$ are included:

$$H_T = \Sigma_{\text{all jets}} E_T + \not{E}_T + E_T^{\text{electron}} \text{ or } p_T^{\text{muon}} \tag{4.2}$$

The H_T variable is representative of the event hard scatter and is used to discriminate $t\bar{t}$ events from the background, as described in the next section and in Chap. 9.

4.7 Data Samples

The analysis presented in this document relies on data accumulated between February 2002 and September 2003, corresponding to an integrated luminosity of 162 pb^{-1}. The primary data set for this analysis is called the W + Jets sample, which contains most of the $t\bar{t}$ events decaying into the lepton + jets mode. The trigger and event selection used to isolate this sample are described below. Other control samples, used to measure the properties of the b-tagger, are also described.

4.7.1 W Sample

The experimental signature of a $t\bar{t}$ event in the lepton+lets decay mode is:

- One W decaying leptonically, producing a single high-momentum electron or muon, and large missing transverse energy $\not{E}_T$ due to the neutrino that escapes the detector.
- Four jets: two jets from the other W decay, and two b jets directly from the top quark decay.

The W leptonic decay gives a clear signature that is used by the hardware and software triggers. Further selections are then made offline.

High-p_T Electron Trigger

At the Level 1 of the trigger, calorimeter towers are gathered in pairs so that the effective $\eta \times \phi$ segmentation is $0.2 \times 15°$. At least one trigger tower is required to have $E_T > 8$ GeV, with an E_{HAD}/E_{EM} ratio less than 0.125. At least one XFT track with $p_T > 8$ GeV/c is required to point to this tower. At Level 2, a clustering algorithm combines the energy deposited in neighboring trigger towers. Towers adjacent to the seed tower found at Level 1 with $E_T > 7.5$ GeV are added to the cluster. The total E_T of the cluster must be larger than 16 GeV. At Level 3, a full event reconstruction and electron identification (as described in the previous sections) is performed. A three-dimensional COT track of $p_T > 9$ GeV must point to a cluster of $E_T > 18$ GeV with $E_{HAD}/E_{EM} < 0.125$.

High-p_T Muon Triggers

The CMUP muon hardware trigger requires hits in the CMU to match hits in the CMP. An XFT track with $p_T > 4\,\mathrm{GeV}/c$ must point to the CMU and CMP hits. Level 2 requires the presence of an XFT track with $p_T > 8\,\mathrm{GeV}/c$, not necessarily matching the muon hits. At Level 3, a fully reconstructed COT track with $p_T > 18\,\mathrm{GeV}/c$ must match a stub in the CMU ($|\Delta x|_{CMU} < 20$ cm) and in the CMP ($|\Delta x|_{CMP} < 10$ cm).

The CMX muon trigger proceeds in a similar way: at Level 1, CMX hits must match CSX hits and an XFT track with $p_T > 8\,\mathrm{GeV}/c$. There is no requirement at Level 2. At Level 3, a fully reconstructed COT track with $p_T > 18\,\mathrm{GeV}/c$ must match a stub in the CMX ($|\Delta x|_{CMU} < 10$ cm).

Offline selection

The first step of the event selection isolates W candidates by identifying events with a single high-momentum muon or electron with $E_T > 20$ GeV and $p_T > 9$ GeV for electrons, and $p_T > 20\,\mathrm{GeV/c}$ for muons. The W identification is completed by requiring that $\not\!\!E_T > 20$ GeV and that the lepton be isolated. The isolation variable I is defined as the ratio of the calorimeter energy in a cone of radius $\Delta R = 0.4$ around the lepton direction (but, in the case of an electron, not including the electron cluster itself) to the lepton energy, and is required to satisfy $I < 0.1$. The isolation requirement is meant to reject leptons from semileptonic decay of heavy flavor hadrons, and leptons faked by hadrons: indeed, in both cases, the lepton candidate is produced in a jet environment and tends to be less isolated than a lepton produced by a W decay.

The second step identifies and counts jets in the W events. Only jets with $E_T >$ 15 GeV and $|\eta < 2|$ are counted. With these requirements, $t\bar{t}$ lepton+jets events are expected to give rise mostly to three or more observed jets, while the sample of W events with one or two jets is expected to contain a negligible fraction of $t\bar{t}$ events. Hence the W+1 and W+2 jet samples are considered control samples, while W events with three or more jets constitute the signal sample for this analysis.

In order to purify the sample, one attempts to identify the b jets in the $t\bar{t}$ events using a secondary vertex b-tagging algorithm (described in Chap. 5 and 6). Events in which at least one of the jets is b-tagged are kept. The sample prior to requiring a b-tagged jet is referred to as the "pretag" sample. Finally, the H_T variable is used to reject the background further, by requiring $H_T > 200$ GeV for events with three jets or more. Chapter 9 describes the event selection optimization and efficiency.

Z^0 and Di-lepton Vetoes

Z^0 boson and top di-lepton decays that contribute to the inclusive high p_T lepton dataset are removed by flagging the presence of a second lepton. Any event with two leptons satisfying the lepton identification is removed. Events where the second lepton is an electron in the plug calorimeter or a muon that fails the CMUP requirement, but has one CMU or CMP muon segment, are also removed. Finally, remaining Z^0 bosons are removed by requiring that there be no second object which forms an invariant mass with the primary lepton between 76 and 106 GeV/c^2. For primary muons the other object is an opposite-signed isolated track with $p_T > 10$ GeV/c. For primary electrons the second object may be such a track, an electromagnetic cluster, or a jet with $E_T > 15$ GeV and $|\eta| \leq 2.0$ that has fewer than 3 tracks and an electromagnetic energy fraction greater than 95%.

4.7.2 Low p_T Electron Sample

A sample of low p_T electrons is used in Chap. 6 to measure the efficiency of the b-tagging algorithm. The trigger and electron identification are the same as for the high p_T electron, but with lower momentum ($p_T > 8$ GeV) and energy ($E_T > 9$ GeV) thresholds .

4.7.3 Jet Samples

Several jet samples are used as control samples to understand the properties of the b-tagging algorithm and the Monte Carlo generators.

- Jet 20, 50, 70, 100 samples:

  The trigger for each of these samples is the same, except for the transverse energy threshold values. For the Jet 20 sample, the trigger at Level 1 requires at least one trigger tower to have $E_T > 5$ GeV. Level 2 requires at least one cluster with $E_T > 20$ GeV. Finally, the Level 3 runs the jet cone algorithm described previously and requires at least one jet with $E_T > 20$ GeV. The trigger for the Jet 50 sample is identical at Level 1; at Level 2 and 3, the threshold is raised to 50 GeV. For Jet 70 and 100, the threshold at Level 1 is raised to 10 GeV. The thresholds at Level 2 and 3 are raised to 70 and 100 GeV, respectively. All of these triggers are heavily prescaled to cope with the very high QCD cross section.

- ΣE_T sample:

  The ΣE_T sample is an independent jet trigger that requires a total transverse energy in the calorimeter at Level 2 and Level 3 larger than 125 GeV.

Chapter 5

Secondary Vertex *b*-Tagging: The SecVtx Algorithm

5.1 Introduction

Identifying heavy flavor jets (*i.e.* jets containing a heavy flavor — bottom or charm — hadron) plays an important role in top physics. While $t\bar{t}$ events produce two b jets from the hadronization of the b quarks from the top decays, most of the non-$t\bar{t}$ processes found in the W + Jets sample do not contain heavy flavor quarks in the final state. Thus identifying b jets significantly reduces the background. Several methods exist. One technique looks for a low momentum electron or muon ("soft lepton tagger") coming from the semileptonic decay of a heavy flavor hadron. This method is limited by the small semileptonic decay branching ratio ($\approx 11\%$) and by the difficulty of reconstructing low momentum leptons in a high-occupancy jet environment. Other methods take advantage of the long life-time of the B (bottom) hadrons: with a mean-life time of the order of 1.5 ps ($c\tau \approx 450$ μm), together with a large relativistic boost, a B hadron in a $t\bar{t}$ event flies for several millimeters (on average) before decaying. The average transverse momentum of a B hadron in a $t\bar{t}$ event is 50 GeV (for a top mass of 175 GeV). For example, for a 50 GeV neutral meson B^0 of mass 5.28 GeV (corresponding to a boost $\beta\gamma = 9.5$) and life-time $c\tau = 460\,\mu m$ with a pseudorapidity $\eta = 0$, the average decay length is 4.4 mm. B hadrons decay to multiple particles, producing a secondary vertex displaced from the primary interaction. Tracks coming from this secondary vertex are reconstructed by the silicon vertex detector with sufficient precision to be distinguished from the prompt tracks produced at the primary interaction. B hadron decay channels are numerous. Most of them involve neutral particles whose trajectories cannot be reconstructed, and/or a C (charm) hadron, which in turn decays to several particles after a finite life-time, producing a tertiary vertex making the kinematics of the decay even more complex. As a result, a full reconstruction of the B hadron is impossible with a high efficiency and more inclusive methods need to be used. The SecVtx algorithm, described in the next sections, selects tracks with a large impact parameter and reconstructs a secondary vertex out of these tracks. The up-side is to be relatively insensitive to the particular B decay involved. The down-side is that C hadrons cannot be distinguished from B hadrons. Although C hadrons have a shorter life-time and decay with lower track multiplicity than B hadrons, they are tagged with a relatively high efficiency as well. Thus the SecVtx algorithm is actually a "heavy flavor tagger". In the rest of this document, the terms b-tagging and heavy flavor tagging are used depending on the context.

The next section describes some characteristics of b jets in $t\bar{t}$ events. In order to distinguish B decay products from the primary interaction, a good measurement of the primary vertex is necessary: Section 5.3 describes the event-by-event primary vertex algorithm. Section 5.4 describes the details of the SecVtx algorithm. In Sec. 5.5, the method used to defined the track selection is discussed.

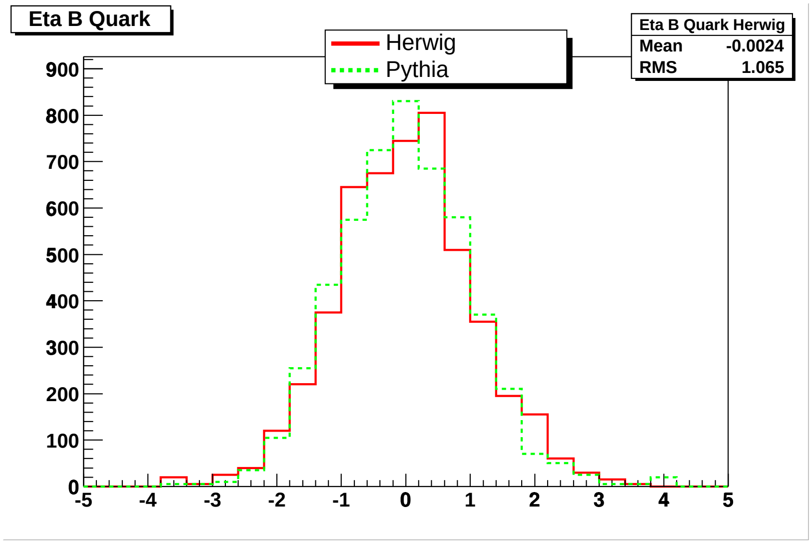


Figure 5.1: *Pseudorapidity of the b quarks from top decays in $t\bar{t}$ lepton+jets Monte Carlo events.*

5.2 Kinematics of b Jets in $t\bar{t}$ Events

In order to illustrate the properties of b jets, samples of $t\bar{t}$ lepton+jets events generated by the PYTHIA and HERWIG Monte Carlo generators are used. Because the top quarks are produced at low momentum relative to their mass, their decay products are not significantly boosted and $t\bar{t}$ events are very spherical. In particular, the b quarks tend to be produced at low pseudorapidity, giving rise to a b jet in the instrumented region of the vertex detector. Figure 5.1 shows the pseudorapidity distribution of the b quarks from top decays in $t\bar{t}$ events: in 96% of the cases, the b quark has a pseudorapidity $|\eta| < 2.2$. Figures 5.2, 5.3, 5.4, 5.5 show the multiplicity, the transverse momentum, the impact parameter, and the impact parameter significance d_0/σ_{d_0}, respectively, of tracks within b jets in $t\bar{t}$ lepton+jets Monte Carlo events. No simulation of the detector is made: instead, the information directly from the event generator is used. Tracks are required to have $p_T > 500$ MeV, to be within $\Delta R = 0.4$ of the primary b quark, and to be in the acceptance of the silicon detector (roughly, this implies a pseudorapidity $|\eta| < 2$). Here perfect tracking reconstruction efficiency is assumed at any pseudorapidity. The uncertainty on the impact parameter σ_{d_0} is simply parametrized as a function of p_T, and the impact parameter is smeared with a Gaussian distribution accordingly. Tracks are categorized into "prompt tracks" (produced at the primary vertex), tracks produced by the B hadron decay, and tracks produced by the sequential C decay, if any. Because the B hadron carries a large fraction of the jet energy, its decay daughters tend to have a larger momentum than the prompt tracks ($\langle p_T \rangle = 3$ GeV/c for prompt tracks, 5.1 GeV/c for tracks directly from the B hadron, and 5.5 GeV/c for tracks from the sequential C hadron), which is helpful in separating them from prompt tracks. Not surprisingly, the impact pa-

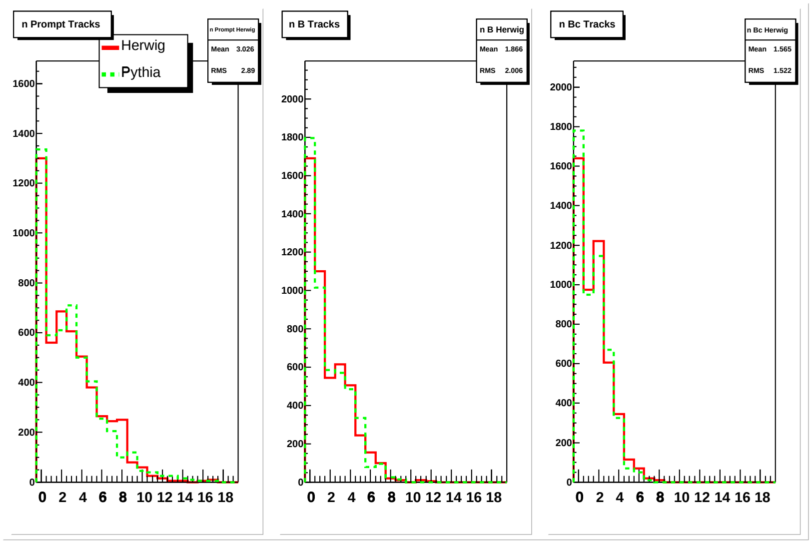


Figure 5.2: *Multiplicity of prompt tracks, tracks from B hadron, and tracks from sequential C hadron (from left to right).*

rameter is much more effective in selecting tracks from the B decay. It is remarkable, however, that a large fraction of the tracks from the B decay are not significantly displaced, due to the exponential shape of the impact parameter distribution. Moreover, some decays lead to very few tracks in the vertex detector, for acceptance reasons and simply because of the charged track multiplicity of the decay. As a result, only 64% (50%) of the b jets contain two tracks (resp. three tracks) or more with $p_T > 0.5$ GeV and $d_0/\sigma_{d_0} > 2.5$. This is an intrinsic limitation of this b-tagging technique.

5.3 Event-by-Event Primary Vertex Finding

In order to roughly estimate the z location of the interaction, the z vertex (described in Sec. 4.1) nearest the identified electron or muon is used[1]. The position of the primary vertex is then determined by fitting together the tracks within a ±1 cm window in z around this vertex. The procedure starts by fitting a vertex using all tracks within the z window and with impact parameter significance (relative to the average beam position) $|d_0/\sigma_{d_0}| < 3$, where σ_{d_0} includes the uncertainty on both the track and the beam position. The transverse profile of the beam at the z of the original vertex estimate is also used as a constraint in the fit. A pruning stage removes tracks which contribute $\chi^2 > 10$ to the fit (or the track with the largest χ^2 contribution if the total fit reduced chi-squared per degree of freedom $\chi^2/\text{ndf} > 5$). After the initial pruning, the fit is repeated using only the remaining tracks until a vertex with no tracks over the χ^2 cut is found. If no tracks survive the pruning stage then

---

[1] For other datasets without high-momentum leptons, such as the jet samples, the vertex which has the highest total scalar sum of transverse momentum is used.

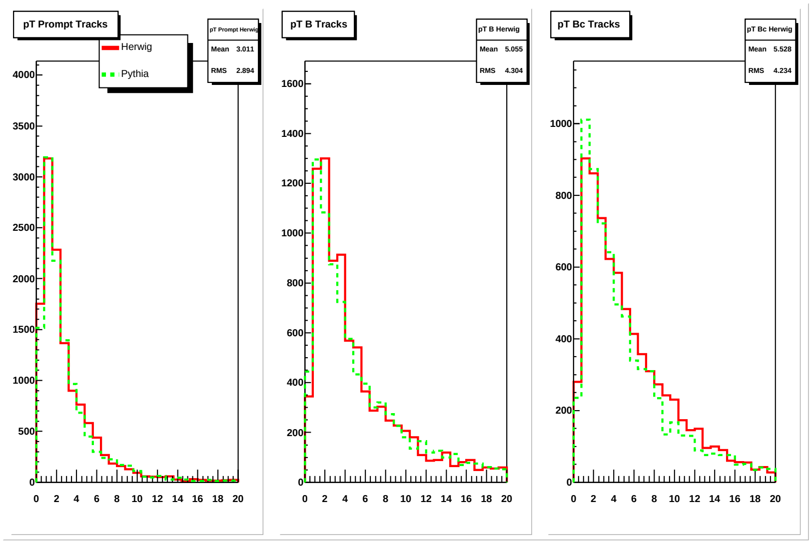


Figure 5.3: *Transverse momentum of prompt tracks, tracks from B hadron, and tracks from sequential C hadron (from left to right).*

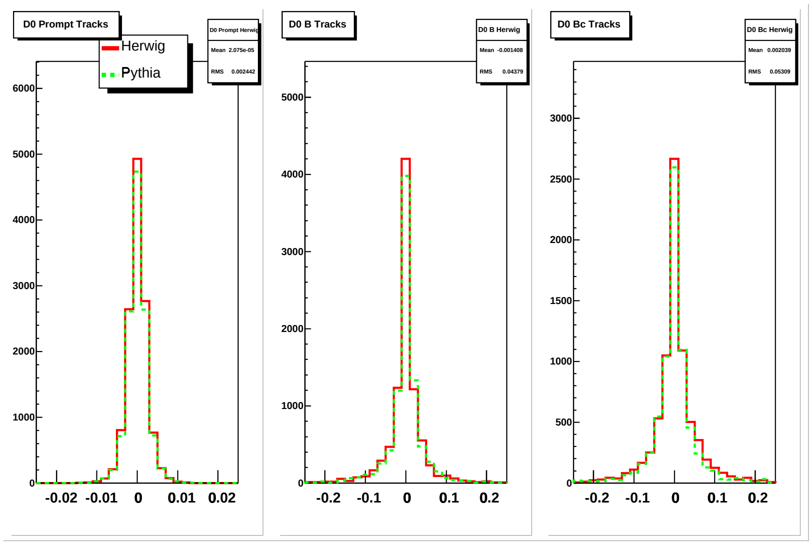


Figure 5.4: *Impact parameter (in cm) of prompt tracks, tracks from B hadron, and tracks from sequential C hadron (from left to right).*

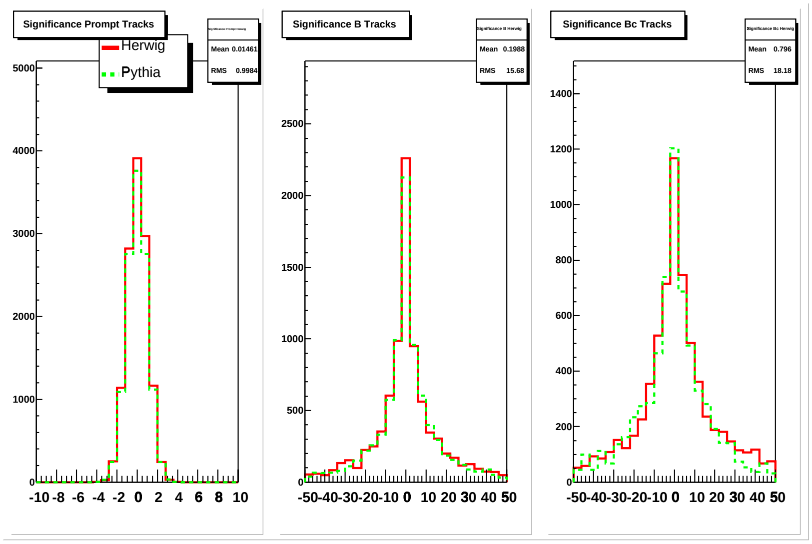


Figure 5.5: *Impact parameter significance* (d_0/σ_{d_0}) *of prompt tracks, tracks from* B *hadron, and tracks from sequential* C *hadron (from left to right).*

the beam profile is used for the primary vertex position estimate. In the W + Jets sample, the uncertainty in the fitted transverse position ranges from 10 to 32 μm depending upon the number of reconstructed tracks and the topology of the event.

5.4 The SecVtx Algorithm

Secondary vertex tagging operates on a per-jet basis, where only tracks within the jet cone of radius $\Delta R = 0.4$ are considered for each jet in the event. A set of cuts involving the transverse momentum, the number of silicon hits attached to the tracks, the quality of those hits, and the χ^2/ndf of the track fit are applied to reject poorly reconstructed tracks. The details of this track selection are described in the next section. Clearly, only jets with at least two of these tracks can produce a displaced vertex; a jet is defined as "taggable" if it contains two selected tracks. Displaced tracks within the jet are selected based on the significance of their impact parameter d_0/σ_{d_0} with respect to the primary vertex and are used as input to the SecVtx algorithm. SecVtx uses a two-pass approach to find secondary vertices:

- The first pass requires at least three tracks with $p_T > 0.5$ GeV/c and $|d_0/\sigma_{d_0}| > 2.5$, out of which it attempts to reconstruct a secondary vertex. At least one of the tracks in the fitted secondary vertex must have $p_T > 1\,\mathrm{GeV}/c$.

- If the first pass is unsuccessful, SecVtx performs a second pass which makes tighter track requirements ($p_T > 1\,\mathrm{GeV}/c$ and $|d_0/\sigma_{d_0}| > 3$, one track with $p_T > 1.5\,\mathrm{GeV}/c$) but requires only two displaced tracks to be found in the jet, and attempts to reconstruct a two-track vertex.

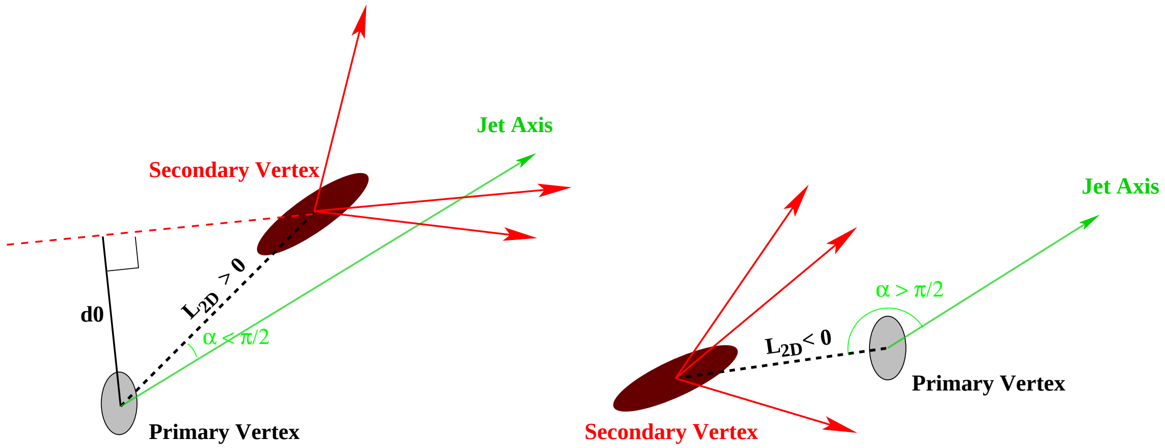


Figure 5.6: *Schematic view of a positive (left) and negative (right) SecVtx tag.*

Once a secondary vertex is found in a jet, the two-dimensional decay length L_{2D} is calculated as the projection onto the jet axis, in the $r \times \phi$ plane only, of the secondary vertex vector pointing from the primary vertex to the secondary vertex. The sign of L_{2D} is defined relatively to the jet direction, specifically by the angle α between the jet axis and the secondary vertex vector (positive for $< 90°$, negative for $> 90°$). Figure 5.8 shows a schematic view of a positive tag (left) and a negative tag (right). Secondary vertices corresponding to the decay of B and C hadrons are expected to have large positive L_{2D} while the secondary vertices from random mis-measured tracks are expected to be less displaced from the primary vertex, and to give a symmetric L_{2D} distribution w.r.t. 0. To reduce the background from the false secondary vertices, a secondary vertex is required to have $L_{2D}/\sigma_{L_{2D}} > 3$ (positive tag) or $L_{2D}/\sigma_{L_{2D}} < -3$ (negative tag), where $\sigma_{L_{2D}}$ is the total estimated uncertainty on L_{2D} including the error on the primary vertex and is typically of the order of 190 μm. Additionally, in order to reject secondary vertices due to material interaction, the secondary vertex must satisfy the following requirements:

- $|L_{2D}| < 5$ cm ($|L_{2D}| < 2$ cm for two-track vertices).
- $|\mathrm{pseudo} - c\tau| < 1$ cm, where $\mathrm{pseudo} - c\tau = \mathrm{L_{2D}} \times \mathrm{M_{sec.vertex}}/\mathrm{p_T^{sec.vertex}}$.

  The mass and momentum of the secondary vertex are reconstructed based on the tracks in the secondary vertex fit.
- $|z_{\mathrm{sec.vertex}} - z_{\mathrm{prim.vertex}}| < 5$ cm.

The negative tags are useful for evaluating the rate of false positive tags ("mistags"), since the L_{2D} distribution of fake tags is expected to be symmetric. Section 8.2 provides more details about mistags. Figure 5.7 shows the L_{2D} distribution of secondary vertices reconstructed by SecVtx for b jets in $t\bar{t}$ Monte Carlo events, after full simulation of the CDF detector. The blue histogram shows all reconstructed vertices, while

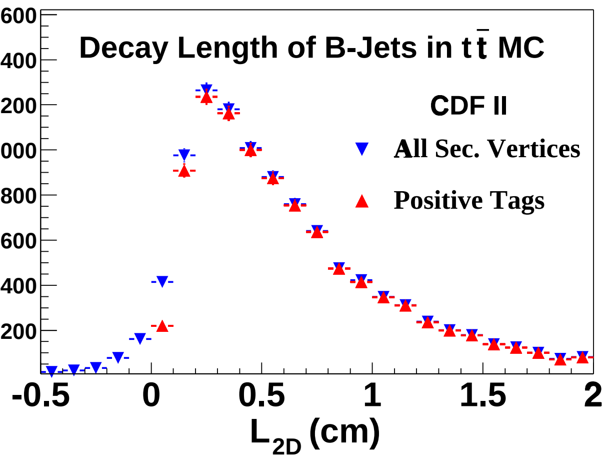


Figure 5.7: *Transverse decay length of secondary vertices reconstructed by SecVtx for b jets in $t\bar{t}$ Monte Carlo events. Blue: all vertices. Red: vertices with $L_{2D}/\sigma_{L_{2D}} > 3$ (positive tags).*

the red histogram shows vertices that pass the $L_{2D}/\sigma_{L_{2D}} > 3$ cut (positive tags). A non-negligible fraction of the vertices are located outside the beam pipe (1.2 cm away from the beam). Figure 5.8 shows a vertex display of an actual data event.

5.5 Track Selection

The vast majority of reconstructed tracks are prompt, meaning that they emanate from the primary interaction point whose location is estimated by the reconstructed primary vertex. Tracks that are displaced from the interaction point can be:

- mis-reconstructed tracks (mostly due to multiple scattering in the material),
- secondary particles produced by nuclear interactions in the detector material,
- the decay product of long life-time particles (B and C hadrons, but also strange particles, mostly K_S and Λ).

We are obviously interested in selecting tracks coming from the decay of a heavy flavor hadron as much as possible. Decays from K_S and Λ are removed explicitly by reconstructing the invariant mass of pairs of tracks. Nuclear interactions are often produced far away from the beam in the detector and at large angle: an upper bound on the impact parameter $|d_0| < 3$ mm removes the majority — but not all — of them. The track selection aims at reducing the fraction of mis-reconstructed tracks and nuclear interaction secondary tracks that pass the impact parameter significance cut described in the previous section. Such tracks tend to have a lower momentum, a larger fit χ^2, fewer attached hits in the silicon detector, or hits of poorer quality than tracks coming from the decay of B or C hadrons. Tracks are selected based on these

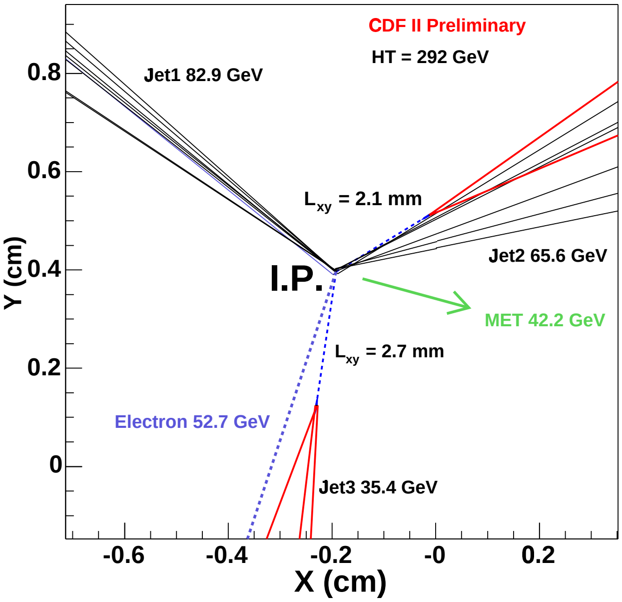


Figure 5.8: *Vertex display of a* W + *3 jets data event with two SecVtx positive tags. Only tracks within a jet cone and passing the selection are shown. Prompt tracks are shown in black; displaced tracks are shown in solid blue, or red if they are part of the final secondary vertex fit.*

criteria. As a first baseline selection, only tracks satisfying the following selection are considered further:

- The track must be within the jet cone:
  $\Delta R = \sqrt{(\eta_{track} - \eta_{jet})^2 + (\phi_{track} - \phi_{jet})^2} < 0.4$
- $p_T > 500$ MeV,
- $|d_0| < 0.3$ cm,
- $|z_0 - z_{PrimaryVertex}| < 5$ cm,
- Not compatible with a K_S, or Λ decay.

In order to measure the purity of the tracks as a function of the selection cuts, a control sample made of generic QCD jet events is used. Such a sample is expected to contain few heavy flavor jets so that, to first order, one can assume that all real, well-reconstructed tracks are prompt, while mis-reconstructed tracks and tracks due to material effect are likely to have a large impact parameter. Thus "real tracks" (signal) are defined as tracks with an impact parameter significance $|d_0/\sigma_{d_0}| < 3$; while fake tracks (background) are defined by $|d_0/\sigma_{d_0}| > 3$. The purity of the track sample is then studied depending on the following variables.

- Track fit χ^2 (normalized to the number of degrees of freedom).
- Number of SVX silicon ϕ (axial) hits.
- Number of **missing** SVX axial hits. Missing hits are identified by extrapolating the track in active modules of the silicon detector where one would expect to find a hit attached to the track, but do not find any.
- Number of **good** silicon hits. A good hits is required not to contain any defective strip, to be made of at most five strips, and not to be shared with any other track satisfying the baseline selection mentioned above.
- The hit pattern: requirement to have a hit on one of the two innermost layers, in some cases.

Figure 5.9 shows the number of missing hits, number of good hits, number of hits, and track χ^2 for all tracks in the control sample satisfying the baseline selection. A large number of classes of tracks is defined, according to bins shown in Table 5.1. For the Pass 1 algorithm (resp. Pass 2), track classes with a signal over background ratio larger than 1 (resp., larger than 2) are accepted. This leads to the selection cuts shown in Table 5.2.

χ^2	< 6; $[6, 12)$; $[12, 18)$; ≥ 18
# hits	3; 4; 5; > 5
# good hits	< 2; 2; 3; 4; > 4
# missing hits	0; 1; 2

Table 5.1: *Bins used to define the track classes (280 classes total). For 3-hit tracks, the hit pattern (i.e. what layers the hits are on) is also considered.*

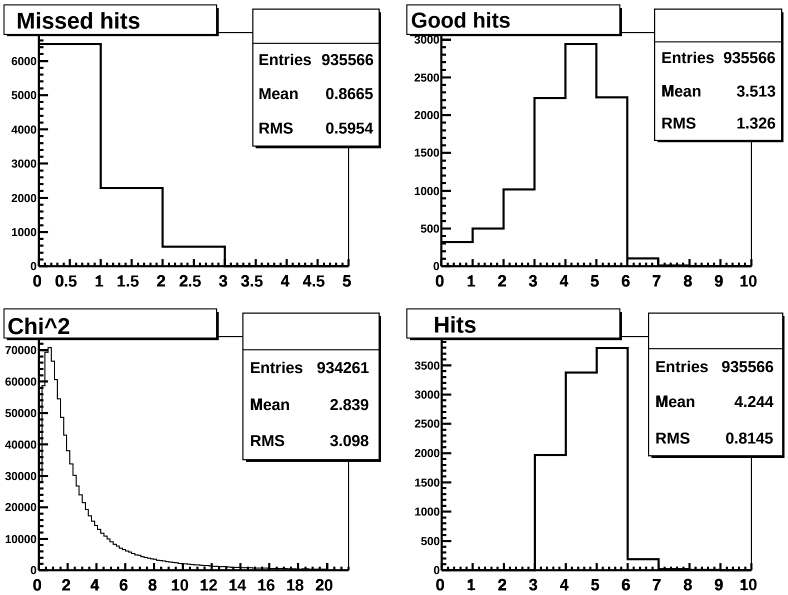


Figure 5.9: *Distribution of the number of missing hits (upper left), number of good hits, number of hits (axial only), and track χ^2 (clockwise) for all tracks in the control sample satisfying the baseline selection [40].*

SecVtx Track Selection

N hit	3 hit	4 hit	$\geq$ 5hit
Pass 1	gd = 3; m = 0; $\chi^2 < 12$ gd = 3; m $\leq$ 2; $\chi^2 < 6$ gd = 2; m = 0; $\chi^2 < 6$	gd = 4; $\chi^2 < 18$; m = 0 gd = 4; $\chi^2 < 12$; m $\leq$ 1 gd = 4; $\chi^2 < 6$; m $\leq$ 2 gd = 3; m = 0; $\chi^2 < 12$ $\chi^2 < 6$; m $\leq$ 1	gd = 5; $\chi^2 < 18$ gd = 4; $\chi^2 < 12$ $\chi^2 < 6$; m $\leq$ 1
Pass 2	gd = 3; m $\leq$ 1; $\chi^2 < 6$ Hit on SVX L0 or L1	gd = 4; $\chi^2 < 6$; m $\leq$ 1 gd = 4; m = 0; $\chi^2 < 6$	gd = 5; $\chi^2 < 12$; m $\leq$ 1 m = 0; $\chi^2 < 6$

Table 5.2: *In addition to the SVX requirements specified above, the associated COT track is required to have at least two axial and two stereo superlayers with at least 6 hits each, and a total of at least 19 axial hits (16 stereo hits). "gd" = number of good hits; "m" = number of missed hit.*

Chapter 6

Performance of the SecVtx *b*-Tagging Algorithm

A precise understanding of the b-tagging algorithm performance is necessary to measure the $t\bar{t}$ production cross section. The efficiency to identify a b jet ("b-tagging efficiency") needs to be taken into account in the $t\bar{t}$ event and heavy flavor backgrounds acceptance, while the wrong identification of light jets as b jets ("mistags") is a source of background.

The measurement of the heavy flavor tagging efficiency is described in Sec. 6.1 to Sec. 6.7. Because it is not possible to measure the tagging efficiency *in situ* on $t\bar{t}$ events, a different strategy has been adopted. A sample of jets whose heavy flavor content can be measured is used to derive the tagging efficiency in the data and a matching sample of Monte Carlo jets is used to determine the tagging efficiency in the simulation. The ratio of efficiencies between data and simulation, named the "tagging scale factor", is then used to tune the tagging efficiency in $t\bar{t}$ Monte Carlo samples when determining the $t\bar{t}$ event acceptance.

The evaluation of the mistag rate is described in Sec. 6.8.

6.1 Data and Monte Carlo Samples

A sample of low-p_T electron data (see Sec. 4.7) is used. This sample is rich in heavy flavor jets containing semileptonic decays of bottom and charm hadrons. Electrons in the events are identified using the selection described in Sec. 4.2. The electron is required to satisfy $E_T > 9\,\mathrm{GeV}$, $p_T > 8\,\mathrm{GeV}/c$, to be non-isolated (isolation $I > 0.1$), and to go through every layer of the silicon detector. Conversions are not vetoed, for a reason that will appear clear in the next section. If several such electrons are identified, the one with highest E_T is selected.

The electron is required to be part of a jet with $E_T > 15$ GeV (*i.e.* the electron is within $\Delta R = 0.4$ of the jet axis), referred to as the "electron-jet" in the following. In case of ambiguity, the jet nearest the electron is chosen. The electron-jet is presumed to contain the semileptonic decay products of a heavy flavor hadron. Finally, events with a jet back-to-back with the electron-jet (referred to as the "away-jet") are selected. The away-jet is required to have $E_T > 15$ GeV and $|\eta| < 1.5$, and the azimuthal angle between the two jets must be larger than 2 rad. If several jets qualify as an away-jet, the one furthest from the electron-jet is selected. The goal of this selection is to isolate a sample of $b\bar{b}$ back-to-back di-jet events in which one of the two b jets contains a semileptonic B decay; the exact composition of the sample is far from this ideal scenario and is discussed later.

In order to produce a Monte Carlo sample matching the data, the HERWIG [23, 24] program was used to generate 2→2 parton events (at tree level, di-jet events), which were passed through a filter requiring an electron with $p_T > 7\,\mathrm{GeV}/c$ and $|\eta| < 1.3$. Events passing this filter were processed using the detector simulation described in Sec. 3.2.9. Finally, a rudimentary trigger simulation was implemented, using the results reported in [41], by randomly discarding events according to the measured trigger efficiency, parametrized in electron p_T and E_T; about 10% of the events were rejected.

Figures 6.1 and 6.2 show a comparison of the Monte Carlo with the data after the event selection. Here, electrons identified as conversions are vetoed, since the data contain many more than the Monte Carlo sample produced. Figure 6.1 shows some event kinematic quantities for events in which the electron-jet is positively tagged by the SecVtx algorithm. The agreement is fair, except for the electron-jet E_T, which is slightly lower in the Monte Carlo. This discrepancy is attributed to conversions that were not removed due to the inefficiency of the conversion removal. Figure 6.2 shows some properties of the SecVtx secondary vertices found in the electron-jet. The number of tracks passing the SecVtx selection is smaller in the data than in the Monte Carlo: because the Monte Carlo does not describe well the underlying event, and does not include minimum-bias multiple interactions, and other interactions due to the beam halo, it does not reproduce the large number of low momentum tracks that go through the tracking system; as a result, the requirements on the silicon tracks (especially in terms of quality of silicon hits) are more efficient in the Monte Carlo than in the data. This directly affects the b-tagging efficiency and scale factor. For the same reason, the data contains many more tags with negative L_{2D} (caused by poorly measured tracks) than the Monte Carlo: thus the mistag rate cannot be estimated from the simulation; it is estimated from the data, as explained in Sec. 6.8.

6.2 Method

Given a perfectly pure sample of b jets, the b-tagging efficiency could be simply estimated by the fraction of b jets that are positively tagged. Unfortunately, the sample is far from being pure. For illustration, Fig. 6.3 shows schematically the composition of the electron-jet. Only about a quarter of the electron-jets are actually heavy flavor jets (a non-negligible fraction of which are c jets). Indeed, most of the identified electrons are either hadrons faking the signature of an electron, or photon conversions. Even if one requires the away-jet to be identified as a b jet (*i.e.* to be positively tagged), the fraction of heavy flavor electron-jets is still far from being 100%, for two reasons: the away-jet may be mistagged, or the physical process may involve a heavy flavor jet only on one side of the event (in the case of gluon splitting, for example). Consequently, it is necessary to measure precisely the sample heavy flavor content. In order to be less sensitive to this measurement, the purer sub-sample of events in which the away-jet is positively tagged is used, so that the tagging efficiency (assuming no mistags for now) is estimated by:

$$\varepsilon = \frac{N_{\mathrm{a}+}^{\mathrm{e}+}}{N_{\mathrm{a}+}} \cdot \frac{1}{F_{HF}^{a+}}, \tag{6.1}$$

where $N_{\mathrm{a}+}^{\mathrm{e}+}$ is the number of events in which both jets are positively tagged, $N_{\mathrm{a}+}$ is the number of events in which the away-jet is positively tagged, and F_{HF}^{a+} is the

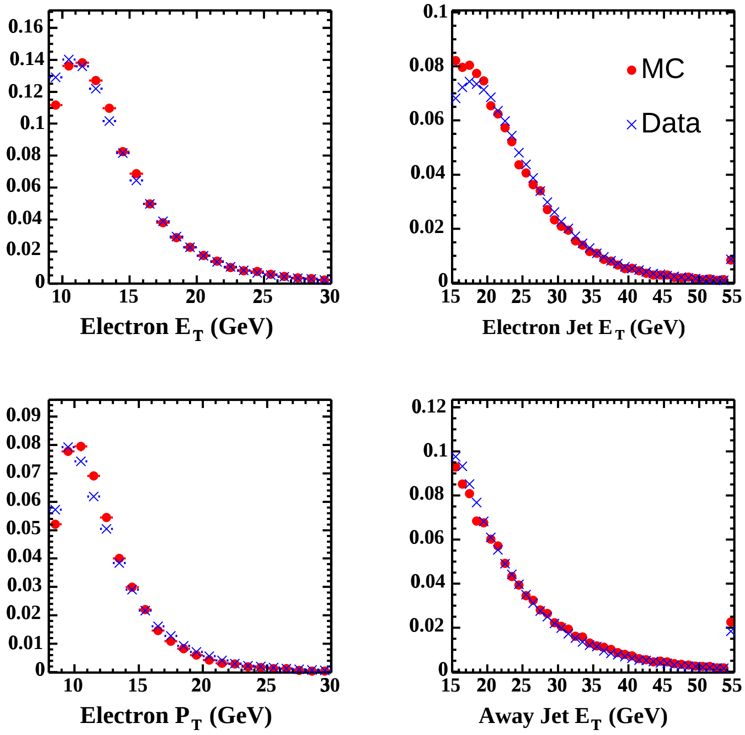


Figure 6.1: *Data/Monte Carlo comparison of some kinematic quantities in events where the electron-jet is positively tagged (identified conversions are vetoed). From top-left, clockwise: electron E_T, electron-jet E_T, away-jet E_T, electron p_T.*

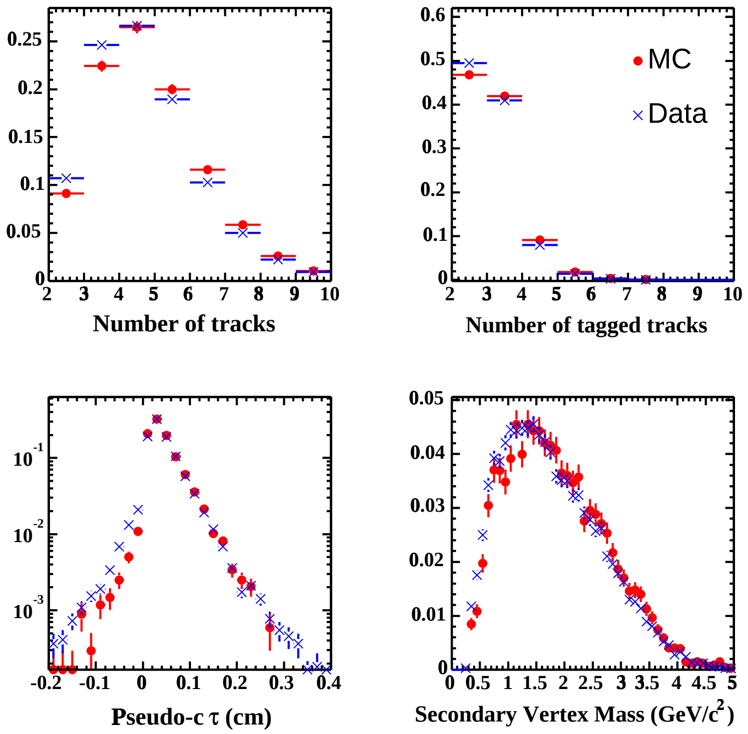


Figure 6.2: *Data/Monte Carlo comparison of some quantities of tagged electron-jets (identified conversions are vetoed). From top-left, clockwise: number of selected tracks, number of tracks in vertex, and vertex mass of positively tagged electron-jets; pseudo-*$c\tau$ *of (positively or negatively) tagged electron-jets.*

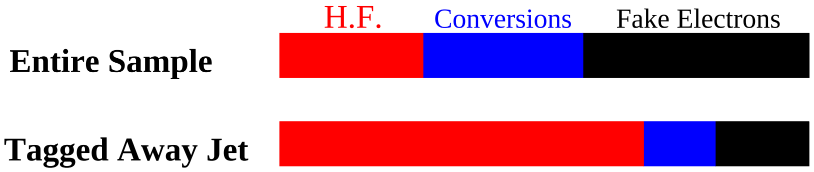


Figure 6.3: *Schematic composition of the electron-jet with (bottom) and without the requirement that the away-jet be positively tagged. Among light electron-jets (conversion electrons and fake electrons), the fraction of conversion electrons is independent of the away-jet heavy flavor property.*

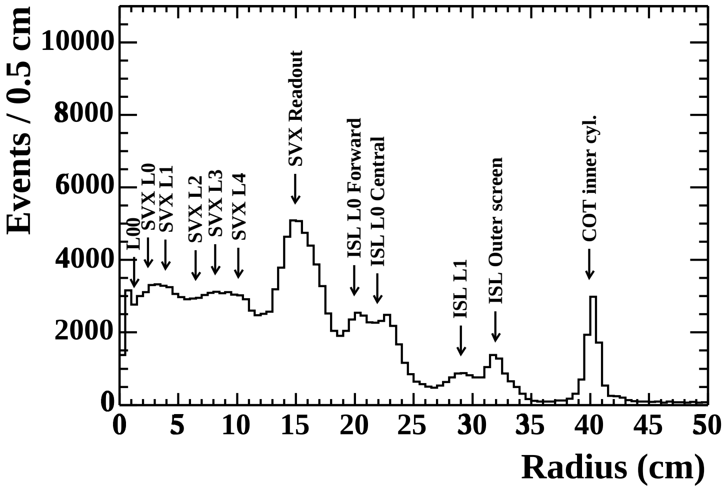


Figure 6.4: *Radius of identified conversions in data.*

fraction of heavy flavor electron-jets when the away-jet is positively tagged. Another advantage of this method is to reduce the fraction of c jets to a negligible level (compared to other sources of systematics). This sub-sample, however, suffers from poor statistics and its heavy flavor content cannot be measured directly. Instead, the heavy flavor fraction is measured in the entire sample, and then extrapolated to the tagged away-jet sub-sample by use of the number of identified conversion electrons. The underlying assumption of this technique is that the heavy flavor production in the away-jet does not depend on whether the identified electron is a fake electron or a conversion electron. The conversion identification is the same as described in Chap. 4. No attempt is made to identify tridents. Figure 6.4 shows the conversion radius, with peaks corresponding to known detector structures.

Appendix A describes in detail the derivation of the efficiency computation. The result is quoted here. Assuming that the number of negative tags is a good estimate of the number of positive mistags, the equation for the efficiency becomes:

$$\varepsilon = \frac{(N^{\mathrm{e}+}_{\mathrm{a}+} - N^{\mathrm{e}-}_{\mathrm{a}+}) - (N^{\mathrm{e}+}_{\mathrm{a}-} - N^{\mathrm{e}-}_{\mathrm{a}-})}{(N_{\mathrm{a}+} - N_{\mathrm{a}-})} \cdot \frac{1}{F^{a+}_{HF}}, \tag{6.2}$$

$$F^{a+}_{HF} = 1 - \frac{\left(\frac{N^{\mathrm{a}+}_{\mathrm{c}} - N^{\mathrm{a}-}_{\mathrm{c}}}{N_{\mathrm{a}+} - N_{\mathrm{a}-}} - \epsilon'_c\right)}{\frac{N_{\mathrm{c}}}{N} - \epsilon'_c} \cdot (1 - F_{HF}), \tag{6.3}$$

where $\epsilon'_c = \frac{N^{\mathrm{e}+}_{\mathrm{c}} - N^{\mathrm{e}-}_{\mathrm{c}}}{N_{\mathrm{e}+} - N_{\mathrm{e}-}}$. A "c" subscript refers to events in which the electron is identified as a conversion. "$a\pm$" and "$e\pm$" refer to the nature (positive or negative) of the tag on the away- (electron-) jet. N is the total number of events (for example, $N^{\mathrm{e}-}_{\mathrm{a}+}$ is the number of events where the electron-jet is negatively tagged and the away-jet is positively tagged). F_{HF} is the electron-jet heavy flavor fraction $F_{HF} = F_b + F_c$, where F_b (F_c) is the fraction of b jets (resp. c jets) in the entire sample, whose measurement is described in the next section.

6.3 Measurement of The Electron-Jet Heavy Flavor Content

Two methods are used to measure the b fraction F_b of the electron-jets:

- Reconstruct $D^0 \to K\pi$ decays.
- Identify muons from semileptonic D decays.

Together with the electron identification, these are signatures of a b decay. In order to estimate F_{HF}, the fraction of c to b jets must be measured as well.

6.3.1 F_b using $D^0 \to K\pi$

One way to measure the b fraction is to count the number of $D^0 \to K\pi$ events in the electron sample. F_b is obtained by the formula:

$$F_\mathrm{b} = \frac{1}{\varepsilon_{D^0}} \cdot \frac{N_{D^0}}{N} \tag{6.4}$$

where ε_{D^0} is the D^0 finding efficiency including the B decay branching ratio, and N_{D^0} is the number of observed D^0 events. From Monte Carlo, ε_{D^0} is found to be 0.01019 ± 0.0005 (statistical error only). However, the branching ratio for $b \to \ell\nu D^0 X$ was found to be 0.061 ± 0.001 in the Monte Carlo, while the PDG value is 0.0690 ± 0.0035. In order to correct for this, the efficiency is scaled by 1.131 ± 0.060 to give 0.01152 ± 0.0008, where the uncertainty includes both MC statistics and the PDG branching ratio error. A 10% systematic uncertainty is conservatively attributed to the D^0 reconstruction efficiency. $D^0 \to K\pi$ decays are selected by looking for pairs of COT tracks that satisfy the following set of cuts:

- Tracks must be within $\Delta R < 0.7$ from the electron-jet,
- The track with opposite charge to the electron is assumed to be the kaon,
- The pion and kaon candidates should have opposite signs,
- $p_T^K > 1.5$ GeV,
- $p_T^\pi > 0.5$ GeV,
- The vertex reconstructed from two candidate tracks must have a positive transverse decay length.

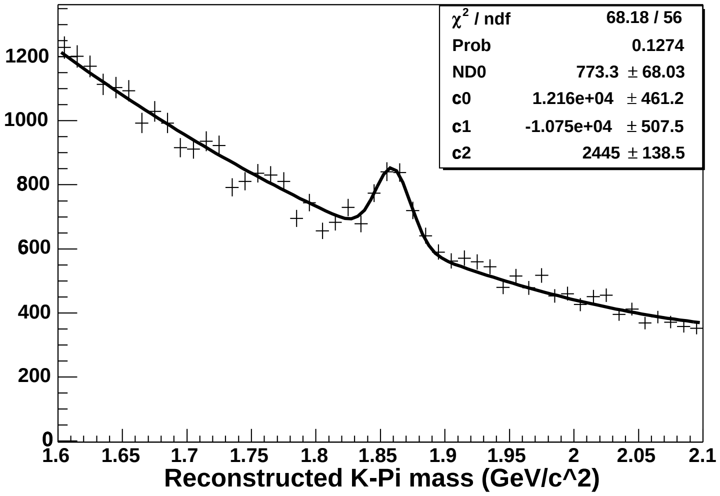


Figure 6.5: *Reconstructed $K\pi$ mass in data.*

Figures 6.5 and 6.6 show the reconstructed $K\pi$ mass distributions for data and Monte Carlo, respectively. From the fitted number of reconstructed $D^0 \to K\pi$ events, F_{b} is determined to be 0.139±0.021; the error includes data and MC statistical uncertainties as well as the branching ratio and reconstruction efficiency systematic uncertainties discussed above.

6.3.2 F_b from Cascade Muons

F_b can be obtained by counting the muons from double-semileptonic sequential decays with the following formula:

$$F_{\mathrm{b}} = \frac{1}{\varepsilon_\mu} \cdot \frac{N_{\mathrm{OS}} - N_{\mathrm{SS}}}{N} \tag{6.5}$$

where N_{OS} (N_{SS}) are the number of opposite (same) sign $\mathrm{e} - \mu$ pairs (N_{SS} is used as a background estimate), and N is the number of electron-jets. The muon detection efficiency, ε_μ, is determined from Monte Carlo and includes the branching fraction,

$$\varepsilon_\mu = \frac{s_{reco} N_\mu^{\mathrm{MC}}}{N^{\mathrm{MC}}} \tag{6.6}$$

where N_μ^{MC} and N^{MC} are the number of muons and electron-jets found in the Monte Carlo, respectively, and s_{reco} is the ratio of the muon stub-finding efficiency between data and Monte Carlo and found to be $s_{reco} = 0.962 \pm 0.007 \pm 0.021$ for CMUP and $s_{reco} = 0.9445 \pm 0.0077 \pm 0.0192$ for CMX muons [42]. The total rate of secondary muons for b electron-jets in the Monte Carlo was found to be 0.089±0.007, in

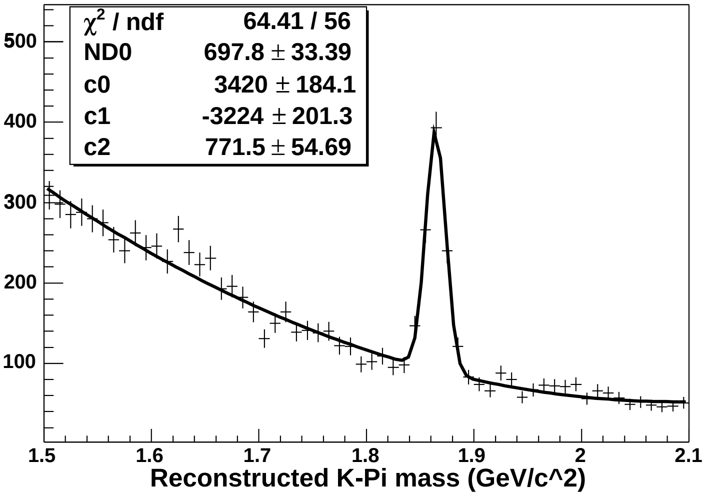


Figure 6.6: *Reconstructed $K\pi$ mass in Monte Carlo. No requirement on the away-jet has been applied, in order to increase statistics.*

class	OS	SS	F_{b}
CMUP	928	234	0.219±0.010±0.022±0.005
CMX	644	229	0.250±0.016±0.036±0.006

Table 6.1: *Numbers of* $\mathrm{e}-\mu$ *pairs and* F_{b} *for the muon method. Errors on* F_b *are statistical (data and MC, resp.), and systematic (reconstruction efficiency.)*

agreement with the value deduced from PDG branching ratios of 0.090±0.012. Both CMUP and CMX muons are used. CMUP muon requirements are $p_t > 4$ GeV/c, CMU $|\Delta x| < 10$ cm, and CMP $|\Delta x| < 15$ cm, while CMX muons are required to have $p_t > 3$ GeV/c and CMX $|\Delta x| < 10$ cm. The muon is required to be within $\Delta R < 0.4$ of the electron-jet for either type.

Table 6.1 summarizes the numbers of opposite and same sign $\mathrm{e}-\mu$ pairs for CMUP and CMX muons as well as F_{b} determined from those numbers. Combining the two types of muons gives $F_b = 0.228 \pm 0.021(\text{stat.}) \pm 0.030(\text{syst.})$, where the systematic error includes 13% from PDG branching ratio uncertainties and the reconstruction efficiency (≈2%).

6.3.3 Measurement of F_c/F_b

Fits of b, c, and light jet templates are made to the secondary vertex mass distribution of positive tags in the electron-jets in order to measure the fraction of each contribution. The fit parameters are the fraction of b jets in positively tagged

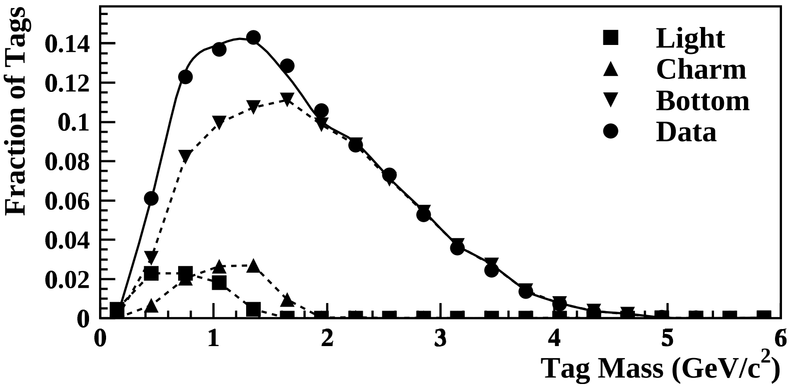


Figure 6.7: *A fit of flavor templates to the SecVtx mass distribution. The round markers are the data and the line shows the sum of the fitted templates.*

electron-jets F_b^{e+}, and the ratio F_c^{e+}/F_b^{e+}. The light-flavor contribution is constrained to $1 - F_b^{e+} - F_c^{e+}$. For the light jet template, various estimates are used, and the difference is included as a systematic uncertainty. The result of one of the fits is shown in Fig. 6.7. The F_c^{e+}/F_b^{e+} ratio is measured to be 0.118±0.017. To obtain F_c/F_b (the ratio without tagging), F_c^{e+}/F_b^{e+} must be corrected for the ratio of b to c tagging efficiency. Without any correction, the simulation gives $\epsilon_b/\epsilon_c = 5.6$. In order to obtain a more realistic estimate, one can calculate the efficiencies after dropping 12% of the selected tracks , which corresponds to the excess of selected tracks between MC and data shown in Fig. 6.13 (see Sec. 6.6); this gives $\epsilon_b/\epsilon_c = 5.2$. Since the latter should be closer to the true value in the data, it is used as the central value and the difference between 5.6 and 5.2 is taken as the uncertainty. F_c^{e+}/F_b^{e+} is multiplied by 5.2±0.4, which gives $F_c/F_b = 0.61 \pm 0.10$. This number is used to compute F_{HF} from F_b.

6.3.4 Determination of F_{HF}

In the preceding sections two values of F_b have been measured: 0.139±0.021 from $D^0 \to K\pi$ reconstruction, and 0.228±0.037 from secondary muon identification. The latter is about two standard deviations larger than the former, which could be due to an underestimation of the $e - \mu$ background. The weighted average of the two methods gives $F_b = 0.161 \pm 0.038$, where the combined uncertainty has been scaled up by a factor 2.09 according to the $\sqrt{\chi^2}$ of the two measurements, in order to reflect the disagreement between the two measurements. Scaling by one plus the F_c/F_b ratio found earlier, $1 + (0.61 \pm 0.10)$, $F_{HF} = 0.259 \pm 0.064$. Table 6.2 summarizes the different sources of uncertainty on F_{HF}.

F_b ($D^0 \to K\pi$)	
D^0 rec. eff.	10
BR	5.1
stat. (data)	8.8
stat. (MC)	4.9
Total $F_b(D^0 \to K\pi)$	15.1
F_b (muon)	
BR	13
Muon rec. eff.	2
stat.(data+MC)	9.2
Total F_b(muon)	16.1
Total F_b(comb.)	23.8
c/b	6.2
Total	24.6

Table 6.2: *Sources of uncertainty on F_b, for each method, and on F_{HF} (relative errors, in %).*

6.4 Result

The F_{HF} value derived above for data and the value $F_{HF}^{MC} = 0.8601$ for Monte Carlo, are used together with the numbers in Table 6.3, and Equ. 6.2 and 6.3 to calculate the efficiencies to tag a jet in data and Monte Carlo. The resulting values averaged over E_T are given in Table 6.4. The efficiencies as a function of the E_T of the jet are shown in Fig. 6.8. The ratio of data to Monte Carlo (scale factor) is also shown as a function of E_T. Several sources of systematic uncertainty, described in the following sections, have been considered and are summarized in Table 6.5. The $F_{HF}^{\mathrm{a}+}$ uncertainty accounts for assumptions made in the calculation of $F_{HF}^{\mathrm{a}+}$ about the tagging efficiency of heavy flavor electron-jets containing a conversion electron pair. The mistag subtraction uncertainty is due to the asymmetry in negative tags *vs.* fake positive tags. The semileptonic uncertainty allows for a possible difference in the scale factor due to the lower charged particle multiplicity of semileptonic B decays. Combining all systematic and statistical errors, a data to Monte Carlo tagging efficiency scale factor of 0.82 ± 0.06 is obtained.

6.5 Mistag Asymmetry

Negative tag subtraction is used extensively, and is based on the assumption that the mistags L_{2D} distribution is symmetric, as well as the approximation that no heavy flavor jet is negatively tagged. This section reviews the uncertainties associated with this assumption.

Quantity	Data	Monte Carlo
N	481301	32180
N_{a+}	12093	2310
N_{a-}	1561	65
N_{e+}	22585	6946
N_{e-}	1400	168
N_{a+}^{e+}	1965	656
N_{a+}^{e-}	51	8
N_{a-}^{e+}	115	10
N_{a-}^{e-}	5	0
N_c	171374	4157
N_c^{a+}	2395	145
N_c^{a-}	509	8
N_c^{e+}	1694	311
N_c^{e-}	274	13

Table 6.3: *Tag counts and conversion yields found in the inclusive electron data and Monte Carlo.*

ε(Data)	0.240 ± 0.007
ε(MC)	0.292 ± 0.010
Scale Factor	0.82 ± 0.06

Table 6.4: *Efficiency to tag a b-jet in the low p_T electron sample for data and Monte Carlo, and data/MC ratio (tagging scale factor). Errors on efficiencies are statistical only. The error on the scale factor includes both statistical and systematic uncertainties.*

F_{HF}	3.5
$F_{HF}^{\mathrm{a}+}$ method	3
Mistags	3
E_T dependence	2.5
Semileptonic	1.2
Stat. (data)	3.2
Stat. (MC)	3.6
Total	7.8

Table 6.5: *Sources of uncertainty on the SecVtx scale factor (relative errors, in %).*

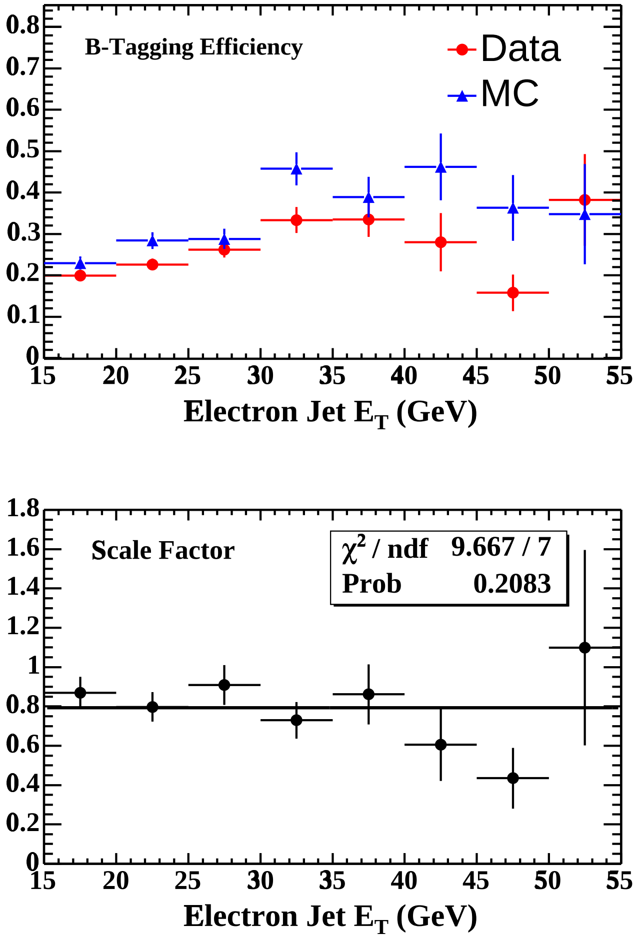


Figure 6.8: *Efficiency to tag a* b *jet as a function of jet* E_T *in the low* p_T *electron sample for data and Monte Carlo (top), and data/MC scale factor (bottom). Errors include statistical uncertainties, and systematic uncertainties due to* F_{HF}.

f_C	1.	1.3	1.5
data (%)	24.02	24.23	24.38
MC (%)	29.46	29.50	29.53

Table 6.6: *Efficiency in data and Monte Carlo for various values of the conversion jet negative tag scale factor ($f_C = 1$ corresponds to default).*

6.5.1 Conversion Jets

Jets in which the electron is identified as coming from a photon conversion are more likely to be positively tagged due to material interaction; they are also less rich in heavy flavor than generic jets. In order to measure the various contributions (heavy flavor, material interaction, “symmetric mistags” due to resolution effects), the pseudo-$c\tau$ distribution of positively tagged electron-jets in which the electron is identified as conversion is fitted with Monte Carlo templates of each contribution (see Chap. 7 for details about the method). Figure 6.9 shows the result of the fit. Material interactions enhance the positive mistag rate by a factor of 1.3 compared to the negative tag rate. To take into account this asymmetry, the efficiency is re-computed including a scale factor for negative tags in conversion electrons, and ϵ'_c becomes:

$$\epsilon'_c = \frac{N_{\mathrm{c}}^{\mathrm{e}+} - f_C . N_{\mathrm{c}}^{\mathrm{e}-}}{N_{\mathrm{e}+} - N_{\mathrm{e}-}}$$

Table 6.6 shows the changes in efficiency for different values of f_C. Changes in Monte Carlo are almost nonexistent because there are few conversion electrons in the sample; changes in the data are less than 1.5%, even with a conservative value of $f_C = 1.5$.

One could also argue that the presence of a conversion may modify the SecVtx tagging efficiency, which would affect $\epsilon'_c = \frac{N_{\mathrm{c}}^{\mathrm{e}+} - N_{\mathrm{c}}^{\mathrm{e}-}}{N_{\mathrm{e}+} - N_{\mathrm{e}-}}$. From Equ. A.12, the average tagging efficiency in conversion electron-jets can be estimated:

$$f' \cdot (\epsilon^c - \epsilon^0) \cdot (\epsilon''_B - \epsilon''_Q) = 0.006 \pm 0.001 \tag{6.7}$$

where $\epsilon''_B - \epsilon''_Q$ is the tagging efficiencies for Heavy Flavor in conversion. Re-writing Equ. A.10 in the case where both jets are tagged gives:

$$f' \cdot (\epsilon^c - \epsilon^0) = \frac{(N_{\mathrm{a}+}^{\mathrm{e}+,\mathrm{c}} - N_{\mathrm{a}+}^{\mathrm{e}-,\mathrm{c}}) - (N_{\mathrm{a}-}^{\mathrm{e}+,\mathrm{c}} - N_{\mathrm{a}-}^{\mathrm{e}-,\mathrm{c}})}{(N_{\mathrm{a}+}^{\mathrm{e}+} - N_{\mathrm{a}+}^{\mathrm{e}-}) - (N_{\mathrm{a}-}^{\mathrm{e}+} - N_{\mathrm{a}-}^{\mathrm{e}-})} - \epsilon^0 = 0.024 + -0.007 \tag{6.8}$$

Dividing 6.7 by 6.8 gives:

$$\epsilon''_B - \epsilon''_Q = 0.25 + -0.08 \tag{6.9}$$

This can be compared with the tagging efficiency computed from single-tag:

$$\epsilon'_B - \epsilon'_Q = \frac{N^{e+} - N^{e-}}{N \cdot F_{HF}} = 0.17 \pm 0.03 \tag{6.10}$$

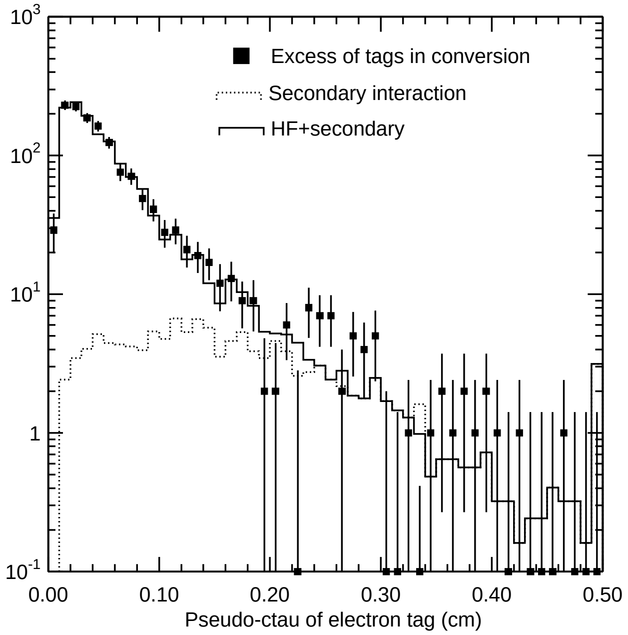


Figure 6.9: *Pseudo-$c\tau$ of positively tagged electron-jets in which the electron is identified as conversion.*

f	1.	0.	2.
data (%)	24.02	24.15	24.03
MC (%)	29.46	29.88	29.04

Table 6.7: *Efficiency in data and Monte Carlo for various values of the generic jet negative tag scale factor ($f = 1$ corresponds to default).*

The two numbers are consistent within uncertainties, but the difference is of the order of 40%. In order to take into account a potential systematic effect, ϵ'_c is scaled by ±20%; the impact on the final result ϵ_B is 3% which is added as a systematic uncertainty.

6.5.2 Generic Jets

The number of negative tags is scaled by a constant factor f, in every negative tag subtraction in the computation. For example, Equ. 6.2 becomes:

$$\varepsilon = \frac{(N^{\mathrm{e}+}_{\mathrm{a}+} - f.N^{\mathrm{e}-}_{\mathrm{a}+}) - (f.N^{\mathrm{e}+}_{\mathrm{a}-} - f^2.N^{\mathrm{e}-}_{\mathrm{a}-})}{(N_{\mathrm{a}+} - f.N_{\mathrm{a}-})} \cdot \frac{1}{F^{a+}_{HF}} \tag{6.11}$$

Expressions for F^{a+}_{HF} and ϵ'_c are modified accordingly. The actual value of f is expected to be of the order of 1.2 in the data; in the Monte Carlo, negative tag rates are much smaller, hence f may be significantly larger. Table 6.7 shows the efficiency for different values of f. $f = 1$ corresponds to the default value. $f = 0$ means that no negative tag subtraction is made. $f = 2$ corresponds to doubling the negative tag subtraction. The effect on the efficiency is very small (about 1.4%), both in data and Monte Carlo, even for such extreme variations of f, because the effect of the various negative subtractions tends to cancel out. Thus the method is quite robust with regard to negative tag subtractions. Based on the results above, a 3% systematic uncertainty is added due to mistags, and 3% due to tagging efficiency in jets with conversions.

6.6 Jet E_T Dependence

The control sample used to measure the scale factor has a relatively low jet E_T spectrum (25 GeV in average) compared to the $t\bar{t}$ sample to which it is applied. Hence the E_T dependence of the scale factor needs to be understood carefully. This E_T dependence was evaluated in three different ways.

- Low-p_T electron sample:

  A fit to the scale factor vs E_T (bottom plot of Fig. 6.8) gives a slope -0.0086±0.0066 per GeV, consistent with 0. Because of the lack of statistics at high E_T, the fit

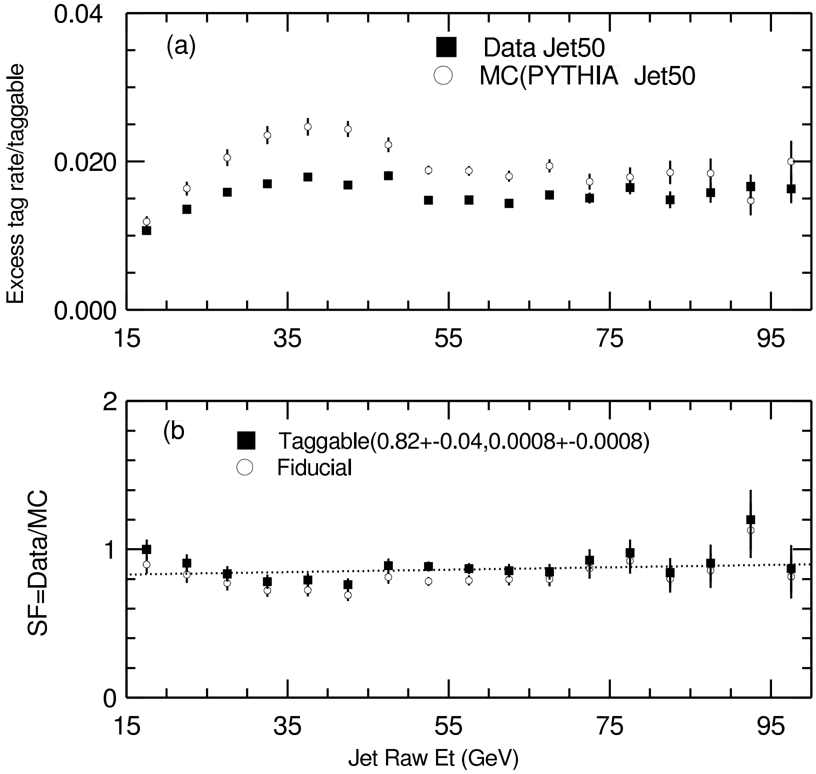


Figure 6.10: *Top: positive excess tag rate in Jet 50 data and* PYTHIA *MC vs jet E_T. Bottom: ratio of data over MC for silicon-fiducial jets, and taggable jets.*

is poor and not conclusive by itself. Besides, a fit with a flat line over the entire range of transverse energy gives a χ^2 of 9.7/7 p.d.f (20.8% probable). Hence no E_T dependence is seen in the direct measurement of the scale factor.

- Generic jet sample:

  The positive tagging rate in the Jet 50 data sample is compared to a PYTHIA [25] Monte Carlo sample (top of Fig. 6.10), after subtraction of the negative tagging rate. The strange behavior below 50 GeV is due to trigger bias in the data and gluon splitting in the Monte Carlo. Above 50 GeV, the agreement between data and Monte Carlo is rather good. This method cannot be used to measure the scale factor absolute value, because the heavy flavor content of this sample is not well known; however, the variation of heavy flavor fraction is sufficiently small over a large range of jet E_T that it can be used to estimate the E_T dependence of the scale factor. The ratio of positive excess rate in data and Monte Carlo is shown at the bottom of Fig. 6.10. A fit of this ratio gives a slope of 0.0008±0.0008 per GeV, consistent with 0.

- E_T dependence due to the number of selected tracks:

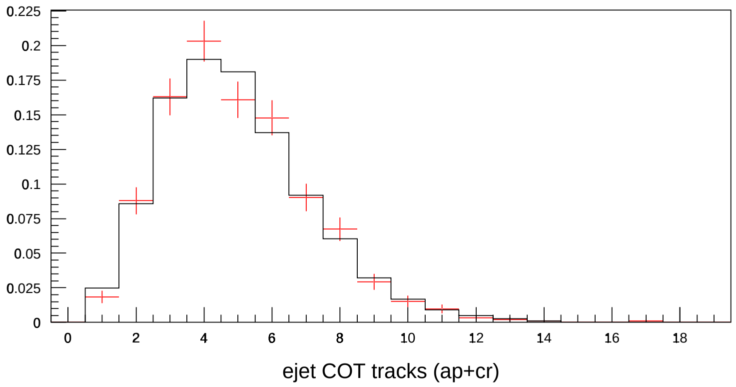


Figure 6.11: *Number of COT tracks in electron-jet for data (solid line) and Monte Carlo.*

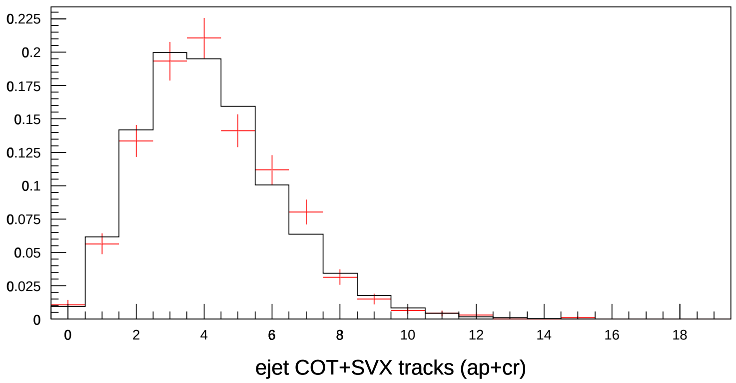


Figure 6.12: *Number of default silicon tracks in electron-jet for data (solid line) and Monte Carlo.*

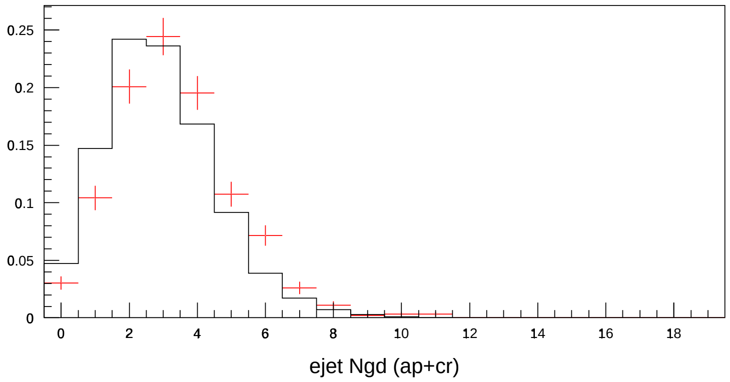


Figure 6.13: *Number of selected silicon tracks in electron-jet for data (solid line) and Monte Carlo. Jets in the Monte Carlo contain in average 12% more selected tracks than in the data.*

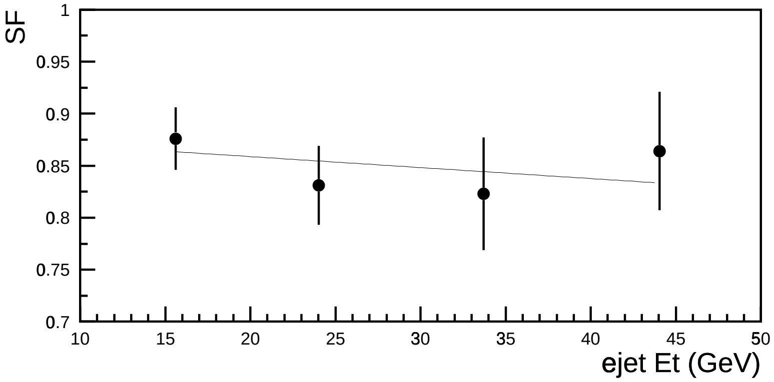


Figure 6.14: *Scale factor relative to the number of selected tracks vs electron-jet* E_T.

Figure 6.11 (resp. 6.12 and 6.13) [43] shows the number of COT tracks (resp. default silicon tracks, and selected silicon tracks) in the electron-jet, for both data and Monte Carlo, after requiring the away-jet to be tagged, and vetoing conversions, in order to have a rather pure heavy flavor jet sample. Data and Monte Carlo are in good agreement in terms of number of COT, and default silicon tracks. However, the Monte Carlo shows an excess of selected silicon tracks, which is, for the most part, responsible for the difference of efficiency in data and Monte Carlo. One can isolate the scale factor due to this effect by measuring the efficiency in Monte Carlo for each number of selected tracks individually, then integrating over the distribution of selected tracks taken from data on one hand, and from Monte Carlo on the other hand. The scale factor computed with this method is shown in Fig. 6.14 vs electron-jet E_T. This distribution is fitted and a slope of -0.0011 ± 0.0021 per GeV, consistent with 0, is measured.

Combining the three estimates, the scale factor E_T dependence is estimated to be: 0.0004 ± 0.0007 per GeV. The three individual results, and the combined one, are consistent with a flat scale factor as a function of jet E_T, from which one concludes that the value of the scale factor averaged over E_T is valid at any E_T. The systematic uncertainty associated with the scale factor E_T dependence is evaluated by varying the scale factor as a function of the jet E_T in the $t\bar{t}$ Monte Carlo sample (see Chap. 9) in two ways:

- Case 1: Scale factor $= 0.82 + 0.0011 \cdot (E_T[\mathrm{GeV}] - 25)$ (combined slope plus one sigma)
- Case 2: Scale factor $= 0.82 - 0.0003 \cdot (E_T[\mathrm{GeV}] - 25)$ (combined slope minus one sigma)

25 GeV corresponds to the average electron-jet E_T in the low-p_T electron sample, energy at which the scale factor is measured. The acceptance of $t\bar{t}$ events varies by 5% between case 1 and case 2, hence a systematic uncertainty of 2.5% is added.

6.7 Semileptonic vs Hadronic decay

The efficiency was measured on a semileptonic decay b jet sample, while $t\bar{t}$ events contain both semileptonic and hadronic b decays. The tagging scale factor may be different for semileptonic decays and hadronic decays, since hadronic decays tend to produce more tracks. In order to evaluate the associated systematic effect, the scale factor is computed by convoluting the efficiency measured in the Monte Carlo by the estimated distribution of selected tracks for semileptonic decays and inclusive decays separately. The distributions of selected tracks multiplicity are estimated with the simulation. Based on Fig. 6.13, one estimates that rejecting 12% of the selected tracks in the simulation gives a good match to the data. The ratio of efficiency before and after rejecting 12% of the selected tracks corresponds to the scale factor due to the excess of selected tracks in the simulation. For semileptonic decays, a scale factor of 0.83 is found, in excellent agreement with the exact measurement. For inclusive decays, the scale factor is 0.84. One concludes that this is not a significant effect and a 1.2% relative systematic uncertainty on the scale factor is added.

6.8 Measurement of the Mistag Rate

A light jet that ends up being positively tagged by SecVtx is referred to as a "mistag". Mistags are caused mostly by random overlap of tracks which are displaced from the primary vertex due to tracking errors, although there are contributions from K_S and Λ decays and nuclear interactions with the detector material (the beam-pipe or the inner silicon layers) as well. Since the Monte Carlo fails at describing them properly, these effects are measured directly from jet data samples without relying on the detector simulation.

Because the SecVtx algorithm is symmetric in its treatment of d_0 and L_{2D} significance, the tracking-related mistags should occur at the same rate for $L_{2D} > 0$ and $L_{2D} < 0$. Therefore, a good estimate of the positive mistag rate due to resolution effect can be obtained from the negative tag rate. Corrections due to material interactions, long-lived light flavor particles, and negatively tagged heavy flavor jets are determined using fits to the pseudo-$c\tau$ spectra of tagged vertices, described in Chap. 7. The sum of these corrections, which increase the mistag rate, is found to be $20 \pm 10\%$ of the negative tag rate.

Samples of jet triggers (see Sec. 4.7) are used to measure the rate of negative tags for taggable jets[1]. The rate is parametrized as a function of four jet variables — E_T, track multiplicity, η, and ϕ — and one event variable ΣE_T, the scalar summed E_T of all jets in the event with $E_T > 8$ GeV and $|\eta| < 2.4$. These parameterized rates are used to obtain the probability that a given jet will be negatively tagged.

The full five-dimensional tag rate matrix was determined using inclusive 20 GeV, 50 GeV, 70 GeV, and 100 GeV jet trigger samples, a total of 11.5 million events. Figure 6.15 shows the integrated negative tag rate per taggable jet as a function of jet E_T and track multiplicity in the jets for all the events in the inclusive jet sample. The negative tag rate ranges from 0.25% to 1.5% as E_T increases from 15 to 160 GeV, prior to the application of the +20% correction discussed above. Detailed cross checks were performed on the tag rate matrix to verify its self-consistency and to check predictability and sample dependence. Both the total tagging rates and the tagging rates as a function of various quantities were used to check how well the matrix predicts the observed data and to estimate systematic errors. Table 6.8 summarizes the differences between the matrix predictions and the observed tag rates in various validation samples. The four jet trigger samples described above were used, along with the ΣE_T sample. The table is divided into two sections. Each row in the table

---

[1] See Chap. 5 for a definition of a taggable jet.

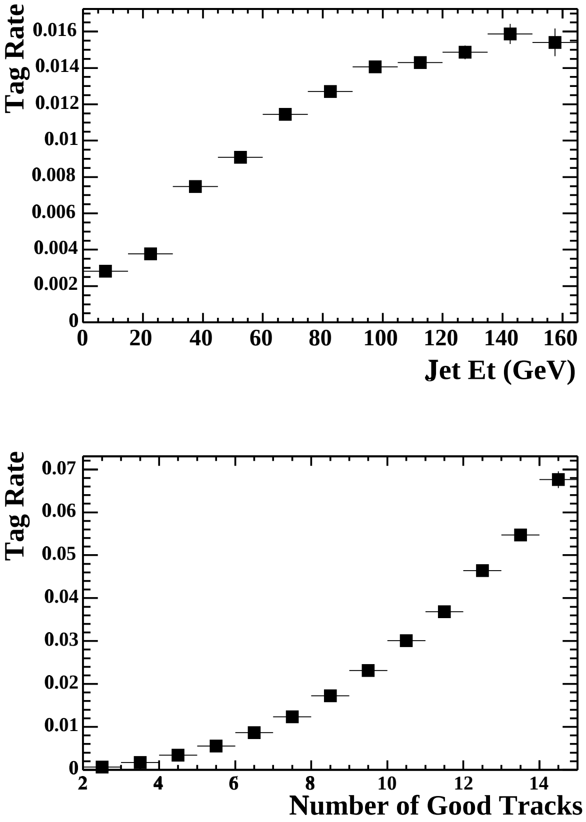


Figure 6.15: *SecVtx negative tag rate as a function of jet E_T and track multiplicity in the inclusive jet data.*

	Negative Tag Rate (%)		
	Obs.	Pred.	Obs./Pred.
Jet20-Jet50	0.728± 0.008	0.677±0.046	1.08± 0.08
Jet50-Jet70	0.958± 0.009	0.930±0.013	1.03± 0.02
Jet50-Jet100	1.219± 0.009	1.151±0.044	1.06± 0.04
Jet50-Et	0.730± 0.005	0.712±0.015	1.03± 0.02
Trigger Jet	0.565± 0.005	0.587±0.005	0.96± 0.01
Non-Trigger Jet	0.659± 0.005	0.640±0.006	1.03± 0.01
ΣE_T	0.712± 0.006	0.726±0.007	0.98± 0.01

Table 6.8: *Differences in predicted and observed negative tagging rates for various samples. The first four rows with labels of the form Sample1-Sample2 compare observed tag rates in Sample 2 to the rates predicted by a matrix made from Sample 1. The last three rows compare the observed tag rates for trigger jets, non-trigger jets, and jets in the ΣE_T sample with predictions from the standard mistag matrix derived from all four jet samples.*

Source	Uncertainty
Trigger jet bias	4%
Sample bias	7%
Statistics	1%
Total	8%

Table 6.9: *Systematic uncertainties assigned to the negative tag rate matrix.*

compares the tag rate predicted from one sample with the observed rate in a second, different sample. The differences in the tag rates of trigger jets and non-trigger jets are well predicted by the matrix. This is mostly due to the inclusion of the jet E_T, η, and ϕ into the matrix binning. The remaining residual difference is taken as a systematic error in the final result. The systematic uncertainties assigned to the tag rate matrix are summarized in Table 6.9. The various contributions are assumed to be uncorrelated and are added in quadrature to find a total systematic uncertainty of 8% on the negative tagging rates, which, combined with the 8% relative uncertainty on the 20% asymmetry correction factor, yields a total mistag rate uncertainty of 11%.

Chapter 7

Understanding the Heavy Flavor Content of the W + Jets Sample

Heavy flavor production in association with a vector boson (*e.g.* $Wb\bar{b}$, $Wc\bar{c}$, Wc) contributes significantly to the $t\bar{t}$ background in the b-tagged W + Jets sample, even though W + light flavor jet production dominates the pretag sample. Several Monte Carlo generators are capable of performing matrix element calculations for W/Z^0 + Jets, even to high jet multiplicity, but these generators use leading-order calculations. As a result, the overall normalization of these calculations has a large theoretical uncertainty. However, the relative contributions of the various diagrams are rather well-defined. Thus a matrix element Monte Carlo program is used to estimate the relative fraction of W + Heavy Flavor production, but the overall normalization of the W + Jets production is measured from the data. Combining the two results gives an estimate of the W + Heavy Flavor signal.

The new event generator ALPGEN [26] is used for this purpose. ALPGEN calculates exact matrix elements at leading order for a large set of parton level processes in QCD and electroweak interactions, taking into account all heavy quark masses, spins, and color flows. Heavy flavor fractions calculated using ALPGEN are calibrated with jet data samples.

7.1 Heavy Flavor Production in Monte Carlo

Events from the ALPGEN matrix element calculation are fed to the HERWIG parton shower program which can produce additional jets from gluon radiation. The matrix element gives a good description of the production of a few, widely separated partons, whereas parton showers are better suited to model the emission of soft collinear gluons. Following a matrix element calculation with a parton showering algorithm would seem to be an ideal combination.

One outstanding issue for such a combined approach is how to avoid double counting in the region of phase space populated by both higher order matrix elements and the parton shower. Specifically, the radiation from the parton shower in a W + n parton Monte Carlo sample can produce jets which cover the part of the phase space also described by the W + (n+1) parton Monte Carlo. Although a rigorous combination prescription has been proposed to avoid such double counting, it has not yet been fully implemented in any of the matrix element Monte Carlo programs [44, 45]. A simple procedure deals with the possible double counting by matching final state partons to reconstructed jets and rejecting events where the showering algorithm has produced a hard parton [46, 47]. Events are rejected if there are extra jets which fail to match to the light partons generated at the matrix element level or if there are missing jets. In the special case of heavy flavor partons, the strict matching criterion is relaxed because two partons may be merged into one jet due to the parton mass. Although it minimizes double counting of generated events, this procedure introduces a new type of systematic uncertainty which depends on the matching criteria and the jet definition.

The matching algorithm is applied at the stable generated particle level, before

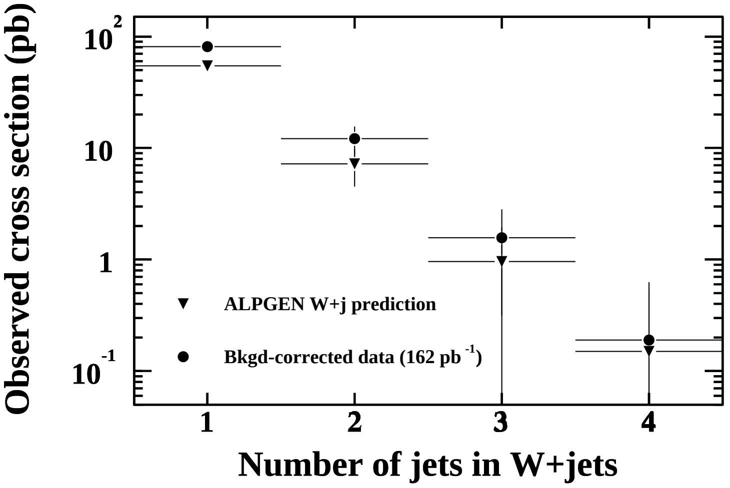


Figure 7.1: *Observed W + Jets cross section compared with the* ALPGEN *W + Jets prediction as a function of number of jets.*

any detector simulation. Stable particles after the parton shower are required to have $p_T > 0.4(0.0)$ GeV/c for charged (neutral) particles and $|\eta| < 3$. The jet clustering is a simple cone clustering scheme where the number of final jets (particles) is reduced by joining the two closest jets (particles) within a cone of radius $\Delta R = 0.4$ into one. Once all possible merging is completed, the jet four-momentum is recalculated using all of the particles inside the jet cone. A stable-particle jet is required to have $E_T > 10$ GeV and $|\eta| < 2.4$, and the matched parton must fall within a cone radius of 0.4.

The following requirements reduce event double counting after the parton shower:

- Reject events in which an extra jet failed to match any parton from the matrix element calculation,
- Ignore matching requirement for heavy flavor partons because the effect of their masses has been treated by the matrix element calculation,
- Keep only the events which pass the strict jet-light parton matching.

Fully exclusive matched events in each matrix element Monte Carlo sample are summed, weighting by the appropriate cross sections. These results are stable in terms of different matching algorithms, cone size, and jet E_T requirement. The combined sample after matching should reproduce well the overall W + Jets cross section. For illustration, the predicted W + Jets cross section, without any acceptance correction, is plotted in Fig. 7.1 with the measurement in the electron and muon channels. The QCD and diboson backgrounds (see Chap. 8) as well as the expected top quark production are subtracted for this measurement. Even though the overall normaliza-

Source Fractions	Uncertainty Wbb	 Wcc	 Wc
Matching criteria	15%	15%	10%
Q^2 scale ($2M_W^2$ to $0.5M_W^2$)	4%	4%	5%
PDF	5%	5%	10%
Jet energy scale	5%	5%	10%
ISR/FSR	10%	10%	10%
b, c masses ($\pm 0.3\,\mathrm{GeV}/c^2$)	6%	10%	
Total	21%	22%	21%

Table 7.1: *Summary of systematic uncertainties in the heavy flavor fraction determination.*

tion of the Monte Carlo does not reproduce the data very well, the jet multiplicity dependences in data and Monte Carlo are in good agreement.

7.2 Measurement of Heavy Flavor Fraction in Simulated W + Jets Events

The heavy flavor fractions for W + Jets events are defined to be the ratio of the observed W + Heavy Flavor and W + Jets cross sections. The matching algorithm operates with particle-level jets, but jets from a full calorimeter simulation provide better agreement with jets in data. A detector-level jet is required to have $E_T > 15\,\mathrm{GeV}$ and $|\eta| < 2$, and a heavy flavor jet is required to match to any b or c parton inside a cone with $\Delta R = 0.4$.

A summary of systematic uncertainties inherent in this heavy flavor fraction measurement is presented in Table 7.1. The matching uncertainty is estimated by remeasuring the heavy flavor fraction after varying the matching cone sizes from 0.4 to 0.7 and E_T from 10 to 15 GeV. Half of the difference is taken as the matching systematic uncertainty. For uncertainties due to Q^2, PDF's, and heavy quark masses, the ratio of the $Wb\bar{b}$ + 1 parton to W + 3 partons cross sections from ALPGEN is estimated, and varied by changing the Q^2 (between $2m_{\mathrm{W}}^2$ and $0.5m_{\mathrm{W}}^2$), parton distribution functions, and the heavy quark mass ($\pm 0.3\,\mathrm{GeV}$). The relative systematic uncertainties in Table 7.1 are applied to all jet multiplicity bins.

7.3 Calibration of Heavy Flavor Fraction Using Jet Data

Due to a statistically limited W + Jets sample, fractions in W + Jets events cannot be verified directly from the data. Fortunately, the jet samples are a large related

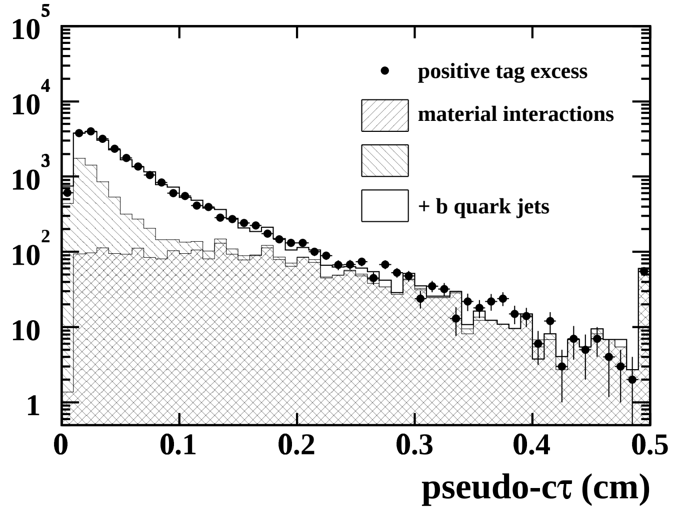


Figure 7.2: *Pseudo-$c\tau$ distribution for jet data, including fitted contributions for the different components of heavy flavor and secondary interactions in light flavor jets.*

class of events which can be used to compare the heavy flavor fractions calculated in Monte Carlo and data. Heavy flavor fractions are measured in both PYTHIA and ALPGEN+HERWIG Monte Carlo jet samples. Events are required to have 2 or 3 jets with $E_T > 15$ GeV and $|\eta| < 2.0$ and at least one jet with $E_T > 20$ GeV to satisfy trigger requirements. Events from the ALPGEN sample must also pass the matching algorithm described above. The samples of $Wb\bar{b}$ and $Wc\bar{c}$ events are further divided into two classes based on the number of visible heavy flavor jets inside the detector ($E_T > 15\,\text{GeV}$ and $|\eta| < 2.4$). Contributions to the jet sample from heavy and light partons are determined by fitting the pseudo-$c\tau$ distribution for tagged jets, thereby discriminating between jets from b, c, and light partons or gluons on a statistical basis. Pseudo-$c\tau$ is defined as $L_{2D} \times M_{\text{vtx}}/p_T^{\text{vtx}}$, where M_{vtx} is the invariant mass of all tracks in the secondary vertex and p_T^{vtx} is the transverse momentum of the secondary vertex 4-vector. Even though the L_{2D} distribution is similar for b and c quarks, the pseudo-$c\tau$ is very different for the two flavors. The fit is made more robust by subtracting the contribution from negative SecVtx tags and fitting the difference only, as shown in Fig. 7.2. Template distributions of the pseudo-$c\tau$ for b and c jets are derived by matching jets to partons in Monte Carlo, and a separate template is created for secondary interactions in light quark jets (including material interactions and long-lived Λ and K_S particles). One might expect the L_{2D} distribution of secondary vertices in light flavor jets and from tracking combinatorics to be symmetric about 0, and one might estimate the number of fake positive tags from light flavor by counting the number of negative tags. Unfortunately, secondary vertices from material interactions or long-lived light flavor particles are more likely to have positive decay lengths than negative decay lengths, and there are some real heavy

E_T (GeV)	$E_T < 25$	$25 < E_T < 35$	$35 < E_T < 45$	$E_T > 45$	All
Taggable	858,643	415,373	128,994	77,632	1,480,642
Pos. - Neg.	12,208	7131	2511	1596	23,446
Negative	3283	1999	803	697	6782
Fitted b's	7937 ± 483	4412 ± 312	1609 ± 131	843 ± 102	$15,147 \pm 507$
Fitted c's	3040 ± 427	1858 ± 276	520 ± 110	407 ± 93	5589 ± 451
Secondary	1284 ± 142	900 ± 102	379 ± 50	324 ± 39	2836 ± 171
ΔN	482 ± 224	431 ± 144	230 ± 59	227 ± 44	1336 ± 365
$\Delta N/N(\%)$	15 ± 7	22 ± 7	29 ± 7	32 ± 7	20 ± 5
b's/Jets (%)	0.92 ± 0.08	1.06 ± 0.10	1.25 ± 0.12	1.09 ± 0.14	1.02 ± 0.06
c's/Jets (%)	0.35 ± 0.06	0.45 ± 0.08	0.40 ± 0.10	0.52 ± 0.13	0.38 ± 0.05

Table 7.2: *Fitted contributions from b, c and secondary interactions / long-lived light flavor particles in data events. The errors on the b and c fraction include the template systematic uncertainty. The ratio $\Delta N/N$ estimates the excess of positive over negative tags in data events, due to secondary interactions and long-lived light flavor particles.*

flavor jets with negative decay lengths. The heavy flavor contribution with negative decay lengths is first estimated from Monte Carlo, and then scaled by a factor of 1.6±0.3 to account for a larger overall observed negative tag contribution in data than in Monte Carlo. The net excess of secondary interactions on the positive side, defined as ΔN, is computed from the secondary contribution fit results, after subtracting the heavy flavor contributions on the negative side. The average correction factor needed to scale the total number of negative tags to obtain the correct number of fake positive tags is 1.2 ± 0.1. The heavy flavor fraction as a function of jet E_T is stable, as shown in Table 7.2, where an uncertainty of 5% (10%) for the b (c) fraction is included due to template uncertainties. These results include the effect of the efficiency scale factor between data and simulation. Measured heavy flavor fractions from the data are consistently 50% higher than the ALPGEN prediction, for both b and c jets, although the PYTHIA calculation seems to match the data more closely. On average the data/ALPGEN ratio is 1.5 ± 0.4, where the uncertainty is dominated by the systematic uncertainties associated with the ALPGEN heavy flavor calculations (Table 7.1).

From these fits alone, it is not clear if the discrepancy is consistent for all production diagrams. Because jets with gluon splitting have a small opening angle, the distribution of $\Delta\phi$ between the two closest jets in an event highlights the contribution from gluon splitting. A sample of events with two tagged jets is selected from the three jet sample and compared to Monte Carlo. The mistag contribution is removed from the double-tagged samples by subtracting events with one or more negative tag. The good agreement, shown in Fig. 7.3, indicates that the gluon splitting contribution relative to other production mechanisms is well-modeled. Another sample with gluon

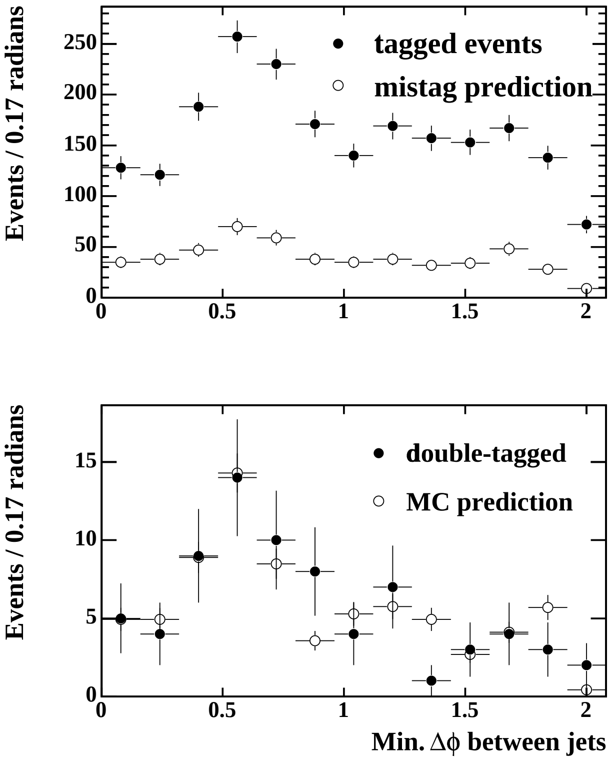


Figure 7.3: *Distribution of closest jets in $\Delta\phi$ for tagged 3-jet data events with fake tag prediction (top) and for double-tagged events in which the tagged jets are also the two closest jets (bottom).*

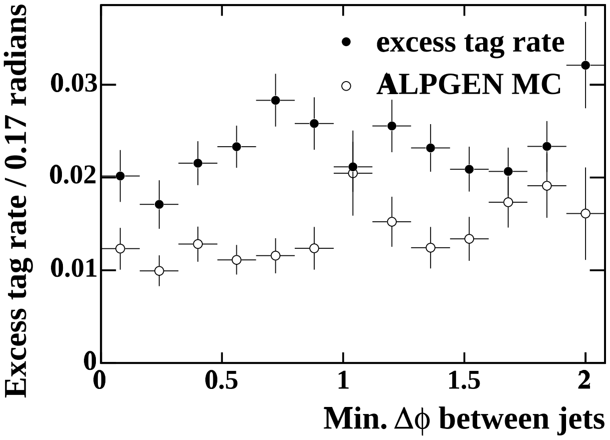


Figure 7.4: *Positive tag excess rate in data and* ALPGEN *Monte Carlo as a function of* $\Delta\phi$*.*

splitting contributions, this time of single-tagged three jet events, can be used to check the dependence of the data/ALPGEN normalization factor. When the excess tag rate, interpreted as the heavy flavor fraction, is plotted as a function of the minimum $\Delta\phi$ between jets (Fig. 7.4), the fractions are flat as a function of $\Delta\phi$ even though the heavy flavor fractions in data are 1.5 times the heavy flavor fractions in Monte Carlo. This consistency disfavors the hypothesis of missing or under-represented heavy flavor production diagrams.

The measured ratio of 1.5 ± 0.4 between the heavy flavor fractions in the ALPGEN/HERWIG samples and the data is not inconsistent with other recent studies, which indicate that a K-factor may be necessary to account for higher-order effects [48]. Based on this calibration with the jet data sample, the expected $Wb\bar{b}$ and $Wc\bar{c}$ contributions derived from ALPGEN are scaled by a factor of 1.5 ± 0.4. Since the Wc contribution is produced through a different diagram, that contribution is not rescaled. Table 7.3 summarizes the one and two b (c) fractions as a function of jet multiplicity, as well as the corresponding SecVtx tagging efficiencies, where the efficiency scale factor, as calculated in Chap. 6, has been included. The 1B (1C) fractions are for events with exactly one jet matched to a b (c) parton, and the 2B (2C) fractions are for events with exactly two jets matched to b (c) partons. These values are used in Chap. 8 to predict the background contribution from W + Heavy Flavor production.

Jet Mult.	1 jet	2 jets	3 jets		$\geq$4 jets	
H_T (GeV)			$H_T > 0$	$H_T > 200$	$H_T > 0$	$H_T > 200$
W + HF fractions (%)						
1B	1.0 ± 0.3	1.4 ± 0.4	2.0 ± 0.5	2.4 ± 0.6	2.2 ± 0.6	2.2 ± 0.6
2B		1.4 ± 0.4	2.0 ± 0.5	2.3 ± 0.6	2.6 ± 0.7	2.6 ± 0.7
1C	1.6 ± 0.4	2.4 ± 0.6	3.4 ± 0.9	3.8 ± 1.0	3.6 ± 1.0	3.5 ± 1.0
2C		1.8 ± 0.5	2.7 ± 0.7	2.9 ± 0.8	3.7 ± 1.0	3.7 ± 1.0
Wc	4.3 ± 0.9	6.0 ± 1.3	6.3 ± 1.3	6.0 ± 1.3	6.1 ± 1.3	5.9 ± 1.3
SecVtx tagging efficiencies (%)						
1B($\geq$ 1tag)	26.8 ± 2.0	27.8 ± 2.2	29.3 ± 2.5	30.9 ± 2.9	24.2 ± 3.3	27.4 ± 3.8
2B($\geq$ 1tag)		48.6 ± 3.2	50.0 ± 3.8	52.6 ± 4.5	50.3 ± 4.9	50.0 ± 5.1
2B($\geq$ 2tags)		9.1 ± 1.4	9.5 ± 1.5	10.4 ± 1.6	8.1 ± 1.4	8.6 ± 1.5
1C($\geq$ 1tag)	6.2 ± 0.9	6.7 ± 1.0	6.1 ± 1.1	6.6 ± 1.3	7.7 ± 1.9	7.5 ± 2.0
2C($\geq$ 1tag)		12.3 ± 1.9	11.6 ± 2.0	12.6 ± 2.5	10.1 ± 2.3	9.6 ± 2.4
2C($\geq$ 2tags)		0.5 ± 0.2	0.4 ± 0.1	0.5 ± 0.2	0.8 ± 0.4	0.9 ± 0.4
Wc ($\geq$ 1tag)	5.8 ± 0.9	6.1 ± 0.9	7.1 ± 1.2	7.6 ± 1.5	5.6 ± 1.6	5.8 ± 1.8

Table 7.3: *Ratio of W + Heavy Flavor production to total W + Jets production, for different jet multiplicities. The heavy flavor ratios include the correction factor* 1.5 ± 0.4 *as measured from the data, and the SecVtx event tagging efficiencies include the scale factor described in Chap. 6.*

Chapter 8

Backgrounds

Several non-$t\bar{t}$ processes are present in the W + Jets sample, even with an event selection optimized to isolate the $t\bar{t}$ signal (Chap. 9). Here is a list of the backgrounds whose contribution needs to be evaluated in order to measure the $t\bar{t}$ production cross section:

- W + Jets: the production of a W boson in association with multiple light jets is a background that is greatly reduced by the use of b jet identification; however, events where a light jet is mistagged are an important source of background.
- W + Heavy Flavor: the production of a W boson in association with heavy flavor jets is an irreducible source of background.
- QCD: jet events can fall into the W + Jets event selection by faking the W leptonic decay signature.
- Z^0 + Jets and Z^0 + Heavy Flavor: if one of the legs of the Z^0 leptonic decay is undetected, such events can fall into the $t\bar{t}$ event selection. Although small, this background is evaluated together with the W backgrounds.
- Other low rate electroweak processes with heavy flavor, such as diboson and single top production also contribute to the background.

The estimation of each background is described in this chapter.

8.1 QCD Background

QCD jet events can fake the W signature in several ways: the lepton can be faked by a hadron, or can come from the semileptonic decay of a heavy flavor hadron. Similarly, the large $\not\!\!E_T$ can be due to a detector mismeasurement or the neutrino of a semileptonic decay. Finally, the b-tag can be due to an actual heavy flavor jet or a mistag. Because such mismeasurements are difficult to reproduce in the simulation, and because the heavy flavor content of such events is not well predicted, one must rely on the data to evaluate this source of background. Since the W and Z^0 backgrounds are calculated by normalizing the inclusive cross section to the number of W + Jets events before tagging, it is necessary to understand the QCD contamination in the pretag sample as well as in the tagged sample.

8.1.1 Pretag QCD Background

Generally, QCD events produce non-isolated leptons and low $\not\!\!E_T$, so that sideband regions with large isolation I and low $\not\!\!E_T$ in the lepton sample contain most of the QCD events. These sideband regions are used to extrapolate the expected QCD contribution in the signal region (low isolation and large $\not\!\!E_T$). The lepton data sample is divided into the following four regions:

1. Region A: $I > 0.2$ and $\not\!\!E_T < 15$ GeV
2. Region B: $I < 0.1$ and $\not\!\!E_T < 15$ GeV
3. Region C: $I > 0.2$ and $\not\!\!E_T > 20$ GeV
4. Region D (W signal region): $I < 0.1$ and $\not\!\!E_T > 20$ GeV

The main assumption of the method is to consider that isolation and $\not\!\!E_T$ are uncorrelated for QCD events, so that the ratio of QCD events at low and high isolation in the low $\not\!\!E_T$ region is the same as in the high $\not\!\!E_T$ region. The number of QCD events in the signal region is given by:

$$N_D^{\mathrm{QCD}} = \frac{N_B \cdot N_C}{N_A} \tag{8.1}$$

The QCD background is calculated separately for the electron and muon channels, and for each jet multiplicity. The contribution of true W + Jets and $t\bar{t}$ events in the sideband regions is estimated using Monte Carlo samples to determine the ratio of W and $t\bar{t}$ in the signal and sideband regions, and are normalized to the observed number of events in the pretag signal region and to the theoretical $t\bar{t}$ production cross section, respectively; the correction amounts to 5-30% depending on the lepton type and jet multiplicity. The number of events found in each region in the data is shown in Tables 8.1 and 8.2 for electrons and muons, respectively. The expected number of events due to W + Jets and $t\bar{t}$, that needs to be subtracted for the QCD background estimate, are also shown. Table 8.3 gives the predicted fraction of QCD events in the pretag signal region. The main source of systematic uncertainty is the underlying assumption that the lepton isolation and $\not\!\!E_T$ are uncorrelated. The region $0.1 < I < 0.2$ (dominated by QCD events) is used to test the prediction based on $I > 0.2$ regions; a 25% systematic uncertainty is added to the pretag QCD background estimate. Because the H_T distribution of QCD events is very similar to the W + Jets one (see Chap. 9), and because the $t\bar{t}$ content of the pretag sample is small, the additional requirement $H_T > 200$ GeV has no significant effect on the pretag QCD fraction.

8.1.2 Tagged QCD Background

Two methods are employed, and combined, to estimate the number of QCD events in the tagged W + Jets sample.

One method evaluates the tagging rate of QCD events in the signal region, and applies this rate to the pretag QCD estimate. Since the heavy flavor content of the QCD events is not well known (in particular, the presence a real lepton in some — but not all — QCD events enhances the heavy flavor content), the tagging rate in region B is used as an estimate of the tagging rate in the signal region D. In order to reduce the statistical uncertainty, the tagging rate *per taggable jet* is measured

Jet multiplicity	1 jet	2 jets	≥ 3 jets
Data			
A (Low $\not\!\!E_T$, High I)	43933	6200	1023
B (Low $\not\!\!E_T$, Low I)	33274	3067	420
C (High $\not\!\!E_T$, High I)	1692	522	159
D (High $\not\!\!E_T$, Low I)	8798	1441	309
Corrections			
A (Low $\not\!\!E_T$, High I)	1.84	0.572	0.594
B (Low $\not\!\!E_T$, Low I)	635	82.0	15.2
C (High $\not\!\!E_T$, High I)	63.6	14.7	6.96

Table 8.1: *Number of electron events in data, and corrections, for pretag events.*

Jet multiplicity	1 jet	2 jets	≥ 3 jets
Data			
A (Low $\not\!\!E_T$, High I)	17125	2598	365
B (Low $\not\!\!E_T$, Low I)	6160	563	81
C (High $\not\!\!E_T$, High I)	737	238	87
D (High $\not\!\!E_T$, Low I)	6554	1010	182
Corrections			
A (Low $\not\!\!E_T$, High I)	1.73	0.54075	0.34058
B (Low $\not\!\!E_T$, Low I)	603	77.6	13.8
C (High $\not\!\!E_T$, High I)	60.2	13.6	6.51

Table 8.2: *Number of muons events in data, and corrections, for pretag events.*

			$H_T > 0$	$H_T > 200$ GeV
Jet multiplicity	1 jet	2 jets	≥ 3 jets	≥ 3 jets
Electrons				
Pretag QCD Frac.	0.14±0.04	0.17±0.04	0.20±0.05	
B Rate	0.0227±0.0011	0.0256±0.0025		
Taggable jets in D	5092	1705	654	426
Tagged Estimate	15.9±4.5	7.4±2.2	3.3±1.0	2.1±0.7
Muons				
Pretag QCD Frac.	0.034±0.009	0.04±0.01	0.08±0.03	
B Rate	0.0295±0.0030	0.0275±0.006		
Taggable jets in D	3746	1180	387	206
Tagged Estimate	3.7±1.1	1.3±0.5	0.87±0.35	0.46±0.19

Table 8.3: *Pretag QCD fraction and tagged QCD background estimates (using the tagging rate in region B).*

in events with two or more jets and then applied to the number of taggable jets in the signal region (including the $H_T > 200$ GeV requirement), and multiplied by the pretag QCD background fraction evaluated in the previous section. Results from this method are shown in Table 8.3. Figure 8.1 shows the tagging rate per taggable jet at high and low $\not\!\!E_T$ vs electron isolation, in the W + 1 jet sample. For $I < 0.1$, the sample is dominated by W + Jets at high $\not\!\!E_T$, hence a lower tagging rate. For $0.1 < I < 0.3$, the agreement between high and low $\not\!\!E_T$ regions rates is good; above 0.3, one could argue that the high $\not\!\!E_T$ region has a larger tagging rate; however the statistics are poor: a 12% systematic uncertainty is assigned for the tagging rate in region B due to a possible bias relative to $\not\!\!E_T$.

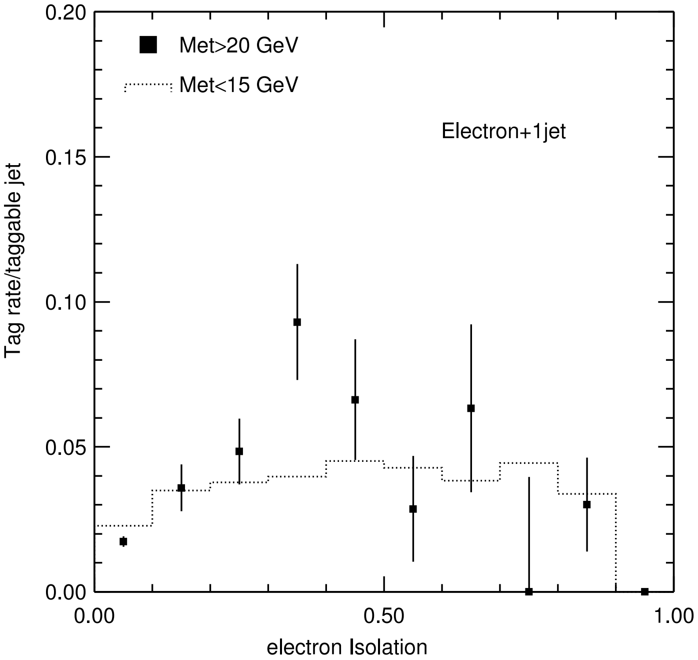


Figure 8.1: *Tagging rate (per taggable jet) at high and low $\not\!\!E_T$ vs electron isolation, for events with 1 jet only.*

Another method of estimating the QCD background uses the isolation distribution of leptons in the *tagged* events from the low $\not\!\!E_T$ region as an unbiased isolation distribution of the QCD events. This isolation distribution can then be used to estimate the number of QCD events at high $\not\!\!E_T$. With the limited statistics in the present data sample, this method is applied with two large isolation bins ($I < 0.1$ and $I > 0.1$), effectively applying Equ. 8.1 to the *tagged* event sample, where the upper isolation cut is lowered from 0.2 to 0.1 in order to gain statistics in the control regions. For the same reason, the B/A ratio is measured in the ≥ 2 jet sample. Figures 8.2 and 8.3 show the electron and muon isolation, respectively, before and after tagging, in the regions of low and high $\not\!\!E_T$. Again, the expected contributions from W + Jets and $t\bar{t}$ processes are subtracted from each sideband region. The event count and corrections due to W + Jets and $t\bar{t}$ contamination are shown in Tables 8.4 and 8.5 for electrons and muons, respectively. Results are shown in Table 8.6. A possible systematic effect associated with the tagged technique is studied by varying the upper isolation cut from 0.1 to 0.5. The resulting dependence of the tagged QCD background on the isolation cut is shown in Fig. 8.4. The dependence on the isolation cut seems small and is well covered by the statistical error from the limited data sample. A 12% systematic uncertainty is conservatively added to this method.

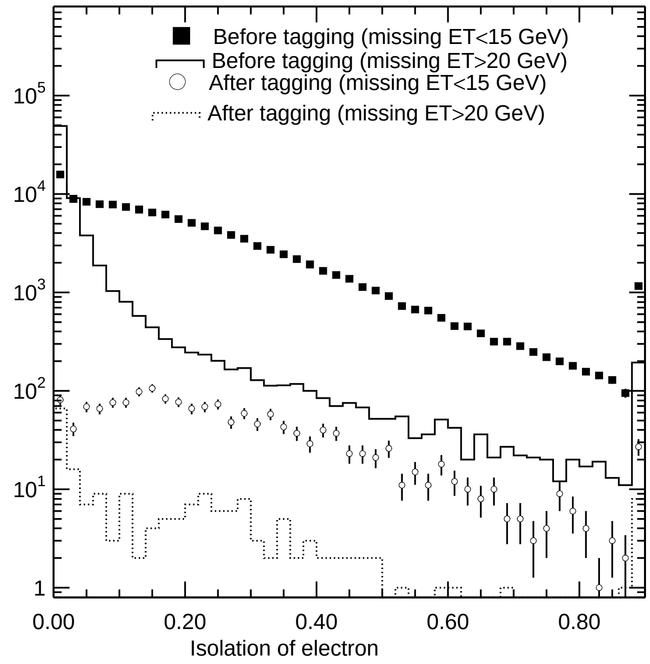


Figure 8.2: *Electron isolation before and after tagging in the high and low* $\not\!\!E_T$ *regions.*

Both background estimates contribute to the weighted average shown in Table 8.7. Figures 8.5 and 8.6 show the tagging rate per taggable jet in regions A, B, C, and D for electrons and muons, for events with 2 or more jets, as a function of H_T. In the signal region (D), the tagging rate increases significantly with H_T, as expected, due to the larger fraction of $t\bar{t}$ events at large H_T. In regions A, B, and C, the tagging rates are independent of H_T within statistical uncertainties, although a slight increase at large H_T is visible, due to the small $t\bar{t}$ contamination in ≥ 3 jet events; computing the tagging rate in events with exactly 2 jets (effectively vetoing $t\bar{t}$events) removes this H_T dependence and gives consistent results. Thus no systematic uncertainty relative to the H_T cut is added. All systematics are treated as uncorrelated. However, for a given source of uncertainty, a 100% correlation between electrons and muons is assumed.

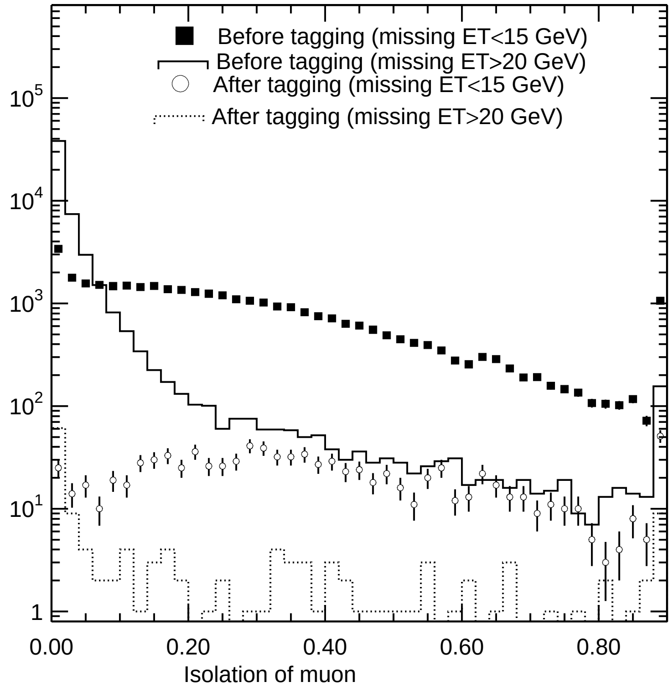


Figure 8.3: *Muon isolation before and after tagging in the high and low $\not\!\!E_T$ regions.*

8.2 Mistags (W + Light Jets)

Mistag background events are W + Jets events where the tagged jet does not result from the decay of a heavy hadron. As described in Sec. 6.8, the mistag rate per jet is parameterized as a function of the number of tracks, the raw jet E_T, the η and ϕ of the jet, and the sum of the E_T for all jets in the event with $E_T > 10$ GeV and $|\eta| < 2.4$. To estimate the size of the mistag background, each jet is weighted with its mistag rate in the pretag sample. The sum of the weights over all jets in the sample is then scaled down by the fraction of pretag events which are due to QCD background, as in Sec. 8.1.1, since these have already been counted in the procedure of Sec. 8.1.2. The low mistag rate per jet means that a negligible number of events have more than one mistagged jet; therefore, the number of mistagged jets is a good approximation of the number of events with at least one mistagged jet. This method is tested by comparing the negative SecVtx tags observed and predicted for the pretag sample as a function of the jet E_T, plotted in Figure 8.7. There is fair agreement in the shape and normalization of the prediction. For the estimate of the number of fake positive tags, the mistag correction factor of 1.2 ± 0.1 described in Sec. 7.3 is applied to account for additional mistags of light jets due to material interactions or long lived light quark hadrons. The results for the mistag estimate are shown in

Jet multiplicity	1 jet	2 jets	≥ 3 jets	
			$H_T > 0$ GeV	$H_T > 200$ GeV
Data				
A (Low $\not\!\!E_T$, High I)	1478	378	86	15
B (Low $\not\!\!E_T$, Low I)	405	83	24	8
C (High $\not\!\!E_T$, High I)	80	46	26	15
D (High $\not\!\!E_T$, Low I)	88	42	35	28
Corrections				
A (Low $\not\!\!E_T$, High I)	0.110	0.0523	0.357	
B (Low $\not\!\!E_T$, Low I)	8.49	3.14	2.45	
C (High $\not\!\!E_T$, High I)	4.43	2.36	3.49	

Table 8.4: *Number of electron events in data, and corrections, for tagged events. Regions A and C are defined with* $I > 0.1$.

Jet multiplicity	1 jet	2 jets	≥ 3 jets	
			$H_T > 0$ GeV	$H_T > 200$ GeV
Data				
A (Low $\not\!\!E_T$, High I)	965	220	65	10
B (Low $\not\!\!E_T$, Low I)	98	14	7	3
C (High $\not\!\!E_T$, High I)	51	30	23	20
D (High $\not\!\!E_T$, Low I)	73	32	22	20
Corrections				
A (Low $\not\!\!E_T$, High I)	0.09985	0.05439	0.24542	
B (Low $\not\!\!E_T$, Low I)	8.5025	3.046	2.002	
C (High $\not\!\!E_T$, High I)	4.3288	2.1249	2.868	

Table 8.5: *Number of muon events in data, and corrections, for tagged events. Regions A and C are defined with* $I > 0.1$.

Table 10.2. The error includes statistical uncertainties from the pretag sample, and takes into account the small effect of correlation between mistag weights that come from the same bin. In addition, there is a 11% systematic uncertainty for the sample dependence of the mistag rate parametrization and the mistag correction factor of 1.2 for material interactions.

8.3 W + Heavy Flavor Backgrounds

The production of W bosons associated with heavy flavor in the processes $Wb\bar{b}$, $Wc\bar{c}$, and Wc is a significant part of the background for the tagged sample. The techniques described in Chap. 7 are used to estimate the fraction of the inclusive W + Jets events which have $Wb\bar{b}$, $Wc\bar{c}$, and Wc. The number of $Wb\bar{b}$, $Wc\bar{c}$, and Wc events is given by multiplying the heavy flavor fractions by the pretag event

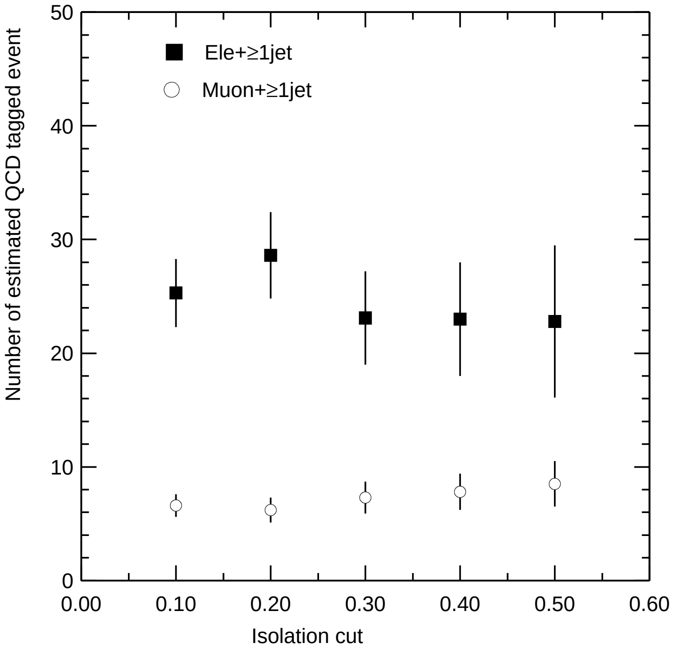


Figure 8.4: *Dependence of the isolation cut for the electron and muon channels on the tagged QCD background estimate.*

			$H_T > 0$	$H_T > 200$ GeV
Jet Mult.	1 jet	2 jets	≥ 3 jets	≥ 3 jets
Electrons				
B/A (tagged)	0.268±0.015	0.219±0.024		
Tagged Estimate	20.3±3.6	9.5±2.1	4.9±1.3	2.6±0.8
Muons				
B/A (tagged)	0.093±0.010	0.056±0.014		
Tagged Estimate	4.3±0.9	1.6±0.5	1.1±0.4	1.0±0.4

Table 8.6: *QCD background estimates (using the tagged event method).*

count, after subtracting the QCD background. Estimates of the tagged background are then obtained by multiplying the tagging efficiencies summarized in Table 7.3.

The pretag W + jets sample includes some contribution from mis-identified Z $\rightarrow \mu^+\mu^-$ events. The heavy flavor fraction for that process is twice as large as for the W events. The extra contribution of heavy flavor from Z^0 events is described in Section 10.1 and given in Table 10.1.

			$H_T > 0$		$H_T > 200$ GeV	
Jet multiplicity	1 jet	2 jets	3 jets	≥ 4 jets	3 jets	≥ 4 jets
Total Estimate	22.9±3.3	10.1±1.7	3.4±0.7	1.4±0.4	1.7±0.4	1.2±0.3

Table 8.7: *Tagged QCD background estimates (combined methods, electrons and muons).*

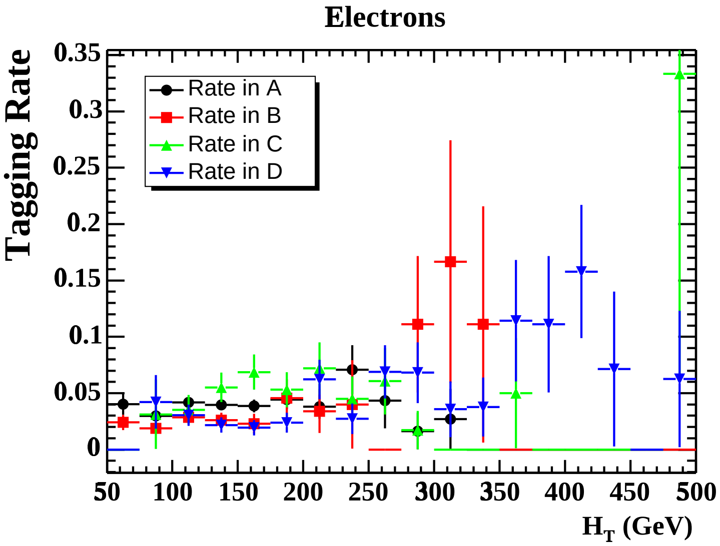


Figure 8.5: *Tagging rate as a function of H_T for electron events with 2 or more jets in regions A, B, C, and D.*

8.4 Other Backgrounds

A number of backgrounds are too small to be measured from data, thus the simulation is used to predict their contribution to the sample. The diboson production processes WW, WZ, and ZZ, in association with jets, can mimic the $t\bar{t}$ signal when one boson decays leptonically and the other decays to a b or c quark jet. The process $Z \rightarrow \tau^+\tau^-$, in association with jets, can mimic the signal when one τ decays leptonically and the other hadronically. Finally, top quarks are also expected to be produced singly with a $t\bar{b}$ final state through s-channel $q\bar{q}$ annihilation, and t-channel W-gluon fusion processes (single top production).

Monte Carlo samples are used to measure the acceptance and tagging efficiency. The Monte Carlo acceptance is corrected for the lepton identification and trigger efficiencies in the same way as for the $t\bar{t}$ acceptance as described in Chap. 9. The tagging efficiency is scaled by the MC/data tagging scale factor (the uncertainty for tagging charm jets as in $W \rightarrow c\bar{s}$ is doubled). The normalization is based on the measured integrated luminosity and the following theoretical cross sections:

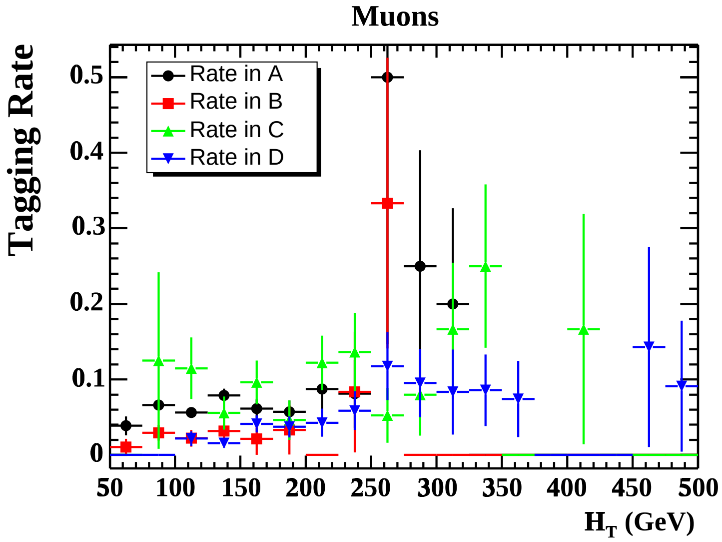


Figure 8.6: *Tagging rate as a function of H_T, for muon events with 2 or more jets in regions A, B, C, and D.*

$\sigma(\text{single top}) = 2.86 \pm 0.09\,\text{pb}$, $\sigma(WW) = 13.25 \pm 0.25\,\text{pb}$, $\sigma(WZ) = 3.96 \pm 0.06\,\text{pb}$, and $\sigma(ZZ) = 1.58 \pm 0.02\,\text{pb}$ [49, 50].

8.5 Background Summary

A complete summary of all of the background contributions is given in Table 10.2. Figure 8.8 shows the contribution of the different backgrounds for each jet bin compared to the number of data events satisfying all of the selection criteria and having at least one positively tagged jet. Good agreement between background and data is found in the one and two jet bins, validating the background estimation. The excess of tags in the three and four jet bins is attributed to the $t\bar{t}$ signal.

Both the mistag and W + Heavy Flavor backgrounds are normalized with the pretag sample, which contains a non-negligible fraction of $t\bar{t}$ events that need to be taken into account: corrections due to $t\bar{t}$ contributions in the pretag sample are discussed in Chap. 10. After this correction, 13.5 ± 1.8 events due to background are expected in the signal region (≥3 jets, $H_T > 200$ GeV).

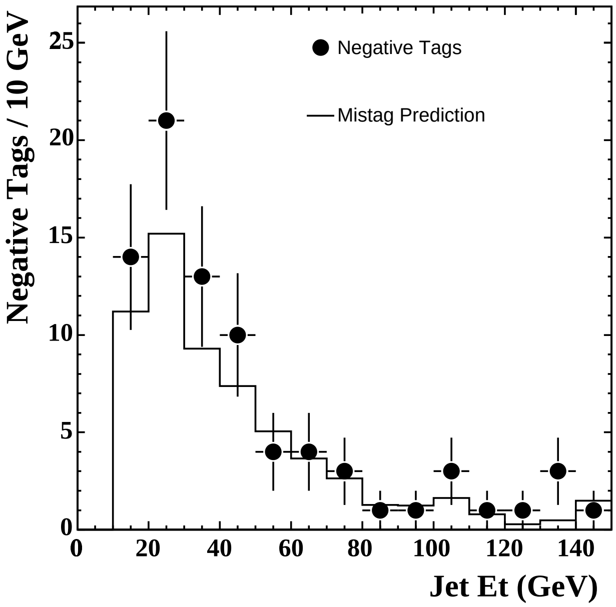


Figure 8.7: *Comparison of observed and predicted negative SecVtx tags versus jet* E_T *in the lepton + jets sample.*

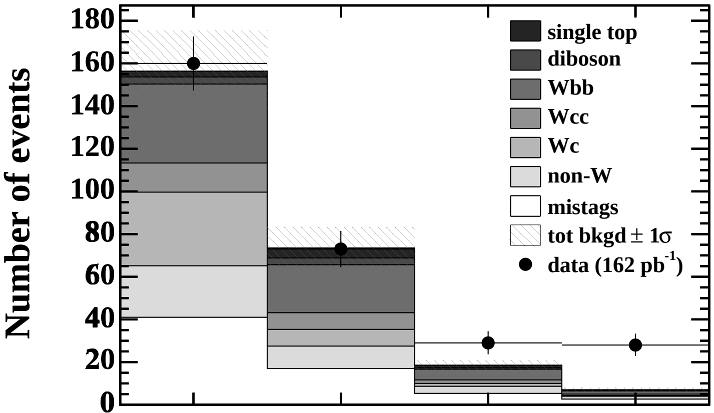


Figure 8.8: *Number of events passing selection criteria with at least one tagged jet and the background prediction.*

Chapter 9

Event Selection Optimization and Acceptance

The event selection described in Sec. 4.7 yields a clean sample of $t\bar{t}$ lepton+jets events: the signal over background ratio is expected to be of the order of 3:1. The following section discusses the event selection optimization using the H_T variable. The rest of the chapter describes the evaluation of the $t\bar{t}$ acceptance of this event selection.

9.1 Event Selection Optimization with the H_T Variable

Several ways of optimizing the event selection were studied in order to maximize the significance of the cross section measurement, and the inclusive event variable H_T was found to have the greatest power to discriminate $t\bar{t}$ signal from background events. The following section discusses the event selection optimization using the H_T variable.

The event quantity H_T is defined as the scalar sum of the transverse energy of all the kinematic objects in the event (transverse momentum for muons), including all jets with $E_T > 8$ GeV and $|\eta| < 2.5$:

$$H_T = \Sigma_{\text{all jets}} E_T + \not{E}_T + E_T^{\text{electron}} \text{ or } p_T^{\text{muon}}$$

Because of the large mass of the top quark, H_T, which is representative of the hard scatter of the event, tends to be significantly larger for $t\bar{t}$ events than for the backgrounds. Figure 9.1 shows the distribution of the H_T variable after all selection cuts have been applied, including b-tagging, for $t\bar{t}$ Monte Carlo, and for the main backgrounds: W + heavy flavor, QCD, and mistags. The W + heavy flavor distributions are taken from ALPGEN Monte Carlo, but all other background shapes are estimated from data.

The ALPGEN Monte Carlo generator is used to estimate the shape of the distribution for the W + heavy flavor background. The QCD background shape is evaluated by selecting pretag events where the lepton is not isolated (isolation $I > 0.2$), while all other kinematic cuts remain unchanged. This subsample is presumably dominated by QCD events with kinematic properties identical to the QCD background events that satisfy the event selection ($I < 0.1$). Each event in the sub-sample is then weighted by the total positive tagging rate measured from the jet sample (see Sec. 8.2). The mistag background shape is estimated from the pretag sample, where each event is weighted by the negative tag rate measured from the jet sample. The background normalization is estimated using methods described in Chap. 8. Other backgrounds (which account for less than 10% of the total background) are included in the overall normalization, with the implicit assumption that their shape is not significantly different from the others. The $t\bar{t}$ contribution is normalized to the theoretical cross section. A subsample of 107 pb^{-1} that was available at the time was used for this optimization study.

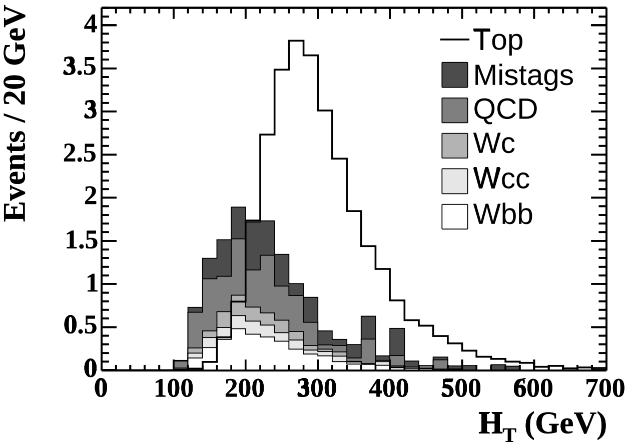


Figure 9.1: *Distribution of the H_T variable for $t\bar{t}$ Monte Carlo, and for various backgrounds normalized to an integrated luminosity of* $107\,\mathrm{pb}^{-1}$.

Figure 9.1 shows that signal and background can be efficiently separated by the use of the H_T variable. Figure 9.2 shows the signal over background ratio and cross section sensitivity as a function of an H_T cut, computed from Fig. 9.1. The statistical sensitivity $(S/\sqrt{S+B})$ is compared to the total sensitivity $(S/\sqrt{S+B+\sigma(B)^2}$, where $\sigma(B)$ is the absolute systematic error on the background estimate). Systematic uncertainties arising from the H_T cut itself are described in Sec. 9.3; they are small enough to be neglected in the optimization process. A cut requiring $H_T > 200$ GeV is found to be optimal: such a cut keeps 96% of the signal and rejects 39% of the background; this improves the signal over background ratio from 2 to 3 and the total significance on the $t\bar{t}$ cross section measurement by 6% for an integrated luminosity of 107 pb^{-1}.

9.2 Acceptance

The acceptance is the fraction of produced $t\bar{t}$ events that are accepted by the triggers and satisfy the selection criteria described in Sec. 4.7. It includes trigger efficiency, reconstruction efficiency, as well as the efficiency of the kinematic cuts, and of the b-tagging algorithm. The branching ratios of the various decay modes of the $t\bar{t}$ pair are included as well. The acceptance is split into various efficiencies and correction factors as follows:

$$\epsilon_{t\bar{t}} = \epsilon_{\mathrm{trig}} \cdot \epsilon_{z_0} \cdot \epsilon_{\mathrm{veto}} \cdot k_{\mathrm{lep-id}} \cdot \epsilon_{\mathrm{tag-event}} \cdot \epsilon_{t\bar{t}}^{\mathrm{MC}} \tag{9.1}$$

where ϵ_{z_0} is the efficiency of the $|z_0| \leq 60$ cm cut, ϵ_{trig} is the trigger efficiency for identifying high p_T leptons, $\epsilon_{t\bar{t}}^{\mathrm{MC}}$ is the fraction of Monte Carlo $t\bar{t}$ events which pass

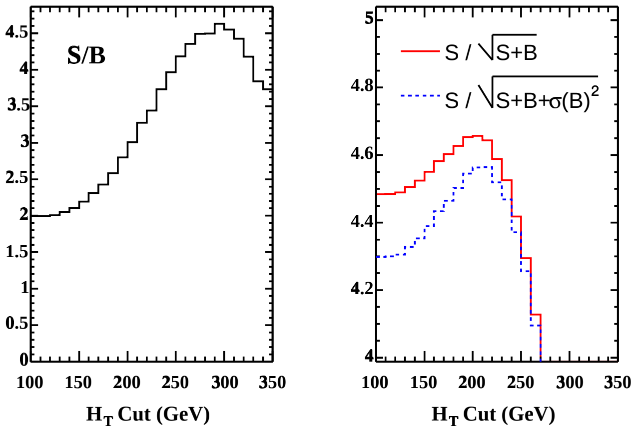


Figure 9.2: *Estimate of S/B, statistical and total significance for* $107\,\mathrm{pb}^{-1}$ *integrated luminosity, as a function of* H_T *cut.*

all the selection cuts (except for b-tagging), and $\epsilon_{\mathrm{tag-event}}$ is the efficiency to tag at least one jet in a $t\bar{t}$ event with the SecVtx b-tagging algorithm. $\epsilon_{t\bar{t}}^{\mathrm{MC}}$ includes the efficiency of the various vetoes (conversion removal, cosmic removal, di-lepton, and Z^0 rejections). $k_{\mathrm{lep-id}}$ is a factor that corrects for the lepton identification efficiency difference between data and Monte Carlo.

A sample of PYTHIA $t\bar{t}$ Monte Carlo events with a top quark mass of $m_t = 175\,\mathrm{GeV}/c^2$ is used to estimate $\epsilon_{\mathrm{tag-event}}$ and $\epsilon_{t\bar{t}}^{\mathrm{MC}}$. All decay modes are allowed, so that contributions from the di-lepton, tau and hadronic channels are taken into account as well as the lepton+jets mode. Tables 9.1, 9.2, and 9.3 show the event count in the MC for each final state of the W pair, and after the selection cuts for CEM, CMUP, and CMX respectively, summarized here:

- Event vertex $|z_0| < 60$ cm.
- Lepton kinematics: basic requirement that an electron candidate with $E_T >$ 20 GeV (or a muon candidate with $p_T >$ 20 GeV) be found.
- Lepton ID: the lepton must satisfy the tight selection requirements described in Chap. 4 and be isolated ($I < 0.1$). Figure 9.3 shows the isolation distribution for electrons and muons that pass the tight lepton selection; the $I < 0.1$ requirement has an efficiency of ≈86% for all leptons.
- Transverse missing energy $\not\!\!E_T >$ 20 GeV. Figure 9.4 shows the $\not\!\!E_T$ distribution after lepton selection.
- At least three jets with $E_T >$ 15 GeV and $|\eta| < 2$.

Cut	e	μ	τ	ee	$\mu\mu$	$\tau\tau$	$e\mu$	$e\tau$	$\mu\tau$	$qqqq$	Total
Initial Sample	51970	52251	51727	4153	4181	4195	8375	8171	8221	162793	356037
$\lvert z_0\rvert < 60$ cm	50223	50523	50002	4026	4051	4061	8117	7924	7952	157385	344264
CEM kinematics	33216	16416	19083	3201	922	1409	4997	5100	2289	64867	151500
CEM ID	17774	61	1428	2402	13	229	3077	3104	252	81	28421
$\not\!\!E_T > 20$ GeV	15865	55	1229	2215	13	201	2834	2875	229	44	25560
≥ 3 jets	14028	43	1102	1027	3	94	523	1217	53	44	18134
Di-Lepton veto	14007	33	1099	767	2	92	308	1161	29	44	17542
Z^0 veto	13699	32	1080	601	1	87	266	1041	22	43	16872
Conversion veto	13657	27	1073	600	1	85	266	1037	21	23	16790
$H_T > 200$ GeV	13118	25	1000	526	1	70	226	888	16	23	15893
≥ 1 b-tag	7009	13	528	282	0	33	117	489	13	6	8490

Table 9.1: *Number of events in the $t\bar{t}$ MC sample passing the CEM electron selection cuts.*

Cut	e	μ	τ	ee	$\mu\mu$	$\tau\tau$	$e\mu$	$e\tau$	$\mu\tau$	$qqqq$	Total
Initial Sample	51970	52251	51727	4153	4181	4195	8375	8171	8221	162793	356037
$\lvert z_0\rvert < 60$ cm	50223	50523	50002	4026	4051	4061	8117	7924	7952	157385	344264
CMUP kinematics	967	14609	2035	68	1887	262	2403	319	2369	3458	28377
CMUP ID	0	10894	791	0	1564	166	1860	142	1869	3	17289
$\not\!\!E_T > 20$ GeV	0	9675	684	0	1442	151	1702	130	1723	1	15508
≥ 3 jets	0	8567	611	0	276	56	650	55	746	1	10962
Di-Lepton veto	0	8562	609	0	165	56	452	35	706	1	10586
Z^0 veto	0	8456	605	0	145	54	394	31	647	1	10333
Cosmic veto	0	8456	605	0	145	54	394	31	647	1	10333
$H_T > 200$ GeV	0	8104	561	0	133	42	361	24	565	1	9791
≥ 1 b-tag	0	4285	320	0	68	18	185	7	318	1	5202

Table 9.2: *Number of events in the $t\bar{t}$ MC sample passing the CMUP muon selection cuts.*

- Di-Lepton events, Z^0 events, conversion candidates (for electrons) and cosmic candidates (for muons) are vetoed.
- Total transverse energy $H_T > 200$ GeV.
- At least one of the jets with $E_T > 15$ GeV and $|\eta| < 2$ must be positively tagged by SecVtx.

the cut at $\not\!\!E_T > 20$ GeV has an efficiency of ≈90%. Figure 9.5 shows the transverse energy of the four leading jets after the lepton selection, the $\not\!\!E_T$requirement, and the vetoes have been applied. The ≥ 3 jets requirement corresponds to demanding the third jet to have $E_T > 15$ GeV; the efficiency of this cut is ≈81%. Finally, Fig. 9.6 shows the jet multiplicity of $t\bar{t}$ events after the full selection has been applied (including b-tagging), with and without the $H_T > 200$ GeV requirement. Clearly, a large fraction of the events contains only three jets, which is why three jet events are included in the selection.

The Monte Carlo acceptance is corrected for several effects that are evaluated from data:

Cut	e	μ	τ	ee	$\mu\mu$	$\tau\tau$	$e\mu$	$e\tau$	$\mu\tau$	$qqqq$	Total
Initial Sample	51970	52251	51727	4153	4181	4195	8375	8171	8221	162793	356037
$\lvert z_0 \rvert < 60$ cm	50223	50523	50002	4026	4051	4061	8117	7924	7952	157385	344264
CMX kinematics	355	5462	667	21	803	77	839	94	894	1136	10348
CMX ID	1	4077	260	0	670	52	641	36	706	1	6444
$\not\!\!E_T > 20$ GeV	1	3619	230	0	616	46	595	34	638	0	5779
≥ 3 jets	1	3226	208	0	122	18	234	12	260	0	4081
Di-Lepton veto	1	3223	207	0	70	18	168	8	244	0	3939
Z^0 veto	1	3189	203	0	59	18	148	8	223	0	3849
Cosmic veto	1	3189	203	0	59	18	148	8	223	0	3849
$H_T > 200$ GeV	1	3037	183	0	53	15	131	6	191	0	3617
≥ 1 b-tag	0	1648	93	0	31	8	78	2	105	0	1965

Table 9.3: *Number of events in the $t\bar{t}$ MC sample passing the CMX muon selection cuts.*

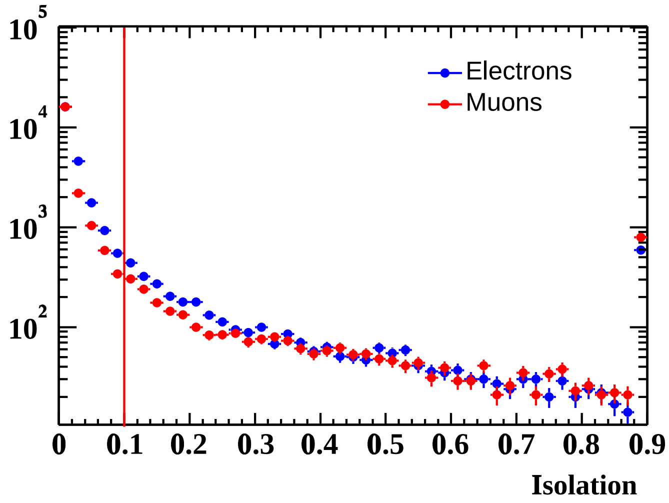


Figure 9.3: *Isolation of the electrons and muons in* PYTHIA *$t\bar{t}$ Monte Carlo events with a top quark mass of $m_t = 175\,\mathrm{GeV}/c^2$ after tight lepton selection. The event selection requires $I > 0.1$.*

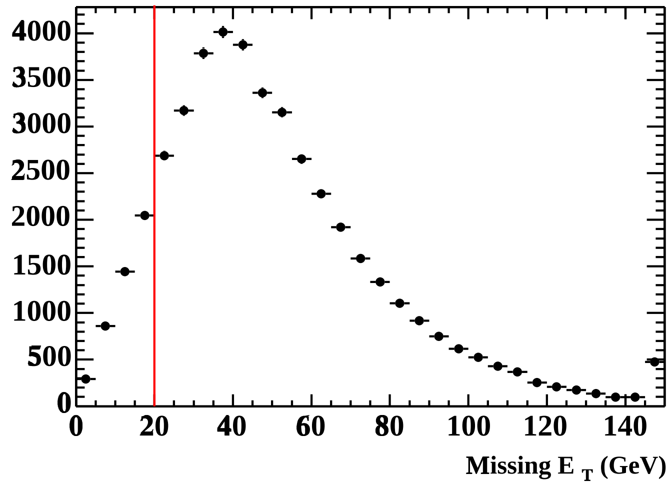


Figure 9.4: *Transverse missing energy in* PYTHIA $t\bar{t}$ *Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after tight lepton selection and* $I > 0.1$. *The event selection requires* $\not\!\!E_T > 20$ *GeV.*

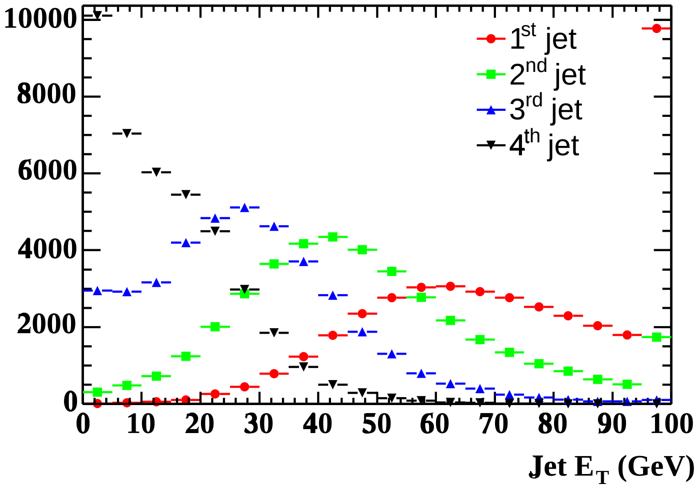


Figure 9.5: *Transverse energy of the first four jets (in decreasing order of transverse energy) satisfying* $|\eta| < 2$ *in* PYTHIA $t\bar{t}$ *Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after tight lepton selection,* $I > 0.1$, *and* $\not\!\!E_T > 20$ *GeV, with and without the* $H_T > 200$ *GeV requirement. The event selection requires the third jet to have* $E_T > 15$ *GeV.*

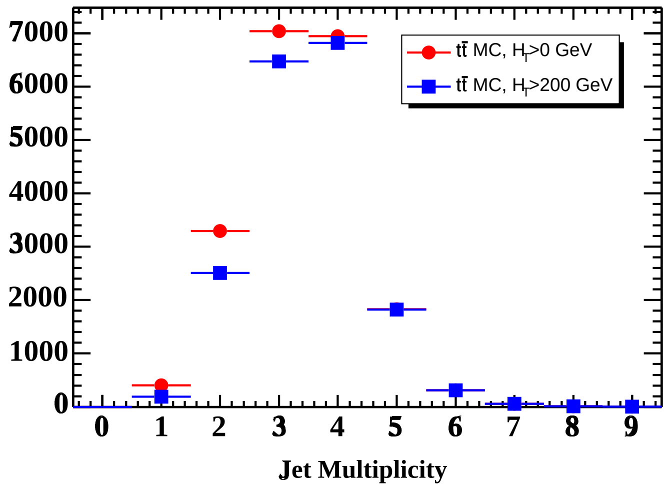


Figure 9.6: *Number of jets with* $E_T > 15$ *GeV and* $|\eta < 2|$ *in* PYTHIA $t\bar{t}$ *Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after tight lepton selection,* $I > 0.1$*, and* $\not\!\!E_T > 20$ *GeV. The event selection requires at least three jets.*

- The b-tagging efficiency is measured from the Monte Carlo sample, but takes into account the b-tagging scale factor of 0.82 ± 0.06 (Chap. 6) by randomly keeping only 82% of all the tags, and discarding the others. The efficiency for tagging at least one jet in a $t\bar{t}$ event (after all other cuts have been applied, including $H_T > 200\,\mathrm{GeV}$) is found to be 53.4±0.3(stat.)±3.2(syst.)%, corresponding to a 6.0% relative uncertainty on the acceptance.

- The efficiency of the z_0 cut is measured from data: $\epsilon_{z_0} = 0.951 \pm 0.005$.

- The trigger efficiency ϵ_{trig} is measured from overlapping independent triggers. It is different for each type of lepton: $\epsilon_{\mathrm{trig}}^{CEM} = 0.9656 \pm 0.0006$, $\epsilon_{\mathrm{trig}}^{CMUP} = 0.887 \pm 0.007$, and $\epsilon_{\mathrm{trig}}^{CMX} = 0.954 \pm 0.006$.

- The factor $k_{\mathrm{lep-id}}$ is evaluated by comparing a data sample of Z^0 events with a PYTHIA Z^0 sample, and found to be 0.965 ± 0.006 for electrons, 0.887 ± 0.009 for CMUP muons, and 1.006 ± 0.008 for CMX muons.

Additional systematic uncertainties due to the electron E_T scale, E_T resolution and the amount of material in the detector amount to 0.3%. For muons, uncertainties due to the p_T scale and resolution and to the muon chambers geometry amount to 1.2%. Systematic uncertainties specific to each type of lepton are listed in Table 9.5. Other systematic uncertainties common to all types of leptons are shown in Table 9.6 and detailed here:

- The tracking efficiency for leptons is assigned a 0.4% uncertainty.

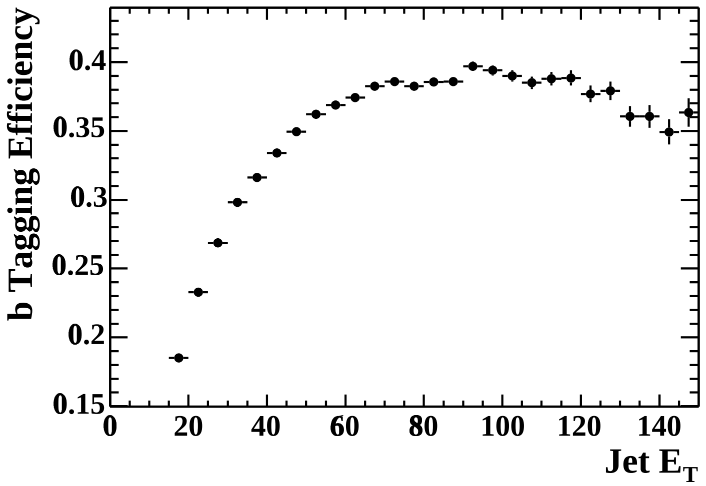


Figure 9.7: *SecVtx b-tagging efficiency for b jets with* $E_T > 15$ *GeV and* $|\eta < 2|$ *in* PYTHIA $t\bar{t}$ *Monte Carlo events with a top quark mass of* $m_t = 175\,\mathrm{GeV}/c^2$ *after full event selection (no requirement on* H_T*) as a function of the jet transverse energy. The jets are required to match a b parton.*

- Lepton isolation: $k_{\mathrm{lep-id}}$ is evaluated in a Z^0 sample with little jet activity, while $t\bar{t}$ events have large jet multiplicity, so that leptons tend to be less isolated in the latter case. The statistically limited Z^0 + Jets sample is used to check the isolation dependence on $k_{\mathrm{lep-id}}$ and a 5% systematic uncertainty is assigned.

- Jet energy scale: based on the single jet energy uncertainty, the jet energies in the $t\bar{t}$ Monte Carlo are shifted by $\pm\sigma$ and the average of the difference from the nominal acceptance (4.9%) is taken as a systematic uncertainty.

- Initial and Final State Radiation (ISR and FSR): in PYTHIA, the parameters relative to ISR and FSR are varied and the change in the acceptance is taken an uncertainty (2.6%).

- PDF: the uncertainty on the PDF is evaluated from the CTEQ6M eigenvector 90% confidence intervals. The difference for the LO CTEQ5L and the NLO CTEQ6M PDFs, as well as the uncertainty on α_S are also included, for a total uncertainty of 2%.

- Monte Carlo generator: the acceptance is measured with both PYTHIA and HERWIG, and half the difference (1.4%) is assigned as a sytematic uncertainty on the choice of the $t\bar{t}$ generator.

The overall acceptance, combining all types of lepton, is 3.84±0.03(stat.)±0.40(syst.)%, including all systematic effects. The systematic uncertainties amount to 8.5% for the

	CEM	CMUP	CMX	Total
MC Sample	344,264	344,264	344,264	344,264
# Events w/o b-tag	15,893	9791	3617	29301
ϵ_{z_0}	0.951 ± 0.005	0.951 ± 0.005	0.951 ± 0.005	-
ϵ_{trig}	0.9656 ± 0.0006	0.887 ± 0.007	0.954 ± 0.006	-
$k_{\mathrm{lep-id}}$	0.965 ± 0.006	0.887 ± 0.009	1.006 ± 0.008	-
Acc. w/o b-tag (%)	4.09±0.36	2.13±0.19	0.959±0.085	7.18±0.61
# Tagged Events	8490	5202	1965	15657
Tag Efficiency (%)	53.4±3.2	53.1±3.2	54.3±3.3	53.4±3.2
Acc. with b-tag (%)	2.19±0.23	1.14±0.12	0.512±0.054	3.84±0.40
Integ. Lumi. (pb^{-1})	162±10	162±10	150±9	

Table 9.4: *Summary table of the $t\bar{t}$ acceptance, for a top quark mass of* $175\,\mathrm{GeV}/c^2$. *Only systematic uncertainties are shown (the statistical uncertainties from the Monte Carlo sample are negligible).*

Quantity	CEM	CMUP	CMX
ϵ_{trig}	0.06	0.8	0.6
$k_{\mathrm{lep-id}}$	0.6	1.1	0.8
Conversion removal	1.4	-	-
Cosmic veto	-	1	1
Other (see text)	0.3	1.2	1.2

Table 9.5: *Relative systematic uncertainties (%) on the signal acceptance relative to CEM electrons, CMUP muons, and CMX muons.*

pretag acceptance, while the event b-tagging efficiency introduces a 6.0% uncertainty. The uncertainty on the luminosity measurement is 5.9%. Table 9.4 summarizes the event selection acceptance for each type of lepton.

9.3 Remarks on the Systematics Associated with the H_T Cut

The $H_T > 200\,\mathrm{GeV}$ increases the systematic uncertainty relative to the jet energy scale from 2.9% to 4.9% for the acceptance. The effect of the H_T cut on the estimate of the W + Jets background, in particular the heavy flavor fraction, is negligible. Finally, the H_T cut slightly increases the dependence of our measurement to the top quark mass: a variation of +5 GeV (-5 GeV) of the top quark mass increases the acceptance by 4.3% (decreases by -1.9%), instead of 3.9% (-1.1%) without the cut. In the cross section measurement, this variation is not treated as a systematic uncertainty but rather the cross section is expressed separately as a function of the

Quantity	Relative error (%)
ϵ_{z_0}	0.5
Tracking Efficiency	0.4
Energy Scale	4.9
PDF	2.0
ISR/FSR	2.6
Lepton Isolation	5.0
Generator	1.4
b-tagging	5.9

Table 9.6: *Relative systematic uncertainties on the signal acceptance which are common to all lepton types.*

top quark mass.

Chapter 10

Measurement of the $\sigma_{t\bar{t}}$ Production Cross Section

10.1 The Z^0 + Jets Samples: Validation of the Background Estimate

The Z^0 + Jets sample provides a good cross check on the background calculations since the $t\bar{t}$ and QCD contamination in this sample are small. The heavy flavor contribution in Z^0 + Jets is expected to be close to that in W + Jets in terms of gluon splitting. However, the gluon fusion diagram $gg \to Z^0 b\bar{b}$ has no counterpart in W + Jets, so that the heavy flavor is enhanced in the Z^0 + Jets production. The same procedures described in Chap. 7 are used to estimate the heavy flavor fractions using the Z^0 + Jets ALPGEN Monte Carlo samples. The fraction of $Zc\bar{c}$+Zc events is approximately twice the fraction of $Wc\bar{c}$+Wc events, and the fraction of $Z^0 b\bar{b}$ events is approximately twice the fraction of $Wb\bar{b}$ events.

Events with a Z^0 boson are selected by identifying oppositely charged e^+e^- and $\mu^+\mu^-$ pairs with an invariant mass between 75 and 105 GeV/c^2. Both leptons are required to pass the tight lepton selection used for the W + Jets analysis. Table 10.1 and Fig. 10.1 show the yield of Z^0 candidates and the number of tagged events observed as a function of jet multiplicity compared to the predictions, estimated in the same way as in Chap. 8 for the W + Jets sample: 14.0 ± 1.9 events are predicted and 18 are observed. A good agreement between data and Monte Carlo is observed at every jet multiplicity.

Some Z^0 + Jets events that fail the standard Z^0 removal (if one leg is not reconstructed) contribute to the W + Jets sample. Based on Monte Carlo, the number of $Z^0 \to \mu^+\mu^-$ events that fall into the W sample is $72 \pm 8\%$ of the number of events observed in $Z^0 \to \mu^+\mu^-$ decay. The contribution of $Z^0 \to e^+e^-$, on the other hand, is negligible, because of the better electron coverage. The W + Jets backgrounds (mistags and W + Heavy Flavor) are normalized with the pretag data sample, so that the Z^0 contribution is already included in the estimate of these backgrounds, except for the fact that the heavy flavor fraction is twice as large in Z^0 + Jets processes than in the W + Jets processes: the expected number of additional tagged events in the W + Jets sample is shown in Table 10.1 and included in the final background estimate.

10.2 The W + Jets Sample: $t\bar{t}$ Production Cross Section

Table 10.2 shows the event count in the pretag and tagged W + Jets samples, together with the background estimates. The production cross section follows from the acceptance and luminosity measurements, and from the background estimate:

$$\sigma_{t\bar{t}} = \frac{N_{\text{observed events}} - N_{\text{background events}}}{\mathrm{A}_{t\bar{t}} \cdot \mathcal{L}} \tag{10.1}$$

Jet multiplicity	Z^0+1 jet	Z^0+2 jets	Z^0+$\geq$3 jets
Pretag sample			
$Z \to e^+e^-$	410	48	10
$Z \to \mu^+\mu^-$	402	59	15
$Z \to \ell^+\ell^-$	812	107	25
Tagged Estimate			
Mistags	2.4 ± 0.2	0.49 ± 0.06	0.23 ± 0.04
$Z^0 b\bar{b}$	1.6 ± 0.4	0.8 ± 0.2	0.26 ± 0.08
$Z^0 c\bar{c}$	4.4 ± 1.3	2.3 ± 0.7	0.8 ± 0.2
$t\bar{t}$ ($\sigma_{t\bar{t}} = 5.6 \pm 1.4$)	0.08 ± 0.02	0.5 ± 0.1	0.13 ± 0.03
Total	8.5 ± 1.7	4.1 ± 0.9	1.4 ± 0.3
Observed Tagged Events	12	3	3
Contribution to Pretag W+Jets	289 ± 35	42 ± 7	11 ± 3
Extra-Contribution to Tagged W+Jets	1.1 ± 0.3	0.6 ± 0.2	0.2 ± 0.1

Table 10.1: *The predicted number of Z^0 + Jets events and the observed number, along with the Z^0 + Jets contribution in the W + Jets sample and the estimate of the resulting extra b-tags in that sample. (The prediction of extra b-tagged events is included in the predicted background summary for the W + Jets sample.)*

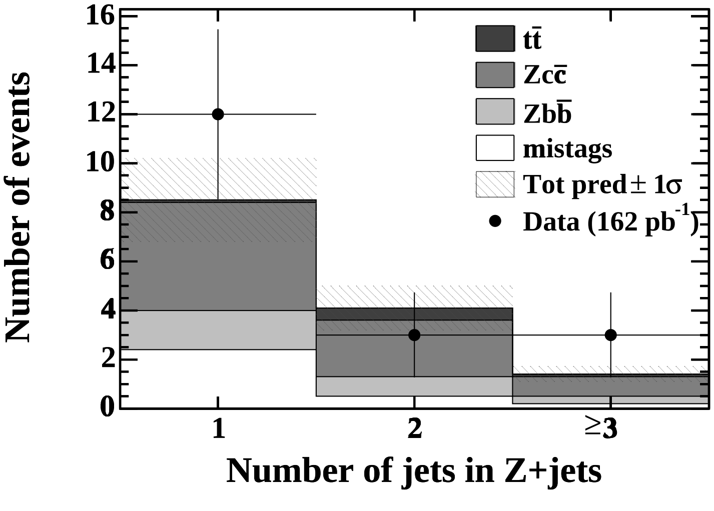


Figure 10.1: *Comparison of the observed and predicted number of events in the b-tagged Z^0 + Jets sample.*

			$H_T > 0$		$H_T > 200$ GeV	
Jet multiplicity	1 jet	2 jets	3 jets	$\geq$ 4 jets	3 jets	$\geq$ 4 jets
Pretag	15314	2448	387	107	179	91
Mistags	40.9 ± 6.1	17.0 ± 2.4	5.2 ± 0.7	2.6 ± 0.4	3.3 ± 0.4	2.3 ± 0.3
$Wb\bar{b}$	37.0 ± 11.2	22.5 ± 6.5	5.0 ± 1.3	1.6 ± 0.5	2.8 ± 0.8	1.4 ± 0.4
$Wc\bar{c}$	13.7 ± 3.4	8.0 ± 2.2	1.6 ± 0.5	0.6 ± 0.2	0.9 ± 0.3	0.5 ± 0.2
Wc	34.5 ± 9.0	7.7 ± 2.0	1.4 ± 0.4	0.3 ± 0.1	0.7 ± 0.2	0.3 ± 0.1
Diboson,$Z \to \tau\tau$	2.2 ± 0.4	2.5 ± 0.4	0.6 ± 0.1	0.1 ± 0.0	0.3 ± 0.1	0.1 ± 0.0
QCD	22.9±3.3	10.1±1.7	3.4±0.7	1.4±0.4	1.7±0.4	1.2±0.3
single top	2.6 ± 0.3	4.6 ± 0.5	1.1 ± 0.1	0.2 ± 0.0	0.8 ± 0.1	0.2 ± 0.0
Z^0+HF	1.1 ± 0.3	0.6 ± 0.2	0.2 ± 0.1		0.10 ± 0.05	
Total	154.9 ± 19.1	73.0 ± 9.8	18.5 ± 2.2	6.9 ± 0.9	10.5 ± 1.3	6.0 ± 0.8
Corrected Total	154.9 ± 19.1	73.0 ± 9.8	23.1 ± 3.0		13.5 ± 1.8	
Data	160	73	29	28	21	27

Table 10.2: *Background summary for the single-tag selection. The total backgrounds are given before and after the correction for $t\bar{t}$ events in the pretag W+jets sample.*

where N_{obs} and N_{bkg} are the number of total observed and background events, respectively, in tagged events with $\geq$ 3 jet bins and $H_T > 200$ GeV; $\epsilon_{t\bar{t}}$ is the signal acceptance (see Table 9.4); and $\mathcal{L}$ is the integrated luminosity. The W + Jets backgrounds (mistags and W + Heavy Flavor) are normalized with the number of events in the pretag sample, which includes a significant contribution from $t\bar{t}$ events; After subtracting the $t\bar{t}$ contribution from the pretag sample in an iterative way, the backgrounds are recalculated ("Corrected Total"); 13.5 background events are expected, while 48 events are observed in the signal sample: this excess is attributed to the $t\bar{t}$ signal. Assuming a top quark mass $m_t = 175\,\text{GeV}/c^2$, the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV is measured:

$$\sigma_{t\bar{t}} = 5.6^{+1.2}_{-1.1}(\text{stat.})^{+0.9}_{-0.6}(\text{syst.})\text{pb} \tag{10.2}$$

The result is consistent with a Standard Model $t\bar{t}$ signal and a top mass of ≈175 GeV, as well as with the Run I measurement, as shown in Fig. 10.3. The acceptance, and therefore the measured cross section, changes with the top quark: Table 10.3 shows the measurement for three values of the top mass (170, 175, and 180 GeV/c^2). Table 10.4 summarizes the sources of systematic uncertainty, compared to the statistical error. Figure 10.2 shows the jet multiplicity of the data events, compared to the expected backgrounds and a $t\bar{t}$ signal assuming a cross section of $\sigma_{t\bar{t}} = 5.6$ pb: the agreement between data and prediction is very good across the entire range of jet multiplicity.

Figures 10.4 to 10.9 show some kinematic properties of the candidates, compared to the expectation. The $t\bar{t}$ contribution is normalized to its theoretical production cross section of 6.7 pb. As a cross check, Figure 10.4 shows the H_T distribution in tagged events with one or two jets: the good agreement confirms a good understanding of the H_T variable. Figure 10.5 shows the H_T distribution of the 57 tagged event candidates with $\geq$ 3 jet: the agreement is excellent and the candidates look very much like

m_t(GeV/c^2)	σ (pb)
170	$5.8^{+1.2}_{-1.1}$(stat.)$^{+0.9}_{-0.6}$(syst.)
175	$5.6^{+1.2}_{-1.1}$(stat.)$^{+0.9}_{-0.6}$(syst.)
180	$5.4^{+1.1}_{-1.0}$(stat.)$^{+0.9}_{-0.6}$(syst.)

Table 10.3: *Measured cross section for different top quark mass assumptions.*

Source	Relative uncertainty on $\sigma_{t\bar{t}}$
Pretag acceptance	8.5%
b-tagging efficiency	6.0%
Background estimate	5.2%
Luminosity	5.9%
Total Systematic	13%
Statistical	21%

Table 10.4: *Relative uncertainties contributing to the $t\bar{t}$ production cross section error.*

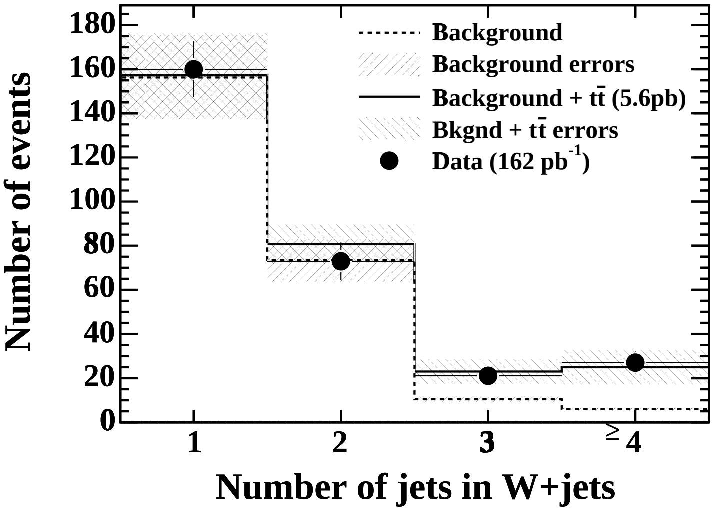


Figure 10.2: *Background and $t\bar{t}$ signal expectation as a function of jet multiplicity.*

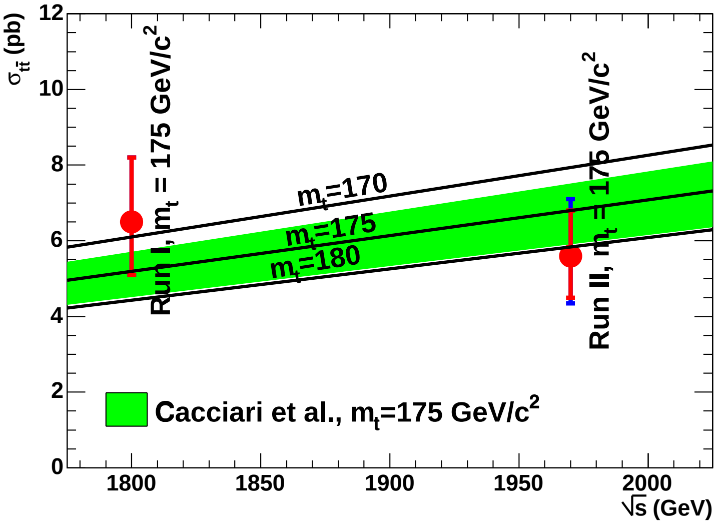


Figure 10.3: *Comparison of the $t\bar{t}$ production cross section measured in Run I at $\sqrt{s}$ = 1.8 TeV [53] (combining measurements using a soft lepton tagger and a secondary vertex tagger in the Lepton+Jets channel), the Run II measurement at $\sqrt{s}$ = 1.96 TeV, and the Standard Model prediction [9] ($m_t = 175\,\mathrm{GeV}/c^2$).*

$t\bar{t}$ events. Similarly, Fig. 10.6 shows the transverse energy of the tagged jets of the 48 event candidates that pass the additional $H_T > 200$ GeV requirement (events with two tagged jets have two entries). Again, the tagged jet E_T spectrum is very $t\bar{t}$-like. Finally, Figures 10.7, 10.8, and 10.9 show the distributions of the transverse decay length L_{2D}, the vertex mass, and the pseudo-$c\tau$ of the SecVtx tags in these jets, all in good agreement with the expectation.

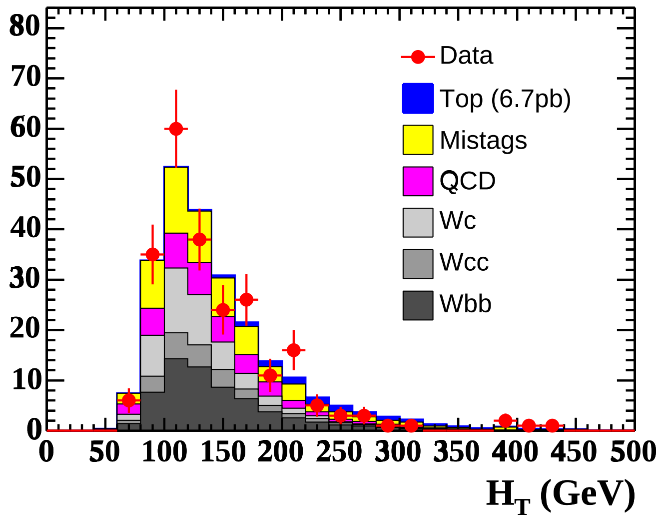


Figure 10.4: *H_T distribution of the tagged candidates with one and two jets in 162* pb^{-1} *of data, compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross section of 6.7 pb).*

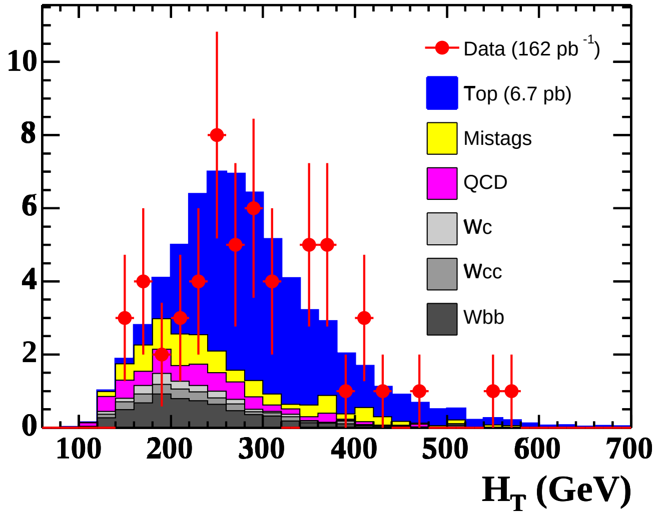


Figure 10.5: *H_T distribution of the 57 tagged candidates in 162* pb^{-1} *of data (three or more jets), compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross section of 6.7 pb).*

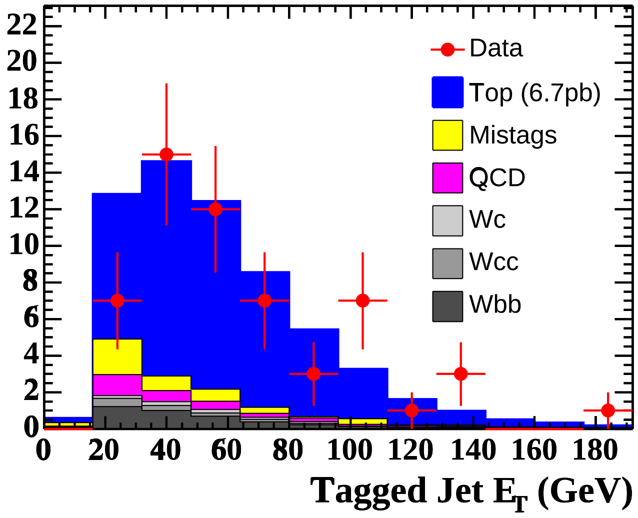


Figure 10.6: *Transverse energy of the tagged jets in the 48 tagged candidates in 162* pb^{-1} *of data (three or more jets, H_T >* 200 *GeV), compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross section of 6.7 pb).*

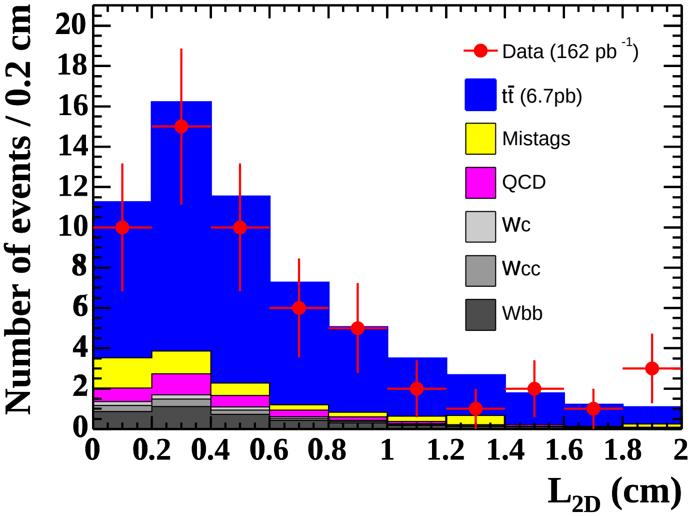


Figure 10.7: *Transverse decay length of the secondary vertices in the 48 tagged candidates in 162* pb^{-1} *of data (three or more jets, H_T >* 200 *GeV), compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross section of 6.7 pb).*

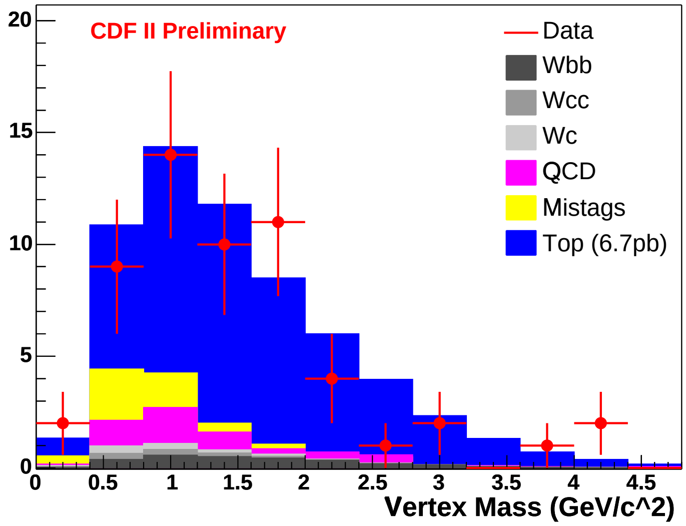


Figure 10.8: *Reconstructed mass of the secondary vertices in the 48 tagged candidates in 162* pb^{-1} *of data (three or more jets,* $H_T > 200$ *GeV), compared to the expected background and* $t\bar{t}$ *signal (normalized to the theoretical cross section of 6.7 pb).*

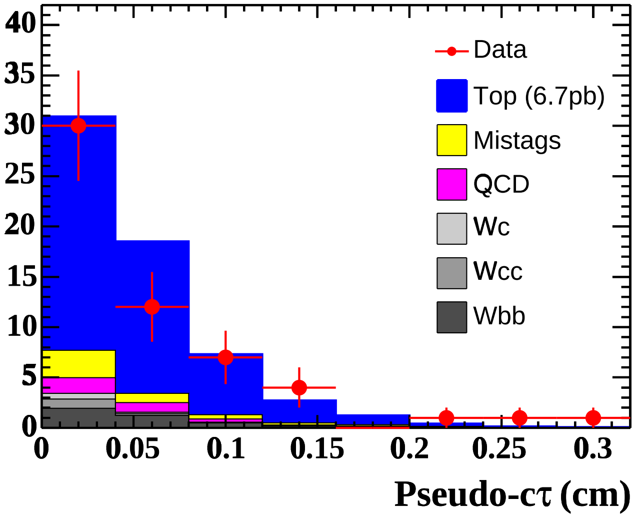


Figure 10.9: *Pseudo-*$c\tau$ *of the secondary vertices in the 48 tagged candidates in 162* pb^{-1} *of data (three or more jets,* $H_T > 200$ *GeV), compared to the expected background and* $t\bar{t}$ *signal (normalized to the theoretical cross section of 6.7 pb).*

Chapter 11

Conclusion and Outlook

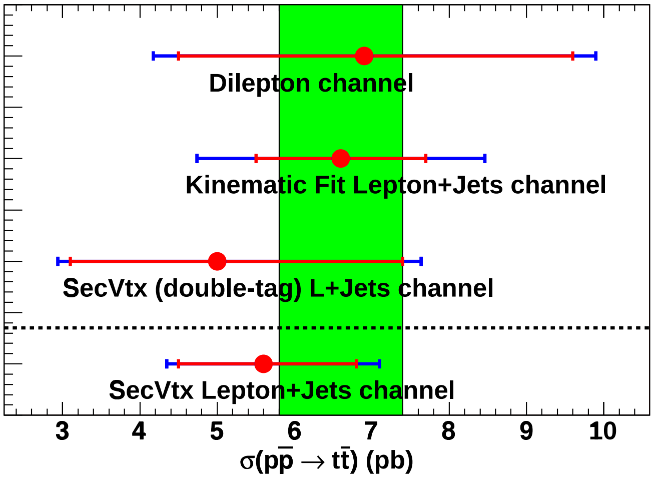


Figure 11.1: *Comparison of the result with some other current CDF Run II $t\bar{t}$ production cross section measurements. The systematic uncertainties (blue) are added in quadrature with the statistical ones (red).*

The $t\bar{t}$ production cross section in $p\bar{p}$ collisions has been measured with a larger data sample and at a higher center-of-mass energy than previously done during the Tevatron Run I. Based on 162 pb^{-1} of data, and assuming a top quark mass $m_t = 175\,\text{GeV}/c^2$, the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV is:

$$\sigma_{t\bar{t}} = 5.6^{+1.2}_{-1.1}(\text{stat.})^{+0.9}_{-0.6}(\text{syst.})\text{pb}$$

This measurement is currently the single most precise measurement of the $t\bar{t}$ production cross section. The result is consistent with a Standard Model $t\bar{t}$ signal and a top mass of $\approx 175\,\text{GeV}/c^2$, as shown in Fig. 10.3. The basic kinematic properties of the candidate events have been studied, and are also in good agreement with the Standard Model. Other measurements using the Run II data accumulated during the same period, but different methods or decay channels, have been realized at Run II. Figure 11.1 compares the measurement described here, with some other such measurements: a measurement similar to this one, using a sample of events with at least two SecVtx-tagged jets [54], a measurement without b-tagging that uses the event kinematics to extract the $t\bar{t}$ signal from the sample [55] (these two methods use the same lepton+jets channel), and a measurement that uses the di-lepton channel [56]. Within uncertainties, they agree with each other and with the Standard Model prediction. Work is in progress to combine these results into a single, improved CDF Run II measurement.

The Tevatron will still have the monopoly on top physics for a few more years until the Large Hadron Collider (LHC) turns on. Much more data are going to be collected at the Tevatron in the meantime, providing a great opportunity to study the top quark. This particular analysis should benefit from a larger data set: even though the statistical uncertainty will not be the dominant one very soon, many systematics are expected to be improved with more data. In particular, the understanding of the lepton isolation in high jet multiplicity events requires a larger Z^0 + Jets sample. The b-tagging efficiency uncertainty should decrease as well with more statistics. The systematic uncertainty due to the jet energy measurement should be significantly improved in the near future, as more effort is put into the detector simulation and the measurement of the top quark mass. Additionally, the b-tagging capability of the silicon detector can be increased by the use of the innermost silicon layer (L00) and forward tracking to extend b-tagging up to $|\eta| = 2$. The electron identification will also be extended up to $|\eta| = 2$ in the future. Even without these improvements in the event reconstruction, a 2 fb^{-1} data set would decrease the statistical uncertainty by a factor 3.5, and one can hope to reach a precision of the order of 10% with this single measurement. Finally, in order to bypass the large systematic uncertainty due to the absolute luminosity measurement, the $t\bar{t}$ production cross section could also be measured relatively to a well-predicted reference cross section, such as the $W^{\pm}$ or Z^0 cross sections.

The LHC will be a "top factory", producing of the order of 8 million $t\bar{t}$ pairs per year at the design low luminosity of $10^{33}/\text{cm}^2/\text{s}$, and will be able to give definite answers about top physics. Many analyses will be limited by systematic effects almost from the beginning. Precision measurements of the top quark mass, width, branching ratios will be possible, allowing to test the Standard Model in the top sector. Moreover, the top signal will be the main source of background for many analyses, such as Higgs and SUSY searches. Thus a precise understanding of the $t\bar{t}$ signal, and of its production cross section in the first place, will be required to pursue searches beyond the Standard Model.

Bibliography

[1] F. Abe *et al.*, CDF Collaboration, Phys. Rev. Lett. 74 2626 (1995);
S. Abachi *et al.*, D0 Collaboration, Phys. Rev. Lett. 74 2632 (1995).

[2] K. Kodama *et al.*, Phys.Lett. B 504, 218 (2001), hep-ex/0012035.

[3] CDF Collaborattion, D0 Collaboration, and Tevatron Electroweak Working Group, hep-ex/0404010.

[4] Review of particle physics, S. Eidelman *et al.*, Phys. Lett. B 592, 1 (2004).

[5] S. Weinberg, Phys. Rev. Lett. 19, 1264 (1967);
A. Salam, *Elementary Particle Theory* (1969);
S.L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D2, 1285 (1970).

[6] P. W. Higgs, Phys. Rev. Lett. 145, 1156 (1966).

[7] LEP Electroweak Working Group, Summer '04,
http://lepewwg.web.cern.ch/LEPEWWG/plots/summer2004/

[8] R. Barate *et al.*, Phys. Lett. B 565, 61 (2003), hep-ex/0306033.

[9] M. Cacciari, S. Frixione, G. Ridolfi, M.L. Mangano, and P. Nason, JHEP 404, 68 (2004).

[10] R.K.Bock and A.Vasilescu, http://rkb.home.cern.ch/rkb/titleD.html

[11] CDF Collaboration, FERMILAB-Pub-96/390-E (1996).

[12] A. Sill *et al.*, Nucl. Instrum. and Methods A 447, 1 (2000).

[13] L. Vacavant, CDF Internal Note 5825.

[14] L. Balka *et al.*, Nucl. Instrum. and Methods A 267, 272 (1988).

[15] G. Apollinari *et al.*, Nucl. Instrum. and Methods A 267, 301 (1988).

[16] M.G. Albrow *et al.*, Nucl. Instrum. and Methods A 480, 524 (2002).

[17] G. Ascoli *et al.*, Nucl. Instrum. and Methods A 268, 33 (1988).

[18] T. Dorigo *et al.*, Nucl. Instrum. and Methods A 461, 560 (2001).

[19] D. Acosta *et al.*, Nucl. Instrum. and Methods A 461, 540 (2001).

[20] P. D. B. Collins, Cambridge University Press (1977).

[21] J. Pumplin *et al.*, JHEP 0207, 012 (2002).

[22] P. Avery, K. Read, and G. Trahern, CLEO Report CSN-212 (1985).

[23] G. Marchesini and B.R. Webber, Nucl. Phys. B 310, 461 (1988).

[24] G. Corcella *et al.*, JHEP 0101, 010 (2001).

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

[26] M.L. Mangano *et al.*, JHEP 07, 001 (2003).

[27] R. Brun and F. Carminati, GEANT Detector Description and Simulation Tool, CERN Programming Library Long Writeup W5013 (1993).

[28] G. Grindhammer, M. Rudowicz, and S. Peters, Nucl. Instrum. and Methods A 290, 469 (1990).

[29] R. Veenhof, A drift-chamber simulation program, http://garfield.web.cern.ch/garfield/

[30] C. Hayes *et al.*, CDF Internal Note 6992.

[31] M. Gold, CDF Internal Note 5871.

[32] F. Abe *et al.*, Phys. Rev. D 45, 1448 (1992).

[33] T. Affolder *et al.*, Phys. Rev. D 64, 032001 (2001).

[34] C. Hayes *et al.*, CDF Internal Note 6992.

[35] C. Chen *et al.*, CDF Internal Note 6394.

[36] W.-M. Yao and K. Bloom, CDF Internal Note 5991.

[37] F. Abe *et al.*, CDF Collaboration, Phys. Rev. D 50, 2966 (1994).

[38] L. Balka *et al.*, Nucl. Instrum. and Methods A 267, 272 (1988).

[39] F. Abe *et al.*, CDF Collaboration, Phys. Rev. L 68, 1104 (1992).

[40] C. Rott *et al.*, CDF Internal Note 6007.

[41] S. Baroiant *et al.*, CDF Internal Note 6257.

[42] K. Bloom and W. Dagenhart, CDF Internal Note 6347.

[43] T. Wright, in CDF Internal Note 6904.

[44] S. Catani *et al.*, JHEP 11, 063 (2001).

[45] F. Krauss *et al.*, JHEP 08, 015, (2002).

[46] M. Mangano, Merging Multijet MEs with shower MCs, http://cepa.fnal.gov/patriot/tt_mc4run2/MCTuning/061104/mlm.pdf (2004).

[47] S. Mrenna and P. Richardson, JHEP 0405, 040 (2004).

[48] J.M. Campbell and J. Huston, arXiv:hep-ph/0405276, submitted to Phys. Rev. D.

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

[50] B.W. Harris *et al.*, Phys. Rev. D 66, 054024 (2002).

[51] A. Bocci *et al.*, CDF Internal Note 4681.

[52] A. Safonov and A. Korytov, CDF Internal Note 5278.

[53] T. Affolder *et al.*, Phys. Rev. D 64, 032002 (2001).

[54] D. Acosta *et al.*, submitted to PRD, CDF Internal Note 7138.

[55] D. Acosta *et al.*, to be submitted to PRD, CDF Internal Note 7288.

[56] D. Acosta *et al.*, submitted to PRL, CDF Internal Note 6882.

[57] I. Volobouev and H. Bachacou, CDF Internal Note 6163.

[58] SVXMon Online Manual
http://cdfcodebrowser.fnal.gov/CdfCode/source/SvxDaqMods/
doc/si_consumer_manual.html

[59] Hans Wenzel and Kaori Maeshima, CDF Internal Note 4835.

[60] http://www-cdfonline.fnal.gov/consumer/home/consumer_home.html

[61] Brent Welch, "Practical Programming in Tcl and Tk", 3rd Edition, Prentice Hall, 2000.

[62] http://cdfcodebrowser.fnal.gov/CdfCode/source/ErrorLogger/doc/html/
0ErrorLogger.html

[63] http://sourceforge.net/projects/blt/

[64] Available from http://www-cdf.lbl.gov/~igv/tcl_extensions/signal_mod1.4.tar.gz.

[65] http://nkek15.fnal.gov/cgi-bin/newproblemdb/DBmain.pl

[66] E. Sexton-Kennedy and M. Shapiro, "Creating Talk_to Menus for AC++", CDF Internal Note 4521.

[67] I. Volobouev, "Speaking Better Tcl to Your AC++ Modules", CDF Internal Note 5731.

[68] http://www-cdfonline.fnal.gov/cdfdb/servlet/RunTypes

[69] M. Bishai *et al.*, "An SVX3D Chip User's Companion", CDF Internal Note 5062.

[70] W. Badgett *et al.*, "Description of the CDF Data Structure for Run II", CDF Internal Note 4152.

[71] J. Conway *et al.*, "Numbering and Naming Convention for SVX II and ISL", CDF Internal Note 4103.

[72] http://www-cdf.lbl.gov/~bachacou/form.html

[73] http://tcl.activestate.com/man/tcl8.0/TclCmd/string.htm#M10

[74] http://fcdfhome.fnal.gov/usr/bachacou/SVXMon/SVXMon_CO_manual/svxmon_CO_instructions.html

Appendix A

Derivation of the b-Tagging Efficiency Estimate

The measurement of the b-tagging efficiency described in Chap. 6 uses identified conversions to estimate the contribution of electrons which are fake or part of a conversion pair. This appendix summarizes the detailed calculation of the b-tagging efficiency.

Most of the electrons in the low-p_T electron data sample are produced by conversions or are fake electrons in light flavor jets. The heavy flavor fraction in the electron side can be enhanced by requiring the away-jet to be tagged, but even so a large fraction of the electron-jets remains light flavored. The heavy flavor production in the jets comes from three subprocesses: direct production, flavor excitation and gluon splitting. For simplicity, the final data sample can be divided into the following four subclasses:

- N_{BB}: the number of events where both sides contain a heavy flavor quark, either c or b,
- N_{BQ}: the number of events where the electron-jet is heavy flavored and the away-jet is light flavored,
- N_{QB}: the number of events where the electron-jet is light flavored (the electron is coming from fakes or conversions) and the away-jet is heavy flavored,
- N_{QQ}: the number of events where both jets are light flavored.

By construction, we have

$$N_{BB} + N_{BQ} + N_{QB} + N_{QQ} = N$$

where N is the total number of events. The heavy flavor contributions in the electron side can be determined using the measurement of heavy flavor fraction:

$$N_{BB} + N_{BQ} = F_{HF} \cdot N$$

The N_{QB} contribution can be determined using the away tags in the conversion electron sample. Finally, the contribution of N_{QQ} is estimated using the mistags in the negative side. Let us define the following notation:

- ϵ'_B: b-tagging efficiency of heavy flavor in the electron-jet,
- ϵ_B: b-tagging efficiency of heavy flavor in the away-jet,
- ϵ'_Q: mistag efficiency in the electron-jets,
- ϵ_Q: mistag efficiency in the away-jets,
- N_{a+}, N_{a-}, N_{e+} and N_{e-}: number of events where the away-jet (the electron-jet) is positively (negatively) tagged (single-tags).
- N^{e+}_{a+}, N^{e-}_{a+}, N^{e+}_{a-} and N^{e-}_{a-}: number of events where both jets are tagged (double-tags) for every combination of positive and negative tag.

Applying the b-tag in the away-jets, the numbers of positive and negative tags are:

$$\epsilon_B \cdot N_{BB} + \epsilon_Q \cdot N_{BQ} + \epsilon_B \cdot N_{QB} + \epsilon_Q \cdot N_{QQ} = N_{a+} \tag{A.1}$$

$$\epsilon_Q \cdot N_{BB} + \epsilon_Q \cdot N_{BQ} + \epsilon_Q \cdot N_{QB} + \epsilon_Q \cdot N_{QQ} = N_{a-}. \tag{A.2}$$

By subtracting Equ. A.2 from Equ. A.1 , we get

$$(\epsilon_B - \epsilon_Q) \cdot (N_{BB} + N_{QB}) = N_{a+} - N_{a-}. \tag{A.3}$$

Applying the second tag on the electron side, the numbers of double-tags are

$$\epsilon'_B \cdot (\epsilon_B - \epsilon_Q) \cdot N_{BB} + \epsilon'_Q \cdot (\epsilon_B - \epsilon_Q) \cdot N_{QB} = N^{e+}_{a+} - N^{e+}_{a-} \tag{A.4}$$

$$\epsilon'_Q \cdot (\epsilon_B - \epsilon_Q) \cdot N_{BB} + \epsilon'_Q \cdot (\epsilon_B - \epsilon_Q) \cdot N_{QB} = N^{e-}_{a+} - N^{e-}_{a-}. \tag{A.5}$$

Subtracting Equ. A.5 from Equ. A.4, we get

$$(\epsilon'_B - \epsilon'_Q) \cdot (\epsilon_B - \epsilon_Q) \cdot N_{BB} = (N^{e+}_{a+} - N^{e-}_{a+}) - (N^{e+}_{a-} - N^{e-}_{a-}). \tag{A.6}$$

From Equ. A.3, we get

$$(\epsilon_B - \epsilon_Q) \cdot N_{BB} = (N_{a+} - N_{a-}) - (\epsilon_B - \epsilon_Q) \cdot N_{QB} \tag{A.7}$$

Substituting Equ. A.7 into Equ. A.6 and rearranging terms, the b-tagging efficiency on the electron-jet is

$$\epsilon'_B - \epsilon'_Q = \frac{(N^{e+}_{a+} - N^{e-}_{a+}) - (N^{e+}_{a-} - N^{e-}_{a-})}{(N_{a+} - N_{a-}) - (\epsilon_B - \epsilon_Q) \cdot N_{QB}}$$

In order to determine N_{QB}, we select the events where the electron is identified as a conversion partner and the away side is b-tagged. The heavy flavor contribution in the away-jets should not depend on whether the electron originated from a photon conversion or a fake. Let us define the following quantities:

- f: the fraction of electrons originating from conversions in light jets in the electron side,
- f': the fraction of electrons originating from conversions in heavy flavor jets in the electron side,
- ϵ^c: the efficiency of the conversion finding algorithm,
- ϵ^0: the error rate of finding a real electron as a part of conversion, which is determined using same sign "conversions",
- n^c: the number of identified conversion electrons.

Applying the conversion finding algorithm to the data sample, the number of conversion electrons is:

$$(f'\epsilon^c + (1-f')\epsilon^0)\cdot(N_{BB}+N_{BQ}) + (f\epsilon^c + (1-f)\epsilon^0)\cdot(N_{QB}+N_{QQ}) = n^c \quad \text{(A.8)}$$

$$f(\epsilon^c-\epsilon^0)\cdot(N_{QB}+N_{QQ}) = n^c - N\epsilon^0 - f'\cdot(\epsilon^c-\epsilon^0)\cdot(N_{BB}+N_{BQ}) \quad \text{(A.9)}$$

By looking for conversions in the tagged electron-jets, we have:

$$(\epsilon'_B-\epsilon'_Q)\cdot(\epsilon^c\cdot f' + \epsilon^0\cdot(1-f'))\cdot(N_{BB}+N_{BQ}) = n^c_{e+} - n^c_{e-}.$$

Since $(\epsilon'_B-\epsilon'_Q)\cdot(N_{BB}+N_{BQ}) = N_{e+}-N_{e-}$, we get:

$$\epsilon^0 + f'\cdot(\epsilon^c-\epsilon^0) = \epsilon'_c \quad \text{(A.10)}$$

where $\epsilon'_c = \frac{(n^c_{e+}-n^c_{e-})}{(N_{e+}-N_{e-})}$. Substituting Equ. A.10 into Equ. A.8 , we have:

$$f\cdot(\epsilon^c-\epsilon^0) = \frac{n^c/N - (\epsilon^0 + (\epsilon'_c-\epsilon^0)\cdot F_{HF})}{1-F_{HF}} \quad \text{(A.11)}$$

Now, we apply the b-tag on the *electron* side in Equ. A.8 and the excess of tags is:

$$f'\cdot(\epsilon^c-\epsilon^0)\cdot(\epsilon'_B-\epsilon'_Q)\cdot(N_{BB}+N_{BQ}) = n^c_{e+} - n^c_{e-} - (N_{e+}-N_{e-})\cdot\epsilon^o, \quad \text{(A.12)}$$

and if we apply the b-tag on the *away* side in Equ. A.8 the excess of tags is:

$$f\cdot(\epsilon^c-\epsilon^0)\cdot(\epsilon_B-\epsilon_Q)\cdot N_{QB} = n^c_{a+}-n^c_{a-}-(N_{a+}-N_{a-})\cdot\epsilon^o - f'\cdot(\epsilon^c-\epsilon^0)\cdot(\epsilon_B-\epsilon_Q)\cdot N_{BB}. \quad \text{(A.13)}$$

Substituting Equ. A.7 and A.10 into Equ. A.13:

$$f\cdot(\epsilon^c-\epsilon^0)\cdot(\epsilon_B-\epsilon_Q)\cdot N_{QB} = n^c_{a+}-n^c_{a-}-(N_{a+}-N_{a-})\cdot\epsilon^0-(\epsilon'_c-\epsilon^0)\cdot((N_{a+}-N_{a-})-(\epsilon_B-\epsilon_Q)\cdot N_{QB}) \quad \text{(A.14)}$$

From Equ. A.11 and Equ. A.14, we get:

$$(\epsilon_B-\epsilon_Q)\cdot N_{QB} = (N_{a+}-N_{a-})\cdot\frac{\frac{n^c_{a+}-n^c_{a-}}{N_{a+}-N_{a-}} - \epsilon'_c}{n^c/N - \epsilon'_c}\cdot(1-F_{HF})$$

Finally, the efficiency can be expressed as:

$$\epsilon'_B - \epsilon'_Q = \frac{(N^{e+}_{a+} - N^{e-}_{a+}) - (N^{e+}_{a-} - N^{e-}_{a-})}{(N_{a+} - N_{a-}) \cdot F^{a+}_{HF}} \tag{A.15}$$

where

$$F^{a+}_{HF} = 1 - \frac{\frac{n^c_{a+} - n^c_{a-}}{N_{a+} - N_{a-}} - \epsilon'_c}{n^c/N - \epsilon'_c} \cdot (1 - F_{HF}) \tag{A.16}$$

Appendix B

List of Event Candidates

Here are listed the b-tagged W+jets event candidates with three jets, and at least four jets, separately. No requirement on the H_T variable is made. The jet quantities are relative to the *tagged* jet (if two jets are tagged in the event, it is listed twice in the table, once for each tagged jet).

B.0.1 3-Jet Events

Run #	Evt #	$\not{E}_T$	H_T	Jet E_T	ϕ	$\eta_{\text{det.}}$	L_{2D}
Electrons							
141618	1968597	64.7447	254.313	101.701	2.60089	0.575363	0.197437
144674	1782954	26.357	264.536	40.2624	2.28309	0.77782	0.694814
148153	6088	54.1679	256.625	71.8762	0.151561	0.184107	0.587188
149387	2551061	22.7792	175.025	37.8804	3.04406	0.700927	1.1859
155145	132579	42.2368	292.402	65.6392	0.524674	-0.714355	0.213425
155145	132579	42.2368	292.402	35.3924	4.52756	-1.15148	0.274205
155345	3194866	68.1073	209.88	46.2693	4.69239	0.615436	0.150715
160541	3336609	112.878	575.155	138.117	3.80199	-0.0646956	2.64821
161678	7014539	28.7786	409.865	187.695	5.6853	-0.275312	0.243108
162178	2690653	50.1795	273.437	63.9978	5.97717	0.741207	0.820473
162663	80178	41.0529	245.523	55.855	0.197616	-0.143647	0.40104
162663	80178	41.0529	245.523	45.9506	4.82608	-0.807448	0.84019
164354	1842551	43.7291	161.269	31.2965	0.262435	0.105041	0.870485
165121	428911	45.4898	381.845	64.3309	3.74214	0.119683	0.155424
165836	566421	34.5337	249.352	83.1627	5.84605	0.404662	1.47331
165902	1487424	72.2613	281.048	40.3018	4.55194	-0.41656	0.275843
165949	434475	21.1366	296.997	121.092	2.77023	0.81694	0.175908
165949	1239675	24.4202	144.285	20.1419	0.182284	0.194252	0.0841145
166037	1912411	23.387	146.686	28.158	1.05047	-0.0594986	0.143249
166567	6377008	42.2395	168.865	38.1018	1.44274	-0.355494	0.314013
166615	6805282	23.462	181.705	20.3877	2.08205	0.257734	0.152756
Muons							
141597	1353293	28.6333	237.114	42.8944	1.81778	0.177422	1.89999
151978	507773	82.1292	234.656	40.8101	5.77172	0.850115	0.348022
160406	154621	23.5261	256.985	40.1872	1.16551	0.923947	0.843269
162519	931923	27.778	208.405	48.8469	0.386276	0.013553	0.246949
160796	2445318	28.7052	153.426	17.0457	5.90518	1.19179	0.350889
164274	2932602	28.5504	223.236	37.8838	4.02992	1.02118	0.816675
165313	1770456	31.3677	306.091	102.711	1.38432	-0.0441787	0.457352
165314	1155563	23.3947	162.975	22.7591	0.620043	-0.776249	0.282697
166567	490760	96.0994	259.308	43.9837	1.61885	0.994469	0.543782
166567	490760	96.0994	259.308	29.6381	2.15572	-0.363614	0.507485
166927	5822870	39.3893	297.68	100.885	0.315039	-0.255028	1.98992

Table B.1: *Positively-tagged jets in W + 3 jet events.*

B.0.2 $\geq$ 4-Jet Events

Run #	Evt #	$\not{E}_T$	H_T	Jet E_T	ϕ	$\eta_{\text{det.}}$	L_{2D}
				Electrons			
153694	1694029	32.1621	318.509	63.6704	0.634712	-0.584663	0.510161
153738	2083102	67.4126	342.491	107.051	3.3847	-0.284402	0.511347
154175	1630925	126.837	364.862	18.8704	3.11091	0.791546	0.0797298
155320	480816	30.6075	251.354	54.1402	3.80981	-0.302445	0.352291
156116	6116596	32.2545	254.985	74.4534	1.20883	-0.598097	1.03947
156116	6116596	32.2545	254.985	57.074	5.69167	-0.900674	0.107979
160153	1270879	59.5783	346.486	53.2456	6.04137	-0.586578	0.122465
161013	111162	88.8547	434.06	109.596	2.36352	-0.0548293	0.10707
161379	494836	23.8865	407.178	32.2888	4.414	0.355498	1.43787
161633	1571961	33.3592	280.546	54.5085	4.23765	-0.0380233	0.614509
161678	5912849	56.3333	361.236	102.865	4.0542	-0.568211	0.38472
162423	261933	29.0737	219.584	39.9638	2.33258	0.398643	0.289171
162986	1538897	59.1495	409.426	46.0505	3.59004	1.25426	0.215135
162986	1538897	59.1495	409.426	31.2757	0.527554	-0.73743	1.82016
164819	2297394	52.0934	196	39.3243	4.06939	0.640204	0.40178
166007	498553	65.6	315.082	70.122	3.97118	-0.533691	0.440997
166614	804529	39.7316	379.213	53.0467	4.15911	-0.668765	0.391993
167551	7969376	66.721	357.638	79.6301	2.51029	-0.6341	0.229068
167551	7969376	66.721	357.638	31.6901	3.46822	1.02481	0.299236
				Muons			
145036	245760	68.5293	318.149	101.894	4.86412	-0.429823	0.617196
152266	3554	41.985	233.775	45.8359	4.49481	0.0776384	0.714783
153693	799494	38.8804	365.316	57.429	4.03046	0.205398	0.430063
160437	280173	34.219	278.999	61.9213	4.60407	1.00186	0.968925
160591	894406	116.777	540.943	70.3086	2.36592	-0.537394	1.09902
161788	361577	73.6752	462.33	129.447	2.57824	0.334251	0.748809
162631	7109631	27.4969	276.802	20.0195	1.16291	0.599294	0.683939
163012	2249546	31.41	360.388	26.4523	5.92551	0.521895	0.172776
166567	11615607	65.4086	348.326	81.5695	5.9693	0.406215	1.36201
166567	11615607	65.4086	348.326	84.1142	4.43329	-1.00317	1.75068
167139	1191211	68.9279	270.058	62.4818	1.86866	-0.855328	0.580815
167139	1191211	68.9279	270.058	40.2173	5.0884	0.482068	0.252367
166805	2534588	47.2278	346.676	131.057	3.28113	-0.639837	0.0625955
167551	3626393	50.5703	271.886	17.5064	4.54702	-0.842267	0.357138

Table B.2: *Tagged jets in $W + \geq 4$ jet events.*

Appendix C

Particle Level Study of an Alternative Jet Reconstruction Algorithm Using Tracking Information

This Monte-Carlo study is based on a $W \to q\bar{q}$ and a $Z^0 \to b\bar{b}$ event data sample generated with PYTHIA [25]. It studies the JCL98 tower classification described in [51] in order to improve the jet resolution in CDF by combining the tracking system and the calorimeter information. A limit is put on the improvement one can expect from such a method as a function of calorimeter segmentation and resolution.

C.1 Introduction

Improving the jet energy resolution in CDF is of great importance for RUN II, especially for the top mass measurement and Higgs search. The discovery potential of the Higgs boson is highly dependent on the b jet energy resolution, and even more on the $b\bar{b}$ dijet mass resolution.

This motivates recent efforts to find a new method of jet reconstruction that combines the calorimeter, the CES, and the tracking system information. The so-called JCL98 method has been implemented and tested using a RUN I γ+jet data sample (see [51]). In this note we present a Monte-Carlo study of an analogous method using a toy detector simulation. The basic idea of the method replaces the hadronic measurement (which has relatively poor resolution) with the tracking information whenever possible. For example, if a charged particle (but no other particle) deposits all its energy in one single calorimeter tower and its momentum is well measured by the COT, it is possible to discard the calorimeter tower energy in the computation of the jet energy, and replace it by the momentum measurement instead. The main limitation of the method arises from overlaps of charged particles, photons, and neutral particles in the same calorimeter tower, which makes it difficult to distinguish between the track-associated particle energy and other particles energy. Other detector effects, such as tower leakage, cracks, or energy from hadrons deposited in the EM calorimeter, increase the confusion: in this note, we do not take these effects into account. The goal of the study is to get a better understanding in a simplified case, to put an upper limit on the improvement one might get using the tracker, and finally to estimate the impact of the calorimeter segmentation and resolution on the method.

C.2 Event Generation

The Monte-Carlo event generator PYTHIA (v. 5.720 [25]) was used to produce $W \rightarrow q\bar{q}$ and $Z^0 \rightarrow b\bar{b}$ events. Underlying event and initial and final state radiation (ISR and FSR) were included (except when explicitly specified). A z vertex position was given randomly to each event, according to a Gaussian distribution of $\sigma = 30$ cm. Λ's and K_S's were all decayed while K_L's, $K^{\pm}$'s and $\pi^{\pm}$'s were assumed to be stable.

C.3 Detector Simulation

A toy detector with the main characteristics of the CDF detector was implemented for the purpose of the study. Only the central part of the detector ($|\eta| < 1.2$) was considered. Magnetic field and calorimeter segmentation were taken into account. In addition to CDF central calorimeter segmentation ($\Delta\eta \times \Delta\phi = 0.1 \times 0.26$ for both electromagnetic (EM) and hadronic (HAD) calorimeters), two alternate calorimeter segmentations were considered : first, the EM calorimeter segmentation was set to $0.05 times 0.05$ (*i.e.* 2×5 finer than CDF) while the HAD remained unchanged; then, both EM and HAD were given a 0.05×0.05 segmentation. Although the latter case is not realistic, it helps in understanding the role of the HAD segmentation.

C.3.1 Projection of Particles into the Towers

Stable particles were determined at the particle level. Neutrinos, mostly produced by b-hadron decays and that carry a non-negligible fraction of the jet energy, were of course not taken into account.

Charged particle trajectories were assumed to be helical due to the tracking system magnetic field (1.4 T) until a radius $R = 150\,\mathrm{cm}$ from the beam axis. Photons and electrons/positrons were projected to the EM calorimeter at the CES radius ($R = 183\,\mathrm{cm}$). Hadrons were projected to the HAD calorimeter ($R = 280\,\mathrm{cm}$). Each particle was attributed to the tower in which it was at the end of the projection and deposited all its energy in the tower (especially, hadrons and muons did NOT deposit any energy in the EM calorimeter). Tower leakage or cracks were not simulated.

Muons were considered separately : the muon detector was assumed to have 100% efficiency; if the muon track was reconstructed within the jet cone, its momentum was simply added to the jet, else it was not taken into account.

C.3.2 Efficiency and Resolution

- Calorimeters :

The calorimeter energies were smeared with a Gaussian distribution. No threshold was included, but negative values given by the Gaussian (which happens most frequently at low energy) were replaced by 0.

Resolutions (E in GeV) :

$\frac{\sigma(E)}{E} = \frac{13\%}{\sqrt{E}} + 2\%$ for the EM calorimeter.

$\frac{\sigma(E)}{E} = \frac{100\%}{\sqrt{E}}$ for the HAD calorimeter.

Some more recent studies claim a resolution of $\frac{\sigma(E)}{E} = \frac{50\%}{\sqrt{E}} + 3\%$ for the HAD calorimeter; although the first value is our default one in this note, we also looked at the effect of such a change in the HAD resolution.

- Tracker :

  The track efficiency was assumed to be 0 for $p_T < 500\,\mathrm{MeV}$ and 0.9 above 500 MeV (see [52]). The resolution was assumed to be perfect and momenta were taken directly from the generator. This assumption is valid since even at energies as high as 10 GeV, the resolution of the tracker is more than an order of magnitude better than the one of the hadronic calorimeter.

- CES efficiency and photon identification :

  The CES response to photons was studied on the minimum bias event Run I Data sample (more than 1 million events). In the following we describe how the CES efficiency for photons was determined. Photons were selected by requiring :

  - an EM tower with $E_T > 100\,\mathrm{MeV}$.

  - no energy deposit in the corresponding HAD tower or in the neighbor EM and HAD towers.

  - no track pointing to the tower or the neighbor towers.

  The probability of having a CES hit above the threshold (250 MeV) in such a tower gives the CES efficiency for photons and is plotted in figure C.1 as a function of the photon E_T.

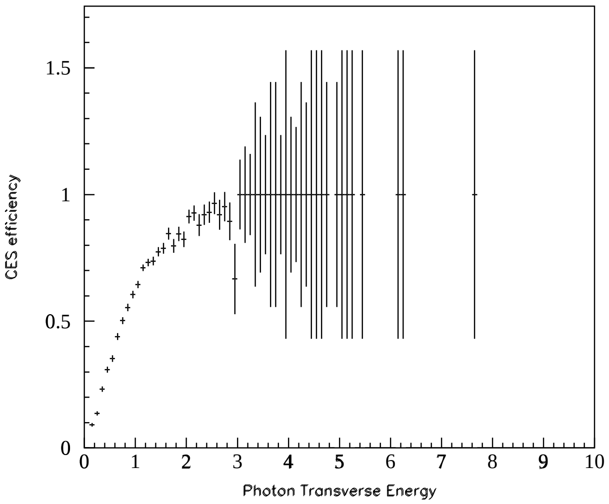


Figure C.1: *CES photon efficiency (Run I minimum bias event data).*

With a good approximation, the CES efficiency is linear from 0 to 1 between $E_T = 0\,\mathrm{GeV}$ and $E_T = 2\,\mathrm{GeV}$, and 1 above 2 GeV : this is what was used in the simulation.

A photon was actually identified by the toy detector if :

- a hit was detected in the CES (using the CES efficiency already mentioned).

- no track was projected into the same tower within $|\Delta z| < 5\,\mathrm{cm}$ (else the photon hit is indistinguishable from the charged particle hit).

C.3.3 Jet Reconstruction

The 2 primary quarks from the W and Z^0 decay were identified at the particle level, each quark giving rise to a jet. A simple cone algorithm was implemented for the purpose of the simulation. The seed-jet axis was the quark axis taken at the generator level and the tower energy centroid was computed iteratively until convergence. A cone radius $R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.7$ was chosen. All the towers with center within the cone were attributed to the jet.

In addition to the jet reconstruction described above, energy clusters were reconstructed at the particle level (*i.e.* particle momenta taken from the generator) in the pseudorapidity range $|\eta| < 2.4$. Particles with $p_T > 1\,\mathrm{GeV}$ were taken as cluster seeds and a cone algorithm (radius 0.7) was used.

C.3.4 Comparison with the Run I CDF Fast Simulation (QFL)

In order to check our toy detector as well as the jet reconstruction, we compared them to the Run I CDF fast simulation (QFL). Figure C.2 shows the dijet mass distribution in $Z^0 \to b\bar{b}$ events with two b-jets in the central detector and no other jet, for our toy detector (with the default $\frac{100\%}{\sqrt{E}}$ HAD resolution and $\frac{50\%}{\sqrt{E}} + 3\%$ HAD resolution) and QFL. All the distributions are scaled to the QFL mean value (89 GeV). The agreement is good.

C.4 Analysis

C.4.1 Event Selection

The following cuts were applied :

- Jet axis in the pseudorapidity range $|\eta| < 0.5$ in order for the jet to be contained completely in the central calorimeter. About half of the events had at least one jet passing the cut.

- $p_T^{jet} > 10\,\mathrm{GeV}$.

- No extra-jet with $p_T^{jet} > 10\,\mathrm{GeV}$ within $|\eta| < 2.4$.

The p_T^{jet} cut and the extra-jet cut have an overall efficiency of about 60%, so that about 30% of the events eventually remained after the selection.

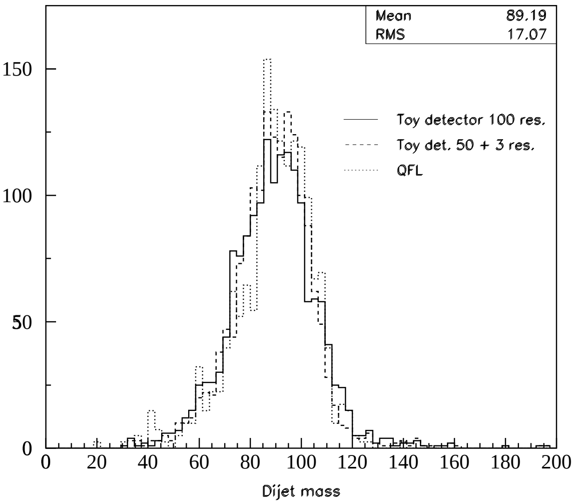


Figure C.2: *Dijet mass distribution in* $Z^0 \rightarrow b\bar{b}$ *events. ISR and FSR on. QFL (dotted line) is compared to the toy detector with CDF segmentation (solid line :* $\frac{100\%}{\sqrt{E}}$ *HAD resolution; dashed line :* $\frac{50\%}{\sqrt{E}} + 3\%$ *HAD resolution).*

C.4.2 Tower Classification

Towers were tagged following JCL98 classification [51] (except that neighbor towers are not considered, since tower leakage is not taken into account here) :

A Gamma Tower is defined as a tower in which at least one photon has been identified, but no track has been projected.

A Track Tower is defined as a tower in which at least one track has been projected, but no photon has been identified.

A Mixed Tower must contain at least one photon and one charged particle.

A Not-Assigned Tower is a tower containing energy but no CES or track hit.

Here is the definition of tower energy combining both tracker and calorimeter (referred to as T&C) used in the rest of the note :

- Track towers :

  $\vec{p}_{tower} = \Sigma\vec{p}_{track} + f^{tr}_{EM} E_{EM}\vec{r}_{EM} + f^{tr}_{HAD} E_{HAD}\vec{r}_{HAD}$

- Mixed towers :

  $\vec{p}_{tower} = \Sigma\vec{p}_{track} + f^{m}_{EM} E_{EM}\vec{r}_{EM} + f^{m}_{HAD} E_{HAD}\vec{r}_{HAD}$

- Gamma towers and Not-Assigned towers (calorimeter only) :

  $\vec{p}_{tower} = E_{EM}\vec{r}_{EM} + E_{HAD}\vec{r}_{HAD}$

where $\vec{r}_{EM}$ (resp. $\vec{r}_{HAD}$) is the unit vector pointing from the vertex to the EM (resp. HAD) cell center. E_{EM} and E_{HAD} are the energies deposited respectively in the EM cell and the HAD cell of the tower.

The 4 parameters f^{tr}_{EM}, f^{tr}_{HAD},f^{m}_{EM} and f^{m}_{HAD} are constants determined at the particle level over the whole sample in order to compensate on average for the 'non-track energy' (*i.e.* energy from neutral particles or charged particles with no reconstructed track) present in Track and Mixed towers. Figure C.3 shows the fraction of energy coming from particles with no associated track (*i.e.* neutral particles or charged particles whose track has not been reconstructed) over the total energy of the tower; the

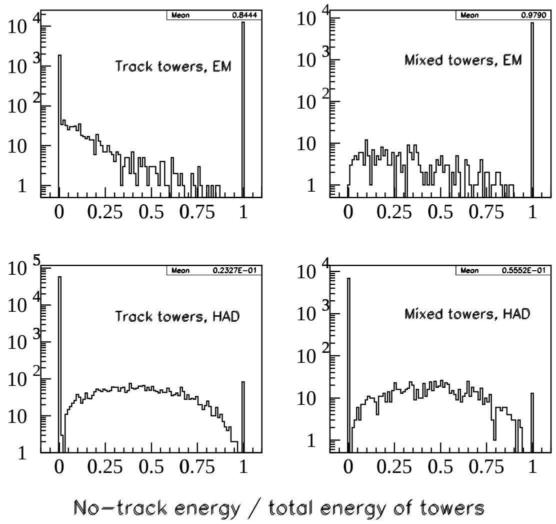


Figure C.3: *Fraction of EM and HAD energy from particles with no associated track in Track and Mixed towers. The mean values determine the parameters used in the T&C definition.* $Z^0 \rightarrow b\bar{b}$ *sample with CDF segmentation. ISR and FSR on. Real efficiencies.*

mean value of the plot determines the parameter in each specific case (Track/Mixed tower; EM/HAD energy). In first approximation, EM energy comes from gammas and must be added to the T&C tower energy, while HAD energy comes mostly from charged particles ($\pi^{\pm}$,$K^{\pm}$,p) and must be replaced by the corresponding track momentum. Nevertheless, electrons, neutral hadrons (mostly K_L, but also neutrons) and non-reconstructed tracks introduce unavoidable fluctuations, that are corrected in average. We will come back to the effects of neutrals in more detail in the next section.

Assuming CDF segmentation and efficiencies as described in the previous section, on the $Z^0 \rightarrow b\bar{b}$ sample, the parameters have the following values :

$$f^{tr}_{EM} = 85\% \qquad f^{m}_{EM} = 98\%$$
$$f^{tr}_{HAD} = 2\% \qquad f^{m}_{HAD} = 5\%$$

In case of a finer segmentation for the EM calorimeter than for the HAD calorimeter, the definition has to be slightly modified. Each tower is composed of one HAD cell and several EM sub-cells. The EM sub-cells are tagged as in the previous case; the HAD cell is tagged according to the total number of tracks and gammas that hit all the EM sub-cells of the tower. For example, if one of the EM sub-cells is tagged as a Gamma cell, and an other as a Track cell, the HAD cell will be tagged as a Mixed cell. The energy definition for a tower becomes :

- Track towers :

$$\vec{p}_{sub-cell} = \Sigma \vec{p}_{track} + f^{tr}_{EM} E_{EM} \vec{r}_{EM} \qquad \vec{p}_{HAD\ cell} = f^{tr}_{HAD} E_{HAD} \vec{r}_{HAD}$$

- Mixed towers :

$$\vec{p}_{sub-cell} = \Sigma \vec{p}_{track} + f^{m}_{EM} E_{EM} \vec{r}_{EM} \qquad \vec{p}_{HAD\ cell} = f^{m}_{HAD} E_{HAD} \vec{r}_{HAD}$$

- Gamma towers and Not-Assigned towers :

$$\vec{p}_{sub-cell} = E_{EM} \vec{r}_{EM} \qquad \vec{p}_{HAD\ cell} = E_{HAD} \vec{r}_{HAD}$$

  and the whole tower momentum is :

$$\vec{p}_{tower} = \Sigma \vec{p}_{sub-cell} + \vec{p}_{HAD\ cell}$$

C.5 Idealized Case

To begin with, we look at the $W \to q\bar{q}$ sample (almost no neutrinos or soft leptons), turn off initial and final state radiation and assume perfect calorimeter resolution as well as perfect photon identification and track reconstruction. Underlying event is included. Figure C.4 shows the ratio of the reconstructed jet p_T to the (generator level) quark p_T, as well as the dijet mass, and compares the Calorimeter-only reconstruction to the T&C reconstruction.

The width and off-set of the distributions in the Calorimeter-only reconstruction is due to several factors :

- magnetic field : loss of low energy charged particles that can not escape the solenoid (threshold at about 100 MeV); in addition to that, the bending of all charged particle trajectories introduces a bias.

- cone effect : all the particles of the jet do not lie in the 0.7 cone. The magnetic field amplifies this effect.

- the underlying event adds some background energy.

The T&C reconstruction introduces some fluctuations that appear as a low tail in the distributions, mainly due to the presence of neutrons and K_L in the HAD calorimeter, and electrons and positrons in the EM calorimeter.

Moving to $Z^0 \to b\bar{b}$ events (figure C.5), a low tail appears even in the Calorimeter-only distributions, due to the presence of neutrinos from B-hadron decays. Figure C.6 shows the ratio of the T&C momentum to the Calorimeter-only momentum with perfect calorimeter resolution (the calorimeter being taken as a reference here), for each tower, in the case of CDF segmentation. Four cases are considered : Mixed and Track towers (*i.e.* presence of photon(s) or not), and presence of neutral(s) or not (here a neutral is any particle with no associated track, *i.e.* either a neutral particle or a charged particle whose track has not been reconstructed). Perfect calorimeter resolution being assumed, the raw energy gives a perfect measurement, except for geometric effects and particle mass effects. Although the former is negligible, the latter is significant for low energy kaons and protons : this effect appears in the low tail of the upper histograms (no neutrals). When a hadron with no associated track is present in a tower (lower plots), the T&C underestimates the true energy : this is the main source of uncertainty arising from the use of the tracking. Entries above 1 are due to electrons associated with gamma energy (whether the photons are identified or not). In the case of perfect efficiencies, there is no neutral EM energy in Track towers, thus the parameter f_{EM}^{tr} is set to 0 and the upper tail vanishes; while for real efficiencies, one has to take into account unidentified photon energy ($f_{EM}^{tr} \neq 0$), introducing a bias each time an electron is present. The peak at 1.85 in Track towers with no neutrals and real efficiencies ($f_{EM}^{tr} = 0.85$) corresponds to towers with a single

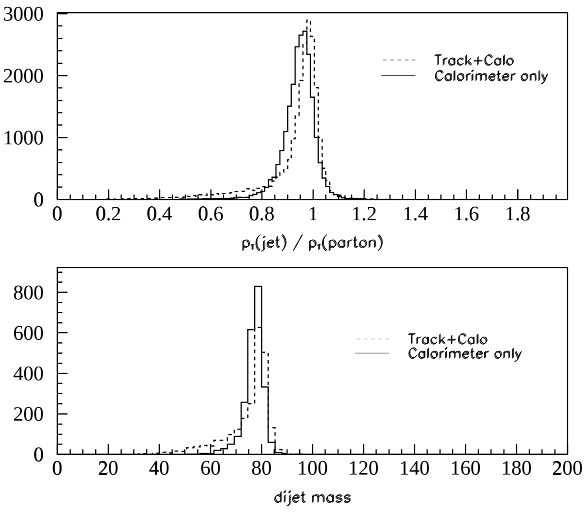


Figure C.4: *Ratio of jet p_T to corresponding quark p_T (top) and dijet mass (bottom). $W \to q\bar{q}$ sample with perfect resolution and efficiencies. CDF segmentation. ISR and FSR off. T&C (dashed line) is compared to the calorimeter only (solid line).*

electron and nothing else.

Figure C.7 shows the number of towers of each type, while figure C.8 shows the fraction of energy carried by each type of tower, in b-jets from $Z^0 \to b\bar{b}$ sample, assuming CDF segmentation and 0.05×0.05 segmentation for both EM and HAD ; ISR and FSR are included. Tables C.1 and C.2 give the average number of towers of each type, as well as the average energy fraction of each type of tower for the two segmentations. In the last column, we considered EM energy only : if only the EM calorimeter has a finer segmentation, the change appears only here. EM energy represents in average about one third of the total energy of the jet.

It has to be noticed that the Mixed tower energy fraction drops from 34% to only 5.4% with the EM and HAD segmentation improvement; from 33% to 4.8% as far as EM energy is concerned. Indeed, a finer segmentation reduces the particle multiplicity in each tower, and thus reduces the number of Mixed towers as well as the amount

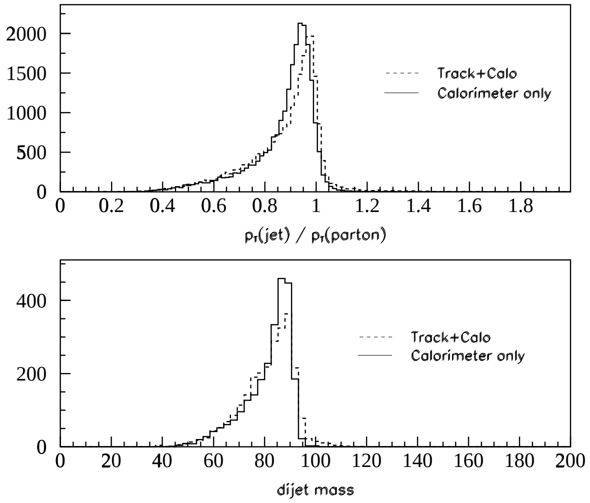


Figure C.5: *Same as figure C.4 on the* $Z^0 \to b\bar{b}$ *sample.*

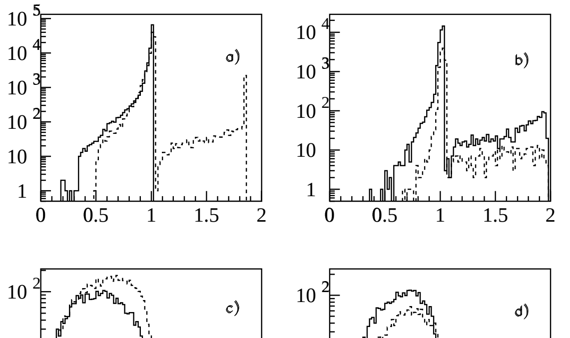


Figure C.6: *Ratio of tower T&C energy to calorimeter only energy (with perfect calorimeter resolution), assuming perfect efficiencies (solid line) and real efficiencies (dashed line).* $Z^0 \to b\bar{b}$ *sample, ISR and FSR on. a)Track towers, no neutral. b)Mixed towers, no neutral. c)Track towers with neutral(s). d)Mixed towers with neutral(s).*

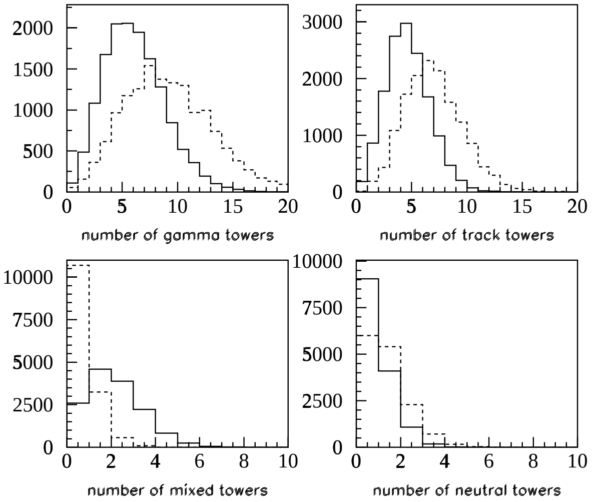


Figure C.7: *Number of towers of each type. Solid line : CDF segmentation. Dashed line : EM and HAD finer segmentation (0.05×0.05). $Z^0 \to b\bar{b}$ sample, ISR and FSR on. Perfect efficiencies are assumed.*

Table C.1: *Number of towers and fraction of energy for each type of tower. CDF segmentation, $Z^0 \to b\bar{b}$ sample, perfect resolution and efficiencies, ISR and FSR on.*

Tower type	# towers	% energy	% EM energy
All	14	100	100
Gamma	6.1	22	61
Track	4.7	38	6.2
Mixed	2.2	34	33
Not-Assigned	1.0	5.8	0

of energy in each of them. This result is important as Mixed towers are responsible for most of the fluctuations.

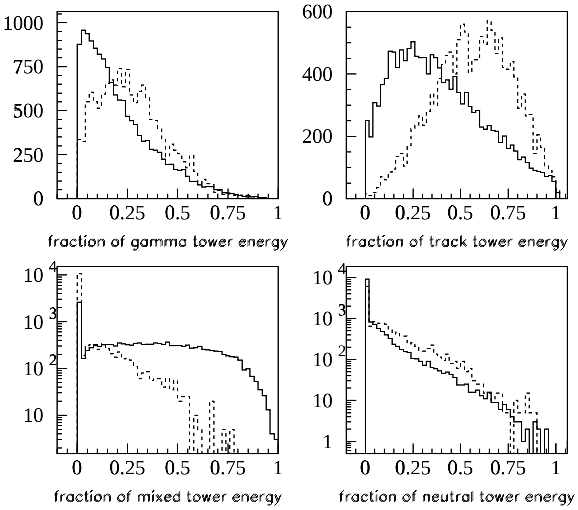


Figure C.8: *Fraction of the jet energy for each type of tower. Solid line : CDF segmentation. Dashed line : EM and HAD finer segmentation (0.05×0.05).* $Z^0 \to b\bar{b}$ *sample, ISR and FSR on. Perfect efficiencies are assumed.*

Table C.2: *Same as table C.1, except segmentation : 0.05×0.05 for both EM and HAD calorimeters.*

Tower type	# towers	% energy	% EM energy
All	17	100	100
Gamma	9.1	27	87
Track	7.0	57	8.9
Mixed	0.8	5.4	4.8
Not-Assigned	1.4	11	0

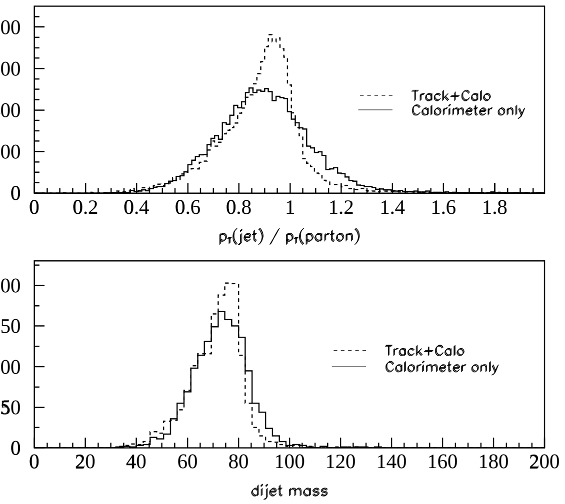


Figure C.9: *$W \to q\bar{q}$ with real resolution and efficiencies. CDF segmentation. ISR and FSR on.*

C.6 Resolution Improvement

We evaluated the resolution for both W and Z^0 samples, and for both jet p_T and dijet mass. Distributions for T&C and Calorimeter-only were scaled to the same mean value in order to make the comparison meaningful, then fit with a Gaussian, starting at about 20% of the distribution height from the bottom. The resolution is defined as the Gaussian σ. The improvement corresponds to the relative decrease in σ between the Calorimeter-only reconstruction and the T&C reconstruction.

ISR and FSR were turned on; resolution and efficiencies were given the parameters described in section 3.2. On the Z^0 sample, two alternate calorimeter segmentations were studied besides the CDF one (finer EM seg.; finer EM and HAD seg.), as well as an improved HAD calorimeter resolution of $\frac{50\%}{\sqrt{E}} + 3\%$ (with CDF seg.)

Figures C.9 to C.12 show the distributions used to measure the resolution (before scaling). A summary of the results is showed in table C.3. The improvement for the dijet mass is less than for the jet : indeed the T&C jet distribution is not Gaussian

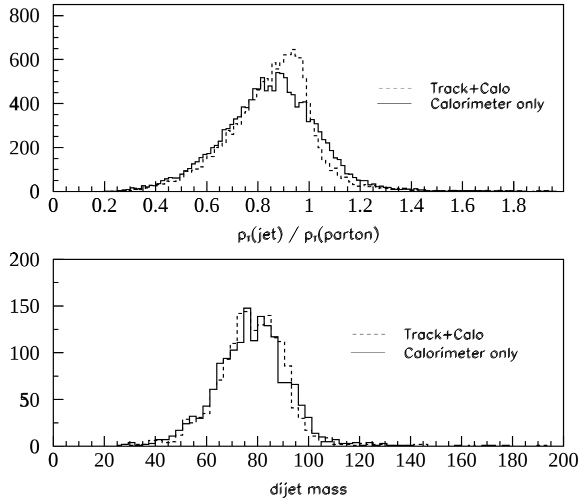


Figure C.10: *$Z^0 \to b\bar{b}$ with real resolution and efficiencies. CDF segmentation. ISR and FSR on.*

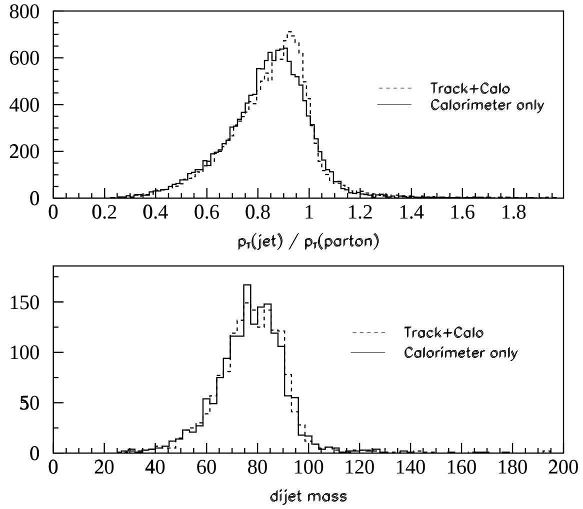


Figure C.11: *same as figure C.10 except HAD resolution* $= \frac{50\%}{\sqrt{E}} + 3\%$.

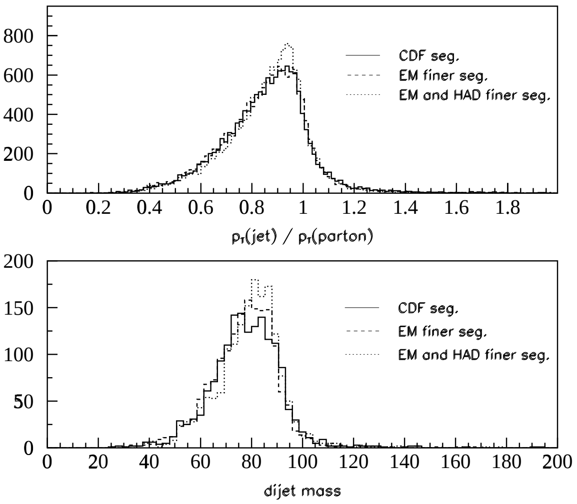


Figure C.12: *$Z^0 \to b\bar{b}$ with real resolution and efficiencies. T&C with 3 different segmentations : CDF seg., finer EM seg., finer EM and HAD seg. (0.05×0.05). ISR and FSR on.*

and has a long low tail which deteriorates the dijet reconstruction. The improvement for b-jets is less than for other jets, due to neutrinos : assuming CDF characteristics, the b-jet improvement (for $Z^0 \to b\bar{b}$) is 24% instead of 39% (for $W \to q\bar{q}$), while the b-dijet mass improvement is 11% instead of 23%. As expected, a better HAD calorimeter resolution makes the use of the tracking system less efficient : 9% instead of 24% improvement for the jet; 4% instead of 11% for the dijet mass (for $Z^0 \to b\bar{b}$). Finally finer segmentations (EM only or both EM and HAD) give a slightly better result by reducing particle multiplicity in each tower and thus fluctuations.

C.7 Conclusion

An upper limit has been put on the improvement in jet resolution one might get using the tracking system in a way similar to JCL98 [51]. With the CDF segmentation, the method is capable of an improvement of up to 39% in the jet resolution, and up

Table C.3: *Gaussian fit σ improvement with T&C reconstruction compared to Calorimeter-only reconstruction (after offset correction).*

	Jet p_T	Dijet mass
CDF segmentation :		
$W \to q\bar{q}$	39%	23%
$Z^0 \to b\bar{b}$	24%	11%
$Z^0 \to b\bar{b}$, better HAD res.	9%	4%
Other segmentations :		
$Z^0 \to b\bar{b}$, finer EM seg.	25%	17%
$Z^0 \to b\bar{b}$, finer EM and HAD seg.	33%	21%

to 23% in the dijet mass resolution for $W \to q\bar{q}$ events. In the case of $Z^0 \to b\bar{b}$, the performance is degraded, because of frequent neutrinos from b-hadron decays (up to 24% for jet, up to 11% for dijet mass); soft-lepton tagging might be able to recover some resolution.

The calorimeter segmentation appears to be of great importance for the method, as the particle multiplicity in each tower decreases with a finer segmentation.

However, some detector effects have been neglected in this study : especially, tower leakage, cracks, and energy from hadrons in the EM calorimeter were not simulated and are not favorable to the method because of the confusion they induce in the energy deposition pattern. It is clear that a very good understanding of the detector response is required if one wants to take full advantage of the tracking system.

Appendix D

Monitoring of the Silicon Vertex Detector with the SVXMon Software

D.1 Introduction

CDF monitors its data quality in real time using the so-called *Consumer Framework* [59]. In this framework, a fraction of events is made available for immediate analysis to a collection of programs known as "consumers" [60]. These programs run on various nodes in the CDF online computing cluster and process events which are continuously served over the network. The results, usually in the form of histograms, are sent to one or more online display programs run by experts or by the shift crew.

SVXMon keeps track of all problems found by the silicon bank unpacker and performs a variety of data integrity checks. For each silicon strip SVXMon accumulates the number of hits and the first four moments of the pulse height distribution. These statistics are used to create plots of occupancies, average pulse heights, distribution shapes, *etc.* The results can be viewed with any degree of detail desired, from layers and barrels down to silicon ladders, chips, and individual strips. The program accumulates statistics both since the beginning of the run and in a recent time window. SVXMon built-in data analysis capabilities can be used to diagnose various silicon tracker and/or DAQ problems online and to detect adverse changes in the running conditions. Offline analysis of SVXMon histograms and error reports allows for efficient tracking of silicon system problems from run to run. The following sections summarize the main features of SVXMon. More detailed documentation can be found elsewhere [57, 58].

D.2 Data Validation

Rather than checking the contents of the raw silicon banks directly, SVXMon works with container objects created by the silicon bank unpacker. The unpacker generates a diagnostic status word for every silicon half-ladder (an electrical module, *a.k.a.* HDI) included in the run, and SVXMon inspects these status words and counts the number of errors seen. The error statistics are maintained per HDI and error type. From time to time SVXMon generates corresponding error messages and sends them to the error logging facility described in Sec. D.3. The frequency of error messages of the same type generated for a given detector is dynamically prescaled in order to reduce the degree of message redundancy but still remind the users about the problems. At the end of each run, a summary message is generated for every HDI and error type with the total error count. The errors detected by the silicon unpacker and the corresponding messages produced by SVXMon are listed in Appendix D.7 together will all other types of SVXMon messages.

In a typical online monitoring configuration, SVXMon also performs a number of other checks of silicon data integrity in every event:

- Identifies readout chips whose pipeline has lost synchronization with the rest of the detector.
- Finds invalid ADC values present in the data stream due to chip malfunctions, optical transmission errors, or problems with pedestal subtraction in the Fiber

Interface Boards (FIBs).

- Monitors error words generated by VME readout buffer cards (VRBs).
- Checks the consistency of several DAQ-related quantities between SVX and L00/ISL (SIXD and ISLD banks).
- Verifies that all detectors included in the run have been read out by the DAQ (and that detectors not included are not read out).

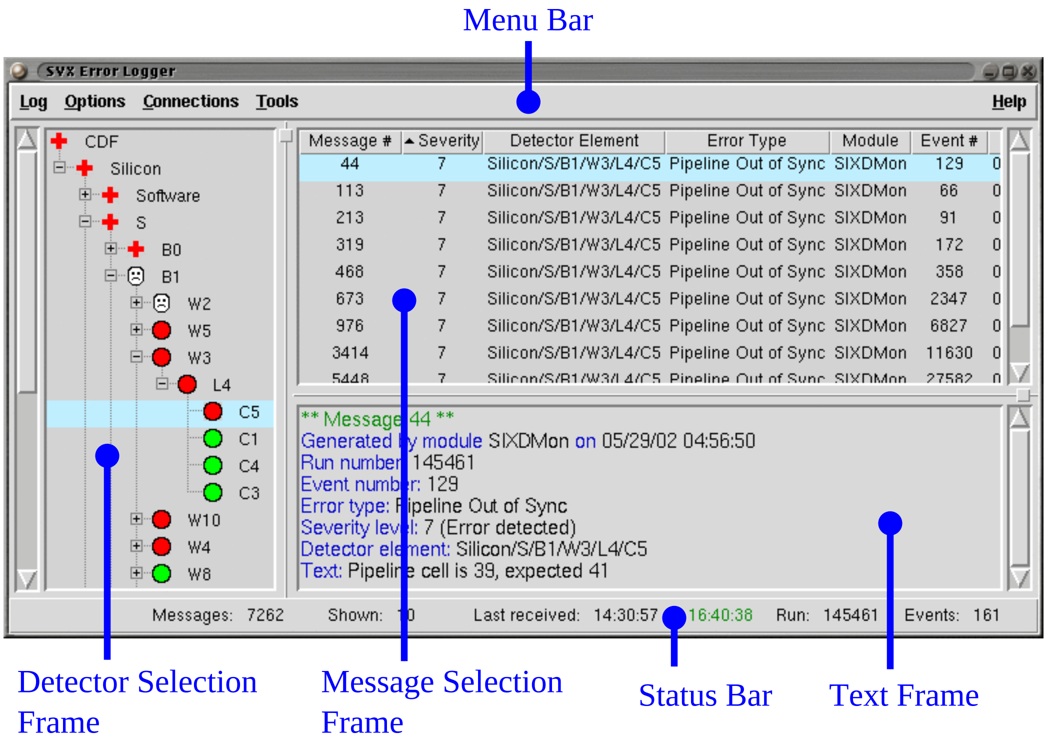
In addition to the deterministic checks listed above, SVXMon can perform an analysis of pulse height distributions and identify misconfigured or malfunctioning readout chips, problems with ladder bias voltage, *etc.* This analysis, however, depends on a set of cuts which requires a substantial amount of tuning. This topic is covered in more detail in Sec. D.5.

D.3 Error Reporting

SVXMon messages are received, reported, and stored by a Tcl/Tk [61] script called "SVX Error Logger". This script provides a convenient and intuitive GUI for collection and analysis of error messages produced by SVXMon and, potentially, by other CDF online monitoring programs. SVX Error Logger receives messages over a TCP/IP link and displays them to the user in a window reminiscent of the Netscape mail reader. The logger incorporates facilities for message viewing, searching, sorting, saving, restoring, distributing, and for generating run summary statistics. The SVX Error Logger program is compatible with the ZOOM Error Logger package [62] via an adapter class called SvxErrorLogger which sets up all necessary communications and performs the message transfers.

D.3.1 SVX Error Logger GUI

The SVX Error Logger runs in a separate window which can be displayed locally or over the network using X Window System. The components of the logger GUI window are illustrated in Fig. D.1. The detector selection frame shows a collapsible



Message #	▲ Severity	Detector Element	Error Type	Module	Event #	
44	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	129	0
113	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	66	0
213	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	91	0
319	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	172	0
468	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	358	0
673	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	2347	0
976	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	6827	0
3414	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	11630	0
5448	7	Silicon/S/B1/W3/L4/C5	Pipeline Out of Sync	SIXDMon	27582	0

```
** Message 44 **
Generated by module SIXDMon on 05/29/02 04:56:50
Run number: 145461
Event number: 129
Error type: Pipeline Out of Sync
Severity level: 7 (Error detected)
Detector element: Silicon/S/B1/W3/L4/C5
Text: Pipeline cell is 39, expected 41
```

Messages: 7262 Shown: 10 Last received: 14:30:57 16:40:38 Run: 145461 Events: 161

Figure D.1: *SVX Error Logger GUI*

hierarchical tree of detector systems and components for which at least one error message has been received. A small icon near the detector name shows the highest error severity for all messages received for this particular detector and all its subdetectors. Messages displayed in the message selection frame can be quickly sorted by number, severity, error type, *etc.*

The status bar at the bottom of the window displays the following (from left to right):

- Current program activity (or an empty string if the program is not doing anything).
- Total number of error messages received.
- The number of error messages shown in the message selection frame. The background color for this number may turn yellow. This means that the set of messages shown in the frame is no longer current. Click on the yellow field to

update the frame.

- The time when the last message arrived over the network. On start-up, this time is initialized to the program start time.
- Current time (in green).
- CDF run number.
- Number of events processed by SVXMon.

A GUI interface allows to search for specific error messages by selecting messages using different message fields and search types. The most recent search criteria will also be applied to the incoming messages until one of the systems in the detector selection frame or "Display Last Message" option are selected. The summary statistics for the accumulated error messages can be viewed by clicking on the "Statistics" entry in the "Tools" menu. The program generates several tables of error counts and pops up a new window which displays these tables.

Error messages about ladders with known problems may be suppressed in the summary by loading a list of bad ladders. The "Known Pb's" entry in the "Log" menu can be used to load the list and the "Hide Known Problems" button turns on/off the masking system. The file containing the list of bad ladders is a simple text file containing one FIB id per line (in hex). Empty lines are ignored. Here is an example of such a list:

```
e500
e501
f140
```

During normal online operation, the list of known problems is updated automatically when SVXMon starts, based on the contents of the Silicon Problem Database [65]. This list is used by SVXMon for masking chips with known problems on the pipeline status map (Sec. D.4.5) and the chip status map (Sec. D.5).

The "Generate HTML" button may be used to write the contents of all tables into an HTML file. A pop-up dialog window allows choice of the file name. This file can be later viewed with a web browser. A different type of summary can be generated by clicking on the "Run Summary" button in the "Tools" menu. This summary is a simple text file which can be later printed on a printer or viewed in a terminal. A pop-up dialog window will allow to choose the file name. A special file name "console" may be used to display the summary information in the text frame instead of writing it to disk. In this summary, the error messages are sorted by the detector element, for example:

```
SB3W7L4
Bit Error (not 0/1) | 1 | This error was present in 1 events out of 156
Unknown Channel | 1 | This error was present in 1 events out of 156
ff or 7f Error | 2 | This error was present in 2 events out of 156
```

After the detector element, all error types encountered for this element are listed together with the number of errors seen (separated by vertical bars) and the text of the last message of this type. For SVXMon, this last message is usually (but not always) a summary message generated at the end of a run. When SVX Error Logger works online in tandem with SVXMon, it automatically generates three text files at the end of each CDF run. The files are named like this:

```
svxmon_RRRR_hhhh_PPPP.errlog
svxmon_RRRR_hhhh_PPPP.errsum
svxmon_RRRR_hhhh_PPPP.errtypes
```

where "RRRR" stands for the run number, "hhhh" for the host name of the machine on which the logger runs (with domain name stripped off), and "PPPP" for the process number of the logger. This naming convention ensures that several error loggers running simultaneously will not use the same file name. The file with extension ".errlog" contains the full list of error messages, the ".errsum" file contains the error summary by detector elements, and the ".errtypes" file stores the error summary by error type.

At the time of this writing, only the summary statistics GUI and the HTML summary support suppression of known problems by detector element. The text summaries always use the full set of messages.

D.4 SVXMon Plots

SVXMon can create four types of plots: long-term, short-term, history, and periodic. Long-term plots display various silicon tracker statistics accumulated since the beginning of the run. Short-term plots display statistics accumulated during the last few minutes of data taking or over a few most recent events. These short-term plots are often called "snapshot plots" in the rest of this document. History plots can show the time history of any quantity for which there is either a long-term or a short-term plot. Periodic plots display sequences of various silicon quantities averaged over a fixed number of events. These plots are similar to history plots in purpose but occupy less memory, and creation of the relevant ROOT histograms is delayed until the end of a job.

Each SVXMon plot is usually placed on its own ROOT canvas which provides some useful auxiliary information such as the CDF run number, the number of events accumulated so far by SVXMon, plot update time, *etc.*

SVXMon plots may be booked either at the beginning of a job in its configuration file or in the middle of a run using an interactive prompt. All plots and histograms are created with a special command named "histo" added to the tcl interpreter by SVXMon. The rest of this section gives a non-exhaustive description of the "histo" command and the objects it creates.

D.4.1 Strip Plots

SVXMon can create several types of plots in which various silicon-related quantities are monitored by strip. One-dimensional plots of some quantity vs the strip number may be created by the following command:

histo strips *detector histoTypes makeSnapshots*

Here, *detector* is a five-element tcl list which specifies one face of a silicon half-ladder [1]. *histoTypes* specifies the quantities to plot. It must be a tcl list which contains one or more of the following keywords:

"occupancy" — Plots strip occupancy in percent vs the strip number.

"vrbocc" — Plots VRB (level 1) strip occupancy in percent (the CDF silicon DAQ may require a configuration adjustment to provide an access to this information).

"nevents" — Plots the number of events in which a strip has registered a hit. This plot differs from the occupancy plot only by normalization.

"mean" — Plots average strip pulse height in ADC counts.

"stdev" — Width of the pulse height distribution in ADC counts.

"skewness" — Pulse height distribution skewness.

"kurtosis" — Pulse height distribution kurtosis.

"bad" — Plots 1 for a "bad" strip, 0 for a "good" strip.

"discarded" — Plots 1 for a discarded strip and 0 for a strip which doesn't have the "discard" tag.

"newbad" — Plots 1 for strips which have the "bad" tag but do not have the "discard" tag. This plot and the next one may become useful in the future if a real-time bad strip diagnostics is implemented in SVXMon.

"newgood" — Plots 1 for strips which have the "discard" tag but do not have the "bad" tag.

"dnoise" — Plots dnoise in ADC counts. Dnoise is the standard deviation of the pulse height difference between adjacent channels divided by $\sqrt{2}$. This plot

---

[1] Many SVXMon commands take an argument (usually called *detector* in this note) which specifies one side of a silicon half-ladder. This argument is a five-element tcl list {*barrel ladderSegment phiWedge layer side*}. Allowed values for *barrel* and *ladderSegment* are "west", "east", and "center". Allowed values for *side* are "z" and "phi". *layer* and *phiWedge* are integers. *layer* should be set to 0 for L00, to 1 for SVX layer 0, and so on, up to 7 for the outer ISL layer.

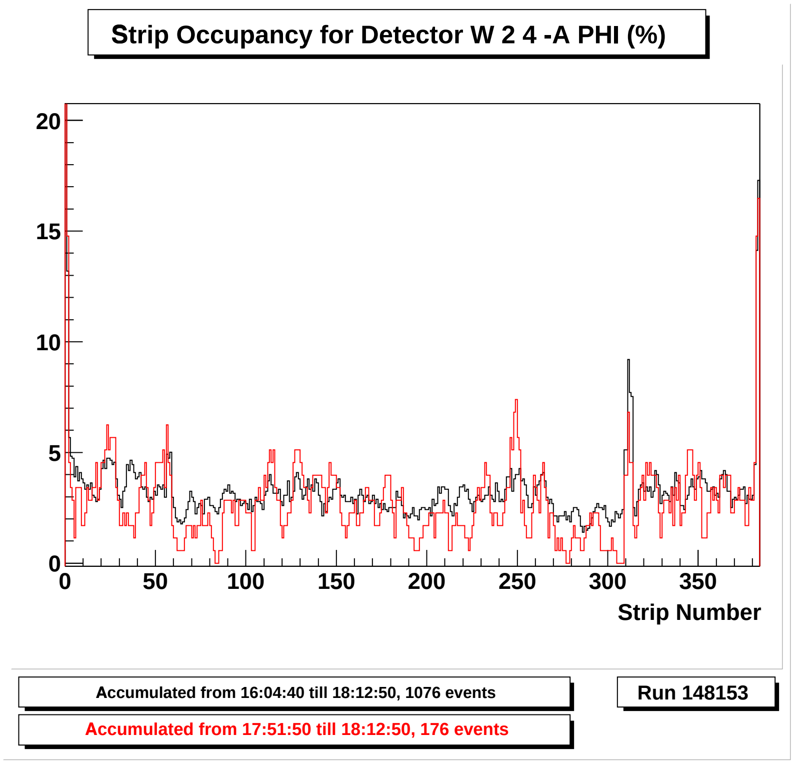


Figure D.2: *Example plot of occupancy vs strip number for the phi side of the SVX half-ladder in the west barrel, west half-barrel, layer 2, wedge 4.*

makes sense only when the half-ladder is used in “read all” mode.

The command will create as many plots as the number of elements in the *histoTypes* list. *makeSnapshots* is a boolean argument which specifies if a short-term plot should be overlayed on top of the long-term one. An example of plots produced by the **histo strips** command is shown in Fig. D.2.

In a typical online monitoring configuration, SVXMon creates plots of strip occupancies, mean charge, and RMS charge for every half-ladder side in SVX/ISL and every z segment in L00. Of course, neither experts nor consumer operators are expected to view about 4000 such plots. Rather, the strip data are processed with SVXMon online (Sec. D.5) and offline (Sec. D.6) quality control algorithms, and the strip-level information for some ladders may be reviewed when problems are detected.

SVXMon can also create 2-d plots of certain quantities vs the strip number on the half-ladder side and another coordinate defined by the value of certain DAQ-related parameter. These plots are booked with the following command:

histo strips2d *paramName detector histoTypes snapshotStrategy*

paramName determines the quantity to use for the y histogram axis (the strip number will be used for the x axis). Valid *paramName* values are:

"adc" — used to produce pulse height distributions
"cellid" — pipeline cell id
"bemode" — back end mode of the SVX readout chip
"bunchx" — SRC bunch crossing number
"tl1a" — time since the last L1 accept

detector is the standard SVXMon half-ladder side specifier. *histoTypes* specifies the list of quantities to histogram. The following keywords may be used in this list: "occupancy", "nevents", "mean", "stdev", "skewness", "kurtosis", and "dnoise". The meaning of the keywords in the same as in the **histo strips** command. Note that specifying "dnoise" as one of the quantities only makes sense when the detector is used in "read all" mode. *snapshotStrategy* is an integer which defines how many snapshots of the 2-d dataset will be used. An example pulse height distribution created with the **histo strips2d** command is shown in Fig. D.3, and a plot of RMS charge for each pipeline cell of a readout chip is shown in Fig. D.4.

The **histo strips2d** command returns a histogram handle command [2]. The only useful thing you can do with the returned handle is to suppress all histogram entries for one of the parameter values. This feature is useful for displaying data from so-called "deadtimeless scans" when two L1 accepts are sent to the silicon front-end at a set of predefined intervals in order to study the effect of DAQ signals on the pedestal and noise. Here is an example of the handle usage:

---

[2] We use the term *handle command* or simply *handle* to describe tcl commands which provide an interface to functions of a particular object rather than a whole class of objects. A typical usage of an SVXMon handle is "$handle configure parameterName" or "$handle cget parameterName" where "parameterName" is a name of some parameter-like object member.

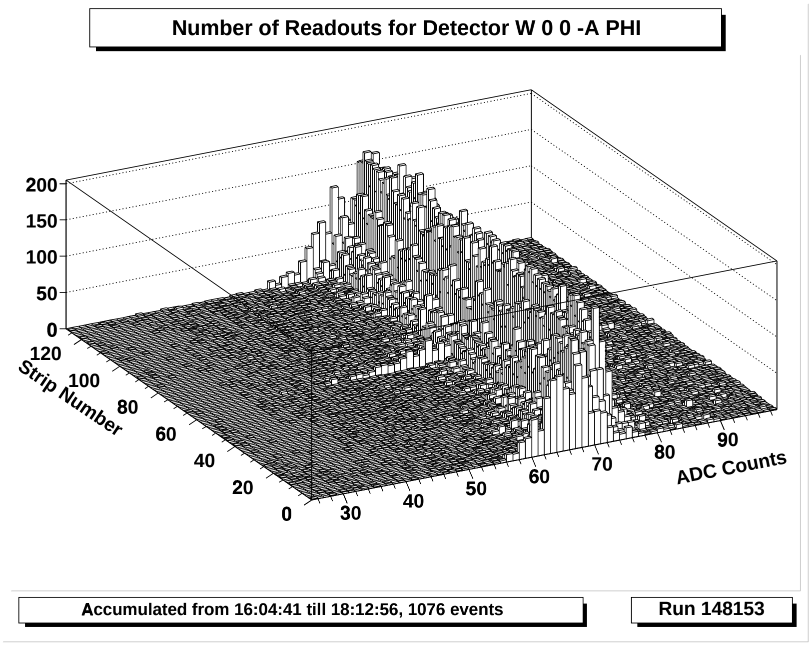


Figure D.3: *Example pulse height distribution for the L00 westmost sensor in wedge 0. The SVXMon command "histo strips2d adc {west west 0 0 phi} {nevents} 0" has been used to book this plot.*

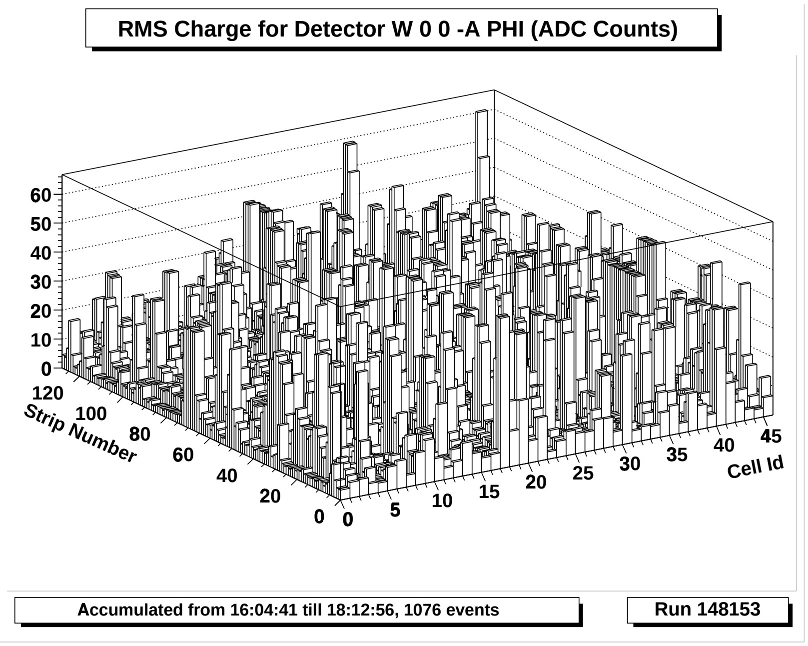


Figure D.4: *Example plot of RMS charge vs the strip and pipeline cell numbers for the L00 westmost sensor in wedge 0. This plot has been booked with SVXMon command "histo strips2d cellid {west west 0 0 phi} {stdev} 0".*

```
set detector {west west 4 2 phi}
set histoTypes [list nevents mean stdev dnoise]
set handle [histo strips2d bunchx $detector $histoTypes 0]
$handle configure skipParamValue 5
```

D.4.2 Occupancy, Mean Charge, *etc.* vs DAQ Parameters

If, for a given half-ladder, you are only interested in the dependence of some silicon quantity (such as occupancy or mean collected charge) on some DAQ parameter and not on the strip number, you can use the **histo daqparam** command:

histo daqparam *paramName detector histoTypes snapshotStrategy*

Histograms booked by this command are essentially sums or averages over all strips of corresponding **histo strips2d** type histograms described in the previous subsection. The command arguments and their meaning is the same as for the **histo strips2d** command. The only exception is that the keyword “nhits” in the list of plot types *histoTypes* should be used instead of the keyword “nevents” in case the user wants to see the total number of channels read out. The command returns a handle which can be used to suppress one of the histogram bins. It works in a similar way to the handle returned by the **histo strips2d** command. Example **histo daqparam** plots are shown in Fig. D.5 and D.6.

Another useful DAQ-related plot can be booked with the command

histo bxtrace

This command takes no additional arguments. It creates a stripchart plot of the bunch crossing number vs the event number which is useful for certain silicon DAQ studies. An example plot is shown in Fig. D.7.

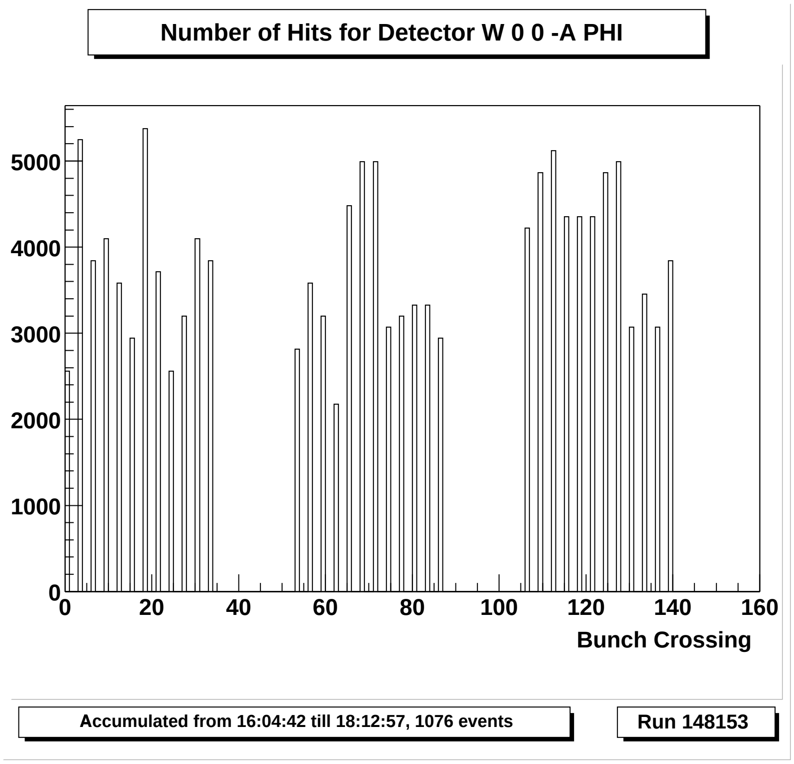


Figure D.5: *Example plot of number of channels read out vs the SRC bunch crossing number for the L00 westmost sensor in wedge 0. Together with the plot on the next figure, it has been booked with the command "histo daqparam bunchx {west west 0 0 phi} {nhits mean} 0". Since L00 is used in "read all" mode, this plot essentially displays the distribution of bunch crossing numbers (up to a constant multiplication factor).*

D.4.3 Chip Plots

Silicon quantities averaged over all good strips in each chip can be viewed on the plots created by the following command:

histo chips *histoTypes makeSnapshots*

This command creates several ROOT canvases which display a set of monitoring plots in a tabular arrangement. Each histogram shows some quantity of interest vs the chip number on a particular barrel, layer, and sensor side. The *histoTypes* argument specifies the list of quantities to plot. The meaning of the keywords in this list is quite similar (but not identical) to the meaning of the corresponding keywords in the strip-level plots:

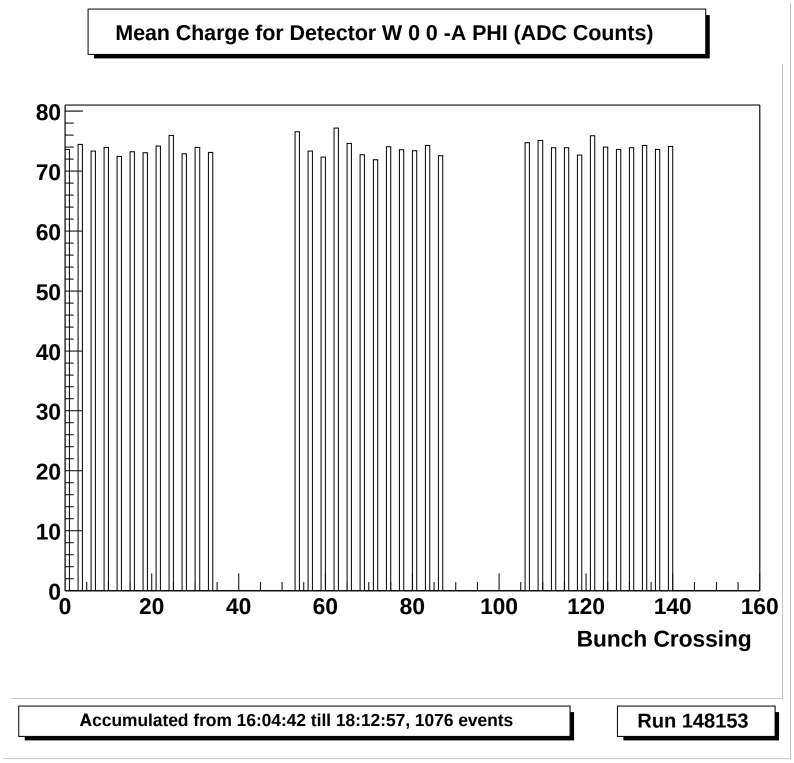


Figure D.6: *Example plot of mean read out charge vs the SRC bunch crossing number for the L00 westmost sensor in wedge 0. Since the mean charge is relatively flat across all bunch crossings present in the run, we can deduce that the front-end preamp reset signal has been safely confined to beam gaps.*

“occupancy” — Plots average strip occupancy for a given chip, in percent.

“vrbocc” — Plots average VRB (level 1) occupancy for each chip, in percent (the CDF silicon DAQ may require a configuration adjustment to provide an access to this information).

“nevents” — Plots the number of hits read out with a given chip. This plot differs from the occupancy plot only by normalization.

“mean” — Plots average pulse height for the hits read out with a given chip.

“stdev” — Plots the width of the pulse height distribution for the hits read out with a given chip. Note that this quantity is different from the average strip noise because it also includes strip-by-strip pedestal variations.

“skewness” — Pulse height distribution skewness.

“kurtosis” — Pulse height distribution kurtosis.

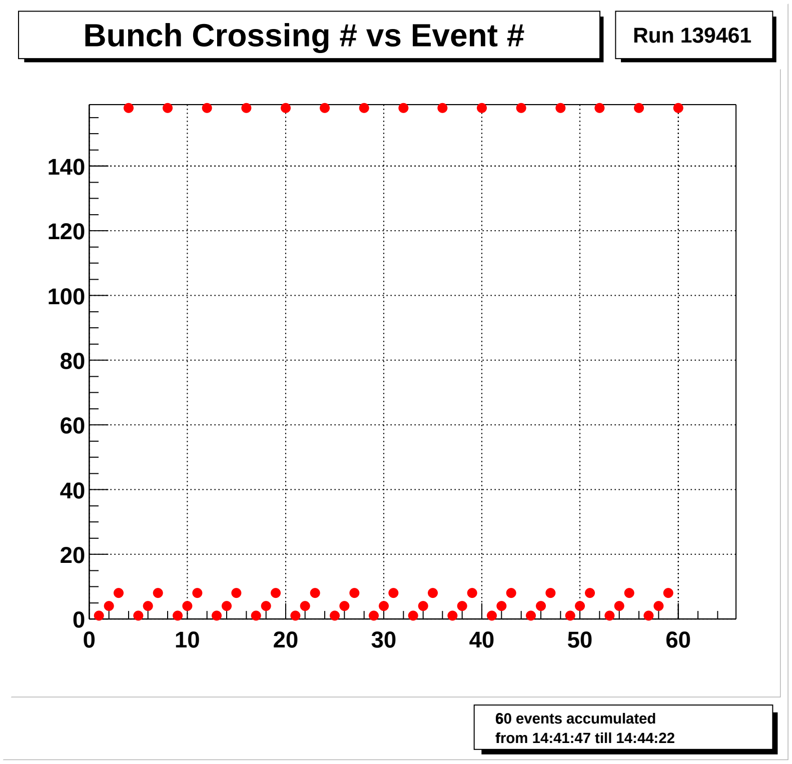


Figure D.7: *Example plot of the bunch crossing number vs the event number.*

“bad” — Plots the number of bad strips connected to a given chip.

“discarded” — Plots the number of discarded strips for a given chip.

“newbad” — Plots the number of strips which have the “bad” tag but don’t have the “discard” tag. This plot and the next one may become useful in the future if real-time bad strip diagnostics is implemented in SVXMon.

“newgood” — Plots the number of strips which have the “discard” tag but don’t have the “bad” tag.

“dnoise” — Plots the width of the distribution of pulse height differences between adjacent channels, divided by $\sqrt{2}$. This quantity is different from the average dnoise because it also takes into account strip-by-strip pedestal variations. This plot makes sense only when the whole silicon system is used in “read all” mode.

makeSnapshots is a boolean argument which should be set to 1 in order to build the short-term plots. These plots will be placed on separate canvases rather than overlayed on top of the long-term ones because the number of plots on a chip canvas is

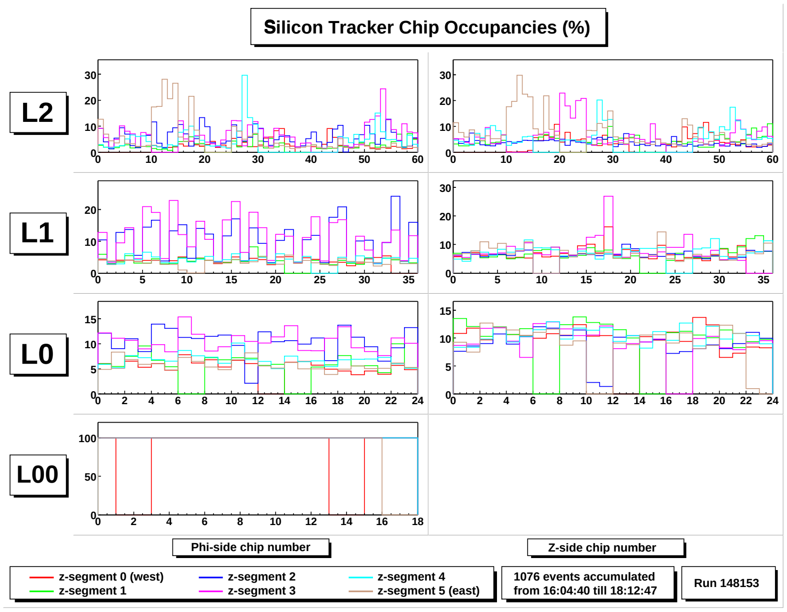


Figure D.8: *Example chip occupancy plots for layers L00, L0, L1 and L2.*

already quite high (up to 48). As it was already mentioned for the strip-level plots, the value of *makeSnapshots* argument should be consistent with the settings of SIXDMon parameters "snapshotStrategy" and/or "vrbOccupancySnapshots" (Sec. **??**).

Example chip occupancy is shown in Fig. D.8. In this plot, the chip numbers on the horizontal axis increase with increasing phi wedge number, and within each wedge chip numbers increase in the readout order.

D.4.4 Tracking Plots

In order to increase the event processing rate, SVXMon does not attempt to run silicon tracking algorithms in a typical online monitoring configuration. However, it can plot certain quantities which characterize silicon tracking performance using the information provided by the Level 2 trigger tracker (SVT) and/or generated by silicon track reconstruction during data processing in the Level 3 trigger[3]. The tracking plots are booked in SVXMon with the following configuration command:

histo tracks *algorithmName*

---

[3] At the time of this writing, silicon tracking is not yet part of Level 3.

This command creates plots which show the number of reconstructed tracks per event per half-ladder (track occupancy), average track residuals, and/or χ^2 for all sensors in the silicon tracker. *algorithmName* is an optional string which describes the algorithm used for track reconstruction. Valid algorithm names are "OutsideInAlg", "Regional", and "SVT".

The **histo tracks** command returns a handle which can be used for tuning of the plot behavior. Valid plot parameters which can be configured by the handle are listed below, together with their default values in parentheses:

nPhiHitMinG, nPhiHitMinB (3, 3)

Two different sets of cuts can be defined: suffix B in parameter names is used for "bad" tracks, suffix G for "good" tracks. Cuts are made on the transverse momentum, p_T, and the number of z and phi side hits of the reconstructed track. Parameters **nPhiHitMinG** and **nPhiHitMinB** specify the minimum number of phi side hits for "good" and "bad" tracks, respectively.

nZHitMinG, nZHitMinB (2, 2)

Minimum number of z hits.

ptHMinG, ptLMinB (0.5, 1.0)

Minimum p_T.

ptHMaxG, ptLMaxB (500.0, 500.0)

Maximum p_T.

debugFlag ("off")

Turns debug printouts on or off.

Examples of such plots are shown in Fig. D.9 and D.10.

D.4.5 Pipeline Plots

The plots which provide information about the coherency of the silicon analog pipeline are created with the tcl configuration command `histo cellid`. Three dif-

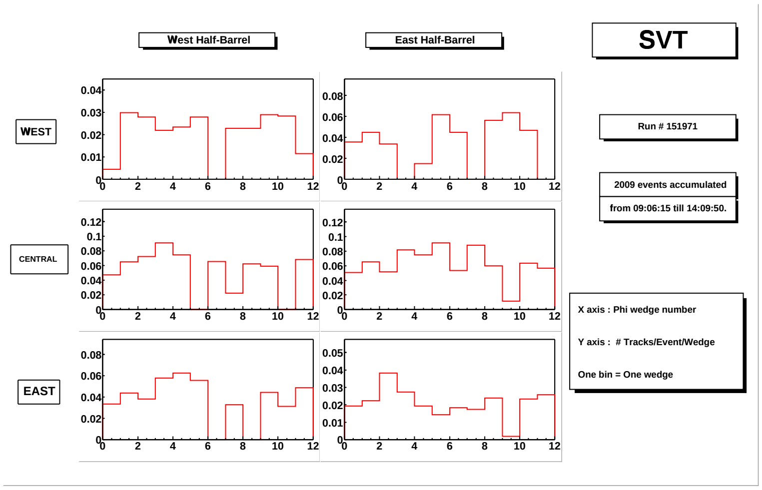


Figure D.9: *SVT Tracking Occupancy.*

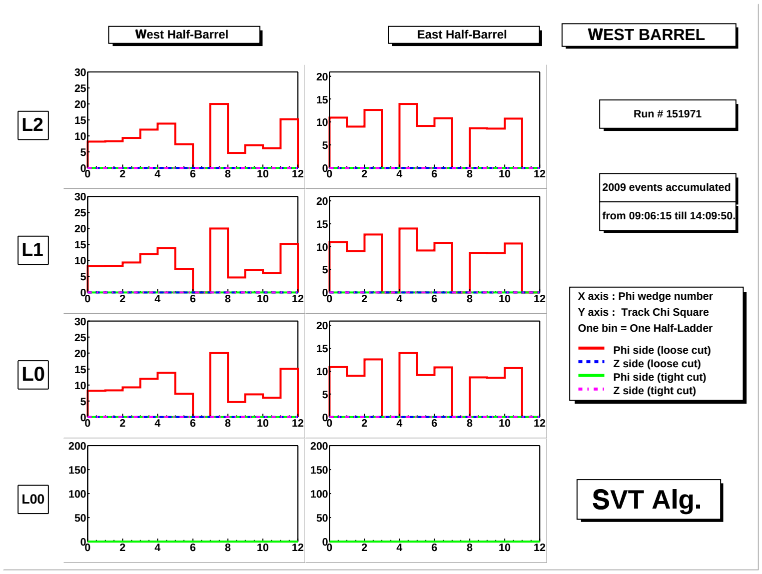


Figure D.10: *Example of track quality plot (here for SVT algorithm). When the SVT algorithm is chosen, this canvas shows the average* χ^2 *of the tracks. For the other algorithms, it shows the average residual of hits on tracks.*

ferent types of plots may be booked:

histo cellid global

Creates a histogram of the “correct” pipeline cell ids and turns the pipeline cell id mismatch monitoring on. The “correct” cell id is defined to be the cell id most often seen in the data stream. Although in theory this definition may fail to identify the "true" correct cell id in case of some pathological failure, in practice it is never a problem because the pipeline stays synchronous through a typical run on well over 50% of the front-end chips. Pipeline cell ids below 0 or above 45 should not be produced by a functional chip even if it is out of sync with the rest of the detector. Such cell ids indicate some kind of a hardware failure, most often it is a sign of an optical transmission error.

histo cellid *detector chipNumSensor*

Creates the histogram of the cell ids encountered in the data stream for a given chip. The canvas also includes a plot of the difference between the chip histogram and the histogram of the "correct" cell id. An example canvas is shown in Fig. D.11. The command arguments have the following meaning: *detector* is the standard SVXMon half-ladder side specifier, and *chipNumSensor* is the chip number on the sensor side, in the readout order. Note that it is not necessary to call the **histo cellid global** command explicitly in case the cell id distribution is booked for at least one chip.

histo cellid map

This command creates a map in which silicon readout chips are color coded according to their pipeline status. Each chip on the map is shown with a small rectangle. The rectangle colors are associated with the pipeline status in the following way:

green — status is "good", pipeline is synchronized.

red — "bad", the chip has lost its synchronization (could also be a persistent optical transmission problem).

yellow — "no data", the bank unpacker was never able to read the pipeline cell id for this chip.

blue — "disabled", either a known problem or the chip was not included in the run.

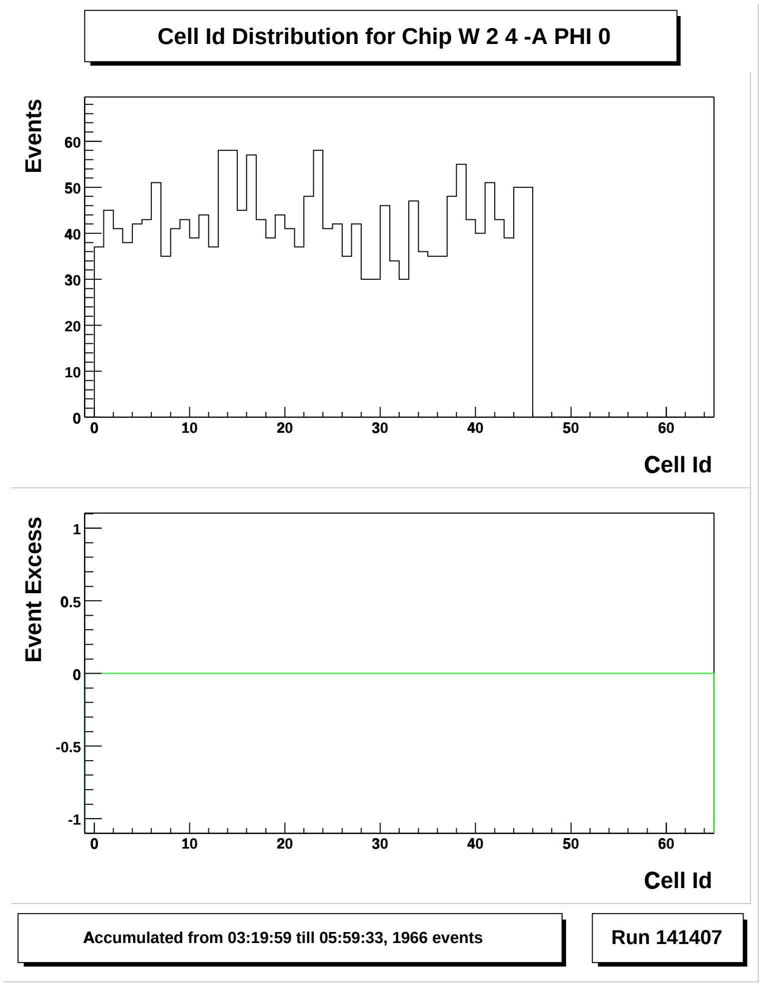


Figure D.11: *Example plot of chip cell id distribution. This chip's pipeline appears to be in sync with the rest of the detector.*

The map is split between two canvases: one canvas is used for SVX and the other for L00/ISL. An example SVX map is shown in Fig. D.12. This command also creates a histogram used to accumulate the distribution of the number of chips whose pipeline has lost synchronization. An example distribution is shown in Fig. D.13. SVXMon is set up to send a request to the main DAQ system to automatically re-initialize the silicon detector if the number of silicon readout chips with desynchronized is above a certain threshold.

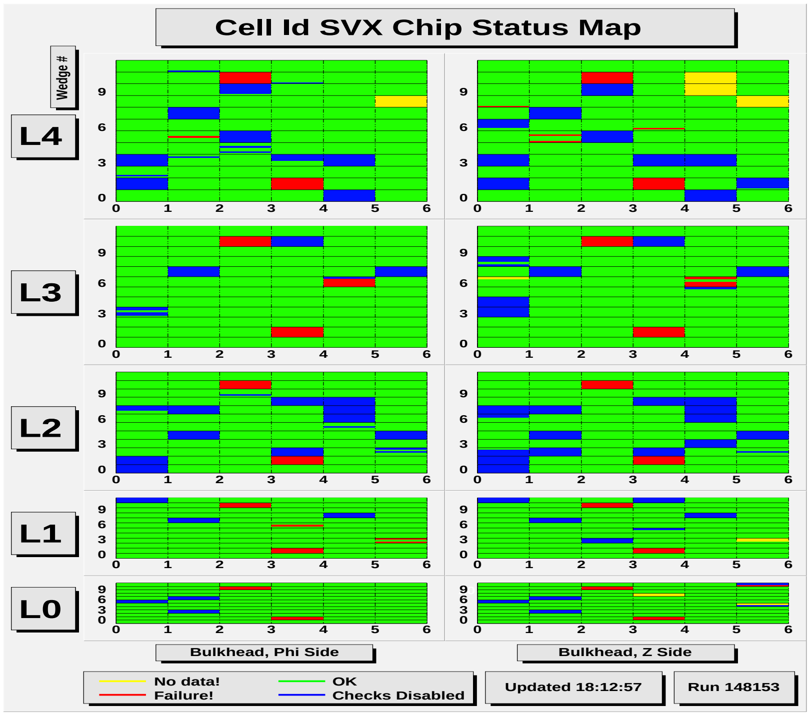


Figure D.12: *Example SVX pipeline status map.*

D.4.6 History Plots

SVXMon can monitor and display the time history of any of its long-term or short-term plots. The quantities shown on the history graphs can be values of selected bins of monitored plots or values of various plot statistics (mean, median, standard deviation, *etc*). The following command creates a history plot:

histo history *plotTitle listOfRecords*

The *plotTitle* argument specifies the title which will be displayed on the plot canvas. *listOfRecords* is a tcl list of history record identifiers. Each identifier has to be obtained beforehand by calling the "svx watch" command. The graphs for all history records in the list are overlayed on one plot. The number of elements in the list should not exceed 14 (this limit is, essentially, the number of distinct plot markers supported by ROOT). Example history plots are shown in Fig. D.14. Please see the description of SIXDMon parameter "historyFrequencyDivider" in Sec. **??** for details about history timing.

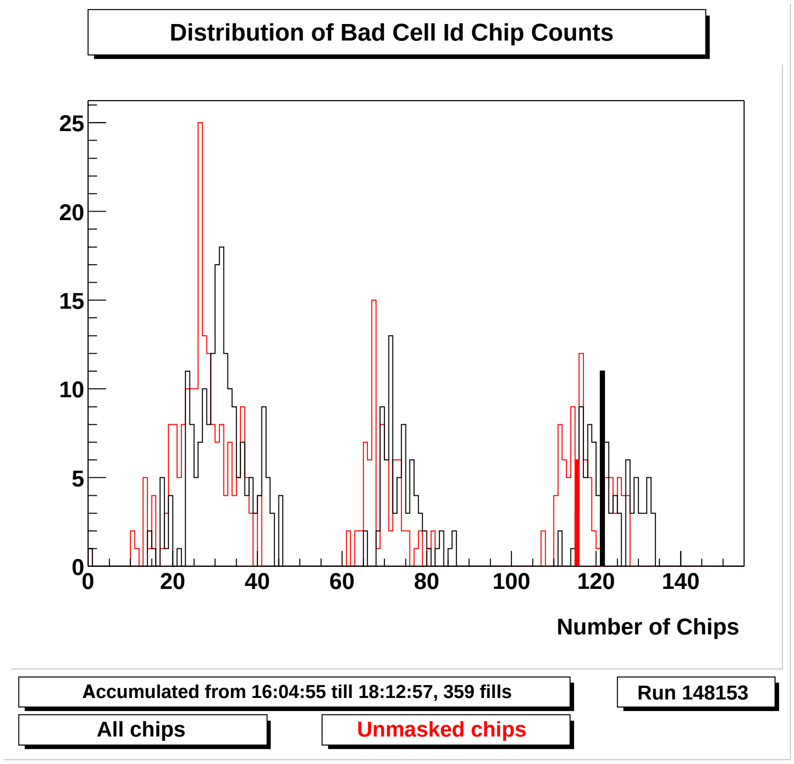


Figure D.13: *Distribution of the number of chips whose pipeline has lost synchronization. One entry is added to this histogram every time SVXMon plots are updated. The two solid bins are the ones most recently updated.*

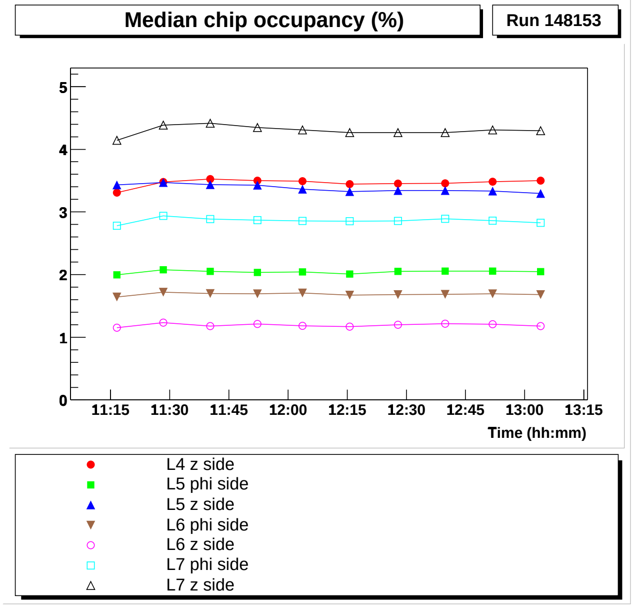


Figure D.14: *Example history plots of median chip occupancy by layer and side.*

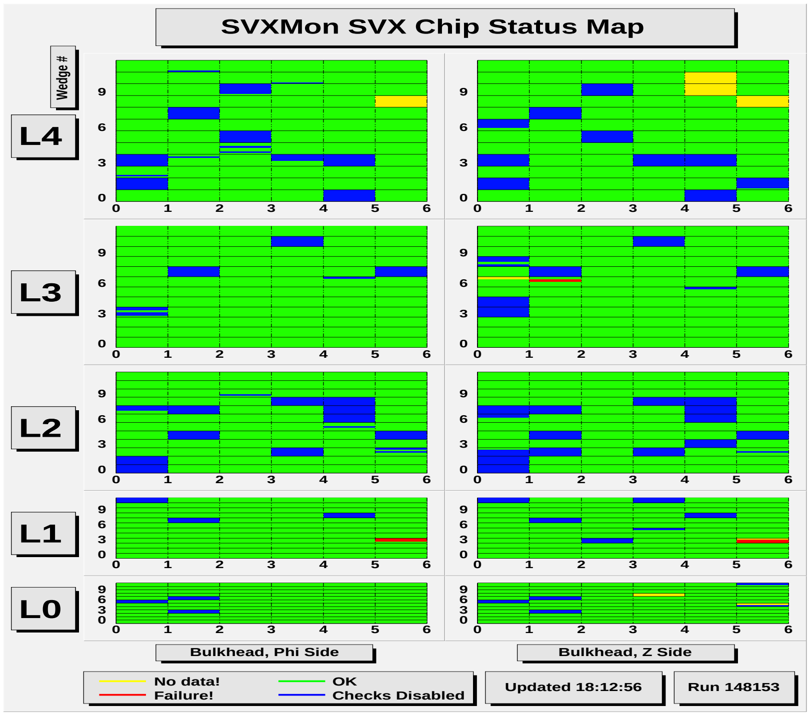


Figure D.15: *Example SVX chip status map.*

D.5 Online Quality Control

SVXMon can be configured to generate error messages when such silicon quantities as occupancies or pulse height averages are not consistent from chip to chip or deviate significantly from their expected values. The basic underlying assumption of the error checking algorithm is that most chips in the silicon system perform well, and the typical values of various quantities should be similar for chips with identical settings and capacitive loads. All chips in the detector are divided into groups based on their layer number, sensor side, dynamic pedestal suppression setting, *etc* (the program supports arbitrary aggregation of readout chips into groups). The median and range are estimated for the distributions of each quantity of interest (such as occupancy) over all member chips in the same group. The chips which deviate significantly from the group medians are reported to the error logger. The problems are also reflected on the chip status map plots. An example map of this type is shown in Fig. D.15.

The automatic chip monitoring facility can be launched from either the configuration file or the SVXMon special prompt by the following command:

histo check *histoTypes makeSnapshots*

Here, *histoTypes* is a tcl list of quantities to monitor. Each element of this list must be chosen from the following set of allowed keywords: “occupancy”, “vrbocc”, “nevents”, “mean”, “stdev”, “skewness”, “kurtosis”, “bad”, “discarded”, “newbad”, “newgood”, and “dnoise”. The keywords have the same meaning as in the “histo chips” command (Sec. D.4.3). *makeSnapshots* is a boolean argument which specifies whether SVXMon can include short-term statistics in its determination of silicon data quality. Set it to 1 in order to use the short-term statistics.

The command returns a handle which is used for tuning the analysis algorithm. The handle supports the following options:

$handle configure *paramName value*

Sets the value of *paramName* to *value* if *paramName* is a valid parameter name and *value* is an acceptable definition for this parameter. This command returns an empty string.

$handle cget *paramName*

Returns the value of parameter *paramName* if *paramName* is a valid parameter name.

$handle parameters

Returns the list of valid parameter names.

$handle paramtable

Prints parameter values to the standard output in a tabular form.

$handle array *action paramName ?valueList?*

An optimized interface for array parameters. It is identical to the TclModule array interface described in detail in Ref. [67].

$handle chipgroup *action groupName ?chipList?*

This command is used to define, enable, disable, or view chip groups, depending on the *action* argument:

$handle chipgroup set *groupName chipList* — defines the set of chips which belong to the named group. The group name may be an arbitrary string, but it must not contain commas (because of the way group cuts are parsed — see the description of "groupcut" parameter later in this section). Each chip in the list of chips is specified as a list {*detector chip*} where *detector* is the five-element half-ladder side specifier and *chip* is the chip number on this half-ladder side in the readout sequence. Each chip may belong to more than one group.

$handle chipgroup list — returns the list of group names defined so far.

$handle chipgroup state *groupName newState* — sets the state of the named group. The *newState* argument must be either "normal" or "disabled". The disabled groups will be ignored by SVXMon, and the program will not produce error messages if some group quantity goes out of limits. This command is useful in order to quickly suppress a flood of messages from a noisy or misconfigured chip group while SVXMon is running. When a new group is created by the **chipgroup set** option, it is alway created in the "normal" state.

$handle chipgroup state *groupName* — shows the state of the named group.

$handle plotgroups

This command creates group plots of the quantities of interest, one bin per group. This command should appear in the configuration file just once, after all group names have already been defined.

The behavior of the monitoring algorithm depends on the values of the following parameters (default values are given in parentheses):

dynamicCuts (0)

The value of this parameter specifies which scheme will be used to determine allowed regions for various chip group statistics. 0 means use static cuts (defined by the **groupcut** parameter) and 1 means that the cut values will be determined dynamically every time a cut is applied by evaluating one of the **limitProcs** tcl procedures.

groupcut (0.0)

This parameter is an array of doubles. This array is used to define cuts for group quantities, such as chip occupancy medians. The indices of this array must be constructed as follows: `groupcut($cutName,$isSnapshot,$groupName)`. `$cutName` is a string which looks like "xxxxx_high" or "xxxxx_low" where xxxxx stands for the name of the quantity monitored (any histogram type keyword accepted by the "histo check" command). `$isSnapshot` must be 1 or 0. This parameter specifies whether the cut will be applied to the statistic accumulated over a short period (when the parameter value is 1) or since the beginning of the run (when the value is 0). `$groupName` is the name of the chip group for which the cut is defined. If some cut is not defined explicitly in this array, it will be automatically set to 0. The cuts defined in this array will be used only if the **dynamicCuts** parameter is set to 0.

limitProcs ("proc dummyErrLimitProc {args} {list −1.0e12 1.0e12}")

This parameter is an array of strings which define tcl procedures used to calculate cuts on group quantities dynamically. The indices of the array are the names of the monitored quantities (any histogram type keyword accepted by the "histo check" command). Each procedure will be called with three arguments: $nEvents, $isSnapshot, and $groupName. It must return a two-element list of doubles {*$lower_limit $upper_limit*}. This mechanism allows for adjustment of cuts during the run depending on the number of events processed, elapsed time, luminosity, *etc.* The cuts defined in this way will be used only if the **dynamicCuts** parameter is set to 1.

chipcut (0.0)

This is an array of doubles used to define cuts for chip quantities, such as the chip occupancy and its deviation from the median occupancy in the group. The array indices look like this: `chipcut($cutName,$isSnapshot,$detector,$chip)`. `$cutName` is a string which must look like "xxxxx_limit_low", "xxxxx_limit_high", "xxxxx_nsigma_low", and "xxxxx_nsigma_high" where xxxxx stands for the name of the monitored quantity. Example:

```
set detector {west west 0 1 phi}
$handle configure chipcut(occup_limit_low,1,$detector,0) 0.5
$handle configure chipcut(occup_nsigma_low,1,$detector,0) 5
$handle configure chipcut(occup_limit_high,1,$detector,0) 10
$handle configure chipcut(occup_nsigma_high,1,$detector,0) 5
```

This sequence of commands instructs SVXMon that the snapshot occupancy of chip 0 on the given ladder should be above 0.5% and below 10%, and it should not deviate from the median snapshot occupancy of the group by more than 5 "sigma". The "sigma" will be in fact determined in a robust manner as the interquartile occupancy range times 0.7413011.

minimalRange (0.001)

Array of doubles. In some degenerate cases (for example, occupancy in read-all mode) the group interquartile range of a monitored quantity may become 0 in which case it becomes impossible to define range multipliers which would allow chips with small deviations from the norm to pass the cuts. In order to be able to allow such chips to pass, we have introduced the lower bound on the range. The array indices should look like this: "xxxxx" (for run histograms) or "inst xxxxx" (for snapshots) where xxxxx is the name of the monitored quantity.

skipChip (0)

Array of boolean values used to enable or disable data quality monitoring chip-by-chip. The indices should look like this: `skipChip($detector,$chip)`. The value of 0 means enable checks. It is a good idea to disable checks for all chips not included in the run.

errorSeverity ("ELerror")

The array of strings which allows the user to change the error logger severity levels of various "Out of Limits" error messages issued when some group quantity gets out of limits. The indices look like this: "Xxxxx Out of Limits" or "Inst xxxxx Out of Limits", where xxxxx is the name of the monitored quantity (the first letter of an index is always in upper case). There are also several special indices: "No Data", "No VRB Occupancies", and "Wrong VRB Data Size". The valid severity levels are "ELincidental", "ELsuccess", "ELinfo", "ELwarning", "ELwarning2", "ELerror", "ELerror2", "ELnextEvent", "ELunspecified", "ELsevere", "ELsevere2", "ELabort", and "ELfatal".

chipSeverity ("ELerror")

The array of strings which allows the user to change the error logger severity levels of various "Out of Limits" error messages issued when some chip quantity gets out of limits. The indices look like this: "Xxxxx Out of Limits" or "Inst xxxxx Out of Limits", where xxxxx is the name of the monitored quantity.

maxEventsNoLadder (5)

The maximum number events which SVXMon will allow to pass before is starts reporting problems about chips not found in the data stream (after bank unpacking). Note that if the chip is included in the run but not found in the data stream, this doesn't automatically mean that there is a real problem. Instead, it could be a chip with high sparsification threshold in a quiet environment.

maxCallsNoLadderVRB (5)

The number of events in the VRB occupancy data must exceed this limit for SVXMon to start reporting that there is no VRB occupancy data for some chip (in case VRB occupancies are monitored at all).

needRunEventsMedians (10)

The number of events which SVXMon must process since the beginning of a run before reporting problems with chip group statistics.

needSnapshotEventsMedians (10)

The number of events which SVXMon must process before reporting problems with short-term chip group statistics. This parameter should be in agreement with the global snapshot schedule so that snapshots are not made more often than **needSnapshotEventsMedians** events.

needRunEventsChips (30)

The number of events which SVXMon must process before reporting problems with chip statistics accumulated since the beginning of the run.

needSnapshotEventsChips (30)

The number of events which SVXMon must process before reporting problems with short-term chip statistics.

padBottomMargin (0.25)

Double. The fractional bottom margin of the pad on which the group plots are placed. It is convenient to have this quantity as a configurable parameter because the optimal bottom margin depends on the length of group names.

labelSize (0.05)

Double. The fractional text size for the group names, as they will be displayed on the plots of group quantities (one group name per bin). It is convenient to have this text size as a parameter because the optimal size depends on the number of chip groups defined.

All parameters except **padBottomMargin** and **labelSize** may be adjusted interactively in the middle of a run by executing the relevant commands at the SVXMon prompt.

In addition to the chip status maps, for each monitored quantity SVXMon creates a set of plots on twelve canvases which display this quantity together with associated upper and lower limits for each chip in the silicon system, one bin per chip. Example plots which display chip occupancy for the west side of the west barrel are shown in Fig. D.16. Whenever some quantity goes out of limits, the color of the plot which

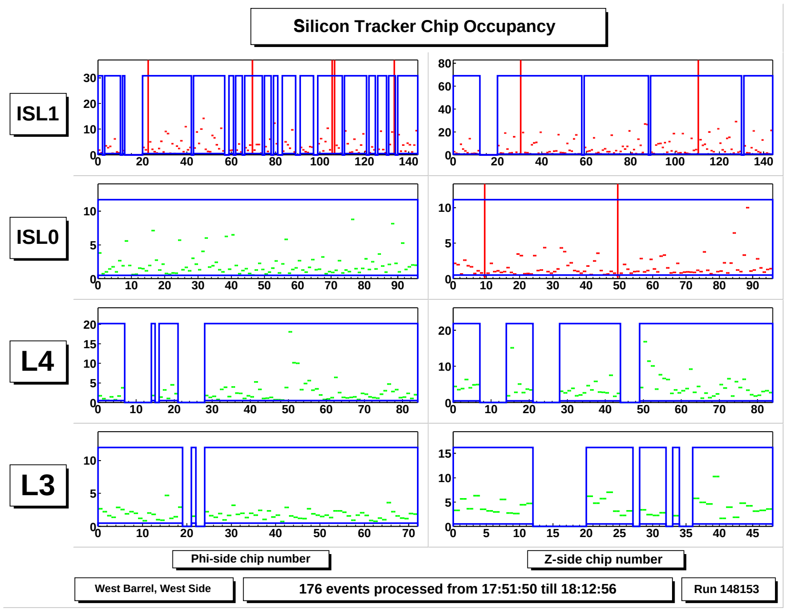


Figure D.16: *Example occupancy monitoring plots for the west side of the west barrel in layers L3 to ISL1. The chip numbers on the horizontal axis increase with the phi wedge number. Within each wedge they increase in readout order. The monitored quantity is drawn in green color if everything is normal, or in red if there is a problem. The acceptable range limits are shown in blue.*

contains the failing chip is changed to red. The chips with problems are marked with large vertical bars so that they are clearly visible, and an error message is sent to the error logger with the location of the failing chip and a description of the problem.

D.6 Offline Histogram Analysis: SvxMonRunCompare

The set of executables and scripts collectively called “SvxMonRunCompare” has been designed to monitor variations in the silicon system performance from run to run by comparing chip-level data. SvxMonRunCompare fills up and analyzes a ROOT ntuple which contains, for each run, several informative statistics for each chip. Such quantities as chip occupancy, mean charge, rms of the pulse height distribution, *etc.* are calculated from the strip-level plots (Sec. D.4.1) stored in the SVXMon root files.

The chip settings needed to ensure similar chip configurations are extracted from the online database.

SvxMonRunCompare can be found in the Run II software repository, package “SvxDaqUtils”.

Ntuple Structure

Every entry in the SvxMonRunCompare ntuple represents a run. The ntuple contains a block with the run information and a block with the chip information for ladders included in the run. Tables D.1 and D.2 describe the contents of each block.

Variable	Source	Description
runNumber	user	Run Number
version	hard-coded	Version of the Software
isAGoodRun	user	0 = Bad ; 1 = Good
nSvxMonEvents	SVXMon	Number of events processed by SVXMon
nChips	SVXMon	Total number of chips included in the run
runtype	database	Type of run as in Ref. [68]
nevents	database	Number of L3 events according to database
extra0	database	Run Control setting variable
modeGlobal	database	Global DPS setting
modeL00	database	L00 DPS setting
modeSVX	database	SVX DPS setting
modeISL	database	ISL DPS setting
rundate	database	Date of data taking

Table D.1: *Description of the run block variables.*

The global dynamic pedestal subtraction (DPS) setting in table D.1 has the following meaning: -1 = undefined, 0 = DPS off, 1 = DPS on, 2 = mixed. For each chip, the occupancy is defined as the median occupancy of the 128 individual strips in the chip. The median is used rather than average in order to reduce the effect of bad strips. The chip mean charge and charge standard deviation are also defined as medians of the corresponding strip quantities. “stdevVar” is a more complicated variable which was chosen for its sensitivity to physics signal compared to noise. It reflects the variation of charge standard deviation from strip to strip. This variation is larger for chips which read out particle signals than for chips which read out only noise because the number of tracks passing through each strip and the deposited charge tend to exhibit non-uniformity which can’t be attributed to noise alone. For each chip, stdevVar is defined as the median of $|\sigma_i - \sigma_{chip}|$ quantities, i=0,...,127, where σ_i is the standard deviation of the charge collected by strip i, and σ_{chip} is the charge standard deviation of the chip as defined above.

Run Comparison

In the course of normal online operation, SVXMon starts SvxMonRunCompare executables at the end of each run in its tcl configuration file. The number of events processed by SVXMon in this run should be sufficient for a reliable estimation of the chip statistics — as a rule of thumb, 200 events is enough. After appending the run to the ntuple (executable `appendSvxMonRun`), SvxMonRunCompare compares this run to a set of recent reference runs already present in the ntuple (executable `compareSvxMonRun`) and produces a log file and a ROOT file containing some interesting histograms. If necessary, it also generates alarms to the CO and the Silicon monitoring experts. A description of the run comparison executables is provided below.

- appendSvxMonRun

  Usage:

  ```
  > appendSvxMonRun runNumber SvxMonFile TreeFile Option
  ```

  Here, “SvxMonFile” is the SVXMon histogram file corresponding to run “runNumber” and “TreeFile” is the ROOT file which contains the run comparison

Variable	Type	Source	Description
barrel	short	database	West=0, Central=1, East=2
halfLadder	short	database	segment (a.k.a bulkhead, or half-barrel)
phiWedge	short	database	Offline wedge number
layer	short	database	Offline layer number (L00=0, SVX L0=1, ...)
side	short	database	ϕ side = 0, Z side = 1
chipNumber	short	database	Chip Number in the readout sequence on the half-ladder side
key	int	database	SiDigiCode key
dpsOn	short	database	DPS. 0=OFF, 1=ON
bandWidth	short	database	Integrator bandwidth (rise time)
brs	short	database	Bias Ratio Select: affects comparator and ramp bias currents
readNN	short	database	Read Nearest Neighbour mode
readAll	short	database	Read All mode
countMod	short	database	Counter Modulo: ADC counter limit
drs	short	database	Driver Resistor Select: sets the output driver resistors
ramPed	short	database	ADC Ramp Pedestal
pDepth	short	database	Pipeline depth
chipId	short	database	Chip Id (in the daisy chain)
threshold	short	database	Threshold
lastChip	short	database	Set to indicate that this is the last chip in the daisy chain
fePol	short	database	Front End Polarity
rOrder	short	database	Read Out order
calDir	short	database	Cal Direction: charge injection polarity
rampDir	short	database	Ramp Direction
compDir	short	database	Comparator Direction
iSel	short	database	Bias Current levels
rampTrim	short	database	Ramp Slope Trim
mean	float	SVXMon	Chip mean charge
occup	float	SVXMon	Chip occupancy
stdev	float	SVXMon	Chip charge standard deviation
stdevVar	float	SVXMon	(see text)

Table D.2: *Description of the chip array block variables.*

ntuple. appendSvxMonRun appends a run to an existing ntuple (option "update"), or creates a new ntuple and appends the run to it (option "create").

- compareSvxMonRun

  Usage:

```
    > compareSvxMonRun runNumber TreeFile \
ParameterFile LogDirectory \
        -n numberOfRuns -m maxRun \
        -p minRun -s scriptName
```

  `compareSvxMonRun` is the main executable of SvxMonRunCompare. It compares the run specified by "runNumber" to a set of reference runs, produces a text log file, a ROOT log file, and, if necessary, alarm log files and an alarm pop-up window.

  ParameterFile: file containing a list of parameters and configuration variables. An example of such a file can be found in Sec. D.6.

  LogDirectory: directory in which the log files will be stored.

  Options -n, -m, and -p determine the set of runs used as reference runs. Option -n is incompatible with options -m and -p. If used, -m and -p have to be used together.

  -n : number of runs to be used as reference runs.

  -m and -p determine the range of runs to be used as reference runs.

  -s : a tcl file that produces a pop-up window when executed if SvxMonRunCompare decides to launch an alarm and if the "pop_alarm" parameter is set to "t" in the parameter file.

  SvxMonRunCompare starts the comparison by choosing a set of reference runs. It selects good runs (isAGoodRun != 0) for which SVXMon processed more than 200 events. With the -n option, it selects the last "numberOfRuns" such runs; with the -m -p options, it selects all such runs in the range [maxRun, minRun].

For each chip present in the run to be compared and for each of the variables "occupancy", "mean", "stdev", and "stdevVar", the program computes the average and rms of the variable over the reference runs, using only the runs in which the chip settings are exactly identical to the settings in the compared run. The deviation in the compared run from the average value is computed and normalized to the rms. In the subsequent description we refer to this normalized deviation as the significance of the variable, $\sigma_{variable}$. Figure D.17 shows the distribution of the four variables for the reference runs together with their values in the compared run.

Diagnostics can be made based on the value and the significance of the four variables. Currently, three different checks are made:

1. "Chip with failure":

   A chip is said to have a failure if any variable value or variable significance is out of the following acceptable ranges:

   limit_low < value < limit_high
   nsigma_low < significance < nsigma_high

   where the cuts have been set in the SvxMonRunCompare parameter file. (see Sec. D.6). In case of a failure, the chip is reported in the text log file and in the ROOT log file (with a canvas such as the one shown in figure D.17).

2. "Unbiased chip":

   By "unbiased chip" we actually mean a chip whose ladder is unbiased. Such a chip is characterized by a higher occupancy value, and lower mean, stdev, and stdevVar values than a biased one. Therefore, we tag a chip as unbiased if

$\sigma_{occupancy}$ > bias_signif_high AND any other variable significance < bias_signif_low

Currently, the cut values are set as follows: bias_signif_high = 6 for occupancy and bias_signif_low = -4 for all other variables.

Any unbiased chip is reported in the text log file and the ROOT log file, and triggers an alarm log file and a popup window on the CO monitors.

3. General data quality cut:

   Figure D.18 shows the occupancy significance for all chips present in the compared run. In this particular run, the distribution is well centered and its rms is close to 1 which means that there is no global discrepancy with the reference runs. We use this histogram (and the corresponding plots for mean, stdev, and stdevVar) to check that, overall, the data in the compared run is not far from the data in the reference run. If, for any variable, either the rms or the absolute value of the distribution mean exceed 1.5, a warning message is issued and stored in the text file. Above 3, a "General Silicon Alarm" goes off.

- removeSvxMonRun

  Usage:

  ```
  > removeSvxMonRun runNumber TreeFile
  ```

  `removeSvxMonRun` removes the run "runNumber" from the ntuple stored in the file "TreeFile".

- modifySvxMonRun

  Usage:

  ```
  > modifySvxMonRun runNumber TreeFile variable newvalue passwd
  ```

  `modifySvxMonRun` modifies the value of the variable "variable" in the entry "runNumber" of the ntuple. This executable is used to tag a run as good or

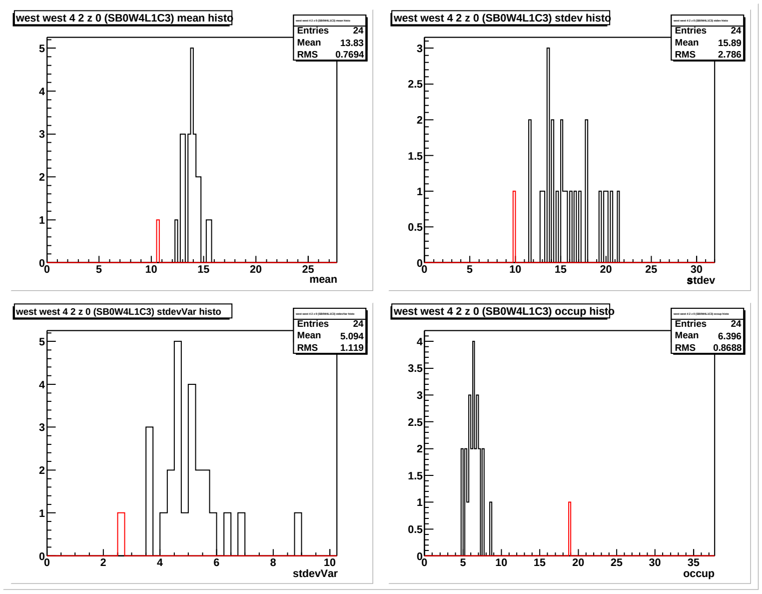


Figure D.17: *Single chip occupancy, mean charge, charge standard deviation, and stdevVar, for reference runs (black line) and the run being compared (red line). The data for this chip looks like its ladder was unbiased, and an alarm should go off.*

bad, and "isAGoodRun" is currently the only variable which can be modified. The only purpose of the password is to prevent uneducated users from running this executable. The correct passwd value is "1975".

Example SvxMonRunCompare Parameter File

```
// This is a lookup table for comparison parameters
// used by SvxMonRunCompare.
// Format for parameters:
//      quantity parameter_name value
// Format for mailing list:
// mail u1@fnal.gov u2@fnal.gov ...
// Format for turning on or off the pop up alarm:
// pop_alarm t     // pop up alarm is on
// pop_alarm f     // pop up alarm is off
// (words separated by spaces only; double slashes "//" are commented)
```

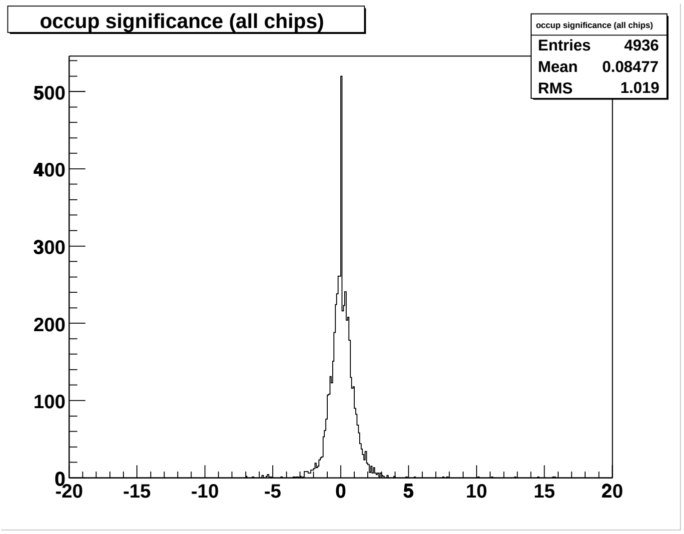


Figure D.18: *Occupancy significance of all chips in the run being compared. The spike at 0 comes from chips that never send any data (due to truncated readout chains). The rest of the distribution is a Gaussian centered at 0, with an rms close to 1: the run shows no deviation from the reference runs.*

```
//
// Author: Henri Bachacou, bachacou@fnal.gov (2002/04/02)


// Parameters for ‘‘chip with failure’’ check:
mean nsigma_high           7.
mean nsigma_low           -7.
mean limit_high          300.
mean limit_low             0.
// Parameters for ‘‘unbiased chip’’ check:
mean bias_signif_high    300.
mean bias_signif_low      -4.


// Same for other variables...
stdev nsigma_high          7.
stdev nsigma_low          -7.
stdev limit_high         300.
```

```
stdev limit_low            0.
stdev bias_signif_high   300.
stdev bias_signif_low     -4.
//
stdevVar nsigma_high       7.
stdevVar nsigma_low       -7.
stdevVar limit_high      300.
stdevVar limit_low         0.
stdevVar bias_signif_high 300.
stdevVar bias_signif_low  -4.
//
occup nsigma_high          7.
occup nsigma_low          -7.
occup limit_high         101.
occup limit_low            0.
occup bias_signif_high     6.
occup bias_signif_low   -300.

// Email list for alarm report:
mail bachacou@fnal.gov

// Turn on the popup window alarm:
pop_alarm t
```

Chip History

The SvxMonRunCompare ntuple is a usual ROOT ntuple. As such, it can be accessed either interactively in a ROOT session, or with the help of some ROOT scripts. Example scripts are available on the online machines:

`start_chip_plotter.C`

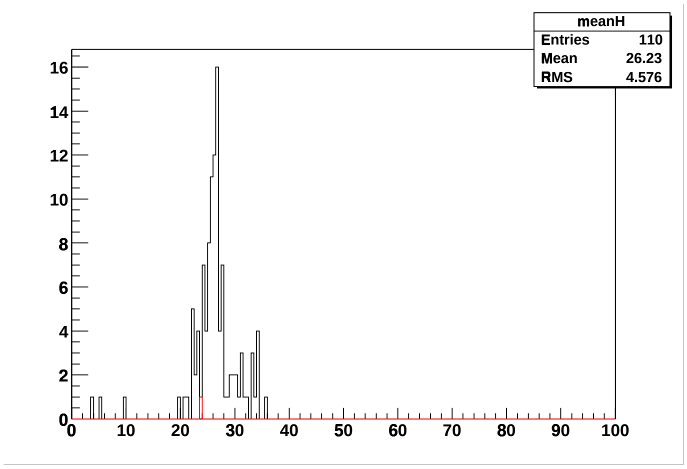


Figure D.19: *Chip Mean Charge for a given run (red line) and a set of reference runs (black line).*

`start_ladder2D_plotter.C`

`start_chip_plotter.C` plots the occupancy, the mean charge, the charge standard deviation, and stdevVar of a single chip for a given run (red line) and for a range of reference runs (black line). The reference runs are "good runs" (isAGoodRun != 0) for which SVXMon processed at least 200 events (nSvxMonEvents >= 200). Figure D.19 shows such a plot for the mean charge of chip SB0W0L0C0 (first chip on phi side). `start_ladder2D_plotter.C` plots one of the four quantities of all chips on one side of a ladder vs the run number. Figure D.20 shows such a plot for the occupancy of ladder SB0W0L0, phi side. In order to start the scripts, type the following commands on any online machine:

```
source ~cdfsoft/cdf2.cshrc
setup cdfsoft2
root start_ladder2D_plotter.C
```

Then follow the instructions.

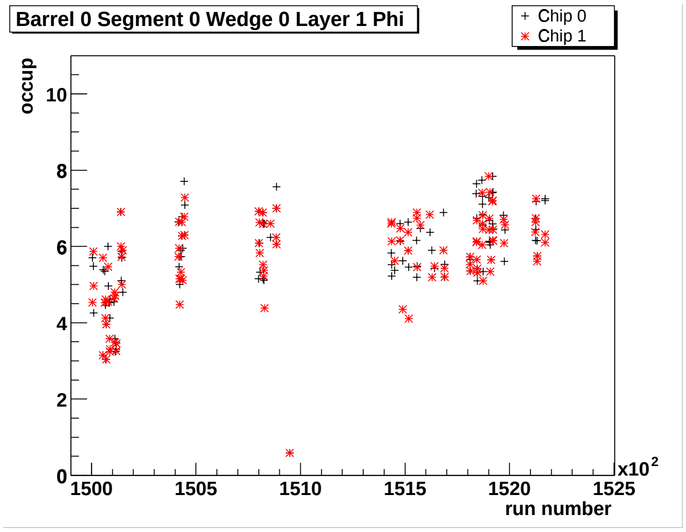


Figure D.20: *Chip Occupancy vs Run Number.*

D.7 SVXMon Messages

SVXMon message types are limited to 20 characters. This restriction is inherited from the ZOOM error logging facility [62] used by SVXMon. The message severity levels are varied according to perceived importance of the detected problem. Severity level 4 ("info") is reserved for summary messages generated at the end of each run.

In the table below, the "Source" column shows where the error diagnostics is generated. In this column, U means that the problem is detected by the unpacker, V means that the error flag is set by VRB hardware, and S means that this particular type of message is generated by SVXMon code. The message types are listed in alphabetical order with the exception of various "Out of Limits" errors. The "Out of Limits" errors are described in detail in Sec. D.5.

Message Type	**Meaning**	**Probable Causes**
0 Events	A plot update was requested before any events were processed	A new run has just begun
Bad ADC Value	The ADC value for some channel is invalid (after bank unpacking)	Hardware error in the chip (the ADC value higher than counter modulo + 1) Optical transmission error Bad pedestal subtraction lookup table in the FIB Mismatch in the data representation used by SVXMon and by the bank unpacker
Bad Chip ID	Chip number in the data is out of range for a given HDI	Optical transmission error
Bad Configuration	SVXMon is unable to configure itself	Online production database is down or unreachable An error in the SVXMon tcl configuration file
Bad FIB Id	Unrecognized FIB id pair	Ids in the FIB register do not conform to the standard convention
Bad Hit Position	Track helix has no intersection with a half-ladder which has a hit on the track	CDF software bug
Bad Parameter Value	Invalid parameter value in the SVXMon configuration file	An error in the SVXMon tcl configuration file
Bad VRB Data Format	The "Data Format Error" bit is set in the VRB error word of the SIXD/ISLD bank	VRB hardware error
Bit 0 High	Bit 0 is stuck high, the unpacker is able to correct the problem	Optical transmission error
Bit 0 Low	Bit 0 is stuck low, the unpacker is able to correct the problem	Optical transmission error

Message Type	Meaning	Probable Causes
Bit 1 High	Bit 1 is stuck high, the unpacker is able to correct the problem	Optical transmission error
Bit 1 Low	Bit 1 is stuck low, the unpacker is able to correct the problem	Optical transmission error
Bit Error (not 0/1)	Stuck bit (not 0 or 1), the unpacker is able to correct the problem	Optical transmission error
Bunch X Out of Sync	SIXD/ISLD bunch crossing numbers are different	FIB hardware error Reformatter error
C2 Termination	Truncation termination character seen (0xc2)	Silicon hybrid drops readout Optical transmission error
Chip Side Error	Wrong chip id in the data stream	Optical transmission error
Configuration Change	The set of silicon ladders included in the system has changed while SVXMon was running	The silicon configuration has been modified
Data Found After EOR	Data found by the bank unpacker after the EOR record in the HDI block	DAQ error
Data Overrun	HDI readout length is 0xffe = 4094 bytes	Silicon hybrid malfunction Optical transmission error
Drop Readout (c2c2)	HDI readout length is 4 bytes long (0xc2c20000)	Silicon hybrid does not read out DOIM failure
Drop Readout (d0 ch)	Channel number 0xd0 = 208 is found in the data stream	?
Dropped Readout	EOR record found in the data stream before all chips have been unpacked	Silicon hybrid drops readout
Duplicate Chan (nn)	Duplicate channel number is found in the data stream (nearest neighbor unpacker)	Optical transmission error (*e.g.*, bit 0 is stuck)
Duplicate Channel	Duplicate channel number is found in the data stream ("classic" unpacker)	Optical transmission error (*e.g.*, bit 0 is stuck)

Message Type	**Meaning**	**Probable Causes**
EOR Not Found	EOR record not found in the data stream	DAQ error
Event Count	SVXMon info message about the number of events with problems dumped to disk for a given run	End of a run
Event Sync Error	The "Event Synchronization Error" bit is set in the VRB error word of the SIXD/ISLD bank	VRB hardware error
Extra HDI	Some HDIs not included in the run (as marked in the hardware database) are found in the data stream	DAQ malfunction FIB id registers are not configured properly
ff or 7f Error	Channel byte differs from expected by more than one high bit	Optical transmission error Silicon chip hardware failure
Glink Frame Error	The "Glink Frame Error" bit is set in the VRB error word of the SIXD/ISLD bank	DAQ error
Glink Sync Error	The "Glink Synch Error" bit is set in the VRB error word of the SIXD/ISLD bank	DAQ error
Illegal Termination	HDI data stream termination character is neither 0xc1 nor 0xc2	Optical transmission error (bit 7 stuck high)
Internal Error	SVXMon detected an inconsistency between its data structures	Bug in SVXMon
Internal VRB Error	The "Internal VRB Error" bit is set in the VRB error word of the SIXD/ISLD bank	VRB hardware error
Invalid Cell Id	The cell id for a given chip is out of range (> 45)	Silicon chip hardware failure Optical transmission error
Invalid Chip Key	Invalid silicon chip identifier (SiChipKey) has been generated from chip position in the data stream	CDF software bug

Message Type	**Meaning**	**Probable Causes**
Invalid EOR	Invalid word in the data stream in place where EOR record is expected	DAQ error
Invalid SRC Command	The "Invalid SRC Command" bit is set in the VRB error word of the SIXD/ISLD bank	DAQ error
Invalid Word	Generic bank unpacking failure	Optical transmission error DAQ error
Missing Chip ID	Chip id is not found in the data stream	Optical transmission error
Missing HDI	Some HDIs included in the run (as marked in the hardware database) are missing in the data stream	DAQ malfunction FIB id registers are not configured properly
No Data	No data unpacked for a given chip in several consecutive SVXMon events	The bank unpacker gives up because of unrecoverable errors upstream from the given chip The silicon configuration has changed, and SVXMon was not restarted The chip was intentionally "sparsified out" by setting high sparsification threshold in the initialization stream
No First Chip	First HDI data word is not a correct chip ID (and not wrong by just one bit)	Optical transmission error Silicon hybrid malfunction
No HDI Info	No HDI blocks inside VRB blocks of the SIXD/ISLD banks	DAQ failure
No Silicon Banks	No SIXD/ISLD banks in the event	Event builder failure
No VRB Info	No VRB blocks inside SIXD/ISLD banks	DAQ failure
No VRB Occupancies	No VRB occupancy data for this chip	DAQ or software failure

Message Type	**Meaning**	**Probable Causes**
Out of Order Chip ID	Out of order chip id encountered in the data stream	Optical transmission error Incorrect chip id setting in the initialization stream
Pipeline Out of Sync	Chip pipeline cell id is different from the "most frequent" one	Synchronization loss on a chip due to noisy front end clock Readout chip hardware failure Optical transmission error
Run Config Invalid	Invalid chip key in the run conditions database	Database-related software bug
Strip Out of Order	Out of order channel number is found in the data stream	Optical transmission error
Stuck Bit	Stuck bit in the chip pipeline cell id	Optical transmission error
Stuck Cell Id	Chip pipeline cell id is stuck at a particular value	Readout chip hardware failure
Suspect Data	The unpacker made a wrong guess about channel numbering while trying to correct a bit error	Optical transmission error
TL1A Out of Sync	Time since Level 1 Accept is different for SIXD and ISLD banks	Reformatter error DAQ error
TS Out of Sync	Trigger supervisor counter is different for SIXD and ISLD banks	DAQ error
Unknown Channel	Fatal unpacker error at a channel number: not duplicate channel, not different from expected channel by one bit, and not all bits high	Readout chip hardware failure Optical transmission error
Unknown Chip Error	Fatal unpacker error at a chip ID: not the next chip ID, not different from expected by one bit, and not all bits high when a channel number is expected	Readout chip hardware failure Optical transmission error

Message Type	**Meaning**	**Probable Causes**
VRBINFO_DEBUG_BX	An info message with a bunch crossing number for a given HDI	The "detailedVrbErrors" flag is turned on in the SVXMon configuration file
VRBs Out of Sync	Either the trigger supervisor counter or time since Level 1 accept is out of sync in several consecutive SVXMon events	See the description of "TL1A Out of Sync" and "TS Out of Sync" errors
Wrong Chip Count	The number of chips in the data stream is wrong for a given HDI	Optical transmission error
Wrong Data Size	Wrong data size for a given ladder in the VRB occupancy data	DAQ or software failure
Wrong HDI Count	The number of HDI blocks in the silicon banks has changed since the previous event	DAQ malfunction
Wrong VRB Count	The number of VRB blocks in the silicon banks has changed since the previous event	DAQ malfunction
*** Out of Limits	Some monitored quantity (occupancy, mean ADC value, RMS ADC value, *etc.*) is out of limits for a given chip. The type of the quantity is provided in place of ***. Please consult Sec. D.5 for further details.	High detector noise Optical transmission errors DAQ failure SVXMon is not configured to perform bad channel suppression