Direct Measurement of the Top Quark Mass at DØ

B. Abbott,[30] M. Abolins,[27] B.S. Acharya,[45] I. Adam,[12] D.L. Adams,[39] M. Adams,[17] S. Ahn,[14] H. Aihara,[23] G.A. Alves,[10] N. Amos,[26] E.W. Anderson,[19] R. Astur,[44] M.M. Baarmand,[44] A. Baden,[25] V. Balamurali,[34] J. Balderston,[16] B. Baldin,[14] S. Banerjee,[45] J. Bantly,[5] E. Barberis,[23] J.F. Bartlett,[14] K. Bazizi,[41] A. Belyaev,[28] S.B. Beri,[36] I. Bertram,[33] V.A. Bezzubov,[37] P.C. Bhat,[14] V. Bhatnagar,[36] M. Bhattacharjee,[44] N. Biswas,[34] G. Blazey,[32] S. Blessing,[15] P. Bloom,[7] A. Boehnlein,[14] N.I. Bojko,[37] F. Borcherding,[14] C. Boswell,[9] A. Brandt,[14] R. Brock,[27] A. Bross,[14] D. Buchholz,[33] V.S. Burtovoi,[37] J.M. Butler,[3] W. Carvalho,[10] D. Casey,[41] Z. Casilum,[44] H. Castilla-Valdez,[11] D. Chakraborty,[44] S.-M. Chang,[31] S.V. Chekulaev,[37] L.-P. Chen,[23] W. Chen,[44] S. Choi,[43] S. Chopra,[26] B.C. Choudhary,[9] J.H. Christenson,[14] M. Chung,[17] D. Claes,[29] A.R. Clark,[23] W.G. Cobau,[25] J. Cochran,[9] L. Coney,[34] W.E. Cooper,[14] C. Cretsinger,[41] D. Cullen-Vidal,[5] M.A.C. Cummings,[32] D. Cutts,[5] O.I. Dahl,[23] K. Davis,[2] K. De,[46] K. Del Signore,[26] M. Demarteau,[14] D. Denisov,[14] S.P. Denisov,[37] H.T. Diehl,[14] M. Diesburg,[14] G. Di Loreto,[27] P. Draper,[46] Y. Ducros,[42] L.V. Dudko,[28] S.R. Dugad,[45] D. Edmunds,[27] J. Ellison,[9] V.D. Elvira,[44] R. Engelmann,[44] S. Eno,[25] G. Eppley,[39] P. Ermolov,[28] O.V. Eroshin,[37] V.N. Evdokimov,[37] T. Fahland,[8] M.K. Fatyga,[41] S. Feher,[14] D. Fein,[2] T. Ferbel,[41] G. Finocchiaro,[44] H.E. Fisk,[14] Y. Fisyak,[7] E. Flattum,[14] G.E. Forden,[2] M. Fortner,[32] K.C. Frame,[27] S. Fuess,[14] E. Gallas,[46] A.N. Galyaev,[37] P. Gartung,[9] T.L. Geld,[27] R.J. Genik II,[27] K. Genser,[14] C.E. Gerber,[14] B. Gibbard,[4] S. Glenn,[7] B. Gobbi,[33] A. Goldschmidt,[23] B. Gómez,[1] G. Gómez,[25] P.I. Goncharov,[37] J.L. González Solís,[11] H. Gordon,[4] L.T. Goss,[47] K. Gounder,[9] A. Goussiou,[44] N. Graf,[4] P.D. Grannis,[44] D.R. Green,[14] H. Greenlee,[14] G. Grim,[7] S. Grinstein,[6] N. Grossman,[14] P. Grudberg,[23] S. Grünendahl,[14] G. Guglielmo,[35] J.A. Guida,[2] J.M. Guida,[5] A. Gupta,[45] S.N. Gurzhiev,[37] P. Gutierrez,[35] Y.E. Gutnikov,[37] N.J. Hadley,[25] H. Haggerty,[14] S. Hagopian,[15] V. Hagopian,[15] K.S. Hahn,[41] R.E. Hall,[8] P. Hanlet,[31] S. Hansen,[14] J.M. Hauptman,[19] D. Hedin,[32] A.P. Heinson,[9] U. Heintz,[14] R. Hernández-Montoya,[11] T. Heuring,[15] R. Hirosky,[17] J.D. Hobbs,[14] B. Hoeneisen,[1,*] J.S. Hoftun,[5] F. Hsieh,[26] Ting Hu,[44] Tong Hu,[18] T. Huehn,[9] A.S. Ito,[14] E. James,[2] J. Jaques,[34] S.A. Jerger,[27] R. Jesik,[18] J.Z.-Y. Jiang,[44] T. Joffe-Minor,[33] K. Johns,[2] M. Johnson,[14] A. Jonckheere,[14] M. Jones,[16] H. Jöstlein,[14] S.Y. Jun,[33] C.K. Jung,[44] S. Kahn,[4] G. Kalbfleisch,[35] J.S. Kang,[20] D. Karmanov,[28] D. Karmgard,[15] R. Kehoe,[34] M.L. Kelly,[34] C.L. Kim,[20] S.K. Kim,[43] A. Klatchko,[15] B. Klima,[14] C. Klopfenstein,[7] V.I. Klyukhin,[37] V.I. Kochetkov,[37] J.M. Kohli,[36] D. Koltick,[38] A.V. Kostritskiy,[37] J. Kotcher,[4] A.V. Kotwal,[12] J. Kourlas,[30] A.V. Kozelov,[37] E.A. Kozlovski,[37] J. Krane,[29] M.R. Krishnaswamy,[45] S. Krzywdzinski,[14] S. Kunori,[25] S. Lami,[44] R. Lander,[7] F. Landry,[27] G. Landsberg,[14] B. Lauer,[19] A. Leflat,[28] H. Li,[44] J. Li,[46] Q.Z. Li-Demarteau,[14] J.G.R. Lima,[40] D. Lincoln,[26] S.L. Linn,[15] J. Linnemann,[27] R. Lipton,[14] Y.C. Liu,[33] F. Lobkowicz,[41] S.C. Loken,[23] S. Lökös,[44] L. Lueking,[14] A.L. Lyon,[25] A.K.A. Maciel,[10] R.J. Madaras,[23] R. Madden,[15] L. Magaña-Mendoza,[11] V. Manankov,[28] S. Mani,[7] H.S. Mao,[14,†] R. Markeloff,[32] T. Marshall,[18] M.I. Martin,[14] K.M. Mauritz,[19] B. May,[33] A.A. Mayorov,[37] R. McCarthy,[44] J. McDonald,[15] T. McKibben,[17] J. McKinley,[27] T. McMahon,[35] H.L. Melanson,[14] M. Merkin,[28] K.W. Merritt,[14] H. Miettinen,[39] A. Mincer,[30] C.S. Mishra,[14] N. Mokhov,[14] N.K. Mondal,[45] H.E. Montgomery,[14] P. Mooney,[1] H. da Motta,[10] C. Murphy,[17] F. Nang,[2] M. Narain,[14] V.S. Narasimham,[45] A. Narayanan,[2] H.A. Neal,[26] J.P. Negret,[1] P. Nemethy,[30] D. Norman,[47] L. Oesch,[26] V. Oguri,[40] E. Oliveira,[10] E. Oltman,[23] N. Oshima,[14] D. Owen,[27] P. Padley,[39] A. Para,[14] Y.M. Park,[21] R. Partridge,[5] N. Parua,[45] M. Paterno,[41] B. Pawlik,[22] J. Perkins,[46] M. Peters,[16] R. Piegaia,[6] H. Piekarz,[15] Y. Pischalnikov,[38] V.M. Podstavkov,[37] B.G. Pope,[27] H.B. Prosper,[15] S. Protopopescu,[4] J. Qian,[26] P.Z. Quintas,[14] R. Raja,[14] S. Rajagopalan,[4] O. Ramirez,[17] L. Rasmussen,[44] S. Reucroft,[31] M. Rijssenbeek,[44] T. Rockwell,[27] M. Roco,[14] N.A. Roe,[23] P. Rubinov,[33] R. Ruchti,[34] J. Rutherfoord,[2] A. Sánchez-Hernández,[11] A. Santoro,[10] L. Sawyer,[24] R.D. Schamberger,[44] H. Schellman,[33] J. Sculli,[30] E. Shabalina,[28] C. Shaffer,[15] H.C. Shankar,[45] R.K. Shivpuri,[13] M. Shupe,[2] H. Singh,[9] J.B. Singh,[36] V. Sirotenko,[32] W. Smart,[14] E. Smith,[35] R.P. Smith,[14] R. Snihur,[33] G.R. Snow,[29] J. Snow,[35] S. Snyder,[4] J. Solomon,[17] P.M. Sood,[36] M. Sosebee,[46] N. Sotnikova,[28] M. Souza,[10] A.L. Spadafora,[23] G. Steinbrück,[35] R.W. Stephens,[46] M.L. Stevenson,[23] D. Stewart,[26] F. Stichelbaut,[44] D.A. Stoianova,[37] D. Stoker,[8] M. Strauss,[35] K. Streets,[30] M. Strovink,[23] A. Sznajder,[10] P. Tamburello,[25] J. Tarazi,[8] M. Tartaglia,[14] T.L.T. Thomas,[33] J. Thompson,[25] T.G. Trippe,[23] P.M. Tuts,[12] N. Varelas,[17] E.W. Varnes,[23] D. Vititoe,[2] A.A. Volkov,[37] A.P. Vorobiev,[37] H.D. Wahl,[15] G. Wang,[15] J. Warchol,[34] G. Watts,[5] M. Wayne,[34] H. Weerts,[27] A. White,[46] J.T. White,[47] J.A. Wightman,[19] S. Willis,[32] S.J. Wimpenny,[9] J.V.D. Wirjawan,[47] J. Womersley,[14] E. Won,[41] D.R. Wood,[31] H. Xu,[5] R. Yamada,[14] P. Yamin,[4] J. Yang,[30] T. Yasuda,[31] P. Yepes,[39] C. Yoshikawa,[16] S. Youssef,[15] J. Yu,[14] Y. Yu,[43] Z.H. Zhu,[41] D. Zieminska,[18] A. Zieminski,[18] E.G. Zverev,[28] and A. Zylberstejn[42]

(DØ Collaboration)

[1] *Universidad de los Andes, Bogotá, Colombia*
[2] *University of Arizona, Tucson, Arizona 85721*

[3]*Boston University, Boston, Massachusetts 02215*
[4]*Brookhaven National Laboratory, Upton, New York 11973*
[5]*Brown University, Providence, Rhode Island 02912*
[6]*Universidad de Buenos Aires, Buenos Aires, Argentina*
[7]*University of California, Davis, California 95616*
[8]*University of California, Irvine, California 92697*
[9]*University of California, Riverside, California 92521*
[10]*LAFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil*
[11]*CINVESTAV, Mexico City, Mexico*
[12]*Columbia University, New York, New York 10027*
[13]*Delhi University, Delhi, India 110007*
[14]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510*
[15]*Florida State University, Tallahassee, Florida 32306*
[16]*University of Hawaii, Honolulu, Hawaii 96822*
[17]*University of Illinois at Chicago, Chicago, Illinois 60607*
[18]*Indiana University, Bloomington, Indiana 47405*
[19]*Iowa State University, Ames, Iowa 50011*
[20]*Korea University, Seoul, Korea*
[21]*Kyungsung University, Pusan, Korea*
[22]*Institute of Nuclear Physics, Kraków, Poland*
[23]*Lawrence Berkeley National Laboratory and University of California, Berkeley, California 94720*
[24]*Louisiana Tech University, Ruston, Louisiana 71272*
[25]*University of Maryland, College Park, Maryland 20742*
[26]*University of Michigan, Ann Arbor, Michigan 48109*
[27]*Michigan State University, East Lansing, Michigan 48824*
[28]*Moscow State University, Moscow, Russia*
[29]*University of Nebraska, Lincoln, Nebraska 68588*
[30]*New York University, New York, New York 10003*
[31]*Northeastern University, Boston, Massachusetts 02115*
[32]*Northern Illinois University, DeKalb, Illinois 60115*
[33]*Northwestern University, Evanston, Illinois 60208*
[34]*University of Notre Dame, Notre Dame, Indiana 46556*
[35]*University of Oklahoma, Norman, Oklahoma 73019*
[36]*University of Panjab, Chandigarh 16-00-14, India*
[37]*Institute for High Energy Physics, 142-284 Protvino, Russia*
[38]*Purdue University, West Lafayette, Indiana 47907*
[39]*Rice University, Houston, Texas 77005*
[40]*Universidade do Estado do Rio de Janeiro, Brazil*
[41]*University of Rochester, Rochester, New York 14627*
[42]*CEA, DAPNIA/Service de Physique des Particules, CE-SACLAY, Gif-sur-Yvette, France*
[43]*Seoul National University, Seoul, Korea*
[44]*State University of New York, Stony Brook, New York 11794*
[45]*Tata Institute of Fundamental Research, Colaba, Mumbai 400005, India*
[46]*University of Texas, Arlington, Texas 76019*
[47]*Texas A&M University, College Station, Texas 77843*

(May 4, 1998)

We determine the top quark mass m_t using $t\bar{t}$ pairs produced in the DØ detector by $\sqrt{s} = 1.8$ TeV $p\bar{p}$ collisions in a 125 pb^{-1} exposure at the Fermilab Tevatron. We make a two constraint fit to m_t in $t\bar{t} \to bW^+\bar{b}W^-$ final states with one W boson decaying to $q\bar{q}$ and the other to $e\nu$ or $\mu\nu$. Likelihood fits to the data yield $m_t(l + \text{jets}) = 173.3 \pm 5.6$ (stat) $\pm$ 5.5 (syst) GeV/c^2. When this result is combined with an analysis of events in which both W bosons decay into leptons, we obtain $m_t = 172.1 \pm 5.2$ (stat) $\pm$ 4.9 (syst) GeV/c^2. An alternate analysis, using three constraint fits to fixed top quark masses, gives $m_t(l + \text{jets}) = 176.0 \pm 7.9$ (stat) $\pm$ 4.8 (syst) GeV/c^2, consistent with the above result. Studies of kinematic distributions of the top quark candidates are also presented.

PACS numbers: 14.65.Ha, 13.85.Qk, 13.85.Ni

Contents

I. INTRODUCTION

The discovery of the top quark by the CDF [1] and DØ [2] collaborations at the Fermilab Tevatron ended the search phase of top quark physics. Since then, emphasis has shifted to determining its properties — especially its large mass (about 200 times that of a proton) and production cross section. Reviews of searches for and the initial observations of the top quark are given in Ref. [3]. Details of the initial DØ top quark search can be found in Ref. [4]. This paper reports on the determination of the top quark mass using all the data collected by the DØ experiment during the 1992–1996 Tevatron runs. This is more than twice as much data as was available for the initial observation. In addition, improvements have been made in event selection, object reconstruction, and mass analysis techniques. The result is a reduction of the statistical and systematic uncertainties by nearly a factor of four. A short paper giving results from this analysis has been published [5]. The CDF collaboration has also recently published an updated top quark mass measurement [6].

The top quark is one of the fundamental fermions in the standard model of electroweak interactions and is the weak-isospin partner of the bottom quark. For a top quark with mass substantially greater than that of the W boson, the standard model predicts it to decay promptly (before hadronization) to a W boson plus a bottom quark with a branching fraction of nearly 100%. A precision measurement of the top quark mass, along with the W boson mass and other electroweak data, can set constraints on the mass of the standard model Higgs boson. It may also be helpful in understanding the origin of quark masses.

In $p\overline{p}$ collisions at a 1.8 TeV center of mass energy, top quarks are produced primarily as $t\overline{t}$ pairs. Each decays into a W boson plus a bottom quark, resulting in events having several jets and often a charged lepton. Due to the large top quark mass, these final state objects tend to have large momenta transverse to the $p\overline{p}$ direction. About 30% of $t\overline{t}$ decays have a single electron or muon (from the decay of one of the W bosons) with a large transverse momentum. Typically, the neutrino that accompanies this electron or muon will also have a large transverse momentum, producing significant missing transverse energy. These characteristics allow for the selection of a sample of "lepton + jets" events with an enriched signal to background ratio. This sample is the basis for the top quark mass analysis reported in this paper. It also comprises a large portion of the data sample used for the measurement of the $p\overline{p} \to t\overline{t}$ production cross section [7]. A similar mass analysis for the final state with two charged leptons plus jets is described in Ref. [8].

Three methods have been used to determine the top quark mass in the lepton + jets channels. Two of them use constrained variable-mass kinematic fits to obtain a best-fit mass value for each event. The top quark mass is then extracted using a maximum likelihood fit to a two-dimensional distribution, with one axis being the best-fit mass, and the other being a variable which discriminates $t\bar{t}$ events from the expected backgrounds. The difference between these two methods is in the discriminant variable and the binning used. The third method uses χ^2 values from fixed-mass kinematic fits. A cut is made using a top quark discriminant to select a sample of events with low background. The expected contribution from the background is subtracted from the distribution of χ^2 versus mass, and the resulting background-subtracted distribution is fit near the minimum to extract the top quark mass.

This paper is organized as follows. Section II briefly describes aspects of the DØ detector essential for this analysis. Section III discusses event selection, including triggers, particle identification, and the criteria used to select the initial event sample. Section IV describes the jet energy corrections. Section V discusses the simulation of $t\bar{t}$ signal and background events. Section VI defines the two discriminants used to separate top quark events from background. Section VII describes the variable-mass kinematic fits to individual events and the likelihood fits used to extract the top quark mass, and gives results from these fits. Section VIII describes the pseudo-likelihood method (which uses fixed-mass kinematic fits), gives results from it, and compares these results with those from the two likelihood methods. Section IX examines some kinematic properties of top quark events. Finally, conclusions are presented in Sec. X.

II. THE DØ DETECTOR

DØ is a multipurpose detector designed to study $p\bar{p}$ collisions at the Fermilab Tevatron Collider. The detector was commissioned during the summer of 1992. The work presented here is based on approximately 125 pb^{-1} of accumulated data recorded during the 1992–1996 collider runs. A full description of the detector may be found in Ref. [9]. Here, we describe briefly the properties of the detector that are relevant for the top quark mass measurement.

The detector was designed to have good electron and muon identification capabilities, and to measure jets and missing transverse energy $\not\!\!E_T$ with good resolution. The detector consists of three major systems: a nonmagnetic central tracking system, a hermetic uranium liquid-argon calorimeter, and a muon spectrometer. A cut away view of the detector is shown in Fig. 1.

The central detector (CD) consists of four tracking subsystems: a vertex drift chamber, a transition radiation detector (not used for this analysis), a central drift chamber, and two forward drift chambers. It measures the trajectories of charged particles and can discriminate between single charged particles and e^+e^- pairs from

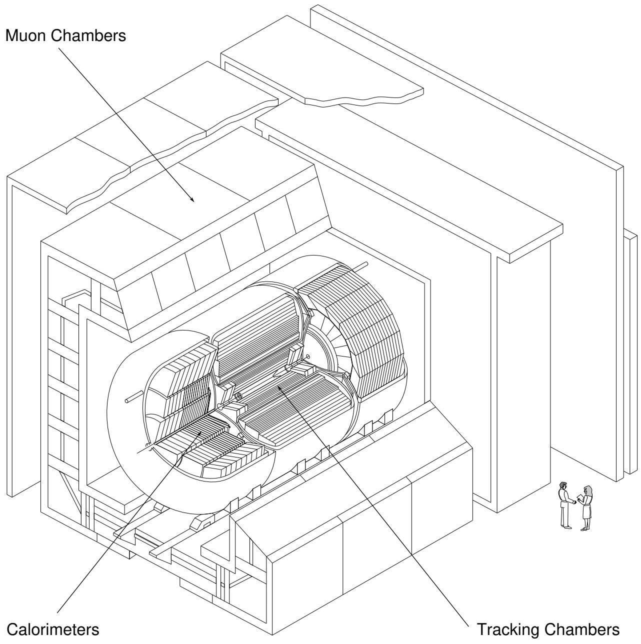

FIG. 1. Cut away isometric view of the DØ detector.

photon conversions by measuring the ionization along their tracks. It covers the region $|\eta| < 3.2$ in pseudorapidity, where $\eta = \tanh^{-1}(\cos\theta)$. (We define θ and ϕ to be the polar and azimuthal angles, respectively.)

The calorimeter is divided into three parts: the central calorimeter (CC) and the two end calorimeters (EC), which together cover the pseudorapidity range $|\eta| < 4.2$. The inner electromagnetic (EM) portion of the calorimeters is 21 radiation lengths deep, and is divided into four longitudinal segments (layers). The outer hadronic portions are 7–9 nuclear interaction lengths deep, and are divided into four (CC) or five (EC) layers. The calorimeters are transversely segmented into pseudoprojective towers with $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. The third layer of the electromagnetic (EM) calorimeter, in which the maximum of EM showers is expected, is segmented twice as finely in both η and ϕ, with cells of size $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$.

Since muons from top quark decays populate predominantly the central region, this work uses only the central portion of the DØ muon system, covering $|\eta| < 1.7$. This system consists of four planes of proportional drift tubes in front of magnetized iron toroids with a magnetic field of 1.9 T and two groups of three planes each of proportional drift tubes behind the toroids. The magnetic field lines and the wires in the drift tubes are oriented transversely to the beam direction. The muon momentum p^μ is measured from the muon's deflection angle in the magnetic field of the toroid.

A separate synchrotron, the Main Ring, lies above the Tevatron and passes through the outer region of the DØ calorimeter. During data-taking, it is used to accelerate protons for antiproton production. Losses from the

Main Ring may deposit energy in the calorimeters, increasing the instrumental background. We reject much of this background at the trigger level by not accepting triggers during injection into the Main Ring, when losses are large. Some triggers are also disabled whenever a Main Ring bunch passes through the detector or when losses are registered in scintillation counters around the Main Ring.

III. EVENT SELECTION

For the purposes of this analysis, we divide the lepton + jets final states into electron and muon channels. We further subdivide these channels based on whether or not a muon consistent with $b \to \mu + X$ is present. We thus have four channels, which will be denoted e+jets, μ+jets, e + jets/μ, and μ + jets/μ.

The event sample used for determining the top quark mass is selected using criteria similar to those used for the $t\bar{t}$ production cross section measurement [7], with the exception of the cuts on the event shape variables $H_T \equiv \sum E_T^{\rm jet}$ and aplanarity. The particle identification, trigger requirements, and event selection cuts are summarized below. Some additional background information about triggering, particle identification, and jet and $\not\!\!E_T$ reconstruction may be found in Ref. [4]. (Note that this reference describes an older version of the analysis; the current electron and muon identification algorithms provide better rejection of backgrounds and increased efficiencies than those used in Ref. [4]. However, much of the information remains applicable.)

A. Particle identification

1. Electrons

Electron identification is based on a likelihood technique. Candidates are first identified by finding isolated clusters of energy in the EM calorimeter with a matching track in the central detector. We then cut on a likelihood constructed from the following four variables:

- The χ^2 from a covariance matrix which measures the consistency of the calorimeter cluster shape with that of an electron shower.
- The electromagnetic energy fraction, defined as the ratio of the portion of the energy of the cluster found in the EM calorimeter to its total energy.
- A measure of the consistency between the track position and the cluster centroid.
- The ionization dE/dx along the track.

To a good approximation, these four variables are independent of each other for electron candidates.

Electrons from W boson decay tend to be isolated, even in $t\bar{t}$ events. Thus, we make the additional cut

$$\frac{E_{\rm tot}(0.4) - E_{\rm EM}(0.2)}{E_{\rm EM}(0.2)} < 0.1, \tag{3.1}$$

where $E_{\rm tot}(0.4)$ is the energy within $\Delta R < 0.4$ of the cluster centroid ($\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) and $E_{\rm EM}(0.2)$ is the energy in the EM calorimeter within $\Delta R < 0.2$.

2. Muons

Two types of muon selection are used in this analysis. The first is used to identify isolated muons from $W \to \mu\nu$ decay. The other is used to tag b-jets by identifying "tag" muons consistent with originating from $b \to \mu + X$ decay.

Besides cuts on the muon track quality, both selections require that:

- The muon pseudorapidity $|\eta^\mu| \le 1.7$.
- The magnetic field integral > 2.0 T $\cdot$ m (equivalent to a momentum change of 0.6 GeV/c).
- The energy deposited in the calorimeter along a muon track be at least that expected from a minimum ionizing particle.

For isolated muons, we apply the following additional selection requirements:

- Transverse momentum $p_T \ge 20$ GeV/c.
- The distance in the $\eta - \phi$ plane between the muon and the closest jet $\Delta R(\mu, j) > 0.5$.

For tag muons, we instead require:

- $p_T \ge 4$ GeV/c.
- $\Delta R(\mu, j) < 0.5$.

3. Jets and missing E_T

Jets are reconstructed in the calorimeter using a fixed-size cone algorithm. We use a cone size of $\Delta R = 0.5$.

Neutrinos are not detected directly. Instead, their presence is inferred from missing transverse energy $\not\!\!E_T$. Two different definitions of $\not\!\!E_T$ are used in the event selection:

- $\not\!\!E_T^{\rm cal}$, the calorimeter missing E_T, obtained from the transverse energy of all calorimeter cells.
- $\not\!\!E_T$, the muon corrected missing E_T, obtained by subtracting the transverse momenta of identified muons from $\not\!\!E_T^{\rm cal}$.

B. Triggers

The DØ trigger system is responsible for reducing the event rate from the beam crossing rate of 286 kHz to the approximately 3–4 Hz which can be recorded on tape. The first stage of the trigger (level 1) makes fast analog sums of the transverse energies in calorimeter trigger towers. These towers have a size of $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ and are segmented longitudinally into electromagnetic and hadronic sections. The level 1 trigger operates on these sums along with patterns of hits in the muon spectrometer. It can make a trigger decision within the space of a single beam crossing (unless a level 1.5 decision is required; see below). After level 1 accepts an event, the complete event is digitized and sent to the level 2 trigger, which consists of a farm of 48 general-purpose processors. Software filters running in these processors make the final trigger decision.

The triggers used are defined in terms of combinations of specific objects (electron, muon, jet, $\not\!\!E_T$) required in the level 1 and level 2 triggers. These elements are summarized below. For more information on the DØ trigger system, see Refs. [4,9].

To trigger on electrons, level 1 requires that the transverse energy in the EM section of a trigger tower be above a programmed threshold. The level 2 electron algorithm examines the regions around the level 1 towers which are above threshold, and uses the full segmentation of the EM calorimeter to identify showers with shapes consistent with those of electrons. The level 2 algorithm can also apply an isolation requirement or demand that there be an associated track in the central detector.

For the latter portion of the run, a "level 1.5" processor was also available for electron triggering. The E_T of each EM trigger tower above the level 1 threshold is summed with the neighboring tower with the most energy. A cut is then made on this sum. The hadronic portions of the two towers are also summed, and the ratio of EM transverse energy to total transverse energy in the two towers is required to be above 0.85. The use of a level 1.5 electron trigger is indicated in the tables below as an "EX" tower.

The level 1 muon trigger uses the pattern of drift tubes with hits to provide the number of muon candidates in different regions of the muon spectrometer. A level 1.5 processor may optionally be used to put a p_T requirement on the candidates (at the expense of slightly increased dead time). In level 2, the full digitized data are available, and the first stage of the full event reconstruction is performed. The level 2 muon algorithm can optionally require the presence of an energy deposit in the calorimeter consistent with that from a muon; this is indicated in the tables below by "cal confirm".

For a jet trigger, level 1 requires that the sum of the transverse energies in the EM and hadronic sections of a trigger tower be above a programmed threshold. Alternatively, level 1 can sum the transverse energies within "large tiles" of size 0.8×1.6 in $\eta \times \phi$ and cut on these sums. Level 2 then sums calorimeter cells around the identified towers (or around the E_T-weighted centroids of the large tiles) in cones of a specified radius ΔR, and imposes a cut on the total transverse energy.

The $\not\!\!E_T$ in the calorimeter can also be computed in both level 1 and level 2. The z position used for the interaction vertex in level 2 is determined from the relative timing of hits in scintillation counters located in front of each EC (level 0).

The trigger requirements used for this analysis are summarized in Tables I–III. These tables are divided according to the three major running periods. Run 1a was from 1992–1993, run 1b was from 1994–1995, and run 1c was during the winter of 1995–1996. Note that not all the triggers listed were active simultaneously, and that differing requirements were used to veto possible Main Ring events. In addition, some of the triggers were prescaled at high luminosity. The "exposure" column in the tables takes these factors into account.

C. Event selection

The first set of cuts used to define the sample for mass analysis is very similar to that used for the cross section analysis [7]:

- An isolated electron or muon with $E_T > 20$ GeV.
- $|\eta^e| < 2.0$ or $|\eta^\mu| < 1.7$.
- At least 4 jets with $E_T > 15$ GeV and $|\eta^{\text{jet}}| < 2.0$.
- $\not\!\!E_T^{\text{cal}} > 25$ GeV for e+jets (untagged) or $\not\!\!E_T^{\text{cal}} > 20$ GeV for μ+jets (both tagged and untagged).
- $\not\!\!E_T > 20$ GeV.

We reject events which contain photons — isolated clusters in the EM calorimeter with shapes consistent with an EM shower and with a poor match to any track in the central detector, and satisfying $E_T > 15$ GeV and $|\eta| < 2$. Three such events are rejected. We also reject events which contain extra isolated high-p_T electrons or which fail additional cuts to remove calorimeter noise and Main Ring effects.

After these cuts, the remaining background is primarily W + jets. About 20% of the background consists of QCD multijet events in which a jet is misidentified as a lepton (estimated using the same methods as used for the cross section measurement).

If a candidate has a tag muon, we require it to pass additional cuts on the direction of the $\not\!\!E_T$ vector. For the e + jets/μ channel, we require

- $\not\!\!E_T > 35$ GeV, if $\Delta\phi(\not\!\!E_T, \mu) < 25°$,

while for the μ + jets/μ channel, we require that the highest-p_T muon satisfy

- $\Delta\phi(\not\!\!E_T, \mu) < 170°$ and

TABLE I. Triggers used during run 1a (1992–1993). "Exposure" gives the effective integrated luminosity for each trigger, taking into account any prescaling.

Name	Exposure (pb^{-1})	Level 1	Level 2	Used by
ELE-HIGH	11.0	1 EM tower, $E_T > 10$ GeV	1 isolated e, $E_T > 20$ GeV	$e+\text{jets}$ $e+\text{jets}/\mu$
ELE-JET	14.4	1 EM tower, $E_T > 10$ GeV, $\|\eta\| < 2.6$ 2 jet towers, $E_T > 5$ GeV	1 e, $E_T > 15$ GeV, $\|\eta\| < 2.5$ 2 jets ($\Delta R = 0.3$), $E_T > 10$ GeV, $\|\eta\| < 2.5$ $\not{E}_T^{\text{cal}} > 10$ GeV	$e+\text{jets}$ $e+\text{jets}/\mu$
MU-JET-HIGH	10.2	1 μ, $\|\eta\| < 2.4$ 1 jet tower, $E_T > 5$ GeV	1 μ, $p_T > 8$ GeV/c 1 jet ($\Delta R = 0.7$), $E_T > 15$ GeV	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$

TABLE II. Same as Table I for run 1b (1994–1995).

Name	Exposure (pb^{-1})	Level 1	Level 2	Used by
EM1-EISTRKCC-MS	93.4	1 EM tower, $E_T > 10$ GeV 1 EX tower, $E_T > 15$ GeV[a]	1 isolated e w/track, $E_T > 20$ GeV $\not{E}_T^{\text{cal}} > 15$ GeV	$e+\text{jets}$ $e+\text{jets}/\mu$
ELE-JET-HIGH	98.0	1 EM tower, $E_T > 12$ GeV, $\|\eta\| < 2.6$ 2 jet towers, $E_T > 5$ GeV, $\|\eta\| < 2.0$	1 e, $E_T > 15$ GeV, $\|\eta\| < 2.5$ 2 jets ($\Delta R = 0.3$), $E_T > 10$ GeV, $\|\eta\| < 2.5$ $\not{E}_T^{\text{cal}} > 14$ GeV	$e+\text{jets}$ $e+\text{jets}/\mu$
MU-JET-HIGH	66.4	1 μ, $p_T > 7$ GeV/c[a], $\|\eta\| < 1.7$ 1 jet tower, $E_T > 5$ GeV, $\|\eta\| < 2.0$[a]	1 μ, $p_T > 10$ GeV/c, $\|\eta\| < 1.7$ 1 jet ($\Delta R = 0.7$), $E_T > 15$ GeV, $\|\eta\| < 2.5$	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$
MU-JET-CAL	88.0	1 μ, $p_T > 7$ GeV/c[a], $\|\eta\| < 1.7$ 1 jet tower, $E_T > 5$ GeV, $\|\eta\| < 2.0$[a]	1 μ, $p_T > 10$ GeV/c, $\|\eta\| < 1.7$, cal confirm 1 jet ($\Delta R = 0.7$), $E_T > 15$ GeV, $\|\eta\| < 2.5$	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$
MU-JET-CENT	48.5	1 μ, $\|\eta\| < 1.0$ 1 jet tower, $E_T > 5$ GeV, $\|\eta\| < 2.0$	1 μ, $p_T > 10$ GeV/c, $\|\eta\| < 1.0$ 1 jet ($\Delta R = 0.7$), $E_T > 15$ GeV, $\|\eta\| < 2.5$	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$
MU-JET-CENCAL	51.2	1 μ, $\|\eta\| < 1.0$ 1 jet tower, $E_T > 5$ GeV, $\|\eta\| < 2.0$	1 μ, $p_T > 10$ GeV/c, $\|\eta\| < 1.0$, cal confirm 1 jet ($\Delta R = 0.7$), $E_T > 15$ GeV, $\|\eta\| < 2.5$	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$
JET-3-MU	11.9	3 jet towers, $E_T > 5$ GeV $\not{E}_T^{\text{cal}} > 20$ GeV	3 jets ($\Delta R = 0.7$), $E_T > 15$ GeV, $\|\eta\| < 2.5$ $\not{E}_T^{\text{cal}} > 17$ GeV	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$
JET-3-MISS-LOW	57.8	3 large tiles, $E_T > 15$, $\|\eta\| < 2.4$ 3 jet towers, $E_T > 7$ GeV, $\|\eta\| < 2.6$	3 jets ($\Delta R = 0.5$), $E_T > 15$ GeV, $\|\eta\| < 2.5$ $\not{E}_T^{\text{cal}} > 17$ GeV	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$
JET-3-L2MU	25.8	3 large tiles, $E_T > 15$, $\|\eta\| < 2.4$ 3 jet towers, $E_T > 7$ GeV, $\|\eta\| < 2.6$	1 μ, $p_T > 6$ GeV/c, $\|\eta\| < 1.7$, cal confirm 3 jets ($\Delta R = 0.5$), $E_T > 15$ GeV, $\|\eta\| < 2.5$ $\not{E}_T^{\text{cal}} > 17$ GeV	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$

[a]This cut was looser than indicated during early portions of the run.

TABLE III. Same as Table I for run 1c (1995–1996).

Name	Exposure (pb^{-1})	Level 1	Level 2	Used by
ELE-JET-HIGH	1.9	1 EM tower, $E_T > 12$ GeV, $\|\eta\| < 2.6$ 2 jet towers, $E_T > 5$ GeV, $\|\eta\| < 2.0$	1 e, $E_T > 15$ GeV, $\|\eta\| < 2.5$ 2 jets ($\Delta R = 0.3$), $E_T > 10$ GeV, $\|\eta\| < 2.5$ $\not{E}_T^{\text{cal}} > 14$ GeV	$e+\text{jets}$ $e+\text{jets}/\mu$
ELE-JET-HIGHA	11.0	1 EM tower, $E_T > 12$ GeV, $\|\eta\| < 2.6$ 2 jet towers, $E_T > 5$ GeV, $\|\eta\| < 2.0$ 1 EX tower, $E_T > 15$ GeV	1 e, $E_T > 17$ GeV, $\|\eta\| < 2.5$ 2 jets ($\Delta R = 0.3$), $E_T > 10$ GeV, $\|\eta\| < 2.5$ $\not{E}_T^{\text{cal}} > 14$ GeV	$e+\text{jets}$ $e+\text{jets}/\mu$
MU-JET-CENT	8.9	1 μ, $\|\eta\| < 1.0$ 1 jet tower, $E_T > 5$ GeV, $\|\eta\| < 2.0$ 2 jet towers, $E_T > 3$ GeV	1 μ, $p_T > 12$ GeV/c, $\|\eta\| < 1.0$ 1 jet ($\Delta R = 0.7$), $E_T > 15$ GeV, $\|\eta\| < 2.5$	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$
MU-JET-CENCAL	11.4	1 μ, $\|\eta\| < 1.0$ 1 jet tower, $E_T > 5$ GeV, $\|\eta\| < 2.0$ 2 jet towers, $E_T > 3$ GeV	1 μ, $p_T > 12$ GeV/c, $\|\eta\| < 1.0$, cal confirm 1 jet ($\Delta R = 0.7$), $E_T > 15$ GeV, $\|\eta\| < 2.5$	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$
JET-3-L2MU	11.3	3 large tiles, $E_T > 15$, $\|\eta\| < 2.4$ 3 jet towers, $E_T > 5$ GeV, $\|\eta\| < 2.0$ 4 jet towers, $E_T > 3$ GeV	1 μ, $p_T > 8$ GeV/c, $\|\eta\| < 1.7$, cal confirm 3 jets ($\Delta R = 0.5$), $E_T > 15$ GeV, $\|\eta\| < 2.5$ $\not{E}_T^{\text{cal}} > 17$ GeV	$\mu+\text{jets}$ $\mu+\text{jets}/\mu$

- $|\Delta\phi(\not{E}_T, \mu) - 90^\circ|/90^\circ < \not{E}_T/(45 \text{ GeV})$.

These cuts remove QCD multijet background events which appear to have a large $\not{E}_T$ due to a mismeasurement of the muon momentum.

For the remaining, untagged, events, we require:

- $E_T^W \equiv |E_T^{\text{lep}}| + |\not{E}_T| > 60$ GeV.
- $|\eta^W| < 2.0$.

For the purpose of these two cuts, we define η^W by assuming that the entire $\not{E}_T$ of the event is due to the neutrino from the decay of the W boson. The longitudinal component of the neutrino momentum p_z^ν is found by using the W boson mass M_W as a constraint. If the transverse mass of the lepton and neutrino $M_T(l\nu)$ is less than M_W, there are two real solutions; the one with the smallest absolute value of p_z^ν is used. Monte Carlo studies show that this is the correct solution about 80% of the time. If $M_T(l\nu) > M_W$ there are no real solutions. In this case, the $\not{E}_T$ is scaled so that $M_T(l\nu) = M_W$. This scaled $\not{E}_T$ is also used for the E_T^W cut (but not for the previous cuts on $\not{E}_T$ alone).

This cut on E_T^W removes a portion of the QCD multijet background. Figure 2 compares the E_T^W distribution for this background to that from Monte Carlo W + jets events.

We show in Fig. 3 the distributions of $|\eta^W|$ for our data and for the Monte Carlo prediction. The data are seen to significantly exceed the prediction of the VECBOS Monte Carlo (described in Sec. V) in the far forward region. The amount of $t\bar{t}$ signal with $|\eta^W| > 2$ is only a few percent ($\approx 3\%$ for $m_t = 175$ GeV/c^2). In addition, a check of the W boson transverse mass and $\not{E}_T$ distributions shows that the QCD multijet background plays no unusually prominent role at high $|\eta^W|$. We note that the VECBOS Monte Carlo, while the best currently available, is only a tree-level calculation of the W+jets process. Particularly in the forward direction, one would expect higher order corrections to play a larger role. To mitigate the effects of this discrepancy, and to further reduce the background, we require $|\eta^W| < 2$. Once this cut is made, the χ_L^2 between the data and prediction is 12.2 for 7 d.o.f., giving a 9% probability. ($\chi_L^2 \equiv 2\sum_i [y_i - N_i + N_i \ln(N_i/y_i)]$, where N is the number of observed events and y is the total number expected from Monte Carlo. This form is appropriate for low statistics [10].) The contribution of this effect to the systematic uncertainty will be discussed in Sec. VII G 2 (and is found to be negligible).

These event selection cuts are summarized in Table IV. When applied to the approximately 125 pb^{-1} of data from the 1992–1996 collider runs, 91 events are selected [11], seven of which have a tag muon. This sample will be referred to as the "precut" sample, and the set of cuts as the "PR" cuts. One additional cut is made to define the final sample. This is based on the χ^2 of a kinematic fit to the $t\bar{t}$ decay hypothesis ($\chi^2 < 10$), and is described in Sec. VII. This final cut reduces the sample to 77 candidate events, of which five are tagged.

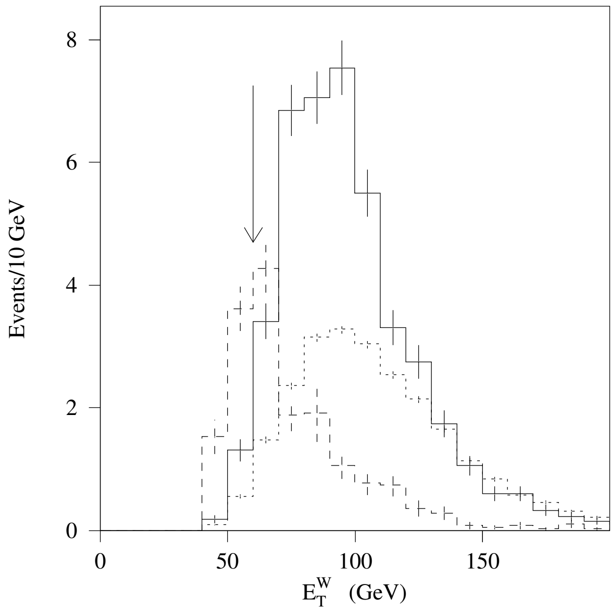

FIG. 2. E_T^W distribution for Monte Carlo W+jets events (solid histogram), QCD multijet background data (dashed histogram), and $t\bar{t}$ signal with $m_t = 175$ GeV/c^2 (dotted histogram). All selection cuts are applied except for the E_T^W cut. The arrow shows the cut value. (The normalizations are taken from the result of the LB fit to the data, as described in Sec. VII E, with channels combined as described in Sec. VII D. The models used to simulate the data are described in Sec. V.)

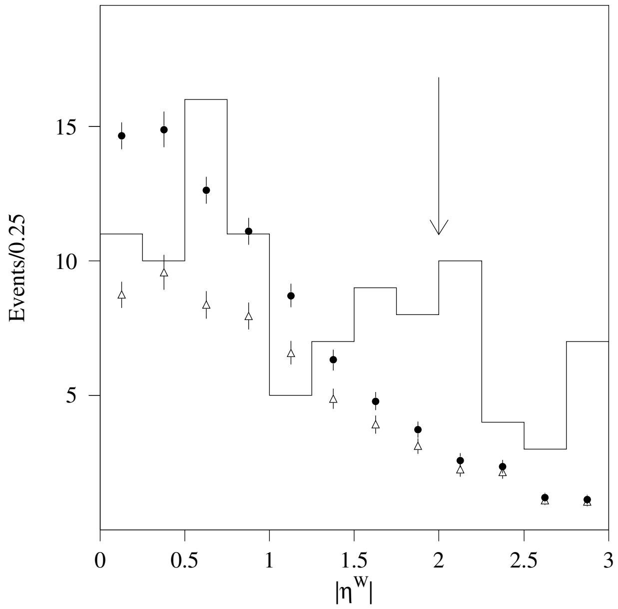

FIG. 3. $|\eta^W|$ distribution for data (histogram), predicted signal plus background (filled circles), and background alone (open triangles). All selection cuts are applied except for the η^W cut. The arrow shows the cut value. (The normalizations are as in Fig. 2.)

TABLE IV. Summary of event selection cuts.

Channel	e + jets	μ + jets	e + jets/μ	μ + jets/μ
Lepton	$E_T^e > 20$ GeV	$p_T^\mu > 20$ GeV/c	$E_T^e > 20$ GeV	$p_T^\mu > 20$ GeV/c
	$\lvert\eta^e\rvert < 2$	$\lvert\eta^\mu\rvert < 1.7$	$\lvert\eta^e\rvert < 2$	$\lvert\eta^\mu\rvert < 1.7$
$\not\!\!E_T$	$\not\!\!E_T > 20$ GeV	$\not\!\!E_T > 20$ GeV	$\not\!\!E_T > 20$ GeV	$\not\!\!E_T > 20$ GeV
	$\not\!\!E_T^{\rm cal} > 25$ GeV	$\not\!\!E_T^{\rm cal} > 20$ GeV		$\not\!\!E_T^{\rm cal} > 20$ GeV
Jets	≥ 4 jets	≥ 4 jets	≥ 4 jets	≥ 4 jets
	$E_T^{\rm jet} > 15$ GeV	$E_T^{\rm jet} > 15$ GeV	$E_T^{\rm jet} > 15$ GeV	$E_T^{\rm jet} > 15$ GeV
	$\lvert\eta^{\rm jet}\rvert < 2.0$	$\lvert\eta^{\rm jet}\rvert < 2.0$	$\lvert\eta^{\rm jet}\rvert < 2.0$	$\lvert\eta^{\rm jet}\rvert < 2.0$
μ Tag	No tag	No tag	Tag required	Tag required
Other	$E_T^W > 60$ GeV	$E_T^W > 60$ GeV	$\not\!\!E_T > 35$ GeV	$\Delta\phi(\not\!\!E_T, \mu) < 170°$
	$\lvert\eta^W\rvert < 2.0$	$\lvert\eta^W\rvert < 2.0$	if $\Delta\phi(\not\!\!E_T, \mu) < 25°$	$\lvert\Delta\phi(\not\!\!E_T, \mu) - 90°\rvert/90° <$
				$\not\!\!E_T/(45\text{ GeV})$
Events passing cuts	43	41	4	3
With $\chi^2 < 10$	35	37	2	3

IV. JET CORRECTIONS AND ENERGY SCALE UNCERTAINTY

To calibrate the energy scale so that data and Monte Carlo (MC) are on an equal footing, we apply a series of energy corrections to the measured objects. These corrections are carried out in three steps. The first of these corrections is done before events are selected and is used by most DØ analyses; the other two corrections are applied during the kinematic fit and are specific to the top quark mass analysis.

A. Standard corrections

For the standard corrections, electromagnetic objects are first scaled by a factor which was chosen to make the invariant mass peak from dielectron events match the Z boson mass as measured by the LEP experiments. (This factor is determined separately for each of the three cryostats of the calorimeter [12].) Next, jet energies are corrected using

$$E(\text{corrected}) = \frac{E(\text{measured}) - O}{R(1-S)}. \quad (4.1)$$

Here, R is the calorimeter response; it is found using E_T balance (as determined from the total $\not\!\!E_T$) in γ + jets events. This determination is done separately and symmetrically for both data and Monte Carlo. O is the offset due to the underlying event, multiple interactions, and noise from the natural radioactivity of the uranium absorber. It is determined by comparing data in which a hard interaction is required to data in which that requirement is relaxed, and by comparing data taken at different luminosities. The term S is the fractional shower leakage outside the jet cone in the calorimeter. It is determined by using single particle showers measured in the test beam to construct simulated showers from MC jets; this leakage is approximately 3% for a 50 GeV jet ($\Delta R = 0.5$) in the central calorimeter. Further details about these corrections may be found in Ref. [13].

TABLE V. Parameters for parton-level jet corrections. $E(\text{corrected}) = (E - A)/B$.

	Light quark jets		Untagged b jets	
η region	A (GeV)	B	A (GeV)	B
$0.0 < \lvert\eta_{\rm det}\rvert < 0.2$	0.322	0.933	-0.672	0.907
$0.2 < \lvert\eta_{\rm det}\rvert < 0.6$	0.635	0.930	-1.34	0.914
$0.6 < \lvert\eta_{\rm det}\rvert < 0.9$	1.86	0.883	0.002	0.868
$0.9 < \lvert\eta_{\rm det}\rvert < 1.3$	1.70	0.933	-0.548	0.904
$1.3 < \lvert\eta_{\rm det}\rvert$	4.50	0.882	2.46	0.859

B. Parton-level corrections

The procedure of the previous section corrects for the portions of showers in the calorimeter which spread outside of the jet cone, but not for any radiation outside of the cone. Thus, the corrected jet energies are systematically lower than the corresponding parton-level energies (i.e., before QCD evolution or fragmentation in the MC). We make a correction to match the scale of the jet energies to that of the unfragmented partons in the MC.

To derive this correction, we use HERWIG [14] $t\bar{t}$ Monte Carlo and match reconstructed jets to the partons from top quark decay. Their energies are then plotted against each other, as in Fig. 4. This relation is observed to be nearly linear. We fit it separately for light quark jets and for untagged b quark jets. The results are given in Table V for different regions in $\eta_{\rm det}$ ($\eta_{\rm det} \equiv$ 'detector-η' $\equiv$ the pseudorapidity corresponding to a particle coming from the geometric center of the detector, rather than from the interaction vertex). Separating the b quark jets allows us to correct, on average, for the neutrinos from b decays. This correction is observed not to depend strongly on the MC top quark mass.

For tagged b quark jets, we have additional information

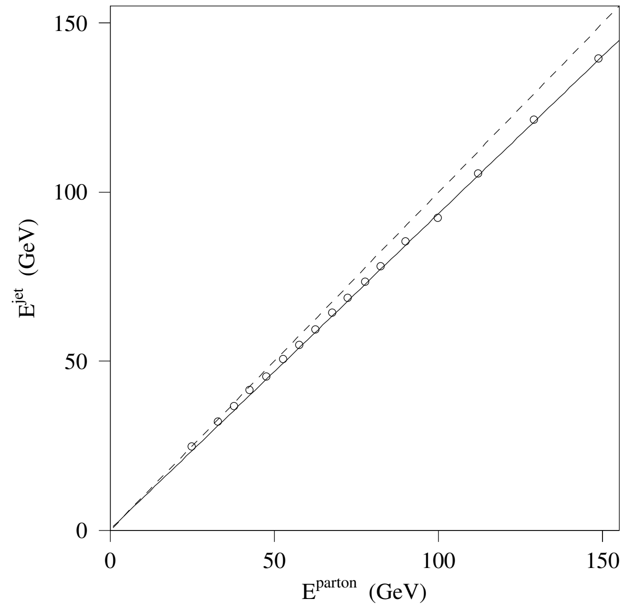

FIG. 4. The measured jet energies for quarks from $W \to q\overline{q}$ in $t\bar{t}$ MC are plotted against the corresponding parton energies. Radiation outside of the jet cone causes the measured jet energy to be lower than the energy at the parton level. The dashed line is drawn along the diagonal, and the solid line is a linear fit to the points. This plot is based on HERWIG fragmentation with $|\eta^{\rm jet}_{\rm det}| < 0.2$.

from the tag muon. However, the momentum spectrum of muons from b quark decay in $t\bar{t}$ events is rather steeply falling; furthermore, the resolution of the muon system is more nearly Gaussian in the inverse momentum $1/p$ than in p. Thus, measurement errors will cause the measured momentum of a tag muon to be biased upwards. We correct for this bias using $t\bar{t}$ MC, as illustrated in Fig. 5. We then further scale the muon momentum to account for the unobserved neutrino, as shown in Fig. 6. The jet itself is corrected using the light quark corrections; the estimated leptonic energy is then added to this corrected jet energy.

C. η-dependent adjustment and energy scale uncertainty

For the final corrections, we study the response of the detector to $\gamma + 1$ jet events, using both data and Monte Carlo. We select events containing exactly one photon with $E_T^\gamma > 20$ GeV, $|\eta^\gamma_{\rm det}| < 1.0$ or $1.6 < |\eta^\gamma_{\rm det}| < 2.5$, and exactly one reconstructed jet of any energy (excluding the photon). We require that the jet satisfy $E_T > 15$ GeV, $|\eta| < 2$, and $|\pi - \Delta\phi(j,\gamma)| < 0.2$ rad. We reject events with Main Ring activity and those which are likely to be multiple interactions. To reject W boson decays, we further require that $\not{E}_T/E_T^\gamma < 1.2$ if $E_T^\gamma < 25$ GeV, or

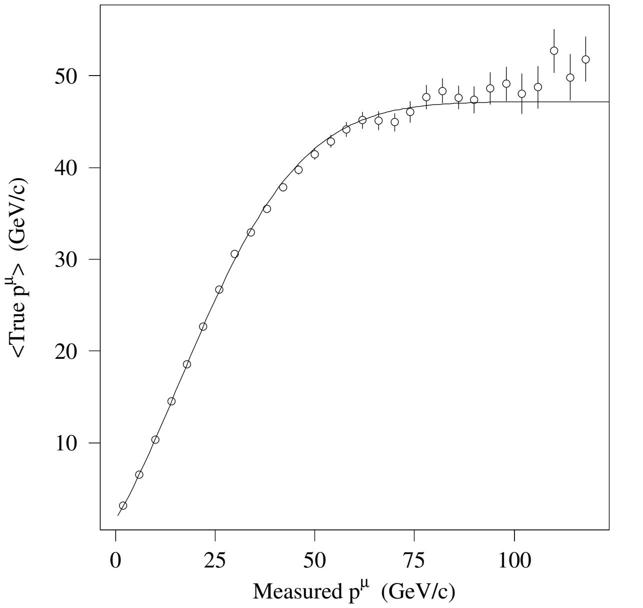

FIG. 5. Correlation between the measured momentum and the true momentum of the tag muon in Monte Carlo $t\bar{t}$ events. The curve is the result of an empirical fit, $47.19[1 - \exp(-0.03398 - 0.01593p^\mu - 0.0005554(p^\mu)^2)]$.

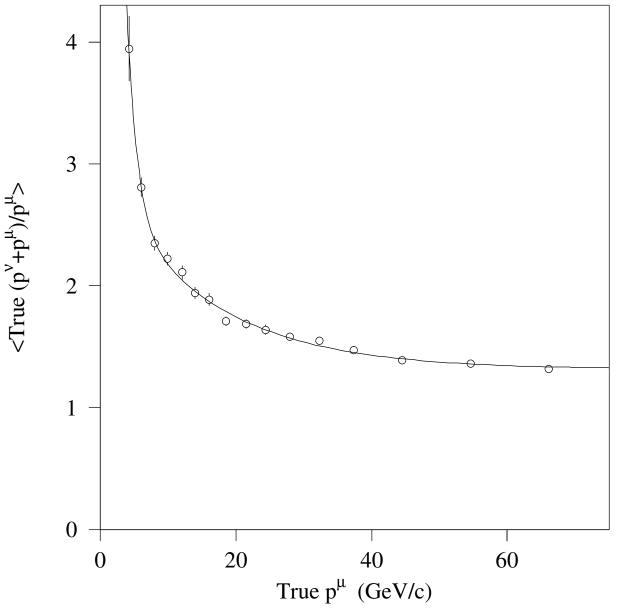

FIG. 6. Correlation between the tag muon momentum and the total leptonic energy from b quark decay in MC $t\bar{t}$ events. The curve is the result of an empirical fit, $1.313 + \exp(3.101 - 0.6528p^\mu) + \exp(0.4622 - 0.06514p^\mu)$.

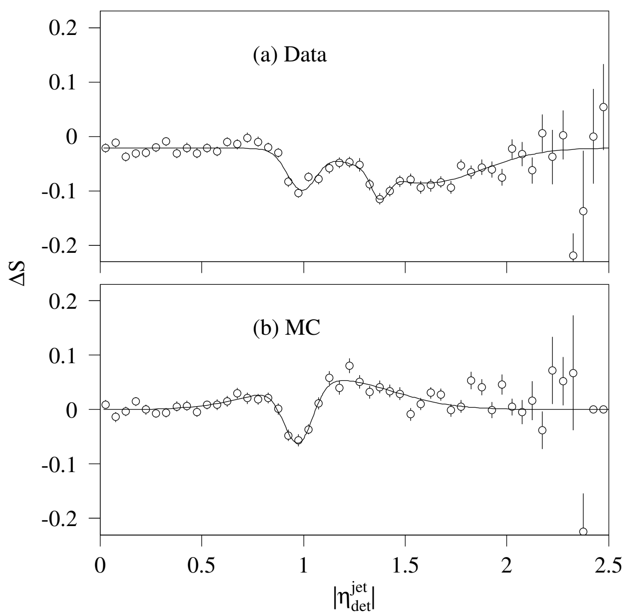

FIG. 7. The energy scale deviation ΔS as a function of $\eta^{\rm jet}_{\rm det}$ for (a) data and (b) Monte Carlo. The curves are empirical multigaussian fits to the points.

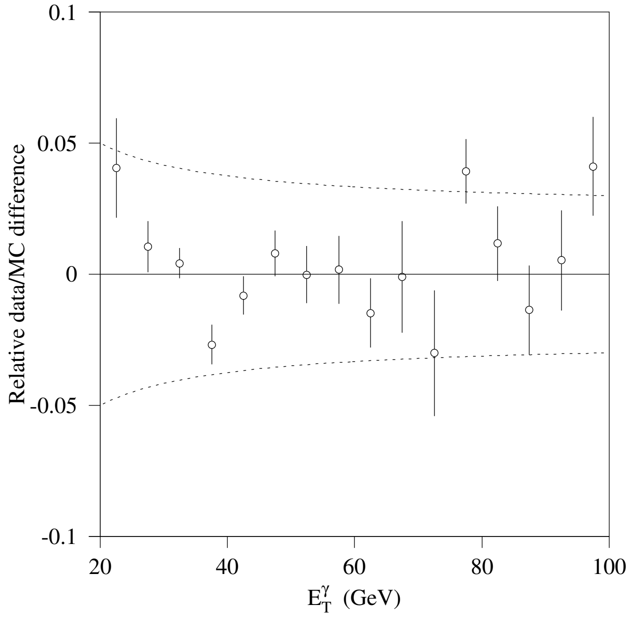

FIG. 8. The relative energy scale difference between data and MC as a function of photon E_T after all jet corrections are applied. The curves are the error band $\pm(2.5\% + 0.5$ GeV).

$\not{E}_T/E_T^{\gamma} < 0.65$ otherwise. With this selection, we compute

$$\Delta S = \left\langle \frac{E_T^{\rm jet} - E_T^{\gamma}}{E_T^{\gamma}} \right\rangle \tag{4.2}$$

and plot it as a function of $\eta^{\rm jet}_{\rm det}$. The result is shown in Fig. 7. This reveals detector inhomogeneities in the transition region between the central and end calorimeters [15]. The curve from Monte Carlo is also seen to have a somewhat different shape than that from data. To remove these effects, we smooth the ΔS distributions by fitting them to the sum of several Gaussians, and scale each jet by $1/(1+\Delta S(\eta^{\rm jet}_{\rm det}))$. This is done separately for data and for Monte Carlo.

To estimate the uncertainty in the relative scale between data and Monte Carlo after all corrections, we derive ΔS as a function of E_T^{γ} (averaging over $\eta^{\rm jet}_{\rm det}$) for both data and MC after all corrections have been applied. The difference of the two is plotted in Fig. 8, along with a band of $\pm(2.5\% + 0.5$ GeV), which we use as our estimate of the systematic uncertainty of the jet energy calibration. (It is the relative data-MC difference that is relevant, rather than the absolute uncertainty, since the final mass is extracted by comparing the data to MC generated with known top quark masses.)

A cross-check of these corrections is provided by $(Z \to ee)$ + jets events. As shown in Fig. 9, the corrected jets satisfactorily balance the Z boson. We also show in Fig. 10 the $W \to q\overline{q}$ and $t \to bq\overline{q}$ masses from $t\overline{t}$ MC before and after the final two corrections. It is seen that the proper masses are recovered.

The accuracy of these corrections depends on how well the Monte Carlo models jet widths. Studies of jets in DØ data show that HERWIG models the transverse energy distribution within jets to within 5–10% [16]. Note, however, that since the determination of the response is done separately for data and for Monte Carlo, any disagreements would, to first order, be removed from the energy scale determination. There can still be second-order effects: for example, if jets in HERWIG were slightly too narrow, and if two jets were to overlap slightly, then the perturbation to the apparent jet energies due to that overlap would be slightly underestimated in the Monte Carlo. For this situation, we calculate that the fraction of the energy of a jet between $R = 0.5$ and $R = 1.0$ of the jet axis which leaks into the nearest jet is about 10%. We further find that this region in R contains about 10% of the total energy of a HERWIG jet. Thus, the leakage of energy from a jet to a neighbor is on the order of 1%. If the fraction of the jet energy outside of $R = 0.5$ is substantially larger in data than in HERWIG, e.g., 20%, a 1% miscalibration would result. This is well within the uncertainty we assign for moderate E_T jets.

V. EVENT SIMULATION

Monte Carlo simulation is used to model the final states expected from top quark decays and their principal physics backgrounds. Although the overall background normalization is estimated using the observed data, the simulation is essential to determine the expected shapes of kinematic distributions.

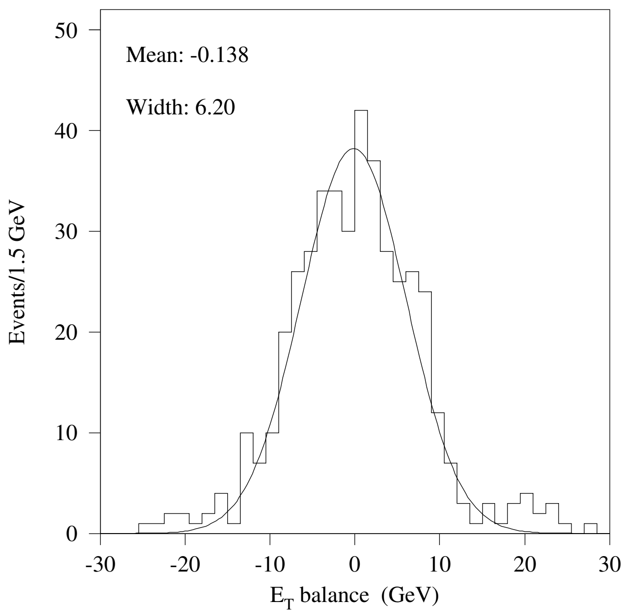

FIG. 9. Transverse energy balance for $(Z \to ee)$ + jets events. The vector $\vec{p}_T^Z + \sum_{\text{jets}} \vec{E}_T^{\text{jet}}$ is projected onto the angle bisector of the two electrons. All jet corrections are applied. The curve is a Gaussian fit to the histogram.

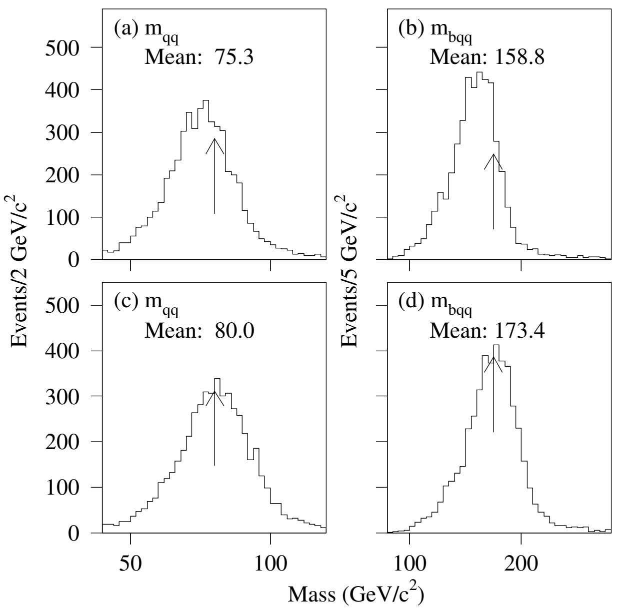

FIG. 10. Masses of $W \to q\bar{q}$ and $t \to bq\bar{q}$ in $t\bar{t}$ MC with $m_t = 175$ GeV/c^2, both (a), (b) with standard corrections only and (c), (d) with all jet corrections. The arrows locate the input W boson and top quark masses.

A. Signal events

Our primary model for $t\bar{t}$ production is the HERWIG generator, version 5.7, with CTEQ3M [17] parton distribution functions. HERWIG models $t\bar{t}$ production starting with the elementary hard process, choosing the parton momenta according to matrix element calculations. Initial and final state gluon emission is modeled using leading log QCD evolution [18]. Each top quark is then decayed to a W boson and a b quark, and final state partons are hadronized into jets. Underlying spectator interactions are also included in the model.

For this analysis, samples are generated with top quark masses between 110 and 230 GeV/c^2. To increase the efficiency in the processing of lepton plus jets events, one of the W bosons is forced to decay to one of the three lepton families. Events with no final state electrons or muons are vetoed, and half of the events in which both W bosons decayed leptonically are discarded in order to preserve the proper branching ratios. The generated events are run through the DØGEANT detector simulation [19,20] and the DØ event reconstruction program.

Additional samples are made using the ISAJET [21] generator to allow for cross-checks.

B. W+jets background

The background due to the production of a W boson along with multiple jets is modeled using the VECBOS [22] event generator. VECBOS supplies final state partons as a result of a leading order calculation which incorporates the exact tree level matrix elements for W and Z boson production with up to four additional partons. To include the effects of additional radiation and the underlying processes, and to model the hadronization of final state partons, the output of VECBOS is passed through HERWIG's QCD evolution and fragmentation stages. Since HERWIG requires information about the color labels of its input partons, it and VECBOS were modified to assign color and flavor to the generated partons. Flavors are assigned probabilistically by keeping track of the relative weights of each diagram contributing to the process. Color labels are simply assigned randomly. To estimate systematic uncertainties, we also generate samples which use ISAJET instead of HERWIG to fragment the VECBOS partons. We test the reliability of the HERWIG and ISAJET simulations of higher order processes by comparing $W+$ four jet events generated using the VECBOS $W+$ four jet process to those generated using the $W+$ three jet process.

Events are generated using the same parton distribution functions assumed for the signal sample. The dynamical scale of the process is set to be the average jet p_T. Systematic uncertainties arising from this choice are estimated by changing the scale to the mass of the W boson in a second sample of events. The background samples are processed through the detector simulation, reconstruction, and event selection in the same manner as for the signal samples.

C. QCD multijet background

The non-W QCD multijet background is estimated, both for the electron and the muon channels, using background-enriched data samples. In the former channels, the sample consists of events containing highly electromagnetic jets failing the electron identification cuts. In the latter, events are selected containing a muon which fails the isolation requirement, but which otherwise passes the muon identification cuts.

VI. TOP DISCRIMINANTS

The key feature that distinguishes top quark events from the W+jets and QCD multijet backgrounds is the fitted mass m_{fit} obtained from kinematic fits of the events to the top quark decay hypothesis. Since the top quark is heavy, the fitted mass tends to be larger for top quark events than for the backgrounds. Therefore, if both the signal to background ratio and the signal are large enough, we should see a clear signal peak in the m_{fit} distribution. However, there is a caveat: this is true only if the cuts to enhance the signal to noise ratio do not significantly distort the fitted mass distributions. Unfortunately, powerful selection variables such as $H_T \equiv \sum E_T^{\text{jet}}$ tend to be highly correlated with the fitted mass. Cuts on them thus introduce severe distortions in m_{fit} which reduce the differences between the distributions for $t\bar{t}$ signal and background, and between the distributions for $t\bar{t}$ signal at different top quark masses, thus impairing the mass measurement.

This distortion of the m_{fit} distribution can be avoided by using variables which are only weakly correlated with the fitted mass. The challenge is to find variables that also provide a useful measure of discrimination between signal and background. After an extensive search of variables that exploit the expected qualitative differences between the kinematics of top quark events and the backgrounds, we have succeeded in finding four variables x_1–x_4 with the desired properties.

This success, however, comes at a price: the discrimination afforded by these variables tends to be weaker than that provided by variables, like H_T, that are mass dependent. But by treating these variables collectively, rather than applying a cut on each separately, we can compensate for their weaker discrimination. It is most effective to combine the variables into a multivariate discriminant $\mathcal{D}(\mathbf{x})$ with the general form

$$\mathcal{D}(\mathbf{x}) \equiv \frac{f_s(\mathbf{x})}{f_s(\mathbf{x}) + f_b(\mathbf{x})}, \tag{6.1}$$

where $\mathbf{x}$ denotes the 4-tuple of mass-insensitive variables and $f_s(\mathbf{x})$ and $f_b(\mathbf{x})$ are functions that pertain to the signal and background, respectively. We choose the functions f_s and f_b so that $\mathcal{D}(\mathbf{x})$ is concentrated near zero for the background and near unity for the signal.

In contrast to previous work [4,7,23] focused on kinematic variables that simply distinguish top quark signal from background, here we identify a subset possessing not only that capability, but also the property of being correlated only weakly with m_{fit}.

In the following sections we describe the variables x_1–x_4 and the two complementary forms we have used for the functions $f_s(\mathbf{x})$ and $f_b(\mathbf{x})$.

A. Variables

The four variables $\{x_1, x_2, x_3, x_4\} \equiv \mathbf{x}$ are defined as follows:

$$\begin{aligned} x_1 &\equiv \not{E}_T \\ x_2 &\equiv \mathcal{A} \\ x_3 &\equiv H_{T2}/H_z \\ x_4 &\equiv \Delta R_{jj}^{\min} E_T^{\min} / E_T^W . \end{aligned} \tag{6.2}$$

Our use of the variable x_1 is motivated by the fact that top quark events have substantial missing transverse energy, due to the neutrino from the leptonically-decaying W boson, while QCD multijet background events do not. Variable x_2 is the aplanarity $\mathcal{A}$ [24], which is defined in terms of the normalized momentum tensor of the jets and the W boson:

$$M_{ab} = \sum_i p_{ia} p_{ib} / \sum_i p_i^2, \tag{6.3}$$

where $\vec{p}_i$ is the three-momentum of the ith object in the laboratory frame, and a, b run over x, y, and z. (For this and the remaining two variables, we use all jets satisfying $E_T^{\text{jet}} > 15$ GeV and $|\eta^{\text{jet}}| < 2$.) If the three eigenvalues of M_{ab} are denoted Q_j such that

$$Q_1 \le Q_2 \le Q_3, \tag{6.4}$$

then

$$\mathcal{A} = \frac{3}{2} Q_1. \tag{6.5}$$

This variable is a measure of the degree to which the final state particles lie out of a plane. In W + jets events, a high p_T W boson recoils against a hadronic system that is typically dominated by a single high p_T jet. In QCD multijet events, two jets, perturbed by gluon radiation, recoil against each other. The signal, by contrast, has a momentum flow that is more spherical. It therefore has a larger aplanarity than do the backgrounds, which have more longitudinal topologies. (The aplanarity for top quark events is expected to decrease with increasing

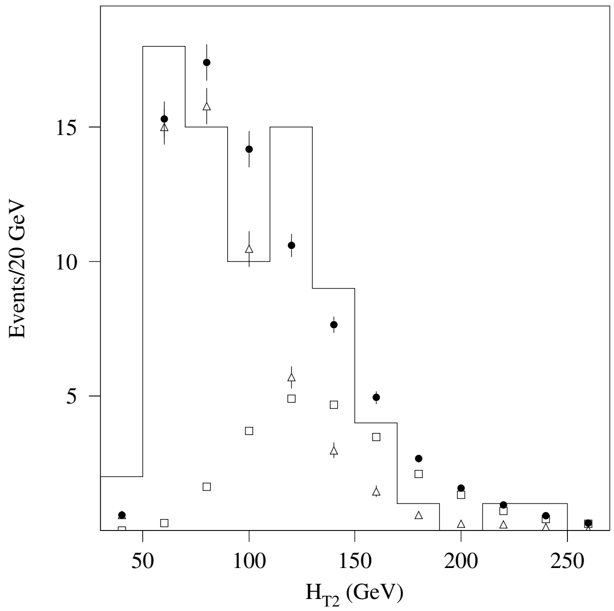

FIG. 11. Plot of H_{T2} for the 77-event candidate sample, compared with the expectation for $m_t = 175$ GeV/c^2 signal plus background (filled circles), signal alone (open squares), and background alone (open triangles). (The normalizations are as in Fig. 2.)

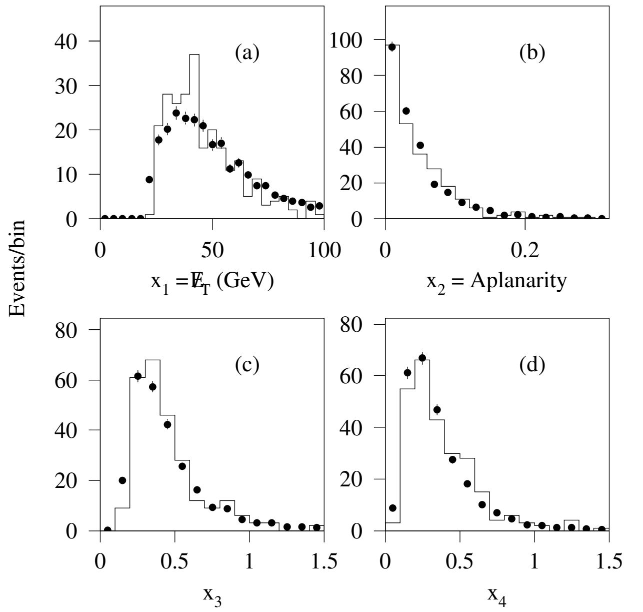

FIG. 12. The variables $x_1 \ldots x_4$ used as input to the top quark discriminants, for W + 3 jet control samples. Histograms are data, and the circles are the expected signal + background mixture.

m_t due to the W boson decay products becoming more collimated. This effect, however, is very small for $m_t <$ 200 GeV/c^2.)

The variable H_T, as noted above, is a powerful discriminant between signal and background. But, since both the signal and background tend to have at least one high p_T jet, we can improve the discrimination somewhat by removing the highest p_T jet from H_T, yielding H_{T2}. A plot of this variable is shown in Fig. 11. This variable, however, is correlated with the fitted mass. Therefore, we divide by another mass-sensitive variable, namely H_z (equal to the sum of $|p_z|$ of the lepton, neutrino, and the jets), in order to reduce that correlation. We thus arrive at variable x_3, which measures the centrality of the events — top quark events being more central than the backgrounds.

The last variable, x_4, is motivated by the observation that the four highest E_T jets in top quark events have a different origin than the jets in W+jets and QCD multijet events. For $t\bar{t}$ events, the four highest E_T jets are mostly from the decay of the $t\bar{t}$ system. These jets tend to be widely separated in $\eta - \phi$ space. For the backgrounds, usually at least one jet is the result of gluon radiation and is therefore somewhat closer to another jet, on average, than the jets in $t\bar{t}$ events. Therefore, we are led to consider the six possible pairs of the four highest E_T jets and take the pair with the minimum separation $\Delta R_{jj}^{\min}$ in $\eta - \phi$ space. We then multiply this minimum separation by the E_T of the lesser jet of the pair, thus constructing a variable akin to the p_T of one jet relative to another. Again, to reduce the correlation with mass, we divide by another mass-sensitive variable, $E_T^W \equiv |E_T^{\rm lep}| + |\not\!\!E_T|$.

We have verified that the variables x_1–x_4 are well modeled by our Monte Carlo calculations. Figure 12 shows the observed distributions of these variables compared with the Monte Carlo predictions for a sample of W+3 jet events, which is dominated by background. In addition, Fig. 13 shows the distributions of these variables for the 77-event candidate sample, compared with Monte Carlo expectations. The Monte Carlo models the data well. We thus use these variables for the multivariate discriminants we now describe.

B. Likelihood discriminant

The correlations among the variables x_1–x_4 are small. Therefore, we assume them to be nearly independent, and write the functions f_s and f_b as

$$f_s(\mathbf{x}) \equiv \prod_{i=1}^{4} s_i^{w_i}(x_i), \qquad (6.6)$$
$$f_b(\mathbf{x}) \equiv \prod_{i=1}^{4} b_i^{w_i}(x_i),$$

where $s_i(x_i)$ and $b_i(x_i)$ are the normalized distributions of variable x_i for signal and background, respectively. These forms reduce to the usual likelihood function for strictly independent variables when the weights $w_i = 1$.

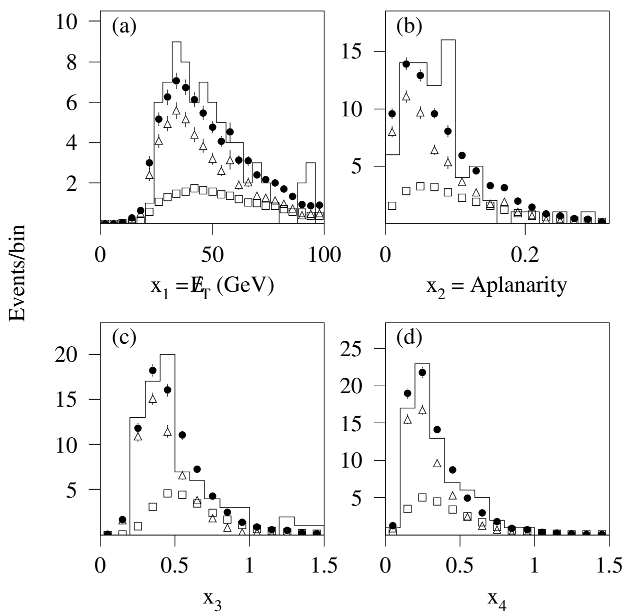

FIG. 13. The variables $x_1 \ldots x_4$ used as input to the top quark discriminants, for the 77-event candidate sample (histogram), $t\bar{t}$ signal plus background for $m_t = 175$ GeV/c^2 (filled circles), signal alone (open squares), and background alone (open triangles). (The normalizations are as in Fig. 2.)

With the weights adjusted slightly away from unity, we can nullify the correlation between $m_{\rm fit}$ and the discriminant $\mathcal{D}_{\rm LB}(\mathbf{x})$ formed from Eqs. (6.1) and (6.6), while maintaining maximal discrimination between high-mass (> 170 GeV/c^2) top events and the background. The subscript "LB" (= "low bias") denotes the fact that cuts on $\mathcal{D}_{\rm LB}$ introduce negligible bias (that is, distortion) in the $m_{\rm fit}$ distributions.

We have found it useful to have a parameterized form for the discriminant $\mathcal{D}_{\rm LB}$. Rather than directly parameterizing the functions f_s and f_b, it is simpler to parameterize the ratio $\mathcal{L} \equiv f_s/f_b$ by using polynomial fits to the four functions $\mathcal{L}_i \equiv s_i(x_i)/b_i(x_i)$ and then computing $\mathcal{L} \equiv \exp \sum_i w_i \ln \mathcal{L}_i$ [25]. We then find $\mathcal{D}_{\rm LB} = \mathcal{L}/(1+\mathcal{L})$.

We also make use of cuts based on $\mathcal{D}_{\rm LB}$ and H_{T2}. All tagged events pass this "LB selection"; for untagged events, we require:

- $\mathcal{D}_{\rm LB} > 0.43$ and
- $H_{T2} > 90$ GeV.

This selection is used in several places to separate the sample into signal-rich and background-rich portions. The cut $\mathcal{D}_{\rm LB} > 0.43$ was chosen to minimize the error on the top quark mass when analyzing Monte Carlo samples. The H_{T2} cut removes very little signal for the top quark masses of interest (see Fig. 11), but provides an easy way of further reducing the background.

We observe 37 events passing this selection, with 31 of them also passing the $\chi^2 < 10$ cut. By scaling down the observed number of events in the precut sample by the efficiency for the LB and $\chi^2 < 10$ cuts as determined from Monte Carlo (taking the signal/background mixture of the precut sample from the cross section analysis), we expect 34.6 ± 2.7 events to pass the LB selection, and 30.5 ± 2.6 events to pass both the LB selection and the $\chi^2 < 10$ cut.

C. Neural network discriminant

The variables x_1–x_4 were chosen to have minimal correlations with the fitted mass. We therefore consider a second, complementary, discriminant in which no attempt is made to nullify the correlation between the discriminant and the fitted mass. We do attempt, however, to account for the small correlations that exist among the variables x_1–x_4. This discriminant, denoted by $\mathcal{D}_{\rm NN}$, is calculated with a neural network (NN) having four input nodes, three hidden nodes, and a single output node, whose value is $\mathcal{D}_{\rm NN}$. The network is trained using the back-propagation algorithm provided in the program JETNET V3.0 [26] using the default training parameters. We use HERWIG $t\bar{t}$ Monte Carlo with $m_t = 170$ GeV/c^2 as the signal, and VECBOS $W +$jets events as the background (equal numbers of each). During training, the target outputs are set to unity for the signal and zero for the background. Under these conditions, the network output approximates the ratio $s(\mathbf{x})/[s(\mathbf{x}) + b(\mathbf{x})]$ [27], where $s(\mathbf{x})$ is the normalized density for the signal and $b(\mathbf{x})$ is the normalized density for the background. Since the correlations among $x_1 \ldots x_4$ are small, as are the correlations with the fitted mass, we should anticipate that the discriminants $\mathcal{D}_{\rm LB}$ and $\mathcal{D}_{\rm NN}$ will provide comparable levels of signal to background discrimination. That this is true is evident, qualitatively, from Fig. 14 which compares the distributions of $\mathcal{D}_{\rm LB}$ and $\mathcal{D}_{\rm NN}$ for top quark events and for the mixture of W+jets and QCD multijet events appropriate for the precuts discussed earlier. The dependence of the discriminants on the top quark mass is indeed small, as shown in Fig. 15. In Fig. 16, we compare the distributions of the two discriminants obtained from the candidate sample to those predicted from Monte Carlo; the agreement is quite good.

Analogous to the LB selection, we will also make use of a cut on $\mathcal{D}_{\rm NN}$. This "NN selection" is defined by $\mathcal{D}_{\rm NN} > 0.6$. This cut value yields roughly the same discrimination as the LB selection. We observe 38 events passing this selection, with 32 of them also passing the $\chi^2 < 10$ cut. Our expectations for these conditions are 32.0 ± 2.0 and 29.0 ± 2.0 events, respectively.

Over the mass range 160–190 GeV/c^2, the efficiencies for the LB and NN selections vary by about 5% and 3%, respectively. This is to be compared with the efficiency for the selection of the standard cross section analysis, which varies by $> 10\%$ over the same range.

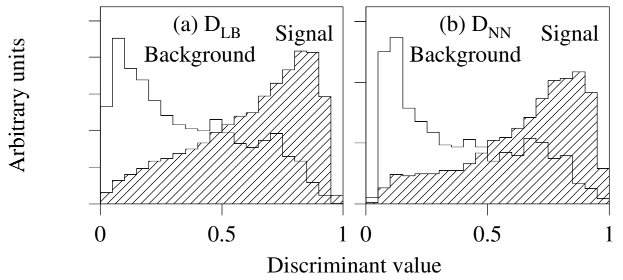

FIG. 14. The discriminant variables (a) $\mathcal{D}_{\mathrm{LB}}$ and (b) $\mathcal{D}_{\mathrm{NN}}$ plotted for the $m_t = 175$ GeV/c^2 $t\bar{t}$ (hatched) sample and the simulated background (unhatched). All histograms are normalized to unity.

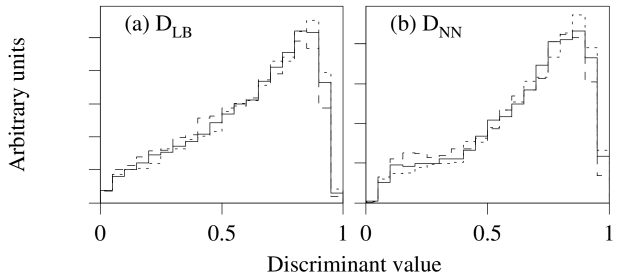

FIG. 15. The discriminant variables (a) $\mathcal{D}_{\mathrm{LB}}$ and (b) $\mathcal{D}_{\mathrm{NN}}$ for $t\bar{t}$ Monte Carlo with $m_t = 150$ GeV/c^2 (dashed lines), $m_t = 175$ GeV/c^2 (solid lines), and $m_t = 200$ GeV/c^2 (dotted lines). All histograms are normalized to unity.

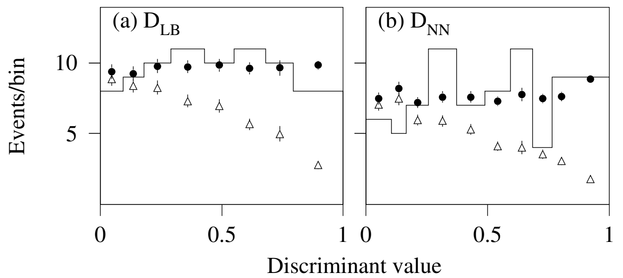

FIG. 16. The discriminant variables (a) $\mathcal{D}_{\mathrm{LB}}$ and (b) $\mathcal{D}_{\mathrm{NN}}$ for the 77-event candidate sample (histogram), $t\bar{t}$ signal plus background (filled circles), and background alone (open triangles). The binnings were chosen such that the predicted signal plus background distribution would be approximately flat.

VII. VARIABLE-MASS FIT

A. Introduction

The method used can be summarized as follows. For each event in the precut sample, we perform a constrained kinematic fit to the hypothesis $t\bar{t} \to l + \mathrm{jets}$ to arrive at a "fitted mass" m_{fit}. Events which fit poorly are discarded. For each event, we also compute a top quark discriminant $\mathcal{D}$ (either $\mathcal{D}_{\mathrm{LB}}$ or $\mathcal{D}_{\mathrm{NN}}$). The events are then entered into a two-dimensional histogram in the $(\mathcal{D}, m_{\mathrm{fit}})$ plane. Similar histograms are also constructed for a sample of background events and for signal Monte Carlo at various top quark masses. For each of these MC masses, we fit a sum of the signal and background histograms to the data histogram. This fit yields a background fraction and a corresponding likelihood value. These likelihood values are then plotted as a function of the top quark mass, and the final result extracted by fitting a quadratic function to their logarithms.

B. Kinematic fit

The goal of the kinematic fit is to constrain a measured event to the hypothesis

$$p\bar{p} \to t\bar{t} + X \to (W^+b)(W^-\bar{b}) + X \to (l\nu b)(q\bar{q}\bar{b}) + X \tag{7.1}$$

(or the charge conjugate) and thus arrive at an estimate m_{fit} of the top quark mass. There is a complication, however, in that when reconstructing the event, we do not know *a priori* which observed jet corresponds to which parton. In fact, due to QCD radiative effects, jet merging and splitting during reconstruction, and jet reconstruction inefficiencies, the observed jets may have no one-to-one correspondence with the unfragmented partons from the $t\bar{t}$ decay. Nevertheless, the fitted mass m_{fit} constructed from the observed jets is correlated with the true top quark mass and can thus be used for a measurement; however, m_{fit} should not be thought of as "the top quark mass" for a particular event.

The inputs to the fit are the kinematic parameters of the lepton, the jets, and the missing transverse energy vector $\vec{\not{E}}_T$. Only the four jets with the largest E_T within $|\eta| < 2.5$ are used in the fit (any additional jets are assumed to be due to initial state radiation). We parameterize electrons and jets in terms of energy E, azimuthal angle ϕ, and pseudorapidity η. For muons, we parameterize the momentum in terms of $k = 1/p$, since the resolution is more nearly Gaussian in that variable. The muon direction is also represented as (ϕ, η). Leptons and light quarks are fixed to zero mass; b quarks are fixed to a mass of 5 GeV/c^2. The transverse momentum of the neutrino is taken to be $\vec{\not{E}}_T$. However, we do not use $\vec{\not{E}}_T$ directly in the fit, as it is correlated with all the other objects in the event. Instead, we use the x and y components of

$$\vec{k}_T = \vec{\not{E}}_T + \vec{E}_T^{\rm lep} + \sum_{4\ {\rm jets}} \vec{E}_T^{\rm jet}. \tag{7.2}$$

This can be thought of as the transverse momentum of the $t\bar{t}$ pair. Note that this is not necessarily a small quantity if the event has more than four jets. One additional variable is needed to uniquely define the event kinematics: we take that to be the z-component of the neutrino momentum p_z^ν. This variable is not measured, but is determined by the fit. This gives a total of 18 variables.

With this parameterization, there are three kinematic constraints which can be applied:

$$\begin{aligned} m(t \to l\nu b) &= m(\bar{t} \to q\bar{q}\bar{b}) \\ m(l\nu) &= M_W \\ m(q\bar{q}) &= M_W. \end{aligned} \tag{7.3}$$

Three constraints and one unmeasured variable allow for a 2C fit.

Since we do not know the correspondence between jets and partons, we try all twelve distinct assignments of the four jets to the partons ($b\bar{b}qq$). (But if the event has a b-tag, only the six permutations in which the tagged jet is used as a b quark are considered.) Once a permutation is chosen, we apply the parton-level and η-dependent jet corrections described in Sec. IV. We apply a loose cut on the hadronic W boson mass before the fit: $40 < m(q\bar{q}) < 140$ GeV/c^2. Permutations failing this cut are rejected without being fit in order to speed up the computation. We arrange the measured variables into a vector $\mathbf{x}^m$ and form the χ^2

$$\chi^2 = (\mathbf{x} - \mathbf{x}^m)^T \mathbf{G} (\mathbf{x} - \mathbf{x}^m), \tag{7.4}$$

where $\mathbf{G}$ is the inverse error matrix. This χ^2 is then minimized subject to the kinematic constraints of Eq. (7.3). The minimization algorithm uses the method of Lagrange multipliers; the nonlinear constraint equations are solved using an iterative technique. (The algorithm used is very similar to that of the SQUAW kinematic fitting program [28]; a detailed description may be found in Ref. [29].) If this minimization does not converge, the permutation is rejected. A permutation is also rejected if $\chi^2 > 10$. For each surviving permutation, this method gives a fitted mass $m_{\rm fit}$ and a χ^2. We pick the $m_{\rm fit}$ value corresponding to the smallest χ^2 as $m_{\rm fit}$ for the event.

There is one additional wrinkle to the above procedure. In order to start each fit, we must specify an initial value for the unmeasured variable p_z^ν. We choose it so that the two top quarks are assigned equal mass. This yields a quadratic equation for p_z^ν. If the solutions are complex, the real part is used. Otherwise, there are two real solutions. Both are tried, and the fit which gives the smaller χ^2 is retained. Note, however, that since p_z^ν does not enter into the χ^2 (its measurement error is effectively infinite), the only effect its initial value can have on the final result is to influence which local minimum the fit will find, should there happen to be more than one. In about 70% of cases, two distinct neutrino solutions yield fit results which differ by less than 5 GeV/c^2.

The error matrix $\mathbf{G}^{-1}$ is taken to be diagonal. The resolutions used are given in Table VI. (The lepton angular resolutions are much smaller than the other resolutions, and can be taken to be effectively zero.) In most cases, these resolutions were derived from $t\bar{t}$ Monte Carlo events by comparing reconstructed objects to generator-level objects.

Results of this procedure on Monte Carlo $t\bar{t}$ samples are shown in Fig. 17. Figure 17(a) shows results using the HERWIG partons directly, before any QCD evolution has taken place. A rather sharp peak is seen; further, about 80% of the time, the permutation with the lowest χ^2 is the one which is actually correct. The residual width seen in the plot is due mainly to the non-zero widths of the W bosons. Figure 17(b) shows results from the same sample, but after QCD evolution and jet fragmentation. The final state particles are clustered together in cones of width $\Delta R = 0.5$ in order to simulate the action of the jet reconstruction algorithm. This distribution is considerably broader. There are fewer events in the hatched plot because it is not always possible to uniquely define the correct permutation. Due to splitting and merging effects, jet finding inefficiencies, and jets falling below the selection threshold, the correct permutation can be uniquely identified in only about 50% of events. In that case, the correct permutation is the lowest χ^2 permutation about 40% of the time. Finally, Fig. 17(c) shows results for a sample which has been through the full detector simulation and reconstruction. The resulting distribution has essentially the same width as that of Fig. 17(b); this indicates that the dominant contribution to the width of this distribution comes from QCD radiation and jet combinatoric effects, and not from the detector resolution.

The (MC) fit χ^2 distributions resulting from the fit to the correct jet permutation are shown in Fig. 18. The distributions agree reasonably well with the expectations for a two degree-of-freedom χ^2, except for a tail at the high end due to non-Gaussian tails in the resolutions. The (MC) $m_{\rm fit}$ distributions for the four channels are shown in Fig. 19.

Figure 20 shows the distributions which result after the jets in each Monte Carlo event are scaled up or down by the per-jet systematic uncertainty of 2.5% + 0.5 GeV. This shifts the fitted mass by approximately ± 3.7 GeV/c^2.

Figure 21 shows the fitted mass distribution for several top quark masses and for the background.

A possible objection to the fit method described here is that it does not take into account the intrinsic widths of the W boson and top quark decays. To investigate this, an alternate fitting method was tried which explicitly incorporates these widths. This method is based on a standard unconstrained minimization package (MINUIT [30]). The quantity minimized is the χ^2 as defined in Eq. (7.4)

TABLE VI. Object resolutions. The operator $\oplus$ denotes a sum in quadrature.

	Energy resolution	$\sigma(\phi)$	$\sigma(\eta)$
Electrons	$\sigma(E_T)/E_T = 0.0157 \oplus 0.072\ \mathrm{GeV}^{1/2}/\sqrt{E_T} \oplus 0.66\ \mathrm{GeV}/E_T$		
Muons	$\sigma(1/p) = C^{\mathrm{a}} \oplus 0.2/p$		
Jets			
$0 < \lvert\eta_{\mathrm{det}}\rvert < 0.8$	$\sigma(E)/E = 0.036 \oplus 1.145\ \mathrm{GeV}^{1/2}/\sqrt{E}$	0.04 rad	0.04
$0.8 < \lvert\eta_{\mathrm{det}}\rvert < 1.4$	$\sigma(E)/E = 0.082 \oplus 1.264\ \mathrm{GeV}^{1/2}/\sqrt{E}$	0.05 rad	0.05
$1.4 < \lvert\eta_{\mathrm{det}}\rvert < 2.0$	$\sigma(E)/E = 0.046 \oplus 1.305\ \mathrm{GeV}^{1/2}/\sqrt{E}$	0.05 rad	0.05
k_T	$\sigma(k_{Tx}) = \sigma(k_{Ty}) = 12$ GeV		

[a]$C = 0.0045/(\mathrm{GeV}/c)$ if the muon track could be matched with a track in the central detector; $C = 0.01/(\mathrm{GeV}/c)$ otherwise.

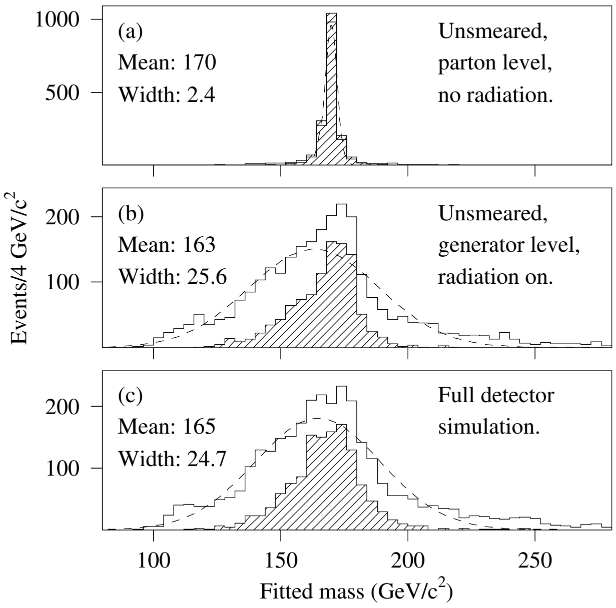

FIG. 17. Tests of kinematic fit method on $t\bar{t}$ Monte Carlo samples ($m_t = 170\ \mathrm{GeV}/c^2$, $e+$jets channel). (a) Using HERWIG partons directly. (b) Final state Monte Carlo particles, after clustering into $R = 0.5$ cones. (c) After full detector simulation and reconstruction. The hatched plots show the results for the correct jet permutation (regardless of whether or not it has the lowest χ^2). Displayed means and widths are from a Gaussian fit, shown by the dashed curve.

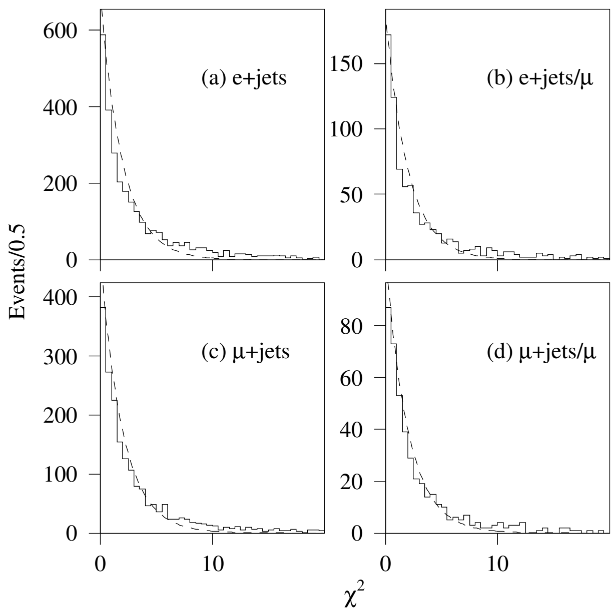

FIG. 18. Fit χ^2 distributions for the correct jet permutation for $t\bar{t}$ Monte Carlo samples ($m_t = 170\ \mathrm{GeV}/c^2$). The dashed curve is the χ^2 distribution for two degrees of freedom, normalized to the area of the histogram.

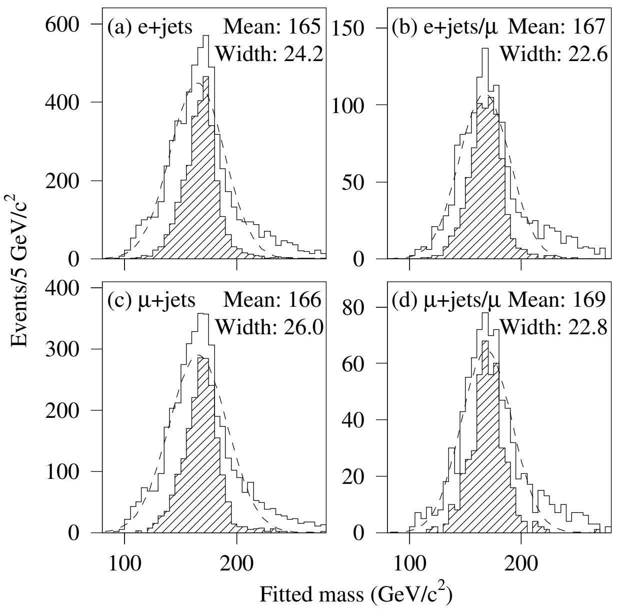

FIG. 19. Fitted mass distributions for $t\bar{t}$ Monte Carlo samples ($m_t = 170$ GeV/c^2) for the jet permutation with the lowest χ^2. Hatched histograms show the results for the correct jet permutation (regardless of whether or not it has the lowest χ^2). Displayed means and widths are from a Gaussian fit, shown by the dashed curve.

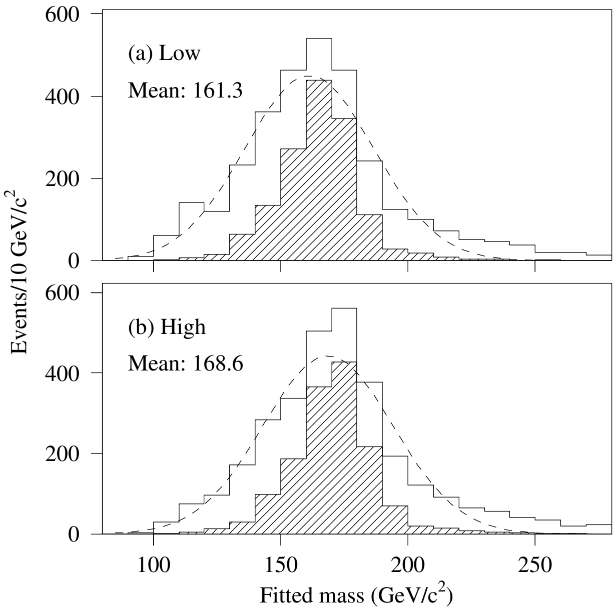

FIG. 20. Fitted mass distributions for $t\bar{t}$ Monte Carlo samples ($m_t = 170$ GeV/c^2, $e+$jets channel). With jets scaled (a) down and (b) up by $2.5\% + 0.5$ GeV. Hatched histograms show the results for the correct jet permutation (regardless of whether or not it has the lowest χ^2). Displayed means are from a Gaussian fit, shown by the dashed curve.

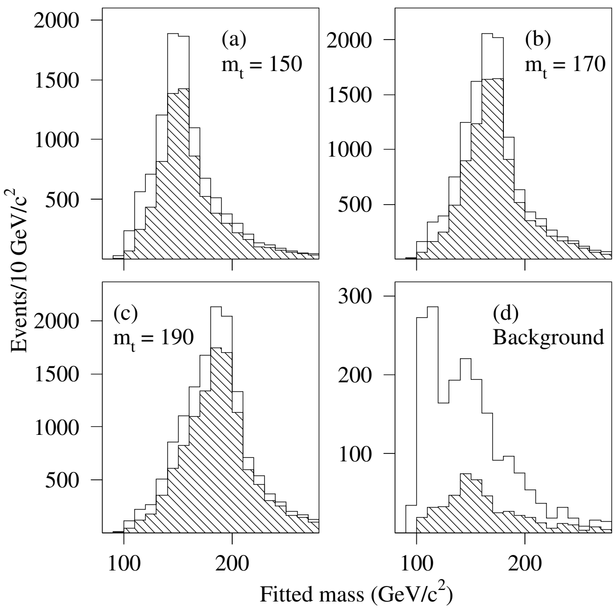

FIG. 21. Fitted mass distributions, all channels combined. Shown is $t\bar{t}$ Monte Carlo with (a) $m_t = 150$ GeV/c^2, (b) $m_t = 170$ GeV/c^2, and (c) $m_t = 190$ GeV/c^2 and (d) background. The hatched distributions are after the LB selection is applied.

with three Breit-Wigner constraint terms added: two for the two W bosons, and one for the top quark mass difference:

$$\begin{aligned}\chi^{2\prime} = \chi^2 &- 2\ln\frac{\Gamma_W^2/4}{\Gamma_W^2/4 + (m(l\nu) - M_W)^2} \\ &- 2\ln\frac{\Gamma_W^2/4}{\Gamma_W^2/4 + (m(q\bar{q}) - M_W)^2} \\ &- 2\ln\frac{\Gamma_t^2}{\Gamma_t^2 + (m(l\nu b) - m(q\bar{q}\bar{b}))^2}.\end{aligned} \tag{7.5}$$

(The factor of 4 difference in the last term comes from convoluting two Breit-Wigner functions centered on $m(l\nu b)$ and $m(q\bar{q}\bar{b})$.) The W boson width is taken to be 2 GeV/c^2. The top quark width is taken to depend on the mass as $\Gamma_t = (\alpha m_t)^3$; the proportionality constant α is set so that $\Gamma_t = 0.6$ GeV/c^2 at $m_t = 140$ GeV/c^2. (Here, $m_t = (m(l\nu b) + m(q\bar{q}\bar{b}))/2$.) These widths are small compared to the experimental resolutions. The results of this procedure are compared to those from the Lagrange-multiplier based fitter in Fig. 22. In most cases, the results are nearly identical, implying that neglecting the widths is not a serious problem. Since this algorithm takes several times longer to execute, it is not used further.

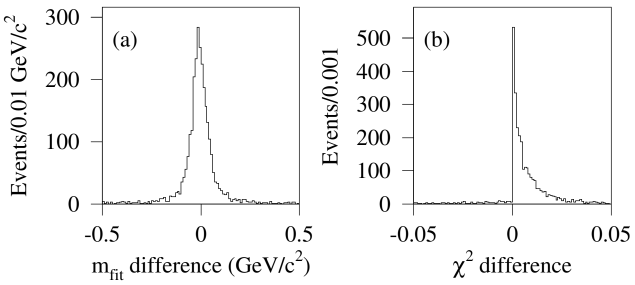

FIG. 22. Differences between the results obtained from the MINUIT-based fitter and the Lagrange-multiplier based fitter for (a) $m_{\rm fit}$ and (b) χ^2. (For $t\bar{t}$ Monte Carlo with $m_t = 170$ GeV$/c^2$, $e+$jets channel.)

C. Likelihood fit

The next problem to be solved is the extraction of the top quark mass from the data sample, which is a mixture of signal and background. This is done using a binned Poisson-statistics maximum-likelihood fit at discrete top quark masses. (The method is described in more detail in Ref. [31].)

We bin the data according to some characteristics of the events. (For this analysis, we will be using $m_{\rm fit}$ and either $\mathcal{D}_{\rm LB}$ or $\mathcal{D}_{\rm NN}$.) Call the number of bins M, the total number of events N, and the number of events in each bin N_j.

We also know the distribution expected for different values of the top quark mass, and also for the background. (This is from Monte Carlo except for the QCD multijet background.) For both the signal and background, we have a distribution of events among the M bins; call the numbers of events in each bin of these distributions A^s_j and A^b_j.

We regard these distributions as drawn from "true" distributions a^s_j and a^b_j, and write the probability for seeing the observed data set D given these parameters as a Poisson likelihood

$$L(D|A,a,p) = \prod_{j=1}^{M} q(N_j, p_s a^s_j + p_b a^b_j)\, q(A^s_j, a^s_j)\, q(A^b_j, a^b_j), \tag{7.6}$$

where q is the Poisson distribution $q(N,a) \equiv e^{-a}a^N/N!$ and p_s and p_b are the signal and background strengths. These strengths can be related to the number of expected events n_s and n_b by $p_s = n_s/(M + \sum_j A^s_j)$, and similarly for n_b. (The M term in the denominator ensures that the sum of the maximum likelihood estimates for n_s and n_b equals N. See Ref. [31] for further discussion. Note that usually $M \ll \sum_j A_j$.) The total number of events expected is thus $n_j = p_s a^s_j + p_b a^b_j$. We eliminate the a_j's from this likelihood by integrating over them; the result is

$$L(D|A,p) = \prod_{j=1}^{M} \sum_{k=0}^{N_j} \frac{p_s^k}{(1+p_s)^{A^s_j+k+1}} \frac{p_b^{N_j-k}}{(1+p_b)^{A^b_j+N_j-k+1}} \times \binom{A^s_j+k}{k} \binom{A^b_j+N_j-k}{N_j-k}. \tag{7.7}$$

Following Ref. [10], we then modify the likelihood by dividing by the constant factor

$$\prod_j q(N_j, N_j). \tag{7.8}$$

This has the effect of making the quantity $-2\ln L$ behave asymptotically like a χ^2 distribution. (Note, however, that for our experiment, the sample size is too small for this asymptotic behavior to be accurately realized.)

We now have a set of signal models, each corresponding to a different top quark mass m_t. For each signal model, we fit it plus the background to the data, yielding n_s and n_b. A maximum likelihood fit is used, based on MINUIT [30]. The minimum value of $-\ln L$ is retained; call this $-\ln L_{\rm min}$. The resulting values of $(m_t, -\ln L_{\rm min})$ then define a likelihood curve as a function of top quark mass.

We also define a statistical uncertainty on $-\ln L_{\rm min}$ due to the finite Monte Carlo statistics. This is done by the simple method of taking in turn each bin j in the input Monte Carlo histograms, varying the contents up or down by $\sqrt{A_j}$, and re-evaluating the likelihood. (To save time, the fit for n_s and n_b is not redone for each variation; early testing showed it to make very little difference.) The resulting variations in $-\ln L_{\rm min}$ for each bin are then added in quadrature. This uncertainty is calculated separately for the signal and background samples; however, any effects from fluctuations in the background sample will be highly correlated from mass point to mass point. Thus, the uncertainties shown on the plots and used in the fit below come from the signal samples only.

The final step is to extract a mass value from this set of $(m_t, -\ln L_{\rm min})$ points. This is done by fitting a quadratic function to the smallest $-\ln L_{\rm min}$ and the four closest points on each side. The points are weighted by the statistical uncertainties assigned to the $-\ln L_{\rm min}$ values. The position of the minimum of this quadratic defines the mass estimate, and its width (where the curve has risen by 0.5) gives an uncertainty estimate. We also want estimates for n_s and n_b. For each mass m_t, we have a separate estimate for n_s and n_b returned from MINUIT. The final estimates of these values are determined by a linear interpolation between the two points bracketing the final m_t estimate. The uncertainties are found in the same manner.

For comparison, some results are also given using 11 points instead of 9 for the polynomial fit, and using a cubic function instead of a quadratic one.

TABLE VII. NN bin definitions.

Bin	$\mathcal{D}_{\mathrm{NN}}$ range
1	0.000 – 0.105
2	0.105 – 0.166
3	0.166 – 0.257
4	0.257 – 0.373
5	0.373 – 0.488
6	0.488 – 0.595
7	0.595 – 0.687
8	0.687 – 0.766
9	0.766 – 0.846
10	0.846 – 1.000

D. Fitting variables and binning

From each event, we derive two variables: the fitted mass m_{fit} and a discriminant $\mathcal{D}$. We use these variables to bin the data into a two-dimensional histogram. The top quark mass is then extracted from a fit to the expectations from Monte Carlo, as described in the previous section.

Two different discriminants and histogram binnings are used. For both binnings, the fitted mass axis has twenty bins of width 10 GeV$/c^2$ over the range 80 to 280 GeV$/c^2$. They differ in the definition of the discriminant axis. For the "LB" analysis, the discriminant axis is divided into two bins, the first bin containing events which fail the LB selection (as defined in Sec. VI B), and the second containing events which pass it. (Recall that all tagged events pass the LB selection.) For the "NN" analysis, the discriminant axis is the NN variable $\mathcal{D}_{\mathrm{NN}}$. (Note that tagging information is not used in forming $\mathcal{D}_{\mathrm{NN}}$.) There are ten unevenly spaced bins, as defined in Table VII. These bin boundaries were chosen so that the expected signal + background distribution populates the bins approximately uniformly. There are thus 40 bins in the LB binning, and 200 bins in the NN binning. Examples of the resulting histograms are shown in Fig. 23.

These histograms are generated separately for each of the four channels. They are then combined using the set of fixed weights given in Table VIII. We derive these numbers by calculating the expected signal and background in each channel using the same techniques as used for the cross section measurement [7] (except that only the precuts are applied). We also combine the histograms for VECBOS $W+$jets background and the QCD multijet background using a fixed QCD fraction of $(22 \pm 5)\%$, derived in the same manner.

E. Fits to data

The results of the kinematic fit for the candidate events are given in Tables IX through XII. (Complete details of the candidate events are available in Ref. [32].) There are

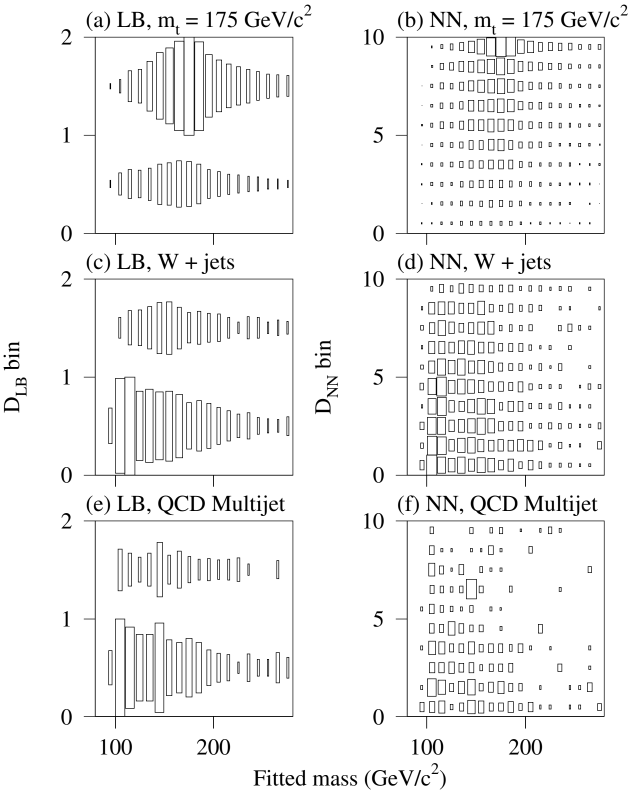

FIG. 23. Monte Carlo histograms for LB and NN analyses for $t\bar{t}$ Monte Carlo with $m_t = 175$ GeV$/c^2$, VECBOS $W+$jets background, and QCD multijet background. More top quark-like events are towards the top of the plots. Box areas are proportional to the histogram contents.

TABLE VIII. Fraction of events expected in each channel after the precuts.

	$e+\text{jets}$	$e+\text{jets}/\mu$	$\mu+\text{jets}$	$\mu+\text{jets}/\mu$
HERWIG $t\bar{t}$				
110–150 GeV/c^2	0.376 ± 0.020	0.085 ± 0.013	0.468 ± 0.025	0.071 ± 0.018
155–170 GeV/c^2	0.418 ± 0.018	0.097 ± 0.011	0.425 ± 0.021	0.059 ± 0.015
172–190 GeV/c^2	0.427 ± 0.016	0.093 ± 0.010	0.409 ± 0.019	0.071 ± 0.013
195–230 GeV/c^2	0.416 ± 0.014	0.097 ± 0.009	0.419 ± 0.018	0.068 ± 0.012
VECBOS	0.531 ± 0.077	0.015 ± 0.017	0.441 ± 0.079	0.013 ± 0.003
QCD	0.443 ± 0.111	0.013 ± 0.030	0.488 ± 0.115	0.056 ± 0.020

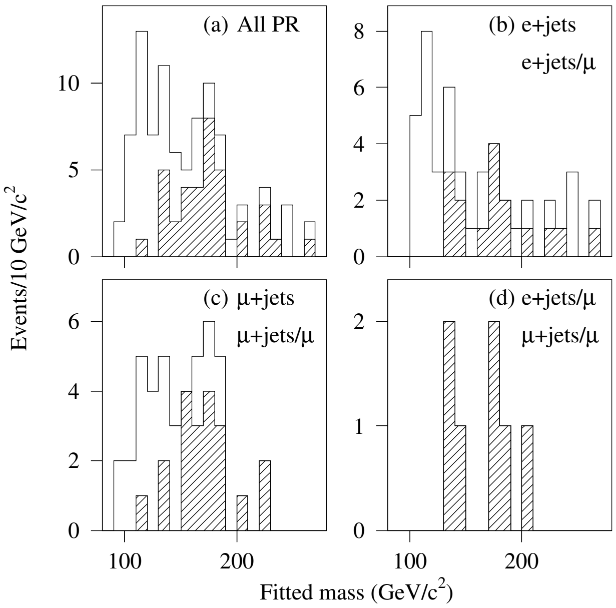

FIG. 24. Fitted mass distributions for candidate events. The hatched histograms show the LB subsample.

91 events passing the precuts (PR). One of these, however, had no successful fits, and is not considered further. Thirty-six of these events then pass the LB selection. The distributions of the fitted masses of these candidates are shown in Fig. 24. When the $\chi^2 < 10$ cut is imposed, there are 77 PR events and 31 LB events. Distributions of their fitted masses are shown in Fig. 25. The χ^2 distribution of the 90 events is shown in Fig. 26. It compares well to the expectation from Monte Carlo.

Results of likelihood fits to the data sample are shown in Table XIII. Several methods of extracting the final top quark mass are tabulated. The labels "quad$N$" and "cub$N$" denote, respectively, N-point quadratic and cubic fits to the negative log likelihood values. The reported central value is the minimum of the fit curve, and the uncertainty indicated is the width of the curve where it has risen by 0.5 from the minimum. For the "avg" fits, the central value is the mean of the likelihood curve (calculated using trapezoidal-rule integration), and the reported uncertainty on the mass is the symmetric interval around the mean containing 68% of the likelihood.

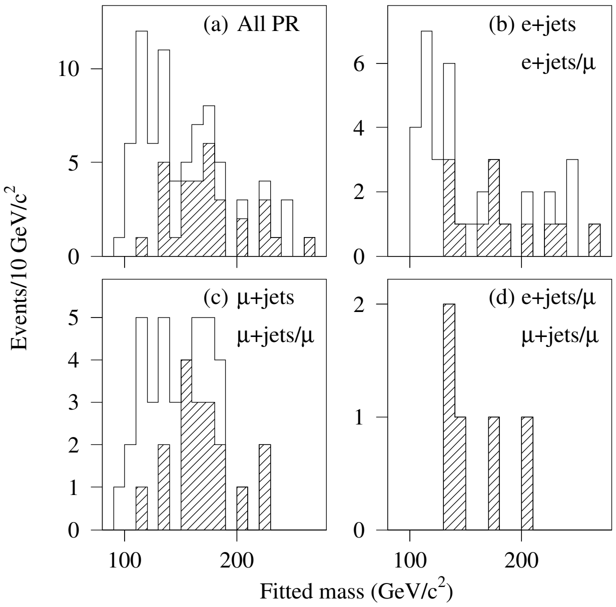

FIG. 25. Fitted mass distributions for candidate events with $\chi^2 < 10$. The hatched histograms show the LB subsample.

TABLE IX. Kinematic fit results and top quark discriminants for events in the $e+$jets channel for the jet permutation with the smallest χ^2. The "Perm" column gives the assignment of the jets to partons, listed in order of decreasing jet E_T: B_l and B_h denote the b quarks associated with the leptonically and hadronically decaying top quarks, respectively, while W denotes the quarks from the hadronically decaying W boson. The fitted mass $m_{\rm fit}$ is in GeV/c^2.

	Run	Event	Perm.	$m_{\rm fit}$	χ^2	$\mathcal{D}_{\rm LB}$	$\mathcal{D}_{\rm NN}$
[c]	62199	15224	B_lWWB_h	265.4	15.9	0.09	0.21
[b][c]	62431	788	WB_hB_lW	241.7	0.23	0.16	0.09
[a][b][c]	63066	13373	B_lWB_hW	206.8	1.35	0.85	0.95
[b][c]	64464	21611	B_hWWB_l	115.7	0.64	0.22	0.31
[a][b][c]	81949	12380	B_lB_hWW	132.7	1.10	0.77	0.82
[b][c]	82024	44002	WB_lB_hW	130.2	0.97	0.06	0.31
[b][c]	82220	20012	B_hWB_lW	120.8	2.53	0.03	0.06
	82996	24461	WB_hWB_l	166.8	31.8	0.73	0.74
	84331	13271	B_hWB_lW	116.8	14.4	0.25	0.27
[b][c]	84890	28925	B_hWB_lW	126.4	0.78	0.06	0.07
[a][b][c]	85917	22	B_lWWB_h	162.3	2.26	0.79	0.81
[b][c]	86518	11716	B_hWWB_l	243.5	0.54	0.18	0.29
[a][b][c]	86601	33128	WB_lB_hW	179.2	0.39	0.43	0.29
[a][b][c]	87063	39091	B_hWB_lW	188.4	0.39	0.58	0.63
[b][c]	87104	25823	WB_hB_lW	119.9	2.11	0.06	0.09
[b][c]	87329	13717	B_hB_lWW	242.1	1.95	0.39	0.23
[b][c]	87446	14294	WWB_hB_l	118.3	1.11	0.59	0.52
	88038	14829	WWB_hB_l	101.0	12.8	0.37	0.28
[c]	88044	9807	WB_lWB_h	145.2	34.0	0.09	0.11
[a][b][c]	88045	35311	WB_hWB_l	178.2	2.71	0.83	0.81
[b][c]	88125	15437	WB_hWB_l	115.9	0.16	0.78	0.74
[b][c]	88463	3627	WB_hWB_l	111.7	9.93	0.16	0.46
[b]	88588	15993	WWB_hB_l	103.4	7.44	0.29	0.30
[a][b][c]	89484	11741	B_hB_lWW	135.0	0.76	0.53	0.58
[b][c]	89550	18042	WWB_hB_l	103.5	0.07	0.30	0.27
[a]	89708	24871	WB_hB_lW	144.6	20.1	0.62	0.74
[a][b][c]	89936	6306	WB_hB_lW	220.4	1.29	0.50	0.68
[a][b][c]	89972	13657	WB_hB_lW	176.7	9.08	0.65	0.77
[b][c]	90108	31611	WB_hWB_l	137.4	0.41	0.21	0.21
[b][c]	90435	32258	B_hB_lWW	154.1	1.05	0.27	0.62
[b][c]	90496	28296	B_hWB_lW	112.9	0.28	0.23	0.19
[b]	90693	8678	B_hB_lWW	105.5	8.98	0.51	0.27
[c]	90795	14246	B_hWB_lW	193.9	12.8	0.09	0.07
[b][c]	90804	6474	WWB_lB_h	114.2	0.64	0.34	0.59
[b][c]	91923	502	WB_hWB_l	162.1	0.14	0.09	0.15
[b][c]	92013	11825	WB_hB_lW	134.1	3.68	0.11	0.15
[b][c]	92217	109	WWB_hB_l	107.8	0.58	0.77	0.82
[b][c]	92278	21744	WB_hWB_l	125.9	7.26	0.17	0.31
[a][b][c]	92673	4679	B_lB_hWW	267.7	1.85	0.92	0.97
[b][c]	94750	4683	B_hWWB_l	201.5	3.63	0.32	0.49
[a]	96329	13811		—		0.54	0.79
[b][c]	96676	79957	WB_lB_hW	224.1	0.47	0.36	0.46
[a][b][c]	96738	27592	B_lWWB_h	236.6	5.68	0.60	0.83

[a]Passes LB selection.
[b]Used in variable-mass analysis.
[c]Used in pseudolikelihood analysis.

TABLE X. Same as Table IX for the $\mu+$jets channel.

	Run	Event	Perm.	$m_{\rm fit}$	χ^2	$\mathcal{D}_{\rm LB}$	$\mathcal{D}_{\rm NN}$
[b][c]	61514	4537	B_hWB_lW	120.8	3.40	0.26	0.59
[a][b][c]	63183	13926	WWB_hB_l	133.7	1.26	0.84	0.83
[a][b][c]	63740	14197	B_lWB_hW	185.3	2.56	0.94	0.96
[b][c]	80703	31477	WB_hB_lW	167.2	0.54	0.24	0.40
[a][b][c]	81909	11966	B_hWB_lW	162.9	1.11	0.67	0.66
[b]	81949	13778	WB_hWB_l	109.2	8.25	0.27	0.25
[b][c]	82639	11573	WB_lWB_h	117.3	2.24	0.35	0.47
[a][b][c]	82694	25595	WB_lWB_h	114.0	2.03	0.56	0.53
[a][b][c]	84696	29253	WB_hB_lW	221.0	1.05	0.74	0.89
[b][c]	84728	18171	B_hB_lWW	136.0	3.65	0.40	0.38
[b][c]	85888	28599	B_hWWB_l	189.6	5.78	0.18	0.09
[a][b][c]	87063	14368	WWB_hB_l	182.1	0.02	0.50	0.72
	87604	14282	B_lWWB_h	90.6	40.6	0.14	0.38
[a][c]	87820	6196	B_hB_lWW	178.0	17.8	0.87	0.97
[a][b][c]	88464	2832	B_hWB_lW	154.1	0.14	0.87	0.93
[a][b][c]	88530	7800	WB_lB_hW	151.2	0.08	0.62	0.60
	88597	1145	WWB_hB_l	124.6	10.2	0.20	0.42
[b][c]	88603	2131	WB_lWB_h	123.7	0.66	0.13	0.17
[b][c]	89751	27345	B_hWWB_l	132.4	1.14	0.15	0.14
[a][b][c]	89943	19016	WB_hB_lW	163.7	0.03	0.65	0.74
[b][c]	90133	14110	WB_hWB_l	169.4	4.88	0.26	0.28
[a][b][c]	90660	20166	WB_lB_hW	222.6	1.28	0.70	0.90
[a][b][c]	90690	12392	B_hWB_lW	153.3	0.58	0.70	0.78
[b][c]	90836	14924	WB_lWB_h	147.4	3.13	0.07	0.08
[b][c]	90864	17697	WB_hWB_l	96.6	0.81	0.44	0.62
[b][c]	91359	15030	WB_hWB_l	118.9	1.81	0.54	0.60
[b][c]	92081	3825	WB_hB_lW	117.7	3.72	0.07	0.40
[b][c]	92082	34466	WB_hB_lW	176.2	0.30	0.31	0.49
[a]	92114	1243	WB_lB_hW	187.0	11.7	0.96	0.96
[a][b][c]	92126	21544	B_lWWB_h	157.2	0.02	0.82	0.91
[b][c]	92142	27042	WB_lB_hW	148.7	4.71	0.24	0.21
[b][c]	92226	34133	WB_hB_lW	140.3	0.49	0.41	0.66
[b][c]	92714	4141	WWB_hB_l	106.4	6.28	0.43	0.59
[a][b][c]	92714	12581	B_lB_hWW	166.3	1.66	0.57	0.66
[b][c]	94750	1147	WWB_lB_h	126.9	0.82	0.32	0.23
[b][c]	96258	2707	B_lWB_hW	171.2	1.02	0.49	0.28
[b][c]	96264	93611	B_hWWB_l	111.7	0.41	0.06	0.14
[b][c]	96280	14555	WB_hB_lW	133.8	0.07	0.69	0.68
[b][c]	96287	20104	WB_hB_lW	182.5	5.64	0.16	0.14
[a][b][c]	96399	32921	B_lB_hWW	172.8	0.28	0.68	0.83
[a][b][c]	96591	39318	B_hB_lWW	174.3	0.94	0.55	0.75

[a]Passes LB selection.
[b]Used in variable-mass analysis.
[c]Used in pseudolikelihood analysis.

TABLE XI. Same as Table IX for the $e+$jets/μ channel.

	Run	Event	Perm.	$m_{\rm fit}$	χ^2	$\mathcal{D}_{\rm LB}$	$\mathcal{D}_{\rm NN}$
[a]	62199	13305	B_lB_hWW	173.2	40.0	0.55	0.61
[a][b][c]	85129	19079	WB_lB_hW	137.0	0.93	0.81	0.85
[a][b][c]	86570	8642	B_hWWB_l	144.5	0.66	0.74	0.29
[a]	89372	12467	B_hWWB_l	186.6	22.1	0.23	0.25

[a]Passes LB selection.
[b]Used in variable-mass analysis.
[c]Used in pseudolikelihood analysis.

TABLE XII. Same as Table IX for the $\mu+\text{jets}/\mu$ channel.

	Run	Event	Perm.	m_{fit}	χ^2	$\mathcal{D}_{\text{LB}}$	$\mathcal{D}_{\text{NN}}$
a b c	58203	4980	WB_hB_lW	138.3	0.25	0.56	0.62
a b c	91712	22	B_hWB_lW	203.3	0.44	0.51	0.44
a b c	92704	14022	WB_hB_lW	175.8	0.11	0.79	0.88

[a]Passes LB selection.
[b]Used in variable-mass analysis.
[c]Used in pseudolikelihood analysis.

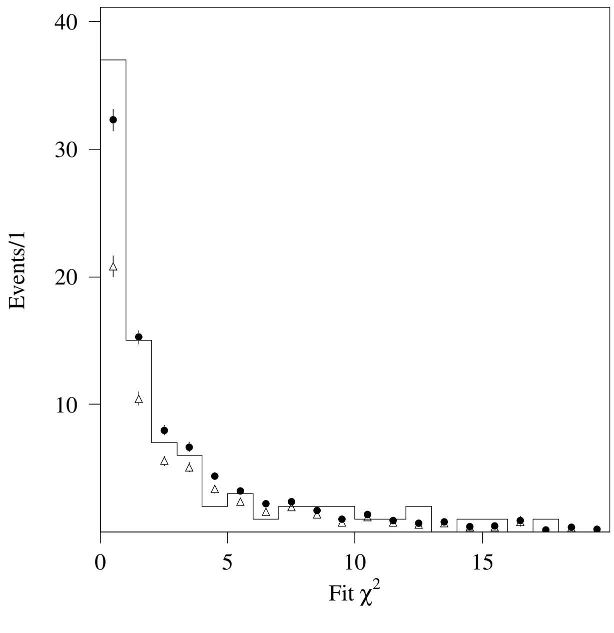

FIG. 26. Fit χ^2 distribution from data (histogram), the expected $t\bar{t}$ signal + background (filled circles), and background alone (open triangles).

Table XIII also shows the result for the "NN2" binning. This is a variant of the NN binning which uses only two bins in $\mathcal{D}_{\text{NN}}$: both the first six bins and the last four bins are coalesced. The result is seen to be consistent with the 10-bin NN analysis.

For our final result, we use the nine-point quadratic fit. This choice is motivated by a desire to use a simple functional form; furthermore, it will be seen in the next section that among the polynomial fits considered, it gives the slope closest to unity when one plots extracted mass versus Monte Carlo input mass. The resulting mass is then 174.0 ± 5.6 GeV$/c^2$ for the LB binning, and 171.3 ± 6.0 GeV$/c^2$ for NN. These fit results are exhibited in Figs. 27–30.

Figure 28 summarizes our results. Note in this figure that $-\ln L$ tends to flatten out away from the minimum. Due to this, we limit the polynomial fit to the central region, where $-\ln L$ is most nearly quadratic. This flattening is related to the fact that we do not impose an external constraint on the number of signal or background events in the likelihood fit. If such a constraint is imposed, as was done in Ref. [2], the $-\ln L$ curve shows less tendency to flatten.

To use more likelihood points in the fit, a functional form which can model this flattening is needed. One such function which we investigated is

$$\begin{aligned} F(x) = -\ln(&P_1 + P_2 g(x-P_5, P_8) \\ &+ P_3 g(x-P_6, 2P_8) \\ &+ P_4 g(x-P_7, 4P_8)), \end{aligned} \tag{7.9}$$

where g is the Gaussian form $g(x,\sigma)=\exp(-(x/\sigma)^2/2)$. We determine the parameters P_1–P_8 by fitting this function (using MINUIT) to the likelihood points over the entire range of 110–230 GeV$/c^2$; the results are plotted in Fig. 28. If we extract from these curves the positions of the minima, the results are $173.6^{+5.6}_{-5.5}$ GeV$/c^2$ for LB and $172.4^{+4.1}_{-4.2}$ GeV$/c^2$ for NN (taking the uncertainty from where the curve rises by 0.5). From this, we conclude that the procedure of fitting a quadratic in the central region does not seriously underestimate the width. In addition, in Monte Carlo studies, $F(x)$ did not perform better on average than the simple quadratic fit; thus, we do not use $F(x)$ for the final mass extraction.

We have explored some additional variations in the definition of the likelihood function. The algorithm of HMCMLL [33] starts with the same likelihood as Eq. (7.6), but eliminates the nuisance parameters a_j^s and a_j^b using a maximum likelihood estimate rather than integration. To be able to compare likelihoods from different Monte Carlo samples, though, we modify the likelihood following the prescription of Ref. [10]:

$$L_\lambda = \frac{L}{\prod_j q(N_j,N_j)\,q(A_j^s,A_j^s)\,q(A_j^b,A_j^b)}. \tag{7.10}$$

The results of this procedure are given in Table XIV. Alternatively, we can eliminate n_s and n_b by integrating

TABLE XIII. Results of fits to the candidate sample, showing the top quark mass m_t and the number of signal and background events n_s and n_b. The labels "quad$N$" and "cub$N$" denote N-point quadratic and cubic fits, while "avg" denotes the mean value of the posterior mass probability distribution. "$-\ln L_{\min}$" is the minimum $-\ln L$ point; $\chi^2_{\rm poly}$ is for the polynomial fit to the likelihood points.

Binning	$-\ln L_{\min}$	Method	m_t (GeV/c^2)	n_s	n_b	$\chi^2_{\rm poly}$
LB	23.1	quad9	$174.0^{+5.6}_{-5.6}$	$23.8^{+8.3}_{-7.8}$	$53.2^{+10.7}_{-9.3}$	4.7
		quad11	$174.3^{+7.5}_{-7.5}$	$23.8^{+8.5}_{-9.1}$	$53.2^{+12.2}_{-9.4}$	29.7
		cub9	$173.7^{+5.7}_{-5.4}$	$23.8^{+8.3}_{-7.8}$	$53.2^{+10.7}_{-9.3}$	4.5
		cub11	$172.4^{+6.4}_{-5.4}$	$23.8^{+8.5}_{-7.8}$	$53.2^{+10.7}_{-9.4}$	14.7
		avg	$175.4^{+7.7}_{-7.7}$	$23.7^{+8.5}_{-9.2}$	$53.3^{+12.4}_{-9.5}$	
NN	74.5	quad9	$171.3^{+6.0}_{-6.0}$	$28.8^{+8.4}_{-9.1}$	$48.2^{+11.4}_{-8.7}$	8.4
		quad11	$170.8^{+6.1}_{-6.1}$	$29.1^{+8.2}_{-9.3}$	$47.9^{+11.6}_{-8.5}$	9.9
		cub9	$173.7^{+3.7}_{-4.7}$	$28.0^{+9.7}_{-8.4}$	$49.0^{+10.7}_{-10.0}$	3.9
		cub11	$172.5^{+4.8}_{-5.5}$	$28.3^{+9.0}_{-8.6}$	$48.7^{+10.9}_{-9.4}$	6.3
		avg	$170.7^{+6.7}_{-6.7}$	$29.2^{+8.4}_{-9.5}$	$47.8^{+11.8}_{-8.7}$	
NN2	29.8	quad9	$172.0^{+5.5}_{-5.5}$	$28.4^{+9.6}_{-9.0}$	$48.6^{+11.3}_{-9.9}$	5.7

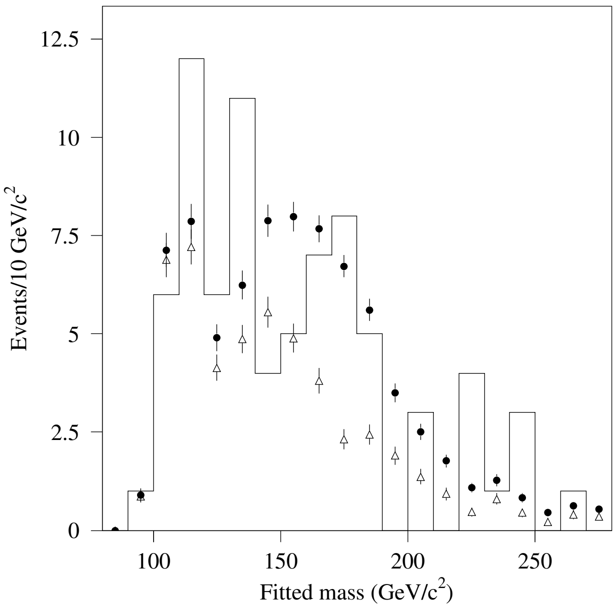

FIG. 27. Fitted mass for all events which pass the precuts and the χ^2 cut. Filled circles are a mixture of $t\bar{t}$ signal and background and open triangles are the background only, both averaged between the results of the LB and NN analyses.

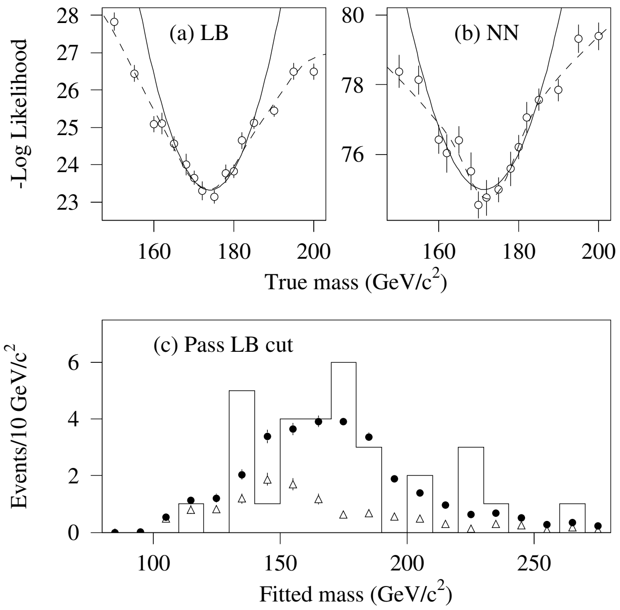

FIG. 28. Summary of the mass fit results, showing the negative log likelihood for (a) the LB and (b) the NN analyses. The solid curve is a quadratic fit to the 9 points around the minimum; the dashed curve is from fitting Eq. (7.9) to all points in the range 110–230 GeV/c^2. (c) Results of the LB fit for events passing the LB selection. The histogram is data, filled circles are a mixture of $m_t = 175$ GeV/c^2 $t\bar{t}$ signal and background, normalized using the results of the LB fit, and open triangles are background only.

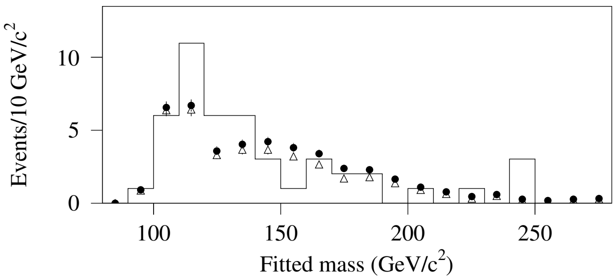

FIG. 29. Results of the LB fit for events failing the LB selection. The histogram is data, filled circles are a mixture of $m_t = 175$ GeV/c^2 $t\bar{t}$ signal and background, normalized using the results of the LB fit, and open triangles are background only.

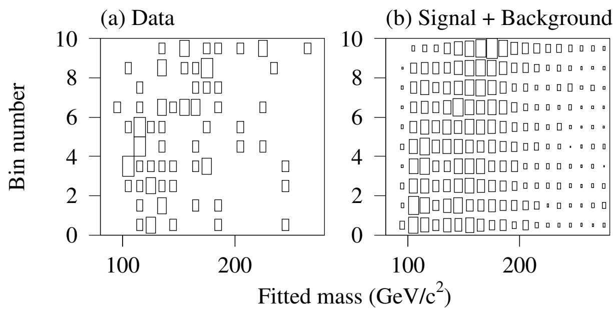

FIG. 30. Results of NN fit: (a) Data, (b) $m_t = 172$ GeV/c^2 $t\bar{t}$ signal plus background, normalized using the results of the NN fit. Box areas are proportional to the histogram contents.

TABLE XIV. Additional fit results.

Method	Binning	$-\ln L_{\min}$	m_t (GeV/c^2)	n_s	n_b
HMCMLL	LB	22.7	$174.1^{+5.6}_{-5.6}$	$23.6^{+7.9}_{-8.2}$	$53.4^{+10.3}_{-9.7}$
	NN	73.1	$172.0^{+5.1}_{-5.1}$	$34.0^{+1.9}_{-14.3}$	$42.6^{+16.0}_{-2.7}$
Integration	LB	17.2	$174.5^{+7.6}_{-7.6}$	$24.9^{+8.1}_{-9.7}$	$54.2^{+12.0}_{-9.9}$
	NN	68.5	$169.8^{+7.3}_{-7.3}$	$30.6^{+8.4}_{-10.2}$	$48.5^{+11.7}_{-9.4}$

over them, rather than by using a maximum likelihood estimate. The results of this are also given in Table XIV. These variations do not have a large effect on the final result.

To further test the stability of these results, we repeat the fits using samples in which one candidate event is removed, for a total of 77 distinct fits. For the LB case, the RMS of the resulting distribution of fits was 0.3 GeV/c^2; the smallest result seen was 173.0 GeV/c^2, and the largest was 174.7 GeV/c^2. For the NN case, the RMS was 0.5 GeV/c^2, the smallest result was 170.1 GeV/c^2, and the largest was 172.5 GeV/c^2.

To summarize the main results of this section, the LB analysis yields $m_t = 174.0 \pm 5.6$ GeV/c^2, and the NN analysis yields $m_t = 171.3 \pm 6.0$ GeV/c^2.

F. Tests with Monte Carlo samples

We test the mass extraction procedure by performing fits to ensembles of Monte Carlo experiments of known composition. The size of the experiments is fixed; the number of background events in each is chosen from a binomial distribution with a fixed mean.

For the first set of tests, the ensembles consist of 1000 experiments with a composition of $\langle n_s \rangle = 26$ and $\langle n_b \rangle = 52$, for an experiment size of $N = 78$ events with a 1:2 signal/background ratio. Results for the LB and NN analyses are shown in Tables XV and XVI. For these tests, the tabulated mean value is from a Gaussian fit to the extracted mass distribution, and the width is the symmetric interval around the mean which contains 68% of the entries. (We estimate the statistical uncertainties on these means and widths to be in the range 0.5–1.0 GeV/c^2.) Note that the 9-point quadratic fit gives the slope closest to unity. Some results for ensembles containing signal only are given in Tables XVII and XVIII.

There are several competing factors which contribute to the mass dependence of the width of the ensemble mass distributions $\sigma(m_t)$ observed in Tables XV and XVI. As m_t increases, the widths of the $m_{\rm fit}$ distributions slowly increase. From this one would expect the $\sigma(m_t)$ to increase with increasing top quark mass. However, we rely on the difference between the signal and background $m_{\rm fit}$ distributions to set the background normalization. This difference is smallest for m_t around 140–150 GeV/c^2; thus, one would expect $\sigma(m_t)$ to be larger in that re-

TABLE XV. Ensemble tests for the LB analysis with 1:2 signal/background, showing means and 68% widths. "Slope" is from a linear fit to the means.

Input Mass (GeV/c^2)	quad9 mean (GeV/c^2)	quad9 width	quad11 mean (GeV/c^2)	quad11 width	cub9 mean (GeV/c^2)	cub9 width	cub11 mean (GeV/c^2)	cub11 width
150	150.4	10.7	150.8	11.1	151.5	10.3	151.9	10.9
155	155.2	9.1	155.3	9.8	155.3	9.0	156.5	8.4
160	160.7	9.2	160.9	9.1	160.9	9.3	161.4	8.3
162	162.6	8.5	162.8	8.5	162.8	9.0	162.9	8.3
165	165.1	9.0	165.3	9.0	165.2	8.7	165.3	8.7
168	168.2	9.3	168.3	9.3	168.1	9.0	168.1	9.0
170	168.9	7.6	169.0	7.7	169.2	7.2	169.1	7.4
172	172.2	7.4	172.2	7.8	172.0	7.4	172.1	7.5
175	174.9	8.4	174.9	8.5	174.9	8.4	174.7	8.3
178	177.6	8.5	177.5	8.5	177.4	8.0	177.2	8.0
180	179.7	8.7	179.6	8.6	179.4	8.2	179.2	8.1
182	181.8	8.1	182.1	8.2	181.3	7.8	181.1	7.5
185	183.9	8.9	183.9	9.1	183.3	8.2	183.2	8.1
190	190.5	9.7	191.1	10.0	189.0	9.0	189.0	8.9
Slope	0.98		0.98		0.94		0.91	

TABLE XVI. Same as Table XV for the NN analysis.

Input Mass (GeV/c^2)	quad9 mean (GeV/c^2)	quad9 width	quad11 mean (GeV/c^2)	quad11 width	cub9 mean (GeV/c^2)	cub9 width	cub11 mean (GeV/c^2)	cub11 width
150	149.0	9.8	150.1	10.8	150.0	8.9	150.8	9.9
155	154.6	9.6	154.6	10.0	155.1	8.6	155.5	8.2
160	159.6	9.5	159.8	9.7	159.6	9.4	160.1	8.7
162	161.8	9.2	162.1	9.0	161.9	9.1	162.3	8.3
165	163.9	9.2	164.4	9.4	163.7	9.2	164.0	8.6
168	167.2	9.7	167.6	10.0	166.9	9.8	167.0	9.8
170	168.3	8.8	168.3	8.2	168.4	8.0	168.3	8.0
172	171.6	8.8	171.5	8.3	171.7	8.4	171.7	8.3
175	174.6	9.3	174.6	9.1	174.5	9.0	174.3	9.0
178	176.6	8.7	176.6	8.8	176.6	8.6	176.6	8.4
180	179.0	9.0	178.9	8.9	178.6	8.7	179.0	8.5
182	181.1	8.9	180.9	9.0	180.8	8.4	180.9	7.8
185	183.0	8.9	182.8	9.1	182.8	8.6	182.5	8.4
190	189.0	9.1	189.0	9.8	188.4	8.5	188.2	8.1
Slope	0.98		0.96		0.95		0.93	

TABLE XVII. Ensemble tests for the LB analysis with $n_s = 26$ events and $n_b = 0$.

Input Mass (GeV/c^2)	quad9 mean (GeV/c^2)	quad9 width	quad11 mean (GeV/c^2)	quad11 width	cub9 mean (GeV/c^2)	cub9 width	cub11 mean (GeV/c^2)	cub11 width
168	168.3	6.7	168.2	6.7	168.4	6.3	168.2	6.5
170	168.9	5.9	168.9	6.2	169.1	5.7	168.9	5.8
172	172.2	6.2	172.2	6.0	172.1	5.9	172.1	5.9
175	175.6	6.6	175.7	6.8	175.5	6.2	175.5	6.4

TABLE XVIII. Same as Table XVII for the NN analysis.

Input Mass (GeV/c^2)	quad9 mean (GeV/c^2)	quad9 width	quad11 mean (GeV/c^2)	quad11 width	cub9 mean (GeV/c^2)	cub9 width	cub11 mean (GeV/c^2)	cub11 width
168	167.7	6.3	168.1	6.8	168.0	5.8	167.9	6.4
170	168.9	6.1	169.0	6.0	169.0	5.6	168.8	5.7
172	172.0	6.1	172.3	6.2	172.0	5.5	172.0	5.9
175	175.6	6.5	175.6	6.7	175.2	6.0	175.3	6.4

TABLE XIX. Results of mass fits to ensembles of Monte Carlo events. The ensembles consisted of 10,000 experiments of 77 events each, with the compositions indicated below.

	Input Mass (GeV/c^2)	$\langle n_s \rangle$	$\langle n_b \rangle$	Mean (GeV/c^2)	Width (GeV/c^2)
LB	175	23.8	53.2	175.0	8.7
NN	172	28.8	48.2	171.6	8.0

gion. Finally, the spacing of the generated Monte Carlo points is finer in the region near 170 GeV/c^2; the available statistics are also larger there. This permits a more accurate determination of the top quark mass in that region, leading to a smaller $\sigma(m_t)$.

Next, we try ensembles with compositions that match the results of the likelihood fit. The results are given in Table XIX. (These and all subsequent results use the "quad9" prescription.) Plots of the mass distributions from these ensembles are shown in Fig. 31. Also shown are the distributions of the pull quantity

$$\text{pull} = \frac{m_t(\text{measured}) - m_t(\text{true})}{\sigma(m_t)}. \tag{7.11}$$

If the uncertainties produced by the mass extraction procedure are correct, these distributions should have unit width, as is indeed observed. In addition, 70% of the 1σ error intervals from the LB ensemble include 175 GeV/c^2, and 69% of those from the NN ensemble include 172 GeV/c^2, as expected.

The minimum $-\ln L$ value for the LB fit was 23.1; for the NN fit, it was 74.5. (A smaller value of $-\ln L$ corresponds to a better fit to the expected distributions.) This quantity is plotted for the LB and NN ensembles in Fig. 32. A $-\ln L$ value larger than that of the data is seen in about 7% of LB experiments and in about 28% of NN experiments.

One can also look at the distribution of statistical uncertainties from ensemble tests. For the data, the statistical uncertainty is 5.6 GeV/c^2 for the LB analysis, and 6.0 GeV/c^2 for the NN analysis. Plots of the statistical uncertainty for the ensemble fits are shown in Fig. 33. An uncertainty smaller than that for the data is seen in about 6% of LB experiments and in about 25% of NN experiments. The correlation between the mass and the uncertainty for the LB ensemble is exhibited in Fig. 34.

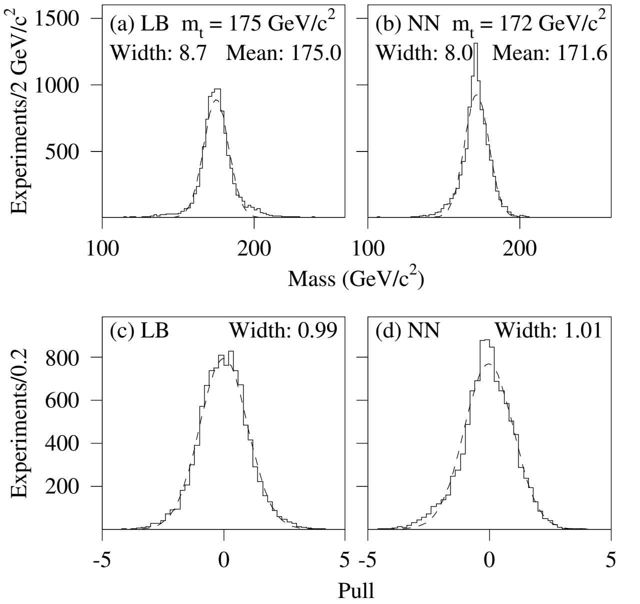

FIG. 31. Mass and pull distributions for 10,000 MC experiment ensembles with compositions matching the fit results. The dashed curves are Gaussian fits. For the mass distributions, the width is the symmetric interval containing 68% of the entries; for the pull distributions, it is from the Gaussian fit.

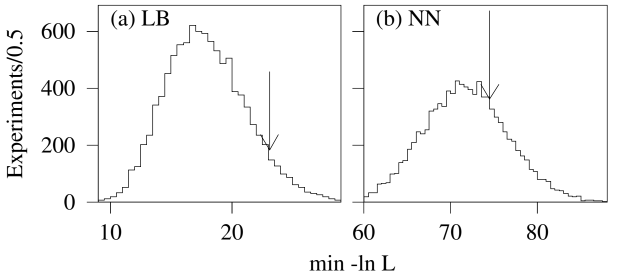

FIG. 32. Minimum $-\ln L$ distributions from the LB and NN ensembles. The arrows show the values corresponding to the data fits.

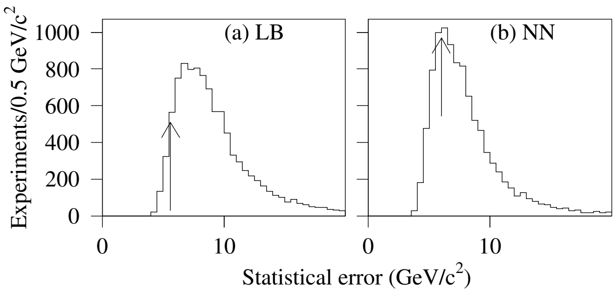

FIG. 33. Statistical uncertainty distributions from the LB and NN ensembles. The arrows show the values corresponding to the data fits.

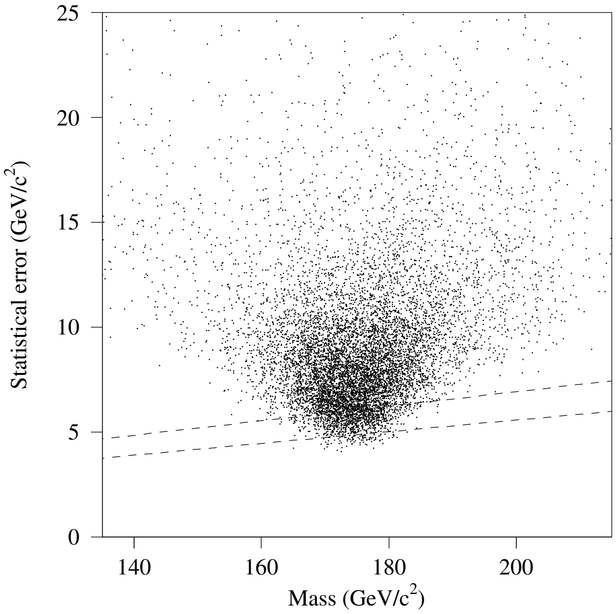

FIG. 34. Scatter plot of masses and statistical uncertainties from the LB ensemble. The dashed lines of constant relative uncertainty delimit the "accurate subset" (see text).

This shows that experiments with a small uncertainty typically yield masses closer to the true value.

It is interesting to examine the ensemble results for that subset of experiments where the extracted statistical uncertainty is similar to that actually obtained. We define this "accurate subset" as follows. First, find the relative uncertainty ($\sigma(m_t)/m_t$) for the result. For LB, this is 0.0322; for NN, it is 0.0350. Then convert these numbers to a percentile in the relative uncertainty distribution. These are 6.0% and 24.9% for LB and NN, respectively. For any ensemble, we then define the accurate subset by looking at its relative uncertainty distribution and selecting those experiments which lie within a range of $\pm 5\%$ around the above percentiles. This is illustrated in Figs. 34–35. This procedure thus selects 10% of the total sample. (The relative uncertainty is used because the statistical uncertainty tends to increase slightly with in-

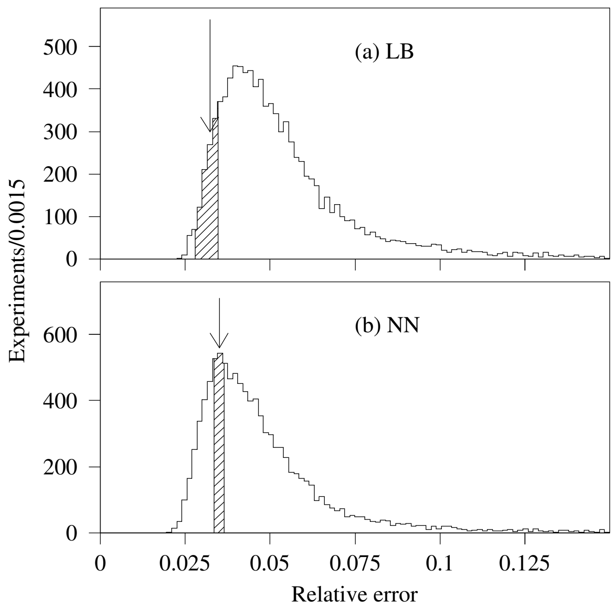

FIG. 35. Relative uncertainty ($\sigma(m_t)/m_t$) distributions from the LB and NN ensembles. The arrows show the value corresponding to the data fits, and the hatched regions show the definitions of the accurate subsets.

creasing mass; therefore, cutting on relative rather than absolute uncertainty results in a less biased subsample.)

There is an additional complication which arises when a cut is made on the statistical uncertainty. The spacing of the generated mass points is finer around $m_t = 175\ \mathrm{GeV}/c^2$. This permits a more accurate determination of the top quark mass in that range. However, this implies that if a small uncertainty is required, the masses of the selected events will be biased towards the region with finer spacing. (Note, however, that as long as a cut on the uncertainty is not made, the uneven MC spacing does not bias the mass. Studies of an even but coarser MC spacing show that adding extra points reduces the statistical uncertainty in the region where the extra points are added, but does not, on average, shift the extracted mass distribution.) Thus, for the accurate subset fits we changed the procedure slightly, adding Monte Carlo points at intervals of 2.5 GeV/c^2 between 130 and 160 GeV/c^2 and also between 185 and 210 GeV/c^2. These additional mass points were constructed by interpolating between the existing MC histograms on either side. The results of these fits with the accurate subset cuts are shown in Fig. 36. The widths are 4.6 GeV/c^2 and 6.0 GeV/c^2 for LB and NN, respectively. This is a further indication that the uncertainty estimates from the likelihood fit are reliable.

The results of the LB and NN analyses can be compared experiment-by-experiment, provided that the ensemble definitions are the same. We use the same ensemble definition as for the first set of tests ($N = 78$ events and a 1:2 signal/background ratio) with $m_t = 175\ \mathrm{GeV}/c^2$. The results for 10,000 experiments are given

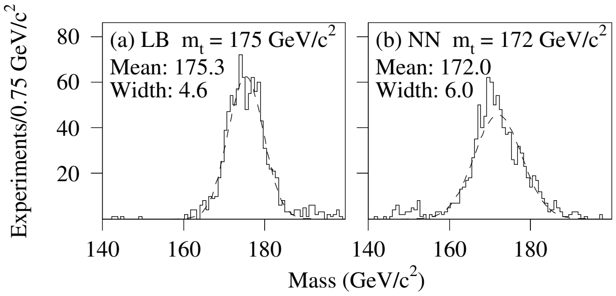

FIG. 36. Mass distributions for accurate subsets of ensembles. The dashed curves are Gaussian fits.

TABLE XX. Comparisons of LB and NN ensembles for $m_t = 175$ GeV$/c^2$ and a 1:2 signal/background ratio. The first line is the mean difference between the results; the second and third lines give the fraction of experiments for which the difference exceeds the observed difference of 2.7 GeV$/c^2$. (Numbers are in GeV$/c^2$.)

	Full ensemble	LB acc. subset	NN acc. subset
$\langle \text{LB} - \text{NN} \rangle$	0.78 ± 0.05	0.34 ± 0.06	0.51 ± 0.09
$(\text{LB} - \text{NN}) > 2.7$	29%	11%	18%
$\lvert\text{LB} - \text{NN}\rvert > 2.7$	45%	16%	28%

in Table XX. It is seen that given the observed statistical uncertainties, a difference between the two analyses of the magnitude seen is expected $\sim 20\%$ of the time.

It is also interesting to look at the correlation between the LB and NN measurements. This can be defined using the ensemble mass distributions of m_{LB} and m_{NN} as

$$\rho = \frac{\langle (m_{\text{LB}} - \langle m_{\text{LB}} \rangle)(m_{\text{NN}} - \langle m_{\text{NN}} \rangle) \rangle}{\sigma_{\text{LB}} \sigma_{\text{NN}}}. \quad (7.12)$$

This is appropriate for Gaussian distributions; however, our distributions typically have a small number of non-Gaussian outliers. To explore the sensitivity of this quantity to these outliers, the following procedure is used.

- For the cuts of interest, plot m_{LB} and m_{NN}. Record the means and RMS widths of these distributions ($\langle m_{\text{LB}} \rangle$, σ_{LB}, $\langle m_{\text{NN}} \rangle$, σ_{NN}).
- Reject experiments which are more than $K\sigma$ from the mean. Specifically, make the additional cut that

$$|m_{\text{LB}} - \langle m_{\text{LB}} \rangle| < K\sigma_{\text{LB}} \quad \text{and} \quad |m_{\text{NN}} - \langle m_{\text{NN}} \rangle| < K\sigma_{\text{NN}}. \quad (7.13)$$

- Replot m_{LB} and m_{NN} with this additional cut, and record the new means and RMS widths ($\langle m_{\text{LB}} \rangle'$, σ'_{LB}, $\langle m_{\text{NN}} \rangle'$, σ'_{NN}).
- Plot (with all cuts) the distribution of

$$\left(m_{\text{LB}} - \langle m_{\text{LB}} \rangle'\right) \cdot \left(m_{\text{NN}} - \langle m_{\text{NN}} \rangle'\right). \quad (7.14)$$

TABLE XXI. Values of correlation parameter ρ.

K	Full Sample	LB acc. subset	NN acc. subset
100	0.62	0.89	0.77
5	0.65	0.89	0.88
4	0.67	0.89	0.89
3	0.70	0.89	0.89
2	0.77	0.87	0.88
1	0.75	0.67	0.78

- Find the mean of this distribution. ρ is then calculated by dividing this mean by $\sigma'_{\text{LB}} \sigma'_{\text{NN}}$.

The results are tabulated for the full sample and for the LB and NN accurate subsets in Table XXI. This is done using the same $m_t = 175$ GeV$/c^2$ ensembles as for the previous comparisons. They do not depend strongly on K within reasonable ranges. To get a single number, we average the $K = 5$ results for the two accurate subset results, giving 0.88. This appears to be a reasonable representation of the accurate subset numbers (within a few percent) for $K \geq 2$. Propagating statistical uncertainties through this calculation gives $\rho = 0.88 \pm 0.04$.

In summary, these ensemble tests show that the masses and uncertainties obtained from the likelihood fit are reliable, and that our observed data set is not particularly unlikely.

G. Systematic uncertainties

1. Energy scale uncertainties

The first major component of the systematic uncertainty is that from the jet energy scale. What is relevant here is the uncertainty in the relative scale between the data and MC, rather than in the absolute scale. This was estimated to be $\pm(2.5\% + 0.5$ GeV) for each jet (see Sec. IV).

We propagate this per-jet uncertainty to the final mass measurement by performing ensemble tests with all the jets in the events comprising the ensemble scaled up or down by the per-jet uncertainty. For these tests, we used large experiment sizes, with $N = 1000$. The results are given in Table XXII and give an uncertainty of about ± 4 GeV$/c^2$. Comparing this with the shifts in the m_{fit} distributions seen after scaling the jets (Fig. 20), we estimate the ratio between a shift in the final extracted mass and a shift in m_{fit} to be about 1.1.

The systematic uncertainty in the electromagnetic energy scale is much smaller than that of the jets, and can be neglected. The systematic uncertainty of the muon momentum measurement is estimated to be 2.5%. The effect of this uncertainty is found to be negligible relative to the jet scale uncertainty.

TABLE XXII. Ensemble means for determining the uncertainty due to jet energy scale. Each experiment consisted of $N = 1000$ events; the signal/background ratios are the same as in Table XIX.

	LB	NN
Input mass	175.0 GeV/c^2	172.0 GeV/c^2
Input $\langle n_s \rangle$	309.1 events	374.0 events
$-2.5\% - 0.5$	170.9 GeV/c^2	167.6 GeV/c^2
Nominal	175.4 GeV/c^2	171.3 GeV/c^2
$+2.5\% + 0.5$	179.4 GeV/c^2	175.2 GeV/c^2
Symmetric Uncertainty	4.2 GeV/c^2	3.8 GeV/c^2

TABLE XXIII. Ensemble means for determining the difference between ISAJET and HERWIG. (All numbers in GeV/c^2.) Each ensemble consisted of $N = 1000$ event experiments with a 1:2 signal/background ratio.

	LB			NN		
m_t	HERWIG	ISAJET	Diff	HERWIG	ISAJET	Diff
150	150.5	151.7	-1.2	149.4	150.4	-1.0
160	161.0	160.9	0.1	159.8	159.4	0.4
170	169.3	170.8	-1.5	168.3	169.0	-0.7
180	180.1	180.1	0.0	179.6	178.9	0.7
190	190.2	190.1	0.1	189.0	188.8	0.2
200	201.9	200.9	1.0	200.5	197.6	2.9

2. Generator dependencies

The next component of the systematic uncertainty is that due to uncertainties in how well the underlying Monte Carlo event generators model reality. We separate this into signal and background components. Of particular concern is the modeling of QCD radiation by the $t\bar{t}$ signal Monte Carlo.

To estimate the uncertainty due to the HERWIG generator, we characterize HERWIG events using variables which are sensitive to the amount of initial and final state radiation (ISR and FSR) in each event. To do this, we match the direction of reconstructed jets with HERWIG partons and use the Monte Carlo parentage information to identify the jets which come from the b quarks and the hadronically-decaying W boson. We consider the four jets with highest E_T $j_1, \ldots j_4$, and define the variables:

- $x \equiv$ Number of jets in $j_1, \ldots j_4$ which do not come from a b quark or the W boson (i.e., jets which are likely to be due to ISR).
- $y \equiv N_j - 4 \equiv$ Number of extra jets of any kind in the event ($N_j \equiv$ number of jets with $E_T > 15$ GeV and $|\eta| < 2.0$).
- $z \equiv$ Number of non-ISR jets in $j_1, \ldots j_4$ which have the same parent as a higher E_T jet (i.e., the number of extra jets due to FSR among the top four).

We take a HERWIG Monte Carlo sample (with m_t = 170 GeV/c^2) and bin it using these variables into a three-dimensional histogram with ranges $0 \le x, y, z \le 2$ (27 bins). For each bin (x, y, z), we plot the fitted masses for all events in that bin, fit them to a Gaussian to form $\langle m_{\rm fit} \rangle (x, y, z)$, and then fit the resulting values to the empirical function

$$G(x, y, z) = m_0 + ux + v \max(0, y - x - z) + wz, \tag{7.15}$$

for fit parameters m_0, u, v, and w. Here, u describes the dependence of $\langle m_{\rm fit} \rangle$ on ISR and v and w describe its dependence on FSR. In particular, the v term describes the dependence of the mass on the number of extra jets which cannot be attributed to either an ISR or FSR jet displacing another jet out of the top four. Additional low E_T jets affect the mass only if they are FSR; thus we group v with w. We compute a population-weighted average of G over all bins; this is seen to agree well with $\langle m_{\rm fit} \rangle$ from the entire sample. Finally, we recalculate this average with (a) u (ISR) increased by 50% and (b) v and w (FSR) increased together by 50%. This gives excursions of 0.69 and 1.74 GeV/c^2, respectively. Adding these in quadrature yields an uncertainty of 1.9 GeV/c^2. (Monte Carlo studies of ensembles constructed of events from individual (x, y, z) bins confirm that, for these variations, the mass resulting from the likelihood fit approximately tracks $\langle m_{\rm fit} \rangle$.)

We have performed several additional cross checks to verify that this is a reasonable estimate of the signal generator uncertainty. The first is simply to compare these results to those from a different event generator, in this case ISAJET. We constructed ensembles from ISAJET events and analyzed them using the MC histograms derived from HERWIG. These are compared to ensembles of HERWIG events in Table XXIII. Taking the six differences in the region 160–180 GeV/c^2 gives a mean of -0.17 GeV/c^2 and a RMS of 0.8 GeV/c^2.

We also vary the QCD coupling strength parameter, $\Lambda_{\rm QCD}$, of the HERWIG $t\bar{t}$ Monte Carlo. The default value of this parameter in HERWIG 5.7 is 0.18 GeV; the current experimental value from the Particle Data Group is $0.21^{+0.04}_{-0.03}$ GeV [3]. Accordingly, we generate additional $t\bar{t}$ Monte Carlo with $\Lambda_{\rm QCD}$ set to 0.15, 0.21, and 0.25 GeV, with m_t = 170 and 175 GeV/c^2 [34]. We then construct ensembles from these samples and process them using the standard analysis. The results are given in Table XXIV. The size of the resulting deviations is on the order of 1 GeV/c^2; they appear to be dominated by Monte Carlo statistics.

We can make another comparison by using a version of HERWIG 5.8 in which final state radiation (FSR) in top quark decays is substantially suppressed. We compare results from ensembles made from this version to those from HERWIG 5.8 with normal radiation. The results are shown in Table XXV. Averaging over LB and NN, this is seen to give an excursion of about 2.15 GeV/c^2. Note

TABLE XXIV. Ensemble tests with $\Lambda_{\rm QCD}$ varied. Ensembles consisted of experiments with $N = 1000$ events and a 1:2 signal/background ratio.

$\Lambda_{\rm QCD}$	$\langle m_{\rm fit}\rangle$ (GeV/c^2)		LB (GeV/c^2)		NN (GeV/c^2)	
(GeV)	$m_t = 170$ GeV/c^2	$m_t = 175$ GeV/c^2	$m_t = 170$ GeV/c^2	$m_t = 175$ GeV/c^2	$m_t = 170$ GeV/c^2	$m_t = 175$ GeV/c^2
0.15	171.0	173.5	170.5	175.2	169.5	174.8
0.18	168.8	173.1	169.2	175.3	168.3	174.5
0.21	170.8	173.6	170.2	174.5	169.5	173.3
0.25	168.7	173.2	168.3	175.7	167.2	175.0

TABLE XXV. Comparison of ensembles constructed using HERWIG 5.8 both with and without FSR suppressed. The ensembles consist of $N = 77$ event experiments. For the LB case, $\langle n_s\rangle = 23.8$, and for NN, $\langle n_s\rangle = 28.8$. For both cases, $m_t = 170$ GeV/c^2.

	$\langle m_{\rm fit}\rangle$ (GeV/c^2)	LB (GeV/c^2)	NN (GeV/c^2)
FSR suppressed	176.0	172.2	172.7
Normal FSR	170.1	170.7	169.9
Difference	5.9	1.5	2.8

that the $m_{\rm fit}$ distribution with FSR suppressed is significantly narrower on the low mass side than distributions with normal radiation. This difference in shape is why the relation between means of $m_{\rm fit}$ and ensemble results is different here than described above.

The results of these cross checks confirm that our estimate for the systematic uncertainty due to the signal generator of 1.9 GeV/c^2 is reasonable.

We have looked at the effects of varying the parton distribution functions, and found them to be negligible. This is to be expected, since the dominant top quark production process at the Tevatron is $q\overline{q} \rightarrow t\overline{t}$ and the proton structure functions for valence quarks are well known in the parton momentum fraction region which is important for top quark production. Using the MRSA' [35] structure functions instead of CTEQ3M produced a change of fitted masses of < 0.1 GeV/c^2.

We also study the effects of varying the VECBOS background model. Besides the sample used for the mass measurement (which uses a Q^2 scale of $\langle p_T^{\rm jet}\rangle^2$ and HERWIG fragmentation), we have samples with a Q^2 scale of M_W^2 and with ISAJET fragmentation. Results from ensembles made from these samples are shown in Table XXVI. (The ensemble compositions were the same as for the jet energy scale tests.) The largest difference seen is about 2.5 GeV/c^2 using the M_W^2 scale with HERWIG fragmentation.

A concern is that the systematic uncertainty assigned here to VECBOS may not adequately reflect the level of agreement between VECBOS and data for η^W in the forward region (Fig. 3). To check this, we reweight the VECBOS events using a smooth function of η^W (a Gaussian) chosen to optimize the agreement between the simulation and the data. When we redo the mass extraction with this reweighted background, the top quark mass shifts by only 0.4–0.5 GeV/c^2, a value much smaller than the uncertainty we attribute to VECBOS. This uncertainty can therefore be neglected.

TABLE XXVI. Ensemble means for determining VECBOS differences. Samples were generated with a VECBOS Q^2 scale of both M_W^2 and $\langle p_T^{\rm jet}\rangle^2$, and using both HERWIG (HW) and ISAJET (IS) for fragmentation. Each experiment consisted of $N = 1000$ events; the signal/background ratios are the same as in Table XIX.

	LB	NN
Input mass	175.0 GeV/c^2	172.0 GeV/c^2
Input $\langle n_s\rangle$	309.1 events	374.0 events
$\langle p_T^{\rm jet}\rangle^2$, HW	175.4 GeV/c^2	171.3 GeV/c^2
M_W^2, HW	177.9 GeV/c^2	173.8 GeV/c^2
$\langle p_T^{\rm jet}\rangle^2$, IS	175.0 GeV/c^2	171.2 GeV/c^2
M_W^2, IS	175.8 GeV/c^2	171.6 GeV/c^2
Max. difference	2.5 GeV/c^2	2.5 GeV/c^2

We also do the fits with the fraction of QCD multijets contributing to the background histogram [$(22 \pm 5)\%$] varied within its uncertainty. The changes to the final extracted mass are < 0.2 GeV/c^2, well below the assigned uncertainty.

3. Noise and multiple interactions

At the luminosities at which most of our data were collected, it is likely that during a single beam crossing, there will be multiple $p\overline{p}$ inelastic interactions (MI). (This is expected about 2/3 of the time.) While these extra interactions rarely give rise to additional high-p_T objects, they do deposit a small amount of additional energy over the entire calorimeter, affecting the jet energy calibration. Additional noise in the calorimeter is produced by the radioactive decay of the uranium absorber. The Monte Carlo samples used for this analysis do not include these effects. To estimate them, we generate a small number of additional Monte Carlo events which include noise, and which are overlaid with one or two additional interactions. The means of the $m_{\rm fit}$ distribution for these samples are given in Table XXVII. Based on the luminosity profile of the collected data, we estimate that in order to represent the data, these samples should be combined in the ratio 0.31 : 0.33 : 0.36. The weighted average of the three means is then 170.5 ± 0.6 GeV/c^2; the shift from the zero additional interaction case is 1.2 ± 0.7 GeV/c^2.

TABLE XXVII. Means of $m_{\rm fit}$ distributions of $t\bar{t}$ Monte Carlo for multiple interaction uncertainty determination. (For the e + jets channel, $m_t = 170$ GeV/c^2.)

	$\langle m_{\rm fit}\rangle$ (GeV/c^2)	Weight
0 additional interactions	169.3 ± 0.4	0.31
1 additional interaction	170.5 ± 1.3	0.33
2 additional interactions	171.6 ± 1.2	0.36

TABLE XXVIII. Uncertainties due to Monte Carlo statistics.

	LB (GeV/c^2)	NN (GeV/c^2)
Signal	0.49	0.99
Background	0.33	0.57
Total	0.6	1.1

Scaling this by the factor 1.1 for the ratio between a shift in final extracted mass and a shift in $m_{\rm fit}$ (Sec. VII G 1) gives an estimated shift due to noise and multiple interactions of 1.3 ± 0.8 GeV/c^2. Since this effect is relatively poorly known and is small compared to other uncertainty sources, we do not attempt to correct the result for this effect, but instead include it as a systematic uncertainty.

4. Monte Carlo statistics

We assess the effect of Monte Carlo statistics on the final result by performing the fit to the data many times, each time smearing the MC histograms used to calculate the likelihood according to Poisson statistics. This is done separately for signal and background. The 68% widths of the resulting mass distributions are given in Table XXVIII.

5. Systematic uncertainty summary

Table XXIX gives a summary of the systematic uncertainties. In addition to the uncertainties already discussed, the mean difference of 0.8 GeV/c^2 between the LB and NN ensemble results from Table XX has been added as a systematic uncertainty, and an additional uncertainty of 1 GeV/c^2 has been added to cover possible small biases in the likelihood fitting method (this is approximately the RMS spread of the different polynomial fits in Table XIII). Note that these two components are of the same order as the estimated uncertainty due to Monte Carlo statistics, and that these small biases are probably due in large part to statistical fluctuations in the Monte Carlo histograms. Nevertheless, we retain these as separate components of the systematic uncertainty in lieu of exploring this further with still larger Monte Carlo samples.

TABLE XXIX. Systematic uncertainty summary.

	LB (GeV/c^2)	NN (GeV/c^2)	Average (GeV/c^2)
Jet energy scale	4.2	3.8	4.0
Generator			
$t\bar{t}$ signal	1.9	1.9	1.9
VECBOS flavors	2.5	2.5	2.5
Noise/MI	1.3	1.3	1.3
Monte Carlo stat.	0.6	1.1	0.85
LB/NN diff	0.8	0.8	0.8
Likelihood fit	1.0	1.0	1.0
Total	5.6	5.4	5.5

The total systematic uncertainties here are slightly smaller than those reported in Ref. [5]. The signal generator uncertainty was 3.3 GeV/c^2, taken from the difference between HERWIG and an older version of ISAJET, and the LB/NN difference was 1.35 GeV/c^2, taken from half the difference of the fit results.

H. Summary

For the final mass result, we combine the results of these two analyses, taking into account their correlation ρ determined earlier. Let $m_{\rm LB}$ and $m_{\rm NN}$ be the two results and $\sigma_{\rm LB}$ and $\sigma_{\rm NN}$ be their uncertainties. Then we form a χ^2 as a function of the combined mass M:

$$\begin{aligned}\chi^2(M) = {} & \frac{1}{\sigma_{\rm LB}^2\sigma_{\rm NN}^2(1-\rho^2)} \\ & \times[\sigma_{\rm NN}^2(M-m_{\rm LB})^2 \\ & -2\rho\sigma_{\rm LB}\sigma_{\rm NN}(M-m_{\rm LB})(M-m_{\rm NN}) \\ & +\sigma_{\rm LB}^2(M-m_{\rm NN})^2].\end{aligned} \tag{7.16}$$

The combined result and its uncertainty is then defined by the minimum of this curve and the points where the curve rises by one unit from the minimum. (Monte Carlo studies of this combination give a width of the pull distribution of 1.11 for the full sample, but 0.76 for the LB accurate subset and 0.97 for the NN accurate subset.) Inserting $m_{\rm LB} = 174.0$ GeV/c^2, $\sigma_{\rm LB} = 5.6$ GeV/c^2, $m_{\rm NN} = 171.3$ GeV/c^2, $\sigma_{\rm NN} = 6.0$ GeV/c^2, and $\rho = 0.88$ (for the accurate subsets) gives

$$M = 173.3 \pm 5.6 \text{ GeV}/c^2. \tag{7.17}$$

The systematic uncertainties of the two methods are averaged, giving a final result of

$$m_t = 173.3 \pm 5.6 \text{ (stat)} \pm 5.5 \text{ (syst) GeV}/c^2. \tag{7.18}$$

VIII. PSEUDOLIKELIHOOD ANALYSIS

A. Introduction

The pseudolikelihood (PL) analysis is an alternate method of extracting the top quark mass, with several important differences from the analyses of the previous section. It thus serves as a nearly independent check of the previous result. In this analysis, we kinematically fit candidate events at a series of fixed top quark masses $m_{\rm fit}$ (3C fits) over the range 100–250 GeV$/c^2$. These fits are done using a different kinematic fitting program (SQUAW [28]) than was used in the previous section. In addition, when looping over jet permutations, we allow the assignment of jets beyond the fourth (in which case at least one of the top four jets is treated as ISR). At each $m_{\rm fit}$, we choose the jet permutation yielding the smallest χ^2, and interpret the resulting plot of $\chi^2/2$ versus $m_{\rm fit}$ as defining a top quark mass "pseudolikelihood" L for a particular event given by

$$L(m_{\rm fit}) = e^{-\chi^2(m_{\rm fit})/2}. \qquad (8.1)$$

We then sum this plot over all candidate events, subtract the expected background contribution, and fit the remainder to a quadratic function to extract the top quark mass. This analysis is performed mainly for signal-enriched subsamples of the entire precut sample (i.e. subsamples passing either the LB or NN selection).

A major motivation for this analysis method is to more fully take into account the information from different jet permutations. For example, the fixed-mass χ^2 plot for one top quark candidate is shown in Fig. 37. The information about both minima in this figure is incorporated directly into the PL analysis, but is not used in the LB and NN likelihood analyses.

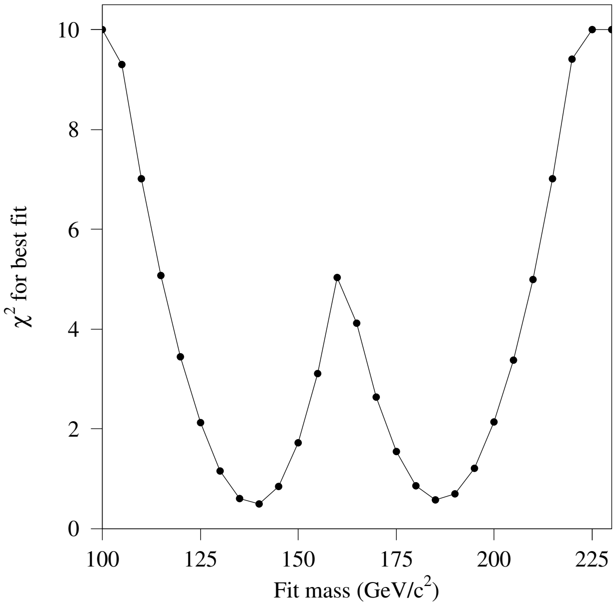

FIG. 37. χ^2 plot for SQUAW fixed-mass fits for event 58203, 4980.

B. PL method

Some examples of $\chi^2/2$ plots for $t\bar{t}$ events are shown in Fig. 38. These are "average $\chi^2/2$" plots: for each $m_{\rm fit}$, we average the $\chi^2/2$ over all events in the sample. The figure shows plots for events generated with both HERWIG and ISAJET for top quark masses from 160 to 190 GeV$/c^2$. The plots from ISAJET are slightly wider than those from HERWIG. We will also need the background shape to subtract the expected background contribution from the data sample. It is determined by combining the average $\chi^2/2$ plot of the VECBOS $W+$jets sample with that of the QCD multijet sample. These plots are shown in Fig. 39. They are broader and have minima at about 150 GeV$/c^2$, lower than those for $t\bar{t}$ events (for $m_t > 160$ GeV$/c^2$). The VECBOS sample uses the average jet transverse momentum Q^2 scale and HERWIG for fragmentation, as in the variable-mass analyses.

The next step is to determine the background normalization. The nominal background fraction in the precut

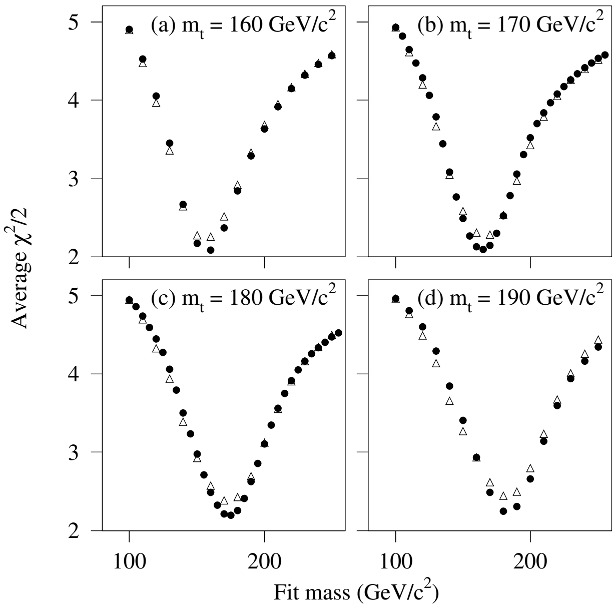

FIG. 38. Average $\chi^2/2$ plots (after LB selection) for HERWIG (filled circles) and ISAJET (open triangles) $t\bar{t}$ events.

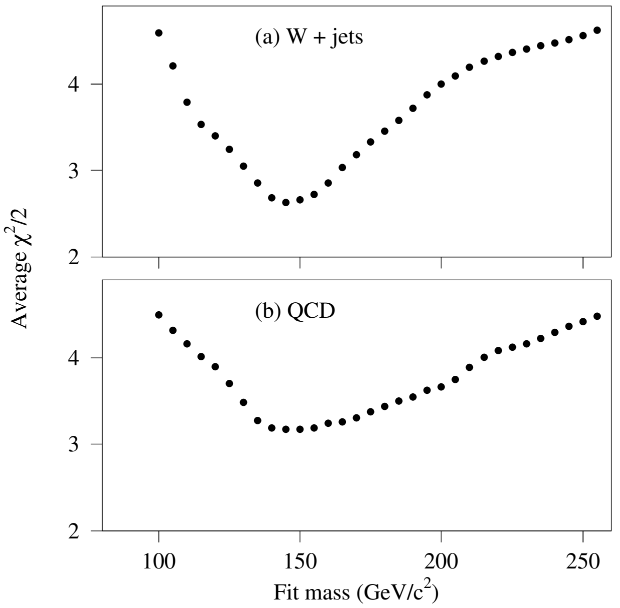

FIG. 39. Average $\chi^2/2$ plots (after LB selection) for (a) VECBOS $W+$jets and (b) QCD multijet background samples.

event sample is found from the cross section analysis to be $\approx 2/3$. One can improve on this nominal background by using properties of the particular sample being analyzed which are sensitive to the background fraction. One such property is the average value of one of the top quark discriminants (either $\mathcal{D}_{\rm LB}$ or $\mathcal{D}_{\rm NN}$). The background fraction can be calculated as

$$\text{BG fraction} = (\mathcal{D}^T - \mathcal{D}^D)/(\mathcal{D}^T - \mathcal{D}^B), \qquad (8.2)$$

where $\mathcal{D}^T$ is the average value expected for $t\bar{t}$ events, $\mathcal{D}^B$ is that expected for background events, and $\mathcal{D}^D$ is that of the sample being analyzed.

We can do an analogous calculation using the $\chi^2/2$ plot. There is, however, a complication, due to the fact that the $\chi^2/2$ plots depend on the top quark mass to a much greater extent than do the likelihood discriminants. Therefore, to get a background from this method, we need a rough estimate of the top quark mass. We find this as follows. For each sample, we construct the average $\chi^2/2$ plot. We compare the plot from data to that predicted from MC signal plus background, with the MC top quark mass varied in 10 steps from 140 to 210 GeV/c^2. We pick the mass which yields the smallest RMS difference with the data.

An additional complication is that, in general, the average $\chi^2/2$ plots for signal and background will cross at some $m_{\rm fit}$. We thus define the variable

$$C = \sum_{m_{\rm fit} > m_{\rm cross}} \chi^2(m_{\rm fit})/2 - \sum_{m_{\rm fit} < m_{\rm cross}} \chi^2(m_{\rm fit})/2, \quad (8.3)$$

where $m_{\rm cross}$ is the point at which the plots cross. ($m_{\rm cross}$ is near 150 GeV/c^2 for top quark masses above 160 GeV/c^2.) We then estimate the background in the same manner as before, using

$$\text{BG fraction} = (C^T - C^D)/(C^T - C^B), \qquad (8.4)$$

where C^T, C^B, and C^D are the values of C from MC signal, background, and the data sample, respectively.

The background fraction for the full precut sample is taken to be the average of three values: the nominal value, the value determined from the top quark discriminants, and the value from the $\chi^2/2$ plot. They are weighted by the squared inverses of their uncertainties.

When analyzing subsets of the precut sample which pass either the LB or the NN selection, we determine the nominal background for the subset by scaling down the background determined from the full precut sample. The subset background fraction is then the weighted average of this nominal background fraction and the fraction estimated from the $\chi^2/2$ plots. The background estimate from the top quark discriminants is not used in this case, as the subset selections tend to make the distributions of these discriminants similar for signal and background. The precut and LB subset background fractions determined from the data are 0.60 and 0.32, respectively.

For each $m_{\rm fit}$, we subtract the $\chi^2/2$ contribution expected for the background from the total. This is evaluated over the range 100–250 GeV/c^2 with a distance between points $\Delta m_{\rm fit} = 10$ GeV/c^2. We then extract the top quark mass and uncertainty using a quadratic fit near the minimum of this background-subtracted $\chi^2/2$ plot. The extracted mass $m_{\rm min}$ is the value at which the fit function has its minimum, and its uncertainty is the deviation that corresponds to an increase of 0.5 units above the minimum. We try to use as many points as possible in the fit provided that the plot remains parabolic over the fit range. The algorithm used to select the fit range is determined empirically by fitting the average $\chi^2/2$ plots for $t\bar{t}$ Monte Carlo events. With $\Delta m_{\rm fit} = 10$ GeV/c^2, at least three points below and two points above the minimum are required; thus, the mass range covered is at least 50 GeV/c^2. If necessary, we add points at the extremes until the value of $\chi^2/2$ exceeds that at the minimum by an amount equal to the number of events in the plot. However, we add points on the high side only if the $\chi^2/2$ values change at an increasing rate, as expected for a parabola. We also do some fits with $\Delta m_{\rm fit} = 5$ GeV/c^2 over the range 100–255 GeV/c^2. In that case, we use at least five points on each side of the minimum.

C. Results of fits to Monte Carlo events

Table XXX contains results of fits to average $\chi^2/2$ plots from MC samples. The mass $m_{\rm min}$ (from a quadratic fit near the minimum) for $t\bar{t}$ Monte Carlo is slightly different from the MC input mass. It has a roughly linear dependence on the input top quark mass, with a slope that is only slightly smaller than that determined from fits with the correct jet assignment. A linear fit to these points gives the following prescription for a "corrected" mass $m_{\rm corr}$:

$$m_{\rm corr} = (m_{\rm min} - 27.0\ {\rm GeV}/c^2)/0.815. \tag{8.5}$$

This relation is used to correct the masses $m_{\rm min}$ obtained from fits.

D. Ensemble studies

We study the performance of the PL method by forming ensembles of simulated experiments consisting of MC events which pass the precuts. These experiments contain $N = 78$ events each, with an average of 26 events from signal and the balance from background. The results are shown in Table XXXI. (All use $\Delta m_{\rm fit} = 10\ {\rm GeV}/c^2$.) The typical uncertainties on the average ensemble masses are about 0.5 GeV/c^2, so the LB and NN subset masses are consistent. We also show in Table XXXII results for ensembles of experiments consisting of 26 signal events and no background. The agreement of the corresponding average mass values between Tables XXXI and XXXII indicates that the background subtraction does not produce a mass bias.

The widths of the $m_{\rm corr}$ distributions for the subset analyses are smaller than those from the entire sample; further, the widths for LB subsets are all smaller than those for the corresponding NN subsets. The widths for the LB subset are smaller because the background for the LB subset is smaller than for the NN subset: at $m_t = 175$ GeV/c^2, the background fraction for LB is 35%, and for NN, it is 42%. Results will therefore be based primarily on LB subset fits. The widths and shifts from the input mass are plotted in Figs. 40 and 41 for the LB subset.

Figure 42 shows the pull distribution (as defined in Eq. (7.11)) for LB subset fits. We find the uncertainty on $m_{\rm corr}$ by dividing the width of the quadratic fit by the slope of the mass correction. A Gaussian fit to the pull distribution for $m_t = 175$ GeV/c^2 has a width of 1.51. Therefore, the corrected uncertainties from quadratic fits typically underestimate the width of the ensemble mass distribution and need to be scaled up by an additional factor of 1.51. (Note that due to the effects of background and incorrect jet assignments, this width is not necessarily unity. But for $t\bar{t}$ events in which the best fit corresponds to the correct jet assignment, the width of the pull distribution is consistent with unity.)

E. Analysis of data sample

We analyze the data for the two subsets defined by the LB and NN selections (see Sec. VI). These subset selections are about 80% and 65% efficient for the $t\bar{t}$ signal, respectively, versus about 30% for background.

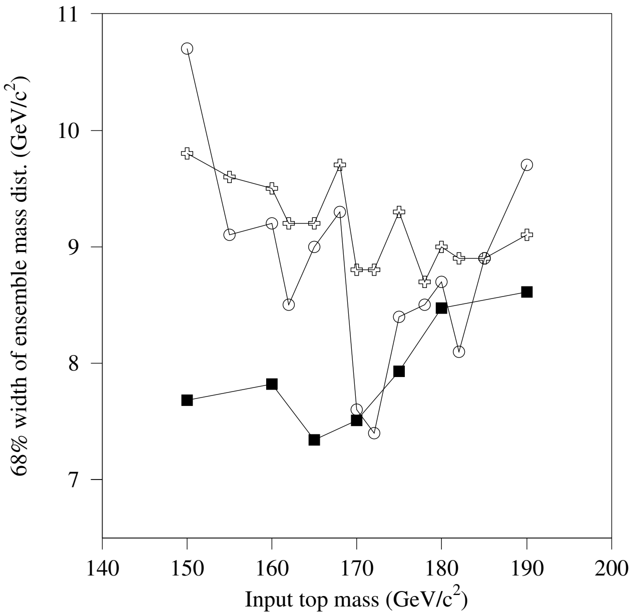

FIG. 40. 68% widths of ensemble mass distributions for different analyses. Squares are for PL fits to the LB subset, circles are for LB variable-mass fits, and plus symbols are for the NN variable-mass fits. Typical uncertainties on the plotted values are between 0.5 and 1.0 GeV/c^2.

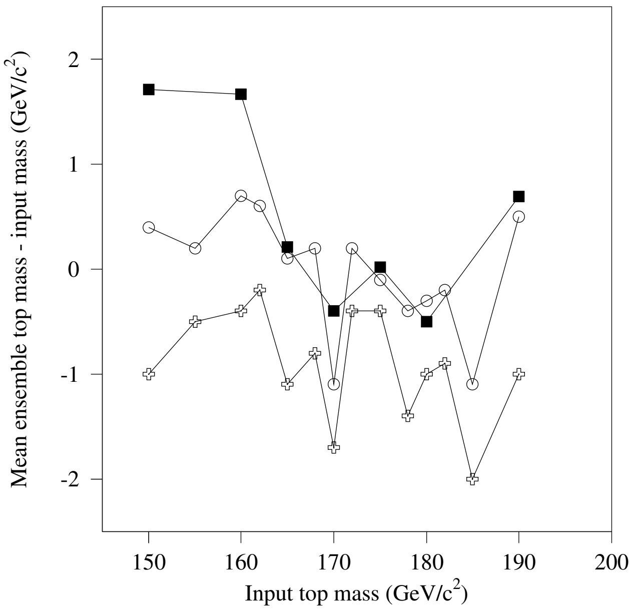

FIG. 41. Same as Fig. 40 for mean ensemble mass deviations.

TABLE XXX. Results of fits to average $\chi^2/2$ plots from MC. $m_{\rm min}$ is the minimum of a quadratic fit to the points, "width" is the width where the fit curve rises by 0.5, and $\langle m_{\rm fit}\rangle$ is the weighted average of the $m_{\rm fit}$ values, where the weights are $e^{-\chi^2/2}$. Entries labeled "jet high" and "jet low" are after scaling jet energies by $\pm(2.5\% + 0.5\ \mathrm{GeV})$.

Sample	HERWIG			ISAJET		
	$m_{\rm min}$ (GeV/c^2)	width (GeV/c^2)	$\langle m_{\rm fit}\rangle$ (GeV/c^2)	$m_{\rm min}$ (GeV/c^2)	width (GeV/c^2)	$\langle m_{\rm fit}\rangle$ (GeV/c^2)
HERWIG						
$m_t = 150$ GeV/c^2	149.5	16.4	159.5			
$m_t = 160$ GeV/c^2	157.7	17.1	165.5	158.9	20.4	165.4
$m_t = 165$ GeV/c^2	161.7	18.4	167.5			
$m_t = 170$ GeV/c^2	164.7	18.9	170.0	166.3	22.0	170.6
$m_t = 175$ GeV/c^2	169.9	19.6	173.1			
jet high	172.3	19.7	175.3			
jet low	166.5	18.4	171.1			
$m_t = 180$ GeV/c^2	173.2	20.5	175.9	172.4	23.9	175.2
$m_t = 190$ GeV/c^2	182.5	21.2	182.3	180.4	25.7	180.4
$m_t = 200$ GeV/c^2	191.3	21.9	188.0	188.7	26.9	185.8
VECBOS						
M_W^2 scale	156.4	29.9	166.2	152.8	28.0	164.0
$\langle p_T^{\rm jet}\rangle^2$ scale	147.1	24.5	160.4	142.2	23.1	157.1
QCD (data)	158.0	33.6	169.4			

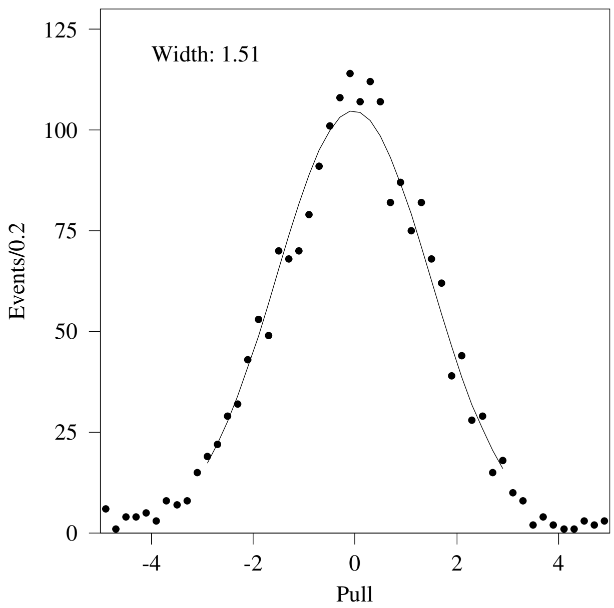

FIG. 42. Pull distribution for LB subset fits to precut ensemble samples with $m_t = 175$ GeV/c^2. The curve is a Gaussian fit to the region -3 to $+3$.

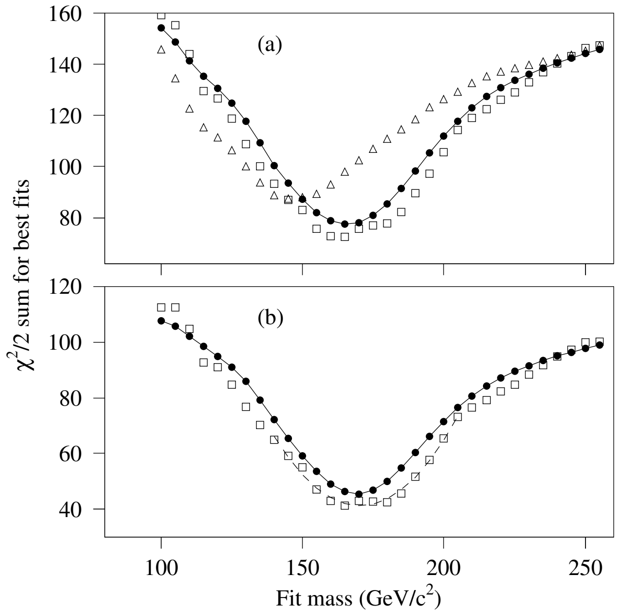

FIG. 43. (a) $\chi^2/2$ plots for the LB subset of the PR sample. Data are the open squares, filled circles are the prediction for a mixture of background and 175 GeV/c^2 top events, and open triangles are the prediction for pure background. The solid line joins the filled circles. (b) Background-subtracted $\chi^2/2$ plot for LB subsets. Data are the open squares, and filled circles are the prediction for 175 GeV/c^2 top events. The dashed curve is a parabola fit near the minimum.

TABLE XXXI. Ensembles with $N = 78$ and a 1:2 signal/background ratio. Entries labeled "jet high" and "jet low" are after scaling jet energies by $\pm(2.5\% + 0.5$ GeV). "Slope" is from a linear fit to the masses. The LB discriminant is used in the background determination for analyses of the precut samples.

	$m_{\rm min}$		$m_{\rm corr}$		Width containing	
	avg. mass (GeV/c^2)	RMS (GeV/c^2)	avg. mass (GeV/c^2)	RMS (GeV/c^2)	68.27% (GeV/c^2)	95.45% (GeV/c^2)
Precut sample, HERWIG						
$m_t = 165$ GeV/c^2	160.0	8.5	163.2	10.4	8.99	22.13
$m_t = 170$ GeV/c^2	163.5	8.4	167.5	10.3	8.85	21.88
$m_t = 175$ GeV/c^2	168.1	8.4	173.1	10.4	9.04	20.98
$m_t = 180$ GeV/c^2	171.8	9.5	177.7	11.7	10.00	22.77
Slope	0.80		0.98			
LB subset, HERWIG						
$m_t = 150$ GeV/c^2	150.6	7.3	151.7	8.9	7.68	16.84
$m_t = 160$ GeV/c^2	158.8	7.4	161.7	9.0	7.82	18.07
$m_t = 165$ GeV/c^2	161.6	7.1	165.2	8.7	7.34	17.27
$m_t = 170$ GeV/c^2	165.2	7.0	169.6	8.6	7.51	17.22
$m_t = 175$ GeV/c^2	169.6	6.7	175.0	8.2	7.93	16.83
jet high	172.6	7.5	178.7	9.2	8.22	18.32
jet low	167.0	8.0	171.7	9.9	8.35	19.73
$m_t = 180$ GeV/c^2	173.3	7.5	179.5	9.2	8.47	18.28
$m_t = 190$ GeV/c^2	182.4	7.7	190.7	9.5	8.61	19.54
Slope	0.78		0.96			
LB subset, ISAJET						
$m_t = 160$ GeV/c^2	158.6	8.9	161.5	10.9	9.23	21.02
$m_t = 170$ GeV/c^2	166.0	8.6	170.5	10.6	9.59	21.57
$m_t = 180$ GeV/c^2	173.0	9.2	179.1	11.3	10.38	22.44
$m_t = 190$ GeV/c^2	180.6	10.0	188.5	12.2	11.38	24.93
Slope	0.73		0.90			
NN subset, HERWIG						
$m_t = 150$ GeV/c^2	149.4	8.3	150.2	10.2	8.55	19.03
$m_t = 160$ GeV/c^2	158.1	8.3	160.8	10.2	8.75	20.21
$m_t = 165$ GeV/c^2	161.1	8.5	164.6	10.4	8.44	19.87
$m_t = 170$ GeV/c^2	164.8	7.8	169.1	9.6	8.41	19.10
$m_t = 175$ GeV/c^2	169.5	7.8	174.8	9.6	8.45	20.50
$m_t = 180$ GeV/c^2	173.3	8.5	179.5	10.5	9.53	21.30
$m_t = 190$ GeV/c^2	182.4	8.7	190.6	10.7	9.67	21.78
Slope	0.81		1.00			

TABLE XXXII. Results of fits to LB subsets using ensembles with $N = 26$ and no background. Entries labeled "jet high" and "jet low" are after scaling jet energies by $\pm(2.5\% + 0.5$ GeV). "Slope" is from a linear fit to the masses.

	$m_{\rm min}$		$m_{\rm corr}$		Width containing	
Input mass (GeV/c^2)	avg. mass (GeV/c^2)	RMS (GeV/c^2)	avg. mass (GeV/c^2)	RMS (GeV/c^2)	68.27% (GeV/c^2)	95.45% (GeV/c^2)
150	150.6	5.0	151.6	6.1	5.96	12.07
160	158.6	5.1	161.5	6.2	6.02	12.56
165	161.6	4.7	165.2	5.8	5.62	12.18
170	165.2	5.0	169.5	6.2	6.15	12.72
175	169.8	5.0	175.2	6.2	6.06	12.51
jet high	172.6	5.3	178.7	6.5	6.41	13.27
jet low	166.9	5.5	171.7	6.7	6.40	13.78
180	173.5	5.6	179.8	6.9	6.95	13.89
190	182.7	5.8	191.0	7.1	6.99	14.40
200	191.0	6.6	201.3	8.0	7.88	16.09
Slope	0.81		1.00			

TABLE XXXIII. Fits to data samples.

Cut	N	Δm_{fit} (GeV/c^2)	m_{min} (GeV/c^2)	m_{corr} (GeV/c^2)	BG fractions precut	subset
LB	32	10.0	171.0 ± 4.6	176.7 ± 8.4	0.60	0.32
LB	32	5.0	170.4 ± 4.3	176.0 ± 7.9	0.60	0.32
NN	33	10.0	164.3 ± 5.5	168.4 ± 10.1	0.65	0.41
Subset common to both PL and variable-mass						
LB	31	10.0	169.0 ± 4.6	174.3 ± 8.5	0.56	0.29
LB	31	5.0	169.8 ± 4.4	175.2 ± 8.0	0.56	0.29
NN	32	10.0	163.0 ± 5.4	166.8 ± 9.9	0.60	0.38

We select the data sample for analysis by requiring that each event have at least one fit with $\chi^2 < 10$. This yields a sample of 78 events, 32 of which pass the LB selection, and 33 of which pass the NN selection, with 27 events in common between these two subsets. (Due to differences in the kinematic fitting, three events in the variable-mass analysis fail the χ^2 cut for 3C SQUAW fits, and four events not in the variable-mass analysis are included in the PL analysis.) Results of fits to these samples are given in Table XXXIII. They are listed for Δm_{fit} values of both 5 and 10 GeV/c^2. A 5 GeV/c^2 increment gives slightly smaller uncertainties. The $\chi^2/2$ plot for the LB subsample is plotted in Fig. 43.

The top quark mass from the NN subset is smaller than that from the LB subset, and has a larger uncertainty. This is due to the fact that the events accepted by the NN selection but rejected by the LB selection tend to be of lower mass than those accepted by LB but rejected by NN. These low mass events are typically rejected from the LB subsample by the $H_{T2} > 90$ GeV cut. Ensemble studies show that the fraction of simulated experiments having an LB-NN difference larger than the 8.3 GeV/c^2 seen in the data is 6%.

If we look at the subset of events selected by both the PL and variable-mass analysis, there are 74 events, with 31 events passing the LB selection and 32 events passing the NN selection. Results of fits to these samples are also given in Table XXXIII.

F. Systematic uncertainties

This section gives estimates of the systematic uncertainties for the PL analysis. The uncertainty in the jet energy scale is $\pm(2.5\% + 0.5$ GeV) per jet (Sec. IV). To estimate the effect of this on m_{corr}, we redo the fits for a $t\bar{t}$ MC sample with all jets scaled up or down by this uncertainty. The results are given in Table XXX. After applying the slope correction, this yields an estimate of ± 3.6 GeV/c^2. Note that this is only valid in the limit of a large number of $t\bar{t}$ events with negligible background. We can also estimate this uncertainty by constructing ensembles with all the jets in the $t\bar{t}$ signal sample scaled up or down. The results are given in Table XXXI; the estimated uncertainty is ± 3.5 GeV/c^2. The same value for this uncertainty would be obtained using the mass shifts from ensemble studies with no background, as given in Table XXXII.

The differences seen in m_{min} between HERWIG events and ISAJET events are shown in Table XXX. The corresponding differences in m_{corr} vary from -1.6 to 2.6 GeV/c^2 over the range m_t = 160–200 GeV/c^2, and have a minimum between 170 and 180 GeV/c^2. We then construct ensembles using ISAJET events and compare these results to those from HERWIG. This is done in Table XXXI. The resulting difference varies from -0.9 to 2.2 GeV/c^2 over the range m_t = 160–190 GeV/c^2, so we assign a systematic uncertainty of 2.2 GeV/c^2 for the signal model.

We estimate the contribution to the systematic uncertainty due to the choice of the VECBOS Q^2 scale and fragmentation method by examining the four different choices listed in Table XXX. One can see that our choice of average jet p_T scale and HERWIG fragmentation represents an intermediate case. The resulting uncertainty in m_t is obtained by constructing ensembles from the different VECBOS parameter choices (but still using the favored choice for background calculation and subtraction). For ensemble samples with m_t = 175 GeV/c^2 events, the average corrected masses for the four choices range from 174.5 to 176.4 GeV/c^2, for a maximum difference of 1.9 GeV/c^2.

Some of the other systematic uncertainty contributions evaluated for the LB and NN analyses (see Table XXIX) cannot be determined in the same way for the PL analysis. The noise and multiple interaction uncertainty is determined from the shift in the mean fitted mass for the variable-mass fits, which are not used in the PL analysis. However, the kinematic fitters used give similar results, so the size of this effect for the PL analysis should be similar to that from the LB and NN variable-mass analyses. The uncertainty due to Monte Carlo statistics is assumed to be negligible. The LB-NN difference can be calculated from the PL ensemble results in Table XXXI. For the 170–180 GeV/c^2 mass range, the mean LB-NN difference is 0.23 GeV/c^2. Finally, the likelihood fit uncertainty contribution can be calculated from the four LB fit values given in Table XXXIII. The RMS of the four LB corrected mass values is 0.9 GeV/c^2. Combining in quadrature these uncertainty contributions with those for the energy scale (3.5 GeV/c^2), signal generator (2.2 GeV/c^2 from the maximum HERWIG-ISAJET difference in the 160–190 GeV/c^2 mass range), and VECBOS flavors (1.9 GeV/c^2) gives a total PL systematic uncertainty of 4.8 GeV/c^2.

G. Summary

Pseudolikelihood analysis of the LB subset of the data gives a top quark mass of 176.0 ± 7.9 (stat) ± 4.8 (syst) GeV/c^2. This is based upon a 14-point quadratic fit (with a mass increment of 5 GeV/c^2) to the

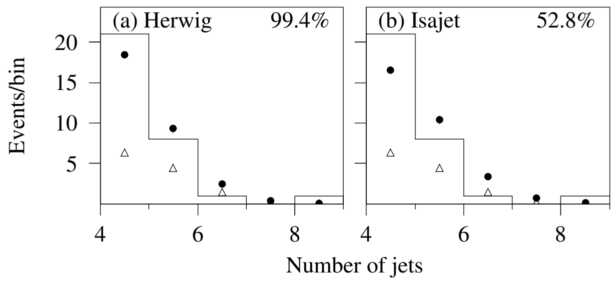

FIG. 44. Number of jets in each event with $E_T > 15$ GeV and $|\eta| < 2$ for (a) HERWIG ($m_t = 170$ GeV$/c^2$) and (b) ISAJET ($m_t = 170$ GeV$/c^2$). The histogram is data, open triangles are expected background, and filled circles are expected signal plus background.

background-subtracted $\chi^2/2$ plot over the range $m_{\text{fit}} =$ 140–205 GeV$/c^2$.

IX. FURTHER KINEMATIC STUDIES

This section presents distributions of additional kinematic quantities derived from the data. In these plots, the data sample is compared to a mixture of $t\bar{t}$ (generated with HERWIG with $m_t = 175$ GeV$/c^2$ unless otherwise specified) and background models. The distributions are shown for the LB subsample and are normalized according to the results of the LB analysis. There are 18.5 signal events and 12.5 background events expected in this subsample. The error bars shown on these plots are from signal and background sample statistics only, and do not include the correlated error in the overall normalization.

To test the compatibility of our predictions with the data, we use a Kolmogorov-Smirnov (K-S) test [36]. The resulting probability is indicated on each plot. Note that binning the data induces an upwards bias in the K-S probabilities. To mitigate this effect, all such probabilities for distributions of continuous variables are calculated using histograms consisting of 10,000 bins.

Figure 44 shows the distribution of the number of jets in each event in the sample. For comparison, the prediction of ISAJET is shown as well as that of HERWIG. (Note that since the number of jets is unavoidably a discrete variable, the K-S probabilities are expected to be biased high.) Figure 45 shows the transverse mass of the lepton and neutrino. The slight rise of the prediction at low m_T^W is due to the QCD multijet background. Figure 46 shows the total transverse momentum k_T (vector sum) of all the objects used in the mass fit. (The full jet corrections are used; however, for this plot only, all untagged jets are corrected using the light quark corrections.) Note that due to the procedure of using only the top four jets for the fit, this is not necessarily the actual transverse momentum of the $t\bar{t}$ system (k_T tends to be somewhat lower, on average).

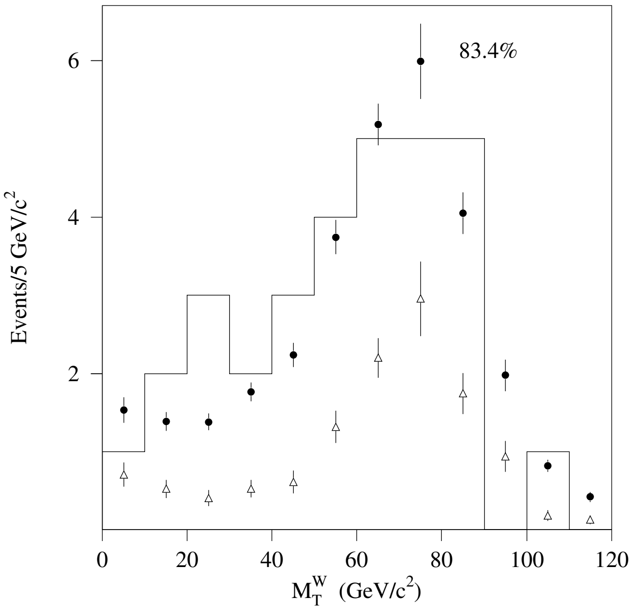

FIG. 45. Transverse mass of the lepton and neutrino. The histogram is data, open triangles are expected background, and filled circles are expected signal plus background.

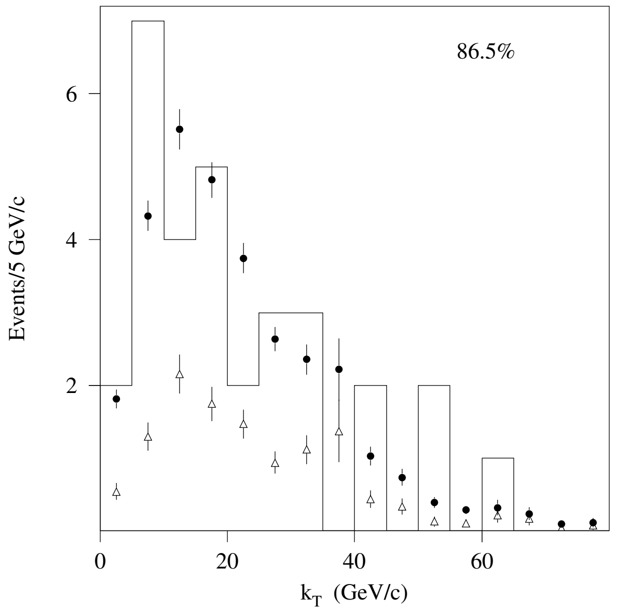

FIG. 46. Total transverse momentum k_T of all objects used in the mass fit (the highest four jets, the lepton, and the $\not{E}_T$). This is a vector sum. The histogram is data, open triangles are expected background, and filled circles are expected signal plus background.

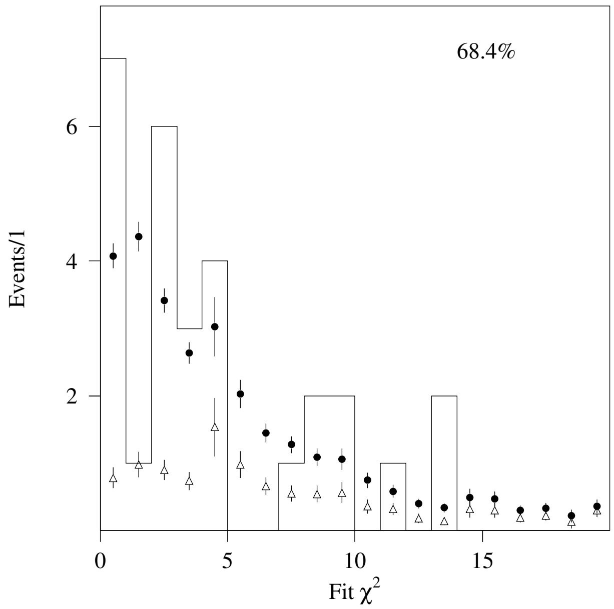

FIG. 47. χ^2 distributions from the 3C fit. The histogram is data (with two overflows), open triangles are expected background, and filled circles are expected signal plus background.

The remaining distributions depend on the results of the kinematic fit. For these, we plot the result corresponding to the jet permutation with the smallest χ^2. We also show the distributions which result if the data and Monte Carlo are refit with the additional constraint that $m_t = 173.3\ \mathrm{GeV}/c^2$. This is now a 3C fit. Note, however, that when making the χ^2 cut to define the sample, the 2C χ^2 is used in all cases; thus, adding the additional constraint does not change the sample definition. The distribution of the 3C fit χ^2 is shown in Fig. 47. There are five events with a 3C fit $\chi^2 > 10$, compared to ≈ 7 expected. They are consistent with a mixture of background and $t\bar{t}$ events where the wrong set of four jets was selected.

Figure 48 shows the invariant mass of the $t\bar{t}$ pair. Figure 49 shows the transverse momenta of the two top quarks, and Fig. 50 shows their pseudorapidity. Figures 51 and 52 show, respectively, the distance in η and ϕ between the two top quarks. The mean of the 13 K-S probabilities we calculate from continuous distributions is $(53\pm9)\%$, consistent with the hypothesis that our predictions for $t\bar{t}$ signal plus background adequately represent our data.

X. CONCLUSIONS

In summary, we measure the top quark mass using lepton + jets events to be $m_t(lj) = 173.3 \pm 5.6$ (stat) $\pm$ 5.5 (syst) GeV/c^2. We have also measured the top quark mass from dilepton events [8], yielding $m_t(ll) = 168.4 \pm 12.3$ (stat) $\pm$ 3.6 (syst) GeV/c^2. We combine

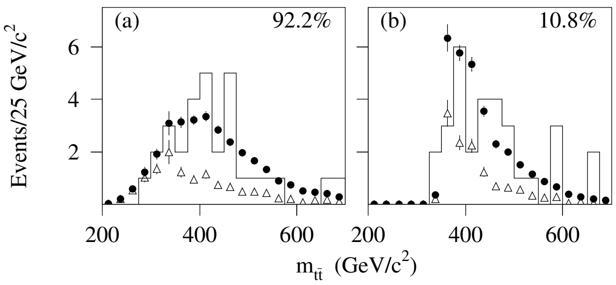

FIG. 48. Invariant mass distribution of the $t\bar{t}$ pair. The histogram is data, open triangles are expected background, and filled circles are expected signal plus background. (a) 2C fit, (b) 3C fit with $m_t = 173.3$ GeV/c^2.

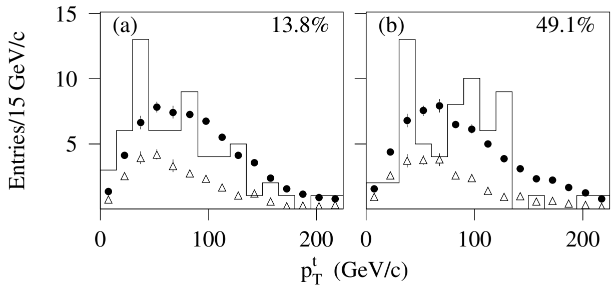

FIG. 49. Same as Fig. 48 for the transverse momenta of the top quarks (two entries per event).

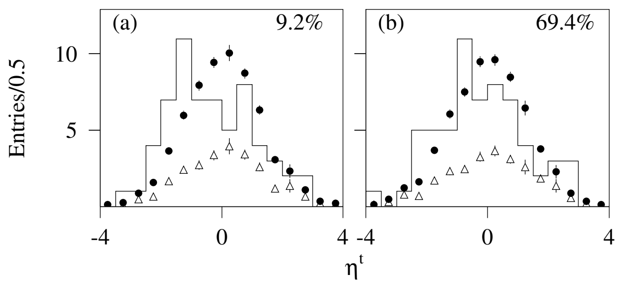

FIG. 50. Same as Fig. 48 for the pseudorapidities of the top quarks (two entries per event).

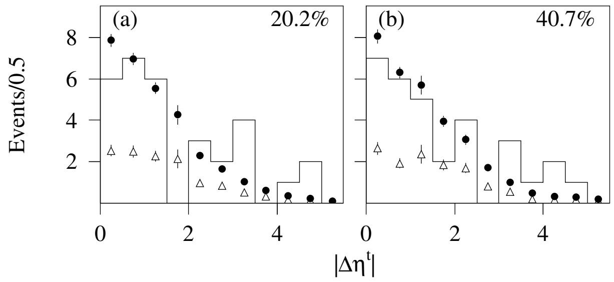

FIG. 51. Same as Fig. 48 for the difference in pseudorapidity η between the two top quarks.

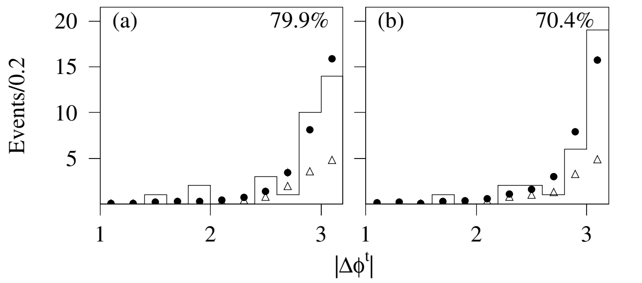

FIG. 52. Same as Fig. 48 for the difference in azimuthal angle ϕ between the two top quarks.

these two values, assuming that the systematics for jet energy scale, multiple interactions, and $t\bar{t}$ signal generator dependencies are fully correlated, and that other systematics are uncorrelated. The result is

$$\begin{aligned} m_t &= 172.1 \pm 5.2 \text{ (stat)} \pm 4.9 \text{ (syst) GeV}/c^2 \\ &= 172.1 \pm 7.1 \text{ GeV}/c^2. \end{aligned} \tag{10.1}$$

In a separate publication [7], we describe the measurement of the $p\bar{p} \to t\bar{t}$ production cross section. The result for $m_t = 172.1$ GeV/c^2 is

$$\sigma(m_t = 172.1 \text{ GeV}/c^2) = 5.6 \pm 1.8\,\text{pb}. \tag{10.2}$$

Our results are plotted in Fig. 53 and are compared to several theoretical calculations of the $t\bar{t}$ production cross section [37]. The agreement of the standard model expectations with our measurement is excellent. We also find agreement between our data and predictions for distributions of various kinematic variables for $t\bar{t}$ decays.

An alternate analysis technique using three constraint fits to fixed top quark masses using the lepton + jets data gives a result of $m_t(lj) = 176.0 \pm 7.9$ (stat) $\pm$ 4.8 (syst) GeV/c^2, consistent with the above result.

ACKNOWLEDGMENTS

We thank the staffs at Fermilab and collaborating institutions for their contributions to this work, and acknowledge support from the Department of Energy and National Science Foundation (U.S.A.), Commissariat à L'Energie Atomique (France), State Committee for Science and Technology and Ministry for Atomic Energy (Russia), CNPq (Brazil), Departments of Atomic Energy and Science and Education (India), Colciencias (Colombia), CONACyT (Mexico), Ministry of Education and KOSEF (Korea), CONICET and UBACyT (Argentina), and CAPES (Brazil).

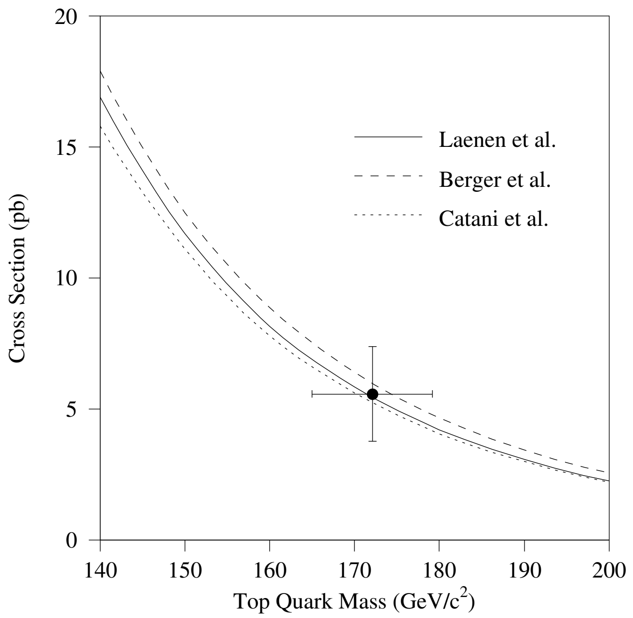

FIG. 53. Comparison of the measured top quark mass and production cross section with theoretical calculations [37].

* Visitor from Universidad San Francisco de Quito, Quito, Ecuador.

† Visitor from IHEP, Beijing, China.

[1] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **74**, 2626 (1995).
[2] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **74**, 2632 (1995).
[3] R. M. Barnett *et al.*, Phys. Rev. D **54**, 1 (1996); C. Campagnari and M. Franklin, Rev. Mod. Phys. **69**, 137 (1997); S. J. Wimpenny and B. L. Winer, Annu. Rev. Nucl. Part. Sci. **46**, 149 (1996).
[4] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. D **52**, 4877 (1995).
[5] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **79**, 1197 (1997).
[6] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **80**, 2767 (1998).
[7] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **79**, 1203 (1997).
[8] DØ Collaboration, B. Abbott *et al.*, Phys. Rev. Lett. **80**, 2063 (1998).
[9] DØ Collaboration, S. Abachi *et al.*, Nucl. Instrum. Methods Phys. Res., Sect. A **338**, 185 (1994).
[10] S. Baker and R. Cousins, Nucl. Instrum. Methods Phys. Res. **221**, 437 (1984).
[11] Note that one event which would have otherwise passed the cuts, event (95653, 10822), was removed from this analysis because it was selected by the dilepton mass analysis [8]. That event contained two electron candidates. However, one of the electrons just grazed the edge of the outer drift chamber, passing through the inner layers but not the outer ones. Thus, although there are hits in the inner layers, a complete track was not recon-

structed. This event also contains a muon near one of its jets, consistent with b quark decay. If this event is treated as a $l+$jets candidate, it has a fit χ^2 of 0.92 and $m_{\rm fit} = 138.7$ GeV/c^2.

[12] DØ Collaboration, J. Kotcher, in *1994 Beijing Calorimetry Symposium, Beijing, China*, edited by H. S. Chen (IHEP, Beijing, 1994); DØ Collaboration, J. Guida, Nucl. Phys. Proc. Suppl. **44**, 153 (1995). The current values of these factors are CC: 1.0443 ± 0.0021, North EC: 1.0463 ± 0.0085, South EC: 1.0230 ± 0.0086.

[13] DØ Collaboration, R. Kehoe, in *Proceedings of the sixth International Conference on Calorimetry in High Energy Physics, Frascati, Italy* (World Scientific, River Edge, NJ, 1996), Fermilab-Conf-96/284-E, 1996.

[14] G. Marchesini *et al.*, Comp. Phys. Comm. **67**, 465 (1992).

[15] The current version of the standard corrections (which was not available in time for this analysis) is better behaved in this regard.

[16] DØ Collaboration, S. Abachi *et al.*, Phys. Lett. B **357**, 500 (1995).

[17] H. L. Lai *et al.*, Phys. Rev. D **51**, 4763 (1995).

[18] G. Altarelli and G. Parisi, Nucl. Phys. B **126**, 298 (1977).

[19] DØ Collaboration, J. Womersley, in *Proceedings of the XXVI International Conference on High Energy Physics, Dallas, Texas*, edited by J. R. Sanford (AIP, New York, 1993), p. 1800.

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

[21] F. Paige and S. Protopopescu, BNL Report 38034, Brookhaven, 1986, (unpublished), release 7.21.

[22] F. A. Berends, H. Kuijf, B. Tausk, and W. T. Giele, Nucl. Phys. B **357**, 32 (1991).

[23] CDF Collaboration, F. Abe *et al.*, Phys. Rev. D **51**, 4623 (1995); CDF Collaboration, F. Abe *et al.*, Phys. Rev. D **52**, 2605 (1995).

[24] V. D. Barger and R. J. N. Phillips, *Collider Physics* (Addison-Wesley, Reading, MA, 1987), p. 281.

[25] Some additional transformations are performed on the x_i's before the fit so that the functions can be better approximated by low order polynomials:

$$\begin{aligned} x'_1 &= \exp\left[-\left(\sqrt{\not{E}_T/(1\ \mathrm{GeV})} - A\right)/B\right] \\ &\quad (A, B = 5, 1.5 \text{ for electrons and } 4.47, 3 \text{ for muons}), \\ x'_2 &= \exp(-11\mathcal{A}), \\ x'_3 &= \ln x_3, \\ x'_4 &= \sqrt{x_4}. \end{aligned}$$

[26] C. Peterson, T. Rögnvaldsson, and L. Lönnblad, Comp. Phys. Comm. **81**, 185 (1994).

[27] E. K. Blum and L. K. Li, Neural Networks **4**, 511 (1991); D. W. Ruck *et al.*, IEEE Trans. Neural Networks **1**, 296 (1990).

[28] O. I. Dahl, T. B. Day, F. T. Solmitz, and N. L. Gould, Technical Report No. P-126, Lawrence Radiation Laboratory, 1968, Berkeley, California (unpublished); J. P. Berge, F. T. Solmitz, and H. D. Taft, Review of Scientific Instruments **32**, 538 (1961).

[29] S. Snyder, Ph.D. thesis, State University of New York at Stony Brook, 1995 (unpublished).

[30] F. James, MINUIT *Minimization Package Reference Manual*, 1994, CERN Program Library Number D506 (unpublished).

[31] P. C. Bhat, H. B. Prosper, and S. Snyder, Phys. Lett. B **407**, 73 (1997).

[32] Detailed information on the candidates used in this analysis is available at `http://www-d0.fnal.gov/public/top/d0-ljets-topmass-candidates.html`. It will also be available from the AIP E-PAPS service.

[33] R. Barlow and C. Beeston, Comp. Phys. Comm. **77**, 219 (1993).

[34] When we increased $\Lambda_{\rm QCD}$ from its nominal value of 0.18 GeV, we also had to adjust the HERWIG quark and gluon virtuality cutoffs `vqcut` and `vgcut` as follows:

$\Lambda_{\rm QCD}$ (GeV)	`vqcut` (GeV/c^2)	`vgcut` (GeV/c^2)
0.18	0.48	0.10
0.21	0.58	0.10
0.25	0.68	0.30

[35] A. Martin, W. Stirling, and R. Roberts, Phys. Lett. B **356**, 89 (1995).

[36] W. T. Eadie *et al.*, *Statistical Methods in Experimental Physics* (North Holland, Amsterdam, 1971).

[37] E. Laenen, J. Smith, and W. L. van Neerven, Phys. Lett. B **321**, 254 (1994); E. Berger and H. Contopanagos, Phys. Lett. B **361**, 115 (1995); E. Berger and H. Contopanagos, Phys. Rev. D **54**, 3085 (1996); S. Catani, M. L. Mangano, P. Nason, and L. Trentadue, Phys. Lett. B **378**, 329 (1996).