FERMILAB-PUB-04-275-E
CDF-7138

Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using Lepton + Jets Events with Secondary Vertex b-tagging

D. Acosta,[16] J. Adelman,[12] T. Affolder,[9] T. Akimoto,[54] M.G. Albrow,[15] D. Ambrose,[43] S. Amerio,[42] D. Amidei,[33] A. Anastassov,[50] K. Anikeev,[31] A. Annovi,[44] J. Antos,[1] M. Aoki,[54] G. Apollinari,[15] T. Arisawa,[56] J-F. Arguin,[32] A. Artikov,[13] W. Ashmanskas,[15] A. Attal,[7] F. Azfar,[41] P. Azzi-Bacchetta,[42] N. Bacchetta,[42] H. Bachacou,[28] W. Badgett,[15] A. Barbaro-Galtieri,[28] G.J. Barker,[25] V.E. Barnes,[46] B.A. Barnett,[24] S. Baroiant,[6] M. Barone,[17] G. Bauer,[31] F. Bedeschi,[44] S. Behari,[24] S. Belforte,[53] G. Bellettini,[44] J. Bellinger,[58] E. Ben-Haim,[15] D. Benjamin,[14] A. Beretvas,[15] A. Bhatti,[48] M. Binkley,[15] D. Bisello,[42] M. Bishai,[15] R.E. Blair,[2] C. Blocker,[5] K. Bloom,[33] B. Blumenfeld,[24] A. Bocci,[48] A. Bodek,[47] G. Bolla,[46] A. Bolshov,[31] P.S.L. Booth,[29] D. Bortoletto,[46] J. Boudreau,[45] S. Bourov,[15] C. Bromberg,[34] E. Brubaker,[12] J. Budagov,[13] H.S. Budd,[47] K. Burkett,[15] G. Busetto,[42] P. Bussey,[19] K.L. Byrum,[2] S. Cabrera,[14] M. Campanelli,[18] M. Campbell,[33] A. Canepa,[46] M. Casarsa,[53] D. Carlsmith,[58] S. Carron,[14] R. Carosi,[44] M. Cavalli-Sforza,[3] A. Castro,[4] P. Catastini,[44] D. Cauz,[53] A. Cerri,[28] C. Cerri,[44] L. Cerrito,[23] J. Chapman,[33] C. Chen,[43] Y.C. Chen,[1] M. Chertok,[6] G. Chiarelli,[44] G. Chlachidze,[13] F. Chlebana,[15] I. Cho,[27] K. Cho,[27] D. Chokheli,[13] M.L. Chu,[1] S. Chuang,[58] J.Y. Chung,[38] W-H. Chung,[58] Y.S. Chung,[47] C.I. Ciobanu,[23] M.A. Ciocci,[44] A.G. Clark,[18] D. Clark,[5] M. Coca,[47] A. Connolly,[28] M. Convery,[48] J. Conway,[6] B. Cooper,[30] M. Cordelli,[17] G. Cortiana,[42] J. Cranshaw,[52] J. Cuevas,[10] R. Culbertson,[15] C. Currat,[28] D. Cyr,[58] D. Dagenhart,[5] S. Da Ronco,[42] S. D'Auria,[19] P. de Barbaro,[47] S. De Cecco,[49] G. De Lentdecker,[47] S. Dell'Agnello,[17] M. Dell'Orso,[44] S. Demers,[47] L. Demortier,[48] M. Deninno,[4] D. De Pedis,[49] P.F. Derwent,[15] C. Dionisi,[49] J.R. Dittmann,[15] P. Doksus,[23] A. Dominguez,[28] S. Donati,[44] M. Donega,[18] J. Donini,[42] M. D'Onofrio,[18] T. Dorigo,[42] V. Drollinger,[36] K. Ebina,[56] N. Eddy,[23] R. Ely,[28] R. Erbacher,[6] M. Erdmann,[25] D. Errede,[23] S. Errede,[23] R. Eusebi,[47] H-C. Fang,[28] S. Farrington,[29] I. Fedorko,[44] R.G. Feild,[59] M. Feindt,[25] J.P. Fernandez,[46] C. Ferretti,[33] R.D. Field,[16] I. Fiori,[44] G. Flanagan,[34] B. Flaugher,[15] L.R. Flores-Castillo,[45] A. Foland,[20]

S. Forrester,[6] G.W. Foster,[15] M. Franklin,[20] J.C. Freeman,[28] H. Frisch,[12] Y. Fujii,[26]
I. Furic,[12] A. Gajjar,[29] A. Gallas,[37] J. Galyardt,[11] M. Gallinaro,[48] M. Garcia-Sciveres,[28]
A.F. Garfinkel,[46] C. Gay,[59] H. Gerberich,[14] D.W. Gerdes,[33] E. Gerchtein,[11] S. Giagu,[49]
P. Giannetti,[44] A. Gibson,[28] K. Gibson,[11] C. Ginsburg,[58] K. Giolo,[46] M. Giordani,[53]
G. Giurgiu,[11] V. Glagolev,[13] D. Glenzinski,[15] M. Gold,[36] N. Goldschmidt,[33] D. Goldstein,[7]
J. Goldstein,[41] G. Gomez,[10] G. Gomez-Ceballos,[31] M. Goncharov,[51] O. González,[46]
I. Gorelov,[36] A.T. Goshaw,[14] Y. Gotra,[45] K. Goulianos,[48] A. Gresele,[4] M. Griffiths,[29]
C. Grosso-Pilcher,[12] U. Grundler,[23] M. Guenther,[46] J. Guimaraes da Costa,[20] C. Haber,[28]
K. Hahn,[43] S.R. Hahn,[15] E. Halkiadakis,[47] A. Hamilton,[32] B-Y. Han,[47] R. Handler,[58]
F. Happacher,[17] K. Hara,[54] M. Hare,[55] R.F. Harr,[57] R.M. Harris,[15] F. Hartmann,[25]
K. Hatakeyama,[48] J. Hauser,[7] C. Hays,[14] H. Hayward,[29] E. Heider,[55] B. Heinemann,[29]
J. Heinrich,[43] M. Hennecke,[25] M. Herndon,[24] C. Hill,[9] D. Hirschbuehl,[25] A. Hocker,[47]
K.D. Hoffman,[12] A. Holloway,[20] S. Hou,[1] M.A. Houlden,[29] B.T. Huffman,[41] Y. Huang,[14]
R.E. Hughes,[38] J. Huston,[34] K. Ikado,[56] J. Incandela,[9] G. Introzzi,[44] M. Iori,[49]
Y. Ishizawa,[54] C. Issever,[9] A. Ivanov,[47] Y. Iwata,[22] B. Iyutin,[31] E. James,[15] D. Jang,[50]
J. Jarrell,[36] D. Jeans,[49] H. Jensen,[15] E.J. Jeon,[27] M. Jones,[46] K.K. Joo,[27] S. Jun,[11]
T. Junk,[23] T. Kamon,[51] J. Kang,[33] M. Karagoz Unel,[37] P.E. Karchin,[57] S. Kartal,[15]
Y. Kato,[40] Y. Kemp,[25] R. Kephart,[15] U. Kerzel,[25] V. Khotilovich,[51] B. Kilminster,[38]
D.H. Kim,[27] H.S. Kim,[23] J.E. Kim,[27] M.J. Kim,[11] M.S. Kim,[27] S.B. Kim,[27] S.H. Kim,[54]
T.H. Kim,[31] Y.K. Kim,[12] B.T. King,[29] M. Kirby,[14] L. Kirsch,[5] S. Klimenko,[16]
B. Knuteson,[31] B.R. Ko,[14] H. Kobayashi,[54] P. Koehn,[38] D.J. Kong,[27] K. Kondo,[56]
J. Konigsberg,[16] K. Kordas,[32] A. Korn,[31] A. Korytov,[16] K. Kotelnikov,[35] A.V. Kotwal,[14]
A. Kovalev,[43] J. Kraus,[23] I. Kravchenko,[31] A. Kreymer,[15] J. Kroll,[43] M. Kruse,[14]
V. Krutelyov,[51] S.E. Kuhlmann,[2] N. Kuznetsova,[15] A.T. Laasanen,[46] S. Lai,[32]
S. Lami,[48] S. Lammel,[15] J. Lancaster,[14] M. Lancaster,[30] R. Lander,[6] K. Lannon,[38]
A. Lath,[50] G. Latino,[36] R. Lauhakangas,[21] I. Lazzizzera,[42] Y. Le,[24] C. Lecci,[25]
T. LeCompte,[2] J. Lee,[27] J. Lee,[47] S.W. Lee,[51] R. Lefèvre,[3] N. Leonardo,[31] S. Leone,[44]
J.D. Lewis,[15] K. Li,[59] C. Lin,[59] C.S. Lin,[15] M. Lindgren,[15] T.M. Liss,[23] D.O. Litvintsev,[15]
T. Liu,[15] Y. Liu,[18] N.S. Lockyer,[43] A. Loginov,[35] M. Loreti,[42] P. Loverre,[49] R-S. Lu,[1]
D. Lucchesi,[42] P. Lujan,[28] P. Lukens,[15] G. Lungu,[16] L. Lyons,[41] J. Lys,[28] R. Lysak,[1]

D. MacQueen,[32] R. Madrak,[20] K. Maeshima,[15] P. Maksimovic,[24] L. Malferrari,[4]
G. Manca,[29] R. Marginean,[38] M. Martin,[24] A. Martin,[59] V. Martin,[37] M. Martínez,[3]
T. Maruyama,[54] H. Matsunaga,[54] M. Mattson,[57] P. Mazzanti,[4] K.S. McFarland,[47]
D. McGivern,[30] P.M. McIntyre,[51] P. McNamara,[50] R. NcNulty,[29] S. Menzemer,[31]
A. Menzione,[44] P. Merkel,[15] C. Mesropian,[48] A. Messina,[49] T. Miao,[15] N. Miladinovic,[5]
L. Miller,[20] R. Miller,[34] J.S. Miller,[33] R. Miquel,[28] S. Miscetti,[17] G. Mitselmakher,[16]
A. Miyamoto,[26] Y. Miyazaki,[40] N. Moggi,[4] B. Mohr,[7] R. Moore,[15] M. Morello,[44]
A. Mukherjee,[15] M. Mulhearn,[31] T. Muller,[25] R. Mumford,[24] A. Munar,[43] P. Murat,[15]
J. Nachtman,[15] S. Nahn,[59] I. Nakamura,[43] I. Nakano,[39] A. Napier,[55] R. Napora,[24]
D. Naumov,[36] V. Necula,[16] F. Niell,[33] J. Nielsen,[28] C. Nelson,[15] T. Nelson,[15] C. Neu,[43]
M.S. Neubauer,[8] C. Newman-Holmes,[15] A-S. Nicollerat,[18] T. Nigmanov,[45] L. Nodulman,[2]
O. Norniella,[3] K. Oesterberg,[21] T. Ogawa,[56] S.H. Oh,[14] Y.D. Oh,[27] T. Ohsugi,[22]
T. Okusawa,[40] R. Oldeman,[49] R. Orava,[21] W. Orejudos,[28] C. Pagliarone,[44] F. Palmonari,[44]
R. Paoletti,[44] V. Papadimitriou,[15] S. Pashapour,[32] J. Patrick,[15] G. Pauletta,[53] M. Paulini,[11]
T. Pauly,[41] C. Paus,[31] D. Pellett,[6] A. Penzo,[53] T.J. Phillips,[14] G. Piacentino,[44]
J. Piedra,[10] K.T. Pitts,[23] C. Plager,[7] A. Pompoš,[46] L. Pondrom,[58] G. Pope,[45]
O. Poukhov,[13] F. Prakoshyn,[13] T. Pratt,[29] A. Pronko,[16] J. Proudfoot,[2] F. Ptohos,[17]
G. Punzi,[44] J. Rademacker,[41] A. Rakitine,[31] S. Rappoccio,[20] F. Ratnikov,[50] H. Ray,[33]
A. Reichold,[41] B. Reisert,[15] V. Rekovic,[36] P. Renton,[41] M. Rescigno,[49] F. Rimondi,[4]
K. Rinnert,[25] L. Ristori,[44] W.J. Robertson,[14] A. Robson,[41] T. Rodrigo,[10] S. Rolli,[55]
L. Rosenson,[31] R. Roser,[15] R. Rossin,[42] C. Rott,[46] J. Russ,[11] A. Ruiz,[10] D. Ryan,[55]
H. Saarikko,[21] S. Sabik,[32] A. Safonov,[6] R. St. Denis,[19] W.K. Sakumoto,[47] G. Salamanna,[49]
D. Saltzberg,[7] C. Sanchez,[3] A. Sansoni,[17] L. Santi,[53] S. Sarkar,[49] K. Sato,[54] P. Savard,[32]
A. Savoy-Navarro,[15] P. Schlabach,[15] E.E. Schmidt,[15] M.P. Schmidt,[59] M. Schmitt,[37]
L. Scodellaro,[42] A. Scribano,[44] F. Scuri,[44] A. Sedov,[46] S. Seidel,[36] Y. Seiya,[40] F. Semeria,[4]
L. Sexton-Kennedy,[15] I. Sfiligoi,[17] M.D. Shapiro,[28] T. Shears,[29] P.F. Shepard,[45]
M. Shimojima,[54] M. Shochet,[12] Y. Shon,[58] I. Shreyber,[35] A. Sidoti,[44] J. Siegrist,[28]
M. Siket,[1] A. Sill,[52] P. Sinervo,[32] A. Sisakyan,[13] A. Skiba,[25] A.J. Slaughter,[15] K. Sliwa,[55]
D. Smirnov,[36] J.R. Smith,[6] F.D. Snider,[15] R. Snihur,[32] S.V. Somalwar,[50] J. Spalding,[15]
M. Spezziga,[52] L. Spiegel,[15] F. Spinella,[44] M. Spiropulu,[9] P. Squillacioti,[44] H. Stadie,[25]

A. Stefanini,[44] B. Stelzer,[32] O. Stelzer-Chilton,[32] J. Strologas,[36] D. Stuart,[9] A. Sukhanov,[16] K. Sumorok,[31] H. Sun,[55] T. Suzuki,[54] A. Taffard,[23] R. Tafirout,[32] S.F. Takach,[57] H. Takano,[54] R. Takashima,[22] Y. Takeuchi,[54] K. Takikawa,[54] M. Tanaka,[2] R. Tanaka,[39] N. Tanimoto,[39] S. Tapprogge,[21] M. Tecchio,[33] P.K. Teng,[1] K. Terashi,[48] R.J. Tesarek,[15] S. Tether,[31] J. Thom,[15] A.S. Thompson,[19] E. Thomson,[43] P. Tipton,[47] V. Tiwari,[11] S. Tkaczyk,[15] D. Toback,[51] K. Tollefson,[34] T. Tomura,[54] D. Tonelli,[44] M. Tönnesmann,[34] S. Torre,[44] D. Torretta,[15] S. Tourneur,[15] W. Trischuk,[32] J. Tseng,[41] R. Tsuchiya,[56] S. Tsuno,[39] D. Tsybychev,[16] N. Turini,[44] M. Turner,[29] F. Ukegawa,[54] T. Unverhau,[19] S. Uozumi,[54] D. Usynin,[43] L. Vacavant,[28] A. Vaiciulis,[47] A. Varganov,[33] E. Vataga,[44] S. Vejcik III,[15] G. Velev,[15] V. Veszpremi,[46] G. Veramendi,[23] T. Vickey,[23] R. Vidal,[15] I. Vila,[10] R. Vilar,[10] I. Vollrath,[32] I. Volobouev,[28] M. von der Mey,[7] P. Wagner,[51] R.G. Wagner,[2] R.L. Wagner,[15] W. Wagner,[25] R. Wallny,[7] T. Walter,[25] T. Yamashita,[39] K. Yamamoto,[40] Z. Wan,[50] M.J. Wang,[1] S.M. Wang,[16] A. Warburton,[32] B. Ward,[19] S. Waschke,[19] D. Waters,[30] T. Watts,[50] M. Weber,[28] W.C. Wester III,[15] B. Whitehouse,[55] A.B. Wicklund,[2] E. Wicklund,[15] H.H. Williams,[43] P. Wilson,[15] B.L. Winer,[38] P. Wittich,[43] S. Wolbers,[15] M. Wolter,[55] M. Worcester,[7] S. Worm,[50] T. Wright,[33] X. Wu,[18] F. Würthwein,[8] A. Wyatt,[30] A. Yagil,[15] U.K. Yang,[12] W. Yao,[28] G.P. Yeh,[15] K. Yi,[24] J. Yoh,[15] P. Yoon,[47] K. Yorita,[56] T. Yoshida,[40] I. Yu,[27] S. Yu,[43] Z. Yu,[59] J.C. Yun,[15] L. Zanello,[49] A. Zanetti,[53] I. Zaw,[20] F. Zetti,[44] J. Zhou,[50] A. Zsenei,[18] and S. Zucchelli[4]

(CDF Collaboration)

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

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

[3]*Institut de Fisica d'Altes Energies, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain*

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

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

[6]*University of California at Davis, Davis, California 95616*

[7]University of California at Los Angeles, Los Angeles, California 90024
[8]University of California at San Diego, La Jolla, California 92093
[9]University of California at Santa Barbara, Santa Barbara, California 93106
[10]Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain
[11]Carnegie Mellon University, Pittsburgh, PA 15213
[12]Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637
[13]Joint Institute for Nuclear Research, RU-141980 Dubna, Russia
[14]Duke University, Durham, North Carolina 27708
[15]Fermi National Accelerator Laboratory, Batavia, Illinois 60510
[16]University of Florida, Gainesville, Florida 32611
[17]Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy
[18]University of Geneva, CH-1211 Geneva 4, Switzerland
[19]Glasgow University, Glasgow G12 8QQ, United Kingdom
[20]Harvard University, Cambridge, Massachusetts 02138
[21]The Helsinki Group: Helsinki Institute of Physics; and Division of High Energy Physics, Department of Physical Sciences, University of Helsinki, FIN-00044, Helsinki, Finland
[22]Hiroshima University, Higashi-Hiroshima 724, Japan
[23]University of Illinois, Urbana, Illinois 61801
[24]The Johns Hopkins University, Baltimore, Maryland 21218
[25]Institut für Experimentelle Kernphysik, Universität Karlsruhe, 76128 Karlsruhe, Germany
[26]High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305, Japan
[27]Center for High Energy Physics: Kyungpook National University, Taegu 702-701; Seoul National University, Seoul 151-742; and SungKyunKwan University, Suwon 440-746; Korea
[28]Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720
[29]University of Liverpool, Liverpool L69 7ZE, United Kingdom
[30]University College London, London WC1E 6BT, United Kingdom
[31]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
[32]Institute of Particle Physics: McGill University, Montréal, Canada H3A 2T8; and University of Toronto, Toronto, Canada M5S 1A7

[33]*University of Michigan, Ann Arbor, Michigan 48109*
[34]*Michigan State University, East Lansing, Michigan 48824*
[35]*Institution for Theoretical and Experimental Physics, ITEP, Moscow 117259, Russia*
[36]*University of New Mexico, Albuquerque, New Mexico 87131*
[37]*Northwestern University, Evanston, Illinois 60208*
[38]*The Ohio State University, Columbus, Ohio 43210*
[39]*Okayama University, Okayama 700-8530, Japan*
[40]*Osaka City University, Osaka 588, Japan*
[41]*University of Oxford, Oxford OX1 3RH, United Kingdom*
[42]*University of Padova, Istituto Nazionale di Fisica Nucleare, Sezione di Padova-Trento, I-35131 Padova, Italy*
[43]*University of Pennsylvania, Philadelphia, Pennsylvania 19104*
[44]*Istituto Nazionale di Fisica Nucleare, University and Scuola Normale Superiore of Pisa, I-56100 Pisa, Italy*
[45]*University of Pittsburgh, Pittsburgh, Pennsylvania 15260*
[46]*Purdue University, West Lafayette, Indiana 47907*
[47]*University of Rochester, Rochester, New York 14627*
[48]*The Rockefeller University, New York, New York 10021*
[49]*Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1, University di Roma "La Sapienza," I-00185 Roma, Italy*
[50]*Rutgers University, Piscataway, New Jersey 08855*
[51]*Texas A&M University, College Station, Texas 77843*
[52]*Texas Tech University, Lubbock, Texas 79409*
[53]*Istituto Nazionale di Fisica Nucleare, University of Trieste/ Udine, Italy*
[54]*University of Tsukuba, Tsukuba, Ibaraki 305, Japan*
[55]*Tufts University, Medford, Massachusetts 02155*
[56]*Waseda University, Tokyo 169, Japan*
[57]*Wayne State University, Detroit, Michigan 48201*
[58]*University of Wisconsin, Madison, Wisconsin 53706*
[59]*Yale University, New Haven, Connecticut 06520*

(Dated: December 20, 2004)

Abstract

We present a measurement of the $t\bar{t}$ production cross section using events with one charged lepton and jets from $p\bar{p}$ collisions at a center-of-mass energy of 1.96 TeV. In these events, heavy flavor quarks from top quark decay are identified with a secondary vertex tagging algorithm. From $162\,\mathrm{pb}^{-1}$ of data collected by the Collider Detector at Fermilab, a total of 48 candidate events are selected, where 13.5 ± 1.8 events are expected from background contributions. We measure a $t\bar{t}$ production cross section of $5.6^{+1.2}_{-1.1}(\mathrm{stat.})^{+0.9}_{-0.6}(\mathrm{syst.})$ pb.

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

I. INTRODUCTION

The top quark is pair-produced in $p\bar{p}$ collisions through quark-antiquark annihilation and gluon-gluon fusion. The measurement of the $t\bar{t}$ cross section tests the QCD calculations for the pair production of a massive colored triplet. These calculations have been performed in perturbation theory to next-to-leading order [1, 2]. Recent work on corrections for soft gluon emission show that their effect on the cross section is small, and that they reduce the theoretical uncertainty arising from the choice of renormalization and factorization scales to less than 5% over the expected range of top masses and parton distribution functions (PDFs). The leading theoretical uncertainties are in the PDFs, arising mostly from the understanding of the gluon distributions at large parton x. The total theoretical uncertainty is approximately 15% [2]. At $\sqrt{s}$=1.96 TeV, the predicted $t\bar{t}$ production cross section is $\sigma_{t\bar{t}} = 6.7^{+0.7}_{-0.9}$ pb at $m_t = 175\,\mathrm{GeV}/c^2$ [2]. For every $1\,\mathrm{GeV}/c^2$ increase in the top mass over the interval $170 < m_t < 190\,\mathrm{GeV}/c^2$, the $t\bar{t}$ cross section decreases by 0.2 pb.

The Standard Model top quark decays to a W boson and a b quark almost 100% of the time. Top quark pair production thus gives rise to two W bosons and two "b jets" from b quark fragmentation. When exactly one W decays leptonically, the $t\bar{t}$ event typically contains a high transverse momentum lepton, missing transverse energy from the undetected neutrino, and four high transverse momentum jets, two of which originate from b quarks. This mode is labelled "W plus jets" or "lepton plus jets." Since the final state branching ratio is directly related to the W branching ratios, the $t\bar{t}$ rate into a particular final state measures both the production and decay properties of the top quark. An unexpected result could thus indicate either a non-standard source of top-like events, or a modification of the top decay branching ratios.

The $p\bar{p}$ collisions for this measurement of $t\bar{t}$ production were produced during Run II of the Fermilab Tevatron. The data were recorded at CDF II, a general purpose detector which combines charged particle tracking, sampling calorimeters, and fine-grained muon detection. Isolating the lepton plus jets decay mode of the top quark builds on the detailed understanding of inclusive leptonic W boson decays in CDF II [3]. The $t\bar{t}$ signature is mimicked by processes in which a W boson is produced in association with several hadronic jets with large transverse momentum. To separate the $t\bar{t}$ events from this background we use precision silicon tracking to b-tag jets containing a secondary vertex from a b hadron decay.

Background contributions from fake Ws, mis-identified secondary vertices and heavy flavor production processes such as $Wb\bar{b}$ are estimated using a combination of Monte Carlo calculations and independent measurements in control data samples. An excess in the number of events which contain a lepton, missing energy, and three or more jets with at least one b-tag is the signal of $t\bar{t}$ production and is used to measure the production cross section $\sigma_{t\bar{t}}$. The dataset defined by this analysis forms the basis for other measurements of top quark properties, such as the top quark mass and the helicity of W bosons produced in top decays.

This measurement builds on the b-tagging techniques employed by CDF at the Tevatron Run I. Then, at $\sqrt{s}$=1.8 TeV, a similar analysis of lepton+jets events with b-tags gave a $t\bar{t}$ cross section of $\sigma_{t\bar{t}} = 5.1 \pm 1.5$ pb [4], compared to an expected value of $\sigma_{t\bar{t}} = 5.2^{+0.5}_{-0.7}$ pb at $m_{\mathrm{t}} = 175\,\mathrm{GeV}/c^2$ [2]. Here, using a larger dataset collected at higher center-of-mass energy as well as improved Monte Carlo tools and detector simulations, we have re-analyzed the heavy flavor fraction in W events and improved our understanding of b-tagging efficiencies, including the contribution of material interactions to fake b tags. In addition, the significance of the measurement is optimized by requiring a large scalar sum of the transverse energies of all objects in the event (H_T), which improves the rejection of background events.

Our analysis complements other recent $t\bar{t}$ cross section determinations at CDF II using dilepton events [5], using lepton plus jets events with b-tags and a kinematically derived estimate of the b-tagged backgrounds [6], using lepton plus jets events with b's identified by their semi-leptonic decays [7] or using lepton plus jets events and kinematic discrimination without b-tagging [8]. The work of Reference [6] is particularly relevant to the measurement described here, in that it uses the same b-tagged event sample, but calculates the backgrounds by appealing to a data control sample available only in the W plus jets selection. Our technique for background estimation is significantly more general, and this paper establishes the ability to use b-tagging in many other kinds of measurements at CDF in the future. We comment further on this matter at the end of Section III.

The organization of this paper is as follows. Sec. II reviews the detector systems and event reconstruction techniques relevant to this measurement. The trigger and sample selections are described in Sec. III. The b-tagging algorithm, its efficiency for tagging b jets, and the understanding of its fake rate are discussed in Sec. IV. The means for estimating backgrounds from processes which produce a W in association with heavy flavor are described in Sec. V. In Sec. VI our understanding of mistags and backgrounds is applied to collate a comprehensive estimate of all tagged contributions to the lepton + jets sample, and this estimate is compared with the data. A cross-check of the background estimation, using the Z + jets sample, is presented in Sec. VII. In Sec. VIII an optimization using the total transverse energy in the event to improve the cross section measurement uncertainty is described, along with the acceptance associated with this event selection. The $t\bar{t}$ production cross section measured in events with at least one b-tagged jet is presented in Sec. IX; the result in events with at least two b-tagged jets is presented in Sec. X. The final results are summarized in Sec. XI.

II. EVENT DETECTION AND RECONSTRUCTION

The CDF II detector is described using a cylindrical coordinate system with the z coordinate along the proton direction, the azimuthal angle ϕ, and the polar angle θ usually expressed through the pseudorapidity $\eta = -\ln(\tan(\theta/2))$. The rectangular coordinates x and y point radially outward and vertically upward from the Tevatron ring, respectively. The detector is approximately symmetric in η and ϕ.

A. Charged Particle Tracking

Drift cell and silicon microstrip systems provide charged particle tracking information in the region $|\eta| \leq 1.0$ and $|\eta| \leq 2.0$, respectively. The tracking systems are contained in a 3.2 m diameter, 5 m long superconducting solenoid which produces a 1.4 T magnetic field aligned coaxially with the $p\bar{p}$ beams, allowing measurement of charged particle momentum transverse to the beamline (p_T).

The Central Outer Tracker (COT) is a 3.1 m long open cell drift chamber which performs 96 track measurements in the region between 0.40 and 1.37 m from the beam axis [9]. Sense wires are arranged in 8 alternating axial and $\pm 2^{\circ}$ stereo “superlayers” with 12 wires each. The position resolution of a single drift time measurement is approximately 140 μm.

Charged particle trajectories are found first as a series of approximate line segments in the individual axial superlayers. Two complementary algorithms are used to associate segments lying on a common circle, and the results are merged to yield a final set of axial tracks. Track segments in the stereo superlayers are associated with axial track segments to reconstruct tracks in three dimensions. COT tracks used in this analysis are required to have at least 3 axial and 3 stereo superlayers with 7 hits per superlayer.

The efficiency for finding isolated high momentum tracks is measured using electrons from $W \to e^{\pm}\nu$ which are identified in the central region $|\eta| \leq 1.1$ using only the calorimetric information for the electron shower and the missing transverse energy (see below). In these events, the efficiency for finding the electron track is found to be $99.93^{+0.07}_{-0.35}\%$ [10], and this is typical for high momentum isolated tracks from either electronic or muonic W decays which are contained in the COT. For high-momentum tracks, the transverse momentum resolution is found to be $\delta p_T/p_T \approx 0.1\% \cdot p_T(\mathrm{GeV})$, the track position resolution at the origin is $\delta z \approx 0.5\,\mathrm{cm}$ in the direction along the beamline and the resolution on the track impact parameter, or distance from the beamline at the track's closest approach in the transverse plane, is $\delta d_0 \approx 350\mu\mathrm{m}$.

A road-based hardware pattern recognition algorithm runs online in the eXtremely Fast Tracker (XFT) to provide track information for triggering [11]. Drift times partitioned into two time bins are used to find the axial segments which are matched in their positions and slopes. An "XFT track" is one which has four matching axial segments on a trajectory. The XFT efficiency is measured in a set of well measured COT tracks which pass through all 4 axial superlayers. The XFT is found to have an average efficiency of $96.7 \pm 0.1\%$ for charged particles with momenta greater than $25\,\mathrm{GeV}/c$.

Inside the inner radius of the COT, a five layer double-sided silicon microstrip detector (SVX) covers the region between 2.5 to 11 cm from the beam axis [12]. Three separate SVX barrel modules are juxtaposed along the beamline to cover a length of 96 cm, approximately 90% of the luminous beam intersection region. Three of the five layers combine an $r - \phi$ measurement on one side and a 90° stereo measurement on the other, and the remaining two layers combine $r - \phi$ with small angle stereo at $\pm 1.2°$. The typical silicon hit resolution is $11\,\mu$m. Additional Intermediate Silicon Layers (ISL) at radii between 19 and 30 cm in the central region link tracks in the COT to hits in SVX.

Silicon hit information is added to reconstructed COT tracks using a progressive "Outside-In" (OI) tracking algorithm. COT tracks are extrapolated into the silicon detector, associated silicon hits are found, and the track is refit with the added information of the silicon measurements. The initial track parameters provide a width for a search road in a given layer. Then, for each candidate hit in that layer, the track is refit and used to define the search road into the next layer. The stepwise addition of the precision SVX information at each layer progressively reduces the size of the search road, while also properly accounting for the additional uncertainty due to multiple scattering in each layer. The search uses the two best candidate hits in each layer to generate a small tree of final track candidates, from which the tracks with the best χ^2 are selected. The efficiency for associating at least three silicon hits with an isolated COT track is $91 \pm 1\%$. The extrapolated impact parameter resolution for high momentum OI tracks is $30\,\mu$m, including the uncertainty in the beam position.

B. Calorimetry for Electrons and Jets

Outside of the tracking systems and the solenoid, segmented calorimeters with projective geometry are used to reconstruct electromagnetic (EM) showers and jets in the rapidity interval [13–15]. The EM and hadronic calorimeters are lead-scintillator and iron-scintillator sampling devices, respectively. The calorimeter is segmented into “towers”, each covering a small range of pseudo-rapidity and azimuth; the full array covers 2π azimuth over the pseudo-rapidity range $|\eta| < 3.6$. The transverse energy $E_T = E\sin\theta$ is measured in each calorimeter tower, where the polar angle is calculated using the measured z position of the event vertex. Proportional and scintillating strip detectors measure the transverse profile of EM showers at a depth corresponding to the shower maximum.

High momentum jets, photons, and electrons leave isolated energy deposits in small contiguous groups of calorimeter towers which can be identified and summed together into an energy “cluster.” For the purpose of triggering, online processors organize the calorimeter tower information into separate lists of clusters for the electromagnetic compartments alone and for the electromagnetic and hadronic compartments combined. Electrons are identified in the central electromagnetic calorimeter (CEM) as isolated, mostly electromagnetic clusters which match with an XFT track, in the pseudorapidity range $|\eta| < 1.1$.

The electron transverse energy is reconstructed from the electromagnetic cluster with a precision $\sigma(E_T)/E_T = 13.5\%/\sqrt{E_T/(\mathrm{GeV})} \oplus 2\%$ [16]. Jets are identified as a group of electromagnetic and hadronic calorimeter clusters which fall within a cone of radius $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \leq 0.4$ [17]. Jet energies are corrected for calorimeter non-linearity, losses in the gaps between towers [18], and multiple primary interactions. The jet energy resolution is approximately $(0.1(E_T/(\mathrm{GeV})+1.0)$ GeV [19].

C. Muon Detection and Reconstruction

For this analysis, muons are detected in three separate subdetectors. Directly outside of the calorimeter, four-layer stacks of planar drift chambers (CMU) detect muons with $p_T > 1.4$ GeV/c which penetrate the five absorption lengths of the calorimeter [20]. Farther out, behind another 60 cm of steel, an additional four layers (CMP) detect muons with $p_T > 2.0$ GeV/c [21]. The two systems cover the same part of the central region $|\eta| \leq 0.6$, although the CMU and CMP have different structures and their geometrical coverages do not overlap exactly. Muons between $0.6 \leq |\eta| \leq 1.0$ pass through at least four drift layers lying on a conic section outside of the central calorimeter; this system (CMX) completes the coverage over the full fiducial region of the COT tracker [21]. The presence of a penetrating muon is reconstructed as a line segment or "stub" in one of the four-layer stacks. Muon candidates are then identified as isolated tracks which extrapolate to the stubs. A track which is linked to both CMU and CMP stubs is called a CMUP muon.

D. Beam Positions and the Primary Interaction Vertex

The event selection depends on reconstructing secondary vertices from b hadron decays. The identification of these decay vertices requires a precise measurement of the primary vertex, the point from which all prompt tracks originate. The primary vertex location in a given event can be found by fitting well-measured tracks to a common point of origin.

The locus of all primary vertices defines the "beamline," the position of the luminous region of the beam-beam collisions through the detector. The beamline can be used as a constraint to refine the knowledge of the primary vertex in a given event. The first estimate of the primary vertices (x_V, y_V, z_V) is binned in the z coordinate. A linear fit to (x_V, y_V) *vs.* z_V yields the beamline of each run section.

The luminous region is long, with $\sigma_z = 29$ cm. The transverse cross section is circular, with a width of approximately $30\,\mu$m at $z = 0$, rising to $\approx 50 - 60\,\mu$m at $|z| = 40$ cm. The beam is neither parallel to nor centered in the detector. At $z = 0$, the beamline is at $(x_V, y_V) \approx (-2.0, 3.9)$ mm, and has a slope of $\approx 5.0\,\mu$m/cm in the horizontal plane and

$\approx 1.7\,\mu$m/cm in the vertical plane. These parameters are rather stable, varying from their mean positions by no more than $\approx 20\%$ during periods of continuous data taking.

At high luminosities, more than one collision can occur on a given bunch crossing; the primary vertices of the collision are typically separated in the z coordinate. For the data analyzed here, there are an average of 1.4 reconstructed vertices per event. The z position of each vertex is calculated from the weighted average of the z coordinates of all tracks within 1 cm of a first iteration vertex, with a typical resolution of $100\,\mu$m.

A final determination uses all of the information above to recalculate a best primary vertex in each candidate event for the b-tagging procedure. This precise calculation, using a beam constraint and OI tracks, is described fully in Sec. IV. As part of the lepton + jets event selection, the events are required to have the reconstructed primary vertex located inside the luminous region ($|z| < 60$ cm).

III. DATA SAMPLES AND EVENT SELECTION

A. Colliding Beam Data

The colliding beam data used in this analysis were recorded during the period March 2002 - August 2003, when the instantaneous Tevatron luminosity ranged from 0.5 to 4.0×10^{31} cm^{-2}s^{-1}.

Cherenkov light detectors in the very forward region ($|\eta| \geq 3.7$) record information on the instantaneous and total integrated luminosity of the Tevatron [22]. The total integrated luminosity for this period is 193.5 pb^{-1}; after quality requirements on the silicon tracking, the data sample used for this analysis amounts to 162 pb^{-1} for CEM electrons and CMUP muons, and 150 pb^{-1} for CMX muons.

For the primary data samples used in this analysis, the detector is triggered on high momentum electrons and muons. The electron hardware triggers require an XFT track with $p_T \geq 8$ GeV/c matched to an EM cluster with $E_T \geq 16$ GeV and the ratio of hadronic to electromagnetic energy less than 0.125. The muon hardware triggers require an XFT track with $p_T \geq 8$ GeV/c matched to muon stubs in the joint CMUP configuration or in the CMX. A complete version of the offline lepton selection is performed online in the last stage of triggering, and repeated in offline processing with updated calibration constants. Other secondary datasets described in Sec. IV use a jet trigger with a certain E_T threshold or an electron trigger with relaxed E_T requirements.

B. Monte Carlo Samples

The understanding of acceptances, efficiencies, and backgrounds relies on detailed simulation of physics processes and the detector response. Most measurements of acceptance and efficiency rely on PYTHIA v6.2 [23] or HERWIG v6.4 [24, 25]. These generators employ leading order matrix elements for the hard parton scattering, followed by parton showering to simulate gluon radiation and fragmentation. Each generator is used in conjunction with the CTEQ5L parton distribution functions [26]. For heavy flavor jets, we interface to QQ v9.1 [27] to provide proper modeling of b and c hadron decays.

The estimate of the b-tagging backgrounds due to higher order QCD processes such as $Wb\bar{b}$ requires special care. This study of backgrounds in the b-tagged sample uses the ALPGEN program [28], which generates high multiplicity partonic final states using exact leading-order matrix elements. The parton level events are then passed to HERWIG and QQ for parton showering and b and c hadron decay. Further discussion of ALPGEN can be found in Sec. V.

The CDF II detector simulation reproduces the response of the detector to particles produced in $p\bar{p}$ collisions. The same detector geometry database is used in both the simulation and the reconstruction, and tracking of particles through matter is performed with GEANT3 [29]. Charge deposition in the silicon detectors is calculated using a simple geometrical model based on the path length of the ionizing particle and an unrestricted Landau distribution. The drift model for the COT uses a parameterization of a GARFIELD simulation, with the parameters tuned to match COT data [9]. The calorimeter simulation uses the GFLASH [30] parameterization package interfaced with GEANT3. The GFLASH parameters are tuned to test beam data for electrons and high-p_T pions, and they are checked by comparing the calorimeter energy of isolated tracks in the collision data to their momenta as measured in the COT. Further detail on the CDF II simulation can be found elsewhere [31].

C. W + Jets Selection

The selection identifies events consistent with the W + jets signature containing a high-momentum electron or muon (hereafter referred to as "lepton," ℓ), large missing transverse energy, and hadronic jets. Parts of this selection, especially the lepton identification and missing transverse energy calculations, are shared with other $t\bar{t}$ cross section analyses (see in particular Ref. [8]). The event selection is summarized below.

The offline electron selection requires an EM cluster with $E_T \geq 20$ GeV matched to a track with $p_T \geq 10\,\mathrm{GeV}/c$. The cluster is required to have an electromagnetic fraction and shower shape consistent with an electron deposit. The extrapolated track is required to match the shower location as measured in the shower maximum strip detector, and to have a momentum consistent with the shower energy. Finally, since the electron from W decay is expected to be isolated from other energy deposits in the calorimeter, the energy in a cone of radius $\Delta R = 0.4$ around the electron cluster, but not including the cluster itself, is measured, and the isolation ratio of the energy in the cone to the energy of the electron is required to be less than 0.1.

Photon conversions in the detector material are a source of electron backgrounds. A conversion is defined as a pair of tracks (one of them the electron) satisfying the following cuts:

- oppositely charged,
- $|\Delta(xy)| < 2\,\mathrm{mm}$, and
- $|\Delta(\cot\theta)| < 0.04$,

where $\Delta(xy)$ is the distance between the tracks in the $r - \phi$ plane at the point where they are parallel in that plane, and $\Delta(\cot\theta)$ is the difference between the cotangents of the polar angles of the two tracks. Electrons that are part of an identified conversion pair are not considered further in the electron selection.

The offline muon selection requires a COT track with $p_T \geq 20\,\mathrm{GeV}/c$ matched to a CMUP or CMX muon stub. The matching is based on the extrapolated track position at the chambers, accounting for the effects of multiple scattering. The energy in the calorimeter tower containing the muon is required to be consistent with the deposition expected from a minimum ionizing particle. Backgrounds from cosmic rays are removed by requiring that the track extrapolates to the origin, and that the minimum ionizing tower energy deposit is within a narrow timing window around the beam crossing.

In these high momentum lepton samples, the signal of the neutrino from $W \to \ell\nu$ is large missing transverse energy, $\not\!\!E_T$. The $\not\!\!E_T$ is calculated as the vector sum of the energy in each calorimeter tower multiplied by the azimuthal direction of the tower. If isolated high momentum muons are found in the event, the $\not\!\!E_T$ is corrected by subtracting the muon energy in the calorimeter and adding the muon p_T to the vector sum. The selection finally requires $\not\!\!E_T \geq 20\,\mathrm{GeV}$.

In addition to the direct $t \to e\nu_e b$ and $t \to \mu\nu_\mu b$ modes, this event selection has a small acceptance for top final states with $W \to \tau\nu$ and a subsequent leptonic τ decay, or with high momentum semi-leptonic b quark decays. These are included in the signal acceptances calculated in Sec. IX.

Z bosons and top dilepton decays that contribute to the inclusive high p_T lepton dataset are removed by flagging the presence of a second lepton. Any event with two leptons satisfying the lepton identification is removed, as well as those events where the second lepton is an electron in the plug calorimeter or a muon that fails the CMUP requirement, but has one CMU or CMP muon segment. Finally, we attempt to remove Z bosons without a well identified second lepton by eliminating events with one lepton and certain second objects which form an invariant mass between 76 and 106 GeV/c^2 with the primary lepton. For primary muons the other object is an opposite-signed isolated track with $p_T > 10$ GeV/c. For primary electrons the second object may be such a track, an electromagnetic cluster, or a jet with $E_T > 15$ GeV and $|\eta| \leq 2.0$ that has fewer than three tracks and an electromagnetic energy fraction greater than 95%. The correction for the residual Z boson contribution to the W + jets sample is described in Section VII. Small contributions from $Z \to \tau\bar{\tau}$ where a τ is tagged are treated as a separate background, and described in Section VI.

The number of jets produced in associated with the leptonically decaying W in the event is measured by selecting jets of cone radius $\Delta R = 0.4$, with $E_T \geq 15\,\mathrm{GeV}$ and $|\eta| \leq 2.0$. The jets are clustered after removing towers associated with the selected isolated electron from the leptonic W decay, and after correcting the tower E_T for the location of the primary vertex z coordinate. The number of events in each jet multiplicity bin is shown in Table I. The overall acceptance $\times$ efficiency of this selection for $t\bar{t}$ events in the lepton+jets channel with three or more jets, including the leptonic branching ratios, is roughly 4% for the electron

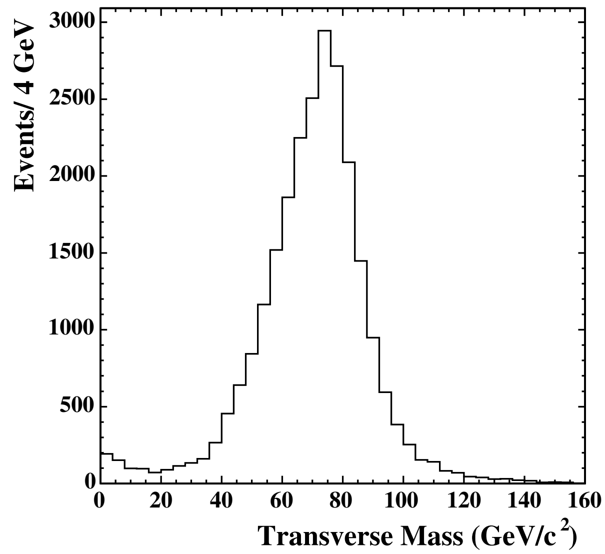

FIG. 1: Transverse mass of the identified lepton and inferred neutrino, consistent with W boson production (162 pb^{-1} data sample).

TABLE I: Number of events selected, before b-tagging, for each jet multiplicity.

	W + 1 jet	W + 2 jets	W + 3 jets	W + 4 jets	W + 3 jets	W + 4 jets
			$H_T > 0$		$H_T > 200$ GeV	
Electrons	8828	1446	241	70	117	63
Muons	6486	1002	146	37	63	28

channel, 2% for muons in the CMUP, and 1% for muons in the CMX.

The presence of the W boson in the selected events is verified by calculating the transverse mass of the lepton and the missing energy: $M_T = \sqrt{(E_T(\ell) + E_T(\nu))^2 - (\overrightarrow{P_T}(\ell) + \overrightarrow{P_T}(\nu))^2}$. The distribution of this variable for all events passing the requirement of a lepton, missing energy, and at least one jet is shown in Fig. 1, and displays the Jacobian edge associated with W production and decay.

As a final optimization step, the selection will incorporate an additional cut on the total transverse energy H_T of all objects in the event. Events from $t\bar{t}$ production have, on average, a significantly greater total transverse energy than background events. The optimization of this requirement and acceptance corrections and uncertainties will be discussed in Sec. VIII.

Because the $t\bar{t}$ signal is expected to contribute significantly to the sample of events with W + 3 jets or W + $\geq$ 4 jets, an excess of observed events over the expected background with those jet multiplicities is assumed to be entirely due to $t\bar{t}$ production. The observed results for events with W + 1 jet or W + 2 jets, where the $t\bar{t}$ contribution is negligible, serve as a check of the background prediction.

In Reference [6], the momentum spectrum of the leading jets in the W + 1 and 2 jet events is shown to be a reasonable model of the backgrounds in the W + 3 or 4 jet events, and is used in deriving a completely independent estimate of the b-tag backgrounds to top production in the W + 3 or 4 jet channels. The estimated background, 18 ± 4 events, is in good agreement with our overall estimate of 23 ± 3 (for $H_T > 0$), derived from an explicit calculation for each contributing background process (see Sec.VI). The independence of these background estimates allows for a combined cross section calculation which will appear in a future paper. We note here that the technique of Reference [6] will work only in the study of top quarks in the W plus jet mode, and we consider it a cross-check on our more general technique for calculating b-tag backgrounds, which will be employed in other b-tagging analyses at CDF.

The final cross section calculation, $\sigma_{t\bar{t}} = (N_{\rm obs} - N_{\rm bkg})/(\epsilon_{t\bar{t}} \times \mathcal{L})$, depends on the product $\epsilon_{t\bar{t}}$ of signal acceptance and selection efficiency, the expected number of non-$t\bar{t}$ background events $N_{\rm bkg}$, and the integrated luminosity $\mathcal{L}$.

IV. SECONDARY VERTEX b-TAGGING

In this Section we describe and discuss the performance of an algorithm to identify jets resulting from heavy quark (b,c) fragmentation. This "SecVtx" algorithm is based upon the algorithm used to discover the top quark [4]. Most of the non-$t\bar{t}$ processes found in the W + jets sample do not contain heavy quarks in the final state. Requiring that one or more of the jets in the event be tagged by SecVtx keeps more than half of the $t\bar{t}$ events while removing approximately 95% of the background.

A. Description of the SecVtx Algorithm

The SecVtx algorithm relies on the displacement of secondary vertices relative to the primary event vertex to identify b hadron decays. The Run II algorithm is essentially unchanged from Run I [4], but the track selection cuts have been re-tuned for the CDF II detector.

In order to select displaced tracks coming from decays of long-lived hadrons, precise knowledge of the collision point is necessary. To find an event-by-event primary vertex, we first identify which of the vertices described in Section II is nearest the identified high-momentum electron or muon. For other datasets without high-momentum leptons, we use the vertex which has the highest total scalar sum of transverse momentum of associated tracks. The position of the primary vertex is then determined by fitting together the tracks within a $\pm 1\,\mathrm{cm}$ window in z around this vertex. The procedure starts by fitting a vertex using all tracks within the z window and with impact parameter significance (relative to the beamline) $|d_0/\sigma_{d_0}| < 3$, where σ_{d_0} includes the uncertainty on both the track and the beamline positions. The transverse profile of the beamline at the z of the original vertex estimate is also used as a constraint in the fit. A pruning stage removes tracks which contribute $\chi^2 > 10$ to the fit (or the track with the largest χ^2 contribution if the total fit reduced chi-squared per degree of freedom $\chi^2/\mathrm{ndf} > 5$). After the initial pruning, the fit is repeated using only the remaining tracks until a vertex with no tracks over the χ^2 cut is found. If no tracks survive the pruning stage then the beamline profile is used for the primary vertex position estimate. In the event sample used for these results the uncertainty in the fitted transverse position ranges from $10 - 32\,\mu\mathrm{m}$ depending upon the number of reconstructed tracks and the topology of the event.

Secondary vertex tagging operates on a per-jet basis, where only tracks within the jet cone are considered for each jet in the event. A set of cuts involving the transverse momentum, the number of silicon hits attached to the tracks, the quality of those hits, and the χ^2/ndf of the final track fit are applied to reject poorly reconstructed tracks. Only jets with at least two of these good tracks can produce a displaced vertex; a jet is defined as "taggable" if it has two good tracks. Displaced tracks in the jet are selected based on the significance of their impact parameter with respect to the primary vertex and are used as input to the SecVtx algorithm. SecVtx uses a two-pass approach to find secondary vertices. In the first pass, using tracks with $p_T > 0.5$ GeV/c and $|d_0/\sigma_{d_0}| > 2.5$, it attempts to reconstruct a secondary vertex which includes at least three tracks (at least one of the tracks must have $p_T > 1\,\mathrm{GeV}/c$). If the first pass is unsuccessful, it performs a second pass which makes tighter track requirements ($p_T > 1\,\mathrm{GeV}/c$ and $|d_0/\sigma_{d_0}| > 3$) and attempts to reconstruct a two-track vertex (one track must have $p_T > 1.5\,\mathrm{GeV}/c$).

Once a secondary vertex is found in a jet, the two-dimensional decay length of the secondary vertex L_{2D} is calculated as the projection onto the jet axis, in the $r-\phi$ view only, of the vector pointing from the primary vertex to the secondary vertex. The sign of L_{2D} is defined relative to the jet direction, specifically by the absolute difference $|\phi|$ between the jet axis and the secondary vertex vector (positive for $< 90^\circ$, negative for $> 90^\circ$). Secondary vertices corresponding to the decay of b and c hadrons are expected to have large positive L_{2D} while the secondary vertices from random mis-measured tracks are expected to be less displaced from the primary vertex. To reduce the background from the false secondary vertices (mistags), a good secondary vertex is required to have $L_{2D}/\sigma_{L_{2D}} > 3$ (positive tag) or $L_{2D}/\sigma_{L_{2D}} < -3$ (negative tag), where $\sigma_{L_{2D}}$, the total estimated uncertainty on L_{2D} including the error on the primary vertex, is estimated vertex-by-vertex but is typically $190\,\mu$m. The negative tags are useful for calculating the false positive tag rate, as detailed in Section VI B. A tagged jet is defined to be a jet containing a good secondary vertex (the SecVtx algorithm will find at most one good vertex per jet).

B. Measurement of Tagging Efficiency

The results described in this paper require a knowledge of the tagging efficiency for $t\bar{t}$ events, *i.e.*, how often at least one of the jets in a $t\bar{t}$ event is positively tagged by SecVtx. Because it is not possible to measure this directly in $t\bar{t}$ events we have adopted a different strategy. A sample of jets whose heavy flavor fraction can be measured is used to derive the per-jet tagging efficiency in the data for that sample. Because the jets in $t\bar{t}$ events will in general have different energies, pseudorapidities, and track multiplicity than the jets in the calibration sample, the measured efficiency cannot be used directly. Instead, a matching sample of Monte Carlo jets is used to determine the tagging efficiency in the simulation for jets like those in the calibration sample, and the ratio of efficiencies between data and simulation (scale factor) is then used to correct the tagging efficiency in $t\bar{t}$ Monte Carlo samples. In other words, the geometrical acceptance and energy dependence of the tagger are taken from the simulation, with the overall normalization determined from the data.

To measure the efficiency for tagging heavy flavor hadrons, we use a sample of low-p_T inclusive electron data which is enriched in semileptonic decays of bottom and charm hadrons. For the matching Monte Carlo sample we use the HERWIG [24] program to generate $2 \rightarrow 2$ parton events, which are passed through a filter requiring an electron with $p_T > 7\,\mathrm{GeV}/c$ and $|\eta| < 1.3$. Events passing this filter are processed using the detector simulation described in Section II.

Electrons in the events are identified using the selection in Section III, except with lower thresholds $E_T > 9\,\mathrm{GeV}$ and track $p_T > 8\,\mathrm{GeV}/c$. Further differences from Section III are that the electrons are required to be non-isolated and conversions are not removed. The electron track must also pass through every layer of the SVX detector.

Along with the electron we require two jets, the "electron jet" and the "away jet." The electron jet is required to have $E_T > 15$ GeV (including the energy of the electron) and to be within 0.4 of the electron in $\eta - \phi$ space (in other words the electron is within the jet cone), and is presumed to contain the decay products of a heavy flavor hadron. The away jet is required to have $E_T > 15$ GeV and $|\eta| < 1.5$, and it must be approximately back-to-back with the electron jet ($\Delta\phi > 2\,\mathrm{rad}$). A total of 481,301 events of the data sample pass these event selection requirements. Figures 2 and 3 show that the Monte Carlo is an adequate representation of the data sample for relevant event selection and tagging variables. The differences can be attributed to the presence of fake electrons in the data which are not completely removed even after requiring a SecVtx tag, and which are not present in the Monte Carlo due to the generator-level electron filter. The discrepancy in Figure 3 on the negative side of the pseudo-$c\tau$ plot shows that the Monte Carlo underestimates the mistag rate observed in the data.

In order to measure the tagging efficiency for electron jets, it is first necessary to characterize their heavy flavor content. Two methods are used to measure the fraction F_b of electron jets which contain a b hadron. The first method is to reconstruct $D^0 \rightarrow K^-\pi^+$ decays within the electron jet and use the invariant mass sidebands to subtract background; this method yields $F_b = 0.139 \pm 0.021$. The second method involves searching for secondary muons within the electron jet resulting from cascade c decays using the same-sign rate to estimate the background; this method gives $F_b = 0.228 \pm 0.037$. Because the agreement is only at the 2σ level, the uncertainty on the weighted average is inflated by 2.09 based on the χ^2 of the two determinations. The combined result of the two measurements is $F_b = 0.161 \pm 0.038$.

The fraction F_c of electron jets which came from a charm quark also contributes to the

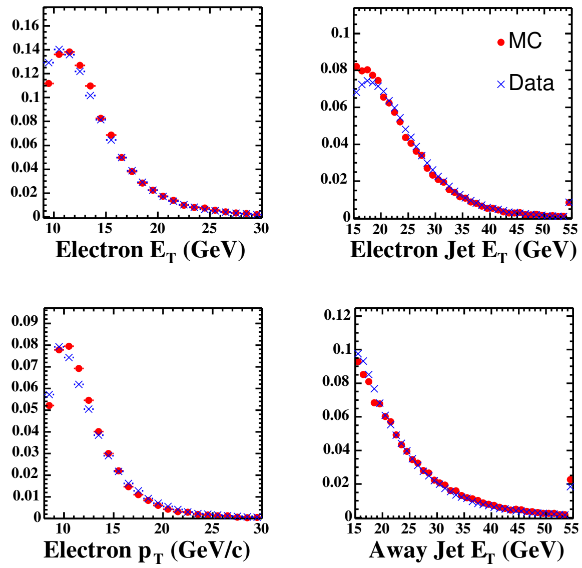

FIG. 2: Data/Monte Carlo comparison of some quantities of tagged electron jets ($L_{2D} > 0$, identified conversions have been removed for plotting purposes). Histograms are normalized to unit area. From top-left, clockwise: electron E_T, electron-jet E_T, away-jet E_T, electron p_T. (The last bin includes all overflow entries.)

total heavy flavor fraction $F_{\mathrm{HF}} = F_b + F_c$. An estimate of the amount of c relative to b in the electron jet is derived from a fit to the invariant mass spectrum of the tracks in the positive tags found in the electron jets. Templates for b, c, and light-flavor jets taken from the Monte Carlo (and also from the data for light-flavor) were fitted to the distribution, as shown in Figure 4, to obtain the ratio of c to b after requiring a positive tag. The result of this fit is $F_c^{\mathrm{tag}}/F_b^{\mathrm{tag}} = 0.118 \pm 0.017$, where the uncertainty is dominated by the systematic error due to varying the light-flavor template. A value for F_c/F_b before any tagging is obtained by multiplying this result by the ratio of tagging efficiencies $\epsilon_b/\epsilon_c = 5.2 \pm 0.4$ predicted by the Monte Carlo, resulting in $F_c/F_b = 0.61 \pm 0.10$. Therefore the total heavy flavor fraction of electron jets F_{HF} is 0.259 ± 0.064.

To measure the tagging efficiency of the heavy flavor electron jets we employ a double-tag technique, requiring that the away jet be tagged by SecVtx. This enhances the heavy flavor fraction of the electron jets and reduces the dependence on F_{HF}, which we were only able to constrain at the 25% level. Another benefit of the double-tag is to reduce the influence of the charm component, so that the resulting heavy flavor tag efficiency is more representative

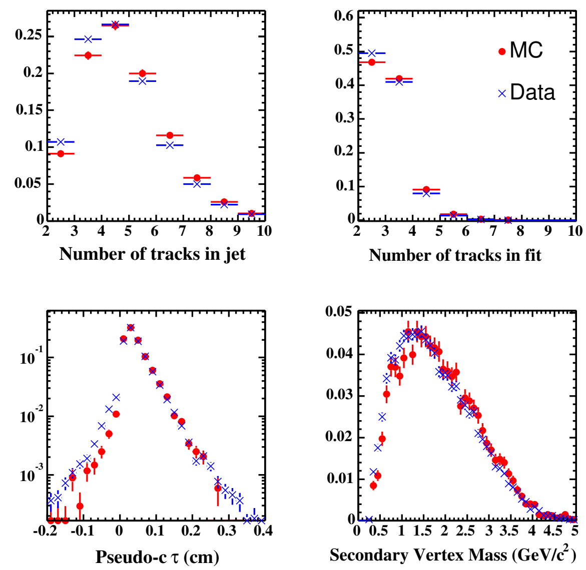

FIG. 3: Data/Monte Carlo comparison of some quantities of tagged electron jets (identified conversions have been removed for plotting purposes). Histograms are normalized to unit area. From top-left, clockwise: number of good tracks in the jet, number of tracks in the tagged vertex, vertex mass of positively tagged electron-jets; pseudo-$c\tau$ of (positively or negatively) tagged electron-jets.

of the b-tagging efficiency. Tagging the away jet reduces the charm from 61% of the bottom component down to around 10%. The difference in the tag efficiency for semileptonic decays, which we measure, and generic heavy hadron decays is used later to estimate a systematic error.

The tagging efficiency for heavy flavor jets containing an electron, derived from the numbers of double- and single-tags, is

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

where $N_{\mathrm{a}+}$ and $N_{\mathrm{a}-}$ are the numbers of positive and negative tagged away jets, and $N^{\mathrm{e}+}_{\mathrm{a}+}$, for example, is the number of events where both electron and away jet are positive tagged.

The factor $F^{\mathrm{a}}_{\mathrm{HF}}$ is the fraction of electron jets containing heavy flavor for events where the away jet is tagged. This number is less than one due to events where the away jet is mistagged or contains heavy flavor due to gluon splitting or flavor excitation, and the electron is either a fake or part of a photon conversion pair. In order to estimate these

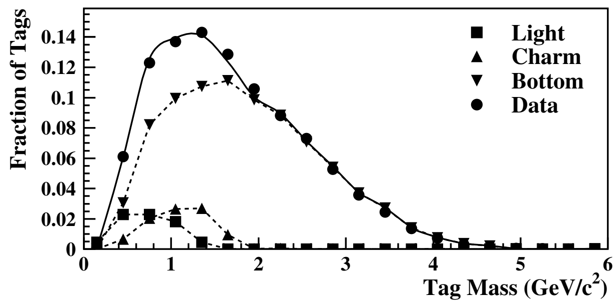

FIG. 4: Fit (solid line) of the relative b and c contributions to the vertex tag mass distribution. Templates for the different flavors are derived from simulation (and the data in the case of light flavor). The error bars for the data are contained within the markers.

effects we use identified conversions (see Section III) to probe the light flavor composition of the electron jets. In this way we write $F_{\rm HF}^{\rm a}$ as

$$F_{\rm HF}^{\rm a} = 1 - \frac{\left(\frac{N_{\rm c}^{\rm a+} - N_{\rm c}^{\rm a-}}{N_{\rm a+} - N_{\rm a-}} - \epsilon_c'\right)}{\frac{N_{\rm c}}{N} - \epsilon_c'} \cdot (1 - F_{\rm HF}), \tag{2}$$

where N is the number of events passing the selection, $\epsilon_c' = \frac{N_{\rm c}^{\rm e+} - N_{\rm c}^{\rm e-}}{N_{\rm e+} - N_{\rm e-}}$, and the c subscript refers to events where the electron was identified as a conversion. A full derivation of this expression can be found in the Appendix.

To illustrate the effectiveness of the conversion finder, Figure 5 shows the estimated radius of the conversion point for identified pairs. Peaks corresponding to known detector structures are clearly visible.

We use the $F_{\rm HF}$ value derived above for data, a value of $F_{\rm HF}^{\rm MC} = 0.861$ for the Monte Carlo (found by counting jets which are matched to a heavy quark), and Equations 1 and 2 (see Appendix) to calculate the efficiencies to tag a heavy flavor jet containing an electron in data and Monte Carlo. The resulting values averaged over jet E_T are given in Table II. The efficiencies as a function of the E_T of the jet are shown in Figure 6. The ratio of data to Monte Carlo efficiencies (scale factor) is also shown as a function of E_T. Additionally, a sample of jet data with one jet having $E_T > 50$ GeV and a corresponding 2→2 PYTHIA Monte Carlo sample have been used to determine that the ratio of jet tag rates is flat over a wider jet E_T range than that spanned by the electron calibration sample. These samples are also used to estimate a systematic uncertainty for extrapolating the scale factor to the higher-E_T jets (typically 40-120 GeV) characteristic of top quark decays.

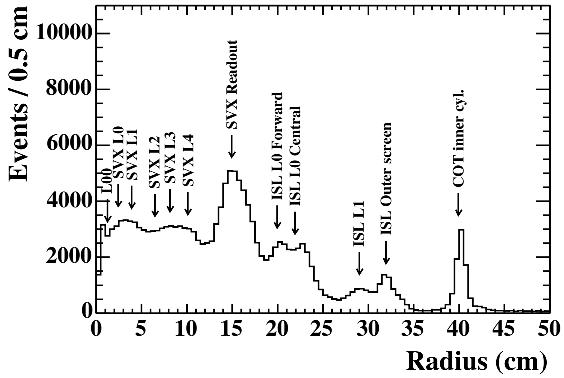

FIG. 5: Radius of identified conversions in data, with location of the silicon detector layers (L00, SVX and ISL), readout system, and ISL and COT main mechanical structures.

TABLE II: Efficiency to tag a heavy flavor electron jet in data and Monte Carlo, and the data/MC ratio (scale factor). Uncertainties on the efficiencies are statistical only; systematic uncertainties on the scale factor are summarized in Table III.

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

Several sources of systematic uncertainty have been considered and are summarized in Table III. The $F^{\mathrm{a}}_{\mathrm{HF}}$ method uncertainty accounts for assumptions made in the calculation of $F^{\mathrm{a}}_{\mathrm{HF}}$ about the tagging efficiency of heavy flavor electron jets containing a conversion electron pair. The mistag subtraction uncertainty is due to the asymmetry in negative tags *vs.* fake positive tags described in the next subsection. E_T dependence was described earlier, and the B-decay uncertainty allows for a possible difference in the scale factor due to the lower charged particle multiplicity of semileptonic B decays compared to all possible decay modes. Combining all systematic and statistical errors we obtain a data to Monte Carlo tagging efficiency scale factor of 0.82 ± 0.06.

A variation of the double-tag technique has also been studied which uses the single-tag rate of electron jets rather than the measurements of F_{HF}. First we write the efficiencies in the data as $\epsilon = SF \times \epsilon_{\mathrm{MC}}$ and $\epsilon^{\mathrm{single}} = SF \times \epsilon^{\mathrm{single}}_{\mathrm{MC}}$, where ϵ is defined in Equation 1 and

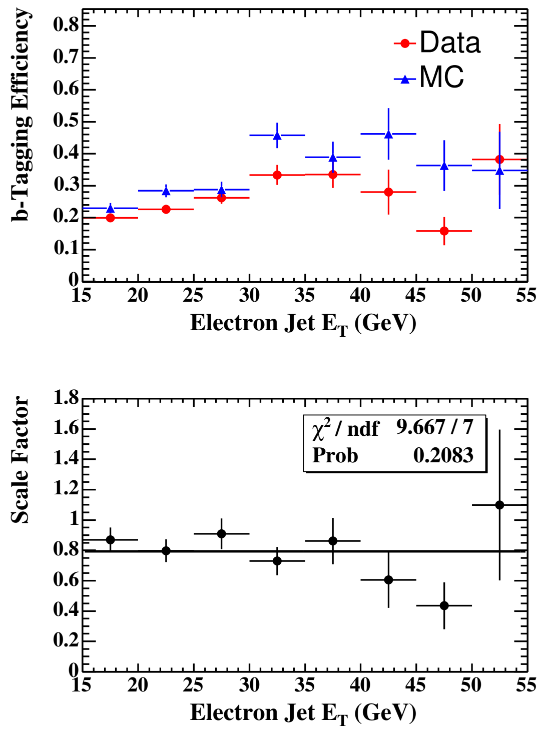

FIG. 6: Efficiency to tag a b-jet as a function of jet E_T in data and Monte Carlo (top), and data/MC scale factor (bottom). The scale factor is consistent with being constant over the E_T range investigated.

TABLE III: Relative uncertainties on the data to Monte Carlo tagging efficiency scale factor, in percent.

Source	uncertainty (%)
F_{HF}	3.5
$F^{\mathrm{a}}_{\mathrm{HF}}$ method	3.0
mistag subtraction	3.0
E_T dependence	2.5
B-decay	1.2
total systematic error	6.2
data statistics	3.2
MC statistics	3.6
total uncertainty	7.8

$\epsilon^{\mathrm{single}} = (N^{e+} - N^{e-})/(F_{\mathrm{HF}}N)$ is the net single-tag efficiency for heavy-flavor electron jets. Although both ϵ and $\epsilon^{\mathrm{single}}$ are tag efficiencies for heavy-flavor jets, they generally differ because the requirement of an away jet tag suppresses the charm content of the sample relative to bottom.

Substituting for ϵ and F_{HF} (using the relation between F_{HF} and $SF \times \epsilon_{\mathrm{MC}}^{\mathrm{single}}$) into Equations A.15 and A.16 allows solution for the efficiency scale factor SF directly in terms of the data tag and conversion rates, and of the MC tag efficiencies ϵ_{MC} and $\epsilon_{\mathrm{MC}}^{\mathrm{single}}$. A result of $SF = 0.81$ is obtained, consistent with the method described above and with similar systematic and statistical errors.

C. Measurement of the Mistag Rate

A "mistag" is defined to be a jet which did not result from the fragmentation of a heavy quark, yet has a SecVtx secondary vertex. Mistags are caused mostly by random overlap of tracks which are displaced from the primary vertex due to tracking errors, although there are contributions from K_S and Λ decays and nuclear interactions with the detector material (the beampipe or the inner silicon layers) as well. Contributions from these effects are measured directly from jet data samples without relying on the detector simulation.

Because the SecVtx algorithm is symmetric in its treatment of d_0 and L_{2D} significance, the tracking-related mistags should occur at the same rate for $L_{2D} > 0$ and $L_{2D} < 0$. Therefore, a good estimate of the positive mistag rate due to resolution effects can be obtained from the negative tag rate. However, some of the negative tags occur in jets which do contain heavy flavor, so that part must be subtracted. In addition, the negative rate will not reflect the mistags due to lifetime or interactions with the detector material. Corrections for all of these effects are determined using fits to the pseudo-$c\tau$ spectra of tagged vertices, described in Section V C. The sum of these corrections is found to be $20 \pm 10\%$ of the negative tag rate, consisting of a subtraction of 20% for removal of the heavy flavor negative tags, and an addition of 40% to account for the mistags due to lifetime and material interactions.

The rate of negative tags for taggable jets is measured in an inclusive sample of jet triggers. The rate is parameterized as a function of four jet variables – E_T, track multiplicity, η, and ϕ – and one event variable ΣE_T, the scalar summed E_T of all jets in the event with $E_T > 10$ GeV and $|\eta| < 2.4$. These parameterized rates are used to obtain the probability that a given jet will be negatively tagged.

The full five-dimensional tag rate matrix was determined using inclusive 20 GeV, 50 GeV, 70 GeV, and 100 GeV jet trigger samples, for a total of 11.5 million events. Figure 7 shows the negative tag rate per taggable jet as a function of jet E_T and track multiplicity (and integrated over the other variables) for all the events in the inclusive jet sample. These rates have not been scaled by the 1.2 ± 0.1 correction discussed above which is applied to convert to an estimate of the positive mistag rate.

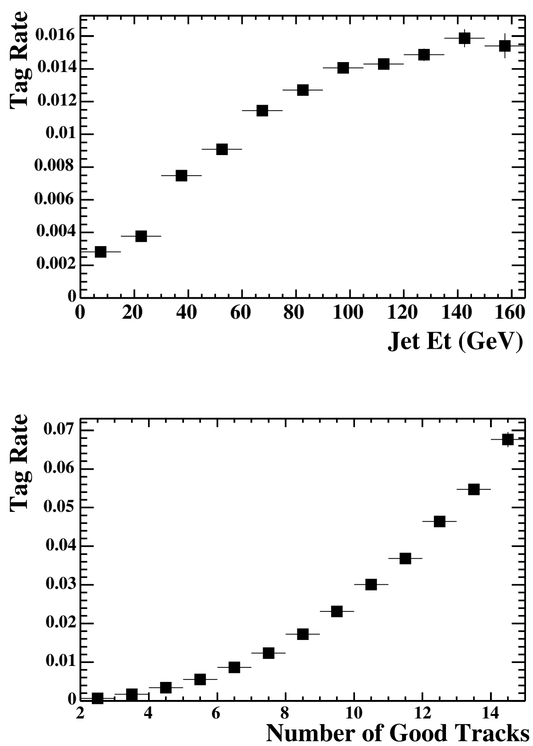

FIG. 7: SecVtx negative tag rate as a function of jet E_T and track multiplicity in the inclusive jet data.

Detailed cross checks were performed on the tag rate matrix to verify its self-consistency and to check predictability and sample dependence. Both the total tag rates and the tag rates as functions of various quantities were used to check how well the matrix predicts the observed data and to estimate systematic errors. Table IV summarizes the differences between the matrix predictions and the observed tag rates in various validation samples. The four jet trigger samples described above were used, along with an independently-triggered sample requiring four jets with $E_T > 15\,\mathrm{GeV}$ and $\Sigma E_T > 125\,\mathrm{GeV}$, referred to as the "SumEt" sample. The table is divided into two sections. Each row in the table compares the tag rate predicted from one sample with the observed rate in a second, different sample.

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

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

The differences in the tag rates of trigger jets and non-trigger jets are well predicted by the matrix. This is mostly due to the inclusion of the jet E_T, η, and ϕ into the matrix binning. The remaining residual difference is taken as a systematic error in the final result.

The systematic uncertainties assigned to the tag rate matrix predictions are summarized in Table V. We assume that the various contributions are uncorrelated and add them in quadrature to find a total systematic uncertainty of 8% on the negative tag rates, which combined with the uncertainty on the correction factor 1.2 ± 0.1 yields a total mistag rate relative uncertainty of 11%.

V. HEAVY FLAVOR CONTRIBUTIONS TO W+JETS

Heavy flavor production in association with a vector boson (*e.g.* $Wb\bar{b}$, $Wc\bar{c}$, Wc) contributes significantly to the non-$t\bar{t}$ background in the b-tagged lepton + jets sample, even though W + light flavor jet production dominates the pretag sample. Several Monte Carlo

TABLE V: Systematic uncertainties assigned to the negative tag rate matrix predictions.

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

generators are capable of performing matrix element calculations for W/Z + jets, even to high jet multiplicity, but these generators use leading-order calculations. As a result, the overall normalization of these calculations has a large theoretical uncertainty, even though the relative contributions of the important diagrams are well-defined.

For this reason, the relative fraction of W + heavy flavor production is calculated in a matrix element Monte Carlo program, and the overall normalization of the W + jets production is measured with collider data. The two results can be combined to estimate the W + heavy flavor background.

For this analysis, we use a new event generator, ALPGEN [28], which calculates exact matrix elements at leading order for a large set of parton level processes in QCD and electroweak interactions. All heavy quark masses, spins and color flows are treated properly inside ALPGEN. Heavy flavor fractions calculated using ALPGEN are calibrated against fractions measured from jet data.

The total W + heavy flavor contribution is estimated by multiplying the number of pretag W + jets events in data, given in Table I, by the calculated W + heavy flavor fraction and the tagging efficiency in Monte Carlo (including the SecVtx efficiency scale factor between data and Monte Carlo). Because the event tagging efficiency depends on the number of heavy flavor jets in the fiducial region $|\eta| < 2.4$, we calculate results separately for the case of 1 and 2 heavy flavor jets.

A. Heavy Flavor Monte Carlo Samples

Parton-level events from the ALPGEN matrix element calculation are fed to the HERWIG parton shower program which generates additional jets from gluon radiation. The matrix element gives a good description of the production of a few, widely separated partons, whereas parton showers are better suited to model the emission of soft collinear gluons. Following a matrix element calculation with a parton showering algorithm provides a better model of the data than does either approach separately.

One outstanding issue for such a combined approach is how to avoid double counting in the region of phase space populated both by higher order matrix elements and the parton shower. Specifically, the radiation from the parton shower in a $W + n$ parton Monte Carlo sample can produce jets which cover the part of the phase space also described by the $W + (n + 1)$ parton Monte Carlo. Although a rigorous combination prescription has been proposed to avoid such double counting, it has not yet been fully implemented in any of the matrix element Monte Carlo programs [32, 33].

A simple procedure deals with the possible double counting by matching final state partons to reconstructed jets and rejecting events where the showering algorithm has produced a hard parton [34, 35]. Events are rejected if there are extra jets which fail to match to the light partons generated at the matrix element level or if there are missing jets. In the special case of heavy flavor partons, the strict matching criteria are relaxed because two partons may be merged into one jet due to the parton mass. Although it minimizes double counting of generated events, this procedure introduces a new type of systematic uncertainty which depends on the matching criteria and the jet definition.

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

The following requirements reduce event double counting after the parton shower: 1) reject events in which an extra jet failed to match any parton from the matrix element calculation, 2) ignore matching requirement for heavy flavor partons because the effect of their masses has been included in the matrix element calculation, and 3) keep only the events which pass the strict jet-light parton matching.

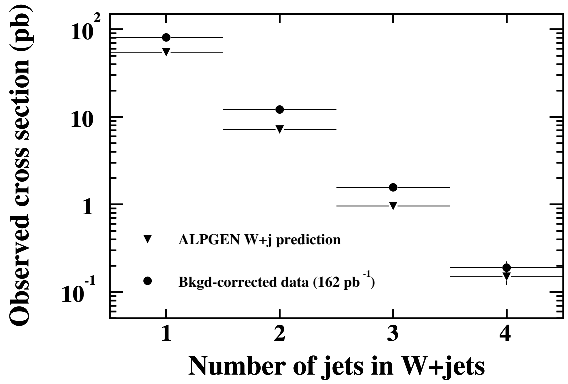

FIG. 8: Observed W + jets cross section compared with the ALPGEN W + jets prediction as a function of number of jets. (Only statistical errors are shown, and the results are not corrected for acceptance.)

Fully exclusive matched events in each matrix element Monte Carlo sample are summed, weighting by the appropriate cross sections. Because the double-counted events have been removed by the matching procedure, this combined sample should reproduce the W + jets data. These results are stable in terms of different matching algorithms, cone size, and jet E_T requirement. The predicted W + jets cross section, without any acceptance correction, is plotted in Figure 8 with the measurement in the electron and muon channels. The non-W and diboson backgrounds as well as the expected contribution from $t\bar{t}$ production are subtracted for this measurement. Even though the overall normalization of the Monte Carlo does not reproduce the data very well, the jet multiplicity dependences in data and Monte Carlo are in good agreement.

B. Heavy Flavor Fraction in Simulated W + jets Events

The heavy flavor fractions for W + jets events, computed using an ALPGEN/HERWIG Monte Carlo sample, are defined to be the ratio of the observed W + heavy flavor and W + jets cross sections.

The matching algorithm operates with particle-level jets, but jets from a full calorimeter simulation provide better agreement with jets in data. A detector-level jet is required to have $E_T > 15\,\mathrm{GeV}$ and $|\eta| < 2$, and a heavy flavor jet is required to match to any b or c parton inside a cone with $\Delta R = 0.4$.

TABLE VI: Summary of systematic uncertainties in the heavy flavor fraction determination.

Source	Uncertainty		
Fractions	$Wb\bar{b}$	$Wc\bar{c}$	Wc
Matching criteria	15%	15%	10%
Q^2 scale ($2M_W^2$ to $0.5M_W^2$)	4%	4%	5%
PDF	5%	5%	10%
Jet energy scale	5%	5%	10%
ISR/FSR	10%	10%	10%
b, c masses ($4.75, 1.55 \pm 0.3\,\mathrm{GeV}/c^2$)	6%	10%	
Total	21%	22%	21%

A summary of systematic uncertainties inherent in this heavy flavor fraction measurement is presented in Table VI. The matching uncertainty is estimated by recomputing the heavy flavor fraction after varying the matching cone sizes (0.4, 0.7) and E_T from 10 to 15 GeV. We take half of the difference in the 4-jet bin as the matching systematic uncertainty. Uncertainties due to the interaction energy scale Q^2, PDFs, and heavy quark masses are calculated by comparing the ratio of the $Wb\bar{b}$ + 1 parton and W + 3 partons cross sections from ALPGEN and estimating the variation by changing the Q^2 (between $2m_W^2$ and $0.5m_W^2$), parton distribution functions (among the 20 eigenvector pairs from CTEQ6M[36]), and the heavy quark mass (± 0.3 GeV). The relative systematic uncertainties in Table VI are applied to all jet multiplicity bins. The final measured heavy flavor fractions for W + jets events can be found in Table VIII.

C. Calibration of Heavy Flavor Fraction Using Jet Data

With the current data sample and a limited number of SecVtx-tagged W + jets data events, it is difficult to verify the ALPGEN heavy flavor fractions in W + jets events directly with data. Fortunately, an inclusive jet sample, without identified W bosons, is a large related class of events whose production processes are described by Feynman diagrams similar to those of W + jets events. In particular, gluon splitting to heavy quark pairs accounts for part of the heavy flavor production in both samples. The inclusive QCD jet sample can

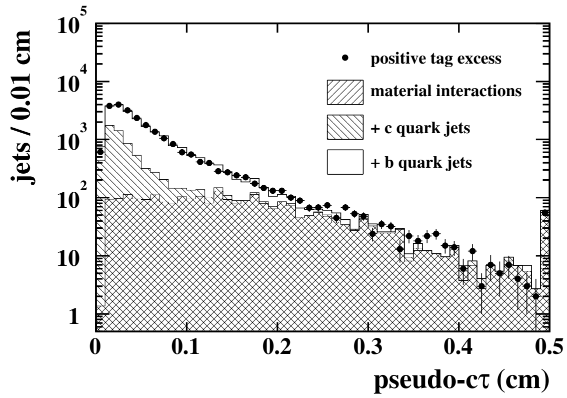

FIG. 9: Pseudo-$c\tau$ distribution for jet data, including fitted contributions for the different components of heavy flavor and secondary interactions in light flavor jets.

be used to compare the heavy flavor fractions calculated in Monte Carlo with results from data. Any discrepancy between heavy flavor fractions in data and Monte Carlo could then be used to adjust the calculated heavy flavor fractions in W + jets events.

Heavy flavor fractions are calculated in both PYTHIA and ALPGEN+HERWIG Monte Carlo jet samples. Events are required to have 2 or 3 jets with $E_T > 15$ GeV and $|\eta| < 2.0$ and at least one jet with $E_T > 20$ GeV to satisfy trigger requirements. Events from the ALPGEN sample must also pass the matching algorithm described in Section V A.

Contributions to the jet data sample from heavy and light partons are determined by fitting the pseudo-$c\tau$ distribution for tagged jets, thereby discriminating between jets from b, c, and light partons or gluons on a statistical basis. Pseudo-$c\tau$ is defined as $L_{\mathrm{2D}} \times M_{\mathrm{vtx}}/p_T^{\mathrm{vtx}}$, where M_{vtx} is the invariant mass of all tracks in the secondary vertex and p_T^{vtx} is the transverse momentum of the secondary vertex four-vector. Even though the L_{2D} distribution is similar for b and c quarks, the pseudo-$c\tau$ is very different for the two flavors.

The fit is made more robust by subtracting the contribution from negative SecVtx tags and fitting the difference only, as shown in Fig. 9. Template distributions of the pseudo-$c\tau$ for b and c jets are derived by matching jets to partons in Monte Carlo, and a separate template is created for secondary interactions in light quark jets, including material interactions and long-lived Λ and K_s^0 particles.

If the signed decay length distribution of secondary vertices in light flavor jets and from tracking combinatorics were symmetric about zero, then the number of fake positive tags

TABLE VII: Fitted contributions from b, c jets and secondary interactions or long-lived light flavor particles in data events. The uncertainties on the b and c fractions are total uncertainties including 5% and 10% uncertainties due to the templates. The ratio $\Delta N/N$ estimates the excess of positive over negative tags in data events, due to secondary interactions and long-lived light flavor particles.

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

from light flavor could be simply estimated by counting the number of negative tags. Unfortunately, secondary vertices from material interactions or long-lived light flavor particles are more likely to have positive decay lengths than negative decay lengths, and there are some real heavy flavor jets with negative decay lengths. The heavy flavor contribution with negative decay lengths is first estimated from Monte Carlo, and then scaled by a factor of 1.6 ± 0.3 to account for a larger overall observed negative tag contribution in data than in Monte Carlo events.

The net excess of secondary interactions on the positive side, ΔN, is computed from the secondary contribution fit results, after subtracting the heavy flavor contributions on the negative side. The resulting average correction factor $\Delta N/N$ needed to scale the number of negative tags to obtain the correct number of fake positive tags is 1.2 ± 0.1.

The heavy flavor fraction as a function of jet E_T is stable, as shown in Table VII, where an uncertainty of 5% (10%) for the b (c) fraction is included due to template uncertainties. These results include the effect of the efficiency scale factor between data and simulation. Measured heavy flavor fractions from the data are consistently 50% higher than the ALPGEN

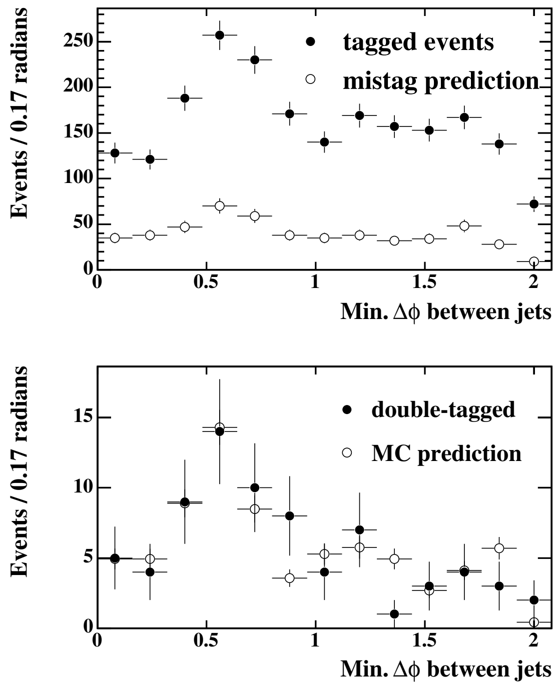

FIG. 10: Distribution of closest jets in $\Delta\phi$ for tagged 3-jet data events with fake tag prediction (top) and for double-tagged events in which the tagged jets are also the two closest jets (bottom).

prediction, for both b and c jets, although the PYTHIA calculation seems to match the data more closely. These heavy flavor fractions are compared with the heavy flavor fractions calculated using ALPGEN inclusive jet Monte Carlo with the matching prescription. On average the data/ALPGEN ratio is 1.5 ± 0.4, where the uncertainty is dominated by the systematic uncertainties associated with the ALPGEN heavy flavor calculations (Table VI). From these fits alone, it is not clear if the discrepancy is consistent for all production diagrams or only for some subsets of gluon splitting to heavy flavor partons.

Because jets with gluon splitting have a small opening angle, the distribution of $\Delta\phi$ between the two closest jets in an event highlights the contribution from gluon splitting. A sample of events with 2 tagged jets is selected from the 3-jet sample and compared to Monte Carlo. The mistag contribution is removed from the double-tagged samples by subtracting events with one or more negative tag. The good agreement, shown in Fig. 10, indicates that the gluon splitting contribution relative to other production mechanisms is well-modelled.

Another sample with gluon splitting contributions, this time of single-tagged 3-jet events, can be used to check the dependence of the data/ALPGEN normalization factor. When the

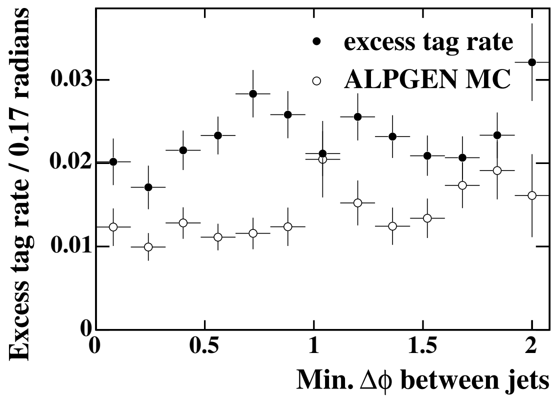

FIG. 11: The positive tag excess rate in data and ALPGEN Monte Carlo as a function of $\Delta\phi$.

excess tag rate, interpreted as the heavy flavor fraction, is plotted as a function of minimum $\Delta\phi$ between jets (Fig. 11), the fractions are flat as a function of $\Delta\phi$ even though the heavy flavor fractions in data are still 1.5 times the heavy flavor fractions in Monte Carlo. This consistency disfavors the hypothesis of missing or under-represented heavy flavor production diagrams.

The measured ratio of 1.5 ± 0.4 between the heavy flavor fractions in the ALPGEN/HERWIG samples and the data is not inconsistent with other recent studies, which indicate that a K-factor may be necessary to account for higher-order effects [37]. Based on this calibration with the jet data sample, we scale the expected $Wb\bar{b}$ and $Wc\bar{c}$ background contributions derived from ALPGEN by a factor of 1.5 ± 0.4. Since the Wc background is produced through a different diagram, that contribution is not rescaled.

Table VIII summarizes the one and two b (c) fractions as a function of jet multiplicity, as well as the corresponding SecVtx tagging efficiencies, where the efficiency scale factor, as calculated in Section IV, has already been applied. The 1B (1C) fractions are for events with exactly one jet matched to a b (c) parton, and the 2B (2C) fractions are for events with exactly two jets matched to b (c) partons. These values are used in Section VI C to predict the background contribution from W + heavy flavor production

TABLE VIII: Ratio of W + heavy flavor production to total W + jet production, for different jet multiplicities. The heavy flavor ratios include the correction factor 1.5 ± 0.4 as measured from jet data, and the SecVtx event tagging efficiencies include the scale factor described in Section IV. These values are used in Section VI C to predict the background contribution from W + heavy flavor production.

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

VI. BACKGROUNDS IN THE TAGGED W + JETS SAMPLE

The non-$t\bar{t}$ events in the W + jets sample are from direct QCD production of heavy flavor without an associated W boson, mistags of light quark jets in W + jets events, W + heavy flavor production, and other low rate electroweak processes with heavy flavor such as diboson and single top production. The estimation of each of these backgrounds is described in turn.

A. Non-W QCD Background

The non-W QCD background is a mixture of events where the lepton does not come from the decay of a W or Z boson. These include lepton and missing energy fakes as well as semileptonic b hadron decays. Since several backgrounds are calculated by normalizing to the number of W + jets events before tagging, it is necessary to understand the level of QCD contamination in the pretag sample. In addition, some of these non-W QCD events may be b-tagged. Both the pretag and tagged contributions are measured directly from data events.

In a leptonic W decay, the lepton is isolated and there is large $\not\!\!E_T$ due to the neutrino, while in non-W events this is not necessarily true. We define the lepton isolation, $\mathcal{I}_{\rm sol}$, as the ratio of energy (not due to the lepton) in the calorimeter in a cone around the lepton direction to the measured electron (muon) energy (momentum). Isolated leptons will have small values of $\mathcal{I}_{\rm sol}$. Sideband regions for lepton isolation and $\not\!\!E_T$ in the high-p_T lepton sample contain mostly non-W events and are used to extrapolate QCD expectations in the signal region. The sideband regions are defined as follows:

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

For the QCD background these two variables are assumed to be mostly uncorrelated: the ratio of non-W events at low and high $\mathcal{I}_{\rm sol}$ values in the low $\not\!\!E_T$ region is the same as in the high $\not\!\!E_T$ region. The number of non-W events in the signal region is estimated by

$$QCD_D = \frac{N_B \times N_C}{N_A}. \tag{3}$$

The contribution of true W and $t\bar{t}$ events in the sideband regions is estimated using Monte Carlo samples to determine the ratio of W and $t\bar{t}$ in the signal and sideband regions, and normalized to the observed number of events in the pretag signal region. The correction is 5-30% depending on the lepton type and event jet multiplicity.

1. *Pretag Backgrounds*

The non-W QCD background is calculated separately for the electron and muon channels, as well as for different jet multiplicities. Table IX gives the predicted QCD background fraction in the signal region. The main source of systematic uncertainty is the underlying assumption that the lepton isolation and $\not\!\!E_T$ are uncorrelated for this background. A study of non-isolated leptons indicates that this assumption adds a 25% systematic uncertainty to the non-W QCD background estimate.

2. *Tagged Backgrounds*

Some of the non-W QCD events are b-tagged and end up in the final event count. One estimate of this contribution applies Equation 3 to the tagged event sample, but this method is limited by the tagged sample size. To increase the number of events, regions A and C are redefined by lowering the isolation boundary to the edge of the signal region, $\mathcal{I}_{\mathrm{sol}} > 0.1$. The precision on this estimate is limited by the number of tagged events in the sideband regions.

A second method scales the pretag QCD fraction by the average tagging rate for QCD events. This method has the advantage of normalizing the background with the larger statistics of the pretag sample, but requires a reliable estimate of the tag rate. The tagging rate in region B for events with two or more jets is applied to the number of taggable jets in the signal region times the pretag QCD background fraction.

Both background estimates contribute to the weighted average shown in Table IX.

B. Mistags

Mistag background events are W + jets events where the tagged jet does not result from the decay of a heavy quark. As described in Sec. IV C, the mistag rate per jet is parameterized as a function of the number of tracks, the raw jet E_T, the η and ϕ of the jet, and the sum of the E_T for all jets in the event with $E_T > 10$ GeV and $|\eta| < 2.4$. To estimate the size of the mistag background, each jet is weighted with its mistag rate in the pretag sample. The sum of the weights over all jets in the sample is then scaled down by the fraction of pretag events which are due to QCD background, as in Sec. VI A 1, since

TABLE IX: non-W QCD background estimate. Results from the tag rate method and the tag sample method are the number of events expected in the b-tagged lepton + jets sample.

			$H_T > 0$		$H_T > 200$ GeV	
Jet multiplicity	1 jet	2 jets	3 jets	≥ 4 jets	3 jets	≥ 4 jets
Electrons						
Pretag non-W QCD Fraction	0.14 ± 0.04	0.17 ± 0.04	0.20 ± 0.05			
Tag Rate Method	16.3 ± 4.7	7.4 ± 2.2	3.2 ± 1.0		2.1 ± 0.6	
Tag Sample Method	21.8 ± 3.8	10.0 ± 2.2	4.9 ± 1.3		2.6 ± 0.8	
Combined Tag Estimate	19.6 ± 3.0	8.7 ± 1.6	2.7 ± 0.6	1.1 ± 0.2	1.3 ± 0.3	1.0 ± 0.3
Muons						
Pretag non-W QCD Fraction	0.034 ± 0.010	0.043 ± 0.011	0.075 ± 0.023			
Tag Rate Method	4.0 ± 1.3	1.2 ± 0.6	0.7 ± 0.3		0.5 ± 0.1	
Tag Sample Method	4.8 ± 1.1	1.8 ± 0.5	1.3 ± 0.4		1.0 ± 0.3	
Combined Tag Estimate	4.5 ± 0.8	1.5 ± 0.4	0.7 ± 0.2	0.2 ± 0.1	0.3 ± 0.1	0.3 ± 0.1
Electron+Muon	24.3 ± 3.5	10.5 ± 1.9	3.4 ± 0.7	1.3 ± 0.3	1.6 ± 0.4	1.2 ± 0.4

these have already been counted in the procedure of Sec. VI A 2. The low mistag rate per jet means that a negligible number of events have more than one mistagged jet; therefore, the number of mistagged jets is a good approximation of the number of events with at least one mistagged jet. This method is tested by comparing the negative SecVtx tags observed and predicted for the pretag sample as a function of the jet E_T, plotted in Figure 12. There is reasonable agreement in the shape and normalization of the prediction.

For the estimate of the number of fake positive tags, the mistag correction factor of 1.2 ± 0.1 described in Sec. V C is applied to account for additional mistags of light quark jets due to material interactions or long lived light quark hadrons. The final results for the mistag estimate are shown in Table XIII. The error includes statistical uncertainties from the pretag sample, including the small effect of correlation between mistag weights that come from the same bin. In addition, an 11% total systematic uncertainty includes uncertainty due to the sample dependence of the mistag rate parameterization and the uncertainty on the mistag correction factor of 1.2 for the positive/negative mistag asymmetry.

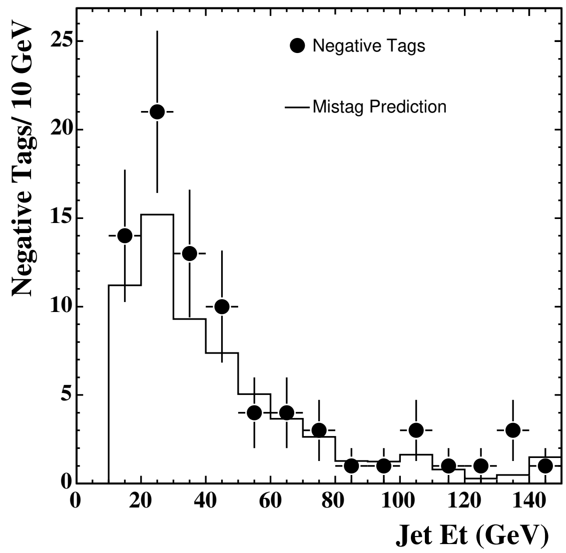

FIG. 12: Comparison of observed and predicted negative SecVtx tags *vs.* jet E_T in the lepton + jets sample.

C. W + Heavy Flavor Backgrounds

The production of W bosons associated with heavy flavor in the processes $Wb\bar{b}$, $Wc\bar{c}$, and Wc is a significant part of the background for the tagged sample. The techniques described in Sec. V are used to estimate the fraction of the inclusive W + jets events which have $Wb\bar{b}$, $Wc\bar{c}$, and Wc. The number of $Wb\bar{b}$, $Wc\bar{c}$, and Wc events is given by multiplying the heavy flavor fractions by the pretag event count, after subtracting the non-W backgrounds. Estimates of the tagged background are then obtained by multiplying the tagging efficiencies summarized in Table VIII.

The pretag W + jets sample includes some contribution from misidentified $Z \to \mu^+\mu^-$ events. The heavy flavor fraction for that process is twice as large as for the W events. The extra contribution of heavy flavor from Z events is described in Section VII and given in Table X. Corrections due to $t\bar{t}$ contributions in the pretag events are discussed in Section IX.

D. Other Backgrounds

A number of backgrounds are too small to be measured directly, thus we use the Monte Carlo to predict their contribution to the sample. The diboson production processes WW, WZ, and ZZ, in association with jets, can mimic the $t\bar{t}$ signal when one boson decays leptonically and the other decays to a taggable b or c quark jet. The process $Z \to \tau^+\tau^-$, in association with jets, can mimic the signal when one τ decays leptonically and the other hadronically. Top quarks are expected to be produced singly with a $t\bar{b}$ final state through s-channel $q\bar{q}$ annihilation, and t-channel W-gluon fusion processes.

We use Monte Carlo samples to measure the acceptance and tagging efficiency. The Monte Carlo acceptance is corrected for the lepton identification and trigger efficiencies as is done for the $t\bar{t}$ acceptance as described in Section VIII. The tagging efficiency is scaled by the MC/data tagging scale factor, with double the uncertainty for tagging charm jets as in $W \to c\bar{s}$. The normalization is based on the measured integrated luminosity and the following theoretical cross sections $\sigma(\text{single top}) = 2.86 \pm 0.09\,\text{pb}$, $\sigma(WW) = 13.25 \pm 0.25\,\text{pb}$, $\sigma(WZ) = 3.96 \pm 0.06\,\text{pb}$, and $\sigma(ZZ) = 1.58 \pm 0.02\,\text{pb}$ [38, 39].

E. Background Summary

A complete summary of all of the background contributions is given in Table XIII. Figure 13 shows the contribution of the different backgrounds for each jet bin compared to the number of data events satisfying all of the selection criteria and having at least one positively tagged jet. We find good agreement between background and data in the one and two jet bins, validating our background calculation. The excess of tags in the three and four jet bins is attributed to $t\bar{t}$. We have already described how the estimates for $Wb\bar{b}$, $Wc\bar{c}$, Wc and mistags, which depend on the number of W pretag events in the data, are corrected for the contribution of QCD backgrounds to the pretag sample. A similar correction needs to be made to account for the real $t\bar{t}$ in the pretag sample. This is done as part of the cross section measurement as described in Section IX. We find the pretag sample to be 10-15% $t\bar{t}$ in the three jet bin and 40-50% $t\bar{t}$ in the four jet bin.

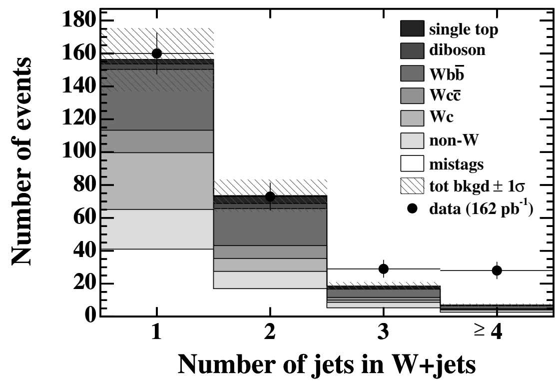

FIG. 13: Number of events passing the selection criteria with at least one tagged jet, and the background prediction for the same selection. The H_T requirement has not been applied.

VII. CROSS CHECK USING THE Z + JETS SAMPLES

An investigation into the Z + jets sample provides a good cross check on our background calculations since the $t\bar{t}$ and non-W QCD contamination in these samples is small.

The heavy flavor contribution in Z + jets is expected to be close to that in W + jets in terms of gluon splitting. However, there is an additional diagram $gg \rightarrow Zb\bar{b}$, which is not present in W + jets. We use the same procedures described in previous sections to estimate the heavy flavor fractions using the Z + jets ALPGEN Monte Carlo samples. The fraction of $Zc\bar{c}$ events (including Zc) is approximately twice the fraction of $Wc\bar{c}$ events, and the fraction of $Zb\bar{b}$ events is approximately twice the fraction of $Wb\bar{b}$ events. The heavy flavor fractions in Z + jets are therefore estimated by multiplying the above factors with the heavy flavor fractions in W + jets listed in Table VIII.

Events with a Z boson are selected by identifying oppositely charged e^+e^- and $\mu^+\mu^-$ pairs with an invariant mass between 75 and 105 GeV/c^2. Both leptons are required to pass the tight lepton selection used for the W + jets analysis in order to collect a pure sample of Z candidates.

Table X lists the yield of Z candidates and the number of tagged events observed as a function of jet multiplicity. The background predictions are also given and are calculated in the same way as in the previous sections for the W + jets sample: 14.0 ± 1.9 events are predicted and 18 are observed in the Z + jets sample (Fig. 14).

Some Z + jets events which fail the standard Z removal contribute to the W + jets

TABLE X: The predicted number of Z + jets events and the observed number, along with the Z + jets contribution left in the W + jets sample and the estimate of the resulting extra b tags in that sample. (The prediction of extra b-tagged events is included in the predicted background summary for the W + jets sample.)

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

sample. The fraction of $Z \to \mu^+\mu^-$ events left in the W sample is about $72 \pm 8\%$ of the number of events observed in $Z \to \mu^+\mu^-$ decay. The contribution of $Z \to e^+e^-$, on the other hand, is negligible. Since those Z events left in the W sample have a higher heavy flavor fraction than the W events, a correction factor accounts for the additional tagged events expected in the W + jets sample.

VIII. EVENT SELECTION OPTIMIZATION AND ACCEPTANCE

The event selection described in Section III, combined with the requirement that at least one jet be positively b-tagged, yields a clean sample of top decays in the lepton plus jets channel; the expected signal over background ratio is of the order of 2:1. Several ways of optimizing the event selection were studied in order to maximize the significance of the cross section measurement, and an inclusive event variable was found to have the greatest power to discriminate $t\bar{t}$ signal from background events. The following section discusses

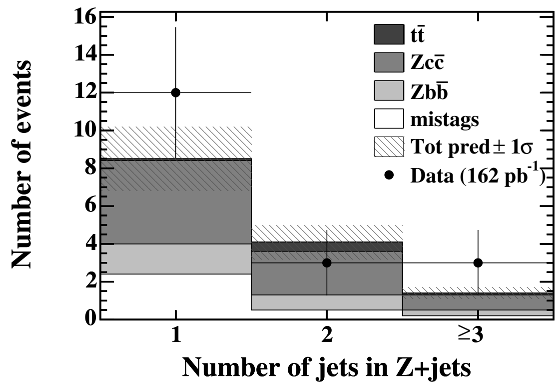

FIG. 14: Comparison of the observed and predicted number of events in the b-tagged Z + jets sample.

event selection optimization using the H_T variable.

A. Optimization with the H_T variable.

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

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

Because of the large mass of the top quark, H_T, which is representative of the hard scatter of the event, tends to be significantly larger for $t\bar{t}$ events than for the backgrounds. Figure 15 shows the distribution of the H_T variable after all selection cuts have been applied, including b-tagging, for $t\bar{t}$ Monte Carlo, and for the main backgrounds: W + heavy flavor, non-W QCD, and mistags. The W + heavy flavor distributions are taken from ALPGEN Monte Carlo, but all other background shapes are estimated from data; the features visible at high H_T in the non-W QCD and mistags distributions are due to poor statistics in the control samples.

The H_T distributions for the three major backgrounds are estimated using methods described in Section VI. The ALPGEN Monte Carlo generator is used to estimate the shape of the distribution for the W + heavy flavor background. The non-W QCD background shape

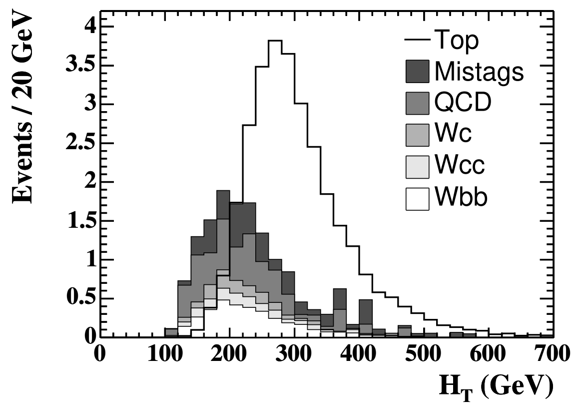

FIG. 15: Distribution of the H_T variable for $t\bar{t}$ Monte Carlo, and for various backgrounds normalized to an integrated luminosity of $107\,\mathrm{pb}^{-1}$.

is evaluated by selecting pretag events where the lepton is not isolated (isolation > 0.2), while all other kinematic cuts remain unchanged. This subsample is presumably dominated by QCD events with kinematic properties identical to the QCD background events that satisfy the event selection (isolation< 0.1). Each event in the sub-sample is then weighted by the total positive tagging rate measured from the jet sample (see Section VI B). The mistag background shape is estimated from the pretag sample, where each event is weighted by the negative tag rate measured from the jet sample. Other backgrounds (which account for less than 10% of the total background) are included in the overall normalization, with the implicit assumption that their shape is not significantly different from the others. The $t\bar{t}$ contribution is normalized to the theoretical cross section.

Figure 15 shows that signal and background can be separated by the use of the H_T variable. Figure 16 shows the signal over background ratio and cross section sensitivity as a function of an H_T cut, computed from figure 15. The statistical sensitivity ($S/\sqrt{S+B}$) is compared to the total sensitivity ($S/\sqrt{S+B+\sigma(B)^2}$, where $\sigma(B)$ is the absolute systematic error on the background estimate). Systematic uncertainties arising from the H_T cut itself are described in Section IX; they are small enough to be neglected in the optimization process. A cut requiring $H_T > 200$ GeV is found to be optimal: such a cut keeps 96% of the signal and rejects 39% of the background; this improves the signal over background ratio from 2 to 3 and the total significance on the $t\bar{t}$ cross section measurement by 6%.

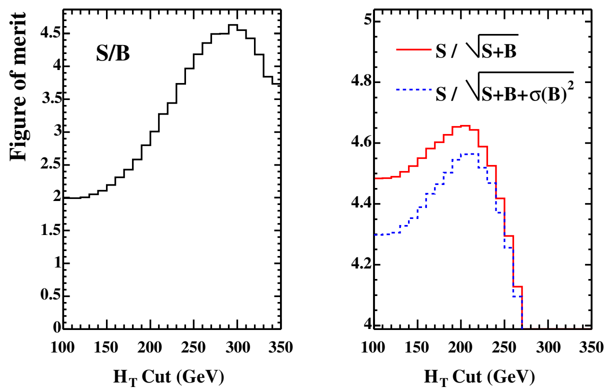

FIG. 16: Estimate of S/B, statistical and total significance for 107 pb^{-1} integrated luminosity, as a function of H_T cut.

B. Acceptance

The acceptance is defined as the fraction of produced $t\bar{t}$ events that satisfy all trigger and selection criteria. It includes trigger efficiency, reconstruction efficiency, as well as the efficiency of the kinematic cuts, and of the b-tagging algorithm. The branching ratios of the various decay modes of the $t\bar{t}$ pair are included as well. The details of the acceptance estimate, in particular its systematic uncertainties, are described elsewhere [8], but the parts specific to this analysis – the b-tagging efficiency, and the H_T variable – are detailed here.

The contributions to the acceptance are divided as follows:

$$\epsilon_{t\bar{t}} = \epsilon_{\text{trig}} \cdot \epsilon_{z_0} \cdot \epsilon_{\text{veto}} \cdot \epsilon_{t\bar{t}}^{\text{MC}} \cdot k_{\text{lep-id}} \cdot \epsilon_{\text{tag-event}} \tag{4}$$

where ϵ_{z_0} is the efficiency of the $|z_0| \leq 60$cm cut, $\epsilon_{t\bar{t}}^{\text{MC}}$ is the fraction of Monte Carlo $t\bar{t}$ events which pass all the selection cuts (except for b-tagging), ϵ_{veto} is the combined efficiency of the various vetoes (conversion removal, cosmic removal, dilepton and Z^0 rejections), ϵ_{trig} is the trigger efficiency for identifying high p_T leptons, and $\epsilon_{\text{tag-event}}$ is the efficiency to tag at least one jet in a $t\bar{t}$ event. $k_{\text{lep-id}}$ is a factor that corrects for the lepton identification efficiency difference between data and Monte Carlo.

The efficiency of the z_0 cut is measured from data and found to be $\epsilon_{z_0} = 0.951 \pm 0.005$. The trigger efficiency ϵ_{trig} is different for each type of lepton trigger: $\epsilon_{\text{trig}}^{CEM} = 0.9656 \pm 0.0006$, $\epsilon_{\text{trig}}^{CMUP} = 0.887 \pm 0.007$, and $\epsilon_{\text{trig}}^{CMX} = 0.954 \pm 0.006$. The factor $k_{\text{lep-id}}$ is evaluated by comparing a data sample of Z + jets events with a PYTHIA Z sample, and found to be 1.00

TABLE XI: Summary table of the $t\bar{t}$ acceptance, for a top quark mass of $175\,\mathrm{GeV}/c^2$.

	CEM	CMUP	CMX	Total
Sample (total)	344,264	344,264	344,264	344,264
# Events w/o b-tag	15,893	9791	3617	29301
Acc. w/o b-tag (%)	4.09±0.03±0.36	2.13±0.02±0.19	0.959±0.016±0.085	7.18±0.04±0.61
# Tagged Events	8490	5202	1965	15657
Tag Efficiency (%)	53.4±0.4±3.2	53.1±0.5±3.2	54.3±0.8±3.3	53.4±0.3±3.2
Acc. with b-tag (%)	2.19±0.02±0.23	1.14±0.01±0.12	0.512±0.009±0.054	3.84±0.03±0.40
Integ. Lumi. (pb^{-1})	162±10	162±10	150±9	

for electrons, 0.95 for CMUP muons, and 0.99 for CMX muons; because there are few $Z+$ jets events at high jet multiplicity, we assign a 5% relative systematic uncertainty on $k_{\mathrm{lep-id}}$.

The efficiency $\epsilon_{t\bar{t}}^{\mathrm{MC}}$ is evaluated using a sample of PYTHIA $t\bar{t}$ Monte Carlo events with top quark mass $m_t = 175\,\mathrm{GeV}/c^2$. Table XI summarizes the event selection acceptance for each type of lepton. The b-tagging efficiency is measured from the same Monte Carlo sample, and takes into account the b-tagging scale factor (section IV) by randomly keeping only 82% of all the tags, and discarding the others. We find the efficiency for tagging at least one jet in a $t\bar{t}$ event (after all other cuts have been applied, including $H_T > 200\,\mathrm{GeV}$) to be 53.4±0.3(stat.)±3.2(syst.)%; the systematic uncertainty comes from the measurement of the b-tagging scale factor, described in Section IV. An important source of uncertainty comes from the measurement of the jet energy, which also affects the $\not\!\!E_T$ and H_T measurements (energy scale): uncertainties relative to the η-dependent detector response, overall energy scale, non-linearity, fraction of energy lost outside the reconstructed jet cone, and multiple interactions are added in quadrature. Multiple interactions are soft interactions that can contribute to the jet energy measurement. They are not included in the simulation; instead, a small average correction is applied to each jet in the data. The energy of each jet is then shifted up and down in the MC by the uncertainty, and half of the difference in the acceptance (4.9%) is taken as a systematic uncertainty. The overall acceptance is 3.84±0.03(stat.)±0.40(syst.)%, including all systematic effects. Table XII summarizes the dominant sources of uncertainty for the acceptance.

TABLE XII: Relative systematic uncertainties on the signal acceptance which are common to all lepton types.

Quantity	Relative error (%)
ϵ_{z_0}	0.5
Tracking Efficiency	0.4
Energy Scale	4.9
PDF	2.0
ISR/FSR	2.6
MC modelling	1.4
Lepton ID	5.0
b-tagging	6.0

IX. CROSS-SECTION FOR SINGLE-TAGGED EVENTS ($\geq$ 1 B-TAGS)

The production cross section follows from the acceptance measurement and the background estimate:

$$\sigma_{t\bar{t}} = \frac{N_{\text{obs}} - N_{\text{bkg}}}{\epsilon_{t\bar{t}} \times \mathcal{L}}, \tag{5}$$

where N_{obs} and N_{bkg} are the number of total observed and background events, respectively, in the $W + \geq 3$ jet bins (see Table XIII); $\epsilon_{t\bar{t}}$ is the signal acceptance (see Table XI); and $\mathcal{L}$ is the integrated luminosity. Many of the predicted backgrounds are based on the number of pretag data events, but that number includes a significant contribution from $t\bar{t}$ events. After subtracting this contribution from the pretag sample, the dependent backgrounds are recalculated. The final background contributions for the single-tag selection are summarized in Table XIII and represented in Figure 17.

The properties of the selected candidate events are consistent with the expectations for $t\bar{t}$ pair production and background contributions. Figures 18 and 19 show the distribution of the event H_T and the tagged jet E_T, and Figure 20 shows the pseudo-$c\tau$ of the tagged jets.

For the optimized selection with the H_T requirement, and for a top quark mass $m_{\text{t}} = 175\,\text{GeV}/c^2$,

$$\sigma_{t\bar{t}} = 5.6^{+1.2}_{-1.1}(\text{stat.})^{+0.9}_{-0.6}(\text{syst.})\text{pb}. \tag{6}$$

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

			$H_T > 0$ GeV		$H_T > 200$ GeV	
Jet multiplicity	W + 1 jet	W + 2 jets	W + 3 jets	W + $\geq$ 4 jets	W + 3 jets	W + $\geq$ 4 jets
Pretag	15314	2448	387	107	179	91
Mistags	40.9 ± 6.1	17.0 ± 2.4	5.2 ± 0.7	2.6 ± 0.4	3.3 ± 0.4	2.3 ± 0.3
$Wb\bar{b}$	37.0 ± 11.2	22.5 ± 6.5	5.0 ± 1.3	1.6 ± 0.5	2.8 ± 0.8	1.4 ± 0.4
$Wc\bar{c}$	13.7 ± 3.4	8.0 ± 2.2	1.6 ± 0.5	0.6 ± 0.2	0.9 ± 0.3	0.5 ± 0.2
Wc	34.5 ± 9.0	7.7 ± 2.0	1.4 ± 0.4	0.3 ± 0.1	0.7 ± 0.2	0.3 ± 0.1
$WW/WZ/ZZ$,$Z \rightarrow \tau\tau$	2.2 ± 0.4	2.5 ± 0.4	0.6 ± 0.1	0.1 ± 0.0	0.3 ± 0.1	0.1 ± 0.0
non-W QCD	24.3 ± 3.5	10.5 ± 1.9	3.4 ± 0.7	1.4 ± 0.4	1.7 ± 0.4	1.2 ± 0.3
single top	2.6 ± 0.3	4.6 ± 0.5	1.1 ± 0.1	0.2 ± 0.0	0.8 ± 0.1	0.2 ± 0.0
Z+HF	1.1 ± 0.3	0.6 ± 0.2	0.2 ± 0.1		0.10 ± 0.05	
Total	156.3 ± 19.1	73.4 ± 9.8	18.5 ± 2.2	6.9 ± 0.9	10.5 ± 1.3	6.0 ± 0.8
Corrected Total	156.3 ± 19.1	73.4 ± 9.8	23.1 ± 3.0		13.5 ± 1.8	
Data	160	73	29	28	21	27

TABLE XIV: Measured cross section for different top quark mass assumptions.

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

The systematic uncertainty is due to uncertainties on the signal acceptance (10% relative), luminosity measurement (6%), and background estimate (5%). The acceptance, and therefore the measured cross section, changes with the top quark mass as shown in Table XIV.

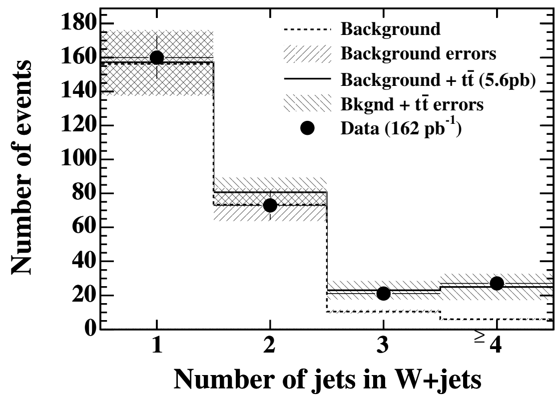

FIG. 17: Background and $t\bar{t}$ signal expectation (based on measured $t\bar{t}$ cross section) as a function of jet multiplicity. Events with three or more jets are required to have $H_T > 200\,\text{GeV}$.

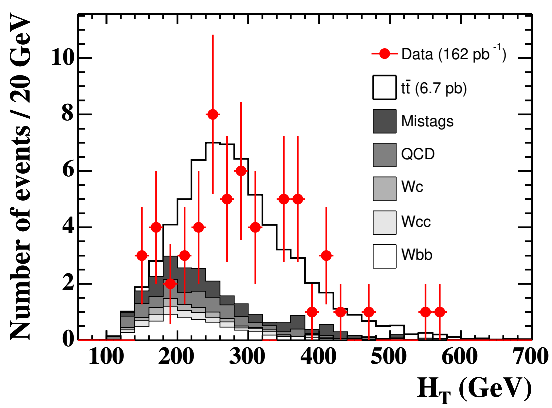

FIG. 18: H_T distribution of the 57 tagged events with three or more jets, compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross-section of 6.7 pb).

X. CROSS CHECK USING DOUBLE-TAGGED EVENTS (≥ 2 b-TAGS)

Each $t\bar{t}$ event contains two energetic b quarks, making it likely that two jets in the event will be tagged. Of the 57 tagged events in the three and four jet bins before the H_T cut, 8 of these are double-tagged events. The double-tag sample provides a cleaner $t\bar{t}$ sample in which to cross-check the cross section with a significantly smaller systematic uncertainty due to the background estimate, although with decreased statistical precision.

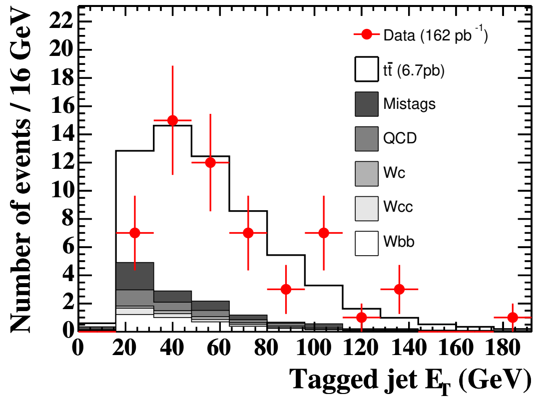

FIG. 19: E_T distribution of the tagged jets in the 57 candidate events, compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross-section of 6.7 pb).

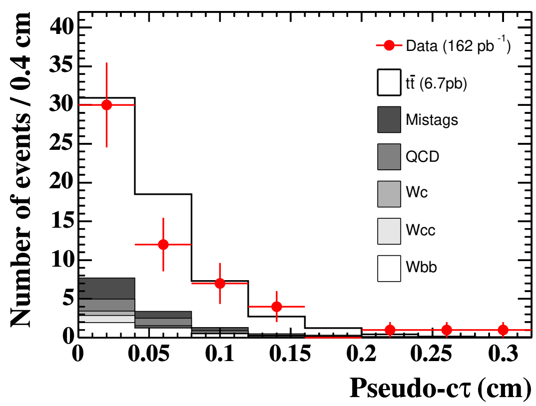

FIG. 20: Pseudo-$c\tau$ distribution of the tagged jets in the 57 candidate events, compared to the expected background and $t\bar{t}$ signal (normalized to the theoretical cross-section of 6.7 pb).

1. Double-Tag Backgrounds

The background estimate for the double-tag selection uses the methods described in Section VI, except for a few additional issues which are specific to the double-tag estimate.

The mistag estimate for double-tags is dominated by events with one real tag of a heavy flavor jet with the second tag coming from the mistag of an additional jet. The mistag estimate is obtained by applying the mistag matrix to the jets in the tagged sample, in contrast to the pretag sample used for the inclusive estimate. Since the tagged sample with three or more jets is dominated by $t\bar{t}$ events, applying the mistag matrix to the entire tagged sample gives an overestimate of the mistag background. Additionally, $Wb\bar{b}$ and $Wc\bar{c}$ events with two heavy flavor jets are not counted as part of the mistag estimate; rather, the mistag estimate is scaled by the fraction of inclusive tagged events which are from mistags, Wc, and non-W QCD background.

The largest background comes from $Wb\bar{b}$ events, where both b quark jets are tagged. This background estimate uses the same heavy flavor fractions and tagging efficiencies given in Section V, and is normalized to the pretag sample of W + jet events. There is a small additional contribution of double-tags in $Wb\bar{b}$ events where the second tag is from mistags of light quark jets, so the mistag matrix is applied to light quark jets in the Monte Carlo sample to account for this additional contribution to the $Wb\bar{b}$ event double-tag efficiency. The $Wc\bar{c}$ background is treated in the same way as $Wb\bar{b}$.

The non-W QCD background estimate uses the same lepton isolation and $\not\!\!E_T$ sideband regions described in Section VI A to estimate the double-tag background from direct production of heavy flavor jets. There are zero double-tagged events with two or more jets in region B ($\mathcal{I}_{\rm sol} < 0.1$ and $\not\!\!E_T < 15$ GeV) compared to 133 single-tagged events, implying a Poisson upper limit of 1.8% at 90% C.L. for a single-tagged QCD event to be double-tagged. This is applied to the inclusive tag QCD background estimate given in Table XIII, and the limit is quoted as the uncertainty on the background estimate of zero double-tagged QCD events.

The double-tag backgrounds for the electroweak and single top processes follow directly from the same Monte Carlo calculations discussed in Section VI D. The only significant contributions come from the s-channel single top process and WZ with $Z \to b\bar{b}$. The total double-tag background estimate is given in Table XV.

2. *Double-Tag Acceptance and Cross Section*

For the double-tag analysis the backgrounds are sufficiently low that we do not apply a cut on H_T. The pretag acceptance uses the same MC sample and lepton identification and trigger efficiency corrections described in Section IX. The pretag efficiencies are $4.32 \pm 0.35\%$ for CEM electron, $2.24 \pm 0.22\%$ for CMUP muon and $1.01 \pm 0.13\%$ for CMX muon $t\bar{t}$ events with three or more jets. The efficiency to double-tag $t\bar{t}$ events with three or more jets is measured from Monte Carlo to be 0.11 ± 0.02 after correcting for the difference in

TABLE XV: Prediction for the number of double-tagged events. Corrected total comes from the $t\bar{t}$ cross section measurement where the pretag sample is corrected for the $t\bar{t}$ contribution. The expected number of $t\bar{t}$ events is calculated using the measured cross section of 5.0 pb.

Jet multiplicity	2 jets	3 jets	$\geq$ 4 jets
Single top	0.40$\pm$ 0.08	0.15$\pm$ 0.03	0.04$\pm$ 0.01
WZ	0.15$\pm$ 0.04	0.02$\pm$ 0.01	0.01$\pm$ 0.01
$Wb\bar{b}$	2.76$\pm$ 0.86	0.64$\pm$ 0.18	0.21$\pm$ 0.06
$Wc\bar{c}$	0.20$\pm$ 0.08	0.05$\pm$ 0.02	0.03$\pm$ 0.01
Mistag/QCD	0.14 $\pm$ 0.04	0.16 $\pm$ 0.04	0.11 $\pm$ 0.03
Total	3.65 $\pm$ 0.97	1.02 $\pm$ 0.23	0.40 $\pm$ 0.09
Corrected Total	3.6$\pm$ 1.0	1.3 $\pm$ 0.3	
$t\bar{t}$ (5.0 pb)	1.0	2.6	4.1
Data	8	3	5

tagging efficiency between data and Monte Carlo.

The total double-tag background estimate is given in Table XV. The cross section for the double-tagged sample is measured using the events in the three and four jet bins as for the inclusive tagged sample, and correcting the pretag sample for the $t\bar{t}$ contribution from double-tags. Eight double-tag events are observed on a background of 1.3 events, implying a cross section of

$$\sigma_{t\bar{t}} = 5.0^{+2.4}_{-1.9}(\text{stat})^{+1.1}_{-0.8}(\text{syst})\text{pb}. \tag{7}$$

The systematic error is due to the following contributions: tagging efficiency (15%), acceptance (7%), luminosity (6%) and backgrounds (5%).

This result gives a consistent cross section measurement in an almost background-free sample. With a larger data sample, this double-tag selection may offer an improved measurement of the $t\bar{t}$ cross section. In addition, the double-tagged sample may prove useful in estimating relative contributions of the different W + jets production diagrams, especially gluon splitting to heavy flavor quark pairs.

XI. CONCLUSIONS

The $t\bar{t}$ production cross section has been measured with vertex-tagged lepton + jets events from $162\,\mathrm{pb}^{-1}$ of data collected at $\sqrt{s} = 1.96\,\mathrm{TeV}$. The selection yields a sample of 48 candidate events with one lepton, large missing transverse energy, large total transverse energy H_T, and three or more jets, where at least one jet has a displaced secondary vertex tag. A total of 13.5 ± 1.8 events are expected from non-$t\bar{t}$ processes. The measured production cross section, assuming a top quark mass of $175\,\mathrm{GeV}/c^2$, is

$$\sigma(p\bar{p} \to t\bar{t}) = 5.6^{+1.2}_{-1.1}(\mathrm{stat.})^{+0.9}_{-0.6}(\mathrm{syst.})\ \mathrm{pb}.$$

Applying the same selection, except for the large transverse energy requirement, to a double-tagged sample (8 observed events with expected background of 1.3 ± 0.3 events) yields a cross section of $5.0^{+2.4}_{-1.9}(\mathrm{stat.})^{+1.1}_{-0.8}(\mathrm{syst.})$ pb. Both results are consistent with the theoretical predictions of $6.7^{+0.7}_{-0.9}$ pb, again assuming $m_t = 175\,\mathrm{GeV}/c^2$ [1, 2].

APPENDIX: DERIVATION OF FORMULAE FOR DOUBLE-TAG METHOD OF DETERMINING EFFICIENCY SCALE FACTOR

The measurement of the tagging efficiency in data employs the double-tag method and uses identified conversions to estimate the contribution of electrons which are fakes or part of a conversion pair. This appendix summarizes the detailed calculation of the tagging efficiency.

Most of the electrons in the inclusive electron data sample (electron $E_T > 8\,\mathrm{GeV}$ with no $\not{E}_T$requirement) are from conversions or fakes in light flavor jets. Away jet tagging enhances the heavy flavor fraction in the electron side, but it still needs a significant light flavor correction. In general, the heavy flavor production in the jets comes from three subprocesses: direct production, flavor excitation and gluon splitting. For simplicity, the final data sample can be divided into the following four subclasses:

- N_{HH}: the number of events where both sides contain a heavy quark, either c or b (it includes the contributions of gluon splitting in both sides),

- N_{HL}: the number of events where the electron side is heavy flavor and the away side is light flavor,

- N_{LH}: the number of events where the electron is coming from fakes and conversions and the away side contains heavy flavor,
- N_{LL}: the number of events where both sides are light flavors.

By construction, we have

$$N_{HH} + N_{HL} + N_{LH} + N_{LL} = N$$

where N is the total number of events passing the final selection. The heavy flavor contributions in the electron side can be determined using the measurement of heavy flavor fraction (see Section IV).

$$N_{HH} + N_{HL} = F_{HF} \cdot N$$

The N_{LH} contribution can be determined using the away tags in the conversion electron sample. Finally, the contribution of N_{LL} is estimated using the mistags in the negative side.

Let us use the following notation to help the derivation of efficiency measurement.

- ϵ'_H: tagging efficiency of heavy flavor in the electron jet,
- ϵ_H: tagging efficiency of heavy flavor in the away jet,
- ϵ'_L: mistag efficiency in the electron jets,
- ϵ_L: mistag efficiency in the away jets,
- N_{a+}, N_{a-}, N_{e+} and N_{e-}: are the number of positive, negative tags in the away jets and in the electron jets,
- N^{e+}_{a+}, N^{e-}_{a+}, N^{e+}_{a-} and N^{e-}_{a-}: are the number of double-tags in the combination of positive or negative tags in electron jet when the away tag is present, either positive or negative.

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

$$\epsilon_H \cdot N_{HH} + \epsilon_L \cdot N_{HL} + \epsilon_H \cdot N_{LH} + \epsilon_L \cdot N_{LL} = N_{a+} \tag{A.1}$$

$$\epsilon_L \cdot N_{HH} + \epsilon_L \cdot N_{HL} + \epsilon_L \cdot N_{LH} + \epsilon_L \cdot N_{LL} = N_{a-}. \tag{A.2}$$

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

$$(\epsilon_H - \epsilon_L) \cdot (N_{HH} + N_{LH}) = N_{a+} - N_{a-}. \tag{A.3}$$

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

$$\epsilon'_H \cdot (\epsilon_H - \epsilon_L) \cdot N_{HH} + \epsilon'_L \cdot (\epsilon_H - \epsilon_L) \cdot N_{LH} = N^{e+}_{a+} - N^{e+}_{a-} \tag{A.4}$$

$$\epsilon'_L \cdot (\epsilon_H - \epsilon_L) \cdot N_{HH} + \epsilon'_L \cdot (\epsilon_H - \epsilon_L) \cdot N_{LH} = N^{e-}_{a+} - N^{e-}_{a-}. \tag{A.5}$$

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

$$(\epsilon'_H - \epsilon'_L) \cdot (\epsilon_H - \epsilon_L) \cdot N_{HH} = (N^{e+}_{a+} - N^{e-}_{a+}) - (N^{e+}_{a-} - N^{e-}_{a-}). \tag{A.6}$$

From Equation A.3, we get

$$(\epsilon_H - \epsilon_L) \cdot N_{HH} = (N_{a+} - N_{a-}) - (\epsilon_H - \epsilon_L) \cdot N_{LH} \tag{A.7}$$

Substituting Equation A.7 into Equation A.6 and rearranging terms, the heavy flavor tagging efficiency on the electron jet is

$$\epsilon'_H - \epsilon'_L = \frac{(N^{e+}_{a+} - N^{e-}_{a+}) - (N^{e+}_{a-} - N^{e-}_{a-})}{(N_{a+} - N_{a-}) - (\epsilon_H - \epsilon_L) \cdot N_{LH}}$$

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

Let us denote the following quantities

- f: the fraction of electrons originating from conversions in no-heavy flavor jets in the electron side,
- f': the fraction of electrons originating from conversions in heavy flavor jets in the electron side,
- ϵ^c: the efficiency of the conversion finding algorithm,
- ϵ^o: the error rate of finding a real electron as a part of conversion, which is determined using the same sign,
- n^c: the number of identified conversion electrons.

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

$$(f' \cdot \epsilon^c + (1-f') \cdot \epsilon^o) \cdot (N_{HH} + N_{HL}) + (f \cdot \epsilon^c + (1-f) \cdot \epsilon^o) \cdot (N_{LH} + N_{LL}) = n^c \quad \text{(A.8)}$$

$$f \cdot (\epsilon^c - \epsilon^o) \cdot (N_{LH} + N_{LL}) = n^c - N \cdot \epsilon^o - f' \cdot (\epsilon^c - \epsilon^o) \cdot (N_{HH} + N_{HL}). \quad \text{(A.9)}$$

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

$$(\epsilon'_H - \epsilon'_L) \cdot (\epsilon^c \cdot f' + \epsilon^o \cdot (1-f')) \cdot (N_{HH} + N_{HL}) = n^c_{e+} - n^c_{e-}.$$

Since $(\epsilon'_H - \epsilon'_L) \cdot (N_{HH} + N_{HL}) = N_{e+} - N_{e-}$, we get

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

where $\epsilon^{c'} = \frac{(n^c_{e+} - n^c_{e-})}{(N_{e+} - N_{e-})}$.

Substituting Equation A.10 into Equation A.8 , we have

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

We apply the tag to the *electron* jet in Equation A.8 and the excess of tags is

$$f' \cdot (\epsilon^c - \epsilon^o) \cdot (\epsilon'_H - \epsilon'_L) \cdot (N_{HH} + N_{HL}) = n^c_{e+} - n^c_{e-} - (N_{e+} - N_{e-}) \cdot \epsilon^o. \quad \text{(A.12)}$$

If we then apply the tag to the *away* jet in Equation A.8 the excess of tags is

$$f \cdot (\epsilon^c - \epsilon^o) \cdot (\epsilon_H - \epsilon_L) \cdot N_{LH} = n^c_{a+} - n^c_{a-} - (N_{a+} - N_{a-}) \cdot \epsilon^o - f' \cdot (\epsilon^c - \epsilon^o) \cdot (\epsilon_H - \epsilon_L) \cdot N_{HH}. \quad \text{(A.13)}$$

Substituting Equations A.7 and A.10 into Equation A.13:

$$f \cdot (\epsilon^c - \epsilon^o) \cdot (\epsilon_H - \epsilon_L) \cdot N_{LH} = n^c_{a+} - n^c_{a-} - (N_{a+} - N_{a-}) \cdot \epsilon^o - (\epsilon^{c'} - \epsilon^o) \cdot ((N_{a+} - N_{a-}) - (\epsilon_H - \epsilon_L) \cdot N_{LH}) \quad \text{(A.14)}$$

From Equation A.11 and Equation A.14, we get

$$(\epsilon_H - \epsilon_L) \cdot N_{LH} = (N_{a+} - N_{a-}) \cdot \frac{\frac{n^c_{a+} - n^c_{a-}}{N_{a+} - N_{a-}} - \epsilon^{c'}}{n^c/N - \epsilon^{c'}} \cdot (1 - F_{HF}).$$

Finally, the efficiency can be expressed as

$$\epsilon'_H - \epsilon'_L = \frac{(N^{e+}_{a+} - N^{e-}_{a+}) - (N^{e+}_{a-} - N^{e-}_{a-})}{(N_{a+} - N_{a-}) \cdot F^a_{HF}} \quad \text{(A.15)}$$

where

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

ACKNOWLEDGMENTS

We acknowledge many fruitful discussions we have had with Michelangelo Mangano. We also thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Particle Physics and Astronomy Research Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; in part by the European Community's Human Potential Programme under contract HPRN-CT-20002, Probe for New Physics.

---

[1] N. Kidonakis and R. Vogt, Phys. Rev. D **68**, 114014 (2003).

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

[3] D. Acosta et al., submitted to Phys. Rev. Lett. (2004), arXiv:hep-ex/0406078.

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

[5] D. Acosta et al., Phys. Rev. Lett. **93**, 142001 (2004).

[6] D. Acosta et al., submitted to Phys. Rev. D (2004), arXiv:hep-ex/0409029.

[7] D. Acosta et al., to be submitted to Phys. Rev. D (2004), Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt(s) = 1.96$ TeV Using Soft Muon Tagging of b-tagged Lepton+jet Events.

[8] D. Acosta et al., to be submitted to Phys. Rev. D (2004), Measurement of the $t\bar{t}$ Cross Section in the Lepton + Jets Channel Using Event Kinematics.

[9] T. Affolder et al., Nucl. Instrum. and Methods A **526**, 249 (2004).

[10] D. Acosta et al., to be submitted to Phys. Rev. D (2004), First Measurements of Inclusive W and Z Cross Sections from Run II of the Tevatron Collider.

[11] E. J. Thomson et al., IEEE Trans. Nucl. Sci. **49**, 1063 (2002).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[28] M. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. Polosa, JHEP **07**, 001 (2003).

[29] R. Brun and F. Carminati, CERN Programming Library Long Writeup **W5013** (1993).

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

[31] E. Gerchtein and M. Paulini, ECONF **C0303241**, TUMT005 (2003), arXiv:physics/0306031.

[32] S. Catani, F. Krauss, R. Kuhn, and B. Webber, JHEP **11**, 063 (2001).

[33] F. Krauss, JHEP **08**, 015 (2002).

[34] M. Mangano (2004), `http://cepa.fnal.gov/patriot/mc4run2/MCTuning/061104/mlm.pdf`.

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

[36] D. Stump et al., JHEP **10**, 046 (2003).

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

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

[39] B. Harris, E. Laenen, L. Phaf, Z. Sullivan, and S. Weinzierl, Phys. Rev. D **66**, 054024 (2002).