arXiv:0803.0739v1 [hep-ex] 5 Mar 2008

Evidence for production of single top quarks

V.M. Abazov[36], B. Abbott[75], M. Abolins[65], B.S. Acharya[29], M. Adams[51], T. Adams[49], E. Aguilo[6], S.H. Ahn[31], M. Ahsan[59], G.D. Alexeev[36], G. Alkhazov[40], A. Alton[64,a], G. Alverson[63], G.A. Alves[2], M. Anastasoaie[35], L.S. Ancu[35], T. Andeen[53], S. Anderson[45], B. Andrieu[17], M.S. Anzelc[53], M. Aoki[50], Y. Arnoud[14], M. Arov[60], M. Arthaud[18], A. Askew[49], B. Åsman[41], A.C.S. Assis Jesus[3], O. Atramentov[49], C. Avila[8], C. Ay[24], F. Badaud[13], A. Baden[61], L. Bagby[50], B. Baldin[50], D.V. Bandurin[59], P. Banerjee[29], S. Banerjee[29], E. Barberis[63], A.-F. Barfuss[15], P. Bargassa[80], P. Baringer[58], J. Barreto[2], J.F. Bartlett[50], U. Bassler[18], D. Bauer[43], S. Beale[6], A. Bean[58], M. Begalli[3], M. Begel[73], C. Belanger-Champagne[41], L. Bellantoni[50], A. Bellavance[50], J.A. Benitez[65], S.B. Beri[27], G. Bernardi[17], R. Bernhard[23], I. Bertram[42], M. Besançon[18], R. Beuselinck[43], V.A. Bezzubov[39], P.C. Bhat[50], V. Bhatnagar[27], C. Biscarat[20], G. Blazey[52], F. Blekman[43], S. Blessing[49], D. Bloch[19], K. Bloom[67], A. Boehnlein[50], D. Boline[62], T.A. Bolton[59], E.E. Boos[38], G. Borissov[42], T. Bose[77], A. Brandt[78], R. Brock[65], G. Brooijmans[70], A. Bross[50], D. Brown[81], N.J. Buchanan[49], D. Buchholz[53], M. Buehler[81], V. Buescher[22], V. Bunichev[38], S. Burdin[42,b], S. Burke[45], T.H. Burnett[82], C.P. Buszello[43], J.M. Butler[62], P. Calfayan[25], S. Calvet[16], J. Cammin[71], W. Carvalho[3], B.C.K. Casey[50], H. Castilla-Valdez[33], S. Chakrabarti[18], D. Chakraborty[52], K. Chan[6], K.M. Chan[55], A. Chandra[48], F. Charles[19,‡], E. Cheu[45], F. Chevallier[14], D.K. Cho[62], S. Choi[32], B. Choudhary[28], L. Christofek[77], T. Christoudias[43], S. Cihangir[50], D. Claes[67], Y. Coadou[6], M. Cooke[80], W.E. Cooper[50], M. Corcoran[80], F. Couderc[18], M.-C. Cousinou[15], S. Crépé-Renaudin[14], D. Cutts[77], M. Ćwiok[30], H. da Motta[2], A. Das[45], G. Davies[43], K. De[78], S.J. de Jong[35], E. De La Cruz-Burelo[64], C. De Oliveira Martins[3], J.D. Degenhardt[64], F. Déliot[18], M. Demarteau[50], R. Demina[71], D. Denisov[50], S.P. Denisov[39], S. Desai[50], H.T. Diehl[50], M. Diesburg[50], A. Dominguez[67], H. Dong[72], L.V. Dudko[38], L. Duflot[16], S.R. Dugad[29], D. Duggan[49], A. Duperrin[15], J. Dyer[65], A. Dyshkant[52], M. Eads[67], D. Edmunds[65], J. Ellison[48], V.D. Elvira[50], Y. Enari[77], S. Eno[61], P. Ermolov[38], H. Evans[54], A. Evdokimov[73], V.N. Evdokimov[39], A.V. Ferapontov[59], T. Ferbel[71], F. Fiedler[24], F. Filthaut[35], W. Fisher[50], H.E. Fisk[50], M. Fortner[52], H. Fox[42], S. Fu[50], S. Fuess[50], T. Gadfort[70], C.F. Galea[35], E. Gallas[50], C. Garcia[71], A. Garcia-Bellido[82], V. Gavrilov[37], P. Gay[13], W. Geist[19], D. Gelé[19], C.E. Gerber[51], Y. Gershtein[49], D. Gillberg[6], G. Ginther[71], N. Gollub[41], B. Gómez[8], A. Goussiou[82], P.D. Grannis[72], H. Greenlee[50], Z.D. Greenwood[60], E.M. Gregores[4], G. Grenier[20], Ph. Gris[13], J.-F. Grivaz[16], A. Grohsjean[25], S. Grünendahl[50], M.W. Grünewald[30], F. Guo[72], J. Guo[72], G. Gutierrez[50], P. Gutierrez[75], A. Haas[70], N.J. Hadley[61], P. Haefner[25], S. Hagopian[49], J. Haley[68], I. Hall[65], R.E. Hall[47], L. Han[7], K. Harder[44], A. Harel[71], R. Harrington[63], J.M. Hauptman[57], R. Hauser[65], J. Hays[43], T. Hebbeker[21], D. Hedin[52], J.G. Hegeman[34], J.M. Heinmiller[51], A.P. Heinson[48], U. Heintz[62], C. Hensel[58], K. Herner[72], G. Hesketh[63], M.D. Hildreth[55], R. Hirosky[81], J.D. Hobbs[72], B. Hoeneisen[12], H. Hoeth[26], M. Hohlfeld[22], S.J. Hong[31], S. Hossain[75], P. Houben[34], Y. Hu[72], Z. Hubacek[10], V. Hynek[9], I. Iashvili[69], R. Illingworth[50], A.S. Ito[50], S. Jabeen[62], M. Jaffré[16], S. Jain[75], K. Jakobs[23], C. Jarvis[61], R. Jesik[43], K. Johns[45], C. Johnson[70], M. Johnson[50], A. Jonckheere[50], P. Jonsson[43], A. Juste[50], E. Kajfasz[15], A.M. Kalinin[36], J.M. Kalk[60], S. Kappler[21], D. Karmanov[38], P.A. Kasper[50], I. Katsanos[70], D. Kau[49], V. Kaushik[78], R. Kehoe[79], S. Kermiche[15], N. Khalatyan[50], A. Khanov[76], A. Kharchilava[69], Y.M. Kharzheev[36], D. Khatidze[70], T.J. Kim[31], M.H. Kirby[53], M. Kirsch[21], B. Klima[50], J.M. Kohli[27], J.-P. Konrath[23], V.M. Korablev[39], A.V. Kozelov[39], J. Kraus[65], D. Krop[54], T. Kuhl[24], A. Kumar[69], A. Kupco[11], T. Kurča[20], J. Kvita[9], F. Lacroix[13], D. Lam[55], S. Lammers[70], G. Landsberg[77], P. Lebrun[20], W.M. Lee[50], A. Leflat[38], J. Lellouch[17], J. Leveque[45], J. Li[78], L. Li[48], Q.Z. Li[50], S.M. Lietti[5], J.G.R. Lima[52], D. Lincoln[50], J. Linnemann[65], V.V. Lipaev[39], R. Lipton[50], Y. Liu[7], Z. Liu[6], A. Lobodenko[40], M. Lokajicek[11], P. Love[42], H.J. Lubatti[82], R. Luna[3], A.L. Lyon[50], A.K.A. Maciel[2], D. Mackin[80], R.J. Madaras[46], P. Mättig[26], C. Magass[21], A. Magerkurth[64], P.K. Mal[82], H.B. Malbouisson[3], S. Malik[67], V.L. Malyshev[36], H.S. Mao[50], Y. Maravin[59], B. Martin[14], R. McCarthy[72], A. Melnitchouk[66], L. Mendoza[8], P.G. Mercadante[5], M. Merkin[38], K.W. Merritt[50], A. Meyer[21], J. Meyer[22,d], T. Millet[20], J. Mitrevski[70], J. Molina[3], R.K. Mommsen[44], N.K. Mondal[29], R.W. Moore[6], T. Moulik[58], G.S. Muanza[20], M. Mulders[50], M. Mulhearn[70], O. Mundal[22], L. Mundim[3], E. Nagy[15], M. Naimuddin[50], M. Narain[77], N.A. Naumann[35], H.A. Neal[64], J.P. Negret[8], P. Neustroev[40], H. Nilsen[23], H. Nogima[3], S.F. Novaes[5], T. Nunnemann[25], V. O'Dell[50], D.C. O'Neil[6], G. Obrant[40], C. Ochando[16], D. Onoprienko[59], N. Oshima[50], N. Osman[43], J. Osta[55], R. Otec[10], G.J. Otero y Garzón[50], M. Owen[44], P. Padley[80], M. Pangilinan[77], N. Parashar[56], S.-J. Park[71], S.K. Park[31], J. Parsons[70], R. Partridge[77], N. Parua[54], A. Patwa[73], G. Pawloski[80], B. Penning[23], M. Perfilov[38], K. Peters[44], Y. Peters[26], P. Pétroff[16], M. Petteni[43], R. Piegaia[1], J. Piper[65], M.-A. Pleier[22], P.L.M. Podesta-Lerma[33,c], V.M. Podstavkov[50], Y. Pogorelov[55], M.-E. Pol[2], P. Polozov[37], B.G. Pope[65], A.V. Popov[39], C. Potter[6], W.L. Prado da Silva[3], H.B. Prosper[49], S. Protopopescu[73], J. Qian[64], A. Quadt[22,d], B. Quinn[66], A. Rakitine[42], M.S. Rangel[2], K. Ranjan[28], P.N. Ratoff[42], P. Renkel[79], S. Reucroft[63], P. Rich[44], J. Rieger[54], M. Rijssenbeek[72],

I. Ripp-Baudot[19], F. Rizatdinova[76], S. Robinson[43], R.F. Rodrigues[3], M. Rominsky[75], C. Royon[18], P. Rubinov[50], R. Ruchti[55], G. Safronov[37], G. Sajot[14], A. Sánchez-Hernández[33], M.P. Sanders[17], A. Santoro[3], G. Savage[50], L. Sawyer[60], T. Scanlon[43], D. Schaile[25], R.D. Schamberger[72], Y. Scheglov[40], H. Schellman[53], T. Schliephake[26], C. Schwanenberger[44], A. Schwartzman[68], R. Schwienhorst[65], J. Sekaric[49], H. Severini[75], E. Shabalina[51], M. Shamim[59], V. Shary[18], A.A. Shchukin[39], R.K. Shivpuri[28], V. Siccardi[19], V. Simak[10], V. Sirotenko[50], P. Skubic[75], P. Slattery[71], D. Smirnov[55], G.R. Snow[67], J. Snow[74], S. Snyder[73], S. Söldner-Rembold[44], L. Sonnenschein[17], A. Sopczak[42], M. Sosebee[78], K. Soustruznik[9], B. Spurlock[78], J. Stark[14], J. Steele[60], V. Stolin[37], D.A. Stoyanova[39], J. Strandberg[64], S. Strandberg[41], M.A. Strang[69], E. Strauss[72], M. Strauss[75], R. Ströhmer[25], D. Strom[53], L. Stutte[50], S. Sumowidagdo[49], P. Svoisky[55], A. Sznajder[3], P. Tamburello[45], A. Tanasijczuk[1], W. Taylor[6], J. Temple[45], B. Tiller[25], F. Tissandier[13], M. Titov[18], V.V. Tokmenin[36], T. Toole[61], I. Torchiani[23], T. Trefzger[24], D. Tsybychev[72], B. Tuchming[18], C. Tully[68], P.M. Tuts[70], R. Unalan[65], L. Uvarov[40], S. Uvarov[40], S. Uzunyan[52], B. Vachon[6], P.J. van den Berg[34], R. Van Kooten[54], W.M. van Leeuwen[34], N. Varelas[51], E.W. Varnes[45], I.A. Vasilyev[39], M. Vaupel[26], P. Verdier[20], L.S. Vertogradov[36], M. Verzocchi[50], M. Vetterli[6,e], F. Villeneuve-Seguier[43], P. Vint[43], P. Vokac[10], E. Von Toerne[59], M. Voutilainen[68,f], R. Wagner[68], H.D. Wahl[49], L. Wang[61], M.H.L.S. Wang[50], J. Warchol[55], G. Watts[82], M. Wayne[55], G. Weber[24], M. Weber[50], L. Welty-Rieger[54], A. Wenger[23,g], N. Wermes[22], M. Wetstein[61], A. White[78], D. Wicke[26], G.W. Wilson[58], S.J. Wimpenny[48], M. Wobisch[60], D.R. Wood[63], T.R. Wyatt[44], Y. Xie[77], S. Yacoob[53], R. Yamada[50], M. Yan[61], T. Yasuda[50], Y.A. Yatsunenko[36], K. Yip[73], H.D. Yoo[77], S.W. Youn[53], J. Yu[78], A. Zatserklyaniy[52], C. Zeitnitz[26], T. Zhao[82], B. Zhou[64], J. Zhu[72], M. Zielinski[71], D. Zieminska[54], A. Zieminski[54,‡], L. Zivkovic[70], V. Zutshi[52], and E.G. Zverev[38]

(The DØ Collaboration)

[1] *Universidad de Buenos Aires, Buenos Aires, Argentina*
[2] *LAFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil*
[3] *Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil*
[4] *Universidade Federal do ABC, Santo André, Brazil*
[5] *Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*
[6] *University of Alberta, Edmonton, Alberta, Canada, Simon Fraser University, Burnaby, British Columbia, Canada, York University, Toronto, Ontario, Canada, and McGill University, Montreal, Quebec, Canada*
[7] *University of Science and Technology of China, Hefei, People's Republic of China*
[8] *Universidad de los Andes, Bogotá, Colombia*
[9] *Center for Particle Physics, Charles University, Prague, Czech Republic*
[10] *Czech Technical University, Prague, Czech Republic*
[11] *Center for Particle Physics, Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic*
[12] *Universidad San Francisco de Quito, Quito, Ecuador*
[13] *LPC, Univ Blaise Pascal, CNRS/IN2P3, Clermont, France*
[14] *LPSC, Université Joseph Fourier Grenoble 1, CNRS/IN2P3, Institut National Polytechnique de Grenoble, France*
[15] *CPPM, IN2P3/CNRS, Université de la Méditerranée, Marseille, France*
[16] *LAL, Univ Paris-Sud, IN2P3/CNRS, Orsay, France*
[17] *LPNHE, IN2P3/CNRS, Universités Paris VI and VII, Paris, France*
[18] *DAPNIA/Service de Physique des Particules, CEA, Saclay, France*
[19] *IPHC, Université Louis Pasteur et Université de Haute Alsace, CNRS/IN2P3, Strasbourg, France*
[20] *IPNL, Université Lyon 1, CNRS/IN2P3, Villeurbanne, France and Université de Lyon, Lyon, France*
[21] *III. Physikalisches Institut A, RWTH Aachen, Aachen, Germany*
[22] *Physikalisches Institut, Universität Bonn, Bonn, Germany*
[23] *Physikalisches Institut, Universität Freiburg, Freiburg, Germany*
[24] *Institut für Physik, Universität Mainz, Mainz, Germany*
[25] *Ludwig-Maximilians-Universität München, München, Germany*
[26] *Fachbereich Physik, University of Wuppertal, Wuppertal, Germany*
[27] *Panjab University, Chandigarh, India*
[28] *Delhi University, Delhi, India*
[29] *Tata Institute of Fundamental Research, Mumbai, India*
[30] *University College Dublin, Dublin, Ireland*
[31] *Korea Detector Laboratory, Korea University, Seoul, Korea*
[32] *SungKyunKwan University, Suwon, Korea*
[33] *CINVESTAV, Mexico City, Mexico*
[34] *FOM-Institute NIKHEF and University of Amsterdam/NIKHEF, Amsterdam, The Netherlands*

[35]Radboud University Nijmegen/NIKHEF, Nijmegen, The Netherlands
[36]Joint Institute for Nuclear Research, Dubna, Russia
[37]Institute for Theoretical and Experimental Physics, Moscow, Russia
[38]Moscow State University, Moscow, Russia
[39]Institute for High Energy Physics, Protvino, Russia
[40]Petersburg Nuclear Physics Institute, St. Petersburg, Russia
[41]Lund University, Lund, Sweden, Royal Institute of Technology and Stockholm University, Stockholm, Sweden, and Uppsala University, Uppsala, Sweden
[42]Lancaster University, Lancaster, United Kingdom
[43]Imperial College, London, United Kingdom
[44]University of Manchester, Manchester, United Kingdom
[45]University of Arizona, Tucson, Arizona 85721, USA
[46]Lawrence Berkeley National Laboratory and University of California, Berkeley, California 94720, USA
[47]California State University, Fresno, California 93740, USA
[48]University of California, Riverside, California 92521, USA
[49]Florida State University, Tallahassee, Florida 32306, USA
[50]Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA
[51]University of Illinois at Chicago, Chicago, Illinois 60607, USA
[52]Northern Illinois University, DeKalb, Illinois 60115, USA
[53]Northwestern University, Evanston, Illinois 60208, USA
[54]Indiana University, Bloomington, Indiana 47405, USA
[55]University of Notre Dame, Notre Dame, Indiana 46556, USA
[56]Purdue University Calumet, Hammond, Indiana 46323, USA
[57]Iowa State University, Ames, Iowa 50011, USA
[58]University of Kansas, Lawrence, Kansas 66045, USA
[59]Kansas State University, Manhattan, Kansas 66506, USA
[60]Louisiana Tech University, Ruston, Louisiana 71272, USA
[61]University of Maryland, College Park, Maryland 20742, USA
[62]Boston University, Boston, Massachusetts 02215, USA
[63]Northeastern University, Boston, Massachusetts 02115, USA
[64]University of Michigan, Ann Arbor, Michigan 48109, USA
[65]Michigan State University, East Lansing, Michigan 48824, USA
[66]University of Mississippi, University, Mississippi 38677, USA
[67]University of Nebraska, Lincoln, Nebraska 68588, USA
[68]Princeton University, Princeton, New Jersey 08544, USA
[69]State University of New York, Buffalo, New York 14260, USA
[70]Columbia University, New York, New York 10027, USA
[71]University of Rochester, Rochester, New York 14627, USA
[72]State University of New York, Stony Brook, New York 11794, USA
[73]Brookhaven National Laboratory, Upton, New York 11973, USA
[74]Langston University, Langston, Oklahoma 73050, USA
[75]University of Oklahoma, Norman, Oklahoma 73019, USA
[76]Oklahoma State University, Stillwater, Oklahoma 74078, USA
[77]Brown University, Providence, Rhode Island 02912, USA
[78]University of Texas, Arlington, Texas 76019, USA
[79]Southern Methodist University, Dallas, Texas 75275, USA
[80]Rice University, Houston, Texas 77005, USA
[81]University of Virginia, Charlottesville, Virginia 22901, USA and
[82]University of Washington, Seattle, Washington 98195, USA

(Dated: March 5, 2008)

We present first evidence for the production of single top quarks in the D0 detector at the Fermilab Tevatron $p\bar{p}$ collider. The standard model predicts that the electroweak interaction can produce a top quark together with an antibottom quark or light quark, without the antiparticle top quark partner that is always produced from strong coupling processes. Top quarks were first observed in pair production in 1995, and since then, single top quark production has been searched for in ever larger datasets. In this analysis, we select events from a 0.9 fb^{-1} dataset that have an electron or muon and missing transverse energy from the decay of a W boson from the top quark decay, and two, three, or four jets, with one or two of the jets identified as originating from a b hadron decay. The selected events are mostly backgrounds such as W+jets and $t\bar{t}$ events, which we separate from the expected signals using three multivariate analysis techniques: boosted decision trees, Bayesian neural networks, and matrix element calculations. A binned likelihood fit of the signal cross section plus background to the data from the combination of the results from the three analysis methods gives a cross section for single top quark production of $\sigma(p\bar{p} \rightarrow tb+X,\ tqb+X) =$

4.7 ± 1.3 pb. The probability to measure a cross section at this value or higher in the absence of signal is 0.014%, corresponding to a 3.6 standard deviation significance. The measured cross section value is compatible at the 10% level with the standard model prediction for electroweak top quark production. We use the cross section measurement to directly determine the Cabibbo-Kobayashi-Maskawa quark mixing matrix element that describes the Wtb coupling and find $|V_{tb} f_1^L| = 1.31^{+0.25}_{-0.21}$, where f_1^L is a generic vector coupling. This model-independent measurement translates into $0.68 < |V_{tb}| \leq 1$ at the 95% C.L. in the standard model.

PACS numbers: 14.65.Ha; 12.15.Ji; 13.85.Qk

Contents

I. INTRODUCTION

A. Single Top Quarks

Top quarks were first observed in top quark – top antiquark pair production via the strong interaction in 1995 [1, 2]. The standard model also predicts that the electroweak interaction can produce a top quark together with a bottom antiquark or a light quark, without the antiparticle top quark partner that is always produced in strong-coupling processes. This electroweak process is generally referred to as single top quark production. Since 1995, the D0 and CDF collaborations have been searching ever larger datasets for signs of single top quark production.

We present here the results of a search for top quarks produced singly via the electroweak interaction from the decay of an off-shell W boson or fusion of a virtual W boson with a b quark. Previously measured top quarks have been produced in pairs from highly energetic virtual gluons via the strong interaction. The cross section for $t\bar{t}$ production at the Fermilab Tevatron proton-antiproton collider (center-of-mass energy = 1.96 TeV) is 6.77 ± 0.42 pb [3] at next-to-leading order (NLO) plus higher-order soft-gluon corrections, for a top quark of mass $m_{\rm top}$ = 175 GeV [4]. The standard model predicts three processes for production of a top quark without its antiparticle partner . These are as follows: (i) the s-channel process $p\bar{p}{\rightarrow}t\bar{b}+X, \bar{t}b+X$ [5, 6, 7], with a cross section of 0.88 ± 0.14 pb [8] at NLO for $m_{\rm top}$ = 175 GeV; (ii) the t-channel process $p\bar{p}{\rightarrow}tq\bar{b}+X, \bar{t}\bar{q}b+X$ [7, 9, 10, 11], with a cross section of 1.98 ± 0.30 pb [8] at the same order in perturbation theory and top quark mass; and (iii) the tW process $p\bar{p}{\rightarrow}tW^-+X, \bar{t}W^++X$ [7, 12], where the cross section at the Tevatron energy is small, 0.08 ± 0.02 pb [12] at LO.

The main tree-level Feynman diagrams for the dominant single top quark production processes are illustrated in Fig. 1. For brevity, in this paper we will use the notation "tb" to mean the sum of $t\bar{b}$ and $\bar{t}b$, and "tqb" to mean the sum of $tq\bar{b}$ and $\bar{t}\bar{q}b$. The analysis reported in this paper searches only for the s-channel process tb and the t-channel process tqb, and does not include a search for the tW process because of its small production rate at the Tevatron.

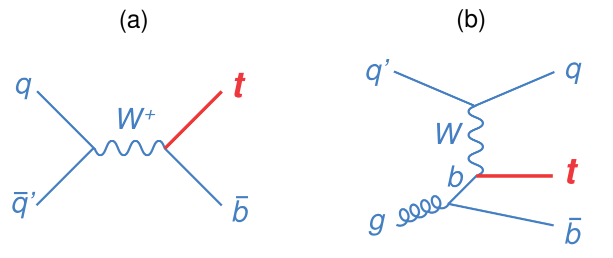

FIG. 1: Main tree-level Feynman diagrams for (a) s-channel single top quark production, and (b) t-channel production.

Top quarks are interesting particles to study since in the standard model their high mass implies a Yukawa coupling to the Higgs boson with a value near unity, unlike any other known particle. They also decay before they hadronize, allowing the properties of a bare quark such as spin to be transferred to its decay products and thus be measured and compared to the standard model predictions. Events with single top quarks can also be used to study the Wtb coupling [7, 13, 14], and to measure directly the absolute value of the quark mixing matrix (the Cabibbo-Kobayashi-Maskawa (CKM) matrix [15, 16]) element $|V_{tb}|$ without assuming there are only three generations of quarks [17, 18]. A measured value for $|V_{tb}|$ significantly different from unity could imply the existence of a fourth quark family or other effects from beyond the standard model [19].

B. Search History

The D0 collaboration has published three searches for single top quark production using smaller datasets. We analyzed 90 pb^{-1} of data from Tevatron Run I (1992–1996 at a center-of-mass energy of 1.8 TeV) which resulted in the first upper limits on single top quark production [20] and we performed a more refined search using neural networks that achieved greater sensitivity [21]. In Run II, we used 230 pb^{-1} of data collected from 2002 to 2004 to set more stringent upper limits [22, 23]. Our best published 95% C.L. upper limits are 6.4 pb in the s-channel (tb production) and 5.0 pb in the t-channel (tqb production). Students in the D0 collaboration have completed ten Ph.D. dissertations on the single top quark search [24]. Our most recent publication [25] presents first evidence for single top quark production using a 0.9 fb^{-1} dataset. We provide a more detailed description of that result here, and also include several improvements to the analysis methods that lead to a final result on the same dataset with slightly higher significance.

The CDF collaboration has published two results from analyzing 106 pb^{-1} of Run I data [26, 27], and one that uses 162 pb^{-1} of Run II data [28]. Their best 95% C.L. upper limits are 14 pb in the s-channel, 10 pb in the t-channel, and 18 pb in the s-channel and t-channel combined. Students in the CDF collaboration have completed seven Ph.D. dissertations on the single top quark search [29].

C. Search Method Overview

The experimental signal for single top quark events consists of one isolated high transverse momentum (p_T), central pseudorapidity (η [30]) charged lepton and missing transverse energy ($\not\!\!E_T$) from the decay of a W boson from the top quark decay, accompanied by a b jet from the top quark decay. There is always a second jet, which originates from a b quark produced with the top quark in the s-channel, or which comes from a forward-traveling up- or down-type quark in t-channel events. Some t-channel events have a detectable b jet from the gluon splitting to $b\bar{b}$. Since there may be significant initial-state or final-state radiation, we include in our search events with two, three, or four jets. We use data collected with triggers that include an electron or a muon, and a jet. In the electron channel, multijet events can fake signal ones when a jet is misidentified as an electron, and we have stringent identification criteria for electrons to reduce this type of background. In the muon channel, $b\bar{b}$+jets events can fake signal ones when one of the b's decays to a muon. We reject much of this background by requiring the muon to be isolated from all jets in the event. Finally, we apply a set of simple selection criteria to retain regions of phase space that single top quark events tend to populate.

We divide the selected events into 12 nonoverlapping samples, referred to as analysis channels, depending on the flavor of the lepton (e or μ), the number of jets (2, 3, 4), and the number of jets identified as originating from b quarks (number of "tagged" jets = 1, 2), because the signal-to-background ratios and fractions of expected signal in each channel differ significantly. The dominant background in most of these channels is W+jets events. We model this background using events simulated with Monte Carlo (MC) techniques and normalized to data before b tagging. We also use an MC model to simulate the background from $t\bar{t}$ events. Finally, we use data events with poorly identified leptons to model the multijet background where a jet is misidentified as an electron, or a muon in a jet from c or b decay is misidentified as a muon from a W boson decay. We apply a neural-network-based b-identification algorithm to each jet in data and keep events with one or two jets that are identified as b jets. We model this b tagging in the MC event samples by weighting each event by the probability that one or more jets is tagged.

After event selection, we calculate multivariate discriminants in each analysis channel to separate as much as possible the expected signal from the background. We then perform a binned likelihood fit of the background model plus possible signal to the data in the discriminant output distributions and combine the results from all channels that improve the expected sensitivity. Finally, we calculate the probability that our data are compatible with background only, use the excess of data over background in each bin to measure the signal cross section, and calculate the probability that the data contains both background and signal produced with at least the measured cross section value.

For each potential analysis channel, the relevant details are the signal acceptance and the signal-to-background ratio. Table I shows the percentage of the total signal acceptance for each jet multiplicity and number of b-tagged jets, and the associated signal-to-background ratios. We used this information to determine that the most sensitive channels have two, three, or four jets, and one or two b tags. In the future, it could be beneficial to extend the analysis to include events with only one jet, b tagged, since the signal-to-background ratios are not bad, and to study the untagged events with two or three jets where there is significant signal acceptance.

TABLE I: Percentage of total selected MC single top quark events for each jet multiplicity and number of b-tagged jets, and the associated signal-to-background ratios, for the electron and muon channels combined. The values shown in bold type are for the channels used in this analysis.

Analysis Channel Statistics					
	1 jet	2 jets	3 jets	4 jets	≥ 5 jets
	s-channel tb				
0 b tags	8%	19%	9%	3%	1%
	1:11,000	1:1,600	1:1,200	1:1,100	1:1,000
1 b tag	6%	**24%**	**12%**	**3%**	1%
	1:270	**1:55**	**1:73**	**1:130**	1:200
2 b tags	—	**9%**	**4%**	**1%**	0%
	—	**1:12**	**1:27**	**1:92**	1:110
	t-channel tqb				
0 b tags	10%	27%	13%	4%	1%
	1:4,400	1:520	1:400	1:360	1:300
1 b tag	6%	**20%**	**11%**	**4%**	1%
	1:150	**1:32**	**1:37**	**1:58**	1:72
2 b tags	—	**1%**	**2%**	**1%**	0%
	—	**1:100**	**1:36**	**1:65**	1:70

D. Differences from Previous Searches

We summarize here the changes and improvements made to the analysis since the previously published D0 result that used 230 pb^{-1} of data [22, 23]. The most important difference is that we have analyzed a dataset four times as large. Other changes include the following: (i) use of an improved model for the t-channel tqb signal from the package SINGLETOP [31], based on COMPHEP [32], which better reproduces NLO-like parton kinematics; (ii) use of an improved model for the $t\bar{t}$ and W+jets backgrounds from the ALPGEN package [33] that has parton-jet matching [34] implemented with PYTHIA [36] to avoid duplicate generation of some initial-state and final-state jet kinematics; (iii) determination from data of the ratio of W boson plus $b\bar{b}$ or $c\bar{c}$ jets to the total rate of W+jets production; (iv) omission of a separate calculation of the diboson backgrounds WW and WZ since they are insignificant; (v) differences in electron, muon, and jet identification requirements and minimum p_T's; (vi) use of a significantly higher efficiency b-tagging algorithm based on a neural network; (vii) splitting of the analysis by jet and b-tag multiplicity so as not to dilute the strength of high-acceptance, good signal-to-background channels by mixing them with poorer ones; (viii) simplification of the treatment of the smallest sources of systematic uncertainty (since the analysis precision is statistics dominated); (ix) use of improved multivariate techniques to separate signal from background; and (x) optimization of the search to find the combined single top quark production from both the s- and t-channels, tb+tqb.

II. THE D0 DETECTOR

The D0 detector [37] consists of three major parts: a tracking system to determine the trajectories and momenta of charged particles, a calorimeter to measure the energies of electromagnetic and hadronic showers, and a system to detect muons, which are the only charged particles that are typically not contained within the calorimeter. The first element at the core of the detector is a tracking system that consists of a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located within a 2 T superconducting solenoidal magnet. The SMT has six barrel modules in the central region, each comprising four layers arranged axially around the beam pipe, and 16 radial disks interspersed with and beyond the central barrels. Ionization charge is collected by $\approx 800,000$ p- or n-type silicon strips of pitch between 50 and 150 μm that are used to measure the positions of the hits. Tracks can be reconstructed up to pseudorapidities [30] of $|\eta^{\rm det}| \approx 3.0$.

The CFT surrounds the SMT with eight thin coaxial barrels, each supporting two doublets of overlapping scintillating fibers of 0.835 mm diameter, one doublet being parallel to the beam axis, and the other alternating by $\pm 3^\circ$ relative to the axis. Visible-light photon counters (VLPCs) collect the light signals from the fibers, achieving a cluster resolution of about 100 μm per doublet layer.

Central and forward preshower detectors contribute to the identification of electrons and photons. The central preshower detector is located just outside of the superconducting coil and the forward ones are mounted in front of the endcap calorimeters. The preshower detectors comprise several layers of scintillator strips that are read out using wavelength-shifting fibers and VLPCs.

Three finely grained uranium/liquid-argon sampling calorimeters constitute the primary system used to identify electrons, photons, and jets. The central calorimeter (CC) covers $|\eta^{\rm det}|$ up to ≈ 1.1. The two end calorimeters (EC) extend the coverage to $|\eta^{\rm det}| \approx 4.2$. Each calorimeter contains an electromagnetic (EM) section closest to the interaction region with aproximately 20 radiation lengths of material, followed by fine and coarse hadronic sections with modules that increase in size with distance from the interaction region and ensure particle containment with approximately six nuclear interaction lengths. In addition to the preshower detectors, scintillators between the CC and EC provide sampling of developing showers in the cryostat walls for $1.1 < |\eta^{\rm det}| < 1.4$.

The three-layer muon system is located beyond the calorimetry, with 1.8 T iron toroids after the first layer to provide a stand-alone muon-system momentum measurement. Each layer comprises tracking detectors and scintillation trigger counters. Proportional drift tubes 10 cm in diameter allow tracking in the region $|\eta^{\rm det}| < 1$, and 1 cm mini drift tubes extend the tracking to $|\eta^{\rm det}| < 2$.

Additionally, plastic scintillator arrays covering $2.7 < |\eta^{\rm det}| < 4.4$ are used to measure the rate of inelastic collisions in the D0 interaction region and calculate the Tevatron instantaneous and integrated luminosities.

We select the events to be studied offline with a three-tiered trigger system. The first level of the trigger makes a decision based on partial information from the tracking, calorimeter, and muon systems. The second level of the trigger uses more refined information to further reduce the rate. The third trigger level is based on software filters running in a farm of computers that have access to all information in the events.

III. TRIGGERS AND DATA

The data were collected between August 2002 and December 2005, with 913 ± 56 pb^{-1} and 871 ± 53 pb^{-1} of good quality events in the electron and muon channels respectively.

As the average instantaneous luminosity of the Tevatron has increased over time, the triggers used to collect the data have been successively changed to maintain background rejection. The requirements at the highest trigger level are the following, with the associated integrated luminosity included in parentheses:

Electron+jets triggers

1. One electron with $p_T > 15$ GeV and two jets with $p_T > 15$ GeV (103 pb^{-1})
2. One electron with $p_T > 15$ GeV and two jets with $p_T > 20$ GeV (227 pb^{-1})
3. One electron with $p_T > 15$ GeV, one jet with $p_T > 25$ GeV, and a second jet with $p_T > 20$ GeV (289 pb^{-1})
4. One electron with $p_T > 15$ GeV and two jets with $p_T > 30$ GeV (294 pb^{-1})

Muon+jets triggers

1. One lower-trigger-level muon with no p_T threshold and one jet with $p_T > 20$ GeV (107 pb^{-1})
2. One lower-trigger-level muon with no p_T threshold and one jet with $p_T > 25$ GeV (278 pb^{-1})
3. One muon with $p_T > 3$ GeV and one jet with $p_T > 30$ GeV (252 pb^{-1})
4. One isolated muon with $p_T > 3$ GeV and one jet with $p_T > 25$ GeV (21 pb^{-1})

5. One muon with $p_T > 3$ GeV and one jet with $p_T > 35$ GeV (214 pb^{-1})

The average efficiency of the electron+jets triggers is 87% for tb events and 86% for tqb events that pass the final selection cuts. The average efficiency of the muon+jets triggers is 87% for tb and 82% for tqb events.

Note that for the electron+jets triggers, the electron usually satisfies one of the jet requirements, and thus there are usually only two independent objects required in each event (one electron and one jet).

IV. EVENT RECONSTRUCTION

Physics objects are reconstructed from the digital signals recorded in each part of the detector. Particles can be identified by certain patterns and, when correlated with other objects in the same event, they provide the basis for understanding the physics that produced such signatures in the detector.

A. Primary Vertices

The location of the hard-scatter interaction point is reconstructed by means of an adaptive primary vertex algorithm [38]. This algorithm first selects tracks coming from different interactions by clustering them according to their z position along the nominal beam line. In the second step, the location and width of the beam in the transverse plane (perpendicular to the beam line) are determined and then used to re-fit tracks, and each cluster of tracks is associated with a vertex using the "adaptive" technique that gives all tracks a weight and iterates the fit. The third and last step consists of choosing the vertex that has the lowest probability of coming from a minimum bias interaction (a $p\bar{p}$ scatter event), based on the p_T values of the tracks assigned to each vertex. The hard-scatter vertex is distinguished from soft-interaction vertices by the higher average p_T of its tracks. In multijet data, the position resolution of the primary vertex in the transverse plane is around 40 μm, convoluted with a typical beam spot size of around 30 μm.

B. Electrons

Electron candidates are defined as clusters of energy depositions in the electromagnetic section of the central calorimeter ($|\eta^{\rm det}| < 1.1$) consistent in shape with an electromagnetic shower. At least 90% of the energy of the cluster must be contained in the electromagnetic section of the calorimeter, $f_{\rm EM} > 0.9$, and the cluster must satisfy the following isolation criterion:

$$\frac{E_{\rm total}(\mathcal{R} < 0.4) - E_{\rm EM}(\mathcal{R} < 0.2)}{E_{\rm EM}(\mathcal{R} < 0.2)} < 0.15, \qquad (1)$$

where E is the electron candidate's energy measured in the calorimeter, and $\mathcal{R} = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ is the radius of a cone defined by the azimuthal angle ϕ and the pseudorapidity η, centered on the electron candidate's track if there is an associated track, or the calorimeter cluster if there is not. Two classes of electrons are subsequently defined and used in this analysis:

- **Loose electron**
  A loose electron must pass the identification requirements listed above. In addition, the energy deposition in the calorimeter must be matched with a charged particle track from the tracking detectors with $p_T > 5$ GeV. Finally, a shower-shape chi-squared, based on seven variables that compare the values of the energy deposited in each layer of the electromagnetic calorimeter with average distributions from simulated electrons, has to satisfy $\chi^2_{\rm cal} < 50$.
- **Tight electron**
  A tight electron must pass the loose requirements, and have a value of a seven-variable EM-likelihood $\mathcal{L} > 0.85$. The following variables are used in the likelihood: (i) $f_{\rm EM}$; (ii) $\chi^2_{\rm cal}$; (iii) $E_T^{\rm cal}/p_T^{\rm track}$, the transverse energy of the cluster divided by the transverse momentum of the matched track; (iv) the χ^2 probability of the match between the track and the calorimeter cluster; (v) the distance of closest approach between the track and the primary vertex in the transverse plane; (vi) the number of tracks inside a cone of $\mathcal{R} = 0.05$ around the matched track; and (vii) the $\sum p_T$ of tracks within an $\mathcal{R} = 0.4$ cone around the matched track. The average tight electron identification efficiency in data is around 75%.

C. Muons

Muons are identified by combining tracks in the muon spectrometer with central detector tracks. Muons are reconstructed up to $|\eta^{\rm det}| = 2$ by first finding hits in all three layers of the muon spectrometer and requiring that the timing of these hits be consistent with the muon originating in the center of the detector from the correct proton-antiproton bunch crossing, thereby rejecting cosmic rays. Secondly, all muon candidates must be matched to a track in the central tracker, where the central track must pass the following criteria: (i) χ^2 per degree of freedom must be less than 4; and (ii) the distance of closest approach between the track and the primary vertex must be less than 0.2 mm if the track has SMT hits and less than 2 mm if it does not. Two classes of muons are then defined for this analysis:

- **Loose-isolated muon**
  A loose muon must pass the identification requirements given above. Loose muons must in addition be isolated from jets. The distance between the muon and any jet axis in the event has to satisfy $\mathcal{R}(\text{muon, jet}) > 0.5$.

- **Tight-isolated muon**
  A tight muon must pass the loose-isolation requirement and additional isolation criteria as follows: (i) the transverse momenta of all tracks within a cone of radius $\mathcal{R} = 0.5$ around the muon direction, except the track matched to the muon, must add up to less than 20% of the muon p_T; and (ii) the energy deposited in a cone of radius $0.1 < \mathcal{R} < 0.4$ around the muon direction must be less that 20% of the muon p_T.

D. Jets

We reconstruct jets based on calorimeter cell energies, using the midpoint cone algorithm [39] with radius $\mathcal{R} = 0.5$. Noisy calorimeter cells are ignored in the reconstruction algorithm by only selecting cells whose energy is at least four standard deviations above the average electronic noise and any adjacent cell with at least two standard deviations above the average electronic noise.

To reject poor quality or noisy jets, we require all jets to have the following: (i) $0.05 < f_{\rm EM} < 0.95$ in the central region, with the lower cut looser in the intercryostat and forward regions; (ii) fraction of jet p_T in the coarse hadronic calorimeter layers < 0.4 in the central region, with looser requirements in the forward regions; and (iii) at least 50% of the p_T of the jet, not including the coarse hadronic layers, matched to energy depositions in towers in Level 1 of the trigger in a cone of radius $\mathcal{R} = 0.5$ around the jet axis in the central region, with looser requirements in the forward regions.

Jet energy scale corrections are applied to convert reconstructed jet energies into particle-level energies. The energy of each jet containing a muon within $\mathcal{R}$(muon, jet) < 0.5 (considered to originate from a semileptonic c- or b-quark decay) is corrected to account for the energy of the muon and the accompanying neutrino (because that energy is not deposited in the calorimeter and so would otherwise be undermeasured). For this correction, it is assumed that the neutrino has the same energy as the muon.

Jets that have the same η and ϕ as a reconstructed electron are removed from the list of jets to avoid double-counting objects.

E. Missing Transverse Energy

Neutrinos carry away momentum that can be inferred using momentum conservation in the transverse plane. The sum of the transverse momenta of undetected neutrinos is equal to the negative of the sum of the transverse momenta of all particles observed in the detector. In practice, we compute the missing transverse energy by adding up vectorially the transverse energies in all cells of the electromagnetic and fine hadronic calorimeters. Cells in the coarse hadronic calorimeter are only added if they are part of a jet. This raw quantity is then corrected for the energy corrections applied to the reconstructed objects and for the momentum of all muons in the event, corrected for their energy loss in the calorimeter.

F. *b* Jets

Given that single top quark events have at least one b jet in the final state, we use a b-jet tagger to identify jets originating from b quarks. In addition to the jet quality criteria described in previous sections, a "taggability" requirement is applied. This requires the jets to have at least two good quality tracks with $p_T > 1$ GeV and $p_T > 0.5$ GeV respectively, that include SMT hits and which point to a common origin. A neural network (NN) tagging algorithm is used to identify jets originating from a b quark. The tagger and its performance in the data is described in detail in Ref. [40]. We summarize briefly here its main characteristics. The NN tagger uses the following variables, ranked in order of separation power, to discriminate b jets from other jets: (i) decay length significance of the secondary vertex reconstructed by the secondary vertex tagger (SVT); (ii) weighted combination of the tracks' impact parameter significances; (iii) jet lifetime probability (JLIP), the probability that the jet originates from the primary vertex; (iv) χ^2 per degree of freedom of the SVT secondary vertex; (v) number of tracks used to reconstruct the secondary vertex; (vi) mass of the secondary vertex; and (vii) number of secondary vertices found inside the jet.

For this analysis, we require the NN output to be greater than 0.775 for the jet to be considered b tagged. The average probability for a light jet in data to be falsely tagged at this operating point is 0.5%, and the average b-tagging efficiency in data is 47% for jets with $|\eta^{\rm det}| < 2.4$.

V. SIMULATED EVENT SAMPLES

A. Event Generation

For this analysis, we generate single top quark events with the COMPHEP-SINGLETOP [31, 32] Monte Carlo event generator. SINGLETOP produces events whose kinematic distributions match those from NLO calculations [8]. The top quark mass is set to 175 GeV, the set of parton distribution functions (PDF) is CTEQ6L1 [41], and the renormalization and factorization scales are $m^2_{\rm top}$ for the s-channel and $(m_{\rm top}/2)^2$ for the t-channel. These scales are chosen such that the LO cross sections are closest to the NLO cross sections [42]. The top quarks and the W bosons from the top quark decays are decayed in SINGLETOP to ensure the spins are properly transfered. PYTHIA [36] is used to add the underlying event, initial-state and final-state radiation, and for hadronization. TAUOLA [43] is used to decay tau leptons, and EVTGEN [44] to decay b hadrons. To calculate the expected number of signal events, these samples are normalized to the NLO cross sections [8] for a top quark mass of 175 GeV: 0.88 ± 0.14 pb for the s-channel and 1.98 ± 0.30 pb for the t-channel.

The W+jets and $t\bar{t}$ samples are generated using ALPGEN [33]. The version we use includes a parton-jet matching algorithm that follows the MLM prescription [34, 35]. For the $t\bar{t}$ samples, the top quark mass is set to 175 GeV, the scale is $m^2_{\rm top} + \sum p^2_T({\rm jets})$, and the PDF set is CTEQ6L1. For the W+jets events, the PDF is also CTEQ6L1 and the scale is $m^2_W + p^2_T(W)$. The W+jets events include separate generation of each jet multiplicity from $W+$ 0 light partons to $W+$ at least 5 light partons for events with no heavy-flavor partons (we refer to these samples as Wjj). Those with $b\bar{b}$ and $c\bar{c}$ partons have separately generated samples with between 0 and 3 additional light partons. The $t\bar{t}$ events include separate samples with additional jets from 0 to 2 light partons.

For the W+jets sets, we remove events with heavy flavor jets added by PYTHIA so as not to duplicate the phase space of those generated already by ALPGEN. The Wcj subprocesses are included in the Wjj sample with massless charm quarks.

Since the W+jets background is normalized to data (see Sec. VII A), it implicitly includes all sources of W+jets, Z+jets, and diboson events with similar jet-flavor composition, in particular Z+jets events where one of the leptons from the Z boson decay is not identified.

The proportions of $Wb\bar{b}$ and $Wc\bar{c}$ in the W+jets model are set by ALPGEN at leading order precision. However, higher order calculations [45, 46, 47] indicate that there should be a higher fraction of events with heavy-flavor jets. We measure a scale factor for the $Wb\bar{b}$ and $Wc\bar{c}$ subsamples using several untagged data samples (with zero b-tagged jets) that have negligible signal content. We obtain:

$$\begin{gathered} \alpha(N_{Wb\bar{b}} + N_{Wc\bar{c}}) + N_{Wjj} + N_{t\bar{t}} + N_{\rm multijets} = N^{\rm zero\text{-}tag}_{\rm data} \\ \alpha = 1.50 \pm 0.45 \end{gathered} \tag{2}$$

where the numbers of events N_i for each background component correspond to the expected number of events after event selection (described in Sec. VI) and background normalization (described in Sec. VII) and removing events with one or more b-tagged jets. Additionally, we check that the same value of $\alpha = 1.5$ is obtained from the complementary W + 1 jet sample, where we require the only jet to be b tagged. Figure 2 illustrates the measurement of the scale factor α.

We examine the distributions expected to suffer the largest shape dependence from higher order corrections, such as the invariant mass of the two leading jets and the p_T of the b-tagged jet, and find good agreement between the shapes of the data and the background model, not only in the signal region, but also in samples enriched with W+jets events.

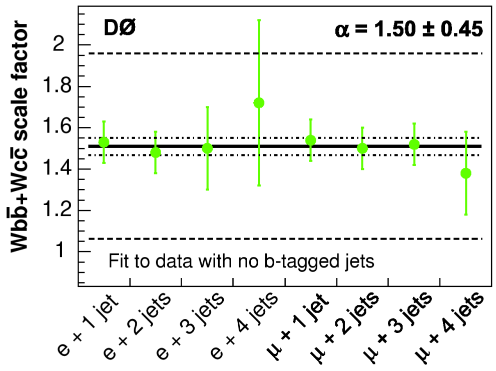

FIG. 2: Measurements of the scale factor α used to convert the fraction of $Wb\bar{b}$ and $Wc\bar{c}$ events in the W+jets background model from leading order to higher order. The points are the measured correction factor in each dataset. The solid line is the average of these values. The dot-dash inner band shows the uncertainty from the fit to the eight data points. The dashed outer line shows the uncertainty on α used in the analysis to allow for the assumption that the scale factor should be the same for $Wb\bar{b}$ and $Wc\bar{c}$, and for small differences in the shapes of distributions between the W + heavy flavor and W + light flavor jets.

Table II shows the cross sections, branching fractions, initial numbers of events, and integrated luminosities of the simulated samples used in this analysis.

B. Correction Factors

We pass the simulated events through a GEANT-based model [48] of the D0 detector. The simulated samples then have correction factors applied to ensure that the reconstruction and selection efficiencies match those found in data. Generally the efficiency to reconstruct, identify, and select objects in the simulated samples is higher than in data, so the following scale factors are used to correct for that difference:

- **Trigger efficiency correction factors**
  The probability for each simulated event to fire the triggers detailed in Sec. III is calculated as a weight applied to each object measured in the event. Electron and jet efficiencies, for all levels of the trigger architecture, are parametrized as functions of p_T and $\eta^{\rm det}$. Muon efficiencies are parametrized as functions of $\eta^{\rm det}$ and ϕ. These corrections are measured using data obtained with triggers different from those used in this search to avoid biases.
- **Electron identification efficiency correction factors**
  We correct each simulated event in the electron channel with a factor that accounts for the differences in electron cluster finding identification,

TABLE II: Cross sections, branching fractions, initial numbers of events, and integrated luminosities of the simulated event samples. Here, ℓ means e, μ, and τ.

	Statistics of the Simulated Samples			
Event Type	Cross Section [pb]	Branching Fraction	Number of Events	Integrated Luminosity [fb^{-1}]
Signals				
$tb \to e$+jets	0.88 ± 0.14	0.1111 ± 0.0022	92,620	947
$tb \to \mu$+jets	0.88 ± 0.14	0.1111 ± 0.0022	122,346	1,251
$tb \to \tau$+jets	0.88 ± 0.14	0.1111 ± 0.0022	76,433	782
$tqb \to e$+jets	1.98 ± 0.30	0.1111 ± 0.0022	130,068	591
$tqb \to \mu$+jets	1.98 ± 0.30	0.1111 ± 0.0022	137,824	626
$tqb \to \tau$+jets	1.98 ± 0.30	0.1111 ± 0.0022	117,079	532
Signal total	$\mathbf{2.86 \pm 0.45}$	$\mathbf{0.3333 \pm 0.0067}$	**676,370**	
Backgrounds				
$t\bar{t} \to \ell$+jets	6.8 ± 1.2	0.4444 ± 0.0089	474,405	157
$t\bar{t} \to \ell\ell$	6.8 ± 1.2	0.1111 ± 0.0089	468,126	620
Top pairs total	$\mathbf{6.8 \pm 1.2}$	$\mathbf{0.5555 \pm 0.0111}$	**942,531**	
$Wb\bar{b} \to \ell\nu bb$	142	0.3333 ± 0.0066	1,335,146	28
$Wc\bar{c} \to \ell\nu cc$	583	0.3333 ± 0.0066	1,522,767	8
$Wjj \to \ell\nu jj$	$18,734$	0.3333 ± 0.0066	8,201,446	1
W+jets total	$\mathbf{19,459}$	$\mathbf{0.3333 \pm 0.0067}$	**11,059,359**	

$f_{\rm EM}$, and isolation efficiencies in the simulation and data. This correction factor is measured in $Z \to ee$ data and simulated events, and parametrized as a function of $\eta^{\rm det}$. A second scale factor is applied to account for the differences between the data and the simulation in the $\chi^2_{\rm cal}$, track matching, and EM-likelihood efficiencies. This second scale factor is also derived from $Z \to ee$ data and simulated events and parametrized as a fuction of $\eta^{\rm det}$ and $\phi^{\rm det}$.

- **Muon identification and isolation efficiency correction factors**
  We correct each simulated event in the muon channel for the muon identification, track match, and isolation efficiencies. The identification correction factor is parametrized as a function of $\eta^{\rm det}$ and ϕ, track match as a function of track-z and $\eta^{\rm det}$, and isolation as a function of the number of jets in the event. These corrections are measured in $Z \to \mu\mu$ data and simulated events.

- **Jet reconstruction efficiency and energy resolution correction factors**
  Simulated jets need to be corrected for differences in the reconstruction and identification efficiency and for the worse energy resolution found in data than in the simulation. The jet energy scale correction is applied to the simulation as in the data, but then simulated jets are corrected for the jet reconstruction efficiency and smeared to match the jet energy resolution found in back-to-back photon+jet events.

- **Taggability and b-tagging efficiency correction factors**
  In data, the taggability and b-tagging requirements are applied directly, as described in Sec. IV F. For simulated samples, *taggability-rate functions* and *tag-rate functions* are applied instead of the direct selection because the modeling of the detector is not sufficiently accurate. The taggability-rate function is parametrized in jet p_T, η, and primary vertex z, and is measured in the selected data sample (Sec. VI) with one loose-isolated lepton. We check that the efficency is the same as in the data sample with one tight-isolated lepton within the uncertainties. The average taggability for central high-p_T jets is around 90%.

  The b-jet efficiency correction is measured in data using a muon-in-jet sample and a b-jet enriched subset where one jet is required to have a small JLIP value, and in an admixture of $Z \to b\bar{b}$ and $t\bar{t}$ simulated events where the b-jets are required to contain a muon. The b-tag efficiency correction for c-quark jets is derived in a combined MC sample with Z boson, multijets, and $t\bar{t}$ decays to c quarks, and assuming that the MC-to-data scale factor is the same as for the b-jet efficiency. The b-tag efficiency correction for light jets is derived from multijet data. All these b-tagging corrections are parametrized as functions of the jet p_T and η. Figure 3 illustrates the tag-rate functions used in this analysis.

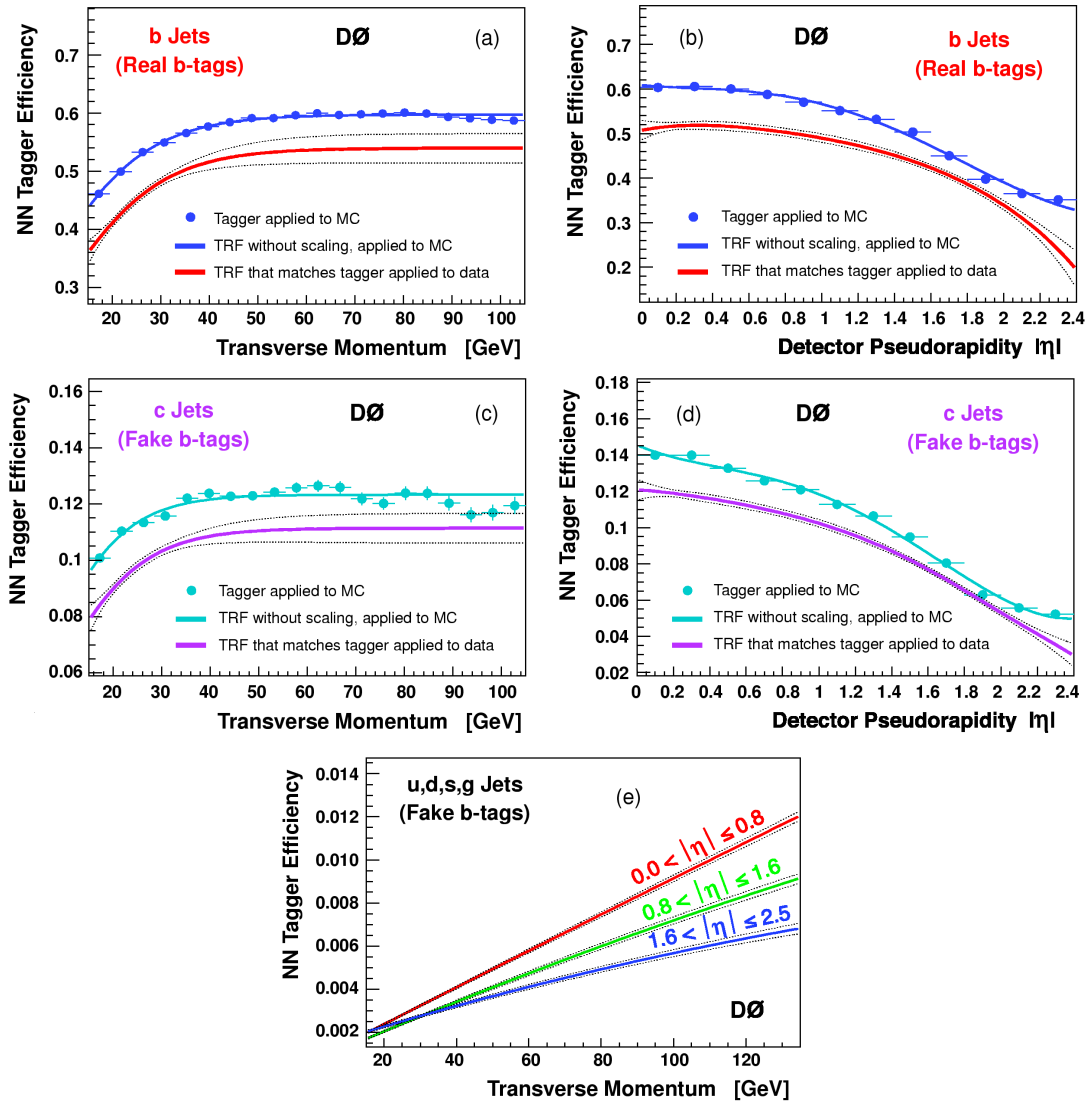

FIG. 3: The tag-rate functions (TRFs) used to weight the MC events according to the probability that they should be b tagged. In plots (a)–(d), the points show the neural network b tagging algorithm (the "tagger") applied directly to the MC events. The upper line that passes through the points is the result of the tag-rate functions, before scaling-to-data, being applied to the MC events to reproduce the result from the tagger. The lower line, with dotted error band, shows the tag-rate functions after they have been scaled to match the efficiency of the NN b tagging algorithm applied to data. In plot (e), the lines show the (scaled) tag-rate functions that are applied to MC events.

VI. EVENT SELECTION

A. Selection Requirements

We apply a loose event selection to find W-like events containing an isolated lepton, missing transverse energy, and two to four jets with high transverse momentum. The samples after this selection, which we call "pretagged," (i.e., *before* tagging has been applied), are dominated by W+jets events, with some $t\bar{t}$ contribution that becomes more significant for higher jet multiplicities. The final selection improves the signal-to-background ratio significantly by requiring the presence of one or two b-tagged jets.

Common selections for both e and μ channels

- Good quality (for data with all subdetectors working properly)
- Pass trigger: offline electrons and muons in the data are matched to the object that fired the appropriate trigger for that run period
- Good primary vertex: $|z_{\rm PV}| < 60$ cm with at least three tracks attached
- Missing transverse energy: $15 < \not\!\!E_T < 200$ GeV
- Two, three, or four jets with $p_T > 15$ GeV and $|\eta| < 3.4$
- Leading jet $p_T > 25$ GeV and $|\eta| < 2.5$
- Second leading jet $p_T > 20$ GeV
- Jet triangle cut $|\Delta\phi(\text{leading jet}, \not\!\!E_T)|$ vs. $\not\!\!E_T$ (see Ref. [23]):
  - from 1.5 to π radians when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 35 GeV when $|\Delta\phi| = \pi$ rad
- One or two b-tagged jets

Electron channel selection

- Only one tight electron with $p_T > 15$ GeV and $|\eta^{\rm det}| < 1.1$
- No tight muon with $p_T > 18$ GeV and $|\eta^{\rm det}| < 2.0$
- No second loose electron with $p_T > 15$ GeV and any $\eta^{\rm det}$
- Electron coming from the primary vertex: $|\Delta z(e, {\rm PV})| < 1$ cm
- Electron triangle cuts $|\Delta\phi(e, \not\!\!E_T)|$ vs. $\not\!\!E_T$ (see Ref. [23]):
  - from 2 to 0 rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 40 GeV when $|\Delta\phi| = 0$ rad
  - from 1.5 to 0 rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 50 GeV when $|\Delta\phi| = 0$ rad
  - from 2 to π rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 24 GeV when $|\Delta\phi| = \pi$ rad

Muon channel selection

- Only one tight muon with $p_T > 18$ GeV and $|\eta^{\rm det}| < 2.0$
- No tight electron with $p_T > 15$ GeV and $|\eta^{\rm det}| < 2.5$
- Muon coming from the primary vertex: $|\Delta z(\mu, {\rm PV})| < 1$ cm
- Muon triangle cuts $|\Delta\phi(\mu, \not\!\!E_T)|$ vs. $\not\!\!E_T$ (see Ref. [23]):
  - from 1.1 to 0 rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 80 GeV when $|\Delta\phi| = 0$ rad
  - from 1.5 to 0 rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 50 GeV when $|\Delta\phi| = 0$ rad
  - from 2.5 to π rad when $\not\!\!E_T = 0$ GeV, and $\not\!\!E_T$ from 0 to 30 GeV when $|\Delta\phi| = \pi$ rad

Some of the selection criteria listed above are designed to remove areas of the data that are difficult to model. In particular, the upper $\not\!\!E_T$ selection gets rid of a few events where the muon p_T fluctuated to a large value. The "triangle cuts" are very efficient in removing multijet events where a misreconstructed jet creates fake missing energy aligned or anti-aligned in azimuth with the lepton or jet.

Background-data selection for measuring the multijet background

- All the same selection criteria as listed above except for the tight lepton requirements
- Electron channel — only one loose-but-not-tight electron
- Muon channel — only one loose-but-not-tight muon

The definitions of loose and tight electrons and muons are in Secs. IV B and IV C.

B. Numbers of Events After Selection

Table III shows the numbers of events in the signal and background samples and in the data after applying the selection criteria. Note that these numbers are just counts of events used later in the analysis, and not signal or background yields after normalizations and corrections have been applied.

VII. BACKGROUND MODEL

A. W+Jets and Multijets Backgrounds

The W+jets background is modeled using the parton-jet matched ALPGEN simulated samples described in Sec. V. This background is normalized to data before b tagging, using a procedure explained below. Because we normalize to data and do not use theory cross sections, small components of the total background from Z+jets and diboson processes (WW, WZ, and ZZ) are implicitly included in the W+jets part of the background model. This simplification does not affect the final

TABLE III: Numbers of events for the electron and muon channels after selection. The MC samples include events coming from τ decays, $\tau \to \ell\nu$ where $\ell = e$ in the electron channel and $\ell = \mu$ in the muon channel.

Numbers of Events After Selection										
	Electron Channel					Muon Channel				
	1 jet	2 jets	3 jets	4 jets	≥ 5 jets	1 jet	2 jets	3 jets	4 jets	5 jets
Signal MC										
tb	6,908	19,465	9,127	2,483	595	3,878	12,852	6,458	1,809	401
tqb	8,971	22,758	12,080	3,797	1,092	8,195	21,066	11,193	3,489	835
Background MC										
$t\bar{t}\to\ell\ell$	7,671	29,537	26,042	12,068	5,396	5,509	24,595	21,803	9,788	3,442
$t\bar{t}\to\ell$+jets	522	5,659	22,477	27,319	14,298	232	3,376	16,293	22,680	8,658
$Wb\bar{b}$	26,611	13,914	9,011	3,848	1,434	27,764	14,488	9,427	3,874	1,204
$Wc\bar{c}$	21,765	13,453	7,562	2,252	591	32,712	19,047	10,141	3,051	663
Wjj	134,660	61,497	34,162	8,290	1,750	147,842	66,201	36,673	9,169	1,502
Pretag data										
Multijets	11,565	6,993	4,043	1,317	431	897	658	462	151	48
Signal data	27,370	8,220	3,075	874	223	17,816	6,432	2,590	727	173
One-tag data										
Multijets	246	322	226	93	34	31	51	49	21	8
Signal data	445	357	207	97	35	289	287	179	100	38
Two-tags data										
Multijets	—	12	15	14	7	—	3	4	1	4
Signal data	—	30	37	22	10	—	23	32	27	10

results because of the low rate from these processes in the final selected dataset, and because the kinematics of the events are similar to those in W+jets events. They are thus identified together with W+jets events by the multivariate discriminants.

The multijet background is modeled using datasets that contain misidentified leptons, as described at the end of Sec. VI A. These datasets provide the shape for the multijet background component in each analysis channel. They are normalized to data as part of the W+jets normalization process.

We normalize the W+jets and multijet backgrounds to data before tagging using the matrix method [49], which lets us estimate how many events in the pretagged samples contain a misidentified lepton (originating from multijet production) and how many events have a real isolated lepton (originating from W+jets or $t\bar{t}$). Two data samples are defined, the *tight* sample, which is the signal sample after all selection cuts have been applied, and the *loose* sample, where the same selection has been applied but requiring only loose lepton quality. The tight data sample, with $N_{\rm tight}$ events, is a subset of the loose data sample with $N_{\rm loose}$ events. The loose sample contains $N_{\rm loose}^{\text{real-}\ell}$ events with a real lepton (signal-like events, mostly W+jets and $t\bar{t}$) and $N_{\rm loose}^{\text{fake-}\ell}$ fake lepton events, which is the number of multijet events in the loose sample.

We measure the probability $\varepsilon_{\text{real-}\ell}$ for a real isolated lepton to pass the tight lepton selection in $Z \to \ell\ell$ data events. The probability for a fake-isolated lepton to pass the tight-isolated lepton criteria, $\varepsilon_{\text{fake-}\ell}$, is measured in a sample enriched in multijet events with the same selection as the signal data but requiring $\not\!\!E_T < 10$ GeV. In the electron channel, these probabilities are parametrized as $\varepsilon_{\text{real-e}}(p_T, \eta)$ and $\varepsilon_{\text{fake-e}}(N_{\rm jets}, \text{trigger period})$. In the muon channel, they are parametrized as $\varepsilon_{\text{real-}\mu}(N_{\rm jets}, p_T)$ and $\varepsilon_{\text{fake-}\mu}(\eta)$. With these definitions, the matrix method is applied using the following two equations:

$$N_{\rm loose} = N_{\rm loose}^{\text{fake-}\ell} + N_{\rm loose}^{\text{real-}\ell} \quad (3)$$

$$\begin{aligned} N_{\rm tight} &= N_{\rm tight}^{\text{fake-}\ell} + N_{\rm tight}^{\text{real-}\ell} \\ &= \varepsilon_{\text{fake-}\ell}\, N_{\rm loose}^{\text{fake-}\ell} + \varepsilon_{\text{real-}\ell}\, N_{\rm loose}^{\text{real-}\ell}, \end{aligned} \quad (4)$$

and solving for $N_{\rm loose}^{\text{fake-}\ell}$ and $N_{\rm loose}^{\text{real-}\ell}$ so that the multijet and the W-like contributions in the tight sample $N_{\rm tight}^{\text{fake-}\ell}$ and $N_{\rm tight}^{\text{real-}\ell}$ can be determined.

The results of the matrix method normalization, which we apply separately in each jet multiplicity bin, are shown in Table IV. The values shown for $\varepsilon_{\text{real-}\ell}$ and $\varepsilon_{\text{fake-}\ell}$ are averages for illustration only. The pretagged background-data sample is scaled to $N_{\rm tight}^{\text{fake-}\ell}$, and the W+jets simulated samples ($Wb\bar{b}$+$Wc\bar{c}$+Wjj) are scaled to $N_{\rm tight}^{\text{real-}\ell}$, after subtracting the expected number of $t\bar{t}$ events in each jet multiplicity bin of the tight sample. These normalization factors are illustrated in Fig. 4.

TABLE IV: Matrix method normalization values in the electron and muon channels for the loose and tight selected samples, and the expected contribution from multijet and W-like events.

Normalization of W+Jets and Multijets Backgrounds to Data										
	Electron Channel					Muon Channel				
	1 jet	2 jets	3 jets	4 jets	$\geq$ 5 jets	1 jet	2 jets	3 jets	4 jets	5 jets
N_{loose}	38,935	15,213	7,118	2,191	654	18,714	7,092	3,054	878	221
N_{tight}	27,370	8,220	3,075	874	223	17,816	6,432	2,590	727	173
$\varepsilon_{\text{real-}\ell}$	0.873	0.874	0.874	0.875	0.875	0.991	0.989	0.987	0.961	0.878
$\varepsilon_{\text{fake-}\ell}$	0.177	0.193	0.188	0.173	0.173	0.408	0.358	0.342	0.309	0.253
$N_{\text{tight}}^{\text{fake-}\ell}$	1,691	1,433	860	256	86	498	329	223	56	10
$N_{\text{tight}}^{\text{real-}\ell}$	25,679	6,787	2,215	618	137	17,319	6,105	2,369	669	162

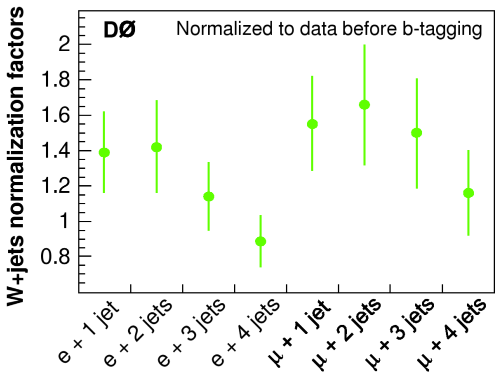

FIG. 4: The factors used to normalize the W+jets background model to pretagged data in each analysis channel.

B. Top-Quark Pairs Background

Background from the $t\bar{t}$ process is modeled using the parton-jet matched ALPGEN simulated samples described in Sec. V. These events are normalized to the theoretical cross section [3] at $m_{\text{top}} = 175$ GeV (chosen to match the value used to generate the samples), which is 6.8 pb.

VIII. SIGNAL ACCEPTANCES

Table V shows the percentage of each signal that remains after selection. We achieve roughly 30% higher acceptances in this analysis compared to our previously published analysis [22, 23] from the use of the more efficient neural network b-tagging algorithm. The total acceptance for the s-channel tb process is $(3.2 \pm 0.4)\%$ and for the t-channel tqb process it is $(2.1 \pm 0.3)\%$.

IX. EVENT YIELDS

We use the term "yield" to mean the number of events of the signal or background in question predicted to be in the 0.9 fb^{-1} of data analyzed here. Tables VI, VII, and VIII show these yields for all signals and backgrounds separated by lepton flavor and jet multiplicity within each table, and by the numbers of b-tagged jets between the tables. Because the W+jets and multijet backgrounds are normalized to data before tagging, the sum of the backgrounds is constrained to equal the number of events observed in the data, as seen in the first table. The yield values shown in these tables have been rounded to integers for clarity, so that the sums of the components will not always equal exactly the values given for these sums. All calculations however have been done with full-precision values.

Table IX summarizes the signals, summed backgrounds, and data for each channel, showing the uncertainties on the signals and backgrounds, and the signal-to-background ratios. Table X shows the signal and background yields summed over electron and muon channels and 1- and 2-tagged jets in the 2-jet, 3-jet, and 4-jet bins, and for the 2-, 3-, and 4-jet bins combined.

Some basic kinematic distributions are shown for electron channel events in Fig. 5 and for muon channel events in Fig. 6. Since the yields are normalized before b tagging, in each case the pretagged distributions are shown in the first row of distributions and the one-tag distributions are shown in the second row.

TABLE V: Signal acceptances after selection.

	Signal Acceptances									
	Electron Channel					Muon Channel				
	1 jet	2 jets	3 jets	4 jets	$\geq$ 5 jets	1 jet	2 jets	3 jets	4 jets	5 jets
Before tagging										
tb	0.55%	1.77%	0.83%	0.23%	0.06%	0.33%	1.36%	0.69%	0.19%	0.05%
tqb	0.52%	1.49%	0.79%	0.25%	0.07%	0.36%	1.17%	0.64%	0.20%	0.05%
One-tag										
tb	0.24%	0.82%	0.39%	0.11%	0.03%	0.15%	0.64%	0.32%	0.09%	0.02%
tqb	0.18%	0.61%	0.34%	0.11%	0.03%	0.13%	0.50%	0.28%	0.09%	0.02%
Two-tags										
tb	—	0.29%	0.14%	0.04%	0.02%	—	0.24%	0.12%	0.03%	0.01%
tqb	—	0.02%	0.05%	0.02%	0.01%	—	0.01%	0.04%	0.02%	0.01%

TABLE VI: Predicted yields after selection and before b tagging.

	Yields Before b-Tagging									
	Electron Channel					Muon Channel				
	1 jet	2 jets	3 jets	4 jets	$\geq$ 5 jets	1 jet	2 jets	3 jets	4 jets	5 jets
Signals										
tb	4	14	7	2	0	3	10	5	1	0
tqb	9	27	14	5	1	6	20	11	3	1
Backgrounds										
$t\bar{t}\rightarrow\ell\ell$	9	35	28	10	4	5	27	22	8	3
$t\bar{t}\rightarrow\ell$+jets	2	26	103	128	67	1	14	71	99	43
$Wb\bar{b}$	659	358	149	42	5	431	312	161	47	10
$Wc\bar{c}$	1,592	931	389	93	10	1,405	1,028	523	131	21
Wjj	23,417	5,437	1,546	343	51	15,476	4,723	1,591	385	85
Multijets	1,691	1,433	860	256	86	498	329	223	58	10
Background Sum	27,370	8,220	3,075	874	223	17,816	6,434	2,592	727	172
Data	27,370	8,220	3,075	874	223	17,816	6,432	2,590	727	173

TABLE VII: Predicted yields after selection for events with exactly one b-tagged jet.

	Yields With One b-Tagged Jet									
	Electron Channel					Muon Channel				
	1 jet	2 jets	3 jets	4 jets	$\geq$ 5 jets	1 jet	2 jets	3 jets	4 jets	5 jets
Signals										
tb	2	7	3	1	0	1	5	2	1	0
tqb	3	11	6	2	1	2	9	5	2	0
Backgrounds										
$t\bar{t}\to\ell\ell$	4	16	13	5	2	2	13	10	4	1
$t\bar{t}\to\ell$+jets	1	11	47	58	30	0	6	32	45	20
$Wb\bar{b}$	188	120	50	14	2	131	110	56	16	4
$Wc\bar{c}$	81	74	36	9	1	64	74	46	13	2
Wjj	175	61	20	5	1	125	58	23	6	2
Multijets	36	66	48	18	7	17	26	24	8	2
Background Sum	484	348	213	110	43	340	286	191	93	30
Data	445	357	207	97	35	289	287	179	100	38

TABLE VIII: Predicted yields after selection for events with exactly two b-tagged jets.

	Yields With Two b-Tagged Jets									
	Electron Channel					Muon Channel				
	1 jet	2 jets	3 jets	4 jets	$\geq$ 5 jets	1 jet	2 jets	3 jets	4 jets	5 jets
Signals										
tb	—	2.3	1.1	0.3	0.1	—	1.9	0.9	0.3	0.1
tqb	—	0.3	0.8	0.4	0.2	—	0.2	0.7	0.4	0.1
Backgrounds										
$t\bar{t}\to\ell\ell$	—	5.5	4.6	1.7	0.7	—	4.6	3.8	1.4	0.5
$t\bar{t}\to\ell$+jets	—	1.7	13.6	21.8	11.7	—	1.0	10.2	18.0	8.1
$Wb\bar{b}$	—	16.2	6.8	1.8	0.3	—	15.3	8.2	2.3	0.6
$Wc\bar{c}$	—	1.6	1.1	0.4	0.1	—	1.6	1.5	0.5	0.1
Wjj	—	0.1	0.1	0.0	0.0	—	0.1	0.1	0.0	0.0
Multijets	—	2.5	3.2	2.7	1.4	—	1.5	1.9	0.4	0.8
Background Sum	—	27.5	29.4	28.4	14.2	—	24.1	25.7	22.7	10.1
Data	—	30	37	22	10	—	23	32	27	10

TABLE IX: Summed signal and background yields after selection with total uncertainties, the numbers of data events, and the signal-to-background ratio in each analysis channel. Note that the signal includes both s-channel and t-channel single top quark processes.

	Summary of Yields with Uncertainties							
	Electron Channel				Muon Channel			
	1 jet	2 jets	3 jets	4 jets	1 jet	2 jets	3 jets	4 jets
Zero-tag								
Signal Sum	9 ± 2	21 ± 4	10 ± 2	3 ± 1	5 ± 1	15 ± 3	7 ± 2	3 ± 1
Bkgd Sum	$26,886 \pm 626$	$7,845 \pm 336$	$2,832 \pm 144$	735 ± 60	$17,476 \pm 515$	$6,124 \pm 351$	$2,375 \pm 178$	610 ± 50
Data	29,925	7,833	2,831	752	17,527	6,122	2,378	599
Signal:Bkgd	1:3,104	1:378	1:286	1:259	1:3,253	1:407	1:320	1:292
One-tag								
Signal Sum	5 ± 1	18 ± 3	9 ± 2	3 ± 1	3 ± 1	14 ± 3	7 ± 2	2 ± 1
Bkgd Sum	484 ± 86	348 ± 61	213 ± 30	110 ± 16	340 ± 63	286 ± 58	191 ± 34	93 ± 15
Data	445	357	207	97	289	287	179	100
Signal:Bkgd	1:95	1:20	1:23	1:38	1:101	1:21	1:26	1:42
Two-tags								
Signal Sum	—	2.6 ± 0.6	1.9 ± 0.4	0.7 ± 0.2	—	2.1 ± 0.5	1.6 ± 0.4	0.6 ± 0.2
Bkgd Sum	—	27.5 ± 6.5	29.4 ± 5.7	28.4 ± 6.0	—	24.1 ± 6.1	25.7 ± 5.5	22.7 ± 5.4
Data	—	30	37	22	—	23	32	27
Signal:Bkgd	—	1:10	1:15	1:39	—	1:12	1:16	1:37

TABLE X: Yields after selection for the analysis channels combined.

	Summed Yields			
	e+μ + 1+2 tags			
	2 jets	3 jets	4 jets	2,3,4 jets
Signals				
tb	16±3	8±2	2±1	25±6
tqb	20±4	12±3	4±1	37±8
Backgrounds				
$t\bar{t}\to\ell\ell$	39±9	32±7	11±3	82±19
$t\bar{t}\to\ell$+jets	20±5	103±25	143±33	266±63
$Wb\bar{b}$	261±55	120±24	35±7	416±87
$Wc\bar{c}$	151±31	85±17	23±5	259±53
Wjj	119±25	43±9	12±2	174±36
Multijets	95±19	77±15	29±6	202±39
Background Sum	686±131	460±75	253±42	1,398±248
Backgrounds+Signals	721±132	480±76	260±43	1,461±251
Data	697	455	246	1,398

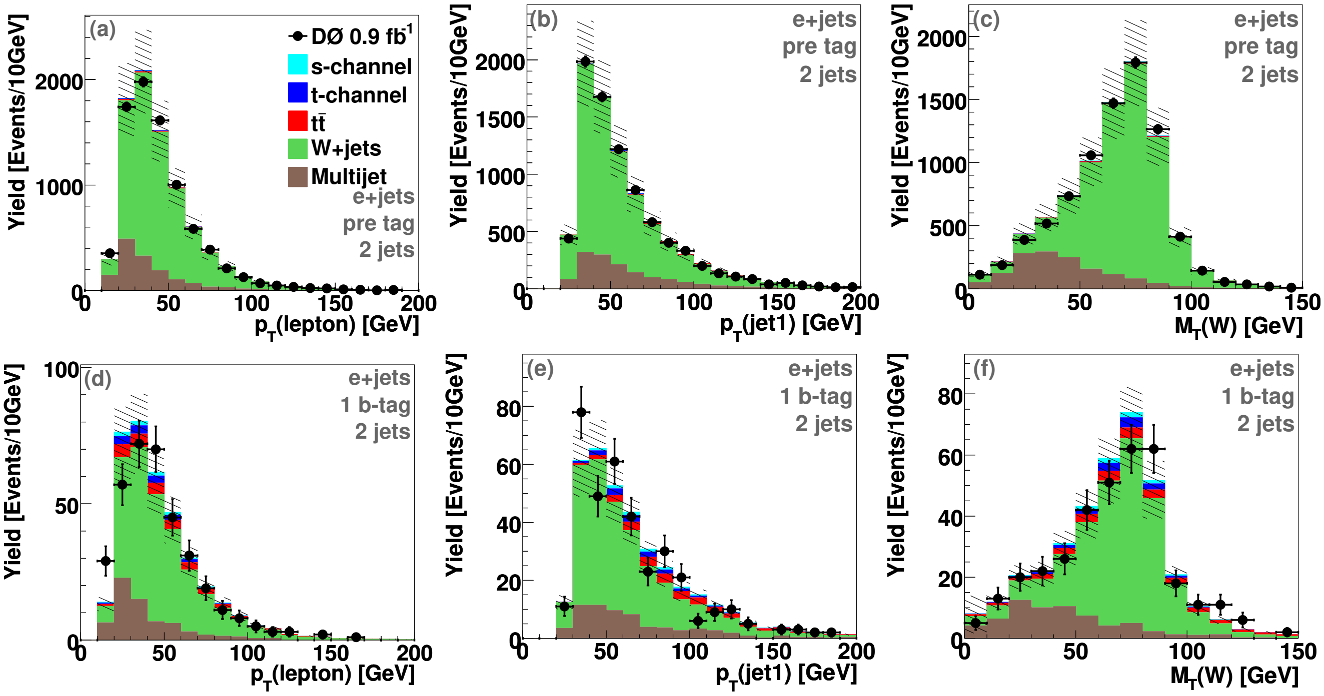

FIG. 5: The first row shows pretagged distributions for the p_T of the electron, the p_T of the leading jet, and the reconstructed W boson transverse mass. The second row shows the same distributions after tagging for events with exactly one b-tagged jet. The hatched area is the $\pm 1\sigma$ uncertainty on the total background prediction.

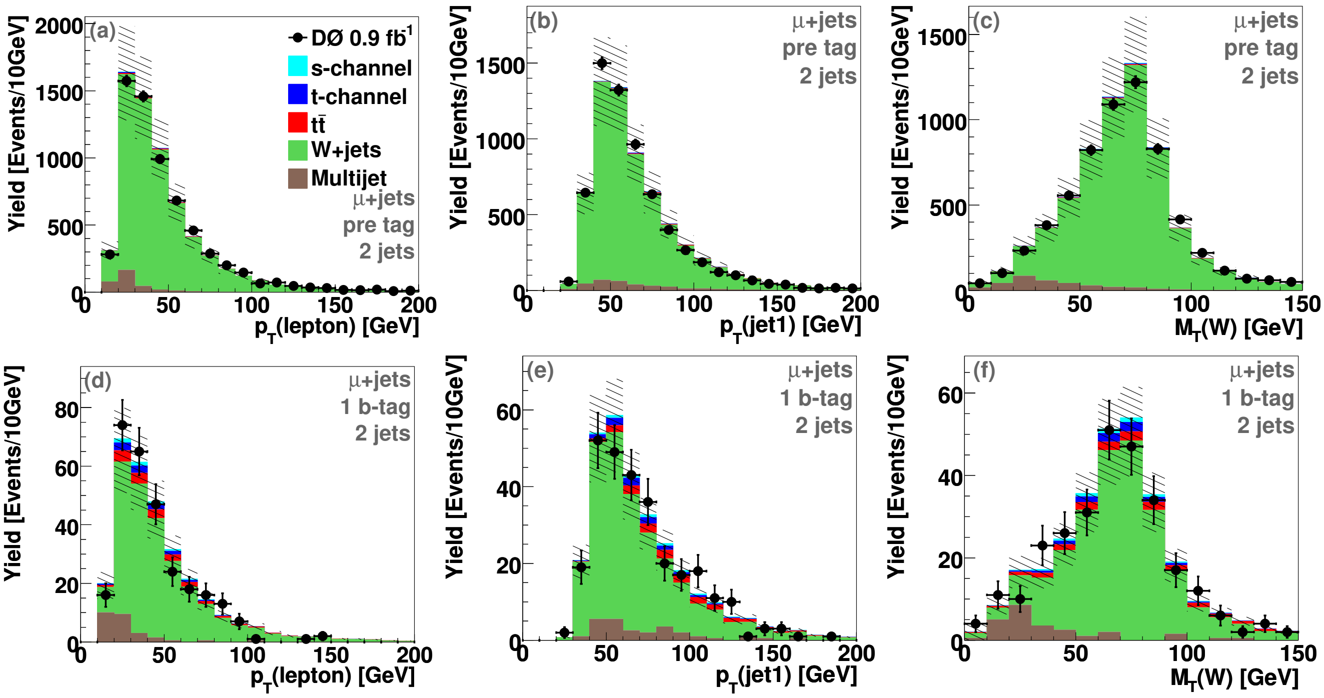

FIG. 6: The first row shows pretagged distributions for the p_T of the muon, the p_T of the leading jet, and the reconstructed W boson transverse mass. The second row shows the same distributions after tagging for events with exactly one b-tagged jet. The hatched area is the $\pm 1\sigma$ uncertainty on the total background prediction.

X. SYSTEMATIC UNCERTAINTIES

We consider several sources of systematic uncertainty in this analysis and propagate them separately for each signal and background source throughout the calculation. Systematic uncertainties enter the analysis in two ways: as uncertainty on the normalization of the background samples and as effects that change the shapes of distributions for the backgrounds and the expected signals. The effect of these uncertainties on the discriminant outputs and how they affect the cross section measurement is described in Sec. XVIII B. Table XI summarizes the relative uncertainties on each of the sources described below.

The first uncertainties listed here affect only the $t\bar{t}$ background normalization.

- **Integrated luminosity**
  At 6.1% [50], this is a small contribution to the $t\bar{t}$ yield uncertainty.
- **Theoretical cross section**
  The uncertainty on the $t\bar{t}$ cross section includes components for the choice of scale and PDF, and also, more significantly, a large component from the top quark mass uncertainty (i.e., using 175 GeV in this analysis when the latest world average value is 170.9 ± 1.8 GeV). The combined uncertainty on the cross section is taken as 18%.

The following uncertainties arise from the correction factors and functions applied to the simulated samples to make them match data, and thus affect both the signal acceptances and the $t\bar{t}$ background yield.

- **Trigger efficiency**
  Functions that represent the trigger efficiency for each object type and trigger level as a function of p_T, $\eta^{\rm det}$, and ϕ are used to weight simulated events. The functions are shifted up and down by one standard deviation of the statistical error arising from the data samples used to calculate the functions and the weight of each event is recalculated. Fixed uncertainties of 3% in the electron channel and 6% in the muon channel are chosen since they encompass all the small variations seen in each analysis channel.
- **Primary vertex selection efficiency**
  The primary vertex selection efficiency in data and the simulation are not the same. We assign a systematic uncertainty of 3% for the difference between the beam profile along the longitudinal direction in data and the simulated distribution.
- **Electron reconstruction and basic identification efficiency**
  The electron reconstruction and basic identification correction factors are parametrized as a function of $\eta^{\rm det}$. The 2% uncertainty in the efficiency accounts for its dependence on variables other than $\eta^{\rm det}$, and as a result of limited data statistics used to determine the correction factors.
- **Electron shower shape, track match, and likelihood efficiency**
  The electron shower shape, track match, and likelihood correction factors are parametrized as a function of $\eta^{\rm det}$ and ϕ. The 5% uncertainty in the efficiency accounts for the dependence on other variables, such as the number of jets and the instantaneous luminosity, and as a result of limited data statistics in determining these correction factors.
- **Muon reconstruction and identification efficiency**
  The correction factor uncertainty of 7% includes contributions from the method used to determine the correction functions, from the background subtraction, and from the limited statistics in the parametrization as a function of the $\eta^{\rm det}$ and ϕ of the muon.
- **Muon track matching and isolation**
  The muon tracking correction functions have an uncertainty that includes contributions from the method used to measure the functions, from the background subtraction, luminosity and timing bias, and from averaging over ϕ and the limited statistics in each bin used to calculate the functions. The muon isolation correction uncertainty is estimated based on its dependence on the number of jets, and covers the dependences not taken into account such as p_T and $\eta^{\rm det}$. The overall value of these uncertainties combined is 2%.
- **Jet fragmentation**
  This systematic uncertainty covers the lack of certainty in the jet fragmentation model (and is measured as the difference between PYTHIA and HERWIG [51] fragmentation) as well as the uncertainty in the modeling of initial-state and final-state radiation. It is 5% for tb and tqb and 7% for $t\bar{t}$.
- **Jet reconstruction and identification**
  The efficiency to reconstruct jets is similar in data and simulated events, but the efficiency of the simulated jets is nevertheless corrected by a parametrization of this discrepancy as a function of jet p_T. We assign a 2% error to the parametrization based on the statistics of the data sample.
- **Jet energy scale and jet energy resolution**
  The jet energy scale (JES) is raised and lowered by one standard deviation of the uncertainty on it and the whole analysis repeated. In the data, the JES uncertainty contains the jet energy resolution uncertainty. But in the simulation, the jet energy resolution uncertainty is not taken into account in the JES uncertainty. To account for this, the energy smearing in the simulated samples is varied by the size of the jet energy resolution. This uncertainty affects the acceptance and the shapes

of the distributions. The value of this uncertainty varies from 1% to 20%, depending on the analysis channel, with typical values between 6% and 10%.

The uncertainty on the W+jets and multijets background yields comes from the normalization to data. The W+jets yield is 100% anticorrelated with the multijets yield.

- **Matrix-method normalization**
  The determination of the number of real-lepton events in data is affected by the uncertainties associated with the determination of the probabilities for a loose lepton to be (mis)identified as a (fake) real lepton, $\varepsilon_{\text{fake-}\ell}$ and $\varepsilon_{\text{real-}\ell}$. The normalization is also affected by the limited statistics of the data sample as described in Sec. VII A. The combined uncertainties on the W+jets and multijets yields vary between 17% and 28%, depending on the analysis channel.
- **Heavy flavor ratio**
  The uncertainty on the scale factor applied to set the $Wb\bar{b}$ and $Wc\bar{c}$ fractions of the W+jets sample, as described in Sec. V, is estimated to cover several effects: dependence on the b-quark p_T, the difference between the zero-tag samples where it is estimated and the signal samples where it is used, and the intrinsic uncertainty on the value of the LO cross section it is being applied to. This uncertainty is 30%. It is included in the matrix method uncertainty described above.

There is one source of uncertainty that affects the signal acceptances, and both the $t\bar{t}$ and W+jets background yields.

- **b-tag modeling**
  The uncertainty associated with the taggability-rate and tag-rate functions is evaluated by raising and lowering the tag rate by one standard deviation separately for both the taggability and the tag rate components and determining the new event tagging weight. These uncertainties originate from several sources as follows: statistics of the simulated event sets; the assumed fraction of heavy flavor in the simulated multijet sample used for the mistag rate determination; and the choice of parametrizations. The b-tag modeling uncertainty varies from 2% to 16%, depending on the analysis channel, and we include the variation on distribution shapes, as well as on sample normalization.

TABLE XI: Summary of the relative systematic uncertainties. The ranges shown represent the different samples and channels.

Relative Systematic Uncertainties	
Integrated luminosity	6%
$t\bar{t}$ cross section	18%
Electron trigger	3%
Muon trigger	6%
Primary vertex	3%
Electron reconstruction & identification	2%
Electron track match & likelihood	5%
Muon reconstruction & identification	7%
Muon track match & isolation	2%
Jet fragmentation	(5–7)%
Jet reconstruction and identification	2%
Jet energy scale	(1–20)%
Tag-rate functions	(2–16)%
Matrix-method normalization	(17–28)%
Heavy flavor ratio	30%
$\varepsilon_{\text{real-e}}$	2%
$\varepsilon_{\text{real-}\mu}$	2%
$\varepsilon_{\text{fake-e}}$	(3–40)%
$\varepsilon_{\text{fake-}\mu}$	(2–15)%

XI. MULTIVARIATE ANALYSES

The search for single top quark production is significantly more challenging than the search for $t\bar{t}$ production. The principal reasons are the smaller signal-to-background ratio for single top quarks and the large overlap between the signal distributions and those of the backgrounds. We therefore concluded from the outset that optimal signal-background discrimination would be necessary to have any chance of extracting a single top quark signal from the available dataset.

Optimal event discrimination is a well-defined problem with a well-defined and unique solution. Given the probability

$$p(S|\mathbf{x}) = \frac{p(\mathbf{x}|S)p(S)}{p(\mathbf{x}|S)p(S) + p(\mathbf{x}|B)p(B)} \tag{5}$$

that an event described by the variables $\mathbf{x}$ is of the signal class, S, the signal can be extracted optimally, that is, with the smallest possible uncertainty [52], by weighting events with $p(S|\mathbf{x})$, or, as we have done, by fitting the sum of distributions of $p(S|\mathbf{x})$ for signal and background to data, as described in Sec. XVIII. In practice, since *any* one-to-one function of $p(S|\mathbf{x})$ is equivalent to $p(S|\mathbf{x})$, it is sufficient to construct an approximation to the discriminant

$$D(\mathbf{x}) = \frac{p(\mathbf{x}|S)}{p(\mathbf{x}|S) + p(\mathbf{x}|B)} \tag{6}$$

built using equal numbers of signal and background events, that is, with $p(S) = p(B)$. Each of the three analyses we have undertaken is based on a different numerical method to approximate the discriminant $D(\mathbf{x})$. From this perspective, they are conceptually identical.

In this paper, we present results from three different multivariate techniques applied to the selected dataset: boosted decision trees (DT) in Sec. XII, Bayesian neural networks (BNN) in Sec. XIII, and matrix elements (ME) in Sec. XIV. The DT analysis approximates the discriminant $D(\mathbf{x})$ using an average of many piecewise approximations to $D(\mathbf{x})$. The BNN analysis uses nonlinear functions that approximate $D(\mathbf{x})$ directly, that is, without first approximating the densities $p(\mathbf{x}|S)$ and $p(\mathbf{x}|B)$. The ME method approximates the densities $p(\mathbf{x}|S)$ and $p(\mathbf{x}|B)$ semi-analytically, starting with leading-order matrix elements, and computes $D(\mathbf{x})$ from them.

The three analyses also differ by the choice of variables used. The basic observables are:

1. missing transverse energy 2-vector ($\not\!\!E_T, \phi$),
2. lepton 4-vector (E_T, η, ϕ), assuming massless leptons,
3. jet 4-vector (E_T, η, ϕ), assuming massless jets, and jet-type, that is, whether it is a b jet or not, for each jet.

These, essentially, are the observables used in the matrix element analysis. The other analyses, however, make use of physically motivated variables [53, 54] derived from the fundamental observables. Of course, the derived variables contain no more information than is contained in the original degrees of freedom. However, for some numerical approximation methods, it may prove easier to construct an accurate approximation to $D(\mathbf{x})$ if it is built using carefully chosen derived variables than one constructed directly in terms of the underlying degrees of freedom. It may also happen that a set of judiciously chosen derived variables, perhaps one larger than the set of fundamental observables, yields better performing discriminants simply because the numerical approximation algorithm is better behaved or converges faster.

The complete set of variables used in the DT and BNN analyses is shown in Table XII. Jets are sorted in p_T and index 1 refers to the leading jet in a jet category: "jet$n$" (n=1,2,3,4) corresponds to each jet in the event, "tag$n$" to b-tagged jets, "untag$n$" to non-b-tagged jets, and "notbest$n$" to all but the best jet. The "best" jet is defined as the one for which the invariant mass $M(W,\text{jet})$ is closest to $m_{\text{top}} = 175$ GeV.

Aplanarity, sphericity, and centrality are variables that describe the direction and shape of the momentum flow in the events [55]. The variable H is the scalar sum of the energy in an event for the jets as shown. H_T is the scalar sum of the transverse energy of the objects in the event. M is the invariant mass of various combinations of objects. M_T is the transverse mass of those objects. Q is the charge of the electron or muon.

A selection of these variables is shown in Figs. 7, 8, and 9 for the sum of all channels: electron plus muon channels, two to four jets, and one or two b-tagged jets. Figure 10 shows distributions for some of the variables from Table XII for SM signals and the background components, normalized to unit area, so that the differences in shapes may be seen.

TABLE XII: Variables used with the decision trees and Bayesian neural networks analyses, in three categories: object kinematics; event kinematics; and angular variables. For the angular variables, the subscript indicates the reference frame. $\star$ indicates variables that were only used for the DT analysis. $\dagger$ indicates variables only used by the BNN analysis.

DT and BNN Input Variables	
Object Kinematics	**Event Kinematics**
$p_T(\text{jet1})$	Aplanarity(alljets,W)
$p_T(\text{jet2})$	Sphericity(alljets,W)
$p_T(\text{jet3})$	Centrality(alljets)†
$p_T(\text{jet4})$	$\not\!\!E_T$
$p_T(\text{best})$	$H(\text{alljets})^\dagger$
$p_T(\text{notbest1})^\star$	$H(\text{jet1,jet2})^\dagger$
$p_T(\text{notbest2})^\star$	$H_T(\text{alljets})$
$p_T(\text{tag1})$	$H_T(\text{alljets}-\text{best})^\star$
$p_T(\text{untag1})$	$H_T(\text{alljets}-\text{tag1})^\star$
$p_T(\text{untag2})$	$H_T(\text{alljets},W)$
$p_T(\ell)^\dagger$	$H_T(\text{jet1,jet2})$
	$H_T(\text{jet1,jet2},W)$
Angular Variables	$M(\text{alljets})$
$\cos(\text{jet1,alljets})_{\text{alljets}}$	$M(\text{alljets}-\text{best})^\star$
$\cos(\text{jet2,alljets})_{\text{alljets}}$	$M(\text{alljets}-\text{tag1})^\star$
$\cos(\text{notbest1,alljets})_{\text{alljets}}$	$M(\text{jet1,jet2})$
$\cos(\text{tag1,alljets})^\star_{\text{alljets}}$	$M(\text{jet1,jet2},W)$
$\cos(\text{untag1,alljets})_{\text{alljets}}$	$M(W,\text{best})$
$\cos(\text{best,notbest1})_{\text{besttop}}$	(i.e., "best" m_{top})
$\cos(\text{best},\ell)_{\text{besttop}}$	$M(W,\text{tag1})$
$\cos(\text{notbest1},\ell)_{\text{besttop}}$	(i.e., "b-tagged" m_{top})
$\cos(Q(\ell)\times\hat{z}(\ell),\ell)_{\text{besttop}}$	$M_T(\text{jet1,jet2})$
$\cos(\text{besttop}_{\text{CMframe}},\ell_{\text{besttop}})$	$p_T(W)^\dagger$
$\cos(\text{jet1},\ell)_{\text{btaggedtop}}$	$p_T(\text{alljets}-\text{best})$
$\cos(\text{jet2},\ell)_{\text{btaggedtop}}$	$p_T(\text{alljets}-\text{tag1})$
$\cos(\text{tag1},\ell)_{\text{btaggedtop}}$	$p_T(\text{jet1,jet2})$
$\cos(\text{untag1},\ell)_{\text{btaggedtop}}$	$M_T(W)$
$\cos(\text{btaggedtop}_{\text{CMframe}},\ell_{\text{btaggedtop}})$	$Q(\ell)\times\eta(\text{untag1})$
$\cos(\text{jet1},\ell)^\dagger_{\text{lab}}$	$\sqrt{\hat{s}}$
$\cos(\text{jet2},\ell)^\dagger_{\text{lab}}$	
$\cos(\text{best},\ell)^\dagger_{\text{lab}}$	
$\cos(\text{tag1},\ell)^\dagger_{\text{lab}}$	
$\mathcal{R}(\text{jet1,jet2})^\star$	

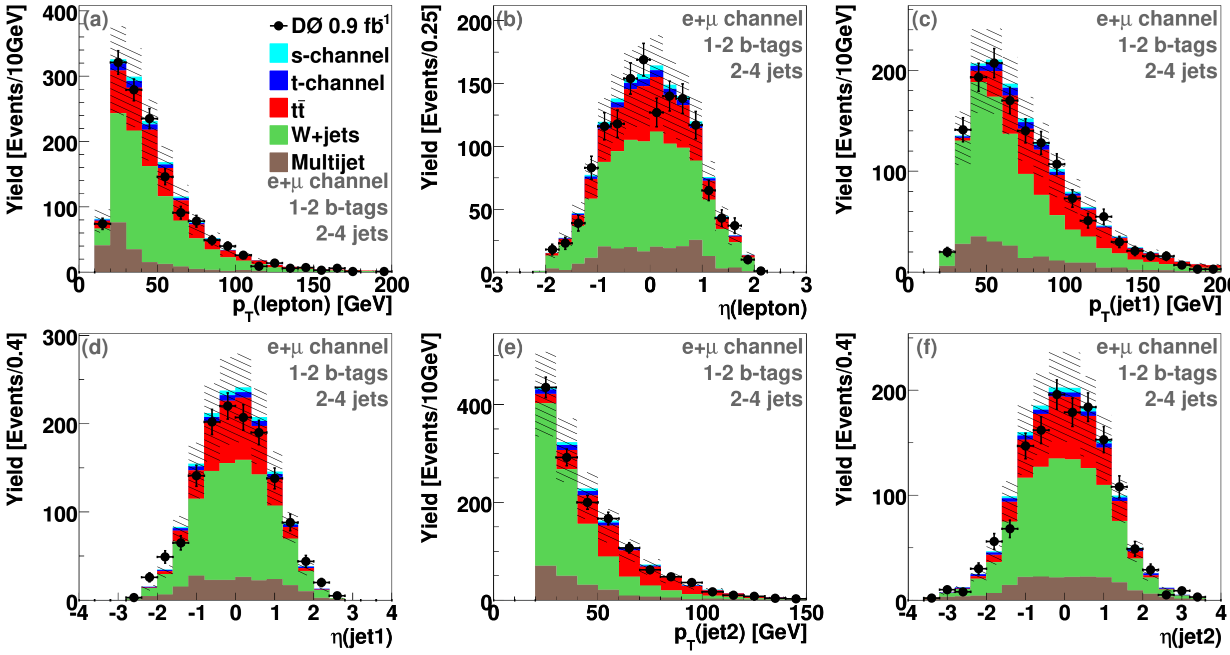

FIG. 7: Comparison of SM signal, backgrounds, and data after selection and requiring at least one b-tagged jet for six discriminating individual object variables. Electron and muon channels are combined. The plots show (a) the lepton transverse momentum and (b) pseudorapidity, (c) the leading jet transverse momentum and (d) pseudorapidity, (e) the second leading jet transverse momentum and (f) pseudorapidity. The hatched area is the $\pm 1\sigma$ uncertainty on the total background prediction.

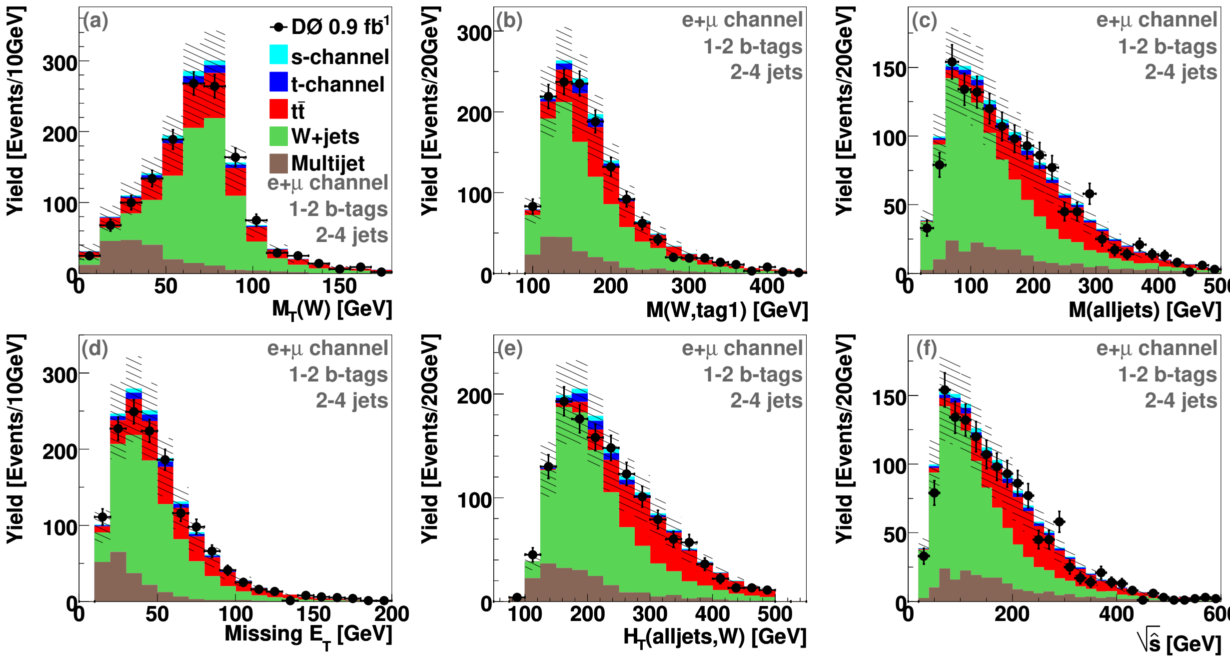

FIG. 8: Comparison of SM signal, backgrounds, and data after selection and requiring at least one b-tagged jet for six discriminating event kinematic variables. Electron and muon channels are combined. Shown are (a) the invariant transverse mass of the reconstructed W boson, (b) the invariant mass of the b-tagged jet and the W boson, (c) the invariant mass of all jets, (d) the missing transverse energy, (e) the scalar sum of the transverse momenta of jets, lepton and neutrino, (f) the invariant mass of all final state objects. The hatched area is the $\pm 1\sigma$ uncertainty on the total background prediction.

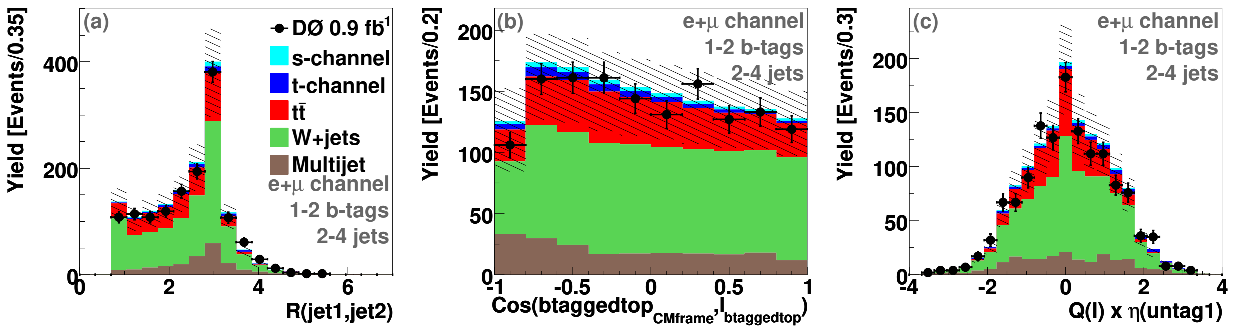

FIG. 9: Comparison of SM signal, backgrounds, and data after selection and requiring at least one b-tagged jet for three discriminating angular correlation variables. Electron and muon channels are combined. Shown are (a) the angular separation between the two leading jets, (b) the cosine of the angle between the reconstructed b-tagged top quark in the center-of-mass rest frame and the lepton in the b-tagged top quark rest frame, and (c) the charge of the lepton multiplied by the pseudorapidity of the leading untagged jet. The hatched area is the $\pm 1\sigma$ uncertainty on the total background prediction.

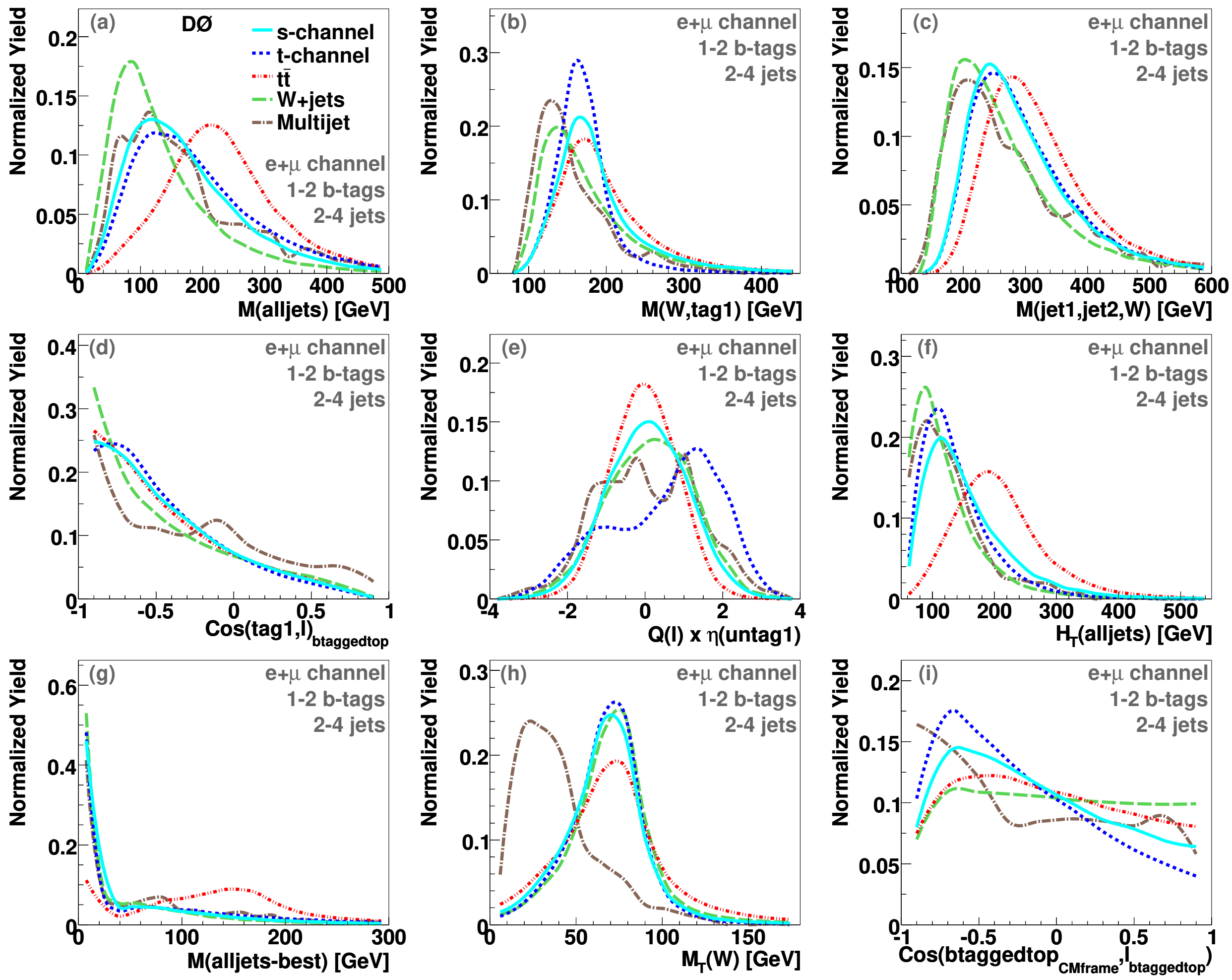

FIG. 10: Shape comparison between the s- and t-channel signals and the backgrounds in the most discriminating variables for the decision tree analysis chosen from the e+2jets/1tag channel. Shown are (a) the invariant mass of all jets, (b) the invariant mass of the b-tagged jet and the W boson, (c) the invariant mass of the two leading jets and the W boson, (d) the cosine of the b-tagged jet and the lepton in the reconstructed b-tagged top quark rest frame, (e) the charge of the lepton multiplied by the pseudorapidity of the leading untagged jet, (f) the scalar sum of the transverse momenta of all the jets, (g) the invariant mass of all jets minus the best jet, (h) the invariant transverse mass of the reconstructed W boson, and (i) the cosine of the angle between the reconstructed b-tagged top quark in the center-of-mass rest frame and the lepton in the b-tagged top quark rest frame. All histograms are normalized to unit area.

XII. BOOSTED DECISION TREES ANALYSIS

A decision tree [56, 57] employs a machine-learning technique that effectively extends a simple cut-based analysis into a multivariate algorithm with a continuous discriminant output. Boosting is a process that can be used on any weak classifier. In this analysis, we apply the boosting procedure to decision trees in order to enhance separation of signal and background.

A. Decision Tree Algorithm

A decision tree classifies events based on a set of cumulative selection criteria (cuts) that define several disjoint subsets of events, each with a different signal purity. The decision tree is built by creating two *branches* at every nonterminal node, i.e., splitting the sample of events under consideration into two subsets based on the most discriminating selection criterion for that sample. Terminal nodes are called *leaves* and each leaf has an assigned purity value p. A simple decision tree is illustrated in Fig. 11. An event defined by variables $\mathbf{x}$ will follow a unique path through the decision tree and end up in a leaf. The associated purity p of this leaf is the decision tree discriminant output for the event: $D(\mathbf{x}) = p$, with $D(\mathbf{x})$ given in Eq. 6.

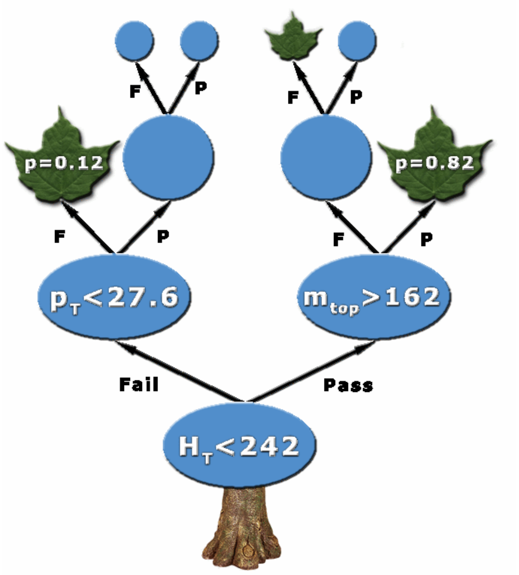

FIG. 11: Graphical representation of a decision tree. Nodes with their associated splitting test are shown as (blue) circles and terminal nodes with their purity values are shown as (green) leaves. An event defined by variables $\mathbf{x_i}$ of which $H_T < 242$ GeV and $m_{\rm top} > 162$ GeV will return $D(\mathbf{x_i}) = 0.82$, and an event with variables $\mathbf{x_j}$ of which $H_T \geq 242$ GeV and $p_T \geq 27.6$ GeV will have $D(\mathbf{x_j}) = 0.12$. All nodes continue to be split until they become leaves. (color online)

One of the primary advantages of decision trees over a cut-based analysis is that events which fail an individual cut continue to be considered by the algorithm. Limitations of decision trees include the instability of the tree structure with respect to the training sample composition, and the piecewise nature of the output. Training on different samples may produce very different trees with similar separation power. The discrete output comes from the fact that the only possible values are the purities of each leaf and the number of leaves is finite.

Decision tree techniques have interesting features, as follows: the tree has a human-readable structure, making it possible to know why a particular event was labeled signal or background; training is fast compared to neural networks; decision trees can use discrete variables directly; and no preprocessing of input variables is necessary. In addition, decision trees are relatively insensitive to extra variables: unlike neural networks, adding well-modeled variables that are not powerful discriminators does not degrade the performance of the decision tree (no additional noise is added to the system).

1. Training

The process in which a decision tree is created is usually referred to as *decision tree training*. Consider a sample of known signal and background events where each event is defined by a weight w and a list of variables $\mathbf{x}$. The following algorithm can be applied to such a sample in order to create a decision tree:

1. Initially normalize the signal training sample to the background training sample such that $\sum w_{\rm signal} = \sum w_{\rm background}$.
2. Create the first node, containing the full sample.
3. Sort events according to each variable in turn. For each variable, the splitting value that gives the best signal-background separation is found (more on this in the next section). If no split that improves the separation is found, the node becomes a leaf.
4. The variable and split value giving the best separation are selected and the events in the node are divided into two subsamples depending on whether they pass or fail the split criterion. These subsamples define two new child nodes.
5. If the statistics are too low in any node, it becomes a leaf.
6. Apply the algorithm recursively from Step 3 until all nodes have been turned into leaves.

Each of the leaves is assigned the purity value

$$p = \frac{s}{s+b}, \tag{7}$$

where s (b) is the weighted sum of signal (background) events in the leaf. This value is an approximation of the discriminant $D(\mathbf{x})$ defined in Eq. 6.

2. Splitting a Node

Consider an impurity measure $i(t)$ for node t. Desirable features of such a function are that it should be maximal for an equal mix of signal and background (no separation), minimal for nodes with either only signal or only background events (perfect separation), symmetric in signal and background purity, and strictly concave in order to reward purer nodes. Several such functions exist in the literature. We have not found a significant advantage to any specific choice and hence use the common "Gini index [58]."

The impurity measure, or Gini index, is defined as:

$$i_{\mathrm{Gini}} = 2p(1-p) = \frac{2sb}{(s+b)^2}, \tag{8}$$

where s (b) is the sum of signal (background) weights in a node. One can now define the decrease of impurity (goodness of split) associated with a split S of node t into children t_P and t_F:

$$\Delta i_{\mathrm{Gini}}(S,t) = i_{\mathrm{Gini}}(t) - p_P \cdot i_{\mathrm{Gini}}(t_P) - p_F \cdot i_{\mathrm{Gini}}(t_F), \tag{9}$$

where p_P (p_F) is the fraction of events that passed (failed) split S. The goal is to find the split S^* that maximizes the decrease of impurity, which corresponds to finding the split that minimizes the overall tree impurity.

3. Boosting

A powerful technique to improve the performance of any weak classifier (defined as any classifier that does a little better than random guessing) was introduced a decade ago: boosting [59]. Boosting was recently used in high energy physics with decision trees by the MiniBooNE experiment [60, 61].

The basic principle of boosted decision trees is to train a tree, minimize some error function, and create a tree T_{n+1} as a modification of tree T_n. The boosting algorithm used in D0's single top quark search is adaptive boosting, known in the literature as AdaBoost [59].

Once a tree indexed by n is built with associated discriminant $D_n(\mathbf{x})$, its associated error ϵ_n is calculated as the sum of the weights of the *misclassified* events. An event is considered misclassified if $|D_n(\mathbf{x}) - y| > 0.5$ where y is 1 for a signal event and 0 for background. The tree weight is calculated according to

$$\alpha_n = \beta \times \ln \frac{1-\epsilon_n}{\epsilon_n}, \tag{10}$$

where β is the boosting parameter. For each misclassified event, its weight w_i is scaled by the factor e^{α_n} (which will be greater than 1). Hence misclassified events will get higher weights. A new tree indexed by $n+1$ is created from the reweighted training sample now working harder on the previously misclassified events. This is repeated N times, where N, the number of boosting cycles, is a parameter specified by the user. The final boosted decision tree result for event i is

$$D(x_i) = \frac{1}{\sum_{n=0}^{N} \alpha_n} \sum_{n=0}^{N} \alpha_n D_n(x_i). \tag{11}$$

In all of our tests, boosting improves performance. Another advantage of boosting decision trees is that averaging produces smoother approximations to $D(\mathbf{x})$. In this analysis 20 boosted trees are used for each analysis channel, which improves the performance by 5 to 10%. The increase in performance saturates in the region of 20 boosting cycles, varying slightly from channel to channel.

4. Decision Tree Parameters

Several internal parameters can influence the development of a decision tree.

- Initial normalization. Step 1 in Sec. XII A 1. In this analysis, we normalize both signal and background such that their sums of weights are both 0.5.
- Criteria to decide when to stop the splitting procedure due to too low statistics (Step 3 in Sec. XII A 1). In this analysis the minimum node size is 100 events per node.
- Impurity function to use to find the best split. We use the Gini index as mentioned in Sec. XII A 2.
- Number of boosting cycles. For this analysis we use 20 boosting cycles.
- Value of the boosting parameter β. We find $\beta = 0.2$ gives the best expected separation.

B. Variable Selection

A list of sensitive variables has been derived based on an analysis of the signal and background Feynman diagrams [42, 53, 54], from studies of single top quark production at next-to-leading order [62, 63], and from other analyses [10, 64]. The variables fall into three categories: individual object kinematics, global event kinematics, and variables based on angular correlations. The complete list of 49 variables is shown in Table XII.

Previous iterations of the single top quark analysis at D0 [21, 22, 23] have always used far fewer input variables. One of the main reasons was that the discriminant was computed with neural networks. Introducing too many variables can degrade the performance of a network, and testing each combination of variables is time-consuming. However, we observe that adding more variables does not degrade the DT performance. If newly introduced variables have some discriminative power, they improve the performance of the tree. If they are not discriminative enough, they are ignored. We tested this empirical observation by training different trees using several subsets of variables from the list of 49 variables. Adding more variables to the training sets never degraded the performance of the trees. Therefore, rather than producing separately optimized lists of variables for each analysis channel, the full list of 49 variables is used in all cases.

C. Decision Tree Training

We train the decision trees on one third of the available simulated events and keep the rest of the events to measure the acceptances. As a cross check, we have also trained on one half and on two thirds of the sample and have found consistent results with those obtained from using only one third. We therefore only present results with one third of the sample used for training.

Three signals are considered:

- s-channel single top quark process only (tb)
- t-channel single top quark process only (tqb)
- s- and t-channel single top quark processes combined (tb+tqb)

For simplicity, and because the decision trees are expected to deal well with all components at once, trees are trained against all backgrounds together rather than making separate trees for each background. The background includes simulated events for $t\bar{t}\rightarrow\ell$+jets, $t\bar{t}\rightarrow\ell\ell$+jets, and W+jets (with three separate components for $Wb\bar{b}$, $Wc\bar{c}$ and Wjj). Each background component is represented in proportion to its expected fraction in the background model. This leads to three different decision trees: (tb, tqb, tb+tqb against $t\bar{t}$, W+jets) for each training. In the tb+tqb training, the s- and t-channel components of the signal are taken in their SM proportions.

Samples are split by lepton flavor, jet multiplicity, and number of b-tagged jets. The current analysis uses the following samples: one isolated electron or muon; 2, 3 or 4 jets; and 1 or 2 b tags. Each sample is treated independently with its own training for each signal, leading to 36 different trees (3 signals $\times$ 2 lepton flavors $\times$ 3 jet multiplicities $\times$ 2 b-tagging possibilities).

XIII. BAYESIAN NEURAL NETWORKS ANALYSIS

A. Introduction

A neural network (NN) $n(\mathbf{x},\mathbf{w})$ [65] is a nonlinear function, with adjustable parameters $\mathbf{w}$, which is capable of modeling any real function of one or more variables [66]. In particular, it can model the discriminant $D(\mathbf{x})$ in Eq. 6. Typically, one finds a single point $\mathbf{w_0}$ in the network parameter space for which $D(\mathbf{x}) \approx n(\mathbf{x},\mathbf{w}_0)$. This can be achieved by minimizing an error function that measures the discrepancy between the value of the function $n(\mathbf{x},\mathbf{w})$ and the desired outcome for variables $\mathbf{x}$: 1 for a signal event and 0 (or -1) if $\mathbf{x}$ pertain to a background event. If the error function is built using equal numbers of signal and background events, the minimization yields the result $D(\mathbf{x}) = n(\mathbf{x},\mathbf{w}_0)$ [67, 68] provided that the function $n(\mathbf{x},\mathbf{w})$ is sufficiently flexible and that a sufficient number of training events are used.

One shortcoming of the minimization is its tendency, unless due care is exercised, to pick a point $\mathbf{w_0}$ that fits the function $n(\mathbf{x},\mathbf{w})$ too tightly to the training data. This over-training can yield a function, $n(\mathbf{x},\mathbf{w}_0)$, that is a poor approximation to the discriminant $D(\mathbf{x})$ (Eq. 6). In principle, the over-training problem can be mitigated, and more accurate and robust estimates of $D(\mathbf{x})$ constructed, by recasting the task of finding the best approximation to $D(\mathbf{x})$ as one of inference from a Bayesian viewpoint [69, 70]. The task is to infer the set of parameters $\mathbf{w}$ that yield the best approximation of $n(\mathbf{x},\mathbf{w})$ to $D(\mathbf{x})$.

Given training data T, which comprise an equal admixture of signal and background events, one assigns a probability $p(\mathbf{w}|T)dw$ to be signal-like or background-like to each point in the parameter space of the network. However, instead of finding the best single point $\mathbf{w_0}$, one averages $n(\mathbf{x},\mathbf{w})$ over every possible point $\mathbf{w}$, weighted by the probability of each point. A Bayesian neural network (BNN) [69, 70] is defined by the function

$$n(\mathbf{x}) = \int n(\mathbf{x},\mathbf{w})\, p(\mathbf{w}|T)\, dw, \tag{12}$$

that is, it is a weighted average over all possible network functions of a given architecture. The calculation is Bayesian because one is performing an integration over a parameter space. If the function $p(\mathbf{w}|T)$ is sufficiently smooth, one would expect the averaging in Eq. 12 to yield a more robust and more accurate estimate of the discriminant $D(\mathbf{x})$ than from a single best point $\mathbf{w_0}$.

There is however a practical difficulty with Eq. 12: it requires the evaluation of a complicated high-dimension integral. Fortunately, this is feasible using sophisticated numerical methods, such as Markov Chain Monte Carlo [69, 71, 72]. We use this method to sample from the posterior density $p(\mathbf{w}|T)$ and to approximate Eq. 12 by the sum

$$n(\mathbf{x}) \approx \frac{1}{K}\sum_{k=1}^{K} n(\mathbf{x},\mathbf{w}_k), \tag{13}$$

where K is the sample size.

We perform the Bayesian neural network calculations for this analysis using the "Software for Flexible Bayesian Modeling" package [73].

1. BNN Posterior Density

Given training event $T = \mathbf{t}, \mathbf{x}$, where $\mathbf{t}$ denotes the *targets* — 1 for signal and 0 for background — and $\mathbf{x}$ denotes the set of associated variables, we construct the posterior probability density $p(\mathbf{w}|T)$ via Bayes' theorem

$$\begin{aligned} p(\mathbf{w}|T) &= \frac{p(T|\mathbf{w})p(\mathbf{w})}{p(T)} \\ &= \frac{p(\mathbf{t}|\mathbf{x},\mathbf{w})p(\mathbf{x}|\mathbf{w})p(\mathbf{w})}{p(\mathbf{t}|\mathbf{x})p(\mathbf{x})} \\ &= \frac{p(\mathbf{t}|\mathbf{x},\mathbf{w})p(\mathbf{w})}{p(\mathbf{t}|\mathbf{x})}, \end{aligned} \tag{14}$$

with $p(\mathbf{x}|\mathbf{w}) = p(\mathbf{x})$. We see that there are two functions to be defined: the likelihood $p(\mathbf{t}|\mathbf{x},\mathbf{w})$ and the prior probability density $p(\mathbf{w})$. For this analysis, the neural network functions have the form

$$n(\mathbf{x},\mathbf{w}) = \frac{1}{1+\exp[-f(\mathbf{x},\mathbf{w})]}, \tag{15}$$

where

$$f(\mathbf{x},\mathbf{w}) = b + \sum_{h=1}^{H} v_h \tanh(a_h + \sum_{i=1}^{I} u_{hi}\, x_i)\,. \tag{16}$$

H is the number of hidden nodes and I is the number of input variables, $\mathbf{x}$. The adjustable parameters $\mathbf{w}$ of the networks are u_{hi} and v_h (the weights) and a_h and b (the biases).

2. BNN Likelihood

If $\mathbf{x}$ are the variables for a event, then the event's probability to be signal is $n(\mathbf{x},\mathbf{w})$; if it is a background event, then its probability is $1 - n(\mathbf{x},\mathbf{w})$. Therefore, the probability of the training event set is

$$p(\mathbf{t}|\mathbf{x},\mathbf{w}) = \prod_{j=1}^{N} n^{t_j}\,(1-n)^{1-t_j}\,, \tag{17}$$

where $t_j = 1$ for signal and $t_j = 0$ for background, and n is the total number of events. The BNN likelihood is proportional to this probability.

3. BNN Prior Density

The last ingredient needed to complete the Bayesian calculation is a prior probability density. This is the most difficult function to specify. However, experience suggests that for each network parameter, a Gaussian centered at the origin of the parameter space produces satisfactory results. Moreover, the widths of the Gaussian should be chosen to favor parameter values close to the origin, since smaller parameter values yield smoother approximations to $D(\mathbf{x})$. Conversely, large parameter values yield jagged approximations. However, since one does not know a priori what widths are appropriate, initially we allowed their values to adapt according to the noise level in the training data. Subsequently, we found that excessive noise in the training data can cause the parameter values to grow too large. Therefore, we now keep the widths fixed to a small set of values determined using single neural networks. This change is an improvement over the method used in Ref. [25].

4. Sampling the BNN Posterior Density

To compute the average in Eq. 13 requires a sample of points $\mathbf{w}$ from the posterior density, $p(\mathbf{w}|T)$. These points are generated using a Markov Chain Monte Carlo method. We first write the posterior density as

$$p(\mathbf{w}|T) = \exp\left[-V(\mathbf{w})\right], \tag{18}$$

where $V(\mathbf{w}) = -\ln p(\mathbf{w}|T)$ may be thought of as a "potential" through which a "particle" moves. We then add a "kinetic energy" term $T(\mathbf{p}) = \frac{1}{2}\mathbf{p}^2$, where $\mathbf{p}$ is a vector of the same dimensionality as $\mathbf{w}$, which together with the potential yields the particle's "Hamiltonian" $\mathcal{H} = T + V$. For a system governed by a Hamiltonian, every phase space point $(\mathbf{w},\mathbf{p})$ will be visited eventually in such a way that the phase space density of points is proportional to $\exp(-\mathcal{H})$. The phase space is traversed by alternating between long deterministic trajectories and stochastic changes in momentum. After every random change, one decides whether or not to accept the new phase space point: the new state is accepted if the energy has decreased, and accepted with a probability less than one if the contrary is true. This algorithm yields a Markov chain $\mathbf{w_1}, \mathbf{w_2}, \ldots \mathbf{w_K}$ of points, which converges to a sequence of points that constitute a faithful sample from the density $p(\mathbf{w}|T)$. In our calculations, each deterministic trajectory comprises 100 steps, followed by a randomization of the momentum. This creates a point that could be used in Eq. 13. However, since the correlation between adjacent points is high, this pair of actions is repeated 20 times, which constitutes one iteration of the algorithm, and a point is saved after each iteration.

B. BNN Training

The Bayesian neural networks are trained separately for each of the 12 analysis channels, with different sets of variables in each channel. The variables are selected using an algorithm called RuleFit [74] that orders them according to their discrimination importance (on a scale of 1 to 100). Variables with discrimination importance greater than 10 are used, which results in the selection of between 18 and 25 variables in the different channels.

For example, the variables for the electron+2jets/1tag channel are shown in Fig. 12. Each network contains a single hidden layer with 20 nodes, with the sample size K set to 100. The number of signal and background events used in the training is 10,000 each.

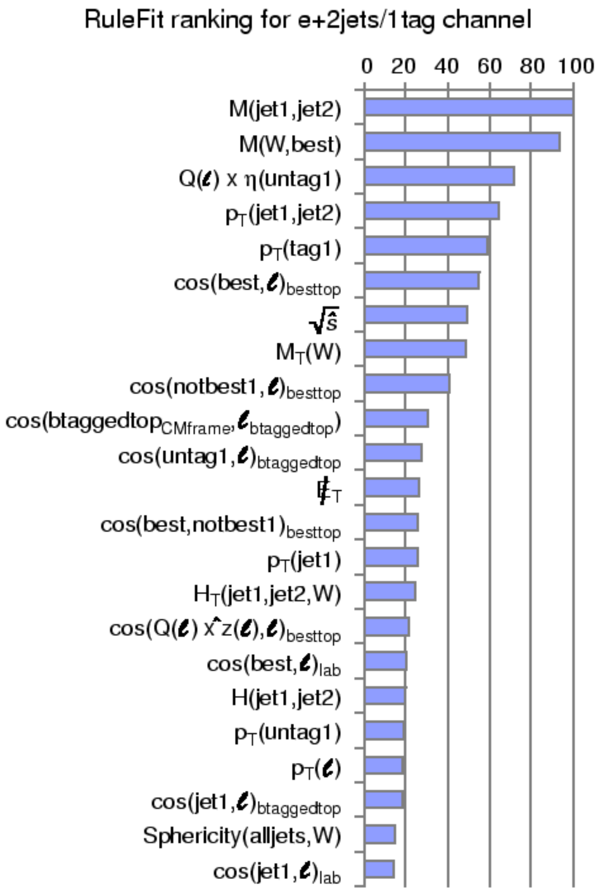

FIG. 12: BNN input variables according to their RuleFit ranking for the electron+2jets/1tag channel.

XIV. MATRIX ELEMENTS ANALYSIS

The main idea behind the matrix element (ME) technique is that the physics of a collision, including all correlations, is contained in the matrix element $\mathcal{M}$, where

$$d\sigma = \frac{(2\pi)^4|\mathcal{M}|^2}{F} d\Phi, \tag{19}$$

and $d\sigma$ is the differential cross section, F is the flux factor, and $d\Phi$ is the Lorentz invariant phase space factor. The ME analysis builds a discriminant directly using Eq. 19, thereby potentially making use of all the available kinematic information in the event. In particular, the method uses

$$p(\mathbf{x}|\text{process}_i) = \frac{1}{\sigma_i}\frac{d\sigma_i}{d\mathbf{x}} \tag{20}$$

where $\mathbf{x}$ is the configuration of the event, and $p(\mathbf{x}|\text{process}_i)$ is the probability density to observe $\mathbf{x}$ given that the physics process is process$_i$ to build the discriminant given in Eq. 6.

For each data and simulated event, two discriminant values are calculated: a t-channel discriminant and an s-channel discriminant. The t-channel discriminant uses the t-channel matrix elements when calculating $p(\mathbf{x}|S)$ as in Eq. 6, while the s-channel discriminant uses s-channel matrix elements. For each analysis channel, these discriminant values are plotted in a two-dimensional histogram, out of which a cross section measurement is extracted, as will be discussed in Sec. XVIII. The ME analysis only uses events with two or three jets and one or two b-tags, and given the two types of leptons, that results in eight independent analysis channels.

The matrix element method was developed by D0 to measure the top quark mass [75] and has been used by D0 [76] and CDF [77, 78, 79] for subsequent measurements. The ME method has also been used to measure the longitudinal W boson helicity fraction in top quark decays [80]. The result detailed here marks the first use of the method to separate signal from background in a particle search [25].

A. Event Probability Density Functions

The event configuration, $\mathbf{x}$, represents the set of reconstructed four-momenta for all selected final state objects, plus any extra reconstruction-level information, such as whether a jet is b tagged, if there is a muon in a jet, the quality of the muon track, and so on. However, the matrix element, $\mathcal{M}$, depends on the parton-level configuration of the event, which we label $\mathbf{y}$. The differential cross section, $d\sigma/d\mathbf{x}$, can be related to the parton-level variant, $d\sigma/d\mathbf{y}$, by integrating over all the possible parton values, using the parton distribution functions to relate the initial state partons to the proton and antiproton, and using a *transfer function* to relate the outgoing partons to the reconstructed objects:

$$\frac{d\sigma}{d\mathbf{x}} = \sum_j \int d\mathbf{y} \left[f_{1,j}(q_1, Q^2)\, f_{2,j}(q_2, Q^2)\, \frac{d\sigma_{\mathrm{HS},j}}{d\mathbf{y}}\, W(\mathbf{x}|\mathbf{y}, j)\, \Theta_{\mathrm{parton}}(\mathbf{y}) \right] \tag{21}$$

where

- $\sum_j$ is the sum over different configurations that contribute to the differential cross section: it is the discrete analogue to $\int d\mathbf{y}$. Specifically, this summation includes summing over the initial parton flavors in the hard scatter collision and the different permutations of assigning jets to partons.
- $\int d\mathbf{y}$ is an integration over the phase space:

$$\int d\mathbf{y} = \int dq_1 dq_2 d^3p_\ell d^3p_\nu d^3p_{q_1} d^3p_{q_2} \dots . \tag{22}$$

  Many of these integrations are reduced by delta functions.
- $f_{n,j}(q, Q^2)$ is the parton distribution function in the proton or antiproton ($n = 1$ or 2, respectively) for the initial state parton associated with configuration j, carrying momentum q, evaluated at the factorization scale Q^2. We use the same factorization scales as used when the simulated samples were generated. This analysis uses CTEQ6L1 [41] leading-order parton distribution functions via LHAPDF [81].
- $d\sigma_{\mathrm{HS}}/d\mathbf{y}$ is the differential cross section for the hard scatter (HS) collision. It is proportional to the square of the leading-order matrix element as given by (c.f., Eq. 19):

$$d\sigma_{\mathrm{HS},j} = \frac{(2\pi)^4 |\mathcal{M}|^2}{4\sqrt{(q_1 \cdot q_2)^2 - m_1^2 m_2^2}} d\Phi \tag{23}$$

  where q and m are the four-momenta and masses of the initial-state partons.
- $W(\mathbf{x}\,|\,\mathbf{y}, j)$ is called the transfer function; it represents the conditional probability to observe configuration $\mathbf{x}$ in the detector given the original parton configuration ($\mathbf{y}$, j). The transfer function is divided into two parts:

$$W(\mathbf{x}\,|\,\mathbf{y}, j) = W_{\mathrm{perm}}(\mathbf{x}\,|\,\mathbf{y}, j)\; W_{\mathrm{reco}}(\mathbf{x}\,|\,\mathbf{y}, j) \tag{24}$$

  where $W_{\mathrm{perm}}(\mathbf{x}\,|\,\mathbf{y}, j)$, discussed in Sec. XIV A 3, is the weight assigned to the given jet-to-parton permutation and $W_{\mathrm{reco}}(\mathbf{x}\,|\,\mathbf{y}, j)$, discussed in Sec. XIV A 4, relates the reconstructed value to parton values for a given permutation.
- $\Theta_{\mathrm{parton}}(\mathbf{y})$ represents the parton-level cuts applied in order to avoid singularities in the matrix element evaluation.

VEGAS Monte Carlo integration is used, as implemented in the GNU Scientific Library [82, 83].

The probability to observe a particular event given a process hypothesis, Eq. 20, also requires the total cross section ($\times$ branching fraction) as a normalization. The total cross section (σ) is just an integration of Eq. 21:

$$\sigma = \int d\mathbf{x}\, \frac{d\sigma}{d\mathbf{x}}\, \Theta_{\mathrm{reco}}(\mathbf{x}). \tag{25}$$

The term $\Theta_{\mathrm{reco}}(\mathbf{x})$ approximates the selection cuts. While conceptually simple, Eq. 25 represents a large integral: 13 dimensions for two-jet events, 17 dimensions for three-jet events other than $t\bar{t}$, and 20 dimensions for $t\bar{t}$ events. However, this integral needs to be calculated only once, not once per event, so the actual integration time is insignificant.

1. *Matrix Elements*

The matrix elements used in this analysis are listed in Table XIII. The code to calculate the matrix elements is taken from the MADGRAPH [84] leading-order matrix-element generator and uses the HELAS [85] routines to evaluate the diagrams. In Table XIII, for the single top quark processes, the top quark is assumed to decay leptonically: $t \to Wb \to \ell^+ \nu b$. For the W + jets processes, the W boson is also assumed to decay leptonically: $W^+ \to \ell^+ \nu$. Charge-conjugate processes are included. The same matrix elements are used for both the electron and muon channels. Furthermore, we use the same matrix elements for heavier generations of incoming quarks, assuming a diagonal CKM matrix. In other words, for the tb process, we use the same matrix element for $u\bar{d}$ and $c\bar{s}$ initial-state partons.

New to the analysis after the result published in Ref. [25] is an optimization of the three-jet analysis channel. For these events, a significant fraction of the background is $t\bar{t} \to \ell$+jets, as can be seen from the yield tables (see, e.g., Table X). While no new processes are added to the two-jet analysis, tqg, $Wcgg$, $Wggg$, and $t\bar{t} \to \ell$+jets are now included in the three-jet analysis.

2. *Top Pairs Integration*

For the $t\bar{t} \to \ell$+jets integration, we cannot assume one-to-one matching of parton to reconstructed object. The final state has four quarks, so one-to-one matching would lead to a four-jet event. We are interested, however, in using the $t\bar{t} \to \ell$+jets matrix element in the three-jet bin.

TABLE XIII: The matrix elements used in this analysis. The numbers in parentheses specify the number of Feynman diagrams included in each process. For simplicity, only processes that contain a positively-charged lepton in the final state are shown. The charge-conjugated processes are also used.

Matrix Elements			
Two Jets		Three Jets	
Name	Process	Name	Process
Signals		**Signals**	
tb	$u\bar{d}\to t\bar{b}$ (1)	tbg	$u\bar{d}\to t\bar{b}g$ (5)
tq	$ub\to td$ (1)	tqg	$ub\to tdg$ (5)
	$\bar{d}b\to t\bar{u}$ (1)		$\bar{d}b\to t\bar{u}g$ (5)
		tqb	$ug\to td\bar{b}$ (4)
			$\bar{d}g\to t\bar{u}\bar{b}$ (4)
Backgrounds		**Backgrounds**	
Wbb	$u\bar{d}\to Wb\bar{b}$ (2)	$Wbbg$	$u\bar{d}\to Wb\bar{b}g$ (12)
Wcg	$\bar{s}g\to W\bar{c}g$ (8)	$Wcgg$	$\bar{s}g\to W\bar{c}gg$ (54)
Wgg	$u\bar{d}\to Wgg$ (8)	$Wggg$	$u\bar{d}\to Wggg$ (54)
		$t\bar{t}\to\ell$+jets	$q\bar{q}\to t\bar{t}\to\ell^+\nu b\bar{u}d\bar{b}$ (3)
			$gg\to t\bar{t}\to\ell^+\nu b\bar{u}d\bar{b}$ (3)

The $t\bar{t}$ events therefore have to "lose" one jet to enter this bin. One way that a jet could be lost is by having its reconstructed p_T be below the selection threshold, which is 15 GeV. Another way to lose a jet is if it is merged with another nearby jet. The jet could also be outside the η acceptance of the analysis with $|\eta| > 3.4$. There is in addition a general reconstruction inefficiency that can cause a jet to be lost, but it is a small effect.

A study of $t\bar{t}\to e$+jets simulated events before tagging shows that 80% of the time when a jet is lost, there is no other jet that passes the selection cuts within $\mathcal{R} < 0.5$, that is, it has not been merged with another jet. The transverse momentum of quarks not matched to a jet passing the selection cuts is peaked at around 15 GeV, indicating that the jet is often lost because it falls below the jet p_T threshold. This study shows that the light-quark jets, which have a softer p_T spectrum, are 1.7 times as likely to be lost owing to the p_T cut as the heavy-quark jets. This observation motivated the following simplification: assume that the lost jet is from a light quark coming from the hadronically decaying W boson. In the most common case, the probability assigned to losing a jet given parton transverse energy $E_T^{\rm parton}$ is the probability that the jet is reconstructed to have $E_T^{\rm reco} < 15$ GeV, which can be calculated from the jet transfer function $W_{\rm jet}$ (discussed in Sec. XIV A 4):

$$\max\left\{\int_0^{15} dE_T^{\rm reco} W_{\rm jet}(E_T^{\rm reco}|E_T^{\rm parton}),\, 0.05\right\}. \qquad (26)$$

A minimum probability of 5% is used to account for other inefficiencies in reconstructing a jet. A random number determines which of the two quarks coming from the W boson is lost for a particular sample point in the MC integration. Other special cases considered are when the two light quarks have $\mathcal{R}(q_1, q_2) < 0.6$, in which case they are assumed to merge, or if the pseudorapidity of the quark is outside our acceptance, in which case it is assumed lost.

3. Assignment Permutations

The (discrete) summation over different configurations incorporated in Eq. 21 includes the summation over the different ways to assign the partons to the jets. A weight for each permutation is included as the $W_{\rm perm}$ part of the transfer function. This analysis uses two pieces of information to determine the weight, namely b tagging and muon charge (the muon from b decay):

$$W_{\rm perm} = W_{b\rm tag} W_{\mu\rm charge}. \qquad (27)$$

The b-tagging weight is assumed to factor by jet:

$$W_{b\rm tag} = \prod_i w_{b\rm tag}({\rm tag}_i \,|\, \alpha_i, p_{Ti}, \eta_i), \qquad (28)$$

where α_i is the flavor of quark i and ${\rm tag}_i$ is true or false depending on whether the jet is b tagged or not. The weights assigned to cases with and without a b tag are:

$$\begin{aligned} w_{b\rm tag}({\rm tag}\,|\,\alpha, p_T, \eta) &= P^{\rm taggable}(p_T,\eta)\varepsilon_\alpha(p_T,\eta) \\ w_{b\rm tag}({\rm notag}\,|\,\alpha, p_T, \eta) &= 1 - P^{\rm taggable}(p_T,\eta)\varepsilon_\alpha(p_T,\eta) \end{aligned}$$

where ε_α is the tag-rate function for the particular quark flavor and $P^{\rm taggable}$ is the taggability-rate function, which is the probability that a jet is taggable.

For the s-channel matrix element and for the $t\bar{t}\to\ell$+jets matrix element, there are both a b quark and a $\bar{b}$ quark in the final state. Furthermore, the matrix element is not symmetric with respect to the interchange of the b and $\bar{b}$ quarks, so it is helpful to be able to distinguish between b jets and $\bar{b}$ jets to make the correct assignment. In the case of muonic decays of the b or $\bar{b}$ quark, it is possible to distinguish between the jets by the charge of the decay muon. One complication is that a charm quark may also decay muonically, and the charge of the muon differs between $b \to c\mu^- \nu$ and $b \to cX\bar{X}' \to s\mu^+\bar{\nu}X\bar{X}'$. However, because $p_T^{\rm rel}$, the muon transverse momentum relative to the jet axis, differs in the two cases, the charge of the muon still provides information. Similarly to $W_{\rm btag}$, we assign the muon charge weight $W_{\mu{\rm charge}}$ based on whether the jet, if it is assumed to be a b or $\bar{b}$ in the given permutation, contains a muon of the appropriate charge. The weight is calculated by the probability that a b or a $\bar{b}$ quark decays directly into a muon given that there is a muon in the jet, parametrized as a function of $p_T^{\rm rel}$ of the muon.

4. Object Transfer Functions

We assume that the parton-level to reconstruction-level transfer function, $W_{\rm reco}$ in Eq. 24, can be factorized into individual per-object transfer functions:

$$W_{\rm reco}(\mathbf{x}\,|\,\mathbf{y}, j) = \prod_i W_{ij}(x_i\,|\,y_i), \tag{29}$$

where $W_{ij}(x_i\,|\,y_i)$ is a transfer function for one object — a jet, a muon, an electron — and x_i and y_i are reconstructed and parton-level information, respectively, for that object. We assume that angles are well measured, so the only transfer functions that are not delta functions are those for energy (for jets and electrons) and $1/p_T$ (for muons). The jet transfer functions, which give the probability to measure a jet energy given a certain parton energy, are parametrized as double Gaussians in four pseudorapidity ranges, for light jets, for b jets with a muon within the jet, and for b jets with no muon in the jet. The electron and muon transfer functions are parametrized as single Gaussians.

The jet and muon transfer functions are measured in PYTHIA $t\bar{t} \to \ell$+jets simulated events. The electron transfer functions are based on the electron resolution measured in single electron and Z boson peak simulated events.

B. Single Top Quark Discriminants

We build separate s-channel and t-channel discriminants, D_s and D_t. The signal probability densities for the various channels are:

$$p(\mathbf{x}\,|\,\mathrm{2jet,s}) = \frac{1}{\sigma_{tb}}\frac{d\sigma_{tb}}{d\mathbf{x}} \tag{30}$$

$$p(\mathbf{x}\,|\,\mathrm{2jet,t}) = \frac{1}{\sigma_{tq}}\frac{d\sigma_{tq}}{d\mathbf{x}} \tag{31}$$

$$p(\mathbf{x}\,|\,\mathrm{3jet,s}) = \frac{1}{\sigma_{tbg}}\frac{d\sigma_{tbg}}{d\mathbf{x}} \tag{32}$$

$$p(\mathbf{x}\,|\,\mathrm{3jet,t}) = \frac{1}{(\sigma_{tqb}+\sigma_{tqg})}\frac{d(\sigma_{tqb}+\sigma_{tqg})}{d\mathbf{x}}. \tag{33}$$

Equation 33 can also be written as:

$$p(\mathbf{x}\,|\,\mathrm{3jet,t}) = w_{tqb}\,p(\mathbf{x}|tqb) + w_{tqg}\,p(\mathbf{x}|tqg), \tag{34}$$

where w_{tqb} and w_{tqg} are the relative yields of the two signal processes. Calculating the yield fractions using Eq. 25, for single-tagged events we use $w_{tqb} = 0.6$ and $w_{tqg} = 0.4$, while for double-tagged events we use $w_{tqb} = 1$ and $w_{tqg} = 0$.

We apply the same methodology of using weights based on yield fraction for the $p(\mathbf{x}|\mathrm{background})$ calculations. We do not use a matrix element for every background that exists, however, so the yield fractions cannot be determined as for the signal probabilities. Some, such as $u\bar{d} \to Wc\bar{c}$, are not included because they have similar characteristics to ones that are included, such as $u\bar{d} \to Wb\bar{b}$. Therefore, we use the yields as determined from the simulated samples and consider what background the matrix elements are meant to discriminate against. We find the performance of the discriminant to be not very sensitive to the chosen weights if the weights are reasonable, and have used the weights given in Table XIV.

TABLE XIV: Background weights chosen for each analysis channel in two-jet and three-jet events.

	Background Fractions			
	1 tag		2 tags	
Weight	Electron	Muon	Electron	Muon
Two-Jet Events				
w_{Wbb}	0.55	0.60	0.83	0.87
w_{Wcg}	0.15	0.15	0.04	0.04
w_{Wgg}	0.35	0.30	0.13	0.09
Three-Jet Events				
w_{Wbbg}	0.35	0.45	0.30	0.40
w_{Wcgg}	0.10	0.10	0.02	0.03
w_{Wggg}	0.30	0.25	0.13	0.10
$w_{t\bar{t} \to \ell+\mathrm{jets}}$	0.25	0.20	0.55	0.47

XV. MULTIVARIATE OUTPUT DISTRIBUTIONS

Discriminant output shapes for signal and different background components are shown in Fig. 13, demonstrating the ability of the three analyses to separate signal from background. The DT discriminant is narrower and more central owing to the averaging effect of boosting (according to Eq. 11). The separation powers of the discriminants shown in Fig. 13 are more directly visualized in Fig. 14.

XVI. ENSEMBLES AND BIAS STUDIES

We have described three sophisticated analyses (DT, BNN, ME), each of which produces a posterior density for the single top quark production cross section. When applied to real data, we obtain well-behaved posterior densities. However, this does not guarantee that these methods are trustworthy and perform as advertised. In order to validate the methods, it is necessary to study their behavior on ensembles of pseudodatasets with characteristics as close as possible to those of the real data. We can use such ensembles to determine, for example, whether an analysis is able to extract a cross section from a signal masked by large backgrounds. We can also determine whether the claimed accuracy is warranted. Moreover, by running the three analyses on exactly the same ensembles, we can study in detail the correlations across analyses and the frequency properties of combined results and their significance.

We generate pseudodatasets from a pool of weighted signal and background events, separately for the electron and muon channels. For example, out of 1.3 million electron events, we calculate a total background yield of 756 events in the selected data. We randomly sample a count N from a Poisson distribution of mean $n = 756$ and select N events, with replacement, from the pool of 1.3 million weighted events so that events are selected with a frequency proportional to their weight. The sample contains the appropriate admixture of signal and background events, as well as the correct Poisson statistics. Moreover, we take into account the fact that the multijets and W+jets sample sizes are 100% anticorrelated. The sample is then partitioned according to the b tag and jet multiplicities, mirroring what is done to the real data. The Poisson sampling, followed by sampling with replacement, is repeated to generate as many pseudodatasets as needed. Each pseudodataset is then analyzed in exactly the same way as real data.

We have performed studies using many different ensembles, of which the most important ones are:

- **Background only (i.e., zero signal) ensemble with systematics** — the background is set to the estimated background yield value; the signal cross section is set to 0 pb; these Poisson-smeared means are further randomized to represent the effects of all systematic uncertainties.
- **Standard model signal ensemble with systematics** — the background is set to the estimated background yield value; the signal cross section is set to the standard model value of 2.86 pb; these Poisson-smeared means are further randomized to represent the effects of all systematic uncertainties.
- **Ensembles with different signal cross sections** — the background is set to the estimated background yield value; the signal cross section is set to a fixed value between 0 pb and a few times the standard model value in each ensemble; only Poisson-smearing for statistical effects is applied.

We use the zero-signal ensemble (with systematics) to calculate the p-value, a measure of the significance of the observed excess. The p-value is the probability that we obtain a measured cross section greater than or equal to the observed cross section, if there were no signal present in the data.

We use the SM signal ensemble (with systematics) to determine the correlations between the three analysis methods so we can combine their results. We also use this ensemble to calculate the compatibility of our measured result with the SM prediction, by determining how many pseudodatasets have a measured cross section at least as high as the result measured with data.

The set of ensembles with different values for the signal cross section is used to assess bias in the cross section measurement, that is, the difference between the input cross section and the mean of the distribution of measured cross sections. For each multivariate analysis, the bias is estimated by applying the entire analysis chain to the ensembles of pseudodatasets that each have a different value for the single top quark cross section. Straight-line fits of the average of the measured cross sections versus the input cross section for the three multivariate analyses are shown in Fig. 15. From this measurement, we conclude that the bias in all three analyses is small. Moreover, when compared with the variances of the ensemble distributions of measured values, the biases are negligible. We thus perform no correction to the expected or measured cross section values.

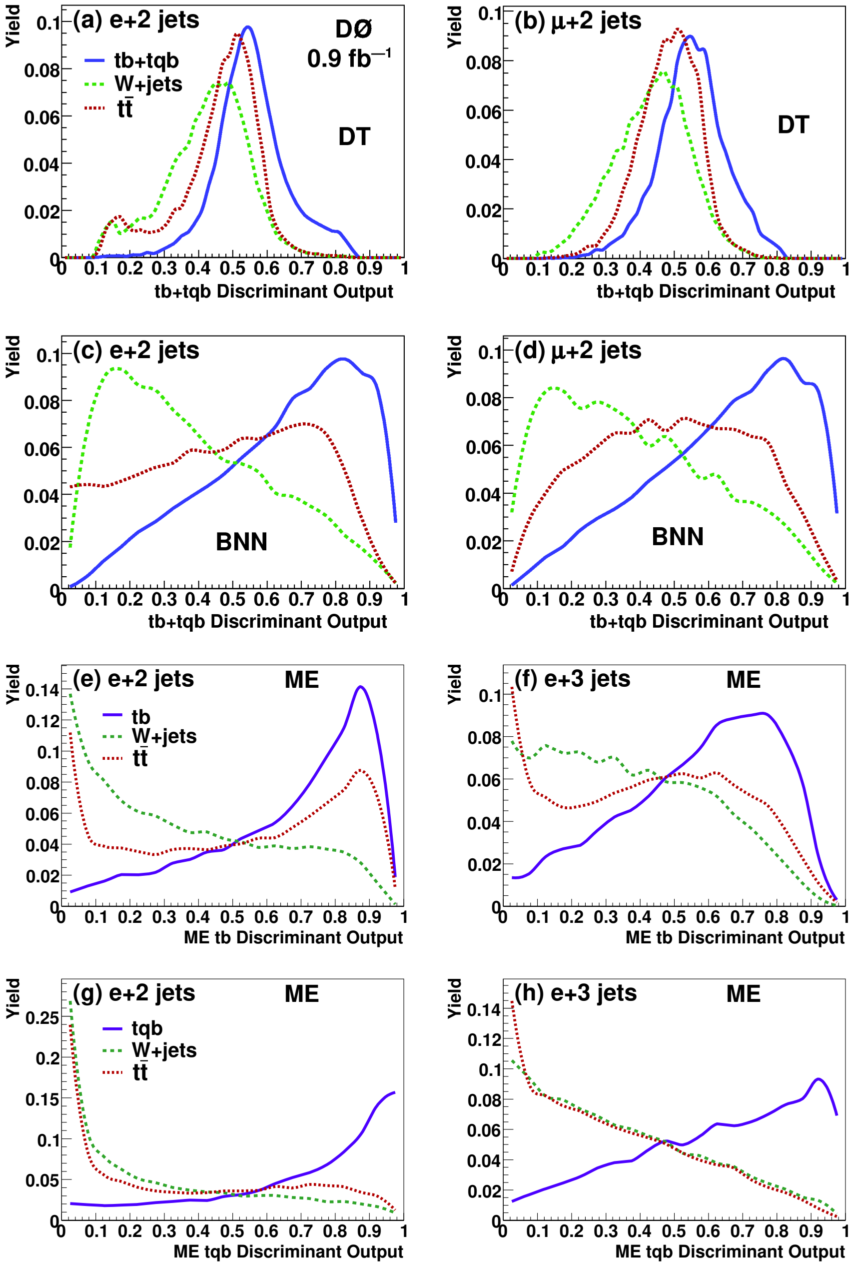

FIG. 13: For plots (a)–(d), DT and BNN discriminant outputs for tb+tqb in the e+jets channel (left column) and μ+jets channel (right column) for events with two jets of which one is b tagged. Plots (e) and (f) show the ME discriminant outputs for tb in the e+jets channel, for two-jet and three-jet events respectively. Plots (g) and (h) show the ME discriminants for tqb in the b tagged e+jets channel, for two-jet and three-jet events respectively. All histograms are normalized to unity.

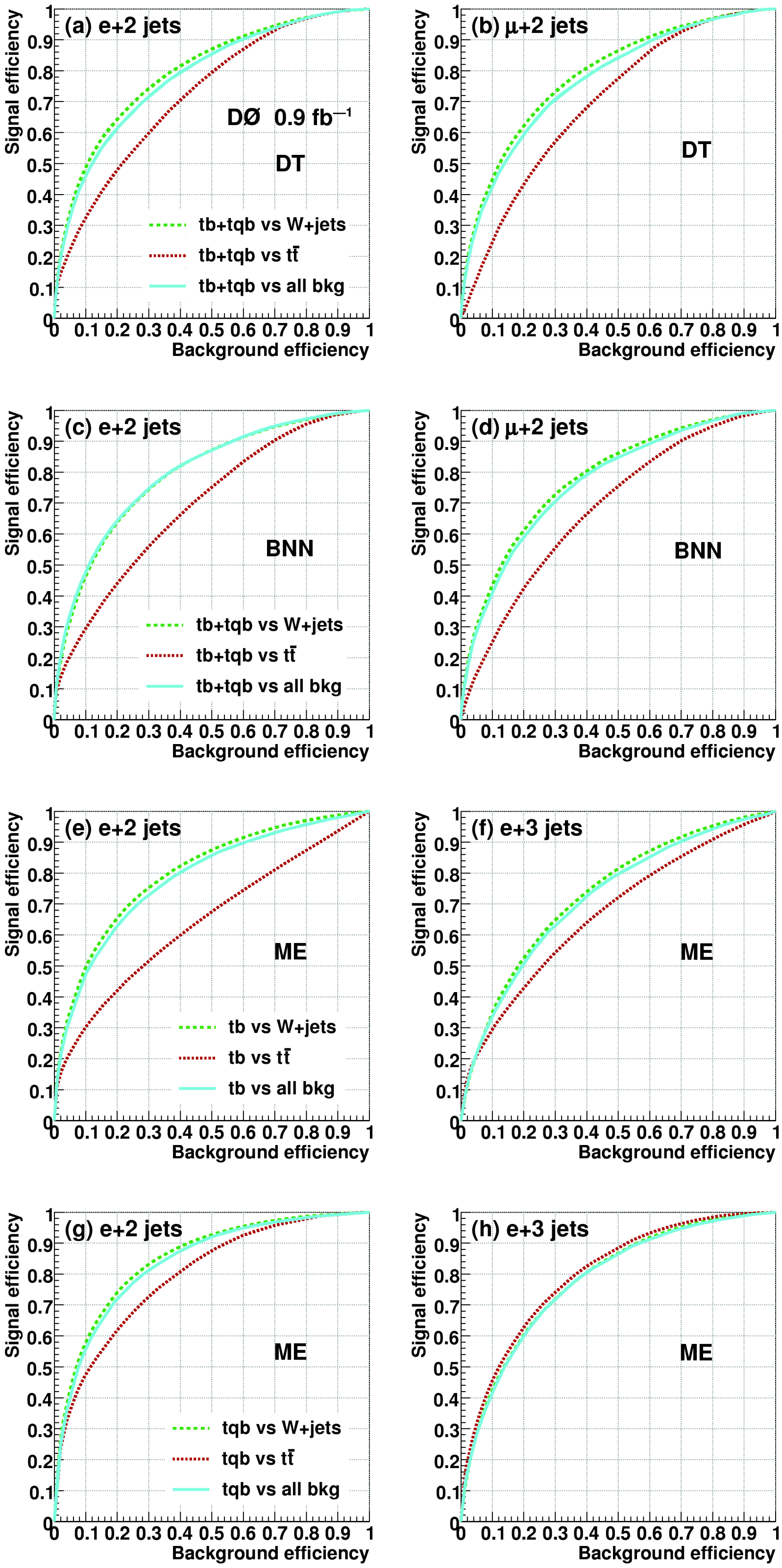

FIG. 14: For plots (a)–(d), DT and BNN $tb+tqb$ signal efficiency versus background efficiency in the e+jets channel (left column) and μ+jets channel (right column) for events with two jets of which one is b tagged. Plots (e)–(h) show the ME signal versus background efficiency for tb signal (third row) and tqb signal (fourth row), for b tagged e+jets events with two jets (left column) and three jets (right column). These curves are derived from the discriminants shown in Fig. 13.

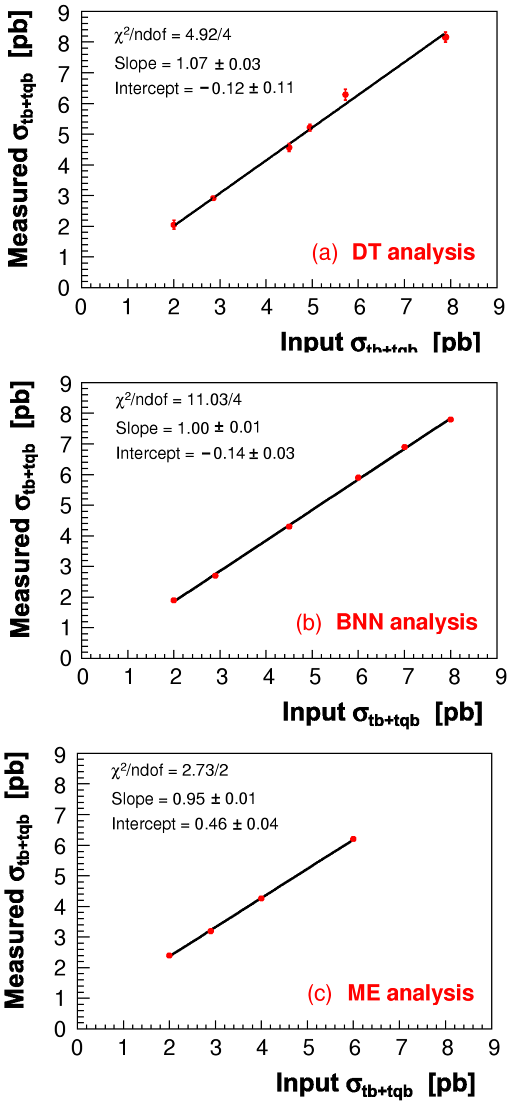

FIG. 15: Ensemble average of measured cross section as a function of the input single top quark cross section for the (a) DT, (b) BNN, and (c) ME analyses.

XVII. CROSS-CHECK STUDIES

In order to check the background model, we apply the multivariate discriminants to two background-dominated samples defined by the following criteria: (i) 2 jets, 1 b tag, and $H_T(\ell, \not\!\!E_T, \text{alljets}) < 175$ GeV for a "W+jets" sample; and (ii) 4 jets, 1 b tag, and $H_T(\ell, \not\!\!E_T, \text{alljets}) > 300$ GeV for a "$t\bar{t}$" sample. The first sample is mostly W+jets and almost no $t\bar{t}$, while the second is mostly $t\bar{t}$ and almost no W+jets.

The tb+tqb decision tree output distributions for these cross-check samples are shown in Fig. 16 and the corresponding Bayesian neural network output distributions are shown in Fig. 17. From these data-background comparisons, we conclude that there is no obvious bias in our measurement. The background model describes the data within uncertainties.

The matrix element analysis does not use four-jet events, so the cross-check samples are defined to have $H_T < 175$ GeV or $H_T > 300$ GeV for any number of jets. Figure 18 shows the s- and t-channel discriminant outputs for two-jet and three-jet events for the $H_T < 175$ GeV cross-check samples. The plots have the electron and muon channels and the one and two b-tag channels combined for increased statistics. Figure 19 shows the same for the $H_T > 300$ GeV samples.

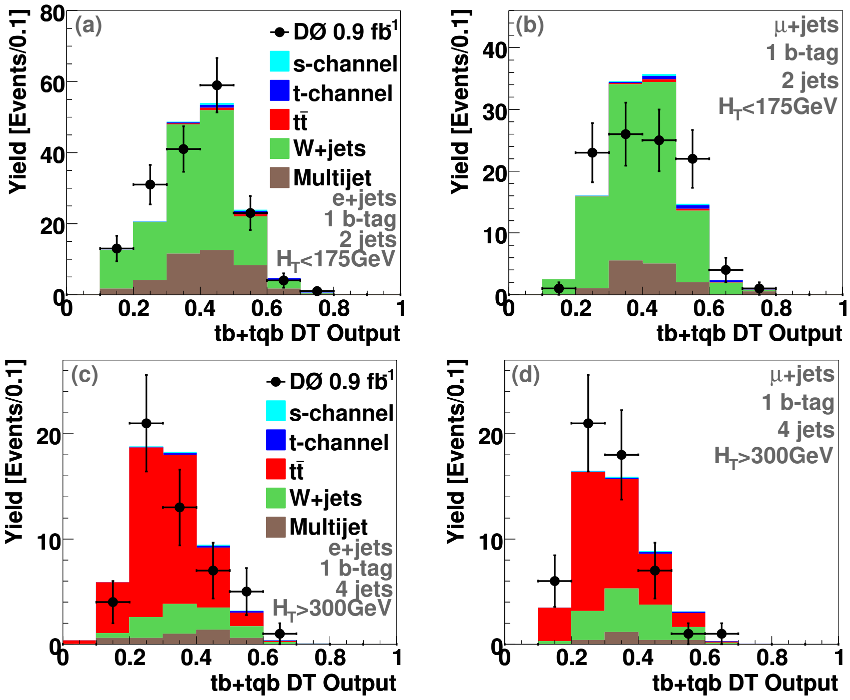

FIG. 16: DT outputs from the W+jets (upper row) and $t\bar{t}$ (lower row) cross-check samples for e+jets events (left column) and μ+jets events (right column).

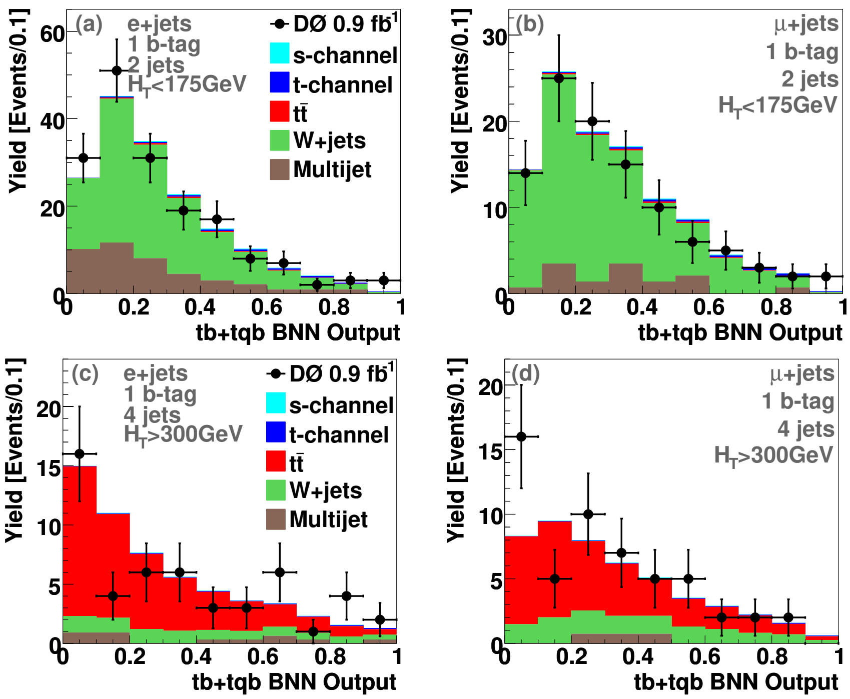

FIG. 17: BNN outputs from W+jets (upper row) and $t\bar{t}$ (lower row) cross-check samples for e+jets events (left column) and μ+jets events (right column).

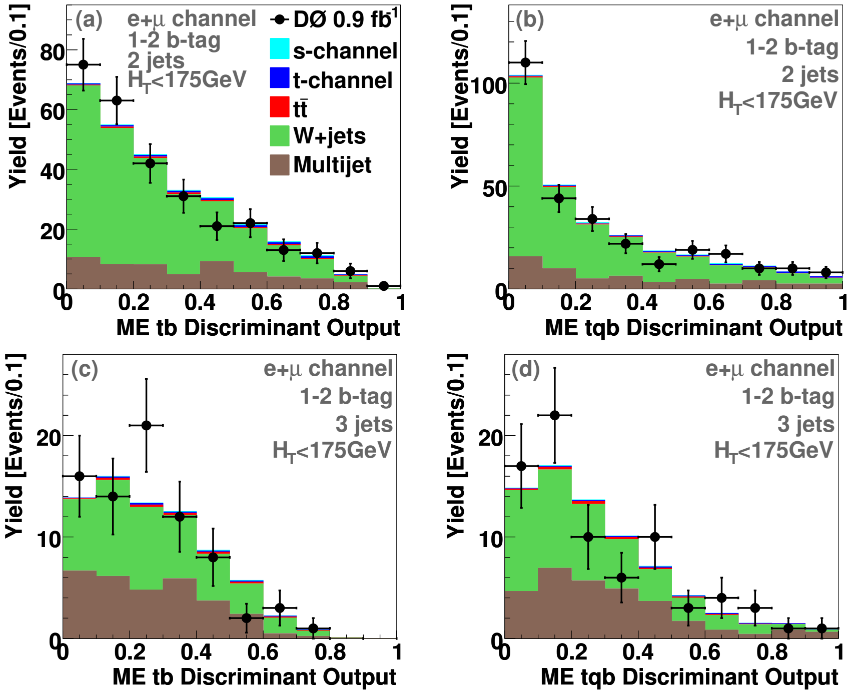

FIG. 18: $H_T < 175$ GeV cross-check plots in two-jet (upper row) and three-jet (lower row) events for the s-channel ME discriminant (left column) and the t-channel ME discriminant (right column). The plots have electrons and muons, one and two b tags combined.

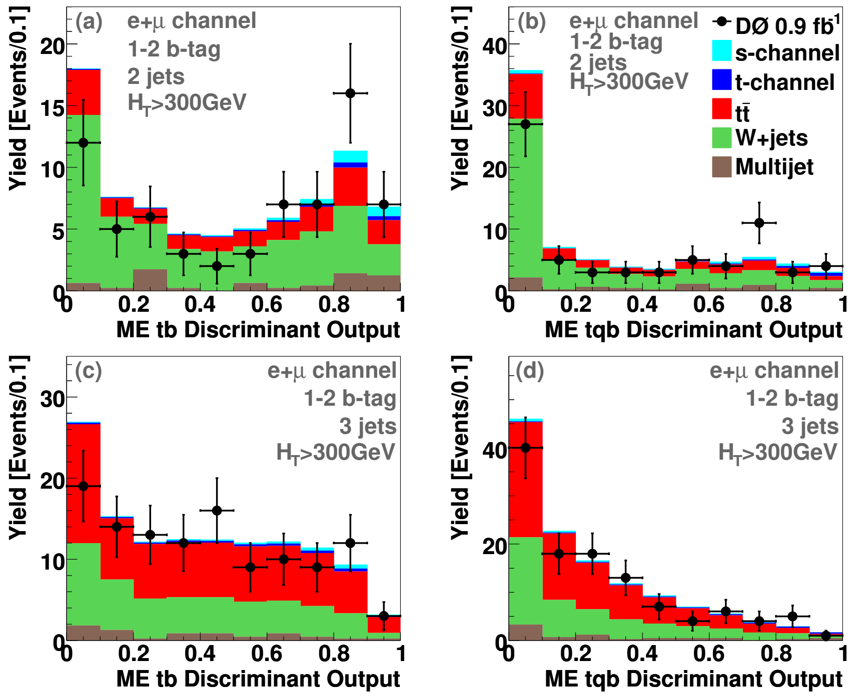

FIG. 19: $H_T > 300$ GeV cross-check plots in two-jet (upper row) and three-jet (lower row) events for the s-channel ME discriminant (left column) and the t-channel ME discriminant (right column). The plots have electrons and muons, one and two b tags combined.

XVIII. CROSS SECTION MEASUREMENTS

We use a Bayesian approach [86, 87] to extract the cross section $\sigma(p\bar{p} \to tb + X, tqb + X)$ from the observed binned discriminant distributions. In principle, the binning of data should be avoided because information is lost. In practice, however, an unbinned likelihood function is invariably approximate because of the need to fit smooth functions to the distributions of the unbinned data. Consequently, the uncertainty in the fits induces an uncertainty in the likelihood function that grows linearly with the number of events. Without study, it is not clear whether an unbinned, but approximate, likelihood function will yield superior results to those obtained from a binned but exact one. Since we have not yet studied the matter, we choose to bin the data and avail ourselves of an exact likelihood function.

A. Bayesian Analysis

For a given bin, the likelihood to observe count D, if the mean count is d, is given by the Poisson distribution

$$L(D|d) = \frac{e^{-d}\, d^D}{\Gamma(D+1)}\,, \tag{35}$$

where Γ is the gamma function. (We write the Poisson distribution in this form to permit the use of noninteger counts in the calculation of expected results. For observed results, the counts are of course integers.) The mean count d is the sum of the predicted contributions from the signal and background sources

$$d = \alpha\, \mathcal{L}\, \sigma + \sum_{i=1}^{N} b_i \equiv a\sigma + \sum_{i=1}^{N} b_i\,, \tag{36}$$

where α is the signal acceptance, $\mathcal{L}$ the integrated luminosity, σ the single top quark production cross section, b_i the mean count (that is, yield) for background source i, N the number of background sources, and $a \equiv \alpha\, \mathcal{L}$ is the effective luminosity for the signal. For analyses in which the signal comprises s- and t-channel simulated events, the latter are combined in the ratio predicted by the standard model. (Without this assumption, the probability of count D would depend on the s- and t-channel cross sections $\sigma_{\rm s}$ and $\sigma_{\rm t}$ explicitly.)

For a distribution of observed counts, the single-bin likelihood is replaced by a product of likelihoods

$$L(\mathbf{D}|\mathbf{d}) \equiv L(\mathbf{D}|\sigma, \mathbf{a}, \mathbf{b}) = \prod_{i=1}^{M} L(D_i|d_i)\,, \tag{37}$$

where $\mathbf{D}$ and $\mathbf{d}$ represent vectors of the observed and mean counts, and $\mathbf{a}$ and $\mathbf{b}$ are vectors of effective luminosity and background yields. The product is over M statistically independent bins: either all bins of a given lepton flavor, b-tag multiplicity, or jet multiplicity, or all bins of a combination of these channels.

From Bayes' theorem, we can compute the posterior probability density of the parameters, $p(\sigma, \mathbf{a}, \mathbf{b}|\mathbf{D})$, which is then integrated with respect to the parameters $\mathbf{a}$ and $\mathbf{b}$ to obtain the posterior density for the single top quark production cross section, given the observed distribution of counts $\mathbf{D}$,

$$p(\sigma|\mathbf{D}) = \frac{1}{\mathcal{N}} \iint L(\mathbf{D}|\sigma, \mathbf{a}, \mathbf{b}) \pi(\sigma, \mathbf{a}, \mathbf{b})\, d\mathbf{a}\, d\mathbf{b}\,. \tag{38}$$

Here, $\mathcal{N}$ is an overall normalization obtained from the requirement $\int p(\sigma|\mathbf{D}) d\sigma = 1$, where the integration is performed numerically up to an upper bound $\sigma_{\rm max}$ when the value of the posterior is sufficiently close to zero. In this analysis, varying $\sigma_{\rm max}$ from 30 pb to 150 pb has negligible effect on the result.

The function $\pi(\sigma, \mathbf{a}, \mathbf{b})$ is the prior probability density, which encodes our knowledge of the parameters σ, $\mathbf{a}$, and $\mathbf{b}$. Since our knowledge of the cross section σ does not inform our prior knowledge of $\mathbf{a}$ and $\mathbf{b}$, we may write the prior density as

$$\pi(\sigma, \mathbf{a}, \mathbf{b}) = \pi(\mathbf{a}, \mathbf{b})\, \pi(\sigma)\,. \tag{39}$$

The prior density for the cross section is taken to be a nonnegative flat prior, $\pi(\sigma) = 1/\sigma_{\rm max}$ for $\sigma \geq 0$, and $\pi(\sigma) = 0$ otherwise. We make this choice because it is simple to implement and yields acceptable results in ensemble studies (see Sec. XVI). The posterior probability density for the signal cross section is therefore

$$p(\sigma|\mathbf{D}) = \frac{1}{\mathcal{N}\sigma_{\rm max}} \iint L(\mathbf{D}|\sigma, \mathbf{a}, \mathbf{b}) \pi(\mathbf{a}, \mathbf{b})\, d\mathbf{a}\, d\mathbf{b}\,. \tag{40}$$

We take the mode of $p(\sigma|\mathbf{D})$ as our measure of the cross section, and the 68% interval about the mode as our measure of the uncertainty with which the cross section is measured. We have verified that these intervals, although Bayesian, have approximately 68% coverage probability and can therefore be interpreted as approximate frequentist intervals if desired.

The integral in Eq. 40 is done numerically using Monte Carlo importance sampling. We generate a large number K of points $(\mathbf{a}_k, \mathbf{b}_k)$ randomly sampled from the prior density $\pi(\mathbf{a}, \mathbf{b})$ and estimate the posterior density using

$$\begin{aligned} p(\sigma|\mathbf{D}) \;&\propto\; \iint L(\mathbf{D}|\sigma, \mathbf{a}, \mathbf{b}) \pi(\mathbf{a}, \mathbf{b})\, d\mathbf{a}\, d\mathbf{b}, \\ &\approx\; \frac{1}{K} \sum_{k=1}^{K} L(\mathbf{D}|\sigma, \mathbf{a}_k, \mathbf{b}_k)\,. \end{aligned} \tag{41}$$

In the presence of two signals, we use the same procedure and calculate the posterior probability density according to Eq. 41, replacing σ by $\sigma_{tb}, \sigma_{tqb}$ everywhere. We also replace the term $a\sigma$ in Eq. 36 by $a_{tb}\sigma_{tb} + a_{tqb}\sigma_{tqb}$, where a_{tb} and a_{tqb} are the effective luminosities of the tb and tqb signals, and σ_{tb} and σ_{tqb} are their cross sections.

The prior density for the cross section $\pi(\sigma)$ in Eq. 36 becomes $\pi(\sigma_{tb}, \sigma_{tqb}) = 1/(\sigma_{tb,\max} + \sigma_{tqb,\max})$ if both σ_{tb} and σ_{tqb} are ≥ 0, and $\pi(\sigma_{tb}, \sigma_{tqb}) = 0$ otherwise. With these two replacements, the posterior probability density becomes a two-dimensional distribution as a function of the two cross sections.

B. Prior Density

The prior density $\pi(\mathbf{a}, \mathbf{b})$ encodes our knowledge of the effective signal luminosities and the background yields (see Sec. X). The associated uncertainties fall into two classes: those that affect the overall normalization only, such as the integrated luminosity measurement, and those that also affect the shapes of the discriminant distributions, which are the jet energy scale and b-tag modeling.

The normalization effects are modeled by sampling the effective signal luminosities $\mathbf{a}$ and the background yields $\mathbf{b}$ from a multivariate Gaussian, with the means set to the estimated yields and the covariance matrix computed from the associated uncertainties. The covariance matrix quantifies the correlations of the systematic uncertainties across different sources of signal and background.

The shape effects are modeled by changing, one at a time, the jet energy scale and b-tag probabilities by plus or minus one standard deviation with respect to their nominal values. Therefore, for a given systematic effect, we create three model distributions: the nominal one, and those resulting from the plus and minus shifts. For each bin, Gaussian fluctuations, with standard deviation defined by the plus and minus shifts in bin yield, are generated about the nominal yield, and added linearly to the nominal yields generated from the sampling of the normalization-only systematic effects. Since effects such as a change in jet energy scale affect all bins coherently, we assume 100% correlation across all bins and sources. This is done by sampling from a zero mean, unit variance Gaussian and using the same variate to generate the fluctuations in all bins.

C. Bayes Ratio

Given two well-defined hypotheses H_0 and H_1 (e.g., the background-only and the signal+background hypotheses), it is natural in a Bayesian context to consider a Bayes factor B_{10},

$$\begin{aligned} B_{10} &= \frac{L(\mathbf{D}|H_1)}{L(\mathbf{D}|H_0)}, \\ &= \frac{\int L(\mathbf{D}|\sigma)\,\pi(\sigma)\,d\sigma}{L(\mathbf{D}|\sigma=0)}, \end{aligned} \tag{42}$$

as a way to quantify the significance of hypothesis H_1 relative to H_0. Here,

$$L(\mathbf{D}|\sigma) = \int\!\!\int L(\mathbf{D}|\sigma, \mathbf{a}, \mathbf{b})\,\pi(\mathbf{a}, \mathbf{b})\,d\mathbf{a}\,d\mathbf{b} \tag{43}$$

is the marginal (or integrated) likelihood and $\pi(\sigma)$ is the cross section prior density, which could be taken as a Gaussian about the standard model predicted value.

Another possible use of a Bayes factor is as an objective function to be maximized in the optimization of analyses; the optimal analysis would be the one with the largest expected Bayes factor. These considerations motivate a quantity akin to a Bayes factor that is somewhat easier to calculate, which we have dubbed a Bayes ratio, defined by

$$\text{Bayes ratio} = \frac{p(\hat{\sigma}|\mathbf{D})}{p(\sigma=0|\mathbf{D})}, \tag{44}$$

where $\hat{\sigma}$ is the mode of the posterior density. The three analyses are optimized using the expected Bayes ratio, which is computed by setting the distribution $\mathbf{D}$ to the expected one.

XIX. RESULTS

A. Expected Sensitivity

Before making a measurement using data, it is useful to calculate the expected sensitivity of these analyses. Furthermore, this expected sensitivity is used to optimize the choice of parameters in the analyses. For each case under consideration we calculate an expected Bayes ratio as defined in Sec. XVIII C. The highest Bayes ratio corresponds to the optimal parameter choice.

Table XV shows the expected Bayes ratio for each possible combination of analysis channels in the DT analysis. It can be seen from the numbers in the table that combining the two single top quark signals (i.e., searching for tb+tqb together) results in the best expected sensitivity. The single-tag two-jet channel contributes the most to this sensitivity, as expected from the high signal acceptance and reasonable signal-to-background ratio, but the addition of the other channels does improve the result; including the poorer ones does not degrade it. While the result from this table refers specifically to the DT analysis, the conclusions hold for all three multivariate techniques. Therefore, from this point onward, the 2–4 jets 1–2 tags result, using electrons and muons in the tb+tqb channel will be considered as default (2–3 jets for the ME analysis).

B. Expected Cross Sections

We measure the expected cross sections for the various channels by setting the number of data events in each

TABLE XV: Expected Bayes ratios from the decision tree analysis, including systematic uncertainties, for many combinations of analysis channels. The best values from all channels combined are shown in bold type.

	Expected Bayes Ratios							
	1–2tags, 2–4jets		$e+\mu$, 2–4jets		$e+\mu$, 1–2tags			All
	e-chan	μ-chan	1 tag	2 tags	2 jets	3 jets	4 jets	channels
tb	1.1	1.1	1.1	1.1	1.2	1.0	1.0	**1.2**
tqb	2.5	1.8	4.5	1.1	3.1	1.5	1.1	**4.7**
tb+tqb	3.2	2.3	6.7	1.3	4.8	1.5	1.1	**8.0**

channel equal to the (noninteger) expected number of background events plus the expected number of signal events (using the SM cross section of 2.86 pb at $m_{\mathrm{top}} =$ 175 GeV), and obtain the following results:

$$\begin{aligned}\sigma^{\exp}(p\bar{p} \to tb + X,\ tqb + X) &= 2.7^{+1.6}_{-1.4}\ \text{pb (DT)}\\ &= 2.7^{+1.5}_{-1.5}\ \text{pb (BNN)}\\ &= 2.8^{+1.6}_{-1.4}\ \text{pb (ME)}.\end{aligned}$$

The expected cross sections agree with the input cross section. The small deviation, less than 10%, is from the nonsymmetric nature of several of the systematic uncertainties, in particular the jet energy scale and b tagging. This effect is also observed in the pseudo-datasets.

C. Measured Cross Sections

The cross sections measured using data with the three multivariate techniques are shown in Fig. 20 where each measurement represents an independent subset of the data, for example, the 2-jet sample with 1 b tag in the electron channel.

The full combination of available channels (the most sensitive case) yields the Bayesian posterior density functions shown in Fig. 21 and cross sections of:

$$\begin{aligned}\sigma^{\mathrm{obs}}(p\bar{p} \to tb + X,\ tqb + X) &= 4.9^{+1.4}_{-1.4}\ \text{pb (DT)}\\ &= 4.4^{+1.6}_{-1.4}\ \text{pb (BNN)}\\ &= 4.8^{+1.6}_{-1.4}\ \text{pb (ME)}.\end{aligned}$$

Figure 22 shows the discriminant outputs from the three multivariate methods. Figure 23 shows the high-discriminant regions for each of these methods, with the signal component normalized to the cross section measured from data. Clearly, a model including a signal contribution fits the data better than does a background-only model.

To further illustrate the excess of data events over background in the high-discriminant region, Fig. 24 shows three variables that are inputs to the DT analysis: invariant mass of lepton+b-tagged jet+neutrino, W transverse mass, and so-called "$Q \times \eta$" (lepton charge times η of the leading untagged jet). They are each shown for low discriminant output, high output, and very high output. The excess of data over a background-only model clearly increases as the discriminant cut is increased.

The DT analysis has also measured the s- and t-channel cross sections separately. The cross sections are found to be:

$$\begin{aligned}\sigma^{\mathrm{obs}}(p\bar{p} \to tb + X) &= 1.0 \pm 0.9\ \text{pb}\\ \sigma^{\mathrm{obs}}(p\bar{p} \to tqb + X) &= 4.2^{+1.8}_{-1.4}\ \text{pb}.\end{aligned}$$

These measurements each assume the standard model value of the single top quark cross sections not being measured, since the s-channel measurement considers the t-channel process as a background and vice versa.

We can remove the constraint of the standard model ratio and form the posterior probability density as a function of the tb and tqb cross sections. This model-independent posterior is shown in Fig. 25 for the DT analysis, using the tb+tqb discriminant. The most probable value corresponds to cross sections of $\sigma(tb) =$ 0.9 pb and $\sigma(tqb) = 3.8$ pb. Also shown are the one, two, and three standard deviation contours. While this result favors a higher value for the t-channel contribution than the SM expectation, the difference is not statistically significant. Several models of new physics that are also consistent with this result are shown in Ref. [88].

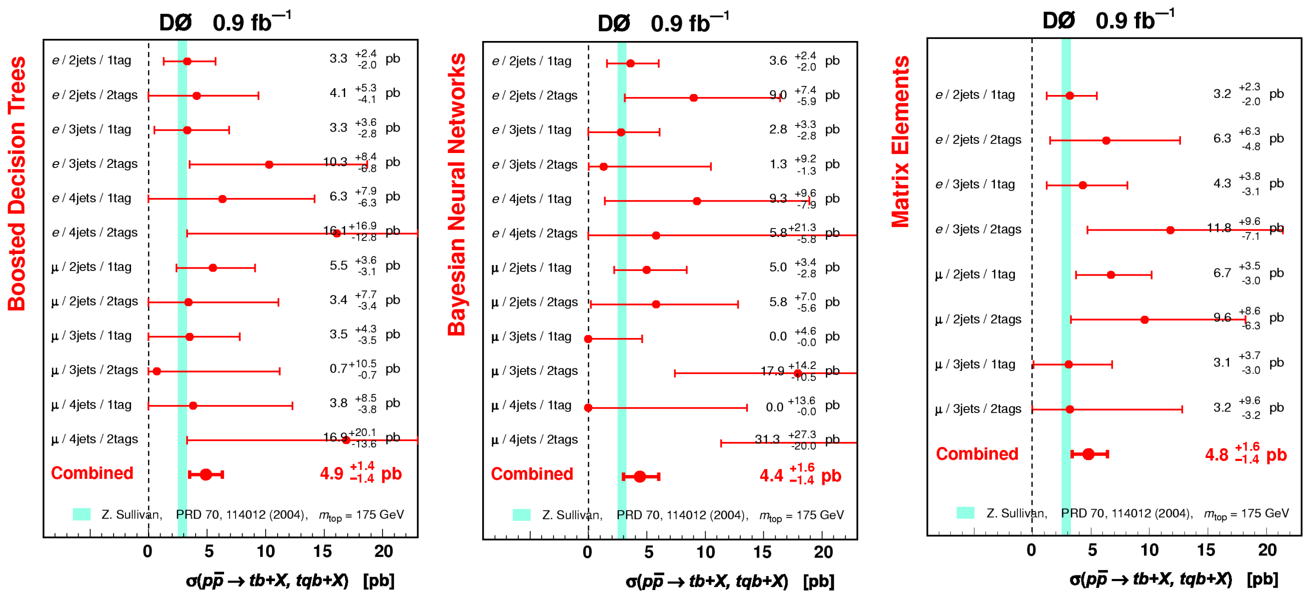

FIG. 20: Summaries of the cross section measurements using data from each multivariate technique. The left plot is DT, the middle one is BNN, and the right plot is ME.

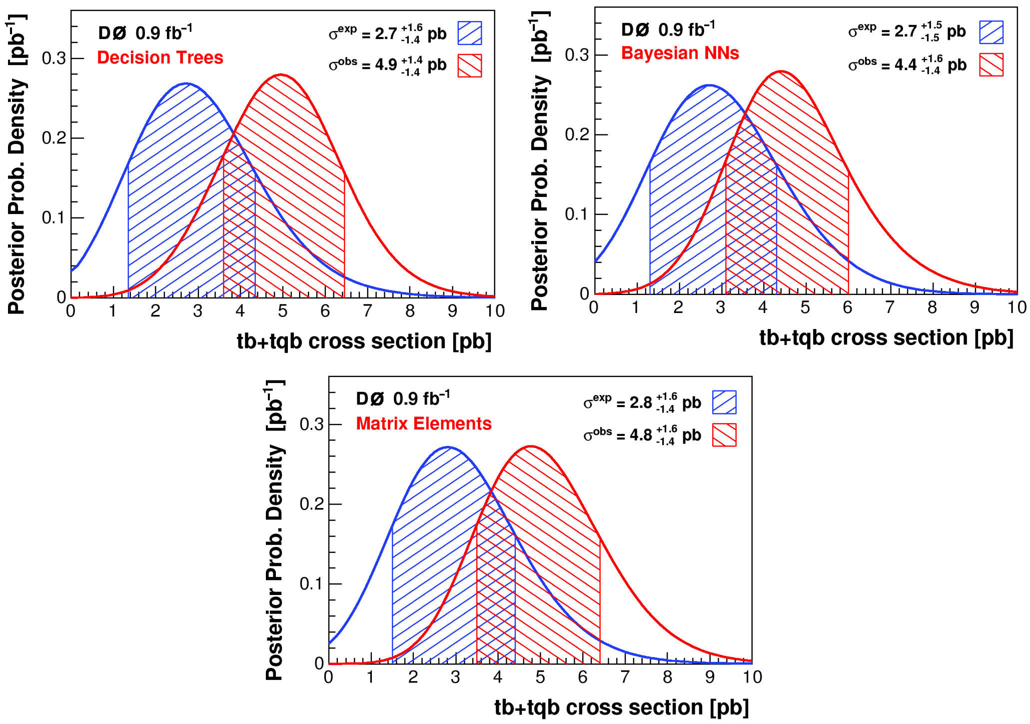

FIG. 21: Expected SM and observed Bayesian posterior density distributions for the DT, BNN and ME analyses. The shaded regions indicate one standard deviation above and below the peak positions.

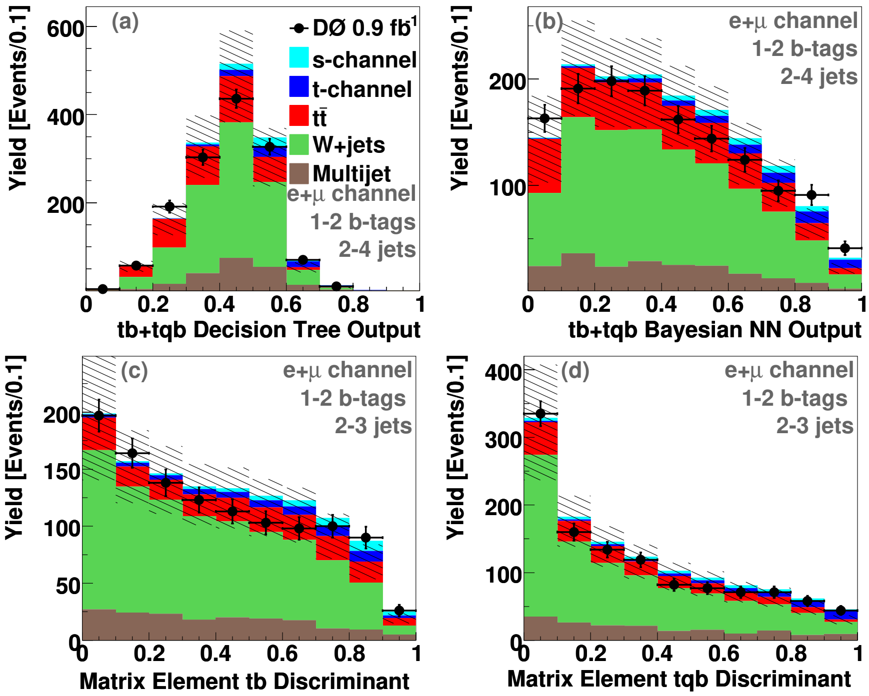

FIG. 22: The discriminant outputs of the three multivariate discriminants: (a) DT, (b) BNN, (c) ME s-channel, and (d) ME t-channel discriminants. The signal component is normalized to the cross section measured from data in each case. The hatched bands show the 1 σ uncertainty on the background.

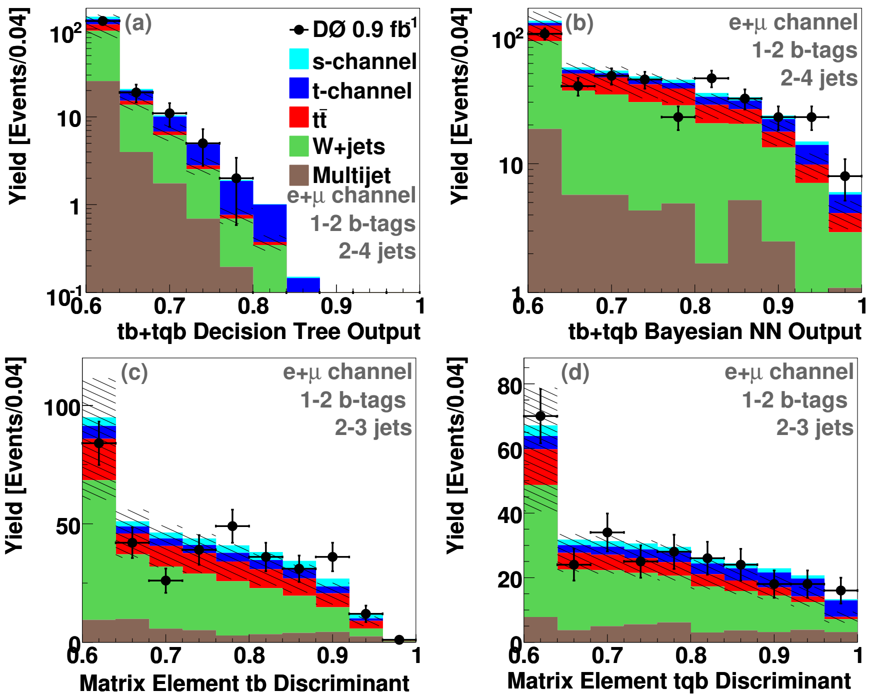

FIG. 23: Zooms of the high-discriminant output regions of the three multivariate discriminants: (a) DT, (b) BNN, (c) ME s-channel, and (d) ME t-channel discriminants. The signal component is normalized to the cross section measured from data in each case. The hatched bands show the 1 σ uncertainty on the background.

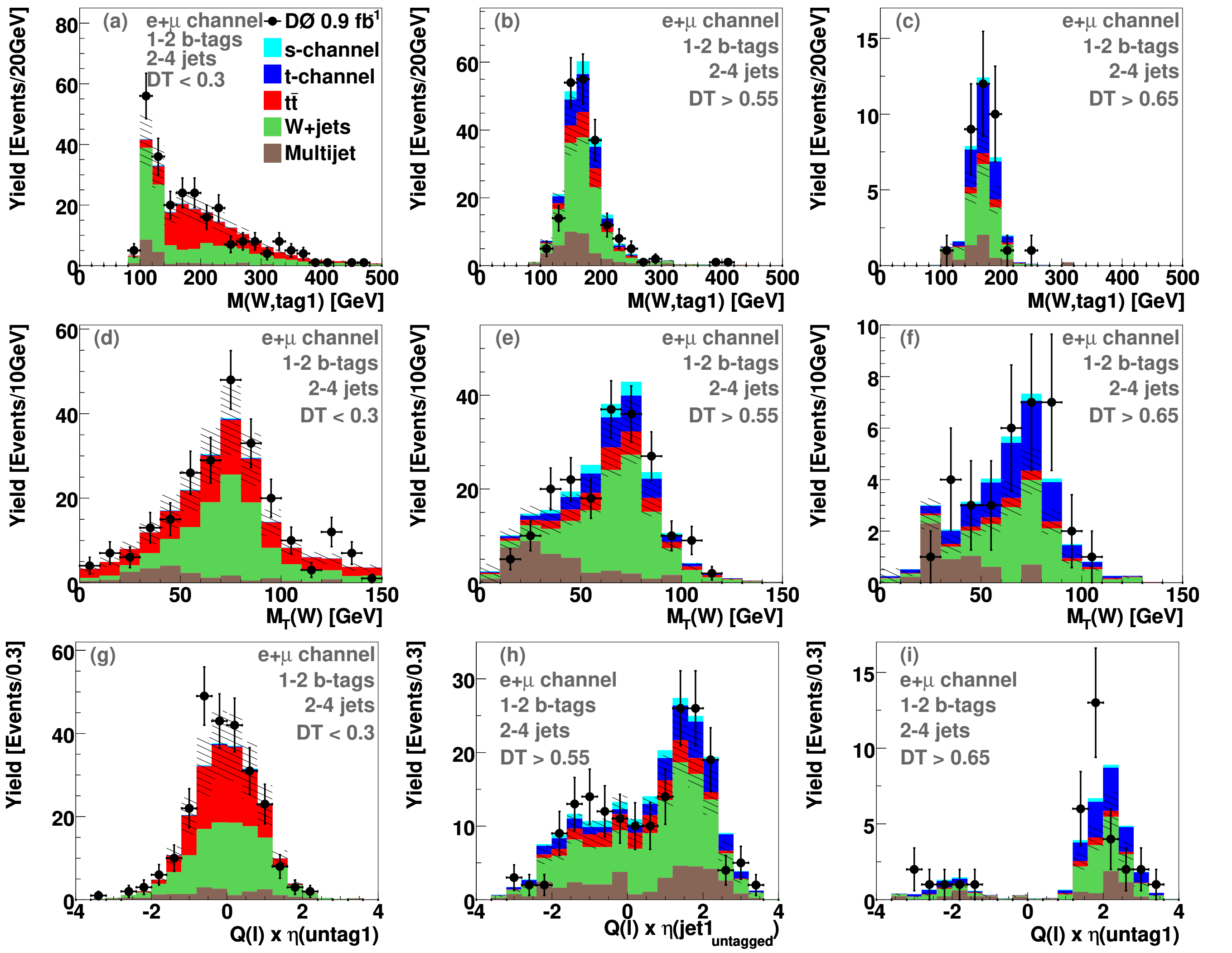

FIG. 24: The b-tagged top quark mass (top row), W boson transverse mass (second row) and Q(lepton)$\times\eta$(untag1) (third row) for the tb+tqb analysis with a low (<0.3, left column), high (>0.55, middle column), and very high (>0.65, right column) DT output, for lepton flavor (e,μ), number of b-tagged jets (1,2), and jet multiplicity (2,3,4) combined. Hatched areas represent the systematic and statistical uncertainties on the background model. The signal cross section is the measured one (4.9 pb).

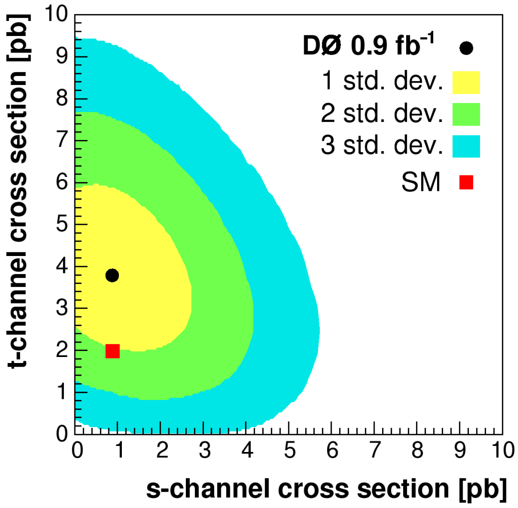

FIG. 25: Posterior probability density as a function of σ_{tb} and σ_{tqb}, when both cross sections are allowed to float in the fit of the $tb+tqb$ DT analysis. Shown are the contours of equal probability density corresponding to one, two, and three standard deviations and the location of the most probable value, together with the SM expectation.

XX. COMBINATION OF RESULTS

Since each multivariate analysis uses the same dataset to measure the single top quark cross section, their results are highly correlated. However, because the correlation is rather less than 100%, one can still gain some additional sensitivity by combining the results. We combine the three cross section measurements, σ_i (i = DT, BNN, ME) using the best linear unbiased estimate (BLUE) method [89, 90, 91]; that is, we take as the new estimate of the cross section the weighted sum

$$\sigma = \sum_i w_i \, \sigma_i, \tag{45}$$

with $\sum_i w_i = 1$, and with the weights chosen so as to minimize the variance

$$\mathrm{Var}(y) = \sum_i \sum_j w_i \, w_j \, \mathrm{Cov}(\sigma_i, \sigma_j), \tag{46}$$

where $\mathrm{Cov}\,(\sigma_i, \sigma_j) \equiv \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$ are the matrix elements of the covariance matrix of the measurements. The variance is minimized when

$$w_i = \frac{\sum_j \mathrm{Cov}^{-1}(\sigma_i, \sigma_j)}{\sum_i \sum_j \mathrm{Cov}^{-1}(\sigma_i, \sigma_j)}, \tag{47}$$

where $\mathrm{Cov}^{-1}(\sigma_i, \sigma_j)$ denotes the matrix elements of the inverse of the covariance matrix. In order to estimate the correlation matrix, each analysis is run on the same ensemble of pseudodatasets, specifically, the SM ensemble with systematics, which comprises 1,900 pseudodatasets common to all three analyses. To estimate the p-value of the combined result, the analyses are run on 72,000 pseudodatasets of the background-only ensemble.

A. Weights, Coverage Probability, and Combined Measurement

We use the SM ensemble with systematics to determine the weights w_i and to check the coverage probability of the confidence intervals calculated as described in Sec. XVIII A. The cross section measurements from this ensemble are shown in Fig. 26 for the individual and combined analyses. The mean and square root of the variance obtained from these distributions give the following results:

$$\begin{aligned}
\sigma^{\text{SM-ens}}(p\bar{p} \to tb + X,\quad tqb + X) \\
&= 2.9 \pm 1.6 \text{ pb} \quad (\text{DT}) \\
&= 2.7 \pm 1.5 \text{ pb} \quad (\text{BNN}) \\
&= 3.2 \pm 1.4 \text{ pb} \quad (\text{ME}) \\
&= 3.0 \pm 1.3 \text{ pb} \quad (\text{Combined}).
\end{aligned}$$

The weights w_i for the three analyses are found to be

- $w_{\mathrm{DT}} = 0.127$,
- $w_{\mathrm{BNN}} = 0.386$,
- $w_{\mathrm{ME}} = 0.488$.

The correlation matrix is

$$\text{Correlation matrix} = \begin{array}{cc} \begin{array}{ccc} \text{DT} & \text{BNN} & \text{ME} \end{array} & \\ \begin{pmatrix} 1 & 0.66 & 0.64 \\ 0.66 & 1 & 0.59 \\ 0.64 & 0.59 & 1 \end{pmatrix} & \begin{array}{l} \text{DT} \\ \text{BNN}\,, \\ \text{ME} \end{array} \end{array} \tag{48}$$

and the one-standard-deviation coverage probability of the (Bayesian) confidence interval is 0.67.

The result from combining the DT, BNN, and DT measurements for the single top quark cross section is

$$\begin{aligned}
\sigma^{\text{obs}}(p\bar{p} \to tb + X,\quad tqb + X) \\
&= 4.7 \pm 1.3 \text{ pb} \quad (\text{Combined}),
\end{aligned}$$

using the values listed at the beginning of Sec. XIX C. Figure 27 summarizes the measurements of the tb+tqb cross section from the individual analyses as well as the combination.

B. Measurement Significance

Having determined the combined result for the single top quark cross section, we can now determine the signal

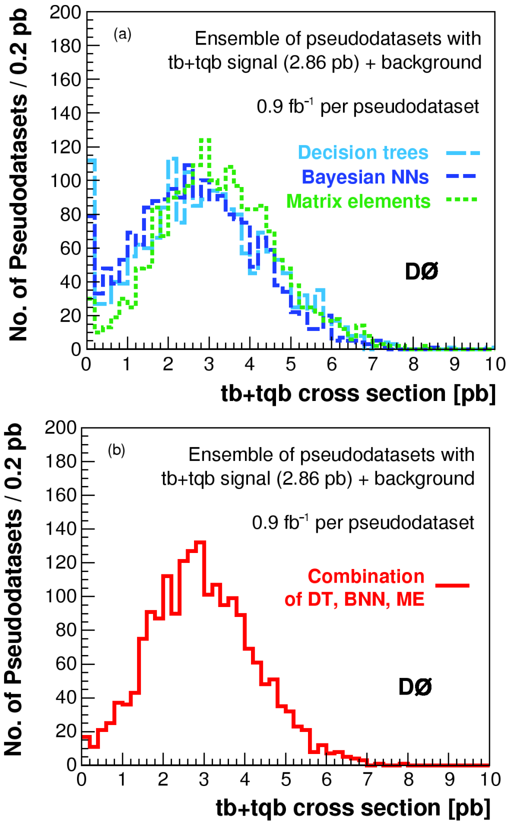

FIG. 26: Distributions of the measured cross sections from (a) the individual analyses, and (b) the combined analysis, using the SM ensemble with systematics.

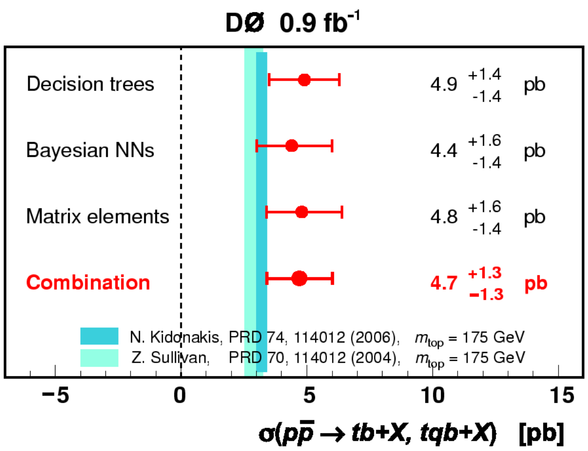

FIG. 27: The measured single top quark cross sections from the individual analyses and their combination.

significance corresponding to this measurement. Distributions of the results from all the analyses are shown in Fig. 28.

The expected p-value (and the associated significance in Gaussian-like standard deviations) is obtained by counting how many background-only pseudodatasets yield a measured cross section greater than the SM value of 2.86 pb. The result is 1.1% or 2.3 standard deviations, as shown in Table XVI.

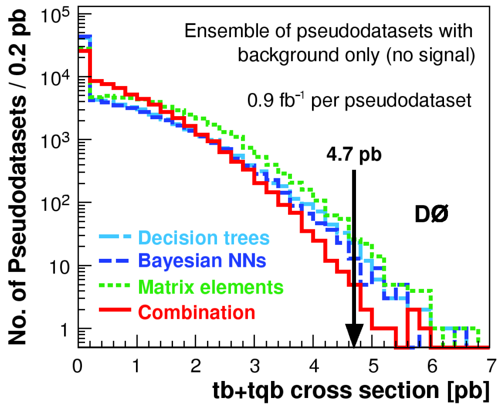

FIG. 28: Distributions of the cross sections measured from data by the three analyses and their combination, using the background-only ensemble. The arrow shows the combined cross section measurement, 4.7 pb.

TABLE XVI: The expected tb+tqb cross sections, p-values, and significances for the individual and combined analyses, using the SM value of 2.86 pb for the single top quark production cross section as the reference point in Fig. 28.

	Expected Results		
Analysis	Expected cross section [pb]	Expected p-value	Expected significance (std. dev.)
DT	2.7	0.018	2.1
BNN	2.7	0.016	2.2
ME	2.8	0.031	1.9
Combined	2.8	0.011	2.3

The observed p-value is similarly calculated by counting how many background-only pseudodatasets result in a cross section above the value of 4.7 pb measured from data. The result is 0.014% or 3.6 standard deviations. The observed cross sections, p-values, and significances from all the analyses are summarized in Table XVII.

Finally, using the SM ensemble with systematics, we quantify the compatibility of our result with the SM expectation by counting how many pseudodatasets result in a cross section with the observed value or higher for each of the analyses. The probabilities for the different analyses are 10% for the DT analysis, 13% for the ME analysis, 13% for the BNN analysis, and 10% for the combined analysis.

TABLE XVII: The cross sections measured from data, p-values, and significances for the individual and combined analyses, the latter two obtained using the background-only ensemble.

Observed Results			
Analysis	Measured cross section [pb]	Measured p-value	Measured significance (std. dev.)
DT	4.9	0.00037	3.4
BNN	4.4	0.00083	3.1
ME	4.8	0.00082	3.2
Combined	4.7	0.00014	3.6

C. Discriminant Comparison

In order to compare the expected performance of the three multivariate techniques, it is instructive to compute a power curve for each method using the two hypotheses H_1 = SM-signal+background and H_0 = background-only. A power curve is a plot of the p-value for the alternative hypothesis H_1 versus the p-value for the null hypothesis H_0. That is, it is a plot of the probability to accept the alternative hypothesis, if true, versus the significance level. The power curves in Fig. 29 show that all three methods, as anticipated given the expected significances of Table XVI, exhibit comparable performance.

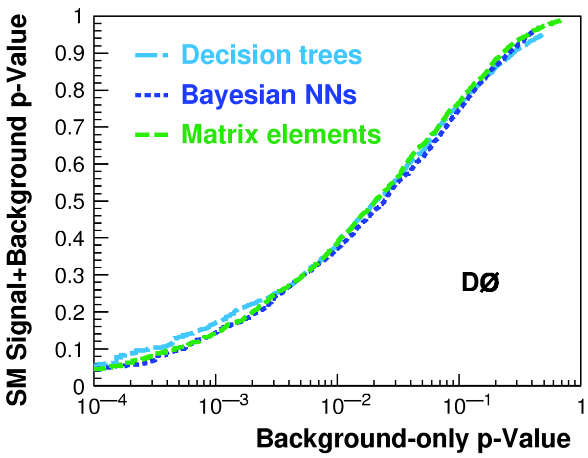

FIG. 29: The p-value computed from the SM-signal+background ensemble versus the p-value from the background-only ensemble for reference cross sections varying monotonically from 0–10 pb.

XXI. $|V_{tb}|$ MEASUREMENT

Within the SM with three generations of quarks, the charged-current interactions of the top quark are of the type $V-A$, and involve a W boson and a down-type quark q ($q = d, s, b$):

$$\Gamma^{\mu}_{Wtq} = -\frac{g}{\sqrt{2}} V_{tq} f_1^L \bar{u}(p_b)\gamma^{\mu} P_L u(p_t), \tag{49}$$

where $|V_{tq}|$ is one of the elements of the 3×3 unitary CKM matrix [15, 16], $f_1^L = 1$ in the SM, and $P_L = (1-\gamma_5)/2$ is the left-handed $(-)$ projection operator. Under the assumption of three generations and a unitary CKM matrix, the $|V_{tq}|$ elements are severely constrained [92]:

$$\begin{aligned} |V_{td}| &= (8.14^{+0.32}_{-0.64}) \times 10^{-3} \\ |V_{ts}| &= (41.61^{+0.12}_{-0.78}) \times 10^{-3} \\ |V_{tb}| &= 0.999100^{+0.000034}_{-0.000004}. \end{aligned} \tag{50}$$

In several extensions of the SM involving, for example, a fourth generation of quarks or an additional heavy quark singlet that mixes with the top quark, the 3×3 CKM matrix is no longer required to be unitary, and $|V_{tb}|$ can be significantly smaller than unity [19].

This paper describes in detail the first direct measurement of $|V_{tb}|$, based on the single top quark production cross section measurement using decision trees [25]. The $|V_{tb}|$ measurement is a relatively straightforward extension of the cross section measurement using the same dataset and analysis infrastructure, since the cross section for single top quark production is directly proportional to $|V_{tb}|^2$. This measurement of $|V_{tb}|$ makes no assumptions on the number of generations or unitarity of the CKM matrix. However, some assumptions are made in the generation of our signal MC samples and the extraction of $|V_{tb}|$ from the cross section measurement. In particular, we assume the following: (i) there are only SM sources of single top quark production; (ii) top quarks decay to Wb; and (iii) the Wtb interaction is CP-conserving and of $V-A$ type. We discuss these assumptions in more detail here.

First, we assume that the only production mechanism for single top quarks involves an interaction with a W boson. Therefore, extensions of the SM where single top quark events can be produced, for example, via flavor-changing neutral current interactions [93] or heavy scalar or vector boson exchange [94], are not considered here.

The second assumption is that $|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2$. In other words, we assume $|V_{ts}|$ and $|V_{td}|$ are negligible compared to $|V_{tb}|$, without making any assumption on the magnitude of $|V_{tb}|$. This is reasonable given the measurements of

$$R = \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2}, \tag{51}$$

by the CDF [95] and D0 [96] collaborations, obtained by comparing the rates of $t\bar{t}$ events with zero, one and two

b-tagged jets. For instance, D0's measurement results in $|V_{td}|^2 + |V_{ts}|^2 = (-0.03^{+0.18}_{-0.16})|V_{tb}|^2$. The requirement that $|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2$ implies that $B(t\to Wb) \simeq 100\%$ and that single top quark production is completely dominated by the Wtb interaction. This assumption is made explicitly when measuring the combined tb+tqb cross section when assuming the SM ratio of σ_{tb}/σ_{tqb} [19], as well as in the generation of single top quark and $t\bar{t}$ simulated samples.

Finally, we assume that the Wtb vertex is charge-parity (CP) conserving and of the $V{-}A$ type as given in Eq. 49, but it is allowed to have an anomalous strength f_1^L. We do not allow for right-handed or tensor couplings that may occur in the most general Wtb vertex [97, 98]. The simulated samples can still be used under the assumption of an anomalous f_1^L coupling: the $t\bar{t}$ cross section and kinematics, as well as the tb and tqb kinematics are completely unaffected. An anomalous value for f_1^L would only rescale the single top quark cross section, allowing it to be larger or smaller than the SM prediction, even under the assumption of $|V_{tb}| = 1$. Therefore, strictly speaking, we are measuring the strength of the $V{-}A$ coupling, i.e., $|V_{tb}f_1^L|$, which is allowed to be > 1. Limiting our measurement to the $[0, 1]$ range implies the additional assumption that $f_1^L = 1$.

A. Statistical Analysis

This measurement uses exactly the same machinery as used to obtain the single top quark cross section posterior. Following standard convention for parameters that multiply the cross section, we choose a prior that is nonnegative and flat in $|V_{tb}|^2$, which means it is flat in the cross section. However, in one of the two cases presented below, we restrict the prior to the SM allowed region [0,1].

B. Systematic Uncertainties

In order to extract $|V_{tb}|$ from the measured cross section, additional theoretical uncertainties [8] need to be considered. These uncertainties are applied separately to the tb and tqb samples in order to take the correlations into account properly. They are listed in Table XVIII. The uncertainty on the top quark mass of 5.1 GeV [75] is used when estimating the $t\bar{t}$ cross section uncertainty and the tb and tqb cross section uncertainties.

C. $|\mathbf{V_{tb}}|$ Result

The measurement for the CKM matrix element is obtained from the most probable value of $|V_{tb}|^2$, given by $|V_{tb}| = \sqrt{|V_{tb}|^2}$, and the uncertainty is computed as $\Delta|V_{tb}| = \Delta|V_{tb}|^2/2|V_{tb}|$. We have used the decision tree result to derive a posterior for $|V_{tb}|$. The posterior without the prior restricted to be only nonnegative gives $|V_{tb}f_1^L|^2 = 1.72^{+0.64}_{-0.54}$, which results in

$$|V_{tb}f_1^L| = 1.31^{+0.25}_{-0.21}.$$

The posterior with the prior restricted to the [0,1] region gives $|V_{tb}|^2 = 1.00^{+0.00}_{-0.24}$, which results in

$$|V_{tb}| = 1.00^{+0.00}_{-0.12}.$$

The corresponding 95% C.L. lower limit on $|V_{tb}|^2$ is 0.46, corresponding to a lower limit of

$$|V_{tb}| > 0.68.$$

The posterior densities for $|V_{tb}|^2$ for each choice of prior are shown in Fig. 30.

TABLE XVIII: Systematic uncertainties on the cross section factor required to extract $|V_{tb}|$.

Additional $\lvert V_{tb}\rvert$ Uncertainties	tb	tqb
Top quark mass	8.5%	13.0%
Factorization scale	4.0%	5.5%
Parton distributions	4.5%	10.0%
α_s	1.4%	0.01%

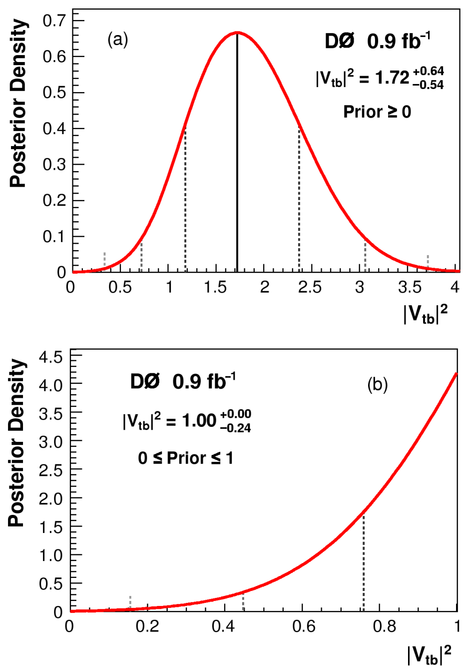

FIG. 30: The posterior density distributions for $|V_{tb}|^2$ for (a) a nonnegative flat prior, and (b) a flat prior restricted to the region [0,1] and assuming $f_1^L = 1$. The dashed lines show the positions of the one, two, and three standard deviation distances away from the peak of each curve.

XXII. SUMMARY

Using approximately 0.9 fb^{-1} of D0 data, we have performed an analysis of events with a single isolated lepton (electron or muon), missing transverse energy, and 2–4 jets (1 or 2 of them b tagged). Using three different multivariate techniques, decision trees, Bayesian neural networks, and matrix elements, we have searched for single top quark events from the s-channel (tb) and t-channel (tqb) processes combined. We measure the cross section to be

$$\sigma\left(p\bar{p} \rightarrow tb + X,\ tqb + X\right) = 4.7 \pm 1.3 \text{ pb}.$$

This corresponds to an excess of 3.6 Gaussian-equivalent standard deviation significance and constitutes the first evidence of a single top quark signal. Ensemble tests have shown this result to be compatible with the standard model cross section with 10% probability.

The decision tree cross section result has been used to extract the first direct measurement of the CKM matrix element $|V_{tb}|$. This result does not assume three-generation unitarity of the matrix. The model independent measurement is

$$|V_{tb} f_1^L| = 1.31^{+0.25}_{-0.21},$$

where f_1^L is a generic left-handed vector coupling. If we constrain the value of $|V_{tb}|$ to the standard model region (i.e., $|V_{tb}| \leq 1$ and $f_1^L = 1$), then at 95% C.L., $|V_{tb}|$ has been measured to be

$$0.68 < |V_{tb}| \leq 1.$$

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC (United Kingdom); MSMT and GACR (Czech Republic); CRC Program, CFI, NSERC and WestGrid Project (Canada); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); CAS and CNSF (China); and the Alexander von Humboldt Foundation.

---

[a] Visitor from Augustana College, Sioux Falls, SD, USA.
[b] Visitor from The University of Liverpool, Liverpool, UK.
[c] Visitor from ICN-UNAM, Mexico City, Mexico.
[d] Visitor from II. Physikalisches Institut, Georg-August-University, Göttingen, Germany.
[e] Also at TRIUMF, Vancouver, B.C., Canada.
[f] Visitor from Helsinki Institute of Physics, Helsinki, Finland.
[g] Visitor from Universität Zürich, Zürich, Switzerland.
[‡] Deceased.

[1] F. Abe *et al.* (CDF Collaboration), Phys. Rev. Lett. **74**, 2626 (1995).
[2] S. Abachi *et al.* (D0 Collaboration), Phys. Rev. Lett. **74**, 2632 (1995).
[3] N. Kidonakis and R. Vogt, Phys. Rev. D **68**, 114014 (2003). The value of the $t\bar{t}$ cross section at $m_{\text{top}} =$ 170 GeV is from N. Kidonakis, private communication.
[4] The most recent world-average measurement of the top quark mass is 170.9 ± 1.8, (hep-ex/0703034), but at the time this analysis was started, the value was 174.3±3.4 GeV, (hep-ex/0507006), and we rounded this to 175 GeV when choosing a mass to use in the analysis.
[5] S. Cortese and R. Petronzio, Phys. Lett. B **253**, 494 (1991).
[6] T. Stelzer and S. Willenbrock, Phys. Lett. B **357**, 125 (1995).
[7] A.P. Heinson, A.S. Belyaev, and E.E. Boos, Phys. Rev. D **56**, 3114 (1997).
[8] Z. Sullivan, Phys. Rev. D **70**, 114012 (2004).
[9] S.S.D. Willenbrock and D.A. Dicus, Phys. Rev. D **34**, 155 (1986).
[10] C.-P. Yuan, Phys. Rev. D **41**, 42 (1990).
[11] R.K. Ellis and S.J. Parke, Phys. Rev. D **46**, 3785 (1992).
[12] T.M.P. Tait, Phys. Rev. D **61**, 034001 (1999).
[13] D.O. Carlson, E. Malkawi, and C.-P. Yuan, Phys. Lett. B **337**, 145 (1994).
[14] E. Malkawi and C.-P.Yuan, Phys. Rev. D **50**, 4462 (1994).
[15] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).
[16] M. Kobayashi and K. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
[17] G.V. Jikia and S.R. Slabospitsky, Phys. Lett. B **295**, 136 (1992).
[18] T. Stelzer, Z. Sullivan, and S. Willenbrock, Phys. Rev. D **58**, 094021 (1998).
[19] J. Alwall, R. Frederix, J.-M. Gérard, A. Giammanco, M. Herquet, S. Kalinin, E. Kou, V. Lemaître, and F. Maltoni, Eur. Phys. J. C **49**, 791 (2007).
[20] B. Abbott *et al.* (D0 Collaboration), Phys. Rev. D **63**, 031101 (2001).
[21] V.M. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **517**, 282 (2001).
[22] V.M. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **622**, 265 (2005).
[23] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. D **75**, 092007 (2007).
[24] D0 Ph.D. dissertations on single top quark physics:
J. McDonald, Florida State University, May 1999;
L. Dudko, Moscow State University, September 2001;
E. Busato, University of Paris VI, April 2005;

M. Agelou, University of Paris VI, May 2005;
S. Jabeen, University of Kansas, January 2006;
B. Clément, IReS de Strasbourg, April 2006;
P. Perea, University of California, Riverside, June 2006;
T. Gadfort, University of Washington, April 2007;
J. Mitrevski, Columbia University, July 2007,
D. Kau, Florida State University, August 2007.

[25] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **98**, 181802 (2007).

[26] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **65**, 091102 (2002).

[27] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **69**, 053003 (2004).

[28] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **71**, 012005 (2005).

[29] CDF Ph.D. dissertations on single top quark physics:
T. Kikuchi, Tsukuba University, March 2000;
S. Wolinski, University of Michigan, 2002;
C. Ciobanu, Ohio State University, August 2002;
B. Stelzer, University of Toronto, May 2005;
T. Walter, University of Karlsruhe, June 2005;
Y. Kemp, University of Karlsruhe, February 2006;
S. Richter, University of Karlsruhe, November 2007.

[30] Pseudorapidity is defined as $\eta = -\ln[\tan(\theta/2)]$, where θ is the polar angle with the origin at the primary vertex. The detector pseudorapidity, or η^{det}, measures θ as the polar angle with origin at the center of the detector.

[31] E.E. Boos, V.E. Bunichev, L.V. Dudko, V.I. Savrin, and V.V. Sherstnev, Phys. Atom. Nucl. **69**, 1317 (2006).

[32] E. Boos, V. Bunichev, M. Dubinin, L. Dudko, V. Ilyin, A. Kryukov, V Edneral, V. Savrin, A. Semenov, and A. Sherstnev (CompHEP Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **534**, 250 (2004).

[33] M.L. Mangano, F. Piccinini, A.D. Polosa, M. Moretti, and R. Pittau, J. High Energy Phys. 07 (2003) 001. We used ALPGEN version 2.05.

[34] S. Höche, F. Krauss, N. Lavesson, L. Lönnblad, M. Mangano, A. Schälicke, and S. Schumann, in *Proceedings of the Workshop on the Implications of HERA for LHC Physics* edited by A. De Roeck and H. Jung (DESY, Hamburg, 2005), p. 288.

[35] M.L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, J. High Energy Phys. 01 (2007) 013.

[36] T. Sjöstrand, L. Lönnblad, S. Mrenna, and P. Skands, hep-ph/0308153. We used PYTHIA version 6.323.

[37] V.M. Abazov *et al.* (D0 Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **565**, 463 (2006).

[38] R. Frühwirth, W. Waltenberger, and P. Vanlaer, CERN-CMS-NOTE-2007-008.

[39] Jets are defined using the iterative seed-based cone algorithm with radius $\mathcal{R} = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.5$, including midpoints as described on pp. 47–77 in G.C. Blazey *et al.*, in *Proceedings of the Workshop on QCD and Weak Boson Physics in Run II*, edited by U. Baur, R.K. Ellis, and D. Zeppenfeld, FERMILAB-PUB-00-297 (2000).

[40] T. Scanlon, Ph.D. thesis, Imperial College, University of London, October 2006.

[41] J. Pumplin *et al.*, J. High Energy Phys. 07 (2002) 012. We used version CTEQ6L1.

[42] E. Boos, L. Dudko, and T. Ohl, Eur. Phys. J. C **11**, 473 (1999).

[43] S. Jadach *et al.*, Comput. Phys. Commun. **76**, 361 (1993). We used TAUOLA version 2.5.

[44] D.J. Lange, Nucl. Instrum. Methods Phys. Res., Sect. A **462**, 152 (2001). We used EVTGEN version 00.00.17.

[45] R.K. Ellis and S. Veseli, Phys. Rev. D **60**, 011501 (1999).

[46] F. Febres Cordero, L. Reina, and D. Wackeroth, Phys. Rev. D **74**, 034007 (2006).

[47] J. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, Phys. Rev. D **75**, 054015 (2007).

[48] R. Brun and F. Carminati, CERN Program Library Long Writeup, Report No. W5013, 1994.

[49] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. D **74**, 112004 (2006).

[50] T. Andeen, B.C.K. Casey, K. DeVaughan, Y. Enari, E. Gallas, D. Krop, R. Partridge, H. Schellman, G.R. Snow, S. Yacoob, and H.D. Yoo, Fermi National Accelerator Laboratory Technical Memorandum No. 2365, (2007).

[51] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, P. Richardson, M.H. Seymour, and B.R. Webber, arXiv:hep-ph/0210213, revised October 2005.

[52] R. Barlow, J. Comp. Phys. **72**, 202 (1987).

[53] L. Dudko, AIP Conf. Proc. **583**, 83 (2001).

[54] E. Boos and L. Dudko, Nucl. Instrum. Methods Phys. Res., Sect. A **502**, 486 (2003).

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

[56] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen, *Classification and Regression Trees* (Wadsworth, Stamford, 1984).

[57] D. Bowser-Chao and D.L. Dzialo, Phys. Rev. D **47**, 1900 (1993).

[58] C. Gini, *Variabilità e Mutabilità* (1912), reprinted in *Memorie di Metodologica Statistica*, edited by E. Pizetti and T. Salvemini, Rome: Libreria Eredi Virgilio Veschi (1955).

[59] Y. Freund and R.E. Schapire, in *Machine Learning: Proceedings of the Thirteenth International Conference*, edited by L. Saitta (Morgan Kaufmann, San Fransisco, 1996) p. 148.

[60] B.P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, Nucl. Instrum. Methods Phys. Res., Sect. A **543**, 577 (2005).

[61] H.-J. Yang, B.P. Roe, and J. Zhu, Nucl. Instrum. Methods Phys. Res., Sect. A **555**, 370 (2005).

[62] Q.-H. Cao, R. Schwienhorst, and C.-P Yuan, Phys. Rev. D **71**, 054023 (2005).

[63] Q.-H. Cao, R. Schwienhorst, J.A. Benitez, R. Brock, and C.-P Yuan, Phys. Rev. D **72**, 094027 (2005).

[64] G. Mahlon and S.J. Parke, Phys. Rev. D **55**, 7249 (1997).

[65] C.M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon Press, Oxford, 1998).

[66] E.K. Blum and L.K. Li, Neural Networks **4**, 511 (1991).

[67] D.W. Ruck, S.K. Rogers, M. Kabrinsky, M.E. Oxley, and B.W. Suter, IEEE Trans. Neural Networks **1**, 296 (1990).

[68] E.A. Wan, IEEE Trans. Neural Networks **1**, 303 (1990).

[69] R.M. Neal, *Bayesian Learning of Neural Networks* (Springer-Verlag, New York, 1996).

[70] P.C. Bhat and H.B. Prosper, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, edited by L. Lyons and M.K. Ünel (Imperial College Press,

London, England, 2006), p. 151.
[71] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth, Phys. Lett. B **195**, 216 (1987).
[72] B.A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004).
[73] R. Neal, "Software for Flexible Modeling and Markov Chain Sampling," `http://www.cs.toronto.edu/~radford/fbm.software.html`.
[74] J.H. Friedman, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology*, edited by L. Lyons and M. K. Ünel (Imperial College Press, London, England, 2006), p. 127, `http://www-stat.stanford.edu/~jhf/R-RuleFit.html`.
[75] V.M. Abazov *et al.* (D0 Collaboration), Nature **429**, 638 (2004).
[76] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. D **74**, 092005 (2006).
[77] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **96**, 152002 (2006).
[78] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. D **74**, 032009 (2006).
[79] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. D **75**, 031105(R) (2007).
[80] V.M. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **617**, 1 (2005).
[81] M.R. Whalley, D. Bourikov, and R.C. Group, in *Proceedings of HERA and the LHC, a Workshop on the Implications of HERA for LHC Physics*, edited by A. De Roeck and H. Jung (DESY, Hamburg, 2005), p. 575, `http://hepforge.cedar.ac.uk/lhapdf`.
[82] G.P. Lepage, J. Comput. Phys. **27**, 192 (1978).
[83] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi, *GNU Scientific Library Reference Manual – Revised Second Edition* (Network Theory Limited, Bristol, 2006), `http://www.gnu.org/software/gsl/`.
[84] F. Maltoni and T. Stelzer, J. High Energy Phys. 02 (2003) 027.
[85] H. Murayama, I. Watanabe, and K. Hagiwara, High Energy Accelerator Research Organization (KEK) Report No. 11, 1991 (unpublished).
[86] I. Bertram, G. Landsberg, J. Linneman, R. Partridge, M. Paterno, and H.B. Prosper, Fermi National Accelerator Laboratory Technical Memorandum No. 2104, (2000), and references therein.
[87] E.T. Jaynes and L. Bretthorst, *Probability Theory: the Logic of Science* (Cambridge University Press, Cambridge, England, 2003).
[88] T. Tait and C.P. Yuan, Phys. Rev. D **63**, 014018 (2001).
[89] L. Lyons, D. Gibaut, and P. Clifford, Nucl. Instrum. Methods Phys. Res., Sect. A **270**, 110 (1988).
[90] R.J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences* (The Manchester Physics Series, John Wiley and Sons, New York, 1989).
[91] G. Cowan, *Statistical Data Analysis* (Oxford University Press, Oxford, 1998).
[92] W.-M. Yao *et al.* (Particle Data Group), J. Phys. G **33**, 1 (2006).
[93] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **99**, 191802 (2007).
[94] V.M. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **641**, 423 (2006).
[95] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **95**, 102002 (2005).
[96] V.M. Abazov *et al.* (D0 Collaboration), Phys. Lett. B **639**, 616 (2006).
[97] G.L. Kane, G.A. Ladinsky, and C.-P. Yuan, Phys. Rev. D **45**, 124 (1992).
[98] K. Whisnant, J.M. Yang, B.-L. Young, and X. Zhang, Phys. Rev. D **56**, 467 (1997).
[99] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. D **75**, 052001 (2007).
[100] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. Lett. **100**, 062004 (2008).