FERMILAB-CONF-07-261-E

Search for New Particles Leading to Z+jets Final States in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV

T. Aaltonen,[23] A. Abulencia,[24] J. Adelman,[13] T. Affolder,[10] T. Akimoto,[55] M.G. Albrow,[17] S. Amerio,[43] D. Amidei,[35] A. Anastassov,[52] K. Anikeev,[17] A. Annovi,[19] J. Antos,[14] M. Aoki,[55] G. Apollinari,[17] T. Arisawa,[57] A. Artikov,[15] W. Ashmanskas,[17] A. Attal,[3] A. Aurisano,[53] F. Azfar,[42] P. Azzi-Bacchetta,[43] P. Azzurri,[46] N. Bacchetta,[43] W. Badgett,[17] A. Barbaro-Galtieri,[29] V.E. Barnes,[48] B.A. Barnett,[25] S. Baroiant,[7] V. Bartsch,[31] G. Bauer,[33] P.-H. Beauchemin,[34] F. Bedeschi,[46] S. Behari,[25] G. Bellettini,[46] J. Bellinger,[59] A. Belloni,[33] D. Benjamin,[16] A. Beretvas,[17] J. Beringer,[29] T. Berry,[30] A. Bhatti,[50] M. Binkley,[17] D. Bisello,[43] I. Bizjak,[31] R.E. Blair,[2] C. Blocker,[6] B. Blumenfeld,[25] A. Bocci,[16] A. Bodek,[49] V. Boisvert,[49] G. Bolla,[48] A. Bolshov,[33] D. Bortoletto,[48] J. Boudreau,[47] A. Boveia,[10] B. Brau,[10] L. Brigliadori,[5] C. Bromberg,[36] E. Brubaker,[13] J. Budagov,[15] H.S. Budd,[49] S. Budd,[24] K. Burkett,[17] G. Busetto,[43] P. Bussey,[21] A. Buzatu,[34] K. L. Byrum,[2] S. Cabrera[q],[16] M. Campanelli,[20] M. Campbell,[35] F. Canelli,[17] A. Canepa,[45] S. Carrillo[i],[18] D. Carlsmith,[59] R. Carosi,[46] S. Carron,[34] B. Casal,[11] M. Casarsa,[54] A. Castro,[5] P. Catastini,[46] D. Cauz,[54] M. Cavalli-Sforza,[3] A. Cerri,[29] L. Cerrito[m],[31] S.H. Chang,[28] Y.C. Chen,[1] M. Chertok,[7] G. Chiarelli,[46] G. Chlachidze,[17] F. Chlebana,[17] I. Cho,[28] K. Cho,[28] D. Chokheli,[15] J.P. Chou,[22] G. Choudalakis,[33] S.H. Chuang,[52] K. Chung,[12] W.H. Chung,[59] Y.S. Chung,[49] M. Cilijak,[46] C.I. Ciobanu,[24] M.A. Ciocci,[46] A. Clark,[20] D. Clark,[6] M. Coca,[16] G. Compostella,[43] M.E. Convery,[50] J. Conway,[7] B. Cooper,[31] K. Copic,[35] M. Cordelli,[19] G. Cortiana,[43] F. Crescioli,[46] C. Cuenca Almenar[q],[7] J. Cuevas[l],[11] R. Culbertson,[17] J.C. Cully,[35] S. DaRonco,[43] M. Datta,[17] S. D'Auria,[21] T. Davies,[21] D. Dagenhart,[17] P. de Barbaro,[49] S. De Cecco,[51] A. Deisher,[29] G. De Lentdecker[c],[49] G. De Lorenzo,[3] M. Dell'Orso,[46] F. Delli Paoli,[43] L. Demortier,[50] J. Deng,[16] M. Deninno,[5] D. De Pedis,[51] P.F. Derwent,[17] G.P. Di Giovanni,[44] C. Dionisi,[51] B. Di Ruzza,[54] J.R. Dittmann,[4] M. D'Onofrio,[3] C. Dörr,[26] S. Donati,[46] P. Dong,[8] J. Donini,[43] T. Dorigo,[43] S. Dube,[52] J. Efron,[39] R. Erbacher,[7] D. Errede,[24] S. Errede,[24] R. Eusebi,[17] H.C. Fang,[29] S. Farrington,[30] I. Fedorko,[46] W.T. Fedorko,[13] R.G. Feild,[60] M. Feindt,[26] J.P. Fernandez,[32] R. Field,[18] G. Flanagan,[48] R. Forrest,[7] S. Forrester,[7] M. Franklin,[22] J.C. Freeman,[29] I. Furic,[13] M. Gallinaro,[50] J. Galyardt,[12] J.E. Garcia,[46] F. Garberson,[10] A.F. Garfinkel,[48] C. Gay,[60] H. Gerberich,[24] D. Gerdes,[35] S. Giagu,[51] P. Giannetti,[46] K. Gibson,[47] J.L. Gimmell,[49] C. Ginsburg,[17] N. Giokaris[a],[15] M. Giordani,[54] P. Giromini,[19] M. Giunta,[46] G. Giurgiu,[25] V. Glagolev,[15] D. Glenzinski,[17] M. Gold,[37] N. Goldschmidt,[18] J. Goldstein[b],[42] A. Golossanov,[17] G. Gomez,[11] G. Gomez-Ceballos,[33] M. Goncharov,[53] O. González,[32] I. Gorelov,[37] A.T. Goshaw,[16] K. Goulianos,[50] A. Gresele,[43] S. Grinstein,[22] C. Grosso-Pilcher,[13] R.C. Group,[17] U. Grundler,[24] J. Guimaraes da Costa,[22] Z. Gunay-Unalan,[36] C. Haber,[29] K. Hahn,[33] S.R. Hahn,[17] E. Halkiadakis,[52] A. Hamilton,[20] B.-Y. Han,[49] J.Y. Han,[49] R. Handler,[59] F. Happacher,[19] K. Hara,[55] D. Hare,[52] M. Hare,[56] S. Harper,[42] R.F. Harr,[58] R.M. Harris,[17] M. Hartz,[47] K. Hatakeyama,[50] J. Hauser,[8] C. Hays,[42] M. Heck,[26] A. Heijboer,[45] B. Heinemann,[29] J. Heinrich,[45] C. Henderson,[33] M. Herndon,[59] J. Heuser,[26] D. Hidas,[16] C.S. Hill[b],[10] D. Hirschbuehl,[26] A. Hocker,[17] A. Holloway,[22] S. Hou,[1] M. Houlden,[30] S.-C. Hsu,[9] B.T. Huffman,[42] R.E. Hughes,[39] U. Husemann,[60] J. Huston,[36] J. Incandela,[10] G. Introzzi,[46] M. Iori,[51] A. Ivanov,[7] B. Iyutin,[33] E. James,[17] D. Jang,[52] B. Jayatilaka,[16] D. Jeans,[51] E.J. Jeon,[28] S. Jindariani,[18] W. Johnson,[7] M. Jones,[48] K.K. Joo,[28] S.Y. Jun,[12] J.E. Jung,[28] T.R. Junk,[24] T. Kamon,[53] P.E. Karchin,[58] Y. Kato,[41] Y. Kemp,[26] R. Kephart,[17] U. Kerzel,[26] V. Khotilovich,[53] B. Kilminster,[39] D.H. Kim,[28] H.S. Kim,[28] J.E. Kim,[28] M.J. Kim,[17] S.B. Kim,[28] S.H. Kim,[55] Y.K. Kim,[13] N. Kimura,[55] L. Kirsch,[6] S. Klimenko,[18] M. Klute,[33] B. Knuteson,[33] B.R. Ko,[16] K. Kondo,[57] D.J. Kong,[28] J. Konigsberg,[18] A. Korytov,[18] A.V. Kotwal,[16] A.C. Kraan,[45] J. Kraus,[24] M. Kreps,[26] J. Kroll,[45] N. Krumnack,[4] M. Kruse,[16] V. Krutelyov,[10] T. Kubo,[55] S. E. Kuhlmann,[2] T. Kuhr,[26] N.P. Kulkarni,[58] Y. Kusakabe,[57] S. Kwang,[13] A.T. Laasanen,[48] S. Lai,[34] S. Lami,[46] S. Lammel,[17] M. Lancaster,[31] R.L. Lander,[7] K. Lannon,[39] A. Lath,[52] G. Latino,[46] I. Lazzizzera,[43] T. LeCompte,[2] J. Lee,[49] J. Lee,[28] Y.J. Lee,[28] S.W. Lee[o],[53] R. Lefèvre,[20] N. Leonardo,[33] S. Leone,[46] S. Levy,[13] J.D. Lewis,[17] C. Lin,[60] C.S. Lin,[17] M. Lindgren,[17] E. Lipeles,[9] A. Lister,[7] D.O. Litvintsev,[17] T. Liu,[17] N.S. Lockyer,[45] A. Loginov,[60] M. Loreti,[43] R.-S. Lu,[1] D. Lucchesi,[43] P. Lujan,[29] P. Lukens,[17] G. Lungu,[18] L. Lyons,[42] J. Lys,[29] R. Lysak,[14] E. Lytken,[48] P. Mack,[26] D. MacQueen,[34] R. Madrak,[17] K. Maeshima,[17] K. Makhoul,[33] T. Maki,[23] P. Maksimovic,[25] S. Malde,[42] S. Malik,[31] G. Manca,[30] A. Manousakis[a],[15] F. Margaroli,[5] R. Marginean,[17] C. Marino,[26] C.P. Marino,[24] A. Martin,[60] M. Martin,[25] V. Martin[g],[21] M. Martínez,[3] R. Martínez-Ballarín,[32] T. Maruyama,[55] P. Mastrandrea,[51] T. Masubuchi,[55] H. Matsunaga,[55] M.E. Mattson,[58] R. Mazini,[34] P. Mazzanti,[5] K.S. McFarland,[49] P. McIntyre,[53] R. McNulty[f],[30] A. Mehta,[30] P. Mehtala,[23] S. Menzemer[h],[11] A. Menzione,[46] P. Merkel,[48] C. Mesropian,[50] A. Messina,[36] T. Miao,[17] N. Miladinovic,[6] J. Miles,[33] R. Miller,[36] C. Mills,[10] M. Milnik,[26] A. Mitra,[1] G. Mitselmakher,[18] A. Miyamoto,[27] S. Moed,[20] N. Moggi,[5] B. Mohr,[8] C.S. Moon,[28]

R. Moore,[17] M. Morello,[46] P. Movilla Fernandez,[29] J. Mülmenstädt,[29] A. Mukherjee,[17] Th. Muller,[26] R. Mumford,[25] P. Murat,[17] M. Mussini,[5] J. Nachtman,[17] A. Nagano,[55] J. Naganoma,[57] K. Nakamura,[55] I. Nakano,[40] A. Napier,[56] V. Necula,[16] C. Neu,[45] M.S. Neubauer,[9] J. Nielsen[n],[29] L. Nodulman,[2] O. Norniella,[3] E. Nurse,[31] S.H. Oh,[16] Y.D. Oh,[28] I. Oksuzian,[18] T. Okusawa,[41] R. Oldeman,[30] R. Orava,[23] K. Osterberg,[23] C. Pagliarone,[46] E. Palencia,[11] V. Papadimitriou,[17] A. Papaikonomou,[26] A.A. Paramonov,[13] B. Parks,[39] S. Pashapour,[34] J. Patrick,[17] G. Pauletta,[54] M. Paulini,[12] C. Paus,[33] D.E. Pellett,[7] A. Penzo,[54] T.J. Phillips,[16] G. Piacentino,[46] J. Piedra,[44] L. Pinera,[18] K. Pitts,[24] C. Plager,[8] L. Pondrom,[59] X. Portell,[3] O. Poukhov,[15] N. Pounder,[42] F. Prakoshyn,[15] A. Pronko,[17] J. Proudfoot,[2] F. Ptohos[e],[19] G. Punzi,[46] J. Pursley,[25] J. Rademacker[b],[42] A. Rahaman,[47] V. Ramakrishnan,[59] N. Ranjan,[48] I. Redondo,[32] B. Reisert,[17] V. Rekovic,[37] P. Renton,[42] M. Rescigno,[51] S. Richter,[26] F. Rimondi,[5] L. Ristori,[46] A. Robson,[21] T. Rodrigo,[11] E. Rogers,[24] S. Rolli,[56] R. Roser,[17] M. Rossi,[54] R. Rossin,[10] P. Roy,[34] A. Ruiz,[11] J. Russ,[12] V. Rusu,[13] H. Saarikko,[23] A. Safonov,[53] W.K. Sakumoto,[49] G. Salamanna,[51] O. Saltó,[3] L. Santi,[54] S. Sarkar,[51] L. Sartori,[46] K. Sato,[17] P. Savard,[34] A. Savoy-Navarro,[44] T. Scheidle,[26] P. Schlabach,[17] E.E. Schmidt,[17] M.P. Schmidt,[60] M. Schmitt,[38] T. Schwarz,[7] L. Scodellaro,[11] A.L. Scott,[10] A. Scribano,[46] F. Scuri,[46] A. Sedov,[48] S. Seidel,[37] Y. Seiya,[41] A. Semenov,[15] L. Sexton-Kennedy,[17] A. Sfyrla,[20] S.Z. Shalhout,[58] M.D. Shapiro,[29] T. Shears,[30] P.F. Shepard,[47] D. Sherman,[22] M. Shimojima[k],[55] M. Shochet,[13] Y. Shon,[59] I. Shreyber,[20] A. Sidoti,[46] P. Sinervo,[34] A. Sisakyan,[15] A.J. Slaughter,[17] J. Slaunwhite,[39] K. Sliwa,[56] J.R. Smith,[7] F.D. Snider,[17] R. Snihur,[34] M. Soderberg,[35] A. Soha,[7] S. Somalwar,[52] V. Sorin,[36] J. Spalding,[17] F. Spinella,[46] T. Spreitzer,[34] P. Squillacioti,[46] M. Stanitzki,[60] A. Staveris-Polykalas,[46] R. St. Denis,[21] B. Stelzer,[8] O. Stelzer-Chilton,[42] D. Stentz,[38] J. Strologas,[37] D. Stuart,[10] J.S. Suh,[28] A. Sukhanov,[18] H. Sun,[56] I. Suslov,[15] T. Suzuki,[55] A. Taffard[p],[24] R. Takashima,[40] Y. Takeuchi,[55] R. Tanaka,[40] M. Tecchio,[35] P.K. Teng,[1] K. Terashi,[50] J. Thom[d],[17] A.S. Thompson,[21] E. Thomson,[45] P. Tipton,[60] V. Tiwari,[12] S. Tkaczyk,[17] D. Toback,[53] S. Tokar,[14] K. Tollefson,[36] T. Tomura,[55] D. Tonelli,[46] S. Torre,[19] D. Torretta,[17] S. Tourneur,[44] W. Trischuk,[34] S. Tsuno,[40] Y. Tu,[45] N. Turini,[46] F. Ukegawa,[55] S. Uozumi,[55] S. Vallecorsa,[20] N. van Remortel,[23] A. Varganov,[35] E. Vataga,[37] F. Vazquez[i],[18] G. Velev,[17] C. Vellidis[a],[46] G. Veramendi,[24] V. Veszpremi,[48] M. Vidal,[32] R. Vidal,[17] I. Vila,[11] R. Vilar,[11] T. Vine,[31] M. Vogel,[37] I. Vollrath,[34] I. Volobouev[o],[29] G. Volpi,[46] F. Würthwein,[9] P. Wagner,[53] R.G. Wagner,[2] R.L. Wagner,[17] J. Wagner,[26] W. Wagner,[26] R. Wallny,[8] S.M. Wang,[1] A. Warburton,[34] D. Waters,[31] M. Weinberger,[53] W.C. Wester III,[17] B. Whitehouse,[56] D. Whiteson[p],[45] A.B. Wicklund,[2] E. Wicklund,[17] G. Williams,[34] H.H. Williams,[45] P. Wilson,[17] B.L. Winer,[39] P. Wittich[d],[17] S. Wolbers,[17] C. Wolfe,[13] T. Wright,[35] X. Wu,[20] S.M. Wynne,[30] A. Yagil,[9] K. Yamamoto,[41] J. Yamaoka,[52] T. Yamashita,[40] C. Yang,[60] U.K. Yang[j],[13] Y.C. Yang,[28] W.M. Yao,[29] G.P. Yeh,[17] J. Yoh,[17] K. Yorita,[13] T. Yoshida,[41] G.B. Yu,[49] I. Yu,[28] S.S. Yu,[17] J.C. Yun,[17] L. Zanello,[51] A. Zanetti,[54] I. Zaw,[22] X. Zhang,[24] J. Zhou,[52] and S. Zucchelli[5]

(CDF Collaboration*)

[1]*Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China*
[2]*Argonne National Laboratory, Argonne, Illinois 60439*
[3]*Institut de Fisica d'Altes Energies, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain*
[4]*Baylor University, Waco, Texas 76798*
[5]*Istituto Nazionale di Fisica Nucleare, University of Bologna, I-40127 Bologna, Italy*
[6]*Brandeis University, Waltham, Massachusetts 02254*
[7]*University of California, Davis, Davis, California 95616*
[8]*University of California, Los Angeles, Los Angeles, California 90024*
[9]*University of California, San Diego, La Jolla, California 92093*
[10]*University of California, Santa Barbara, Santa Barbara, California 93106*
[11]*Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain*
[12]*Carnegie Mellon University, Pittsburgh, PA 15213*
[13]*Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637*
[14]*Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia*
[15]*Joint Institute for Nuclear Research, RU-141980 Dubna, Russia*
[16]*Duke University, Durham, North Carolina 27708*
[17]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510*
[18]*University of Florida, Gainesville, Florida 32611*
[19]*Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy*
[20]*University of Geneva, CH-1211 Geneva 4, Switzerland*
[21]*Glasgow University, Glasgow G12 8QQ, United Kingdom*
[22]*Harvard University, Cambridge, Massachusetts 02138*
[23]*Division of High Energy Physics, Department of Physics, University of Helsinki and Helsinki Institute of Physics, FIN-00014, Helsinki, Finland*
[24]*University of Illinois, Urbana, Illinois 61801*

[25]The Johns Hopkins University, Baltimore, Maryland 21218
[26]Institut für Experimentelle Kernphysik, Universität Karlsruhe, 76128 Karlsruhe, Germany
[27]High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305, Japan
[28]Center for High Energy Physics: Kyungpook National University, Taegu 702-701, Korea; Seoul National University, Seoul 151-742, Korea; SungKyunKwan University, Suwon 440-746, Korea
[29]Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720
[30]University of Liverpool, Liverpool L69 7ZE, United Kingdom
[31]University College London, London WC1E 6BT, United Kingdom
[32]Centro de Investigaciones Energeticas Medioambientales y Tecnologicas, E-28040 Madrid, Spain
[33]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
[34]Institute of Particle Physics: McGill University, Montréal, Canada H3A 2T8; and University of Toronto, Toronto, Canada M5S 1A7
[35]University of Michigan, Ann Arbor, Michigan 48109
[36]Michigan State University, East Lansing, Michigan 48824
[37]University of New Mexico, Albuquerque, New Mexico 87131
[38]Northwestern University, Evanston, Illinois 60208
[39]The Ohio State University, Columbus, Ohio 43210
[40]Okayama University, Okayama 700-8530, Japan
[41]Osaka City University, Osaka 588, Japan
[42]University of Oxford, Oxford OX1 3RH, United Kingdom
[43]University of Padova, Istituto Nazionale di Fisica Nucleare, Sezione di Padova-Trento, I-35131 Padova, Italy
[44]LPNHE, Universite Pierre et Marie Curie/IN2P3-CNRS, UMR7585, Paris, F-75252 France
[45]University of Pennsylvania, Philadelphia, Pennsylvania 19104
[46]Istituto Nazionale di Fisica Nucleare Pisa, Universities of Pisa, Siena and Scuola Normale Superiore, I-56127 Pisa, Italy
[47]University of Pittsburgh, Pittsburgh, Pennsylvania 15260
[48]Purdue University, West Lafayette, Indiana 47907
[49]University of Rochester, Rochester, New York 14627
[50]The Rockefeller University, New York, New York 10021
[51]Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1, University of Rome "La Sapienza," I-00185 Roma, Italy
[52]Rutgers University, Piscataway, New Jersey 08855
[53]Texas A&M University, College Station, Texas 77843
[54]Istituto Nazionale di Fisica Nucleare, University of Trieste/ Udine, Italy
[55]University of Tsukuba, Tsukuba, Ibaraki 305, Japan
[56]Tufts University, Medford, Massachusetts 02155
[57]Waseda University, Tokyo 169, Japan
[58]Wayne State University, Detroit, Michigan 48201
[59]University of Wisconsin, Madison, Wisconsin 53706
[60]Yale University, New Haven, Connecticut 06520

(Dated: June 15, 2007)

We present the results of a search for new particles that lead to a Z boson plus jets in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using the Collider Detector at Fermilab (CDF II). A data sample with a luminosity of 1.06 fb^{-1} collected using Z boson decays to ee and $\mu\mu$ is used. We describe a completely data-based method to predict the dominant background from standard-model Z+jet events. This method can be similarly applied to other analyses requiring background predictions in multi-jet environments, as shown when validating the method by predicting the background from W+jets in $t\bar{t}$ production. No significant excess above the background prediction is observed, and a limit is set using a fourth generation quark model to quantify the acceptance. Assuming $BR(b' \to bZ) = 100\%$ and using a leading-order calculation of the b' cross section, b' quark masses below 268 GeV/c^2 are excluded at 95% confidence level.

*With visitors from [a]University of Athens, 15784 Athens, Greece, [b]University of Bristol, Bristol BS8 1TL, United Kingdom, [c]University Libre de Bruxelles, B-1050 Brussels, Belgium, [d]Cornell University, Ithaca, NY 14853, [e]University of Cyprus, Nicosia CY-1678, Cyprus, [f]University College Dublin, Dublin 4, Ireland, [g]University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom, [h]University of Heidelberg, D-69120 Heidelberg, Germany, [i]Universidad Iberoamericana, Mexico D.F., Mexico, [j]University of Manchester, Manchester M13 9PL, England, [k]Nagasaki Institute of Applied Science, Nagasaki, Japan, [l]University de Oviedo,

I. INTRODUCTION

This paper presents a search for new particles decaying to Z gauge bosons created in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV with the CDF II detector at the Fermilab Tevatron, extending and complementing other work with such final states [1–4]. A variety of extensions to the standard model predict new particles with couplings to Z bosons [5–9]. We wish to discover or rule out these types of models, while maintaining model independence in the search. That is, while these theories offer guidance about the possible characteristics of physics beyond the standard model, they do not necessarily correspond to what actually exists in nature, and so the analysis is not tailored to specific models.

Of course, some assumptions are necessary in choosing how to discriminate between the standard model background and new signals. We examine final states with Z bosons and additional jets. In particular, we focus on final states in which there are at least 3 jets, each with at least 30 GeV of transverse energy E_T. This assumption was motivated by studying the optimal kinematic selection of a specific model, the fourth generation model [5]. In the fourth generation model, an additional pair of heavy quarks is added to the standard model's three. The production mechanisms of the new down-type quark (called the b') would be identical to that of the top quark, with pair-production having the largest cross section. Depending on its mass, the direct tree-level decays of the b' could be either kinematically forbidden or heavily Cabibbo-suppressed. These situations could give rise to a large branching ratio of $b' \to bZ$ via a loop diagram. While the selection was chosen as the optimal set of kinematic cuts using this model as a signal, this analysis constrains all models with Z+3 jet final states.

The dominant background for this final state is from standard model Z production with jets from higher order QCD processes. A leading order calculation of this background is insufficient. Use of higher order calculations is complicated because it involves hard-scattering matrix elements in combination with soft non-perturbative QCD processes. Recent NLO predictions [10] have been used [11] with the aid of Monte Carlo simulations to account for the non-perturbative overlap. Any such method requires validation with data. In this paper, we develop a different approach that uses the data as more than a validation tool, and uses it alone for the background estimation. In this approach, we extrapolate the jet transverse energy distributions from a low energy control region of the data into the high energy signal region.

E-33007 Oviedo, Spain, [m]University of London, Queen Mary College, London, E1 4NS, England, [n]University of California Santa Cruz, Santa Cruz, CA 95064, [o]Texas Tech University, Lubbock, TX 79409, [p]University of California, Irvine, Irvine, CA 92697, [q]IFIC(CSIC-Universitat de Valencia), 46071 Valencia, Spain,

This paper is organized as follows. Section II contains a brief overview of the portions of the CDF II detector relevant to this measurement. Section III lists the trigger requirements, and describes and motivates the signal sample selections. Section IV lists the backgrounds. Section V describes, validates, and applies the method of predicting the dominant background. In Sec. VI the predictions for the remaining backgrounds are described. In Sec. VII we present the results of the search, and conclude in Sec. VIII.

II. THE CDF II DETECTOR

The CDF II detector is described in detail elsewhere [12]; here, only the portions required for this analysis are described. We first describe the coordinate system conventions. In the CDF coordinate system, the origin is the center of the detector, and the z axis is along the beam axis, with positive z defined as the proton beam direction. The x axis points radially outward from the Tevatron ring, leaving the y axis direction perpendicular to the earth's surface with positive direction upward. Spherical coordinates are used where appropriate, in which θ is the polar angle (zero in the positive z direction), ϕ is the azimuthal angle (zero in the positive x direction), and the pseudorapidity η is defined by $\eta \equiv -\ln[\tan(\theta/2)]$. At hadron colliders, transverse energies and momenta are usually the appropriate physical quantities, defined by $E_T \equiv E \sin\theta$ and $p_T \equiv p \sin\theta$ (where E is a particle's energy and p is the magnitude of a particle's momentum).

A tracking system is situated directly outside the beam pipe and measures the trajectories and momenta of charged particles. The innermost part of the tracking system is the silicon detector, providing position measurements on up to 8 layers of sensors in the radial region $1.3 < r < 28$ cm and the polar region $|\eta| \lesssim 2.5$. Outside of this detector lies the central outer tracker (COT), an open-cell drift chamber providing measurements on up to 96 layers in the radial region $40 < r < 137$ cm and the polar region $|\eta| \lesssim 1$. Directly outside of the COT a solenoid provides a 1.4 T magnetic field, allowing particle momenta to be obtained from the trajectory measurements in this known field.

Surrounding the tracking system, segmented electromagnetic (EM) and hadronic calorimeters measure particle energies. In the central region, the calorimeters are arranged in a projective barrel geometry and cover the polar region $|\eta| < 1.2$. In the forward region, the calorimeters are arranged in a projective "end-plug" geometry and cover the polar region $1.2 < |\eta| < 3.5$. Two sets of drift chambers, one directly outside the hadronic calorimeter and another outside additional steel shielding, measure muon trajectories in the region $|\eta| < 0.6$; another set of drift chambers similarly detects muons in the region $0.6 < |\eta| < 1$. Muon scintillators surround these drift chambers in the region $|\eta| < 1$ for trig-

ger purposes. A luminosity measurement is provided by Cherenkov detectors in the region $3.7 < |\eta| < 4.7$ via a measurement of the average number of $p\bar{p}$ collisions per crossing [13].

Collision events of interest are selected for analysis offline using a three level trigger system, with each level accepting events for processing at the next level. At level 1, custom hardware enables fast decisions using rudimentary tracking information and a simple counting of reconstructed objects. At level 2, trigger processors enable decisions based on partial event reconstruction. At level 3, a computer farm running fast event reconstruction software makes the final decision on event storage.

III. DATA SAMPLE AND EVENT SELECTION

We first describe the baseline Z selection, and then describe the kinematic selection used to discriminate the potential signal from the standard model background. The kinematic selection is chosen and backgrounds are predicted *a priori*, before looking in the signal region. While remaining as data-driven as possible throughout the analysis, Monte Carlo simulation is used in some studies, consistency checks, and for illustration purposes. In all cases, the Monte Carlo events are generated with PYTHIA [14] and the detector responses are modeled with a GEANT simulation [15] as described in [16].

A. Baseline Z Selection

The data sample consists of $Z \to ee$ and $Z \to \mu\mu$ candidate events collected using single electron and muon triggers. The electron trigger requires at least one central electromagnetic energy cluster with $E_T > 18$ GeV and a matching track with $p_T > 9$ GeV/c. The muon trigger requires at least one central track with $p_T > 18$ GeV/c with matching hits in the muon drift chambers. The average integrated luminosity of these data samples is 1.06 fb^{-1} [17].

Z candidate events are selected offline by requiring at least one pair of electron or muon candidates both with $p_T > 20$ GeV/c and invariant mass in the range $81 < M_{\ell\ell} < 101$ GeV/c^2. The electron and muon identification variables are described in detail in Refs. [16, 18]. The selection is described briefly here. To increase efficiency, only one of the lepton pair has stringent identification requirements (the "tight" candidate), while on the other lepton the identification requirements are relaxed (the "loose" candidate).

"Loose" electron candidates consist of well-isolated EM calorimeter clusters with low energy in the hadronic calorimeter; in the central part of the detector ($|\eta| < 1.2$) well-measured tracks from the COT are required; in the forward parts of the detector ($|\eta| > 1.2$) no track is required, but the shower shape in the EM calorimeter is required to be consistent with that expected from electrons.

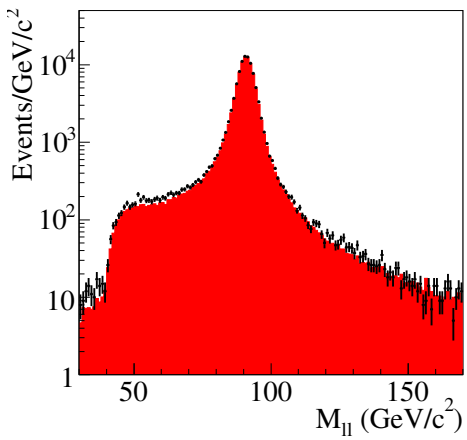

FIG. 1: Distribution of $M_{\ell\ell}$ of $Z \to ee$ and $Z \to \mu\mu$ data (black points and errors) using the baseline Z selection described in the text. Overlaid are standard model $Z \to ee$ and $Z \to \mu\mu$ Monte Carlo events, normalized to the number of events expected with the given luminosities using the expected cross section of 250 pb.

"Tight" electron candidates have all the requirements of "loose" candidates, and are additionally required to be central ($|\eta| < 1.2$), to have a shower shape consistent with that expected from electrons, to have calorimeter position and energy measurements consistent with its matching track, and to have no nearby tracks consistent with that expected in electrons from photon conversions.

"Loose" muon candidates consist of well-measured tracks in the COT and well-isolated EM and hadronic calorimeter clusters with minimal energy deposits. "Tight" muon candidates have all the requirements of "loose" candidates, and are additionally required to have matching hits in the muon drift chambers.

Finally, all electron and muon pairs are required to be consistent with originating from the same z vertex and to have a time-of-flight difference (as measured by the COT) inconsistent with that expected for cosmic rays. They are also required to be separated in ϕ by an angle greater than 5° to remove two lepton candidates misreconstructed from a single lepton.

Using this selection, the distribution of $M_{\ell\ell}$ is plotted and compared to standard model Z Monte Carlo simulation in Fig. 1.

B. Kinematic Selection

The analysis focuses on topologies with large numbers of highly energetic jets in the final state, for which the signal (from the decay of heavy objects) can be better separated from standard model Z+jet production. Jets are clustered using the "MIDPOINT" clustering algorithm [19] with a cone size of 0.4 radians. Corrections are ap-

plied to extrapolate the jet energies back to the parton level using a generic jet response [20]. Jets are required to have $|\eta| < 2$.

The following discriminators are used:

$$N_{\rm jet}^X = \text{Number of jets in the event with } E_T > X \text{ GeV}$$
$$J_T^X = \text{Scalar sum of } E_T \text{ of jets in the event with } E_T > X \text{ GeV}$$

The thresholds X as well as the cut values on these variables are determined by optimization [21]. In the optimization we use the figure of merit $S/(1.5+\sqrt{B})$ (where S is the expected number of signal events and B is the expected number of background events) to quantify the sensitivity as a compromise between best discovery and best limit potential [22, 23]. In the low background region ($B \ll 1$), maximizing this figure of merit is equivalent to maximizing the signal efficiency. In the high background region ($B \gg 1$), this figure of merit has the same behavior as $S/\sqrt{B}$. For the optimization study, $p\bar{p} \to b'\bar{b}'$ Monte Carlo simulations with a range of masses are used as the signal S. Standard model Z Monte Carlo simulations are used for the background B.

In order to be sensitive to a range of masses, we must take into account the generic behavior of new signals: as mass increases the cross section decreases while the transverse energy spectra become harder. Therefore, to be optimally sensitive to higher mass signals, we cut at larger values of $N_{\rm jet}$ and J_T thus removing more of the background to give sensitivity to the lower cross sections.

For the sake of simplicity, we desire that our selection only changes gradually with mass and uses the same E_T threshold on all jets. With a simple selection, the data-based background prediction method becomes easier. To confirm that this desire for simplicity does not considerably reduce the search sensitivity, and to understand what cut values and thresholds to use, we first establish a "target" selection. The "target" selection is defined as the selection with the highest sensitivity when placing cuts on the individual jet E_T's and J_T. This is found by scanning through all possible cuts on J_T^{10} (that is, J_T is calculated with a 10 GeV threshold on the jets) and all possible E_T thresholds for up to 4 jets (ordered by E_T), and finding the point with the optimal sensitivity. In this scan, step sizes of 10 GeV are used for the jet E_T thresholds, and a step size of 50 GeV is used for J_T^{10}. This scan is done independently for b' masses in the range $100 \leq m_{b'} \leq 350$ GeV/c^2 with a step size of 50 GeV/c^2.

The optimal points found by this scan for a b' mass of 150 GeV/c^2 are shown in column 2 of Table I. These cut values give the best possible sensitivity at this mass point when placing cuts on the individual jet E_T's and J_T^{10}. Again, we wish to choose a simple selection that gradually changes as a function of mass, and use the target sensitivities at all mass points for comparison. Based on the optimal target points for b' masses in the range $100 \leq m_{b'} \leq 350$ GeV/c^2, we choose the simpler requirements of $N_{\rm jet}^{30} \geq 3$ and $J_T^{10} > m_{b'}c^2$. The sensitivity of the simple requirements is compared to the target sensitivity in column 3 of Table I for the 150 GeV/c^2 mass point.

Variable	Values from scan	Values of simple selection
$E_T^{\rm jet\ 1}$ thresh.:	50	30
$E_T^{\rm jet\ 2}$ thresh.:	30	30
$E_T^{\rm jet\ 3}$ thresh.:	30	30
$E_T^{\rm jet\ 4}$ thresh.:	20	0
J_T^{10} cut:	0	150
$N_{\rm sig}$:	48.5	75.5
$N_{\rm bkg}$:	2.60	13.8
$S/(1.5+\sqrt{B})$:	15.6	14.5

TABLE I: Optimal point compared with the simple selection of $N_{jet}^{30} \geq 3$ and $J_T^{10} > 150$, for the $m_{b'} = 150$ GeV/c^2 mass point. Here, $N_{\rm sig}$ is the number of signal events expected in 1fb^{-1} after the given selection using b' Monte Carlo simulations. $N_{\rm bkg}$ is the number of background events expected in 1fb^{-1} after the given selection using standard model Z Monte Carlo simulations. In this optimization study, 2.7×10^5 standard model Z events were used; 1500 signal events were used (both counted before jet selection).

From the table it is apparent that, for $m_{b'} = 150$ GeV/c^2, the sensitivity of the simple cuts is only negligibly less than the target sensitivity. We find the same to be true for all mass points studied, except for the $m_{b'} = 100$ GeV/c^2 mass point. In that case, however, the sensitivity of the simple cuts is still adequate because of the larger cross sections for lower mass particles [24]. In addition, low masses near 100 GeV/c^2 are less interesting as they are already more tightly excluded [25]. Thus, we conclude that the simpler selection of $N_{\rm jet}^{30} \geq 3$ and $J_T^{10} > m_{b'}c^2$ is nearly optimal for the mass range of interest.

In the above, J_T was calculated using a 10 GeV E_T threshold on the jets. For the purposes of the background estimation, it is simpler to use the same E_T threshold on J_T as one uses on the $N_{\rm jet}$ variable. Therefore, a 30 GeV threshold is used when calculating J_T. This was found to give a small decrease in sensitivity in the b' model with the benefit of a gain in simplicity.

The kinematic jet selection was found to be optimal when using the fourth generation model as the signal. When optimizing using the figure of merit $S/(1.5+\sqrt{B})$, the optimal point is independent of the normalization of the signal. That is, any model with a different cross section but the same kinematic distributions will give the same optimal point. In addition, the shape of the kinematic distributions are mostly determined by the b' mass. We therefore expect that this selection is nearly optimal for all models with heavy particles produced in pairs and decaying to Z+jet. In general, this selection is sensitive to any model with high E_T jets in the final state. It may not be optimal for an arbitrary model, but designing a simple selection that is optimal for the entire

class of Z+high E_T jet models is not possible.

In this optimization, we assumed new signals would lead to final states consisting of a Z boson and many high E_T jets. Of course, some assumption about signal characteristics must be made in order to understand how to separate signal from background. These assumptions will naturally reduce the model independence of the search. There is a trade-off between the specificity of these assumptions and the sensitivity to a particular model. For example, in nearly all new physics models with Z boson final states, the transverse momentum spectrum of the Z is harder than for standard model Z production. This is because, in these models, the Z is usually a decay product of a massive particle. One would conclude that the Z transverse momentum is a very model-independent variable, and therefore well-motivated. However, we find, in the b' model sensitivity study, that the jet kinematic requirements have much higher sensitivity than the Z transverse momentum. The cost of this sensitivity is a loss of generality: with this assumption we are no longer sensitive to Z final states without high E_T jets. The sensitivity of the b' model can be further enhanced by requiring b jets using displaced vertices (because of the $b' \to bZ$ decay), again with a cost to generality. In our analysis, as a compromise between model independence and sensitivity, we choose to require additional jets in the event.

To summarize, after selecting $Z \to ee$ and $Z \to \mu\mu$ events, the kinematic selection is:

- $N_{\rm jet}^{30} \geq 3$, and
- $J_T^{30} > m_{b'}c^2$.

That is, Z events with $N_{\rm jet}^{30} \geq 3$ are selected, and the J_T^{30} distribution is scanned for an excess. Step sizes of 50 GeV are used.

IV. BACKGROUNDS

In the signal region described above, there are potential backgrounds from the following sources:

- single-Z production in conjunction with jets,
- multi-jet events, where two jets fake leptons,
- cosmic rays coincident with multi-jet events,
- WZ+jets, where the W decays to jets,
- ZZ+jets, where one of the Z's decays to jets,
- WW+jets, where both W's decay to leptons, and
- $t\bar{t}$+jets, where both W's decay to leptons.

The dominant background is from standard model single-Z production in conjunction with jets. Since beyond leading-log order diagrams make potentially large contributions to events with $N_{\rm jet}^{30} \geq 3$, calculation of this background from theoretical first principles is extremely difficult, and therefore would require careful validation with data. Rather than using data as merely a validation tool we take a different approach, and instead measure the background directly from data, and with data alone. The following section is devoted to describing this prediction technique for the dominant background from Z+jet. As this technique has not been applied previously, it is explained thoroughly, with careful validation studies described. The remaining backgrounds are estimated in Sec. VI.

V. DATA-BASED Z+JET BACKGROUND PREDICTION TECHNIQUE

Given the above selection, there are two tasks: the total number of background events with $N_{\rm jet}^{30} \geq 3$ must be predicted, and the shape of the J_T^{30} distribution after this cut must be predicted. When combined, these two components give the full normalized J_T^{30} distribution prediction. The background for events with $N_{\rm jet}^{30} \geq 3$ and any J_T^{30} cut can be obtained from this distribution. The method for predicting each of the two components is described separately in the following two sections.

In each of the prediction methods, fits to various jet E_T distributions are used. A parameterization that describes the shapes of these jet E_T distributions well is therefore required. The parameterization used is:

$$f(E_T) = p_0 \frac{e^{-E_T/p_1}}{(E_T)^{p_2}}, \tag{1}$$

where the p_i are fitted parameters. This parameterization was motivated by observations in Monte Carlo simulations, control regions of data, and phenomenological studies that: at low E_T, the jet E_T shape follows a power law function; at high E_T, it follows an exponential decay function. The above parameterization satisfies these limiting behaviors. With the above convention, the parameter p_1 has dimensions of energy, the parameter p_2 is dimensionless, and both parameters are positive. Further discussion and motivation for this parameterization is provided in [18].

A. Number of Events with $N_{\rm jet}^{30} \geq 3$

In order to predict the total number of events with $N_{\rm jet}^{30} \geq 3$, we use the jet E_T distributions in the $N_{\rm jet}^{30} \leq 2$ control regions. Since jets are counted above an E_T threshold (in this case 30 GeV), the $N_{\rm jet}$ distribution is completely determined from the jet E_T distributions. To illustrate this, and to describe the method, standard model $Z \to \mu\mu$ Monte Carlo simulations are used. After validation with control samples, the method is applied to the Z data.

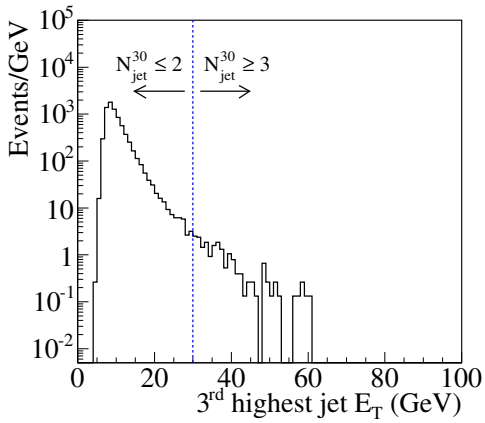

FIG. 2: E_T distribution of the third highest E_T jet in standard model $Z \to \mu\mu$ Monte Carlo simulations. Events with $N_{\rm jet}^{30} \leq 2$ have $E_T < 30$ GeV; events with $N_{\rm jet}^{30} \geq 3$ have $E_T > 30$ GeV.

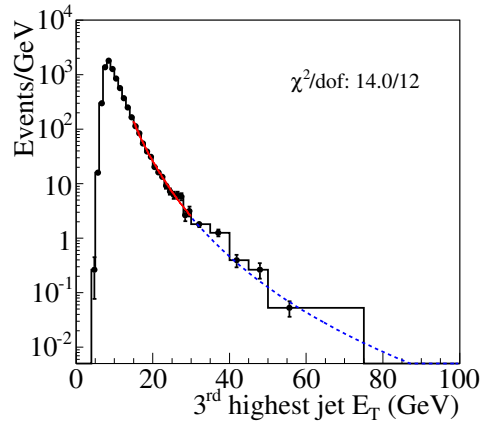

FIG. 3: E_T distribution of the third highest E_T jet in standard model $Z \to \mu\mu$ Monte Carlo events. The distribution is fit to Eq. (1) in the range $15 < E_T < 30$ GeV, and extrapolated to the $E_T > 30$ GeV region.

In Fig. 2, the E_T distribution of the third highest jet is shown. By construction, a cut on $N_{\rm jet}^{30} \leq 2$ separates this distribution into two regions. This distribution can be fit in the $E_T < 30$ GeV region and extrapolated to the $E_T > 30$ GeV region to get the expected number of background events with $N_{\rm jet}^{30} \geq 3$.

We fit the parameterization from Eq. (1) to the jet E_T distribution of Fig. 2, and show the results in Fig. 3 [26]. The fit matches well the broad features of the distribution above 30 GeV. The number of events with $N_{\rm jet}^{30} \geq 3$ is then predicted by integrating the fitted distribution from 30 GeV to infinity. The fit prediction obtained with this method (with its uncertainty from fit parameter error propagation described in Sec. V C) is 116^{+10}_{-13} events (with the number of generated Monte Carlo events having an equivalent luminosity of 7 fb^{-1}). The number of events observed in the simulated data with $N_{\rm jet}^{30} \geq 3$ is 152. In this case, the extrapolation predicts the background to within $31 \pm 16\%$. The level of consistency will be evaluated further in the validation studies with data in Sec. V D.

This method, using the jet E_T distributions to predict integrals of the $N_{\rm jet}$ distribution, can clearly be extended to other analyses as well. For illustration purposes only we describe other examples here, still using standard model $Z \to \mu\mu$ Monte Carlo simulation. Consider predicting the total number of events with $N_{\rm jet}^{80} \geq 1$ (that is, we require at least one jet with an E_T threshold of 80 GeV). In this case, a fit to the highest E_T jet distribution below 80 GeV can be extrapolated to above that threshold, as in Fig. 4. (Note that the highest E_T distribution in this figure is harder than the third highest E_T jet distribution, as one expects when ordering the jets by E_T). It is clear that the extrapolation describes the distribution reasonably well.

If we instead wish to predict the number of events with $N_{\rm jet}^{40} \geq 1$, we must fit the same E_T distribution below 40 GeV and extrapolate it to above that threshold, also shown in Fig. 4. It is clear that the extrapolation does not describe the high E_T portion of the distribution well. There is a large systematic uncertainty present in extrapolations that use such a small portion of the distribution that the shape can not be reliably obtained. This can be mitigated by raising the E_T threshold, unless the shape of the jet E_T distribution at high E_T can be otherwise constrained. In the case examined in this analysis, we fit the third highest E_T jet (which has a softer E_T distribution than the highest E_T jet) in the region $E_T < 30$ GeV. We have checked that the data in this region constrains the shape sufficiently with validation studies using control samples of data and Monte Carlo simulations, described later in Sec. V D.

From the above, it is apparent that one can estimate the background for events with $N_{\rm jet}^{X} \geq n$ by fitting the E_T distribution of the $n^{\rm th}$ highest E_T jet in the region $E_T < X$ and extrapolating the fit to the region $E_T > X$, as long as the fit region $E_T < X$ constrains the shape sufficiently.

B. J_T Shape Determination

We now describe the method used to determine the shape of the J_T^{30} distribution of events with $N_{\rm jet}^{30} \geq 3$. After finding the shape, it is then normalized to the number of events with $N_{\rm jet}^{30} \geq 3$ found by the above method. We again use standard model $Z \to \mu\mu$ Monte Carlo events to explain the method, and later will apply it to data.

Since J_T^{30} is simply the sum of the individual jet transverse energies above 30 GeV, if the E_T distributions of jets for events with $N_{\rm jet}^{30} \geq 3$ are known, the J_T^{30} distri-

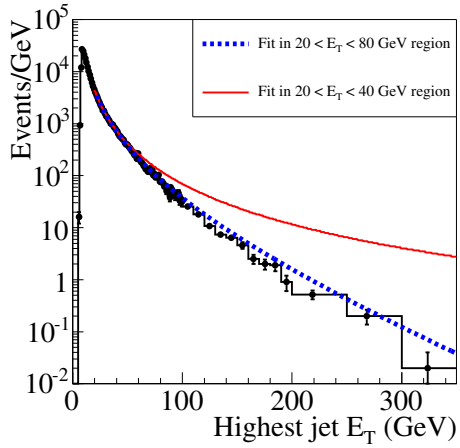

FIG. 4: E_T of the highest E_T jet in standard model $Z \to \mu\mu$ Monte Carlo events. The distribution is fit to Eq. (1) in the region $20 < E_T < 80$ GeV (dotted line), and again in the region $20 < E_T < 40$ GeV (solid line).

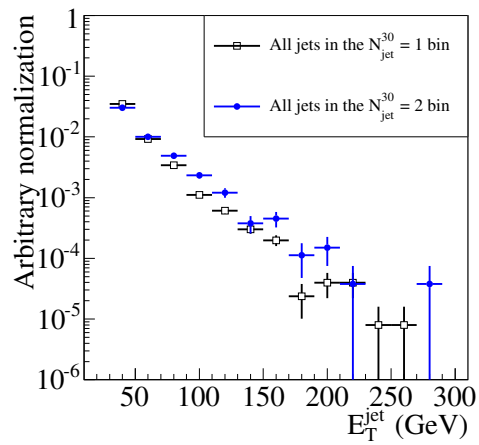

FIG. 5: E_T distribution of jets in $N_{\rm jet}^{30} = 1$ events (open squares) and in $N_{\rm jet}^{30} = 2$ events (solid circles) in $Z \to \ell\ell$ data. Events with higher $N_{\rm jet}^{30}$ have harder E_T spectra.

bution can be predicted for these events. We extrapolate the shape of these jet E_T distributions from the jet E_T distributions of $N_{\rm jet}^{30} \leq 2$ events. In order to do such an extrapolation, we must understand the variation of the jet E_T distribution as a function of $N_{\rm jet}^{30}$.

The E_T distributions of all jets in events with $N_{\rm jet}^{30} = 1$ and 2, normalized to have equal area, is shown in Fig. 5 using $Z \to \ell\ell$ data. The general shape is similar, though jets in $N_{\rm jet}^{30} = 2$ events have a slightly harder tail at high E_T. We model this by fitting to each jet E_T distribution (using Eq. (1)) and extrapolating the fit parameters to $N_{\rm jet}^{30} \geq 3$ events. To avoid simultaneously extrapolating two fit parameters we only extrapolate the exponential parameter (p_1), as this parameter governs the high E_T behavior in our parameterization. In order to extrapolate only this parameter, we fit the $N_{\rm jet}^{30} = 1$ E_T spectrum allowing both parameters to float freely, then fix the power law parameter (p_2) in the fit to the $N_{\rm jet}^{30} = 2$ E_T spectrum. We then extrapolate the p_1 parameter of Eq. (1) linearly as a function of $N_{\rm jet}^{30}$, from their fitted values at $N_{\rm jet}^{30} = 1$ and $N_{\rm jet}^{30} = 2$ into the region $N_{\rm jet}^{30} \geq 3$.

Figures 6 and 7 show the fits of the spectra for events with 1 and 2 jets. Figure 8 shows the linear extrapolation of the exponential parameters. For illustration, the exponential parameter obtained from a fit to the E_T distribution in $N_{\rm jet}^{30} = 3$ events (again fixing the power law parameter to that found in the $N_{\rm jet}^{30} = 1$ events) is shown on the same figure. The extrapolation reasonably predicts the parameter for events with $N_{\rm jet}^{30} = 3$ [27].

This dependence of the jet E_T spectra on $N_{\rm jet}^{30}$ is modeled as described by our parameter extrapolation, allowing us to predict the shapes of the jet E_T spectra for events with $N_{\rm jet}^{30} \geq 3$. The J_T^{30} distribution is now almost completely determined. Only an estimate for the

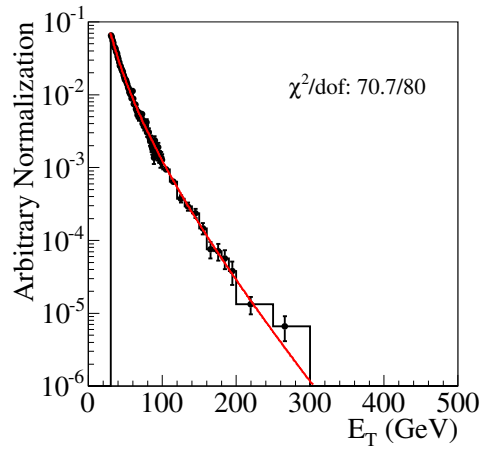

FIG. 6: E_T distribution of jets in $N_{\rm jet}^{30} = 1$ events in standard model $Z \to \mu\mu$ Monte Carlo events. The distribution is fit to Eq. (1) in the range $E_T > 30$.

relative fractions of events with 3, 4, 5, ... jets is needed. For this, we use an exponential fit parameterization, fit to the $N_{\rm jet}^{30}$ distribution in the region $N_{\rm jet}^{30} \leq 2$, and use this shape in the $N_{\rm jet}^{30} \geq 3$ region. This fit is shown in Fig. 9. There is no theoretical motivation for an exponential shape; we merely use it as an estimate, and verify that the J_T^{30} prediction does not strongly depend on the chosen parameterization. As the total number of events with $N_{\rm jet}^{30} \geq 3$ is already constrained using the method from Sec. V A, the dependence of the J_T^{30} distribution on the exponential parameterization of the $N_{\rm jet}^{30}$ distribution is small.

Finally, given the above shapes, it is straightforward to

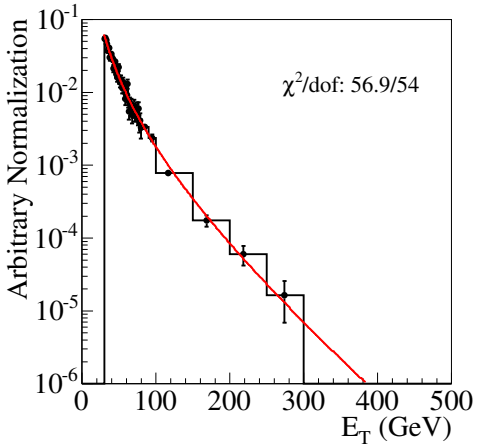

FIG. 7: E_T distribution of jets in $N_{\rm jet}^{30} = 2$ events in standard model $Z \to \mu\mu$ Monte Carlo events. The distribution is fit to Eq. (1) in the range $E_T > 30$, with the parameter p_2 fixed to that obtained from Fig. 6.

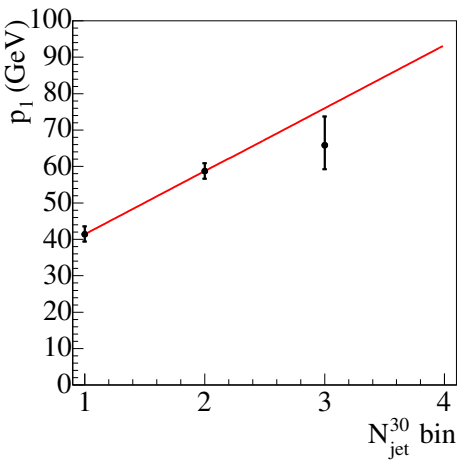

FIG. 8: The extrapolation of the exponential parameter p_1 vs. $N_{\rm jet}^{30}$ in standard model $Z \to \mu\mu$ Monte Carlo events.

make a simple Monte Carlo program that samples these shapes to get the J_T^{30} distribution. The steps required to make this J_T^{30} prediction are:

1. For each event, generate the number of jets by randomly sampling the predicted $N_{\rm jet}^{30}$ distribution in the range $\{3, 4, 5, ...\}$.

2. Take the appropriate jet E_T distribution for this number of jets after extrapolating the exponential fit parameter. Independently sample this jet E_T distribution for each jet.

3. Sum these jets to obtain the J_T^{30}.

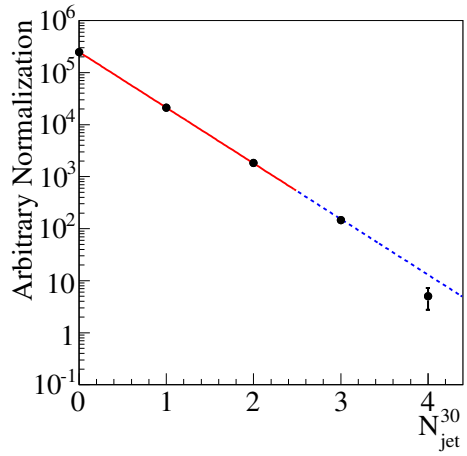

FIG. 9: $N_{\rm jet}^{30}$ distribution in standard model $Z \to \mu\mu$ Monte Carlo events, fit to an exponential in the range $N_{\rm jet}^{30} \leq 2$. This shape is used to estimate the relative fractions of events with 3, 4, 5, ... jets.

The process is repeated as necessary until the J_T^{30} shape is obtained to the desired level of statistical precision.

On step 2, the jet E_T shapes are independently sampled; however, there is potentially some correlation between the individual jet energies. Including this correlation in the J_T^{30} shape prediction would have the effect of making the tail at large values of J_T^{30} slightly harder. In the validation studies in Sec. V D we verify that the correlation is below the level necessary to affect the fit prediction. To understand this further, in Fig. 10, we plot the E_T of one the jets versus the other in events with $N_{\rm jet}^{30} = 2$ in the $Z \to \ell\ell$ data. There is no correlation evident in the plot; in the 663 events with $N_{\rm jet}^{30} = 2$, only a small correlation of 25% is found, indicating that independently sampling the E_T distribution is a reasonable approximation.

C. Uncertainties on Fit Prediction

There are two sources of uncertainty on the mean background prediction: the statistical uncertainty from the finite amount of data in the fits, and the systematic uncertainty from imperfect modeling of the various shapes in the fits.

1. *Statistical Uncertainty on Fit Prediction*

The third highest E_T jet normalization fit predicts the total number of events with $N_{\rm jet}^{30} \geq 3$, using the parameter values at the minimum $-\log L$, where L is the likelihood (or equivalently, the maximum likelihood). The 1σ uncertainty on the number of events is simply obtained

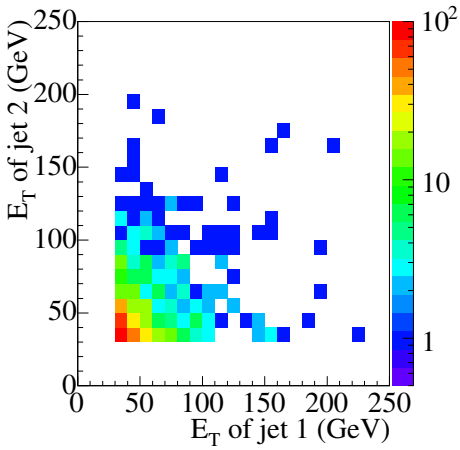

FIG. 10: The E_T of a random jet vs. the E_T of the other, using jets with $N_{\rm jet}^{30} = 2$ in $Z \to \ell\ell$ data.

from its values at the minimum $-\log L + \frac{1}{2}$. Since the total number of events with $N_{\rm jet}^{30} \geq 3$ is given by a single fit, its uncertainty is easily determined with this method.

The J_T^{30} prediction is obtained by extrapolating the behavior of multiple distributions, and to estimate its shape uncertainty we vary each fit parameter independently within its uncertainty (output by the fit) and re-do the extrapolation procedure. The individual uncertainties are combined in quadrature to obtain the total uncertainty. The normalization error is then added in quadrature as well to obtain the uncertainty on the fully-normalized J_T^{30} distribution.

2. *Systematic Uncertainty on Fit Prediction*

As the background from Z+jet events is determined from a fit to the data, the only source of systematic uncertainties is mis-parameterization of those data. If the data were poorly parameterized, fitting a subset of the data would give a large change in the background prediction. We therefore estimate the size of the mis-parameterization uncertainties by changing the range of each fit and re-doing the fit procedure to obtain the J_T^{30} normalization and shape prediction. Both uncertainties, that on the total number of events with $N_{\rm jet}^{30} \geq 3$ (from the third highest E_T jet fit), and that on the J_T^{30} shape, are estimated in this way. The variations from each fit range change are then added in quadrature to obtain the full uncertainty. The fit range changes are summarized in Table II. The "$\pm 1\sigma$" range changes are chosen to give sufficient coverage when observed in control samples of data.

Finally, using the technique and the uncertainties developed above in the Monte Carlo simulation, we can demonstrate that the method is self-consistent by checking the normalized J_T^{30} prediction for events with $N_{\rm jet}^{30} \geq 3$ matches that observed in Monte Carlo events. This comparison is shown in Fig. 11. The observed distribution agrees well with the prediction.

Distribution	nominal range	"-1σ" range	"$+1\sigma$" range
Third highest E_T jet	$(15, 30)$ GeV	$(15, 26)$ GeV	$(17, 30)$ GeV
$N_{\rm jet}^{30} = 1$ jet E_T	$(30, \infty)$ GeV	$(30, 150)$ GeV	$(70, \infty)$ GeV
$N_{\rm jet}^{30} = 2$ jet E_T	$(30, \infty)$ GeV	$(30, 80)$ GeV	$(50, \infty)$ GeV
$N_{\rm jet}^{30}$ shape	$[0, 2]$ jets	$[0, 1]$ jets	$[1, 2]$ jets

TABLE II: Nominal fit ranges and the fit range changes used to estimate systematic uncertainties. The nominal fit range of each distribution is shown in the second column. The third and fourth columns show the ranges used to estimate the uncertainty from a mis-parameterization of that distribution.

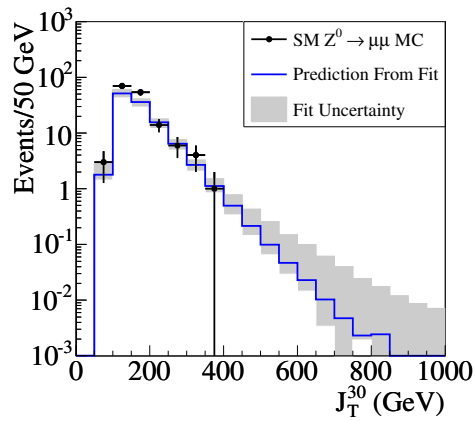

FIG. 11: The prediction for the J_T^{30} distribution (blue line) of standard model Z Monte Carlo and its uncertainty (gray band), compared to the actual distribution (black points with errors).

D. Validation of Technique

Having demonstrated and described the procedure for obtaining the Z+jet background using Monte Carlo simulation, its validation, done predominantly in data, is now described. The Z+jet data cannot be used as a validation sample because of potential signal bias, so we must test on other data samples. We use two sets of multi-jet data as background-only validation samples, and W+jet data as a background sample containing a real heavy quark signal from $t\bar{t}$ production. Finally, we do signal-injection studies with Monte Carlo simulations to understand the effect of signal bias on the fit procedure.

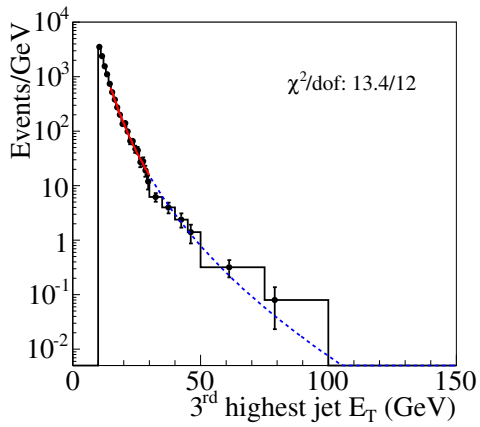

FIG. 12: E_T distribution of the third highest E_T jet in "X"+jet events selected with the jet triggers as described in the text. The distribution is fit to Eq. (1) in the $15 < E_T < 30$ GeV region and extrapolated to the $E_T > 30$ GeV region.

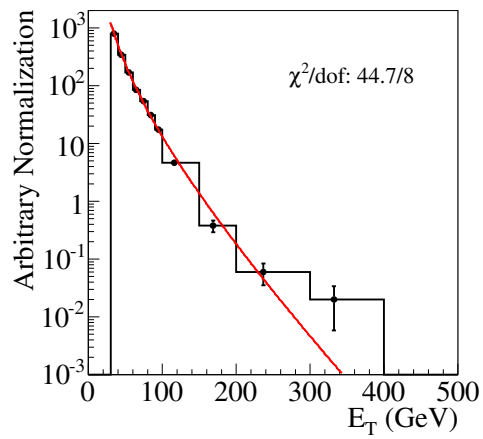

FIG. 13: E_T distribution of jets in $N_{\rm jet}^{30} = 1$ "X"+jet events, selected with the jet triggers as described in the text. The distribution is fit to Eq. (1) in the $E_T > 30$ GeV region.

1. *Multi-Jet Data*

The Z+jet background extrapolation only requires information about the jet E_T distributions, and not the Z. It should therefore perform similarly well not only for Z+jet events, but "X"+jet events, provided that the "X" has a similar transverse momentum spectrum to the Z. For example, if the "X" has a minimum p_T threshold, the E_T distributions of the jets will be sculpted such that they no longer follow the power law $\times$ exponential parameterization of Eq. (1).

We first obtain "X"+jet events from multi-jet data dominated by QCD interactions using prescaled jet triggers that require at least one jet with $E_T > 20$ GeV [28]. An "X" is then constructed by picking two random jets in the event, requiring they both have $E_T > 20$ GeV (to match the electron and muon p_T cuts), and requiring $M_X > 70$ GeV/c^2 to remove the invariant mass turn-on. The invariant mass is not further restricted to the region $81 < M_X < 101$ GeV/c^2 to maximize statistics; in any case the J_T^{30} distribution is observed to not depend on M_X in this sample.

Given this "X" selection, the remaining jets in the event are used to validate the procedure. Figure 12 shows the third highest E_T jet distribution. We extrapolate this distribution above 30 GeV using Eq. (1). A prediction of 97 ± 27 (statistical uncertainty only) events with $N_{\rm jet}^{30} \geq 3$ is obtained. 80 events are observed. This is consistent within the uncertainties. To quantitatively evaluate the level of consistency we calculate the probability to measure the observed number of events or higher given the background prediction, as well as convert this probability to units of standard deviations [29]. This calculation gives a corresponding probability of 0.73; this is a 0.6σ level of consistency.

We now predict the J_T^{30} shape. Figures 13 and 14 show the fits to the jet E_T spectra for events with $N_{\rm jet}^{30} = 1$ and 2. We extrapolate the parameter p_1 using the plot in Fig. 15 to events with $N_{\rm jet}^{30} \geq 3$. The $N_{\rm jet}^{30}$ shape is taken from the fit in Fig. 16. Using these ingredients, the simple Monte Carlo program is used to obtain the J_T^{30} shape, which is normalized to the prediction of 97 events with $N_{\rm jet}^{30} \geq 3$. The prediction and total uncertainty is shown overlaid with the actual distribution in "X"+jet data in Fig. 17. The distribution clearly agrees well within the uncertainty envelope.

Because the J_T^{30} uncertainties in each bin are correlated, an independent data/background comparison in each bin is not straightforward. Rather, we test the shape agreement once using the (arbitrarily chosen) region of $J_T^{30} > 200$ GeV. Above 200 GeV, $19.7^{+9.2}_{-9.0}$ events are expected and 20 events are observed.

The background extrapolation method can accurately predict the normalization and shape of the J_T^{30} distribution in the jet triggered sample. However, because of the prescale, this sample has relatively low statistics despite the large cross section of QCD multi-jet processes. To obtain a higher statistics sample of multi-jet data, we can use the electron triggers, which are not prescaled. In this sample we construct an "X" by pairing the triggered electron with a "fake" electron, which is an EM calorimeter cluster that is reconstructed as an electron but fails the low hadronic energy requirement. "X" events selected in this way are dominated by QCD dijet events. Again, $M_X > 70$ GeV/c^2 is required to remove the invariant mass turn-on. Additionally the invariant mass region $81 < M_X < 101$ GeV/c^2 is vetoed to remove real $Z \to ee$ events. Figure 18 shows the plot of the invariant mass before these requirements.

Given this "X" selection, the remaining jets in the event are used to validate the procedure. Figure 19 shows

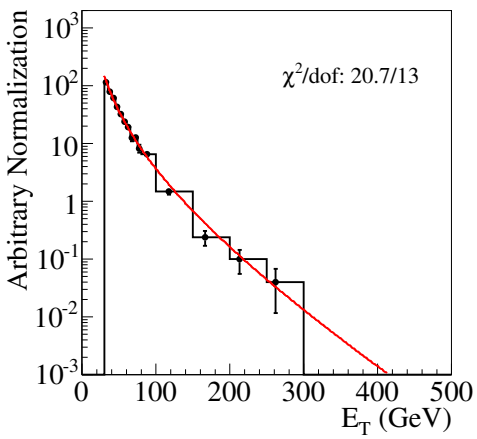

FIG. 14: E_T distribution of jets in $N_{\rm jet}^{30} = 2$ "X"+jet events selected with the jet triggers as described in the text. The distribution is fit to Eq. (1) in the $E_T > 30$ GeV region with the parameter p_2 fixed to that obtained from the fit in Fig. 13.

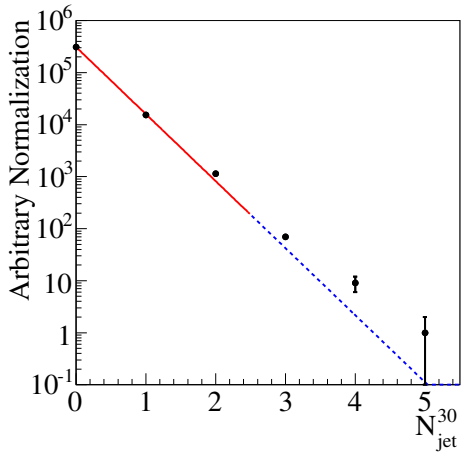

FIG. 16: $N_{\rm jet}^{30}$ distribution in "X"+jet events selected with the jet triggers as described in the text. The distribution is fit to an exponential in the range $N_{\rm jet}^{30} \leq 2$.

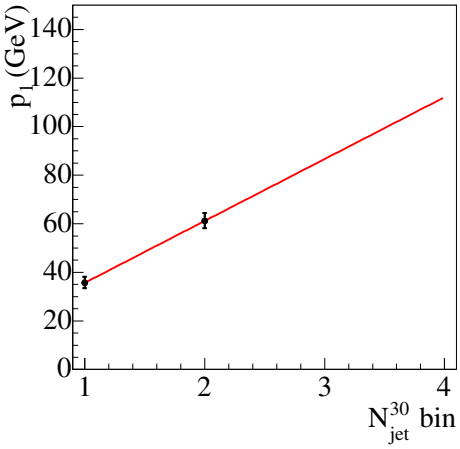

FIG. 15: The extrapolation of the exponential parameter p_1 vs. $N_{\rm jet}^{30}$ in "X"+jet events selected with the jet triggers as described in the text.

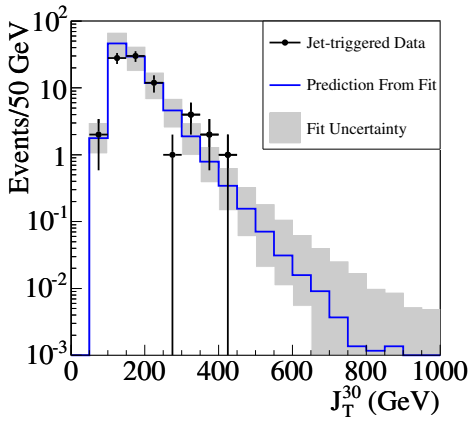

FIG. 17: The prediction (blue line) and uncertainty (gray band) for the J_T^{30} distribution of "X"+jet events selected with the jet triggers as described in the text. The prediction is compared to the actual distribution (black points with errors). The observation agrees with the prediction.

the third highest E_T jet distribution. We extrapolate this distribution above 30 GeV using Eq. (1). A prediction of 4427^{+354}_{-310} (statistical uncertainty only) events with $N_{\rm jet}^{30} \geq 3$ is obtained. 4509 events are observed. Approximating the Poisson distribution of the number of observed events as a Gaussian, this is a 0.23σ level of consistency.

The J_T^{30} shape is predicted using the previously described procedure of extrapolating the jet E_T distributions from events with $N_{\rm jet}^{30} = 1$ and 2 to $N_{\rm jet}^{30} \geq 3$. The normalized prediction and its uncertainty are compared to the actual distribution in the data in Fig. 20. The distribution agrees well within the uncertainty envelope. Above 200 GeV, 1412^{+477}_{-212} events are expected; 1128 events are observed, for a -1.3σ level of consistency. The background prediction is compared to the number of observed events as a function of the J_T^{30} cut in Table III. The prediction agrees well over the entire J_T^{30} distribution.

We have seen that the background extrapolation performs well enough in this high-statistics validation sample. Because of the high-statistics, this sample can be divided into subsamples and test the prediction method many times over. The electron-triggered multi-jet data is divided into 50 subsamples to check the background estimation with a sample size similar to that expected in the Z+jet data.

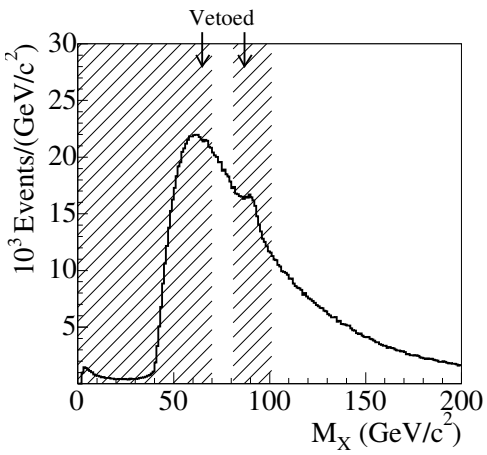

FIG. 18: Distribution of M_X in "X"+jet events selected from the electron triggers as described in the text. The shaded regions are removed; that is, events with $M_X > 70$ GeV/c^2 are selected, and the $81 < M_X < 101$ GeV/c^2 region is vetoed.

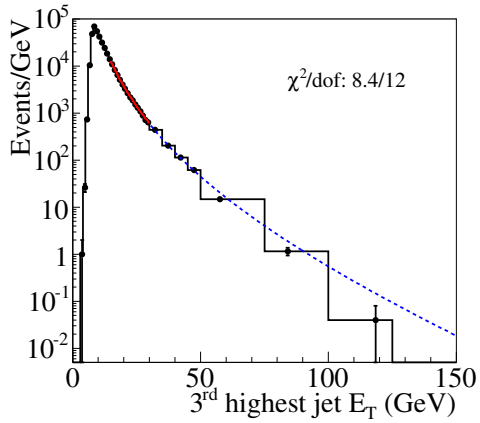

FIG. 19: E_T distribution of the third highest E_T jet in "X"+jet events selected with the electron triggers as described in the text. The distribution is fit to Eq. (1) in the $15 < E_T < 30$ GeV region and extrapolated to the $E_T > 30$ GeV region.

To validate the third highest E_T jet extrapolation, we evaluate the consistency between the fit prediction and the observation in each subsample. The pull distribution from these calculations is observed to be consistent with a Gaussian with mean 0 and width of 1, indicating that the mean prediction and the uncertainties are correctly calculated for the $N_{\rm jet}^{30} \geq 3$ prediction. On average, the background prediction is $3 \pm 5\%$ low relative to the data. That is, the background prediction underestimates the background, but by an amount consistent with zero. This is consistent with the fit done in standard model Z Monte Carlo simulation in Sec. V A, in which the background prediction was $31 \pm 16\%$ low relative to the data.

Minimum J_T^{30} cut	Total Bkg. (events)	Data (events)
50	4430^{+1270}_{-600}	4509
100	4380^{+1250}_{-590}	4463
150	2810^{+830}_{-360}	2602
200	1410^{+480}_{-210}	1128
250	667^{+281}_{-133}	436
300	$312^{+172}_{-81.8}$	170
350	$146^{+106}_{-47.4}$	62
400	$68.7^{+64.8}_{-26.2}$	27
450	$32.8^{+38.9}_{-14.3}$	15
500	$16.2^{+23.3}_{-8.4}$	6
550	$7.9^{+14.5}_{-4.5}$	3
600	$3.9^{+8.8}_{-2.5}$	0

TABLE III: The "X"+jet data (selected with the electron triggers as described in the text) vs. J_T^{30}, compared with the background prediction.

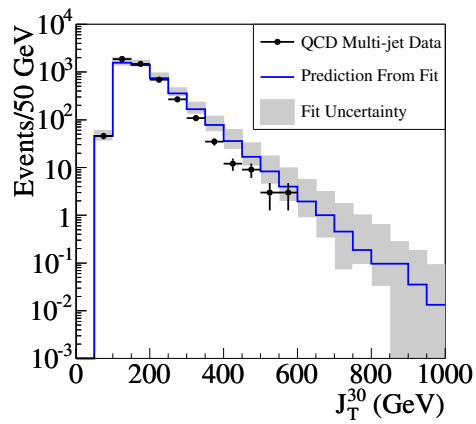

FIG. 20: The prediction (blue line) and uncertainty (gray band) for the J_T^{30} distribution of "X"+jet events selected with the electron triggers as described in the text. The prediction is compared to the actual distribution (black points with errors). The observation agrees with the prediction, with a maximum fluctuation downward of 1.9σ. The data are below the prediction for several point because the shape uncertainty is correlated between bins.

To validate the J_T^{30} shape prediction, in each subsample we evaluate the consistency between the fit prediction and the observation using a cut of $J_T^{30} > 200$ GeV. In this case, the resulting pull distribution was inconsistent with a Gaussian with mean 0 and width 1. We find that the background prediction overestimates the number of observed events, and that the uncertainty is overly conservative, after correcting for this bias. On average, the background prediction is $23 \pm 7\%$ high relative to the data. However, we find that this bias is covered by the uncertainties, with an average uncertainty on the background prediction of 47%. To clarify, these biases are only present in the J_T^{30} shape prediction, and not in the

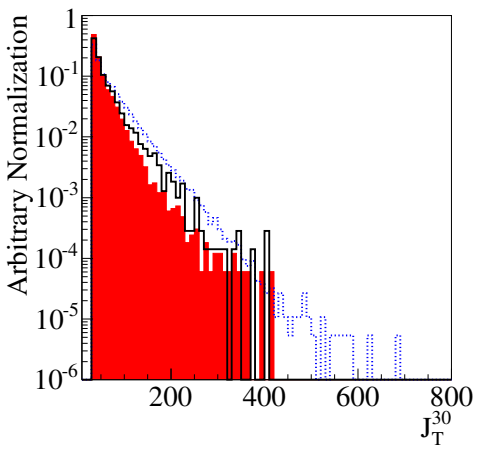

FIG. 21: The J_T^{30} distribution without the $N_{\rm jet}^{30} \geq 3$ requirement in the Z+jet data (black line), compared to "X"+jet data selected with the jet triggers (red histogram) and to "X"+jet data selected with the electron triggers (dotted blue line).

$N_{\rm jet}^{30} \geq 3$ prediction.

To compare the jet kinematics in each of the validation samples (both the "X" events selected from jet triggers and the "X" events selected from the electron triggers) to the Z+jet data, the J_T^{30} distribution of each is plotted, without the $N_{\rm jet}^{30} \geq 3$ requirement, in Fig. 21. The overall shape of each is the same, although they are slightly different—for example, electron-triggered "X"+jet data have a harder spectrum. However, the background estimation takes these differences into account in the fit procedure.

These validations show that the fit prediction method correctly calculates the background when there is no signal present. To verify that it calculates the background correctly in the presence of signal, we use W+jet data.

2. *W+jet Data*

The tree-level single W diagrams and the physics that gives rise to additional jets is similar to Z+jet production, and so similar behavior in the W+jet data is expected. However, in the W+jet data, in addition to the single-W production there is also a heavy quark signal from the top quark, producing W bosons via $t\bar{t} \rightarrow WWb\bar{b}$. This sample provides a useful and interesting validation of the method—it is a real data sample that can test whether or not the background fit procedure performs properly in the presence of a signal similar to that of the search.

W events in the $W \rightarrow \mu\nu$ channel are selected by requiring exactly one "tight" muon and missing transverse energy ($\not\!\!E_T$). The $\not\!\!E_T$ is measured using the vector sum of the calorimeter tower transverse energies and the muon p_T. $\not\!\!E_T > 25$ GeV is required. Since only a single muon is required, this is the so-called "lepton+jets" channel of the top quark selected with only kinematic information, and without tagging b-jets [31].

Using this W+jet selection, we test the extraction of the top signal for events with $N_{\rm jet}^{30} \geq 3$ using only data as a validation of the method for predicting the Z+jet background. We expect standard model W+jet to be the dominant background for $t\bar{t}$ after the $N_{\rm jet}^{30}$ requirement. In single W+jet Monte Carlo simulation with no $t\bar{t}$ component, the method does predict the actual Monte Carlo distribution well. We then apply the same method to the W+jet data, fitting the third highest E_T jet distribution to Eq. (1) in Fig. 22. In this case, the extrapolation does not describe the data well.

The extrapolation predicts 439^{+20}_{-20} (stat.) $^{+30}_{-24}$ (syst.) events; 762 events are observed. We make the hypothesis that this excess is due to the top quark, and test this by checking that the cross section is consistent with that expected for $t\bar{t}$. The excess of the data above the background gives the number of $t\bar{t}$ candidates, 323^{+34}_{-34} (stat.) $^{+30}_{-24}$ (syst.). Using $t\bar{t}$ Monte Carlo events gives an estimate for the product of acceptance and efficiency of $3.41 \pm 0.02\%$. The luminosity of the muon-triggered sample is 1.04 fb^{-1}. A cross section of 9 ± 1 pb (stat. uncert. only) [30] is therefore obtained. The proximity to the previous measured cross section in this channel at CDF using 194 pb^{-1}, 6.6 ± 1.1 (stat.) $\pm$ 1.5 (syst.) pb [31], indicates that the excess is consistent with the background+$t\bar{t}$ hypothesis, and that the fit procedure is accurately predicting the background from single W+jet production in the presence of signal.

A prediction is now made for the J_T^{30} shape of the W+jet background. Figures 23 and 24 show the fits to the jet E_T spectra for events with $N_{\rm jet}^{30} = 1$ and 2; Fig. 25 shows the parameter p_1 extrapolation; Fig. 26 shows the $N_{\rm jet}^{30}$ shape fit. We use these shapes to obtain the J_T^{30} shape and errors, add the expected contribution from $t\bar{t}$ using Monte Carlo simulation (normalized to the "measured" cross section of 9 pb), and compare this to the actual distribution in data in Fig. 27. The observed data are well described by the total J_T^{30} prediction, verifying that the fit procedure can predict the J_T^{30} shape of the background in the presence of signal.

While the predicted shape of the J_T^{30} distribution agrees with the data well (after adding the expected contribution from $t\bar{t}$), the total uncertainty on the background prediction becomes extremely large at high J_T^{30}. The J_T^{30} distribution for $t\bar{t}$ peaks near 200 GeV, where the uncertainty is small, but it is instructive to understand the reason for the increased uncertainty at very large J_T^{30}. This large error is completely dominated by a poor parameterization of the E_T distribution of jets in $N_{\rm jet}^{30} = 2$ events. Since, in Fig. 24, the fitted parameterization poorly describes the data, changing the range from nominal (our method for determining the size of the mis-parameterization uncertainty) will make a large difference in the fit. However, this is not a problem with

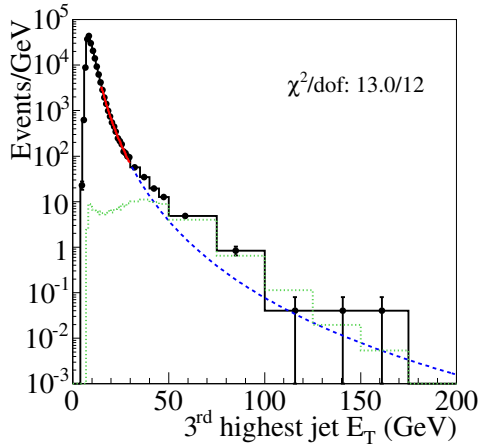

FIG. 22: E_T distribution of the third highest E_T jet in W+jet events (black line and points). The distribution is fit to Eq. (1) in the $15 < E_T < 30$ GeV region and extrapolated to the $E_T > 30$ GeV region. The dotted green line shows the contribution from $t\bar{t}$ at the "measured" cross section of 9 pb. There is very little contribution from $t\bar{t}$ within the fit region. The extrapolated distribution is inconsistent with the background-only hypothesis, but consistent with the background plus $t\bar{t}$ hypothesis.

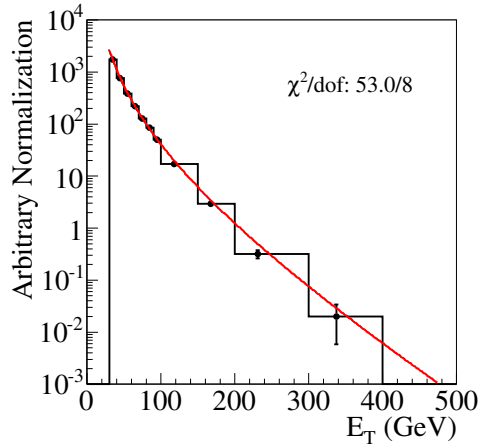

FIG. 23: E_T distribution of jets in $N_{\rm jet}^{30} = 1$ W+jet events. The distribution is fit to Eq. (1) in the $E_T > 30$ GeV region.

the parameterization in Eq. (1), because if the same spectrum is fit without fixing the power law parameter to the value observed in events with $N_{\rm jet}^{30} = 1$, the quite reasonable fit, shown in Fig. 28, is obtained. That is, the parameterization still describes the $N_{\rm jet}^{30} = 2$ E_T spectrum well, but our method of fixing the power law parameter in this fit to that observed from the $N_{\rm jet}^{30} = 1$ E_T spectrum does not describe the behavior of the changing jet E_T distributions as a function of $N_{\rm jet}^{30}$ well in this sample. In the other

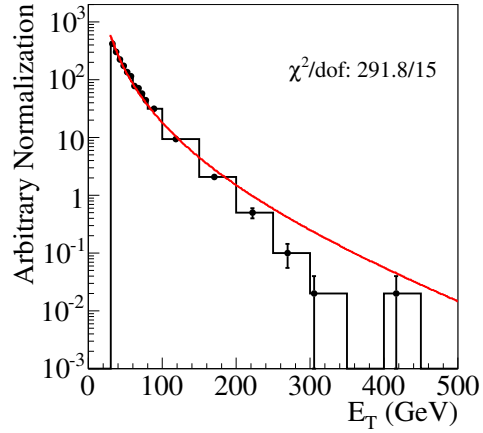

FIG. 24: E_T distribution of jets in $N_{\rm jet}^{30} = 2$ W+jet events. The distribution is fit to Eq. (1) in the $E_T > 30$ GeV region with the parameter p_2 fixed to that obtained from the fit in Fig. 23.

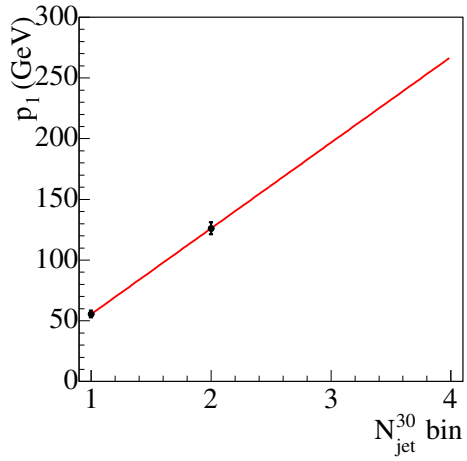

FIG. 25: The extrapolation of the exponential parameter p_1 vs. $N_{\rm jet}^{30}$ in W+jet events.

validation samples in data and Monte Carlo simulations, and particularly in the fits of the Z+jet data, we find no such large systematic effect from a mis-parameterization in the $N_{\rm jet}^{30} = 2$ E_T distribution. This issue therefore does not affect this analysis, but it suggests the background prediction procedure could be enhanced with a more sophisticated parameter extrapolation, perhaps by extrapolating both parameters p_1 and p_2 simultaneously.

3. *Signal Injection Studies*

The studies in data indicate the fit method adequately predicts the background, without and with the presence

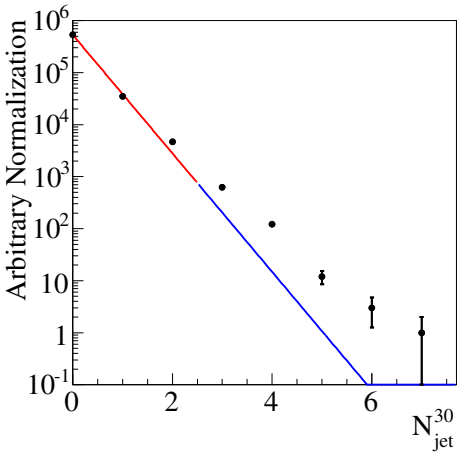

FIG. 26: $N^{30}_{\rm jet}$ distribution in W+jet events. The distribution is fit to an exponential in the range $N^{30}_{\rm jet} \leq 2$.

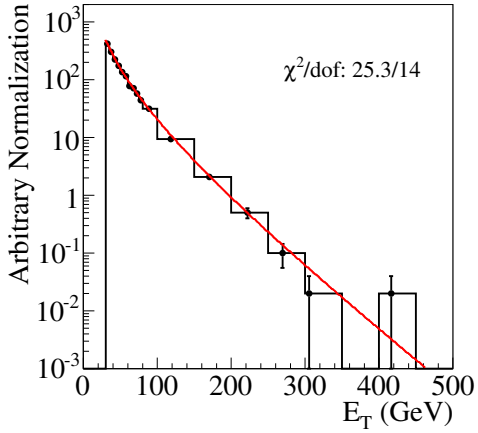

FIG. 28: E_T distribution of jets in $N^{30}_{\rm jet} = 2$ W+jet events. The distribution is fit to Eq. (1) in the $E_T > 30$ GeV region without fixing the parameter p_2.

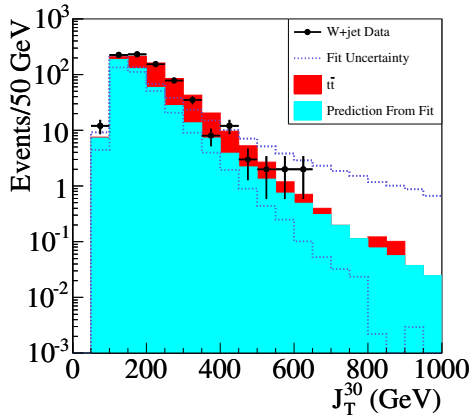

FIG. 27: The prediction (cyan histogram) and uncertainty (dotted lines) for the J^{30}_T distribution of W+jet events. The expectation from $t\bar{t}$ is added to the prediction. The data (points with errors) agree with the background plus $t\bar{t}$ hypothesis.

of signal. We would also like to understand at what point, if any, signal contamination causes an unacceptably large change to the background prediction. That is, we need to verify that the background extrapolation does not "fit-away" the signal, as the jet E_T distributions may be substantially changed if there is a large amount of signal in the fitted regions.

To study this effect we use standard model Z Monte Carlo events with $b' \to bZ$ Monte Carlo events added at a variety of signal masses. An equivalent luminosity of 1 fb^{-1} of Monte Carlo events is used to understand the effect with the approximate amount of statistics that is present in the data. For this study $BR(b' \to bZ) = 100\%$ is assumed; reducing this branching ratio will only reduce the effect of a signal bias.

For example, the predicted J^{30}_T distributions, generated with and without $m_{b'} = 200$ GeV/c^2 Monte Carlo signal events added to the Z+jet background fit, are shown in Fig. 29. The difference between the background predictions with and without signal is small compared to the actual number of Monte Carlo events, indicating that signal does not bias the fit to a large degree at this mass point.

As expected, as the b' mass increases the fit becomes less biased from the presence of signal; as the b' mass decreases, the fit becomes more biased. At a b' mass of 150 GeV/c^2, we found an increase in signal bias, but sensitivity to this mass point is still retained (at a significance of 4.8σ). At a b' mass of 100 GeV/c^2, however, we found that the signal was completely fit away. We therefore do not set limits below 150 GeV/c^2. We note that this search is still sensitive to models with masses near 100 GeV/c^2, as long as the cross sections are sufficiently small as to not bias the fit. In general, though, lower masses produce more signal contamination than higher masses, as both the cross sections are larger and the E_T distributions have larger fractions within the fit regions. Sensitivity to these lower masses could be increased by lowering E_T thresholds and $N_{\rm jet}$ cuts, and applying similar fit procedures with the altered selection.

E. Application of Technique to the Signal Sample

We now apply the fit technique to the combined $Z \to ee$ and $Z \to \mu\mu$ data to predict the background from Z+jet final states. The third highest E_T jet distribution is shown in Fig. 30, with events that have $N^{30}_{\rm jet} \geq 3$ removed. We fit in the region $15 < E_T < 30$ GeV, and extrapolate to the region $E_T > 30$ GeV. We pre-

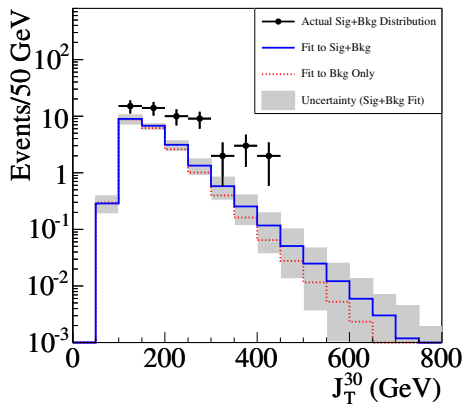

FIG. 29: Prediction for the J_T^{30} distribution in standard model $Z \to \mu\mu$ events, with and without the presence of a 200 GeV/c^2 b' signal introduced. The difference between the two predictions is small compared to the excess of signal at large J_T^{30}.

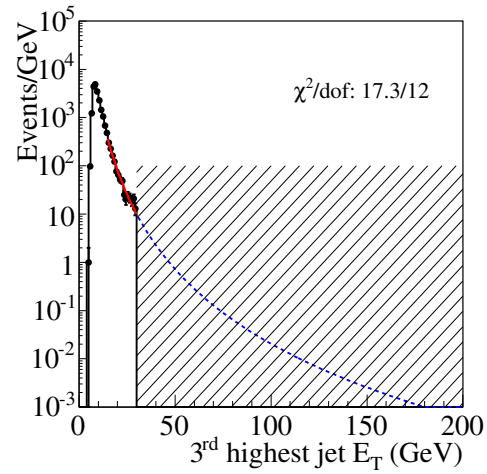

FIG. 30: E_T distribution of the third highest E_T jet in $Z \to ee$ and $Z \to \mu\mu$ events with $N_{\rm jet}^{30} \leq 2$. The distribution is fit to Eq. (1) in the $15 < E_T < 30$ GeV region and extrapolated to the $E_T > 30$ GeV region. Events with $N_{\rm jet}^{30} \geq 3$ (equivalent to $E_T > 30$ GeV, the hatched region) are removed from the distribution.

dict $72.2^{+9.8}_{-11.1}$ events with $N_{\rm jet}^{30} \geq 3$.

To obtain the J_T^{30} shape of the Z+jet background, we fit the jet E_T distributions of events with $N_{\rm jet}^{30} = 1$ and 2, and linearly extrapolate the fit parameter p_1 to events with $N_{\rm jet}^{30} \geq 3$. The fit to the $N_{\rm jet}^{30} = 1$ jet E_T spectrum is shown in Fig. 31, the fit to the $N_{\rm jet} = 2$ jet E_T spectrum in Fig. 32, and the extrapolation of the fit parameter in Fig. 33. The fit to the $N_{\rm jet}^{30}$ distribution in the 0, 1, and 2 jet bins in Fig. 34 is used as an estimate of the shape of the $N_{\rm jet}^{30}$ distribution in the 3 and higher jet bins. With these ingredients, the simple Monte Carlo program is used to obtain the expected J_T^{30} shape, which is then normalized to the prediction for the total number of $N_{\rm jet}^{30} \geq 3$ background events, $72.2^{+9.8}_{-11.1}$. The J_T^{30} distribution prediction and its total statistical+systematic uncertainty is shown in Fig. 35.

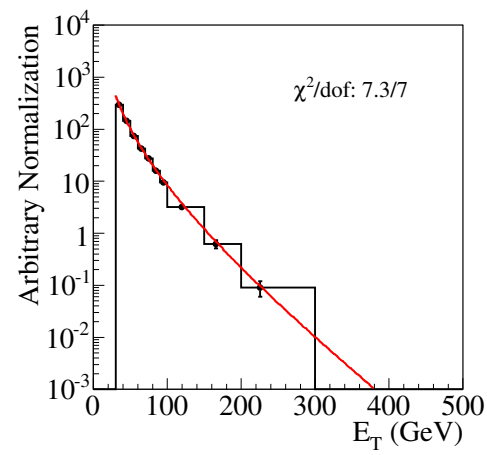

FIG. 31: E_T distribution of jets in $N_{\rm jet}^{30} = 1$ $Z \to ee$ and $Z \to \mu\mu$ events. The distribution is fit to Eq. (1) in the $E_T > 30$ GeV region.

VI. REMAINING BACKGROUNDS

After having estimated the contribution from Z+jet with the above technique, the remaining backgrounds listed in Sec. IV are now estimated.

The second background, multi-jet fakes, has approximately the same shape as the Z+jet background, and is therefore included in the fit procedure. This shape similarity is demonstrated when validating the procedure using multi-jet data in Sec. V D 1 above. Since this background is already included in the Z+jet background estimate, no further determination of it is needed.

Nonetheless, its size is independently measured to confirm that it is small relative to the Z+jet background. To obtain an upper bound on the multi-jet background, the sidebands of the $M_{\ell\ell}$ distribution for events with $N_{\rm jet}^{30} \geq 3$ are used. We attribute all of the events in the sidebands to multi-jet fakes, and interpolate from the sidebands into the $81 < M_{\ell\ell} < 101$ GeV/c^2 region. Using this method, less than 11 ± 2 events from multi-jet fakes are predicted. The small size relative to the Z+jet background, $72.2^{+9.8}_{-11.1}$, indicates that this background is relatively unimportant.

While the third background, from multi-jet events occurring simultaneously with cosmic rays, is also included in the fit procedure as the jet E_T spectra are similar to the Z+jet background, its size is again independently

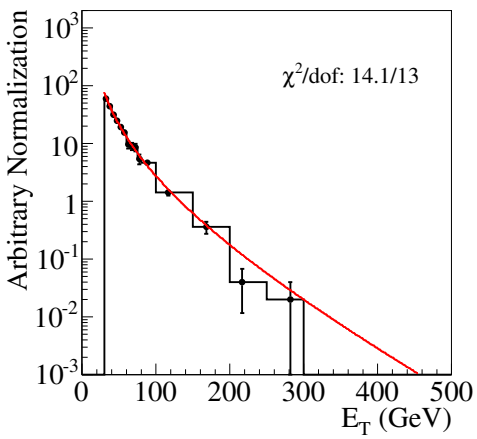

FIG. 32: E_T distribution of jets in $N_{\rm jet}^{30} = 2$ $Z \to ee$ and $Z \to \mu\mu$ events. The distribution is fit to Eq. (1) in the $E_T > 30$ GeV region with the parameter p_2 fixed to that obtained from the fit in Fig. 31.

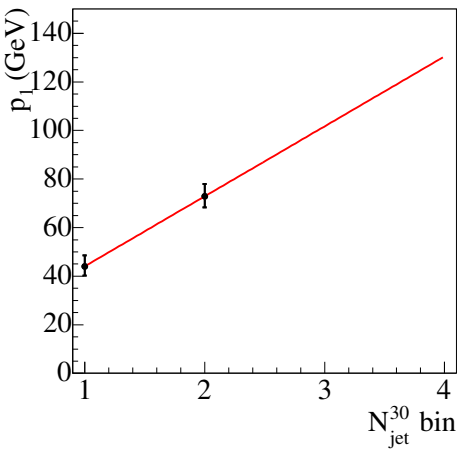

FIG. 33: The extrapolation of the exponential parameter p_1 vs. $N_{\rm jet}^{30}$ in $Z \to ee$ and $Z \to \mu\mu$ events.

measured. This background is rejected using timing information from the COT. That information is also used to estimate this background using the number of events rejected with the timing cut, combined with a measurement of the rejection efficiency in a sample of cosmic rays with high-purity. We find a negligible background.

The remaining backgrounds are not included in the fit procedure since they contain jets from the decays of massive particles and so the jet E_T spectra do not follow the parameterization in Eq. (1). They can be estimated with Monte Carlo simulations normalizing to the expected standard model cross sections. All remaining backgrounds are negligible relative to the Z+jet background, the largest being from WZ, with an estimated

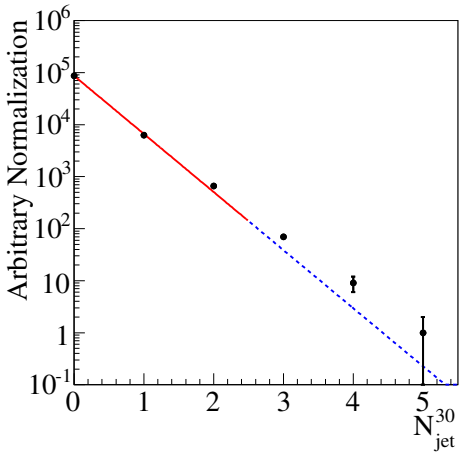

FIG. 34: $N_{\rm jet}^{30}$ distribution in $Z \to ee$ and $Z \to \mu\mu$ events. The distribution is fit to an exponential in the range $N_{\rm jet}^{30} \leq 2$.

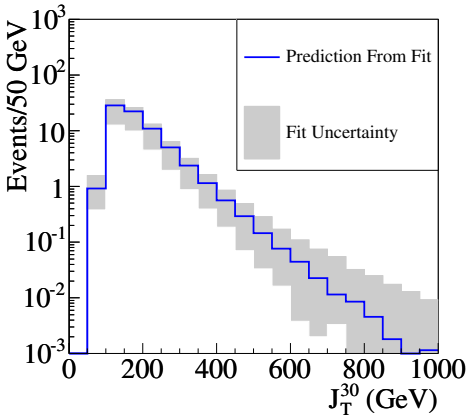

FIG. 35: The prediction (blue line) and uncertainty (gray band) for the J_T^{30} distribution of $Z \to ee$ and $Z \to \mu\mu$ events.

contribution of 1.6 ± 0.1 events. Each of the background contributions to the $N_{\rm jet}^{30} \geq 3$ region is summarized in Table IV. As the backgrounds from WZ, ZZ, and $t\bar{t}$ are negligible compared to the Z+jet background, they are excluded in the background estimation vs. J_T^{30}.

VII. RESULTS

We now compare the background prediction to the observation in the Z+jet data. From the third highest E_T jet extrapolation, $75.3^{+9.8}_{-11.1}$ events with $N_{\rm jet}^{30} \geq 3$ are predicted, and 80 events are observed. In Fig. 36, the extrapolation is shown overlaid with the data. The data agree with the extrapolation well. The predicted J_T^{30} distribution is compared to that observed in data in

Process	Background
Z+jet	$72.2^{+9.8}_{-11.1}$
Multi-jet fakes	$< 11 \pm 2$ (included in Z+jet fit)
Cosmics	negligible
WZ	1.6 ± 0.1
ZZ	0.7 ± 0.1
$t\bar{t}$	0.8 ± 0.1
Total	$75.3^{+9.8}_{-11.1}$

TABLE IV: Summary of all backgrounds after selecting events with $N^{30}_{\rm jet} \geq 3$, independent of J^{30}_T.

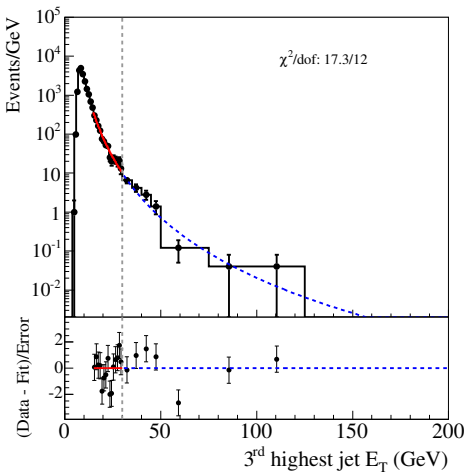

FIG. 36: E_T distribution of the third highest E_T jet in $Z \to ee$ and $Z \to \mu\mu$ events. The fit from Fig. 30 is overlaid. The fit extrapolation matches the distribution above 30 GeV well.

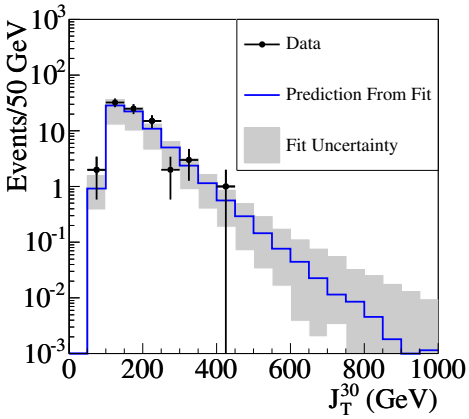

FIG. 37: The J^{30}_T prediction and uncertainty from Fig. 35 compared to the observed distribution (black points and errors) in $Z \to ee$ and $Z \to \mu\mu$ events with $N^{30}_{\rm jet} \geq 3$. The prediction agrees well with the data.

Minimum J^{30}_T cut	Total Bkg. (events)	Data (events)
50	$72.2^{+17.9}_{-41.3}$	80
100	$71.3^{+17.3}_{-40.7}$	78
150	$42.8^{+9.6}_{-24.8}$	46
200	$20.6^{+5.6}_{-12.6}$	21
250	$9.7^{+3.6}_{-6.2}$	6
300	$4.7^{+2.3}_{-3.1}$	4
350	$2.3^{+1.5}_{-1.6}$	1
400	$1.2^{+1.0}_{-0.9}$	1
450	$0.6^{+0.7}_{-0.5}$	0
500	$0.3^{+0.5}_{-0.3}$	0

TABLE V: The data compared to the Z+jet background fit prediction vs. J^{30}_T.

Fig. 37. Again, the data agree with the prediction quite well. The predicted and observed number of events integrated above various J^{30}_T cut values are listed in Table V. We search for an excess above the prediction at each J^{30}_T cut value. Even when ignoring the systematic uncertainties, the maximum difference upward has a significance of $+0.9\sigma$; the maximum difference downward has a significance of -1.4σ.

Given that there is no significant excess present in the data, a cross section limit is set using the fourth generation model. At each b' mass, the counting experiment is evaluated with the requirement $J^{30}_T > m_{b'}c^2$. The limit is set at a 95% confidence level by integrating a likelihood obtained using a Bayesian technique that smears the Poisson-distributed background with Gaussian acceptance and mean background uncertainties [32]. The background and its uncertainty are taken from the fit prediction (listed in Table V); the product of acceptance and efficiency is taken from Monte Carlo simulation, with correction factors applied to match the observed efficiency of leptons in $Z \to \ell\ell$ data. The uncertainty on the product of acceptance and efficiency is 10%, with the dominant source from a jet energy scale uncertainty of 6.7% [20], the second dominant from a luminosity uncertainty of 5.9%, and the remainder from Monte Carlo event statistics and imperfect knowledge of lepton identification efficiencies [16], parton distribution functions [33], and initial and final state radiation.

The 95% confidence level cross section limit as a function of mass is shown in Fig. 38. In models with different acceptances, the acceptances of the fourth generation model (for these values, see Appendix A) simply need to be factored out and the acceptances of those models should be included.

To set a mass limit on the fourth generation model, the b' cross section is calculated at leading order using PYTHIA, with the assumption that $BR(b' \to bZ) = 100\%$. With this assumption, the mass limit observed is $m_{b'} > 268$ GeV/c^2. The previous search on this model in the bZ channel obtained a limit of $m_{b'} > 199$ GeV/c^2 [2], with a selection catered to the specific b' model by tagging b-jets using displaced vertices.

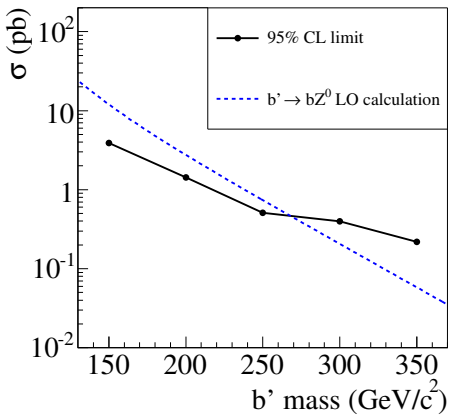

FIG. 38: Cross section limit vs. b' mass, set at a confidence level of 95%. In the acceptance calculation $BR(b' \to bZ) \equiv \beta = 100\%$ was assumed. If $\beta < 100\%$, the acceptance would scale by the factor $1-(1-\beta)^2$, since the b' is produced in pairs and only one of them is required to decay to a Z with our selection. In addition, non-Z decays could change the acceptance of the $N_{\rm jet}^{30} \geq 3$ cut.

VIII. CONCLUSION

We have presented the results of a search for new particles decaying to Z bosons and jets. We developed and validated a new technique to predict the dominant background from the data alone. This technique complements the phenomenological-based method of predicting backgrounds via Monte Carlo calculations of higher-order matrix elements and non-perturbative soft parton showers. The technique presented here has advantages of not requiring careful tuning of phenomenological parameters when comparing to data and not requiring the many resource-consuming iterations of Monte Carlo detector simulations. The speed with which it can be applied makes it an attractive tool for calculation of backgrounds in jet-rich environments at future experiments, including those at the Large Hadron Collider.

In the application of the technique on CDF Z+jet data, no significant excess above background was seen. A cross section limit was therefore set on a fourth generation model as a function of mass. A mass limit of $m_{b'} > 268$ GeV/c^2 using a leading-order b' cross section calculation with the assumption that $BR(b' \to bZ) = 100\%$ was set at a 95% confidence level.

APPENDIX A: ACCEPTANCE OF b' MODEL

In Table VI the acceptance times efficiency to select $b' \to bZ$ events (assuming $BR(b' \to bZ) = 100\%$) after the kinematic cuts is shown. As these acceptances include a factor from $BR(Z \to \ell\ell)$, they are maximally $BR(Z \to ee) + BR(Z \to \mu\mu) = 6.7\%$.

b' mass (GeV)	Acceptance (%)
150	1.05
200	1.44
250	1.61
300	1.66
350	1.77

TABLE VI: Acceptances to select $b' \to bZ$ events versus mass, after applying the $N_{\rm jet}^{30} \geq 3$ and $J_T^{30} > m_{b'}c^2$ requirements. These include a factor from the branching ratio of $Z \to ee$ and $Z \to \mu\mu$. If this factor is removed, the acceptances range from 8–14%. $BR(b' \to bZ) = 100\%$ was assumed.

Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Comisión Interministerial de Ciencia y Tecnología, Spain; the European Community's Human Potential Programme; the Slovak R&D Agency; and the Academy of Finland.

[1] D. Acosta et al. [CDF Collaboration], Phys. Rev. Lett. **94**, 091803 (2005).
[2] T. Affolder et al. [CDF Collaboration], Phys. Rev. Lett. **84**, 835 (2000).
[3] V. M. Abazov et al. [DØ Collaboration], Phys. Rev. D **74**, 011104 (2006).
[4] A. L. Scott [CDF Collaboration], Int. J. Mod. Phys. A **20**, 3263 (2005).
[5] P.H. Frampton, P.Q. Hung, and M. Sher, Phys. Rept. **330**, 263 (2000).

[6] S. Ambrosanio et al., Phys. Rev. D **54**, 5395 (1996).

[7] G. Anderson et al., Phys. Rev. D **61**, 095005 (2000).

[8] J. D. Bjorken, S. Pakvasa, and S. F. Tuan, Phys. Rev. D **66** 053008 (2002).

[9] M.J. Strassler and K.M. Zurek, arXiv:hep-ph/0605193.

[10] M. Krämer, S. Mrenna, and D.E. Soper, Phys. Rev.D **73** 014022 (2006).

[11] M. D'Onofrio [CDF and D0 Collaborations], FERMILAB-CONF-06-224-E

[12] D. Acosta, et al. [CDF Collaboration], Phys. Rev. D **71**, 032001 (2005);
C.S. Hill, Nucl. Instrum. Methods A **530**, 1 (2004);
A. Sill, et al., Nucl. Instrum. Methods A **447**, 1 (2000);
A. Affolder, et al., Nucl. Instrum. Methods A **453**, 84 (2000);
T. Affolder et al., Nucl. Instrum. Methods A **526**, 249 (2004);
D. Acosta et al., Nucl. Instrum. Methods A **518**, 605 (2004);
L. Balka, et al., Nucl. Instrum. Methods A **267**, 272 (1988);
S.R. Hahn, et al., Nucl. Instrum. Methods A **267**, 351 (1988);
S. Bertolucci, et al., Nucl. Instrum. Methods A **267**, 301 (1988);
M. Albrow et al., Nucl. Instrum. Methods A **480**, 524 (2002);
G. Apollinari et al., Nucl. Instrum. Methods A **412**, 515 (1998);
G. Ascoli et al., Nucl. Instrum. Methods A **268**, 33 (1988);
T. Dorigo et al., Nucl. Instrum. Methods A **461**, 560 (2001);
D. Acosta et al., Nucl. Instrum. Methods A **494**, 57 (2002);
W. Ashmanskas, et al., Nucl. Instrum. Methods A **518**, 532 (2004).
E.J. Thomson, et al., IEEE Trans. on Nucl. Science **49**, 1063 (2002).

[13] D. Acosta et al., Nucl. Instrum. Methods Phys. Res. A **494** 57 (2002); S.Klimenko, J. Kongisberg, and T.M. Liss, FERMILAB-FN-0741.

[14] T. Sjostrand et al., Comput. Phys. Commun. **135**, 238 (2001). PYTHIA version 6.216 was used.

[15] R. Brun and F. Carminati, CERN Programming Library Long Writeup W5013 (1993). GEANT version 3.21/14 was used.

[16] A. Abulencia et al. [CDF Collaboration], FERMILAB-PUB-05-360-E [arXiv:hep-ex/0508029].

[17] The integrated luminosity of the electron-triggered sample is 1.08 fb^{-1}; that of the muon-triggered sample is 1.04 fb^{-1}; the uncertainty on each is 5.9%, described in [13]. In the averaging, we weight each luminosity with the acceptance of a 200 GeV b' signal after applying the $N_{\mathrm{jet}}^{30} \geq 3$ cut described in Sec. III B. After applying this cut, the acceptances of the $b' \to Z \to ee$ and $b' \to Z \to \mu\mu$ channels are approximately the same, so the weights are approximately equal.

[18] A.L. Scott, Ph.D. thesis, University of California, Santa Barbara (2007).

[19] S.D. Ellis et al., eConf **C010630** P513 (2001), [arXiv:hep-ph/0111434].

[20] A. Bhatti et al., Nucl. Instrum. Methods Phys. Res. A **566**, 375 (2006).

[21] In this section, it is useful to make a distinction between a "threshold" and a "cut." We define a "threshold" as a requirement on a jet E_T; we define a cut as a requirement on a variable made from individual jets. So, in the equation $N_{\mathrm{jet}}^{30} \geq 3$, we say that we have a 30 GeV threshold on the jet E_T, and we have a cut of at least 3 jets.

[22] G. Punzi, *In the Proceedings of PHYSTAT2003: Statistical Problems in Particle Physics, Astrophysics, and Cosmology, Menlo Park, California, 8-11 Sep 2003, pp MODT002* [arXiv:physics/0308063].

[23] The full sensitivity figure of merit obtained from Eq. (8) of [22] is more complicated than that used here. The one used here, $S/(1.5+\sqrt{B})$, is a reduced formula for the full figure of merit, in the case where the significance of discovery and the limit confidence level (both optimized for with this figure of merit) correspond to 3σ probabilities, and where the number of signal and background events can be described by Gaussian (as opposed to Poisson) distributions.

[24] For the $m_{b'} = 100$ GeV/c^2 mass point, the target selection has a signal of 450 on a background of 229, giving a figure of merit $S/(1.5+\sqrt{B}) = 27$. With the simple cuts, there is a signal of 64 on a background of 20, giving a figure of merit $S/(1.5+\sqrt{B}) = 11$. While the sensitivity figure of merit is smaller, a signal of 64 on a background of 20 is still adequate for a discovery.

[25] W.-M. Yao et al. (Particle Data Group), J. Phys. G **33**, 1 (2006).

[26] We use unbinned likelihood-maximization fits. In practice, rather than maximizing the likelihood L, the quantity $-\log L$ is minimized. When comparing unbinned fits with binned histograms, we place the x-value of each bin at the average of the entries in that bin.

[27] As we do not expect many events in data with $N_{\mathrm{jet}}^{30} \geq 4$, a detailed checking of the parameter extrapolation to these jet multiplicities is not necessary. We have verified that the extrapolation is consistent with this distribution with the statistics present in Monte Carlo. In addition, this extrapolation is implicitly validated when the validation of the method as a whole is done in Sec. V D.

[28] Since the cross section for jet events with $E_T > 20$ GeV is extremely large, not all events of this type are able to be kept by the data acquisition system. Only a fraction of these events are kept; the inverse of the fraction of events kept is known as the prescale. For the single jet trigger with $E_T > 20$ GeV, the total prescale used is approximately 500.

[29] This probability calculation is done by integrating the distribution of the expected number of events above the observed value. For the distribution of the expected number of events, we use a Poisson distribution with a mean equal to the fit prediction (97 in this case) smeared with a Gaussian centered at zero with a width equal to the background prediction's uncertainty (± 27 in this case). If the number of data events is higher than the background prediction, we use the upper uncertainty in the Gaussian smearing; we use the lower uncertainty if the data is below the background prediction. Additionally we convert this probability to units of standard deviation by inverting:

$$\int_n^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = p$$

.

[30] The statistical uncertainty is the statistical uncertainty on the background prediction added in quadrature with the statistical uncertainty on the number of observed events, $\sqrt{762}$. We do not give a full systematic uncertainty, as we do not evaluate the uncertainty on the acceptance of $t\bar{t}$. The systematic uncertainty on the cross section from the background prediction alone is 1.5 pb.

[31] D. Acosta et al. [CDF Collaboration], Phys. Rev. D **72**, 052003 (2005).

[32] Section 4 of J. Conway, *Proceedings of Workshop on Confidence Limits, Geneva, Switzerland, 17-18 Jan 2000*, FERMILAB-CONF-00-048-E.

[33] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. Nadolsky and W. K. Tung, JHEP **0207**, 012 (2002).