# Search for B_{s}^{0}→μ^{+}μ^{-} and B^{0}→μ^{+}μ^{-} decays with the full CDF Run II data set

T. Aaltonen *et al.* (CDF Collaboration)

Phys. Rev. D **87**, 072003 — Published 4 April 2013

DOI: 10.1103/PhysRevD.87.072003

# Search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with the full CDF Run II data set


T. Aaltonen,[21] S. Amerio,[40] D. Amidei,[32] A. Anastassov[x],[15] A. Annovi,[17] J. Antos,[12] G. Apollinari,[15] J.A. Appel,[15] T. Arisawa,[53] A. Artikov,[13] J. Asaadi,[48] W. Ashmanskas,[15] B. Auerbach,[2] A. Aurisano,[48] F. Azfar,[39] W. Badgett,[15] T. Bae,[25] A. Barbaro-Galtieri,[26] V.E. Barnes,[44] B.A. Barnett,[23] P. Barria[hh],[42] P. Bartos,[12] M. Bauce[ff],[40] F. Bedeschi,[42] S. Behari,[15] G. Bellettini[gg],[42] J. Bellinger,[55] D. Benjamin,[14] A. Beretvas,[15] A. Bhatti,[46] K.R. Bland,[5] B. Blumenfeld,[23] A. Bocci,[14] A. Bodek,[45] D. Bortoletto,[44] J. Boudreau,[43] A. Boveia,[11] L. Brigliadori[ee],[6] C. Bromberg,[33] E. Brucken,[21] J. Budagov,[13] H.S. Budd,[45] K. Burkett,[15] G. Busetto[ff],[40] P. Bussey,[19] P. Butti[gg],[42] A. Buzatu,[19] A. Calamba,[10] S. Camarda,[4] M. Campanelli,[28] F. Canelli[oo],[11, 15] B. Carls,[22] D. Carlsmith,[55] R. Carosi,[42] S. Carrillo[m],[16] B. Casal[k],[9] M. Casarsa,[49] A. Castro[ee],[6] P. Catastini,[20] D. Cauz,[49] V. Cavaliere,[22] M. Cavalli-Sforza,[4] A. Cerri[f],[26] L. Cerrito[s],[28] Y.C. Chen,[1] M. Chertok,[7] G. Chiarelli,[42] G. Chlachidze,[15] K. Cho,[25] D. Chokheli,[13] M.A. Ciocci[hh],[42] A. Clark,[18] C. Clarke,[54] M.E. Convery,[15] J. Conway,[7] M .Corbo,[15] M. Cordelli,[17] C.A. Cox,[7] D.J. Cox,[7] M. Cremonesi,[42] D. Cruz,[48] J. Cuevas[z],[9] R. Culbertson,[15] N. d'Ascenzo[w],[15] M. Datta[qq],[15] P. De Barbaro,[45] L. Demortier,[46] M. Deninno,[6] F. Devoto,[21] M. d'Errico[ff],[40] A. Di Canto[gg],[42] B. Di Ruzza[q],[15] J.R. Dittmann,[5] M. D'Onofrio,[27] S. Donati[gg],[42] M. Dorigo[nn],[49] A. Driutti,[49] K. Ebina,[53] R. Edgar,[32] A. Elagin,[48] R. Erbacher,[7] S. Errede,[22] B. Esham,[22] R. Eusebi,[48] S. Farrington,[39] J.P. Fernández Ramos,[29] R. Field,[16] G. Flanagan[u],[15] R. Forrest,[7] M. Franklin,[20] J.C. Freeman,[15] H. Frisch,[11] Y. Funakoshi,[53] A.F. Garfinkel,[44] P. Garosi[hh],[42] H. Gerberich,[22] E. Gerchtein,[15] S. Giagu,[47] V. Giakoumopoulou,[3] K. Gibson,[43] C.M. Ginsburg,[15] N. Giokaris,[3] P. Giromini,[17] G. Giurgiu,[23] V. Glagolev,[13] D. Glenzinski,[15] M. Gold,[35] D. Goldin,[48] A. Golossanov,[15] G. Gomez,[9] G. Gomez-Ceballos,[30] M. Goncharov,[30] O. González López,[29] I. Gorelov,[35] A.T. Goshaw,[14] K. Goulianos,[46] E. Gramellini,[6] S. Grinstein,[4] C. Grosso-Pilcher,[11] R.C. Group[52],[15] J. Guimaraes da Costa,[20] S.R. Hahn,[15] J.Y. Han,[45] F. Happacher,[17] K. Hara,[50] M. Hare,[51] R.F. Harr,[54] T. Harrington-Taber[n],[15] K. Hatakeyama,[5] C. Hays,[39] J. Heinrich,[41] M. Herndon,[55] A. Hocker,[15] Z. Hong,[48] W. Hopkins[g],[15] S. Hou,[1] R.E. Hughes,[36] U. Husemann,[56] J. Huston,[33] G. Introzzi[mm],[42] M. Iori[jj],[47] A. Ivanov[p],[7] E. James,[15] D. Jang,[10] B. Jayatilaka,[15] E.J. Jeon,[25] S. Jindariani,[15] M. Jones,[44] K.K. Joo,[25] S.Y. Jun,[10] T.R. Junk,[15] M. Kambeitz,[24] T. Kamon[25],[48] P.E. Karchin,[54] A. Kasmi,[5] Y. Kato[o],[38] W. Ketchum[rr],[11] J. Keung,[41] B. Kilminster[oo],[15] D.H. Kim,[25] H.S. Kim,[25] J.E. Kim,[25] M.J. Kim,[17] S.B. Kim,[25] S.H. Kim,[50] Y.K. Kim,[11] Y.J. Kim,[25] N. Kimura,[53] M. Kirby,[15] K. Knoepfel,[15] K. Kondo*,[53] D.J. Kong,[25] J. Konigsberg,[16] A.V. Kotwal,[14] M. Kreps,[24] J. Kroll,[41] M. Kruse,[14] T. Kuhr,[24] M. Kurata,[50] A.T. Laasanen,[44] S. Lammel,[15] M. Lancaster,[28] K. Lannon[y],[36] G. Latino[hh],[42] H.S. Lee,[25] J.S. Lee,[25] S. Leo,[42] S. Leone,[42] J.D. Lewis,[15] A. Limosani[t],[14] E. Lipeles,[41] H. Liu,[52] Q. Liu,[44] T. Liu,[15] S. Lockwitz,[56] A. Loginov,[56] D. Lucchesi[ff],[40] J. Lueck,[24] P. Lujan,[26] P. Lukens,[15] G. Lungu,[46] J. Lys,[26] R. Lysak[e],[12] R. Madrak,[15] P. Maestro[hh],[42] S. Malik,[46] G. Manca[a],[27] A. Manousakis-Katsikakis,[3] F. Margaroli,[47] P. Marino[ii],[42] M. Martínez,[4] K. Matera,[22] M.E. Mattson,[54] A. Mazzacane,[15] P. Mazzanti,[6] R. McNulty[j],[27] A. Mehta,[27] P. Mehtala,[21] C. Mesropian,[46] T. Miao,[15] D. Mietlicki,[32] A. Mitra,[1] H. Miyake,[50] S. Moed,[15] N. Moggi,[6] C.S. Moon[aa],[15] R. Moore[pp],[15] M.J. Morello[ii],[42] A. Mukherjee,[15] Th. Muller,[24] P. Murat,[15] M. Mussini[ee],[6] J. Nachtman[n],[15] Y. Nagai,[50] J. Naganoma,[53] I. Nakano,[37] A. Napier,[51] J. Nett,[48] C. Neu,[52] T. Nigmanov,[43] L. Nodulman,[2] S.Y. Noh,[25] O. Norniella,[22] L. Oakes,[39] S.H. Oh,[14] Y.D. Oh,[25] I. Oksuzian,[52] T. Okusawa,[38] R. Orava,[21] L. Ortolan,[4] C. Pagliarone,[49] E. Palencia[f],[9] P. Palni,[35] V. Papadimitriou,[15] W. Parker,[55] G. Pauletta[kk],[49] M. Paulini,[10] C. Paus,[30] T.J. Phillips,[14] G. Piacentino,[42] E. Pianori,[41] J. Pilot,[36] K. Pitts,[22] C. Plager,[8] L. Pondrom,[55] S. Poprocki[g],[15] K. Potamianos,[26] F. Prokoshin[cc],[13] A. Pranko,[26] F. Ptohos[h],[17] G. Punzi[gg],[42] N. Ranjan,[44] I. Redondo Fernández,[29] P. Renton,[39] M. Rescigno,[47] T. Riddick,[28] F. Rimondi*,[6] L. Ristori[42],[15] A. Robson,[19] T. Rodriguez,[41] S. Rolli[i],[51] M. Ronzani[gg],[42] R. Roser,[15] J.L. Rosner,[11] F. Ruffini[hh],[42] A. Ruiz,[9] J. Russ,[10] V. Rusu,[15] A. Safonov,[48] W.K. Sakumoto,[45] Y. Sakurai,[53] L. Santi[kk],[49] K. Sato,[50] V. Saveliev[w],[15] A. Savoy-Navarro[aa],[15] P. Schlabach,[15] E.E. Schmidt,[15] T. Schwarz,[32] L. Scodellaro,[9] F. Scuri,[42] S. Seidel,[35] Y. Seiya,[38] A. Semenov,[13] F. Sforza[gg],[42] S.Z. Shalhout,[7] T. Shears,[27] P.F. Shepard,[43] M. Shimojima[v],[50] M. Shochet,[11] I. Shreyber-Tecker,[34] A. Simonenko,[13] P. Sinervo,[31] K. Sliwa,[51] J.R. Smith,[7] F.D. Snider,[15] V. Sorin,[4] H. Song,[43] D. Sperka,[55] M. Stancari,[15] R. St. Denis,[19] B. Stelzer,[31] O. Stelzer-Chilton,[31] D. Stentz[x],[15] J. Strologas,[35] Y. Sudo,[50] A. Sukhanov,[15] I. Suslov,[13] K. Takemasa,[50] Y. Takeuchi,[50] J. Tang,[11] M. Tecchio,[32] P.K. Teng,[1] J. Thom[g],[15] E. Thomson,[41] V. Thukral,[48] D. Toback,[48] S. Tokar,[12] K. Tollefson,[33] T. Tomura,[50] D. Tonelli[f],[15] S. Torre,[17] D. Torretta,[15] P. Totaro,[40] M. Trovato[ii],[42] F. Ukegawa,[50] S. Uozumi,[25] F. Vázquez[m],[16] G. Velev,[15] C. Vellidis,[15] C. Vernieri[ii],[42] M. Vidal,[44] R. Vilar,[9] J. Vizán[ll],[9] M. Vogel,[35] G. Volpi,[17] P. Wagner,[41] R. Wallny,[8] S.M. Wang,[1] A. Warburton,[31] D. Waters,[28] W.C. Wester III,[15] D. Whiteson[b],[41] A.B. Wicklund,[2] S. Wilbur,[11] H.H. Williams,[41] J.S. Wilson,[32] P. Wilson,[15] B.L. Winer,[36] P. Wittich[g],[15] S. Wolbers,[15] H. Wolfe,[36] T. Wright,[32] X. Wu,[18] Z. Wu,[5] K. Yamamoto,[38]

D. Yamato,[38] T. Yang,[15] U.K. Yang[r],[11] Y.C. Yang,[25] W.-M. Yao,[26] G.P. Yeh,[15] K. Yi[n],[15] J. Yoh,[15] K. Yorita,[53] T. Yoshida[l],[38] G.B. Yu,[14] I. Yu,[25] A.M. Zanetti,[49] Y. Zeng,[14] C. Zhou,[14] and S. Zucchelli[ee6]

(CDF Collaboration[†])

[1]*Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China*
[2]*Argonne National Laboratory, Argonne, Illinois 60439, USA*
[3]*University of Athens, 157 71 Athens, Greece*
[4]*Institut de Fisica d'Altes Energies, ICREA, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain*
[5]*Baylor University, Waco, Texas 76798, USA*
[6]*Istituto Nazionale di Fisica Nucleare Bologna, [ee] University of Bologna, I-40127 Bologna, Italy*
[7]*University of California, Davis, Davis, California 95616, USA*
[8]*University of California, Los Angeles, Los Angeles, California 90024, USA*
[9]*Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain*
[10]*Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*
[11]*Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA*
[12]*Comenius University, 842 48 Bratislava, Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia*
[13]*Joint Institute for Nuclear Research, RU-141980 Dubna, Russia*
[14]*Duke University, Durham, North Carolina 27708, USA*
[15]*Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[16]*University of Florida, Gainesville, Florida 32611, USA*
[17]*Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy*
[18]*University of Geneva, CH-1211 Geneva 4, Switzerland*
[19]*Glasgow University, Glasgow G12 8QQ, United Kingdom*
[20]*Harvard University, Cambridge, Massachusetts 02138, USA*
[21]*Division of High Energy Physics, Department of Physics,*
*University of Helsinki and Helsinki Institute of Physics, FIN-00014, Helsinki, Finland*
[22]*University of Illinois, Urbana, Illinois 61801, USA*
[23]*The Johns Hopkins University, Baltimore, Maryland 21218, USA*
[24]*Institut für Experimentelle Kernphysik, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany*
[25]*Center for High Energy Physics: Kyungpook National University,*
*Daegu 702-701, Korea; Seoul National University, Seoul 151-742,*
*Korea; Sungkyunkwan University, Suwon 440-746,*
*Korea; Korea Institute of Science and Technology Information,*
*Daejeon 305-806, Korea; Chonnam National University,*
*Gwangju 500-757, Korea; Chonbuk National University, Jeonju 561-756,*
*Korea; Ewha Womans University, Seoul, 120-750, Korea*
[26]*Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*
[27]*University of Liverpool, Liverpool L69 7ZE, United Kingdom*
[28]*University College London, London WC1E 6BT, United Kingdom*
[29]*Centro de Investigaciones Energeticas Medioambientales y Tecnologicas, E-28040 Madrid, Spain*
[30]*Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[31]*Institute of Particle Physics: McGill University, Montréal, Québec H3A 2T8,*
*Canada; Simon Fraser University, Burnaby, British Columbia V5A 1S6,*
*Canada; University of Toronto, Toronto, Ontario M5S 1A7,*
*Canada; and TRIUMF, Vancouver, British Columbia V6T 2A3, Canada*
[32]*University of Michigan, Ann Arbor, Michigan 48109, USA*
[33]*Michigan State University, East Lansing, Michigan 48824, USA*
[34]*Institution for Theoretical and Experimental Physics, ITEP, Moscow 117259, Russia*
[35]*University of New Mexico, Albuquerque, New Mexico 87131, USA*
[36]*The Ohio State University, Columbus, Ohio 43210, USA*
[37]*Okayama University, Okayama 700-8530, Japan*
[38]*Osaka City University, Osaka 588, Japan*
[39]*University of Oxford, Oxford OX1 3RH, United Kingdom*
[40]*Istituto Nazionale di Fisica Nucleare, Sezione di Padova-Trento, [ff] University of Padova, I-35131 Padova, Italy*
[41]*University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[42]*Istituto Nazionale di Fisica Nucleare Pisa, [gg] University of Pisa,*
*[hh] University of Siena and [ii] Scuola Normale Superiore, I-56127 Pisa,*
*Italy, [mm] INFN Pavia and University of Pavia, I-27100 Pavia, Italy*
[43]*University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA*
[44]*Purdue University, West Lafayette, Indiana 47907, USA*
[45]*University of Rochester, Rochester, New York 14627, USA*
[46]*The Rockefeller University, New York, New York 10065, USA*
[47]*Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1,*
*[jj] Sapienza Università di Roma, I-00185 Roma, Italy*

[48] Texas A&M University, College Station, Texas 77843, USA
[49] Istituto Nazionale di Fisica Nucleare Trieste/Udine; [nn] University of Trieste, I-34127 Trieste, Italy; [kk] University of Udine, I-33100 Udine, Italy
[50] University of Tsukuba, Tsukuba, Ibaraki 305, Japan
[51] Tufts University, Medford, Massachusetts 02155, USA
[52] University of Virginia, Charlottesville, Virginia 22906, USA
[53] Waseda University, Tokyo 169, Japan
[54] Wayne State University, Detroit, Michigan 48201, USA
[55] University of Wisconsin, Madison, Wisconsin 53706, USA
[56] Yale University, New Haven, Connecticut 06520, USA

We report on a search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays using proton-antiproton collision data at $\sqrt{s} = 1.96$ TeV corresponding to 10 fb$^{-1}$ of integrated luminosity collected by the CDF II detector at the Fermilab Tevatron collider. The observed number of $B^0$ candidates is consistent with background-only expectations and yields an upper limit on the branching fraction of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 4.6\times10^{-9}$ at 95% confidence level. We observe an excess of $B_s^0$ candidates. The probability that the background processes alone could produce such an excess or larger is 0.94%. The probability that the combination of background and the expected standard model rate of $B_s^0 \to \mu^+\mu^-$ could produce such an excess or larger is 6.8%. These data are used to determine a branching fraction $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (1.3^{+0.9}_{-0.7}) \times 10^{-8}$ and provide an upper limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.1 \times 10^{-8}$ at 95% confidence level.

PACS numbers: 13.20.He 13.30.Ce 12.15.Mm 12.60.Jv

## I. INTRODUCTION

The study of decays of the $B_s^0$ meson (with a quark content of $\overline{b}s$) and the $B^0$ meson ($\overline{b}d$) into a dimuon pair ($\mu^+\mu^-$) has long been of great interest as a test of the standard model (SM) of particle physics. These flavor-changing neutral-current (FCNC) decays occur in the SM only through weak-interaction-mediated loop processes whose amplitudes are suppressed via the Glashow-Iliopoulos-Maiani mechanism [1] and by helicity conservation. In the SM, the branching fractions for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ are predicted to be $(3.2 \pm 0.3) \times 10^{-9}$ and $(1.1 \pm 0.1) \times 10^{-10}$, respectively [2]. Note that the inclusion of charge conjugate modes is implied throughout this paper. Non-SM particles in the loop processes or non-SM coupling mechanisms can significantly alter the rate of these decays so that measurements of their branching fractions serve as powerful tools to probe for the effects of new physics beyond the SM. For example, in the minimal supersymmetric standard model the $B_s^0 \to \mu^+\mu^-$ decay rate is proportional to $(\tan\beta)^6$ [3–5], where $\tan\beta$ is the ratio of the vacuum expectation values of the two Higgs fields. The decay rate can be enhanced relative to the SM by over two orders of magnitude at large $\tan\beta$ values. The search is also sensitive to supersymmetry (SUSY) in cosmologically consistent scenarios [6–9]. Other models, such as R-parity violating SUSY [6], littlest Higgs model with T-parity [10], or models with extra dimensions [11, 12], predict large effects independent of the value of $\tan\beta$. Substantial negative interference effects can suppress the $B_s^0 \to \mu^+\mu^-$ branching fraction by as much as a factor of three in portions of the SUSY parameter space [13]. In the absence of an observation, limits on $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are complementary to limits provided by direct searches in constraining the new-physics parameter-space. The status of constraints on new physics in a variety of different models and in a model independent treatment are discussed in Ref. [14].

Recent results on $\mathcal{B}(B_s^0(B^0) \to \mu^+\mu^-)$ include limits from the ATLAS [15], CMS [16], D0 [17], and CDF [18]

*Deceased
†With visitors from [a] Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09042 Monserrato (Cagliari), Italy, [b] University of California Irvine, Irvine, CA 92697, USA, [c] University of California Santa Barbara, Santa Barbara, CA 93106, USA, [d] University of California Santa Cruz, Santa Cruz, CA 95064, USA, [e] Institute of Physics, Academy of Sciences of the Czech Republic, 182 21, Czech Republic, [f] CERN, CH-1211 Geneva, Switzerland, [g] Cornell University, Ithaca, NY 14853, USA, [h] University of Cyprus, Nicosia CY-1678, Cyprus, [i] Office of Science, U.S. Department of Energy, Washington, DC 20585, USA, [j] University College Dublin, Dublin 4, Ireland, [k] ETH, 8092 Zürich, Switzerland, [l] University of Fukui, Fukui City, Fukui Prefecture, Japan 910-0017, [m] Universidad Iberoamericana, Lomas de Santa Fe, México, C.P. 01219, Distrito Federal, [n] University of Iowa, Iowa City, IA 52242, USA, [o] Kinki University, Higashi-Osaka City, Japan 577-8502, [p] Kansas State University, Manhattan, KS 66506, USA, [q] Brookhaven National Laboratory, Upton, NY 11973, USA, [r] University of Manchester, Manchester M13 9PL, United Kingdom, [s] Queen Mary, University of London, London, E1 4NS, United Kingdom, [t] University of Melbourne, Victoria 3010, Australia, [u] Muons, Inc., Batavia, IL 60510, USA, [v] Nagasaki Institute of Applied Science, Nagasaki 851-0193, Japan, [w] National Research Nuclear University, Moscow 115409, Russia, [x] Northwestern University, Evanston, IL 60208, USA, [y] University of Notre Dame, Notre Dame, IN 46556, USA, [z] Universidad de Oviedo, E-33007 Oviedo, Spain, [aa] CNRS-IN2P3, Paris, F-75205 France, [bb] Texas Tech University, Lubbock, TX 79609, USA, [cc] Universidad Tecnica Federico Santa Maria, 110v Valparaiso, Chile, [dd] Yarmouk University, Irbid 211-63, Jordan, [ll] Universite catholique de Louvain, 1348 Louvain-La-Neuve, Belgium, [oo] University of Zürich, 8006 Zürich, Switzerland, [pp] Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114 USA, [qq] Hampton University, Hampton, VA 23668, USA, [rr] Los Alamos National Laboratory, Los Alamos, NM 87544, USA

experiments. The most sensitive result is from the LHCb experiment [19], which reported an excess of $B_s^0 \to \mu^+\mu^-$ events and measured $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 3.2^{+1.5}_{-1.2} \times 10^{-9}$ and an upper limit for the $\mathcal{B}(B^0 \to \mu^+\mu^-)$ within a factor of about nine of the SM $B^0 \to \mu^+\mu^-$ rate. The previous CDF result reported $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.8^{+1.1}_{-0.9} \times 10^{-8}$ and sets a two-sided interval at 90% C.L. of $4.6 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.9 \times 10^{-8}$ and an upper limit of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.5 \times 10^{-8}$ at 90% C.L. The $B_s^0$ and $B^0$ results from all experiments are compatible and indicate that there is no strong enhancement in the $B_s^0 \to \mu^+\mu^-$ decay rate. Further measurements of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are likely to constrain strongly new physics models predicting significant deviations from the SM predictions.

We report on a search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays using the complete Run II data set of $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV collected by the upgraded Collider Detector at Fermilab (CDF II) and corresponding to an integrated luminosity of 10 fb$^{-1}$. Because the previous CDF analysis [18], using 7 fb$^{-1}$ of integrated luminosity, reported an excess of $B_s^0 \to \mu^+\mu^-$ signal events, the same analysis methodology is applied to the full available data set. The sensitivity of the analysis reported here is improved with respect to that reported in Ref. [18] due to the 24% increase in event-sample size. All other aspects of the analysis have remained the same.

## II. EXPERIMENTAL SETUP: THE CDF II DETECTOR

The CDF II detector is a general-purpose detector [20–22] with cylindrical symmetry (Fig. 1) designed to detect products of $p\overline{p}$ collisions at a center-of-mass energy of $\sqrt{s} = 1.96$ TeV. A cylindrical coordinate system is used to describe particle trajectories. The $z$ axis is defined as the direction of the proton beam. Besides the azimuthal angle $\phi$, radius relative to the beam line $r$, and polar angle $\theta$, we define a pseudorapidity $\eta = -\ln(\tan(\theta/2))$. The transverse momentum, $p_T$, represents the component of a particle's momentum in the plane perpendicular to the beam axis, $p_T = p \sin\theta$.

The most important subdetectors for this analysis are briefly described below and include the tracking system and the muon system. Additional subsystems such as the calorimeters and luminosity detector system also play a role in the analysis. The calorimeters are used in part of the particle identification process, while information from the luminosity detector system is used in some of the background estimations. A more detailed description of the CDF II detector can be found in Ref. [20].

### A. Tracking system

The tracking system consists of silicon microstrip detectors, a multi-wire open-cell drift chamber, and a solenoidal superconducting magnet. The innermost

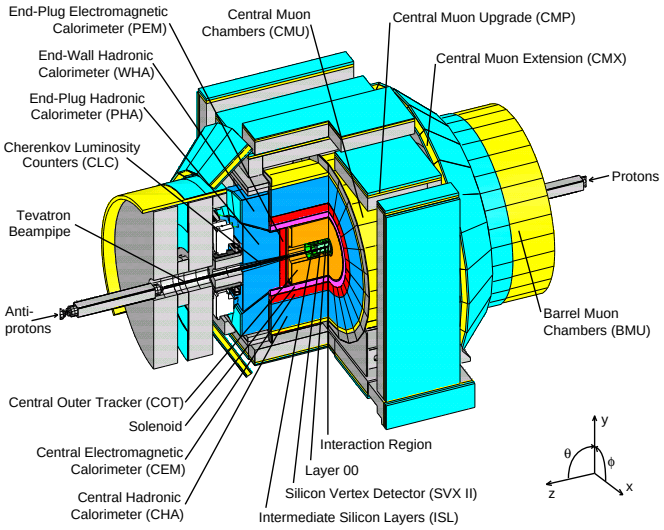


FIG. 1: Cutaway isometric view of the CDF II detector.

tracking system, L00, is a single-sided silicon microstrip system mounted on the beam pipe [23]. Outside L00 is the SVXII detector, with five layers of double-sided silicon microstrip sensors [24]. One side of each sensor provides azimuthal ($r\phi$) information while the opposite side provides longitudinal information ($rz$). The SVXII hit resolution is 11 $\mu$m, while the impact parameter resolution for charged particles with $p_T > 2.0$ GeV/$c$ is about 40 $\mu$m, which includes a 35 $\mu$m contribution due to the size of the $p\overline{p}$ luminous region. The association of SVXII $rz$ hits allows measurement of the $z$ coordinate of charged-particle trajectories (tracks) at the $p\overline{p}$ interaction point with 70 $\mu$m resolution. The combination of excellent $r\phi$ and $z$ resolution allows precise determination of the three-dimensional spacepoint defined by the $B_s^0(B^0) \to \mu^+\mu^-$ decay vertex and the rejection of background from pairs of random muon-candidates that accidentally meet the selection requirements (combinatorics).

Outside the silicon subsystems is the COT [25], an open-cell multi-wire drift chamber divided into eight concentric superlayers. The superlayers themselves are divided in $\phi$ into supercells, each containing 12 sense wires. In addition to charged-particle trajectories, the COT also measures the ionization $dE/dx$ per unit path-length for particle identification. For this analysis the $dE/dx$ information is mainly used to help reject kaons.

Surrounding the COT is a superconducting solenoidal magnet producing a 1.4 T magnetic field parallel to the beam axis. The $p_T$ resolution of the COT, $\sigma_{p_T}/p_T^2 \approx 0.15\%$ (GeV/$c$)$^{-1}$ [25], is determined by comparing the curvature of inward- and outward-going tracks of cosmic-ray events. The absolute momentum scale is determined using $J/\psi$, $\Upsilon$, and $Z$-boson resonances, where the resonances decay into two muons [26].

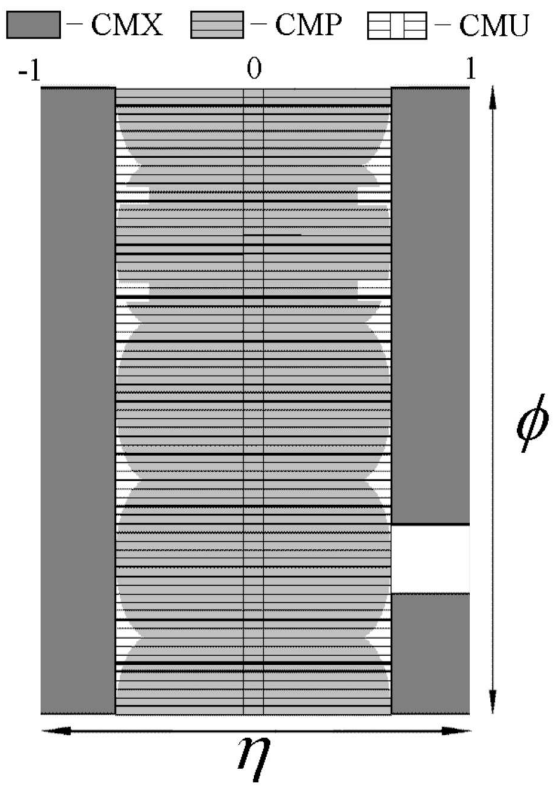


FIG. 2: Muon system $\phi$ and $\eta$ coverage.

### B. Muon system

Outside the solenoidal magnet are electromagnetic and hadronic calorimeters, which in turn are surrounded by the muon systems consisting of multi-layer single-wire drift chambers and scintillators. The drift chambers are used to reconstruct muon-track segments (stubs) while the scintillators are used for timing information to match muon candidates to the correct $p\overline{p}$ collision crossing. The $\eta$ and $\phi$ coverage of the muon subdetectors used in this analysis is shown in Fig. 2.

In the central region, the cylindrical central muon chambers (CMU) [27] provide coverage up to $|\eta| < 0.6$. Because material corresponding to 5.5 interaction lengths lies between the $p\overline{p}$ luminous region and the CMU, a muon must have a minimum $p_T$ of 1.4 GeV/$c$ (range-out threshold) to reach the CMU. The CMU is subdivided in 24 wedges in $\phi$, each with four layers of drift chambers, and lies immediately outside the central hadronic calorimeter.

Beyond the CMU are additional central chambers with nearly the same $\eta$ coverage known as the CMP. Additional steel absorber with a thickness of 2.3 interaction lengths is placed between the CMU and CMP yielding a range-out threshold $p_T$ of 2.2 GeV/$c$. The CMP forms a box around the cylindrical CMU and is comprised of four layers of drift chambers, and a layer of scintillator.

The CMX [20] detector extends the muon system coverage to higher pseudorapidity, $0.6 < |\eta| < 1.0$. The CMX consists of two arches at each end of the detector, with additional upper and lower sets of chambers. The CMX consists of eight layers of drift chambers arranged in conic sections; the geometry is such that particles can traverse six of the eight layers on average. The range-out threshold $p_T$ for the CMX is about 2.0 GeV/$c$.

In this analysis muons are required to have either a CMU or CMX stub and have the stub matched to an extrapolated track from the tracking system. Further information, such as $z$ position and stub angle, is used in a multivariate likelihood discriminant for muon identification. Information from the CMP detector is used, if available, to identify high-purity muons. The muons are paired into either a CMU-CMU (CC) or a CMU-CMX (CF) channel. The selection criteria for these two channels are discussed in the next section.

### C. Muon trigger

At the Tevatron $p\overline{p}$ crossings occurred every 396 ns (2.5 MHz) and peak instantaneous luminosities of $4 \times 10^{32}$ cm$^{-2}$s$^{-1}$ were achieved, creating an event rate of 1.7 MHz. An online system of custom hardware boards and software algorithms (the "trigger" system) was employed to reduce the data rate to about 100 events per second, which were recorded to tape for later analysis. The CDF trigger system was divided into three consecutive levels with increasing granularity, sophistication, and precision. Trigger level one (L1) was evaluated for every $p\overline{p}$ crossing and used coarse track, calorimeter, and muon-stub information to identify potentially interesting events that were then passed to trigger level two. Trigger level two (L2) used more precise calorimeter and muon-stub information to eliminate events poorly reconstructed in trigger L1. Accepted events were passed to trigger level three (L3), a CPU farm performing full event reconstruction and identifying the most interesting events to record to tape for later analysis. The data sets used in the present analysis were collected with a set of L1, L2, and L3 triggers that required a pair of muon candidates.

In L1, muons were identified by matching a track reconstructed in the COT to a muon stub reconstructed in one of the muon systems. The track reconstruction was performed by a custom-built system [28] that achieved a $p_T$ resolution of $\sigma_{p_T}/p_T^2 = 1.7\%/(\text{GeV}/c)$. Custom electronic boards identified muon stubs in each of the CMU, CMX, and CMP systems and performed a coarse matching to the L1 tracks [29]. Events were required to have two separate CMU-track matches or one CMU-track match and one CMX-track match. The particles associated to the tracks were required to have opposite electric charge. In L2 the track muon-stub matches were confirmed using more sophisticated algorithms and improved resolutions. In order to remove through-going cosmic-ray muons and backgrounds from dijet events that generate falsely-identified muon candidates (fake muons),

only events with a dimuon opening-angle less than 120 degrees in the plane transverse to the beam line were passed to L3. The full event reconstruction employed in L3 performed a full track fit and required CMU muon candidates to have $p_T > 1.5$ GeV/$c$, CMX muon candidates to have $p_T > 2.0$ GeV/$c$, the scalar sum $p_T$ for the two muon candidates to exceed 5 GeV/$c$, the dimuon mass to be less than 6 GeV/$c^2$, and the difference in the $z$ coordinates of the muon tracks at the point of closest approach to the beam line to satisfy $|\Delta z_0| < 5$ cm. At the highest instantaneous luminosities, the accept rate of this dimuon trigger path was too high and events were randomly discarded with a frequency that depended on the instantaneous luminosity. This reduction in rate discarded approximately 10% of the total dimuon candidate events. In the full Run II data set collected by CDF, 822 740 (498 443) events satisfied the CMU-CMU (CMU-CMX) trigger path with $M_{\mu^+\mu^-} > 4.669$ GeV/$c^2$ and formed the initial data sample for the CC (CF) channel.

## III. MONTE CARLO SIMULATION

We employ Monte Carlo (MC) simulations of $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays together with a full CDF detector simulation to estimate signal efficiencies not measurable with data control samples. In addition to the $B_s^0(B^0) \to \mu^+\mu^-$ MC sample we also produce a sample of $B^+ \to J/\psi\ K^+$ simulated events for modeling cross-checks. A MC sample of $B_s^0 \to \mu^+\mu^-$ is generated using Pythia [30] and EvtGen [31] with the underlying event modeling tuned to reproduce minimum-bias events in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV [32]. We generate simulations of $b\overline{b}$ pair production and their subsequent hadronization. One of the resulting $B$ hadrons is required to be a $B_s^0$ or $B^0$ meson that decays to two muons. There are no requirements on the second $B$ hadron, which is allowed to decay inclusively. The MC events are also run through a detailed CDF II detector simulation [33] that accounts for resolution and occupancy effects in all the subdetector systems. The MC events are required to meet all the baseline requirements discussed in Sec. IV A with the exception of the $dE/dx$ and muon likelihood requirements, which are omitted because data-driven estimates of their efficiency are used instead. We weight the $B_s^0$-meson $p_T$ and isolation (cf. Sec. IV A) distributions to match the measured spectra obtained from $B^+ \to J/\psi\ K^+$ data and $B_s^0 \to J/\psi\ \phi$ data, respectively. The $B^+$ $p_T$ spectrum is expected to be similar to that of the $B_s^0$ for $p_T > 4.0$ GeV/$c$ and provides a significantly larger sample size. The isolation spectra, however, may differ between $B^+$ and $B_s^0$ mesons due to the participation of $\overline{u}$ and $\overline{s}$ quarks in the hadronization processes, respectively. For the $B^0 \to \mu^+\mu^-$ search the $B^0$-meson $p_T$ and isolation distributions are weighted using $B^+ \to J/\psi\ K^+$ data.

We use simulated $B_s^0 \to \mu^+\mu^-$ decays to estimate a mass resolution of 24 MeV/$c^2$ for events passing the baseline and vertex requirements described in Sec. IV A. Comparisons between data and MC using $J/\psi \to \mu^+\mu^-$ and $B^+ \to J/\psi\ (\to \mu^+\mu^-)\ K^+$ samples reveal a 10% discrepancy in mass resolution, which is propagated as a systematic uncertainty and negligibly affects the efficiencies discussed in Sec. V.

## IV. EVENT SELECTION

This analysis searches for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays using the full 10 fb$^{-1}$ CDF II data set. The same analysis methods are used for both decays. The signal-search dimuon-mass range is adapted to the different $B^0$ and $B_s^0$ pole masses and corresponds to $\pm 2.5$ times the two-track mass resolution. The branching fractions are measured relative to a $B^+ \to J/\psi\ (\to \mu^+\mu^-)\ K^+$ normalization mode. This mode, together with directly produced $J/\psi \to \mu^+\mu^-$ decays, are used to estimate signal efficiencies and perform consistency checks. Initially we apply baseline requirements (described in Sec. IV A) on all data and MC samples. An artificial neural network (NN) classifier is then applied to enhance the expected signal-to-background-ratio. In order to avoid inadvertent biases, the data in an extended mass-signal region, $5.169 < M_{\mu^+\mu^-} < 5.469$ GeV/$c^2$, are kept hidden until all selection criteria are finalized. An unbiased optimization of the analysis is performed using mass-sideband data as a model of the combinatorial background and MC events as a model of the peaking backgrounds and signal.

Major background processes include Drell-Yan dimuon production ($q\bar{q} \to \mu^+\mu^-$) processes through virtual $\gamma$ and $Z$-boson states, double semileptonic decay ($b\bar{b}, c\bar{c} \to \mu^+\mu^- X$), and sequential semileptonic decay ($b \to c\mu^- X \to s\mu^-\mu^+ X'$) of $b$ and $c$ quarks. A combination of a semileptonic decay and a fake muon or two fake muons from two-body hadronic decays of $B$ hadrons ($B \to h^+h'^-$ where $h$ and $h'$ are charged pions and kaons), can also be a source of background. Fake muons are tracks from pions and kaons that have a matching muon stub and are falsely identified as muons. The backgrounds can be divided into a combinatorial dimuon background and a peaking $B \to h^+h'^-$ background, which are estimated separately. Backgrounds are studied in detail in statistically-independent control samples with various baseline requirements inverted or relaxed to enhance the background contribution.

The search for a $B_s^0(B^0) \to \mu^+\mu^-$ signal is performed in bins of NN output. The NN binning is determined by an *a priori* optimization, discussed in detail in Sec. VIII, that uses the expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ limit in absence of signal as a figure of merit, resulting in eight NN bins. Additionally, the signal region is divided into five mass bins centered on the world average $B_s^0$ and $B^0$ masses. This yields a total of 80 single-bin counting experiments corresponding to the CC and CF topologies, each with eight NN bins, and five mass bins.

Once the signal efficiencies and background estimates are well understood, a thorough statistical analysis of the result is performed. The sections below discuss the analysis methodology and results in more detail.

### A. Baseline event selection

Except where specifically discussed, all samples used in this analysis are required to pass a set of baseline requirements that consist of kinematic-, particle-identification-, and vertex-related requirements discussed in this section.

Muon-candidate tracks are required to be matched with a muon identified by the trigger and have $p_T > 2.0$ GeV/$c$ and $p_T > 2.2$ GeV/$c$ for CMU and CMX muons, respectively. Tracks are required to be fiducial to the COT, by demanding that the absolute value of the $z$ coordinate be less than 155 cm at the COT exit radius ($r = 136$ cm). This ensures that tracks traverse the full radial extent of the COT. Tracks are also required to have $r\phi$ hits in at least three L00+SVXII layers.

A likelihood method [34] together with a $dE/dx$ based selection [35] is used to further suppress contributions from hadrons misidentified as muons. The muon likelihood is based on matching muon stubs to COT tracks in the $\phi$ and $z$ coordinates, energy deposits in the electromagnetic and hadronic calorimeters, information concerning the subsystems in which the muon is identified, and kinematic information. We require a muon likelihood with a value greater than 0.1, which is approximately 99% efficient for signal, while rejecting 50% of the combinatorial background, which contains a significant fraction of hadrons misidentified as muons. The $B \to h^+h'^-$ decays are efficiently rejected as discussed in detail in Sec. VI B.

A calibration for the $dE/dx$ measurement is applied to ensure stability over the tracker volume and operational conditions. This calibration corrects for effects dependent on instantaneous luminosity, local density of tracks, and kinematic information. We make the requirement $\ln\left(\frac{dE/dx_{\rm o}}{dE/dx_{\rm e}}\right) > -0.83$, where ln is the natural logarithm, $dE/dx_{\rm o}$ is the observed $dE/dx$ after the calibration has been applied, and $dE/dx_{\rm e}$ is the expected $dE/dx$ estimated using the observed particle's momentum and the muon mass hypothesis. This requirement is nearly 100% efficient for muons, while rejecting about 50% of kaons. Estimations of the signal efficiency and background rejection for the choice of likelihood and $dE/dx$ selection criteria are made using $J/\psi \to \mu^+\mu^-$ events compared to combinatorial background from the dimuon-mass sidebands and kaons from $B^+ \to J/\psi\ K^+$ decays.

Muon pairs are required to have an invariant mass in the range $4.669 < M_{\mu^+\mu^-} < 5.969$ GeV/$c^2$. Reconstructed $B$-meson candidates must also have $|\eta| < 1$ and $p_T > 4.0$ GeV/$c$. We reconstruct the $p\overline{p}$ interaction point for each event by refitting track-helix parameters of all charged particles, after excluding the muon candidates, that have $z_0$ within 1 cm of the dimuon average $z_0$ and have $p_T > 0.5$ GeV/$c$ to a common space point (vertex). If the fit fails, the primary vertex is determined by the beam line position, estimated using COT+SVXII information, at the average $z_0$ of the dimuon pair.

In addition to the above requirements, we fit the two muon tracks to a common secondary vertex. Several demands on secondary-vertex-related variables are made. We define a three-dimensional displacement length $L_{3D} = \vec{p}(B) \cdot \vec{x}_B/|\vec{p}(B)|$, where $\vec{p}(B)$ is the $B$-candidate-momentum vector estimated as the vector sum of the muon momenta and $\vec{x}_B$ is the secondary-vertex position vector determined relative to the primary-vertex position. We estimate a proper decay time $t = L_{3D}M_{\mu^+\mu^-}/|\vec{p}(B)|$, where $M_{\mu^+\mu^-}$ is the dimuon invariant mass, which in turn is used to define the proper decay-length $\lambda = ct$. The baseline requirements demand that the measured proper decay-length of the $B$ candidate, with its uncertainty, $\sigma_\lambda$, satisfy $0 < \lambda < 0.3$ cm and $\lambda/\sigma_\lambda > 2$; the secondary vertex fit $\chi^2/N_{\rm dof}$, where $N_{\rm dof}$ is the number of degrees of freedom, must be less than 15; the three-dimensional displacement length and its uncertainty $\sigma_{L_{3D}}$ satisfy $L_{3D} < 1.0$ cm and $\sigma_{L_{3D}} < 0.015$ cm; the three-dimensional opening angle between $\vec{x}_B$ and $\vec{p}(B)$, $\Delta\Omega$, satisfy $\Delta\Omega < 0.7$ rad; and the $B$-candidate track-isolation, $I$, satisfy $I > 0.50$. The isolation is defined as $I = p_T(B)/(p_T(B) + \sum_i p_T(i))$, where the sum goes over all charged particles with $p_T > 500$ MeV/$c$ and within an $\eta\phi$ cone centered around the $B$-meson momentum with radius $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 1.0$. The trigger and baseline requirements result in a total of 60 842 CC and 64 495 CF muon pairs, shown in Fig. 3.

For the final selection, we define search regions around the known $B_s^0$ and $B^0$ masses [36]. These regions correspond to approximately $\pm 2.5\sigma_m$, where $\sigma_m \approx 24$ MeV/$c^2$ is the two-track mass resolution estimated from $B_s^0 \to \mu^+\mu^-$ MC events satisfying the trigger and baseline requirements. The sideband regions $5.009 < M_{\mu^+\mu^-} < 5.169$ GeV/$c^2$ and $5.469 < M_{\mu^+\mu^-} < 5.969$ GeV/$c^2$ are used to estimate combinatorial backgrounds. Dimuon candidates with mass smaller than $M_{\mu^+\mu^-} = 5.009$ GeV/$c^2$ are not used for background estimations due to $b \to \mu^+\mu^- X$ [37] contributions but are used for the NN training. It was verified that inclusion of these candidates does not significantly affect the discriminating capabilities of the NN. Backgrounds from $B \to h^+h'^-$ decays, which peak in the signal mass region, are estimated separately.

### B. Normalization decay mode

A sample of $B^+ \to J/\psi\ K^+$ events serves as a normalization decay mode. The $B^+ \to J/\psi\ K^+$ sample is collected using the same dimuon triggers and selection requirements as used for the signal sample so that common systematic uncertainties are suppressed. The kaon

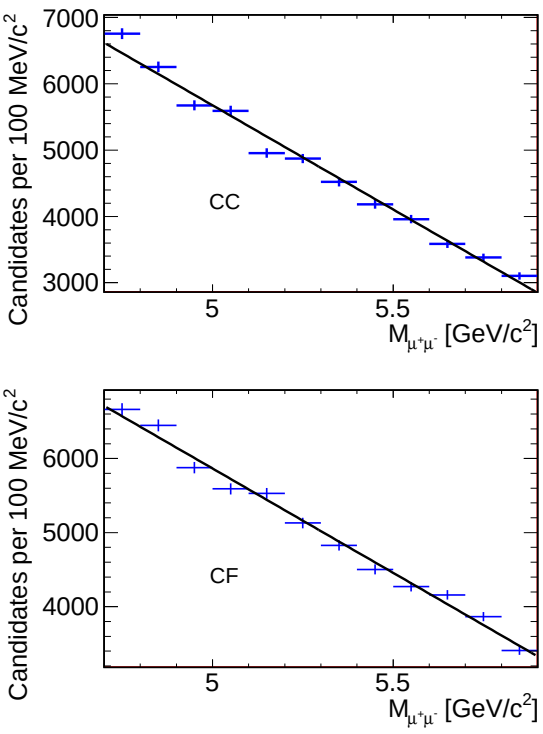


FIG. 3: Dimuon invariant mass distribution for events satisfying the baseline requirements for the $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ search sample with linear fit overlaid for the CC (top) and CF (bottom) channels.

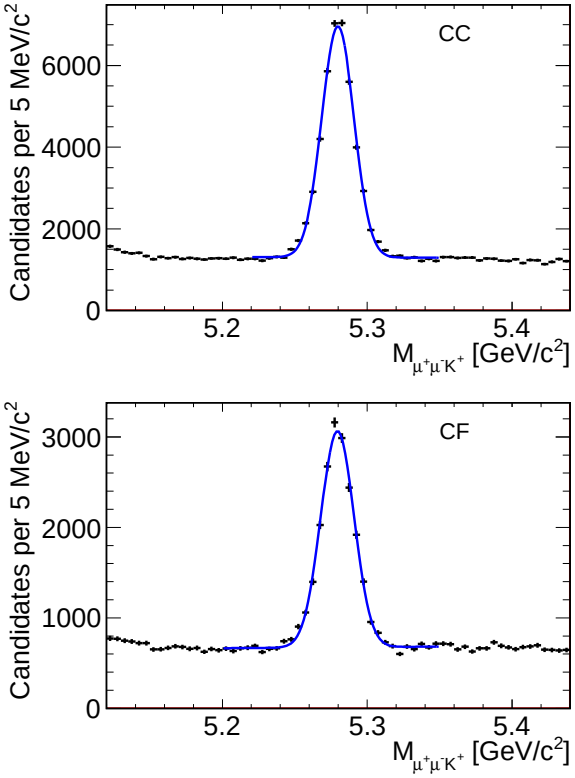


FIG. 4: The $\mu^+\mu^-K^+$ invariant mass distribution for events satisfying the requirements for the $B^+ \to J/\psi\ K^+$ sample with fit overlaid for the CC (top) and CF (bottom) channels.

candidate must satisfy the same COT and L00+SVXII requirements as the muon candidates and must have $p_T > 1$ GeV/$c$, a regime for which the COT tracking efficiency is well understood. For each kaon candidate the $\mu^+$, $\mu^-$, and $K^+$ tracks are constrained to originate from a common vertex in three dimensions. The $\chi^2$ probability of the vertex fit is required to be greater than $10^{-5}$. Additionally, the dimuon invariant mass is required to be consistent with the world average $J/\psi$ mass ($3.017 < M_{\mu^+\mu^-} < 3.177$ GeV/$c^2$). The $\mu^+\mu^-K^+$ mass distribution of candidates satisfying these criteria is shown in Fig. 4. A signal mass region, $|M_{\mu^+\mu^-K^+} - M_{B^+}| < 35$ MeV/$c^2$, is used together with mass sidebands to estimate a $B^+ \to J/\psi\ K^+$ yield of $28\,081 \pm 219$ and $12\,144 \pm 153$ in the CC and CF channels, respectively, using simple sideband subtraction. These yields include a 0.14% correction for $B^+ \to J/\psi\,\pi^+$ contributions. The correction factor is determined by comparing the relative geometric acceptance, reconstruction efficiency, mass window efficiency, and branching fraction of $B^+ \to J/\psi\ K^+$ and $B^+ \to J/\psi\,\pi^+$ decays. The uncertainties on the $B^+ \to J/\psi\ K^+$ yields are due to the limited size of the $\mu^+\mu^-K^+$ sample. The shape of the $\mu^+\mu^-K^+$ mass distribution is parametrized using a Gaussian summed with a first-order polynomial and a fit to the data yields a Gaussian mean of $5280 \pm 11$ MeV/$c^2$ and $5274 \pm 11$ MeV/$c^2$ for CC and CF, respectively, consistent with the world average $B^0$ mass.

### C. The neural network discriminator

We search for $B_s^0 \to \mu^+\mu^-$ ($B^0 \to \mu^+\mu^-$) decays in a narrow 120 MeV/$c^2$ mass window centered around the world average $B_s^0$ ($B^0$) mass. After application of the baseline selection criteria the data sample is dominated by combinatorial background and has a signal-to-background ratio of approximately $10^{-4}$. We significantly improve the discrimination power between signal and combinatoric background by combining kinematic, isolation-related, and lifetime-related variables using a neural-network classifier. A key feature of the NN is that it is designed to be independent of the dimuon mass. This allows an estimation of the combinatorial background in the signal mass region by interpolation from mass sidebands after the NN has been applied to the data.

Fourteen variables describing the measured kinematics-, isolation-, and lifetime-related properties of the signal are selected to construct a NN discriminant $\nu_N$. These variables are selected based on a study of the physics characteristics of the signal and background using training samples described below. Variables that are poorly modeled by the MC or

correlated to dimuon mass are excluded from the NN. The signal and background samples are separated into CC and CF data sets that are exclusively independent. Trainings are performed separately using the same input variables for the CC and CF data sets.

We chose the NEUROBAYES NN package [38, 39] for the construction of the NN. Using this multivariate analysis technique, we achieve a background rejection of 99.9%, while maintaining a signal efficiency of 40%.

#### 1. Background training sample

We define two dimuon mass sidebands, $4.669 < M_{\mu^+\mu^-} < 5.169$ GeV/$c^2$ (lower SB) and $5.469 < M_{\mu^+\mu^-} < 5.969$ GeV/$c^2$ (upper SB), which are used to construct a combinatorial background sample used as one of the NN training samples. Although this analysis is based on 10 fb$^{-1}$, the NN optimization was done, *a priori*, during the previously published analysis based on 7 fb$^{-1}$ [18]. A total of 36 329 CC and 39 657 CF mass-sideband events survive the baseline requirements in 7 fb$^{-1}$ of integrated luminosity and are used as the background sample for the NN training and testing.

#### 2. Signal training sample

The simulated $B_s^0 \to \mu^+\mu^-$ signal sample used for the NN training is described in Sect. III. The dimuon mass range used for this signal sample corresponds to the search region defined in Sec. IV A. A sub-sample of MC signal events equal in size to the sum of the background samples is randomly chosen for NN training while the full MC sample is used to estimate the efficiency of the neural network.

#### 3. Input parameters

Candidates from combinatorial backgrounds, when compared to signal events, tend to have smaller dimuon mass, shorter proper-decay-lengths, a softer $p_T$ spectrum, and a higher density of tracks near the $B_s^0$ candidate. We investigated a variety of kinematic, isolation, and lifetime variables as inputs to the NN. After removing variables that were poorly modeled or were found to cause a correlation between $\nu_N$ and the dimuon mass, an initial set of 20 discriminating variables remained. Multiple neural networks were trained varying the number and combination of input parameters employed. The performances of the trained networks were compared in the plane of combinatorial-background acceptance versus signal acceptance. A NN using 14 input variables was found to offer excellent background discrimination and is employed as the final discriminant in this analysis. The 14 variables employed are listed here in order of descending discrimination power between signal and background:

**$\Delta\Omega$:** Three-dimensional angle between the $B_s^0$ momentum and the vector pointing from the primary to secondary vertex.

***I*:** Isolation of the $B_s^0$ candidate as defined in Sec. IV A.

**Larger $d_0(\mu)$:** For the muon pair, the impact parameter of the muon with the larger value.

**$d_0(B_s^0)$:** Impact parameter of $B_s^0$ candidate with respect to the primary event vertex.

**$L_T/\sigma_{L_T}$:** Significance of the transverse decay length $L_T$, where $L_T = \vec{p_T}(B) \cdot \vec{x}_T/|\vec{p_T}(B)|$ and $\vec{x}_T$ is the secondary-vertex-position vector relative to the primary-vertex position in the plane transverse to the beam line.

**$\chi^2$:** $\chi^2$ per degree of freedom of the secondary-vertex fit.

**$L_{3D}$:** Three-dimensional vertex displacement as defined in Sec. IV A.

**Lower $p_T(\mu)$:** For the muon pair, the transverse momentum of the muon with the lower value.

**Significance of smaller $d_0(\mu)$:** $d_0(\mu)/\sigma_{d_0(\mu)}$ of the muon with smaller impact parameter, where $\sigma_{d_0(\mu)}$ is the estimated uncertainty of $d_0(\mu)$.

**$\lambda/\sigma_\lambda$:** Significance of $\lambda$.

**Smaller $|d_0(\mu)|$:** For the muon pair, the impact parameter of the muon with the smaller value.

**$\lambda$:** Three-dimensional proper decay length defined in Sec. IV A.

**$\Delta\Omega_T$:** Angle between the $B_s^0$ momentum and the vector pointing from the primary to secondary vertex in the plane transverse to the beam line.

**Significance of larger $d_0(\mu)$:** $d_0(\mu)/\sigma_{d_0(\mu)}$ of the muon with larger impact parameter.

When available, the silicon $rz$ tracking information strongly discriminates against combinatorial and partially reconstructed backgrounds. Tracks from combinatorial background are less likely to originate from a common vertex in three dimensions and are suppressed by the vertex-fit quality information used in the multivariate discriminate. Similarly, precise three-dimensional reconstruction of the primary and secondary vertices allows the comparison of the $B$-candidate flight direction and the secondary-vertex vector, which rejects both combinatorial and partially-reconstructed background. Comparisons of the distributions of background and signal samples for the input variables with the greatest discriminating power are shown in Fig. 5.

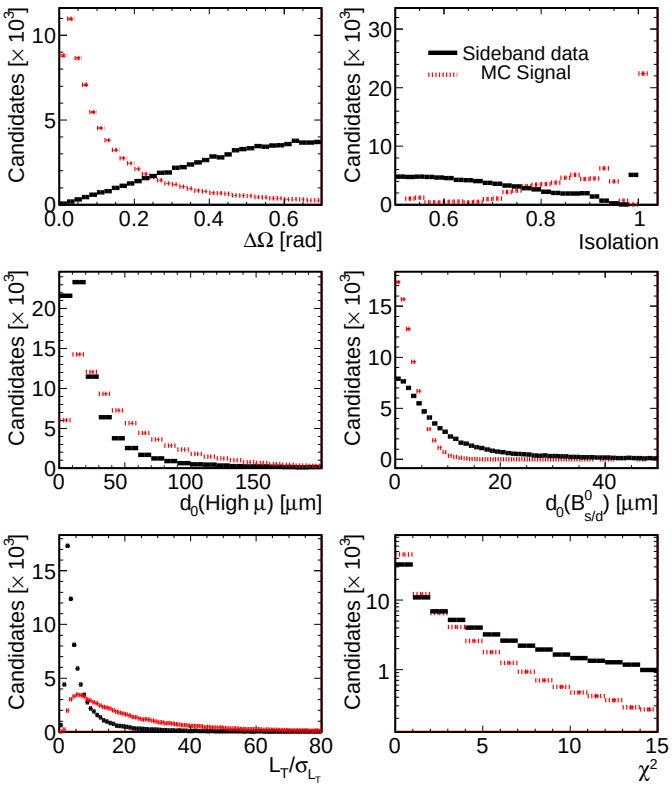


FIG. 5: Comparison of the combinatorial background distribution (solid), taken from dimuon mass sideband events in the first 7 fb$^{-1}$ of data, to the signal distribution (dashed), taken from MC events, for the six most discriminating of the NN input variables.

#### 4. *Discriminant output*

When training the NN, 80% of each training sample is used for the actual training, while the remaining 20% is used for validation and over-training tests. The trained NN takes the input parameters for every event and returns an output value $\nu_N$ in the range [0 (background-like), 1 (signal-like)]. The combined CC and CF $\nu_N$ distribution is shown in Fig. 6 for background and signal separately. For various $\nu_N$ requirements, the resulting signal and background efficiencies are given in Table I.

### D. Neural network consistency checks

We perform several consistency checks of the NN, including tests for $M_{\mu^+\mu^-}$-$\nu_N$ correlations, tests of over-training, and studies of MC mis-modeling of the $\nu_N$ signal distribution.

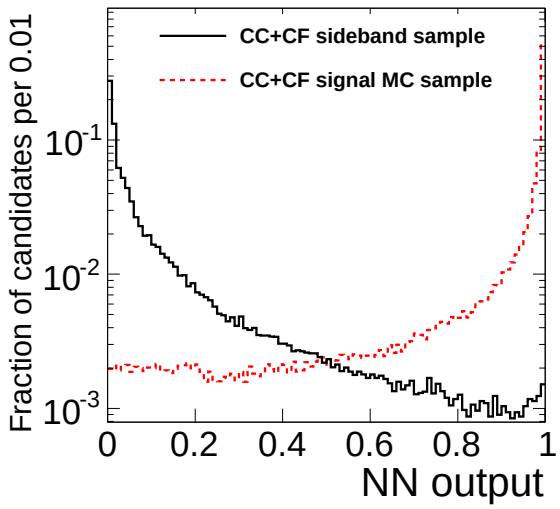


FIG. 6: Distributions of $\nu_N$ for signal and background samples. The background sample consists of dimuon mass sideband events from the first 7 fb$^{-1}$ of data.

#### 1. *Check for correlations between $\nu_N$ and dimuon mass*

We estimate the dominant combinatorial background in the dimuon mass signal region by linear interpolation from the dimuon mass sideband region. An unbiased estimate of the resulting combinatorial background requires that $\nu_N$ is independent of the dimuon mass. We perform several studies of the $M_{\mu^+\mu^-}$-$\nu_N$ dependence. For example, we divide the sideband sample into an "inner" region and an "outer" region according to dimuon mass. The inner region is defined as $5.002 < M_{\mu^+\mu^-} < 5.169$ GeV/$c^2$ and $5.496 < M_{\mu^+\mu^-} < 5.636$ GeV/$c^2$, and is used as a "signal" sample. The outer sample is formed by the remaining events in the sideband, and is treated as a "background" sample. Using these "inner signal" and "outer background" samples, the NN training is repeated with the same set of 14 input variables listed in Sec. IV C 3 and the resulting $\nu_N$ distributions are compared. This check is based on the observation that event properties are nearly identical for events in the inner and outer regions and thus only differ by dimuon mass. The resulting $\nu_N$ distributions for the inner and outer samples are compared in Fig. 7 for the CC and CF channels separately. No significant difference in the $\nu_N$ distribution is found, indicating that $\nu_N$ is independent of dimuon mass. This is strong evidence that the NN cannot use the 14 input variables to infer anything about the dimuon mass and that $M_{\mu^+\mu^-}$ and $\nu_N$ are uncorrelated.

As a further investigation of $M_{\mu^+\mu^-}$-$\nu_N$ dependencies, we study the average value of $\nu_N$ as a function of the dimuon mass. We use both the signal-sideband data, with two oppositely-charged muons and $\lambda > 0$

TABLE I: Efficiency for various $\nu_N$ requirements, prior to MC weighting, signal (Sig.) and background (Bgd.) events for the CC and CF trainings separately. Uncertainties include only the statistical component.

| | CC training | | CF training | |
|---|---|---|---|---|
| $< \nu_N$ | Sig. (%) | Bgd (%) | Sig. (%) | Bgd. (%) |
| 0.999 | $17.72 \pm 0.25$ | $0.01 \pm 0.01$ | $11.18 \pm 0.21$ | $0.01 \pm 0.01$ |
| 0.998 | $29.13 \pm 0.33$ | $0.03 \pm 0.01$ | $24.71 \pm 0.33$ | $0.02 \pm 0.01$ |
| 0.997 | $34.34 \pm 0.37$ | $0.05 \pm 0.01$ | $30.97 \pm 0.38$ | $0.03 \pm 0.01$ |
| 0.996 | $38.69 \pm 0.39$ | $0.07 \pm 0.01$ | $37.37 \pm 0.43$ | $0.05 \pm 0.01$ |
| 0.995 | $42.18 \pm 0.42$ | $0.08 \pm 0.02$ | $42.34 \pm 0.47$ | $0.07 \pm 0.01$ |
| 0.994 | $44.79 \pm 0.43$ | $0.10 \pm 0.02$ | $44.54 \pm 0.48$ | $0.08 \pm 0.01$ |
| 0.993 | $46.77 \pm 0.45$ | $0.12 \pm 0.02$ | $45.74 \pm 0.49$ | $0.10 \pm 0.02$ |
| 0.992 | $48.38 \pm 0.46$ | $0.13 \pm 0.02$ | $46.60 \pm 0.50$ | $0.11 \pm 0.02$ |
| 0.991 | $49.99 \pm 0.47$ | $0.14 \pm 0.02$ | $47.57 \pm 0.50$ | $0.11 \pm 0.02$ |
| 0.990 | $51.50 \pm 0.48$ | $0.16 \pm 0.02$ | $48.47 \pm 0.51$ | $0.12 \pm 0.02$ |
| 0.980 | $60.87 \pm 0.53$ | $0.32 \pm 0.03$ | $57.59 \pm 0.57$ | $0.25 \pm 0.03$ |
| 0.970 | $64.29 \pm 0.55$ | $0.44 \pm 0.03$ | $62.14 \pm 0.60$ | $0.37 \pm 0.03$ |
| 0.960 | $66.64 \pm 0.57$ | $0.52 \pm 0.04$ | $64.90 \pm 0.62$ | $0.48 \pm 0.03$ |
| 0.950 | $68.91 \pm 0.58$ | $0.63 \pm 0.04$ | $67.17 \pm 0.64$ | $0.55 \pm 0.04$ |

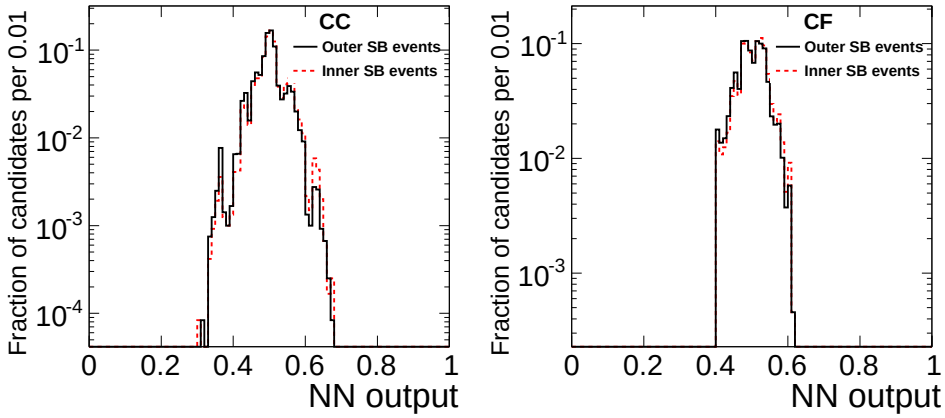


FIG. 7: Comparison of the $\nu_N$ distribution of inner sideband events to outer sideband events using the full 10 fb$^{-1}$ of data and a custom-trained NN as described in Sec. IV D 1. The CC and CF channels are shown separately.

(OS+), and a control sample consisting of events with two oppositely-charged muons but with proper-decay length $\lambda < 0$ (OS−). The OS− control sample is discussed further in Sec. VI C and is dominated by prompt combinatorial background. The resulting distributions, shown in Fig. 8, are consistent with a flat line, again indicating that $M_{\mu^+\mu^-}$ and $\nu_N$ are independent.

To check for correlations of $\nu_N$ with small variations of mass – of the order of the mass difference between $B^0$ and $B_s^0$– a separate $B^0$ MC sample is produced. The resulting $\nu_N$ distribution is compared with that obtained from the $B_s^0$ MC sample in Fig. 9. No significant difference between the $B^0$ and $B_s^0$ $\nu_N$ distributions is found, again indicating the $M_{\mu^+\mu^-}$ and $\nu_N$ are independent.

We conclude that the choice of discriminating variables yields a NN classifier with excellent signal-to-background discrimination, while remaining independent of $M_{\mu^+\mu^-}$ and leaving the shape of the dimuon mass distribution unchanged.

### 2. *Check for NN over-training*

A portion of the background and signal training samples is set aside for internal validation tests. These tests include metrics provided by the NEUROBAYES package that are sensitive to over-training. All the metrics show that no over-training occurred for any of the neural nets used in this analysis. As a further test of possible over-training, we repeat the NN optimization using 33% and 50% of the input background sample in the training. The resulting $\nu_N$ distribution for sideband events is compared among these two trainings and the default training in Fig. 10 and no significant differences are observed. We

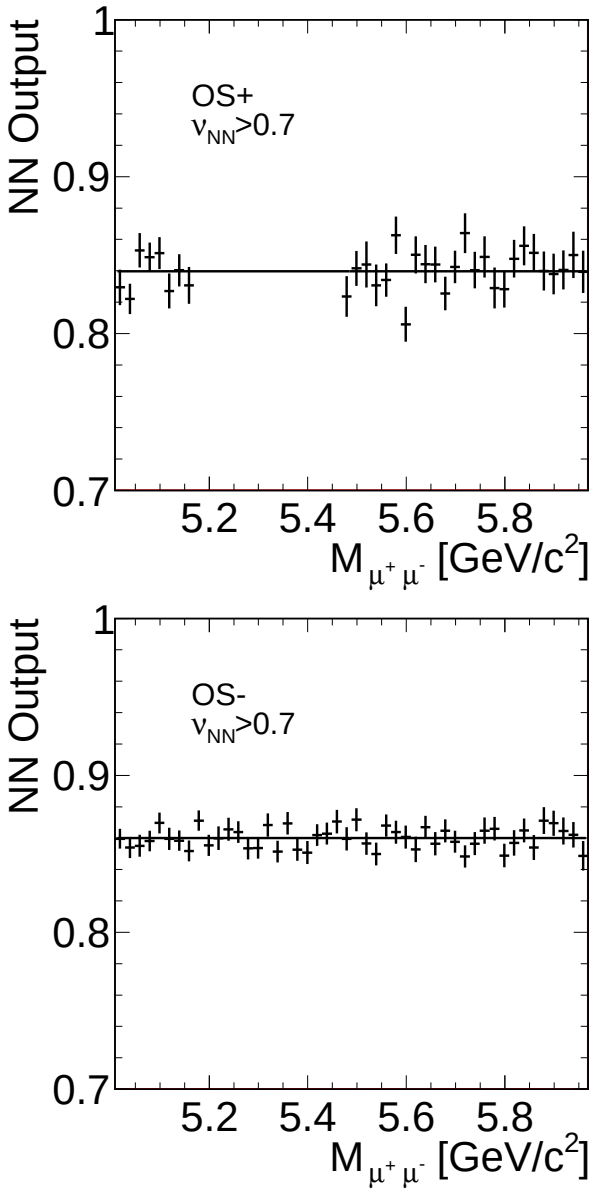


FIG. 8: Result of the $M_{\mu^+\mu^-}$-$\nu_N$ correlation check using the first 7 fb$^{-1}$ of data. Top: correlation between average $\nu_N$ and dimuon mass in the OS+ sample for which we keep the $B_s^0 \to \mu^+\mu^-$ + $B^0 \to \mu^+\mu^-$ signal region blinded. Bottom: correlation between average $\nu_N$ and dimuon mass in background-dominated OS− sample.

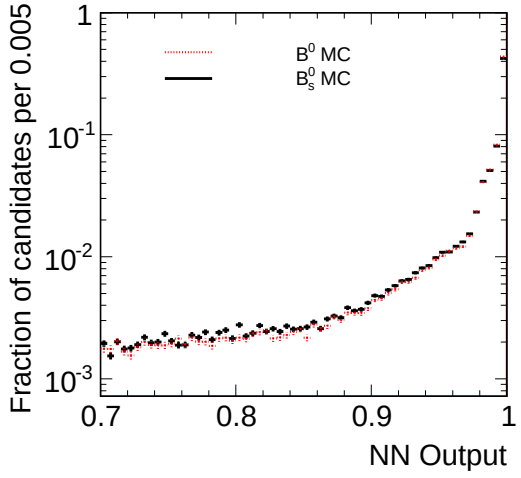


FIG. 9: Comparison of the $\nu_N$ distribution for simulated samples of $B_s^0$ and $B^0$ signal events. The distributions have been normalized to the same area over the entire NN output range, $0 < \nu_N < 1$.

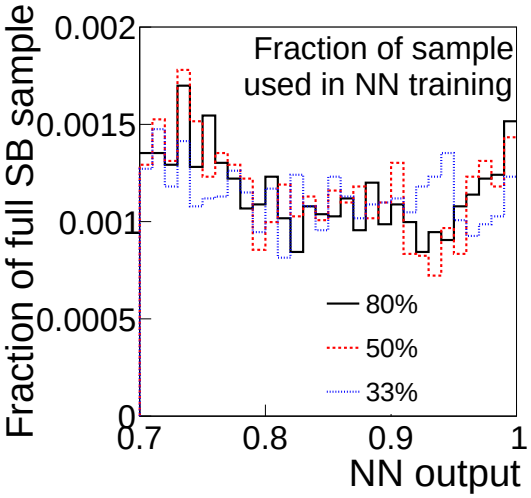


FIG. 10: The $\nu_N$ distributions for networkks trained with 33%, 50%, and 80% of the input background sample taken from the dimuon mass sideband events reconstructed in 7 fb$^{-1}$ of data.

conclude there is no evidence of NN over-training.

### 3. *Validation of the discriminant distributions*

We check the MC modeling of the 14 input variables and additional kinematic and lifetime variables using the sample of $B^+ \to J/\psi K^+$ events. Distributions from $B^+ \to J/\psi\ K^+$ data and MC are compared in Figs. 11 and 12; in order to better mimic the resolutions relevant for $B^0 \to \mu^+\mu^-$ decays, the vertex variables use only the two muons from the $J/\psi$ of the $B^+ \to J/\psi\ K^+$ decay while the $B$-hadron momentum variables, $p(B)$ and $p_T(B)$, and isolation variables use the three-track information. The $\nu_N$ distribution obtained from the $B^+ \to J/\psi\ K^+$ MC is compared to that obtained using the sideband-subtracted $B^+ \to J/\psi\ K^+$ data in Figures 13 and 14 for the CC and CF channels, respectively. The small discrepancies observed are used to assign sys-

tematic uncertainties as discussed in Sec. V D.

## V. DETERMINING THE $B_s^0 \to \mu^+\mu^-$ BRANCHING FRACTION

The branching fraction $\mathcal{B}(B_s^0(B^0) \to \mu^+\mu^-)$ is determined using Eq. (1), where $\alpha_{B_s^0(B^0)}$ is the geometrical and kinematic acceptance of the triggers employed to collect the dimuon data set; $\epsilon^{\rm trig}_{B_s^0(B^0)}$ is the trigger efficiency for $B_s^0(B^0)$ hadrons decaying to $\mu^+\mu^-$ within the acceptance; $\epsilon^{\rm reco}_{B_s^0(B^0)}$ is the efficiency of the reconstruction, baseline, and mass requirements for $\mu^+\mu^-$ pairs satisfying the trigger requirements; and $\epsilon^{\rm NN}_{B_s^0(B^0)}$ is the efficiency of the NN selection for events satisfying the trigger and baseline requirements and a given set of $\nu_N$ requirements. The equivalent efficiencies and acceptance for the normalization mode are indicated with the $B^+$ subscript. No $\epsilon^{\rm NN}_{B^+}$ term appears since the NN is not applied to the normalization mode. The $B_s^0(B^0) \to \mu^+\mu^-$ and $B^+ \to J/\psi\ K^+$ acceptance and efficiencies are estimated separately to account for kinematic differences arising from the differing $B$-hadron decays. The $b$-quark fragmentation-fraction-ratio $f_u/f_s$, the relevant product branching fractions for the normalization mode $\mathcal{B}(B^+ \to J/\psi\ K^+ \to \mu^+\mu^- K^+)$, and their uncertainties are taken from Ref. [36]. For the $B^0$ search, $f_s$ is replaced by $f_d$ and the fragmentation ratio is set to unity. Due to their differing sensitivities, the CC and CF channels are treated separately and then combined to yield the final result. Normalizing the observed signal rate to the rate of an abundant, well-known, and kinematically-similar decay results in a significant reduction to the total uncertainty since systematic effects largely cancel in the acceptance and efficiency ratios of Eq. (1).

$$\mathcal{B}(B_s^0(B^0) \to \mu^+\mu^-) = \frac{N_{B_s^0(B^0)}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B_s^0(B^0)}} \cdot \frac{\epsilon^{\rm trig}_{B^+}}{\epsilon^{\rm trig}_{B_s^0(B^0)}} \cdot \frac{\epsilon^{\rm reco}_{B^+}}{\epsilon^{\rm reco}_{B_s^0(B^0)}} \cdot \frac{1}{\epsilon^{\rm NN}_{B_s^0(B^0)}} \cdot \frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\psi\ K^+ \to \mu^+\mu^- K^+) \quad (1)$$

We define the single-event-sensitivity (SES) as the branching fraction determined from Eq. (1) when setting $N_{B_s^0(B^0)} = 1$. The SES approximates the smallest signal branching-fraction to which the analysis is sensitive. The methods used to estimate the inputs to Eq. (1) are described below and the results are summarized in Table II. Combining the CC and CF channels yields a SES for the $B_s^0 \to \mu^+\mu^-$ search of $1.4 \times 10^{-9}$ with an 18% total uncertainty. The combined SES for the $B^0 \to \mu^+\mu^-$ search is a factor of 3.5 smaller with a reduced uncertainty of about 12% since the fragmentation ratio $f_d/f_u$ is not relevant.

### A. Acceptance

The acceptances are determined using $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\ K^+$ MC simulations for $B_s^0$ and $B^+$ mesons that satisfy $|y| < 1.0$ and $p_T(B_s^0) > 4$ GeV/$c$, where $y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)$ is the rapidity. Both muons must satisfy the fiducial and kinematic requirements of the trigger discussed in Sec. II C. Muons are required to have $p_T > 2.0$ GeV/$c$ if detected in the CMU, and $p_T > 2.2$ GeV/$c$ if detected in the CMX and their trajectories must extrapolate to the active fiducial-volumes of the muon, COT, and L00+SVXII systems. The kaon in the normalization channel must have $p_T > 1.0$ GeV/$c$ and its trajectory must extrapolate to the active fiducial-volumes of the COT and L00+SVXII systems. Effects from COT-track reconstruction, multiple scattering, and stub-track matching are included in the reconstruction efficiencies discussed below. Systematic uncertainties on the acceptances are assessed by varying the $b$-quark mass, fragmentation modeling, and renormalization and factorization scales in the MC generation by one standard-deviation-uncertainty and quantifying the resulting change in the acceptance ratio. Additionally, we assign variations of the acceptance due to changes in the size of the $p\overline{p}$ luminous region as a systematic uncertainty. The observed differences are added in quadrature. The final acceptance ratio including statistical and systematic uncertainties is given in Table II for the CC and CF channels separately.

### B. Trigger efficiencies

The trigger efficiency is measured separately for the L1, L2, and L3 dimuon triggers using data control samples collected using unbiased trigger selections identifying a high-purity sample of dimuon events that satisfy the criteria of the trigger under study. The efficiency is determined from the fraction of events in this sample for which also the trigger under study fires. Dimuon data collected requiring only trigger L1 are used to measure the L2 and L3 efficiencies, which exceed 99%. The L1 dimuon trigger efficiency is determined as the product of the L1 efficiency for each muon separately. The single-muon L1 efficiencies are measured using a tag-and-probe method on samples of high-purity $J/\psi \to \mu^+\mu^-$ events

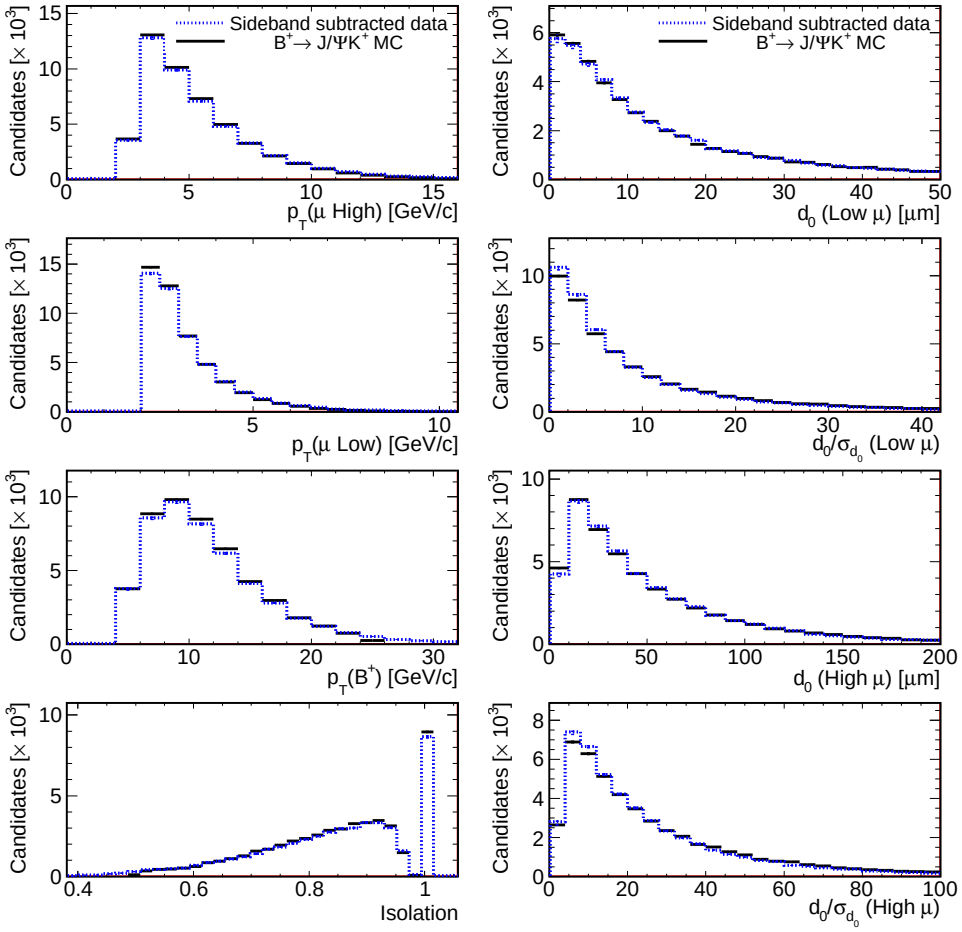


FIG. 11: A comparison of $B^+ \to J/\psi\ K^+$ sideband-subtracted data to simulated $B^+ \to J/\psi\ K^+$ events for a variety of kinematic and lifetime-related distributions, including some used as input to the neural net.

collected with a trigger L1 that requires only one muon. The muon firing the L1 trigger used to collect the $J/\psi$ control sample is identified as the "tag", while the second muon from the $J/\psi$ decay is unbiased with respect to the trigger whose efficiency we are studying and can be used a "probe". Dimuon mass-sideband subtraction is used to remove effects of the small background present in the sample. The single-muon L1 efficiencies are parametrized as a function of the date the data were recorded and the track $p_T$, $|\eta|$, and $\phi$ for CMU and CMX muons separately. The parametrization by date accounts for significant changes in detector operating conditions arising from variations in trigger configuration, COT performance, or Tevatron beam parameters. The parametrization in $p_T$ describes the rapidly changing L1 efficiency near the trigger $p_T$ threshold. The parametrization in $|\eta|$ describes changes in the L1 efficiency due to the increased ionization path-lengths of tracks traversing the COT at large $|\eta|$, which may increase the probability for the corresponding hits to exceed the noise threshold and fire the trigger. The parametrization in $\phi$ is primarily important for a small amount (0.2 fb$^{-1}$) of early data for which the COT gain was temporarily degraded in the bottom portion of the chamber [40]. The resulting single-muon L1 efficiency is about 96% for the muons relevant for this analysis and plateaus at about 99% for muons with $p_T > 5$ GeV/$c$. The dimuon L1 efficiency is estimated by convolution of the single-muon efficiencies with the ($p_T^{\mu^+}$, $|\eta^{\mu^+}|$, $\phi^{\mu^+}$, $p_T^{\mu^-}$, $|\eta^{\mu^-}|$, $\phi^{\mu^-}$) distribution obtained from $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\ K^+$ MC for events in the geometrical and kinematic acceptance of the trigger. This convolution yields a dimuon L1 efficiency of about 93% for both the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\ K^+$ samples in each of the CC and CF channels. The total systematic uncertainty for the trigger L1 efficiency is in the 1–2% range and is dominated by variations of the single-muon efficiency as a function of the isolation of the muon and by differences between muons detected in the $\eta > 0$ and $\eta < 0$ volumes of the detector. Double-muon correlations are studied and found to be negligible. When propagating the single-muon uncertainties to the dimuon efficiency, the uncertainties are taken to be 100% correlated between the two muons.

The total trigger efficiency is determined as the prod-

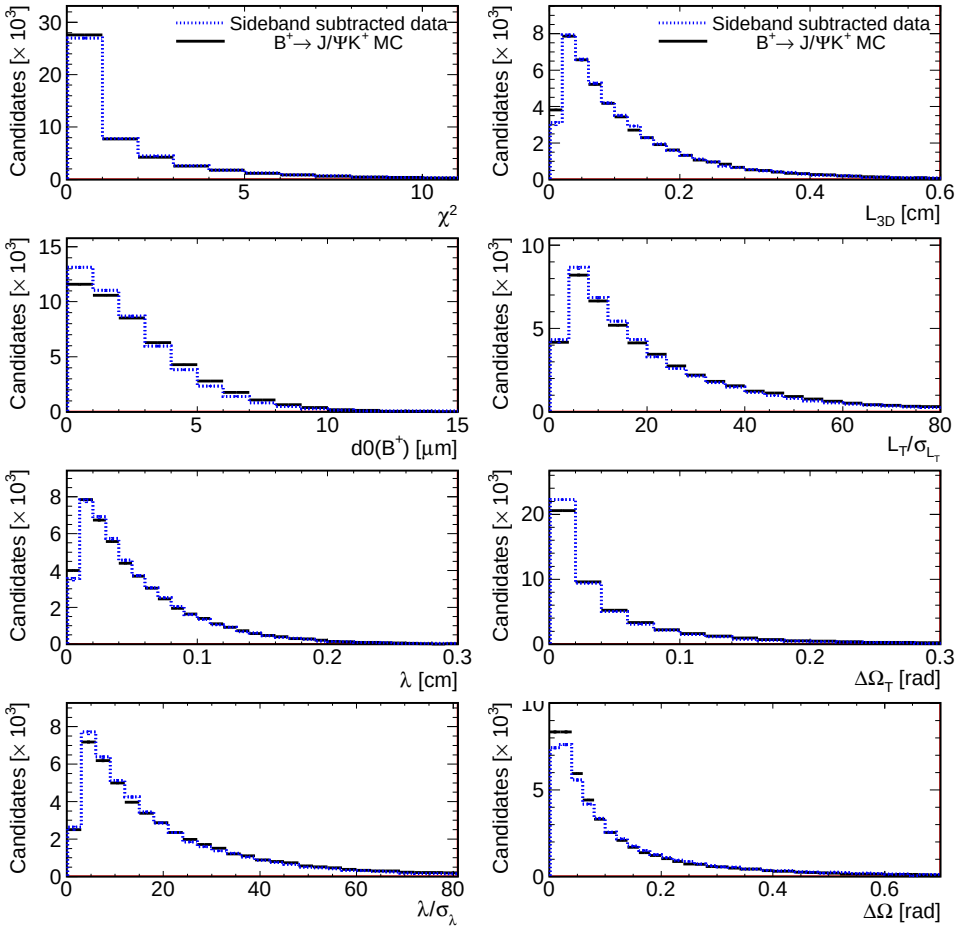


FIG. 12: A comparison of $B^+ \to J/\psi\ K^+$ sideband-subtracted data to simulated $B^+ \to J/\psi\ K^+$ event for several variables used as input to the neural net.

uct of the L1, L2, and L3 trigger efficiencies for the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\ K^+$ modes separately. The total uncertainty of the efficiency ratio is much less than 0.01 and is estimated by treating the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\ K^+$ total uncertainties as 100% correlated. The final trigger efficiency ratios for the CC and CF channels and their associated uncertainties are shown in Table II.

### C. Reconstruction efficiencies

The total reconstruction efficiency is factorized into the efficiency for COT-track reconstruction, muon-stub reconstruction, association of L00+SVXII hits to the COT track, the dimuon-vertex reconstruction, and the dimuon-mass reconstruction efficiencies. For the normalization mode, the kaon COT, the kaon silicon, and the $B^+$ vertex reconstruction efficiencies are also evaluated.

#### 1. *COT track reconstruction efficiency*

The probability for identifying a single track in the COT is evaluated by embedding MC tracks into real data events and measuring the fraction of those tracks that are successfully reconstructed and satisfy the CDF standard-track-quality requirements. Hits from MC charged-particle tracks are inserted into events from beam data. In readout channels where MC hits overlap with hits in the event record, the hits are merged. The MC is tuned so that quantities such as deposited charge per hit and single-hit position resolution match those observed in the data. The dependencies of these quantities on various parameters, such as the track $\eta$, the electric charge of the track, and the track isolation, are also accurately reproduced in the MC. For charged particles fully fiducial to the COT and with $p_T > 1.5$ GeV/$c$, the COT reconstruction is consistent with being fully efficient. Systematic uncertainties include variations of the MC parameters affecting the COT hit distributions and account for small ($< 1\%$) observed variations as a function of track isolation, $p_T$, and $\eta$. The dimuon efficiency is taken to be

TABLE II: A summary of the inputs used in Eq. (1) to determine $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in the CC and CF channels separately. The relative uncertainties are given parenthetically. The uncertainties for the trigger efficiency ratios are significantly smaller than 1% and are denoted by 0.00. The single-event-sensitivities (SES) for $\nu_N > 0.70$ are given in the last row. Combining CC and CF gives a SES for the $B_s^0 \to \mu^+\mu^-$ search of $1.4 \times 10^{-9}$ ($\pm$18%). The combined SES for the $B^0 \to \mu^+\mu^-$ search is $3.9 \times 10^{-10}$ ($\pm$12%).

| | CC | | CF | |
|---|---|---|---|---|
| $\alpha_{B^+}/\alpha_{B_s^0}$ | $0.31 \pm 0.02$ | ($\pm$6%) | $0.20 \pm 0.01$ | ($\pm$7%) |
| $\epsilon^{\rm trig}_{B^+}/\epsilon^{\rm trig}_{B_s^0}$ | $1.00 \pm 0.00$ | ( - ) | $0.98 \pm 0.00$ | ( - ) |
| $\epsilon^{\rm reco}_{B^+}/\epsilon^{\rm reco}_{B_s^0}$ | $0.85 \pm 0.06$ | ($\pm$8%) | $0.84 \pm 0.06$ | ($\pm$9%) |
| $\epsilon^{\rm NN}_{B_s^0}(\nu_N > 0.70)$ | $0.92 \pm 0.04$ | ($\pm$4%) | $0.86 \pm 0.04$ | ($\pm$4%) |
| $\epsilon^{\rm NN}_{B_s^0}(\nu_N > 0.995)$ | $0.46 \pm 0.02$ | ($\pm$5%) | $0.47 \pm 0.02$ | ($\pm$8%) |
| $N_{B^+}$ | $28081 \pm 219$ | ($\pm$1%) | $12144 \pm 153$ | ($\pm$1%) |
| $f_u/f_s$ | $3.55 \pm 0.47$ | ($\pm$13%) | $3.55 \pm 0.47$ | ($\pm$13%) |
| $\mathcal{B}(B^+ \to J/\psi\ (\to \mu^+\mu^-)\ K^+)$ | $(6.01 \pm 0.21) \times 10^{-5}$ | ($\pm$4%) | | |
| SES ($\nu_N > 0.70$) | $2.2 \times 10^{-9}$ | ($\pm$18%) | $3.3 \times 10^{-9}$ | ($\pm$18%) |

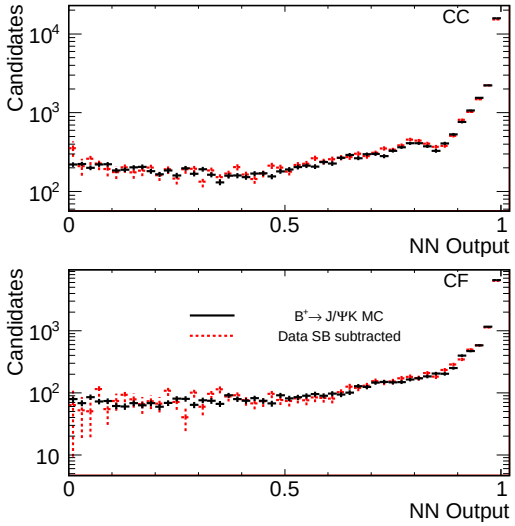


FIG. 13: Comparison of the $\nu_N$ distribution for $B^+ \to J/\psi\ K^+$ sideband-subtracted data to that obtained using simulated $B^+ \to J/\psi\ K^+$ events for the CC and CF channels separately. The MC is normalized to the observed number data events.

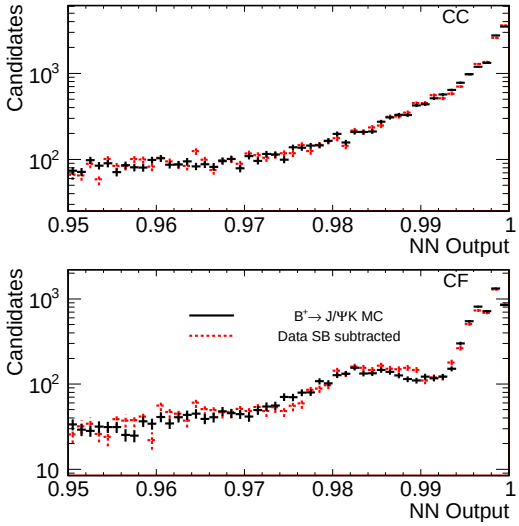


FIG. 14: Comparison of the $\nu_N > 0.95$ distribution for $B^+ \to J/\psi\ K^+$ sideband-subtracted data to that obtained using simulated $B^+ \to J/\psi\ K^+$ events for the CC and CF channels separately. The MC is normalized to the observed number of data events over the range $0 < \nu_N < 1$.

the square of the single-track efficiency and the resulting dimuon efficiency ratio is $1.00 \pm 0.01$ for both the CC and CF channels.

### 2. *Muon stub reconstruction efficiency*

The muon-stub reconstruction efficiencies are determined from $J/\psi \to \mu^+\mu^-$ events using a tag-and-probe method similar to that described in Sec. V B. The "tag" muon candidate is required to have a COT track satisfying the relevant trigger and COT baseline requirements of Sec. IV A matched to a muon stub, while the other muon is only required to pass the COT baseline requirements. The muon-stub reconstruction efficiency is determined using the fraction of events for which the second muon is also matched to a muon stub. The efficiency is estimated both with data and MC samples and the efficiency measured with data is divided by the efficiency measured with MC to account for geometric losses already included in the acceptance estimate. As a consistency check of the efficiency estimates, the same procedure is used to determine the efficiency of high-$p_T$ muons using a sample of high-purity $Z^0 \to \mu^+\mu^-$ decays. The difference in efficiencies is taken as the systematic uncertainty. The ratio of dimuon efficiencies for both the CC and CF is

$1.00 \pm 0.03$.

### 3. Muon ID efficiency

The combined efficiency of the $dE/dx$ and muon-likelihood requirements is determined using a $J/\psi \to \mu^+\mu^-$ sample by comparing the signal yield with and without the application of these muon identification criteria. The single-muon efficiency is determined as a function of $p_T$. The total dimuon efficiency is evaluated by convolution of the $(p_T^{\mu^+},\ p_T^{\mu^-})$ distribution from $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\ K^+$ MC for events within the acceptance and satisfying the trigger, COT, and muon-stub reconstruction requirements. This efficiency is cross-checked *in situ* using $B^+ \to J/\psi\ K^+$ data events. The difference between the efficiency in the $B^+ \to J/\psi\ K^+$ MC events and in the $B^+ \to J/\psi\ K^+$ sideband-subtracted data is assigned as a systematic uncertainty. The final dimuon efficiency ratio for the muon identification requirements is $1.01 \pm 0.03$ for both the CC channel and CF channels.

### 4. L00+SVXII association efficiency

The efficiency of associating L00 and SVXII hits to muon tracks reconstructed in the COT is estimated using $J/\psi \to \mu^+\mu^-$ events in a manner similar to that described in Sec. V C 2. The $J/\psi$ sample includes only muons reconstructed as a track in the COT matched to a muon stub in the CMU or CMX and surviving the muon identification requirements. The efficiency for associating at least three $r\phi$ hits from the L00+SVXII silicon layers is evaluated as the fraction of this sample for which the muon tracks satisfy that criteria. The variation of the efficiency on track $p_T$ , isolation, and azimuthal dimuon opening angle is used to assign systematic uncertainties. The resulting dimuon efficiency ratio is $1.00 \pm 0.03$ for both the CC and CF channels.

### 5. Dimuon vertex efficiency

The efficiency of the dimuon-vertex requirements specified in Sec. IV A is estimated using simulated $B^+ \to J/\psi\ K^+$ and $B_s^0 \to \mu^+\mu^-$ samples. The resulting efficiency is found to be consistent with the efficiency determined using sideband-subtracted $B^+ \to J/\psi\ K^+$ data events, thus verifying the accuracy of the MC modeling. The resulting ratio of efficiencies for the dimuon-vertex requirements is $0.99 \pm 0.01$ for the CC and CF channels.

### 6. Dimuon mass efficiency

The efficiency of the dimuon-mass requirements is estimated using simulated $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi\ K^+$ events surviving the baseline, vertex, and trigger requirements. Comparisons of the mean and width of the invariant mass distribution using data and MC samples of $J/\psi \to \mu^+\mu^-$ and $B^+ \to J/\psi\ (\to \mu^+\mu^-)\ K^+$ events reveal discrepancies at the 10% level for the width, which are used to assign systematic uncertainties. Since the signal-search mass-windows correspond to $\pm 2.5\sigma_m$ and have high efficiency, the systematic uncertainties negligibly affect the efficiency ratio, which is 1.00.

### 7. $B^+ \to J/\psi\ K^+$ reconstruction efficiency

In addition to the dimuon efficiencies discussed above, the total reconstruction efficiency for the $B^+ \to J/\psi\ K^+$ events also includes the efficiency for reconstructing the kaon as a COT track, for associating L00+SVXII hits to the COT track, and for the $\mu^+\mu^-K^+$ vertex requirements. The kaon COT efficiency is estimated using the method of Sec. V C 1 and is $0.964 \pm 0.016$ for both the CC and CF channels. This value is lower than that for muons due to inefficiencies resulting from kaon interaction with matter and due to the lower transverse-momentum threshold ($p_T^K > 1.0$ GeV/$c$) employed for the kaons. The uncertainty includes variations of the relevant kaon-matter interaction cross sections; of the material modeling in the MC; and efficiencies as a function of kaon isolation, $p_T$, and $\eta$. The efficiency for associating at least three $r\phi$ hits from L00+SVXII to a good kaon COT track is evaluated using $B^+ \to J/\psi\ K^+$ data events. The $B^+ \to J/\psi\ K^+$ signal yield is compared before and after applying the L00+SVXII requirements on the kaon track and the ratio is used as a measure of the efficiency. The $\mu^+\mu^-$ pair in the events must satisfy all the relevant dimuon baseline requirements of Sec. IV A, including the silicon requirements discussed in Sec. V C 4 and the dimuon vertex requirements of Sec. V C 5. The resulting efficiencies are $0.942 \pm 0.002$ and $0.948 \pm 0.003$ for the CC and CF channels, respectively. The efficiency of the baseline requirements relevant for the $\mu^+\mu^-K^+$ vertex is also directly determined using sideband-subtracted $B^+ \to J/\psi\ K^+$ data and is $0.938 \pm 0.006$ and $0.919 \pm 0.010$ for the CC and CF channels, respectively.

### 8. Total reconstruction efficiency

The total dimuon-reconstruction efficiency-ratio is the product over all the dimuon efficiency ratios described above. The final efficiency ratio relevant for Eq. (1) is obtained by including the product of the kaon reconstruction efficiencies and the $B^+ \to J/\psi\ K^+$ vertex efficiency. The resulting ratios for the CC and CF channel and their associated total uncertainties are shown in Table II.

### D. Efficiency of the NN selection

The final NN selection criteria divide the surviving events into eight bins in $\nu_N$. The bin boundaries are determined in an optimization described in Sec. VIII. The efficiency associated with each $\nu_N$ bin is estimated from simulated $B_s^0 \to \mu^+\mu^-$ events meeting all other selection criteria. In Table II the efficiency summed over all eight bins ($\nu_N > 0.70$) and for the highest bin alone ($\nu_N > 0.995$) are shown for the CC and CF channels. For the likelihood fits described in Sec. VIII, the efficiency given in Table III is used in each bin separately. The NN efficiency determined from the $B^0 \to \mu^+\mu^-$ MC sample is consistent with the results of Table III. Recall that the NN is not applied in selecting the $B^+ \to J/\psi\, K^+$ sample relevant for the normalization in Eq. (1).

The MC modeling of the NN efficiency is checked by comparing the signal efficiencies for $B^+ \to J/\psi\, K^+$ decays reconstructed in data and MC using the $B^+ \to J/\psi\, K^+$ MC sample described in Sec. IV D 3. The comparisons are shown in Figs. 13 and 14 and are quantified in Table IV. Some difference in performance is observed and likely arises from radiation-damage-induced silicon-sensor degradation that is not completely simulated. For most NN bins the difference between MC and sideband-subtracted data does not exceed 2.5 times the associated statistical uncertainty. The most significant deviation occurs in the $\nu_N > 0.995$ bin and is 3.4% and 7.0% for the CC and CF, respectively. These differences are assigned as systematic uncertainties to this bin.

The distributions of simulated $B^+ \to J/\psi\, K^+$ and $B_s^0 \to \mu^+\mu^-$ events are weighted to match the $p_T$ and isolation distributions from $B^+ \to J/\psi\, K^+$ and $B_s^0 \to J/\psi\, \phi$ data, respectively. A systematic uncertainty of 4% is assigned as determined by varying the observed $B$-meson $p_T$ and isolation distributions within their statistical uncertainties, repeating the weighting, and quantifying the resulting changes in the NN efficiencies.

### E. Standard model signal expectations

The expected SM $B_s^0 \to \mu^+\mu^-$ signal yield for each NN bin is given in Table V and is estimated using Eq. (1), the SM value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ [2], the quantities from Table II, and the NN efficiencies of Table III to solve for $N_{B_s^0}$. Combining all NN bins, approximately 1.4 and 1.0 SM $B_s^0 \to \mu^+\mu^-$ events are expected in the CC and CF channels, respectively. The expected SM $B^0 \to \mu^+\mu^-$ yield is a factor of $(f_s/f_u) \times (\mathcal{B}(B_s^0 \to \mu^+\mu^-)/\mathcal{B}(B^0 \to \mu^+\mu^-)) \approx 9$ smaller.

## VI. BACKGROUND ESTIMATION

The background falls into two classes. The dominant source of background in the $B_s^0 \to \mu^+\mu^-$ search comes from accidental combinations of muon candidates that meet the selection requirements (combinatorial background). In addition, a peaking background from $B \to h^+h'^-$ decays, where $h$ and $h'$ are either a pion or kaon, contributes. These two-body charmless $B$ decays are a more significant background for the $B^0 \to \mu^+\mu^-$ search due to the downward shift in $M_{\mu^+\mu^-}$ caused by assuming the muon mass for both charged particles. The two classes of backgrounds are estimated for each $(\nu_N, M_{\mu^+\mu^-})$ bin for the CC and CF channels separately.

### A. Combinatorial backgrounds

The combinatorial background is estimated using data events in the $5.009 < M_{\mu^+\mu^-} < 5.169$ GeV/$c^2$ and $5.469 < M_{\mu^+\mu^-} < 5.969$ GeV/$c^2$ sidebands. A fit to a straight line for the sum of sideband events over all eight $\nu_N$ bins, $\nu_N > 0.70$, is shown in Fig. 15 and is used to determined a fixed slope. For each $\nu_N$ bin the mass sidebands are fit to a straight line using the fixed slope but with a free floating normalization. The resulting function is integrated over the relevant mass signal region to estimate the combinatorial background in that $\nu_N$ and $M_{\mu^+\mu^-}$ bin. The fits in each NN bin are shown in Figures 16 and 17 for the CC and CF channels, respectively. This fixed-slope methodology significantly reduces the uncertainty on the combinatorial background for the highest $\nu_N$ bins and is possible because $M_{\mu^+\mu^-}$ and $\nu_N$ are independent (cf. Sec. IV D 1). We assign a 6% systematic uncertainty associated with the statistical uncertainty on the fixed slope. The statistical uncertainty from the normalization is also propagated into the background-estimate uncertainties and is a dominant contribution to the total uncertainty in the most sensitive NN bins. The final combinatorial background estimates are given in Tables VI and VII.

Fully reconstructed $B \to h^+h'^-$ and partially reconstructed $b \to \mu^+\mu^- X$ decays can have kinematic properties similar to $B_s^0(B^0) \to \mu^+\mu^-$ decays and thus can obtain large $\nu_N$ values. If these background processes contribute significantly to the sideband regions, they would invalidate the fixed-sloped methodology. The estimated $B \to h^+h'^-$ contribution to the sideband regions is less than 0.1 event in the CC and CF channels each. The partially reconstructed decays are constrained to have $M_{\mu^+\mu^-} < M_{B_s^0}$, and the lower edge of the lower sideband (5.009 GeV/$c^2$) is chosen to largely eliminate these events. To account for the possibility that these background processes are affecting the combinatorial background estimates in the highest $\nu_N$ bins, an additional systematic uncertainty is assessed for the three highest bins using alternative fits to the $M_{\mu^+\mu^-}$ sideband distribution. An alternative straight-line fit is performed as described above except that all parameters of the fit are left floating. A second alternative fit is performed over the extended dimuon-mass sideband region, $4.669 <$

TABLE III: The NN efficiency for each $\nu_N$ bin for the CC and CF channels after $p_T$ and isolation weighting. The uncertainties include only the statistical component.

| NN bin | CC | CF |
|---|---|---|
| $0.700 < \nu_N < 0.760$ | 2.2±0.1 | 2.3±0.1 |
| $0.760 < \nu_N < 0.850$ | 4.1±0.1 | 4.4±0.1 |
| $0.850 < \nu_N < 0.900$ | 2.9±0.1 | 3.4±0.1 |
| $0.900 < \nu_N < 0.940$ | 4.5±0.1 | 4.7±0.1 |
| $0.940 < \nu_N < 0.970$ | 8.3±0.1 | 6.2±0.1 |
| $0.970 < \nu_N < 0.987$ | 10.9±0.1 | 10.2±0.1 |
| $0.987 < \nu_N < 0.995$ | 12.5±0.1 | 8.5±0.1 |
| $0.995 < \nu_N < 1.000$ | 46.1±0.3 | 46.8±0.3 |

TABLE IV: Relative difference in NN bin efficiency between $B^+ \to J/\psi\ K^+$ data and MC. A positive (negative) difference indicates that the MC efficiency is higher (lower) than the data efficiency. The differences normalized to the associated statistical uncertainty are given in parenthesis.

| NN bin | CC | CF |
|---|---|---|
| $0.700 < \nu_N < 0.760$ | –8.3% (–1.6$\sigma$) | –5.3% (–0.7$\sigma$) |
| $0.760 < \nu_N < 0.850$ | –8.5% (–2.3$\sigma$) | –7.9% (–1.4$\sigma$) |
| $0.850 < \nu_N < 0.900$ | 4.0% (+0.9$\sigma$) | –8.2% (–1.3$\sigma$) |
| $0.900 < \nu_N < 0.940$ | –0.5% (–0.1$\sigma$) | 2.4% (+0.5$\sigma$) |
| $0.940 < \nu_N < 0.970$ | 0.1% (+0.1$\sigma$) | –6.1% (–1.4$\sigma$) |
| $0.970 < \nu_N < 0.987$ | 2.9% (+1.1$\sigma$) | 0.3% (+0.1$\sigma$) |
| $0.987 < \nu_N < 0.995$ | 4.4% (+2.1$\sigma$) | –4.1% (–1.0$\sigma$) |
| $0.995 < \nu_N < 1.000$ | 3.4% (+2.6$\sigma$) | 7.0% (+3.7$\sigma$) |

$M_{\mu^+\mu^-} < 5.169$ GeV/$c^2$ and $5.469 < M_{\mu^+\mu^-} < 5.969$ GeV/$c^2$. By decreasing the lower edge of the lower sideband, a non-negligible number of partially reconstructed decays are admitted to the sideband sample in the highest $\nu_N$ bins and an exponential fit function is used. The alternative fits are shown in Figs. 18 and 19 for the CC and CF channels, respectively. For each alternative, the resulting combinatorial-background estimate is compared to the default estimates in Tables VI and VII and the largest observed difference is assigned as an additional systematic uncertainty for each bin. In the highest $\nu_N$ bins, these systematic uncertainties are of the same magnitude as the statistical uncertainty on the normalization. The final systematic uncertainties range from ±19% (±3%) to ±43% (±42%) for the CC (CF) channel.

### B. Peaking backgrounds

Background from $B \to h^+h'^-$ decays must be estimated separately since they produce a peak in the $M_{\mu^+\mu^-}$ distribution and are not included in the combinatorial-background estimates described in Sec. VI A. Decays involving $B$ baryons, such as $\Lambda_b \to p\pi^-$, are more heavily suppressed than the $B$-meson background due to the significantly lower rate at which protons reach the muon detectors and satisfy the muon-identification requirements and due to smaller production cross-sections [35].

The $B \to h^+h'^-$ contribution to the dimuon-mass signal region is estimated using data to determine the pion- and kaon-misidentification rates and the methods of Sec. V A to determine the remaining acceptances and efficiencies. The resulting $B \to h^+h'^-$ background is about a factor of ten smaller than the combinatorial background in the $B_s^0$ search, while comprising about half the total background in the $B^0$ search.

The probability for pions and kaons to meet the muon identification requirements is extracted with a pure sample of kaons and pions from $D^*$-tagged $D^0 \to K^-\pi^+$ decays. These decays yield two same-sign pions, one from the $D^{*+} \to D^0\pi^+$ decay and one from the subsequent $D^0$ decay, and an oppositely signed kaon. These charge correlations are used to identify the pions and kaons unambiguously without needing to rely on any particle identification criteria. The $D^*$-tagged event sample is collected using the first 7 fb$^{-1}$ of data with a trigger that requires two charged particles displaced from the beam line, each with $p_T > 2$ GeV/$c$, that reconstruct to a secondary vertex [41]. We further require that the trigger particles have opposite charge and satisfy the baseline tracking criteria described in Sec. IV A. A kinematic fit is performed, constraining the two tracks to a common vertex, that must satisfy $\chi^2/N_{\rm dof} < 15$. The resulting $D^0 \to K^-\pi^+$ candidate must have $|\eta| < 1$, an impact parameter less than 100 $\mu$m, and $1.77 < M_{K\pi} < 1.97$ GeV/$c^2$. For surviving events we associate a third track, the soft pion from the $D^{*+} \to D^0\pi$ decay, with $p_T > 0.4$ GeV/$c$, $|z_0| < 1.5$ cm, and $|d_0| < 600$ $\mu$m and form a three-track vertex that must satisfy $144 < M_{K\pi\pi} - M_{K\pi} < 147$ MeV/$c^2$. The resulting sample of kaons and pions from the $D^0$ decay

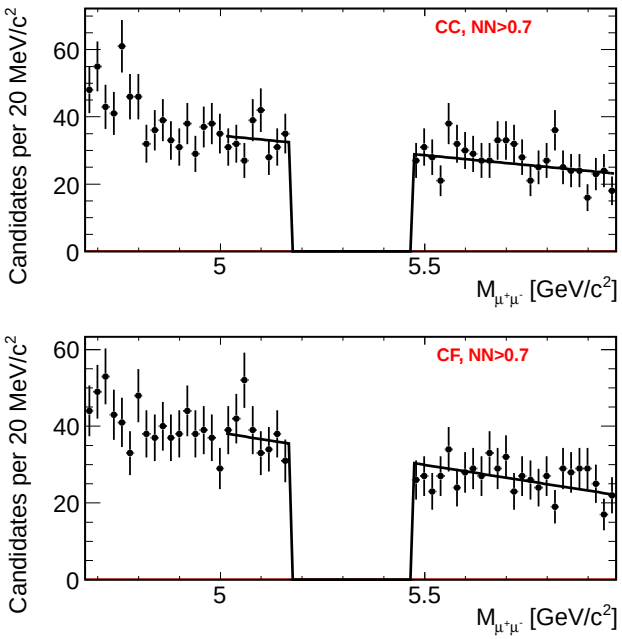


FIG. 15: Dimuon-mass distributions for $\nu_N > 0.7$ for the CC and CF channels with extended signal-region blinded. The slopes from these fits are fixed and used to estimate the combinatorial background in each NN bin.

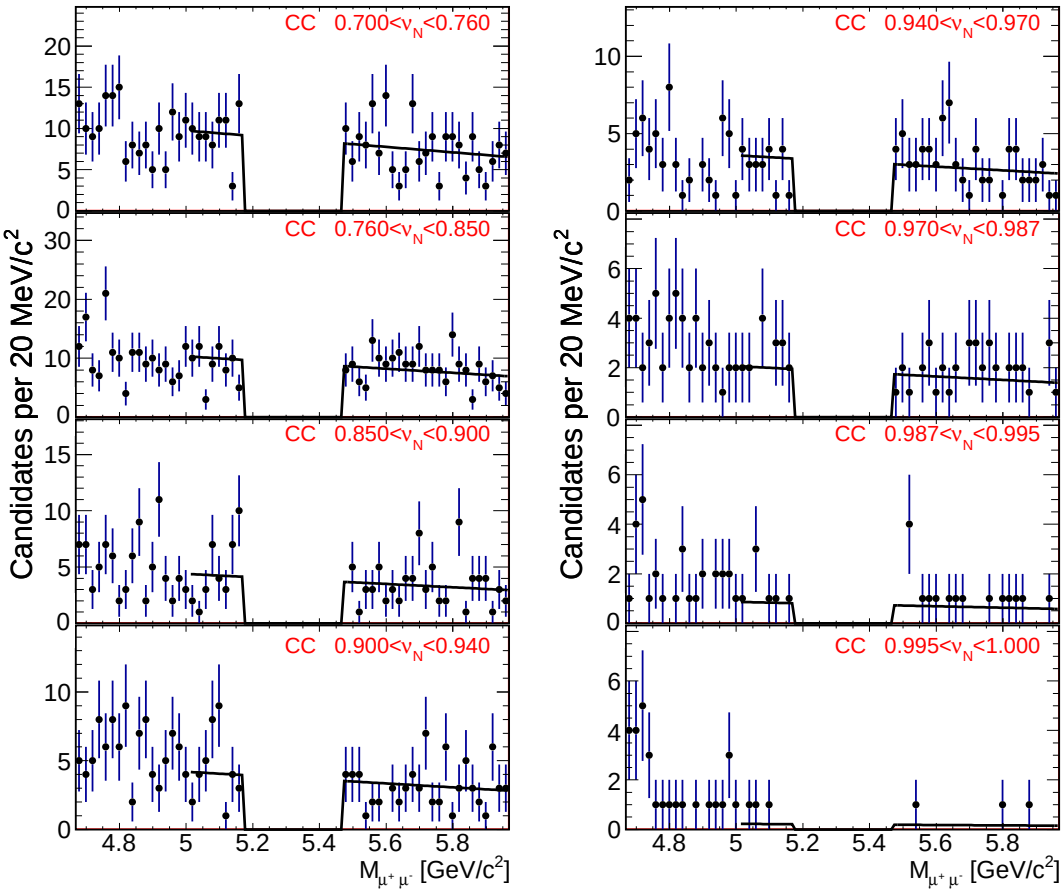


FIG. 16: Dimuon-mass distributions with fit overlaid for each of the eight NN bins for the CC channel with the extended signal-region blinded. The slope of each curve is taken from the fit in Fig. 15 while the normalization is determined in each bin separately.

TABLE V: The SM expected $B_s^0 \to \mu^+\mu^-$ signal contribution in each NN bin for the CC and CF channels separately.

| NN bin | CC | CF |
|---|---|---|
| $0.700 < \nu_N < 0.760$ | $0.04 \pm 0.01$ | $0.03 \pm 0.01$ |
| $0.760 < \nu_N < 0.850$ | $0.07 \pm 0.01$ | $0.05 \pm 0.01$ |
| $0.850 < \nu_N < 0.900$ | $0.05 \pm 0.01$ | $0.04 \pm 0.01$ |
| $0.900 < \nu_N < 0.940$ | $0.07 \pm 0.01$ | $0.05 \pm 0.01$ |
| $0.940 < \nu_N < 0.970$ | $0.10 \pm 0.02$ | $0.07 \pm 0.01$ |
| $0.970 < \nu_N < 0.987$ | $0.13 \pm 0.02$ | $0.11 \pm 0.02$ |
| $0.987 < \nu_N < 0.995$ | $0.20 \pm 0.04$ | $0.09 \pm 0.02$ |
| $0.995 < \nu_N < 1.000$ | $0.75 \pm 0.13$ | $0.52 \pm 0.10$ |

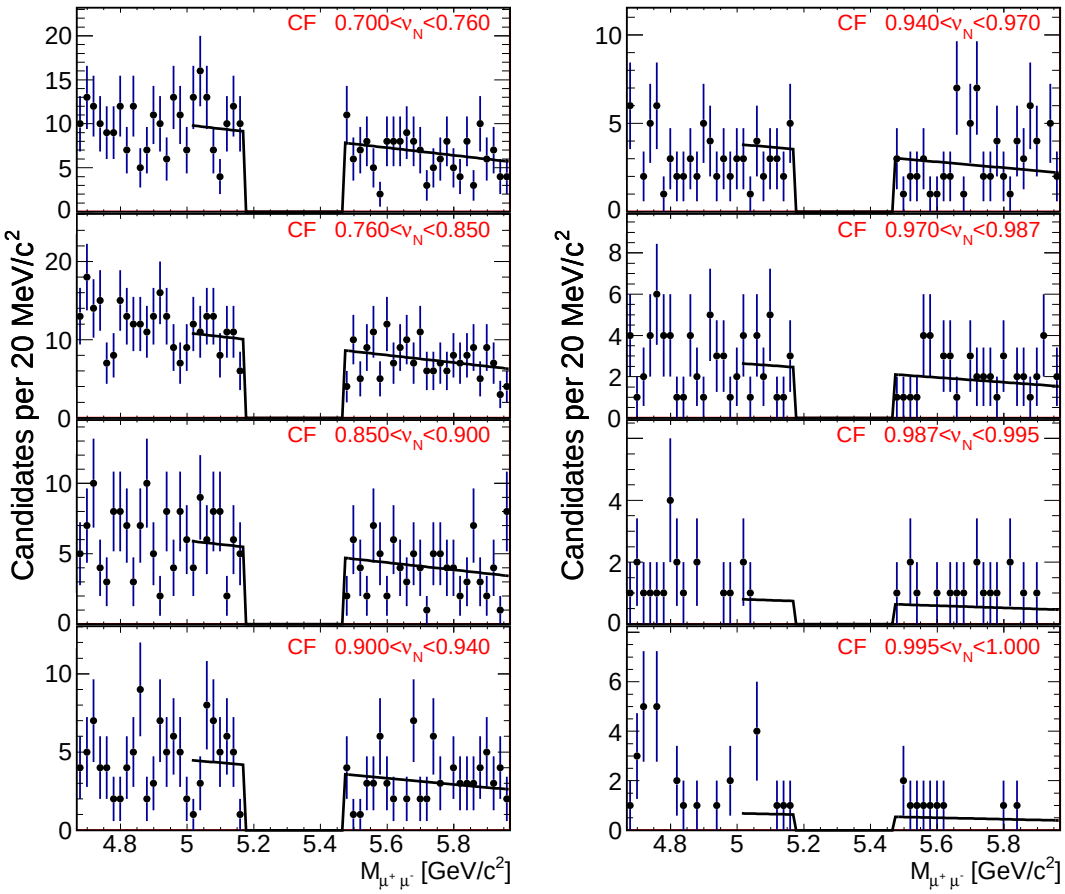


FIG. 17: Dimuon-mass distributions with fit overlaid for each of the eight NN bins for the CF channel with the extended signal-region blinded. The slope of each curve is taken from the fit in Fig. 15 while the normalization is determined in each bin separately.

TABLE VI: Estimated number of combinatorial background events for the $B_s^0$ dimuon-mass signal-region for each $(M_{\mu^+\mu^-}, \nu_N)$ bin and the associated statistical uncertainty.

| NN bin \ Mass bins (GeV/$c^2$) | 5.310−5.334 | 5.334−5.358 | 5.358−5.382 | 5.382−5.406 | 5.406−5.430 |
|---|---|---|---|---|---|
| **CC** | | | | | |
| $0.700 < \nu_N < 0.760$ | 10.42±0.72 | 10.33±0.71 | 10.23±0.70 | 10.14±0.70 | 10.04±0.69 |
| $0.760 < \nu_N < 0.850$ | 11.02±0.74 | 10.92±0.74 | 10.82±0.73 | 10.72±0.72 | 10.62±0.71 |
| $0.850 < \nu_N < 0.900$ | 4.69±0.46 | 4.65±0.45 | 4.61±0.45 | 4.56±0.44 | 4.52±0.44 |
| $0.900 < \nu_N < 0.940$ | 4.49±0.45 | 4.45±0.44 | 4.41±0.44 | 4.37±0.43 | 4.33±0.43 |
| $0.940 < \nu_N < 0.970$ | 3.85±0.41 | 3.81±0.41 | 3.78±0.40 | 3.74±0.40 | 3.71±0.39 |
| $0.970 < \nu_N < 0.987$ | 2.21±0.30 | 2.19±0.30 | 2.17±0.30 | 2.14±0.30 | 2.12±0.29 |
| $0.987 < \nu_N < 0.995$ | 0.92±0.19 | 0.91±0.19 | 0.91±0.19 | 0.90±0.19 | 0.89±0.19 |
| $0.995 < \nu_N < 1.000$ | 0.24±0.10 | 0.24±0.10 | 0.24±0.10 | 0.23±0.10 | 0.23±0.10 |
| **CF** | | | | | |
| $0.700 < \nu_N < 0.760$ | 10.18±0.72 | 10.05±0.71 | 9.93±0.70 | 9.80±0.69 | 9.68±0.68 |
| $0.760 < \nu_N < 0.850$ | 11.21±0.76 | 11.08±0.75 | 10.94±0.74 | 10.80±0.73 | 10.66±0.72 |
| $0.850 < \nu_N < 0.900$ | 6.11±0.54 | 6.03±0.53 | 5.96±0.52 | 5.88±0.52 | 5.81±0.51 |
| $0.900 < \nu_N < 0.940$ | 4.65±0.46 | 4.59±0.46 | 4.54±0.45 | 4.48±0.44 | 4.42±0.44 |
| $0.940 < \nu_N < 0.970$ | 3.94±0.42 | 3.90±0.42 | 3.85±0.41 | 3.80±0.41 | 3.75±0.40 |
| $0.970 < \nu_N < 0.987$ | 2.74±0.35 | 2.71±0.34 | 2.67±0.34 | 2.64±0.34 | 2.61±0.33 |
| $0.987 < \nu_N < 0.995$ | 0.83±0.19 | 0.82±0.18 | 0.81±0.18 | 0.80±0.18 | 0.79±0.18 |
| $0.995 < \nu_N < 1.000$ | 0.71±0.17 | 0.70±0.17 | 0.69±0.17 | 0.68±0.17 | 0.67±0.16 |

TABLE VII: Estimated number of combinatorial background events for the $B^0$ dimuon-mass signal-region for each $(M_{\mu^+\mu^-}, \nu_N)$ bin and the associated statistical uncertainty.

| NN bin \ Mass bin (GeV/$c^2$) | 5.219−5.243 | 5.243−5.267 | 5.267−5.291 | 5.291−5.315 | 5.315−5.339 |
|---|---|---|---|---|---|
| **CC** | | | | | |
| $0.700 < \nu_N < 0.760$ | 10.78±0.74 | 10.69±0.74 | 10.59±0.73 | 10.50±0.72 | 10.40±0.72 |
| $0.760 < \nu_N < 0.850$ | 11.41±0.77 | 11.30±0.76 | 11.21±0.76 | 11.10±0.75 | 11.00±0.74 |
| $0.850 < \nu_N < 0.900$ | 4.85±0.47 | 4.81±0.47 | 4.77±0.46 | 4.72±0.46 | 4.68±0.46 |
| $0.900 < \nu_N < 0.940$ | 4.64±0.46 | 4.60±0.46 | 4.56±0.45 | 4.52±0.45 | 4.48±0.44 |
| $0.940 < \nu_N < 0.970$ | 3.98±0.42 | 3.95±0.42 | 3.91±0.42 | 3.88±0.41 | 3.84±0.41 |
| $0.970 < \nu_N < 0.987$ | 2.28±0.32 | 2.26±0.31 | 2.24±0.31 | 2.22±0.31 | 2.20±0.30 |
| $0.987 < \nu_N < 0.995$ | 0.95±0.20 | 0.95±0.20 | 0.94±0.20 | 0.93±0.20 | 0.92±0.19 |
| $0.995 < \nu_N < 1.000$ | 0.25±0.10 | 0.25±0.10 | 0.24±0.10 | 0.24±0.10 | 0.24±0.10 |
| **CF** | | | | | |
| $0.700 < \nu_N < 0.760$ | 10.65±0.75 | 10.52±0.74 | 10.40±0.73 | 10.27±0.72 | 10.15±0.72 |
| $0.760 < \nu_N < 0.850$ | 11.73±0.80 | 11.60±0.79 | 11.46±0.78 | 11.32±0.77 | 11.18±0.76 |
| $0.850 < \nu_N < 0.900$ | 6.39±0.56 | 6.31±0.55 | 6.24±0.55 | 6.16±0.54 | 6.09±0.53 |
| $0.900 < \nu_N < 0.940$ | 4.87±0.48 | 4.81±0.48 | 4.75±0.47 | 4.70±0.47 | 4.64±0.46 |
| $0.940 < \nu_N < 0.970$ | 4.13±0.44 | 4.08±0.44 | 4.03±0.43 | 3.98±0.43 | 3.94±0.42 |
| $0.970 < \nu_N < 0.987$ | 2.87±0.36 | 2.83±0.36 | 2.80±0.35 | 2.77±0.35 | 2.73±0.35 |
| $0.987 < \nu_N < 0.995$ | 0.87±0.20 | 0.86±0.19 | 0.85±0.19 | 0.84±0.19 | 0.83±0.19 |
| $0.995 < \nu_N < 1.000$ | 0.74±0.18 | 0.73±0.18 | 0.72±0.18 | 0.71±0.17 | 0.70±0.17 |

has a purity of $> 99\%$ and is used to estimate the efficiency of the muon-identification requirements for kaons and pions, or fake rates.

The pion (kaon) fake rate is determined as the ratio of the number of $D^0$ candidates with a pion (kaon) track that satisfies the muon identification criteria of Sec. IV A to the total number of $D^0$ candidates. The fake rates for $\pi^+$, $\pi^-$, $K^+$, and $K^-$ are measured as a function of instantaneous luminosity and $p_T$ for central and forward muons separately. The $\pi^+$ and $\pi^-$ rates are found to be consistent and are combined to yield an average $\pi^\pm$ fake rate. The number of events in which one track meets the muon selection criteria is estimated by fitting the $M_{K\pi}$ mass distribution to a function that is the sum of a Gaussian distribution and first-order polynomial. The number of events where both tracks fail the muon-selection criteria is determined by fitting the $M_{K\pi}$ distribution to the sum of two Gaussian distributions and a first-order polynomial. Figures 20 and 21 show the $M_{K\pi}$ distributions with fits overlaid for a lower and higher kaon $p_T$ bin for the central-muon-detector system combining all luminosity bins.

Time-dependent changes in fake rates can occur due to changes in the accelerator performance affecting the instantaneous luminosity or due to differences in detector performance associated with aging or changes in the operational configuration. The instantaneous luminosity was found to be the primary source of the fake-rate time dependence. We perform a consistency check to investigate other sources. The fake rates, binned in $p_T$ and instantaneous luminosity, are applied to the $D^*$-tagged sample as weights. The sum of weights is then compared to the actual number of fakes in bins of calendar date resulting in differences of up to 20%. This 20% difference is assigned as a systematic uncertainty and accounts for the largest contribution to the fake-rate uncertainty. In

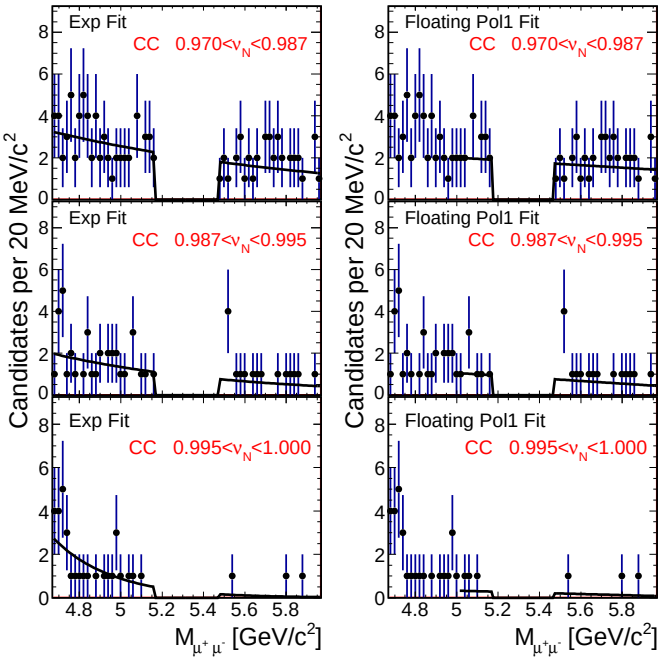


FIG. 18: Dimuon mass-sideband distributions with alternative fits overlaid for the three highest NN bins in the CC channel with the extended signal region blinded.

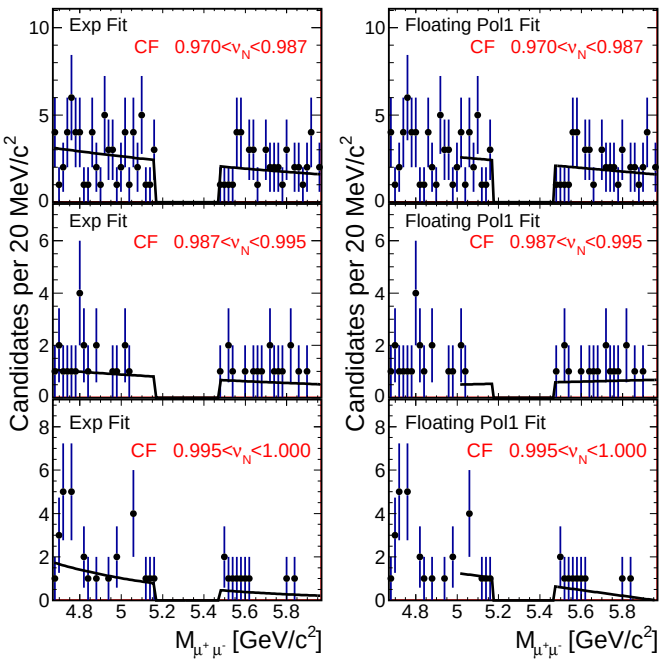


FIG. 19: Dimuon mass-sideband distributions with alternative fits overlaid for the three highest NN bins in the CF channel with the extended signal region blinded.

the final determination of the total $B \to h^+h'^-$ contribution a weighted average of fake rates is used, based on the instantaneous-luminosity profile of the dimuon-mass-sideband events. The luminosity-averaged fake rates are shown in Fig. 22.

The expected number of peaking background events for a specific two-body hadronic decay is given by

$$N_b = F_{sb}\,\frac{\mathcal{B}_b}{S_s}\,\frac{\epsilon_b^{\rm reco}}{\epsilon_{B_s^0}^{\rm reco}},$$

where $F_{sb}$ is the ratio of relevant fragmentation fractions depending on which signal and background channels are being evaluated (i.e., $B_s^0 \to h^+h'^-$ or $B^0 \to h^+h'^-$) and on which signal-search mass-window is selected (i.e., $B_s^0 \to \mu^+\mu^-$ or $B^0 \to \mu^+\mu^-$). For $B^0$ hadronic backgrounds in the $B^0$ search and for $B_s^0$ hadronic backgrounds in the $B_s^0$ search this factor equals unity. For $B^0$ hadronic backgrounds in the $B_s^0$ search region $F_{sb} = f_d/f_s$, while for $B_s^0$ hadronic backgrounds in the $B^0$ search $F_{sb} = f_s/f_d$. The value for $f_d/f_s$ is taken from Ref. [36] and has a 13% uncertainty. The value of the branching fraction $\mathcal{B}_b$ for a specific background mode and its associated uncertainty are also taken from Ref. [36]. For unobserved $B \to h^+h'^-$ processes, the current branching-fraction upper-limits are used and a 100% uncertainty is assigned. The $S_s$ term corresponds to the single-event-sensitivity for $B_s^0 \to \mu^+\mu^-$ as defined in Sec. V and is taken from Eq. (1). The last term corrects $S_s$ for the differing reconstruction efficiencies between $B_s^0 \to \mu^+\mu^-$ ($\epsilon_{B_s^0}^{\rm reco}$) and the relevant $B \to h^+h'^-$ decay ($\epsilon_b^{\rm reco}$). In particular, the muon-stub reconstruction and identification efficiencies discussed in Secs. V C 2 and V C 3 are replaced by the relevant double-track pion and kaon fake-rate estimates. The double-track fake rate is estimated from the convolution of the single-track fake rates with the $(p_T^{h^+}, p_T^{h^-})$ spectra obtained from a sample of $B \to h^+h'^-$ MC events meeting the baseline criteria and using the instantaneous-luminosity distribution observed in the $M_{\mu^+\mu^-}$ sideband events. A systematic uncertainty of 35% is assigned based on the 20% single-track uncertainty. The efficiency for the reconstructed invariant mass falling into a given $M_{\mu^+\mu^-}$ bin is corrected for differences between $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h'^-$ decays using simulated $B \to h^+h'^-$ events surviving the baseline criteria weighted by their double-track fake rates. All other relevant efficiencies are found to be consistent between $B_s^0 \to \mu^+\mu^-$ and $B \to h^+h'^-$ decays. The final peaking background estimates are given in Tables VIII and IX.

The $\nu_N$ distribution is assumed to be the same for both $B \to h^+h'^-$ and $B_s^0 \to \mu^+\mu^-$ since they feature similar kinematic properties and the $\nu_N$ does not use any muon-identification criteria as input variables. We verify that the $p_T$ dependence of the fake rates negligibly affects the $\nu_N$ distribution.

### C. Background estimate checks with control samples

The methods used to predict the background rates are validated using statistically-independent background-

TABLE VIII: The $B \to h^+h'^-$ background estimates and their total uncertainty for the $B_s^0$ signal window for each $(M_{\mu^+\mu^-}, \nu_N)$ bin. The contributions are negligibly small in all the lower $\nu_N$ bins. Uncertainties less than 0.001 are given as 0.000 in the table.

| NN bin \ Mass bin (GeV/$c^2$) | 5.310−5.334 | 5.334−5.358 | 5.358−5.382 | 5.382−5.406 | 5.406−5.430 |
|---|---|---|---|---|---|
| **CC** | | | | | |
| $0.700 < \nu_N < 0.760$ | 0.003±0.000 | 0.001±0.000 | - | - | - |
| $0.760 < \nu_N < 0.850$ | 0.006±0.000 | 0.002±0.000 | 0.001±0.000 | - | - |
| $0.850 < \nu_N < 0.900$ | 0.005±0.001 | 0.002±0.000 | 0.001±0.000 | - | - |
| $0.900 < \nu_N < 0.940$ | 0.007±0.001 | 0.003±0.000 | 0.001±0.000 | - | - |
| $0.940 < \nu_N < 0.970$ | 0.011±0.001 | 0.003±0.000 | 0.001±0.000 | - | - |
| $0.970 < \nu_N < 0.987$ | 0.013±0.002 | 0.005±0.001 | 0.002±0.000 | 0.001±0.000 | - |
| $0.987 < \nu_N < 0.995$ | 0.019±0.002 | 0.007±0.001 | 0.002±0.000 | 0.001±0.000 | - |
| $0.995 < \nu_N < 1.000$ | 0.074±0.010 | 0.026±0.003 | 0.009±0.001 | 0.003±0.000 | 0.001±0.000 |
| **CF** | | | | | |
| $0.700 < \nu_N < 0.760$ | 0.001±0.000 | - | - | - | - |
| $0.760 < \nu_N < 0.850$ | 0.002±0.000 | 0.001±0.000 | - | - | - |
| $0.850 < \nu_N < 0.900$ | 0.002±0.000 | 0.001±0.000 | - | - | - |
| $0.900 < \nu_N < 0.940$ | 0.002±0.000 | 0.001±0.000 | - | - | - |
| $0.940 < \nu_N < 0.970$ | 0.003±0.000 | 0.001±0.000 | - | - | - |
| $0.970 < \nu_N < 0.987$ | 0.004±0.001 | 0.002±0.000 | 0.001±0.000 | - | - |
| $0.987 < \nu_N < 0.995$ | 0.004±0.001 | 0.002±0.000 | 0.001±0.000 | - | - |
| $0.995 < \nu_N < 1.000$ | 0.021±0.003 | 0.009±0.001 | 0.003±0.000 | 0.002±0.000 | - |

TABLE IX: The $B \to h^+h'^-$ background estimates and their total uncertainty for the $B^0$ signal window for each $(M_{\mu^+\mu^-}, \nu_N)$ bin. Uncertainties less than 0.001 are given as 0.000 in the table.

| NN bin \ Mass bin (GeV/$c^2$) | 5.219−5.243 | 5.243−5.267 | 5.267−5.291 | 5.291−5.315 | 5.315−5.339 |
|---|---|---|---|---|---|
| **CC** | | | | | |
| $0.700 < \nu_N < 0.760$ | 0.015±0.002 | 0.013±0.001 | 0.011±0.001 | 0.006±0.001 | 0.002±0.000 |
| $0.760 < \nu_N < 0.850$ | 0.027±0.003 | 0.027±0.003 | 0.019±0.002 | 0.011±0.002 | 0.004±0.001 |
| $0.850 < \nu_N < 0.900$ | 0.022±0.002 | 0.019±0.002 | 0.014±0.002 | 0.008±0.001 | 0.003±0.000 |
| $0.900 < \nu_N < 0.940$ | 0.030±0.003 | 0.029±0.003 | 0.022±0.003 | 0.013±0.002 | 0.004±0.001 |
| $0.940 < \nu_N < 0.970$ | 0.047±0.005 | 0.039±0.004 | 0.031±0.004 | 0.016±0.002 | 0.005±0.001 |
| $0.970 < \nu_N < 0.987$ | 0.060±0.006 | 0.052±0.006 | 0.040±0.005 | 0.023±0.003 | 0.009±0.001 |
| $0.987 < \nu_N < 0.995$ | 0.084±0.008 | 0.083±0.009 | 0.061±0.008 | 0.033±0.004 | 0.011±0.001 |
| $0.995 < \nu_N < 1.000$ | 0.325±0.032 | 0.298±0.031 | 0.221±0.028 | 0.126±0.017 | 0.050±0.006 |
| **CF** | | | | | |
| $0.700 < \nu_N < 0.760$ | 0.004±0.000 | 0.004±0.000 | 0.003±0.000 | 0.002±0.000 | 0.001±0.000 |
| $0.760 < \nu_N < 0.850$ | 0.007±0.001 | 0.008±0.001 | 0.006±0.001 | 0.004±0.001 | 0.001±0.000 |
| $0.850 < \nu_N < 0.900$ | 0.006±0.001 | 0.006±0.001 | 0.005±0.001 | 0.003±0.000 | 0.001±0.000 |
| $0.900 < \nu_N < 0.940$ | 0.008±0.001 | 0.008±0.001 | 0.007±0.001 | 0.004±0.001 | 0.001±0.000 |
| $0.940 < \nu_N < 0.970$ | 0.011±0.001 | 0.011±0.001 | 0.009±0.001 | 0.005±0.001 | 0.002±0.000 |
| $0.970 < \nu_N < 0.987$ | 0.017±0.002 | 0.018±0.002 | 0.015±0.002 | 0.008±0.001 | 0.003±0.000 |
| $0.987 < \nu_N < 0.995$ | 0.014±0.001 | 0.015±0.002 | 0.012±0.002 | 0.007±0.001 | 0.002±0.000 |
| $0.995 < \nu_N < 1.000$ | 0.078±0.008 | 0.084±0.009 | 0.065±0.008 | 0.038±0.005 | 0.014±0.002 |

dominated data samples. The control samples are designed to reproduce the salient features of the combinatorial and $B \to h^+h'^-$ backgrounds. Given a signal sample that consists of two opposite-charge muons with $\lambda > 0$, we form four independent control samples,

**OS−:** opposite-sign muon pairs, passing the baseline requirements with $\lambda < 0$;

**SS+:** same-sign muon pairs with $\lambda > 0$ and relaxed trigger-matching to improve event sample size;

**SS−:** same-sign muon pairs with $\lambda < 0$ and relaxed trigger-matching to improve event sample size;

**FM+:** opposite-sign fake-muon-enhanced pairs, in which at least one track of which is required to *fail* the muon likelihood or $dE/dx$ requirement with $\lambda > 0$.

The OS− sample is representative of combinatorial backgrounds with short lifetime, which have a symmetric lifetime distribution around zero. The same-sign samples are dominated by events in which a muon from a semileptonic decay of a $B$ hadron is combined with a muon from the sequential semileptonic decay $b \to cX \to \mu X$ of the other $\overline{B}$ meson in the event and by events in which muons are combined from non-sequential processes. The FM+ sample is enriched in $B \to h^+h'^-$ background due to the reversal of the muon-identification requirements. To mimic the $\Delta\Omega$ and $\lambda$ distributions of the $\lambda > 0$ samples, we apply the transformations $\lambda \to -\lambda$ and $\Delta\Omega \to \pi - \Delta\Omega$ to the $\lambda < 0$ samples.

The background contribution to each control sample is estimated using the same methods as described in Sec. VI A and VI B. For the OS− and SS± samples, only the combinatorial background is estimated due to the dominance of this component over the peaking back-

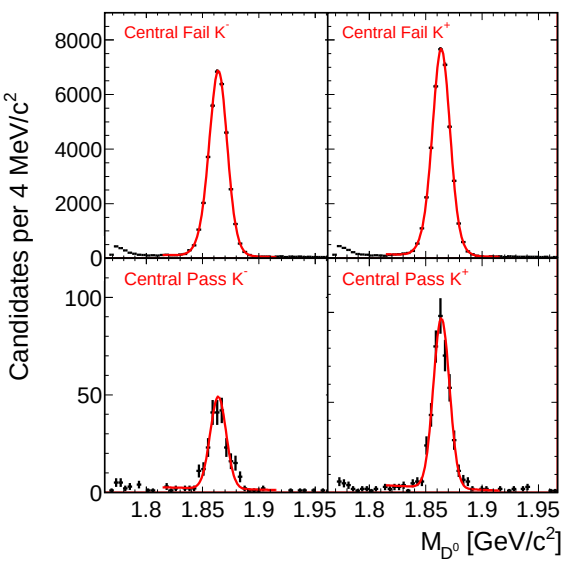


FIG. 20: The $M_{K\pi}$ distributions with fits overlaid for central kaons with $2.0 < p_T < 2.8$ GeV/$c$. The top panels contain the distributions for $D^0$ candidates with kaons that fail the muon requirements, while the bottom panels contain the distributions for $D^0$ candidates with kaons that meet the muon ID requirements. Distributions for $K^-$ ($K^+$) are on the left (right).

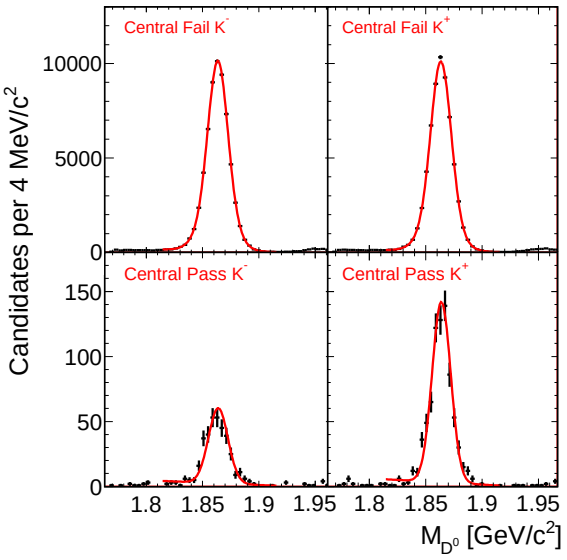


FIG. 21: The $M_{K\pi}$ distributions with fits overlaid for central kaons with $6.0 < p_T < 8.0$ GeV/$c$. The top panels contain the distributions for $D^0$ candidates with kaons that fail the muon requirements, while the bottom panels contain the distributions for $D^0$ candidates with kaons that meet the muon ID requirements. Distributions for $K^-$ ($K^+$) are on the left (right).

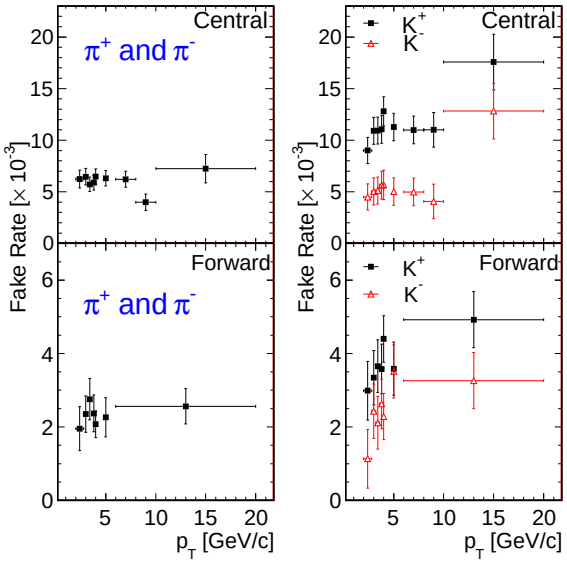


FIG. 22: Fake rates as a function of $p_T$ and averaged over instantaneous luminosity for central (top) and forward (bottom) muon detectors. The left panels show the fake rates for pions while the right panels show the fake rate for kaons.

ground. For the FM+ sample, we estimate both the combinatorial and the $B \to h^+h'^-$ background. New fake-rates are evaluated using the relaxed muon-identification criteria and the method described in Sec. VI B. In all cases the backgrounds are evaluated for the extended $M_{\mu^+\mu^-}$ signal region for each $\nu_N$ bin and then compared to the observed number of events. The Poisson probability for making an observation at least as large as found in the extended signal region, given the predicted background and its systematic uncertainty, is calculated. These probabilities are expected to be uniformly distributed between 0 and 100% for a set of independent samples. The resulting comparisons for all $\nu_N$ bins and control samples are shown in Table X.

No large deviation between the predicted number of background events and the observed number of events is seen, providing confidence in the methods employed to estimate the background rates.

## VII. SYSTEMATIC UNCERTAINTIES

The systematic uncertainties are either related to uncertainties in efficiencies, acceptances, normalization factors, peaking-background estimates, or combinatorial-background estimates. Table XI summarizes all systematic uncertainties.

The dominant systematic uncertainty among the efficiencies, acceptances, and normalization factors is the 13% uncertainty on the ratio of $b$-quark fragmentation fractions, $f_u/f_s$ [36]. The second largest systematic uncertainty is about a factor of two smaller and is due to the acceptance ratio (Sec. V A).

TABLE X: A comparison of the predicted (Pred.) and observed (Obs.) number of events in the extended $M_{\mu^+\mu^-}$ signal region as a function of $\nu_N$ bin for the various control samples. The uncertainties correspond to the uncertainty on the mean of the background prediction. The Poisson probability (Prob.) for making an observation at least as large as the observed yield is also shown. In cases where no events are observed, the probability is actually the Poisson probability to observe zero events assuming a Poisson mean equal to the predicted mean.

| | | CC | | | CF | | |
|---|---|---|---|---|---|---|---|
| Sample | NN bin | Pred. | Obs. | Prob.(%) | Pred. | Obs. | Prob.(%) |
| OS− | $0.700 < \nu_N < 0.760$ | 268.8±14.3 | 249 | 82.3 | 261.8±13.9 | 241 | 84.1 |
| | $0.760 < \nu_N < 0.850$ | 320.8±16.1 | 282 | 95.1 | 399.0±18.5 | 397 | 53.2 |
| | $0.850 < \nu_N < 0.900$ | 150.3±9.9 | 156 | 36.5 | 193.5±11.4 | 185 | 68.3 |
| | $0.900 < \nu_N < 0.940$ | 146.2±9.7 | 158 | 23.0 | 177.4±10.8 | 183 | 37.7 |
| | $0.940 < \nu_N < 0.970$ | 146.2±9.7 | 137 | 72.9 | 156.8±10.1 | 143 | 81.2 |
| | $0.970 < \nu_N < 0.987$ | 100.4±7.8 | 98 | 58.3 | 112.6±8.2 | 110 | 58.3 |
| | $0.987 < \nu_N < 0.995$ | 78.8±6.8 | 59 | 97.0 | 53.3±5.4 | 68 | 6.5 |
| | $0.995 < \nu_N < 1.000$ | 41.2±4.8 | 42 | 47.2 | 48.2±5.1 | 44 | 70.0 |
| SS+ | $0.700 < \nu_N < 0.760$ | 4.8±1.2 | 3 | 81.8 | 0.9±0.5 | 3 | 8.9 |
| | $0.760 < \nu_N < 0.850$ | 3.6±1.0 | 5 | 30.6 | 5.1±1.2 | 5 | 55.4 |
| | $0.850 < \nu_N < 0.900$ | 2.4±0.8 | 5 | 12.2 | 0.9±0.5 | 6 | 0.2 |
| | $0.900 < \nu_N < 0.940$ | 1.5±0.7 | 3 | 21.3 | 0.9±0.5 | 1 | 56.8 |
| | $0.940 < \nu_N < 0.970$ | 1.5±0.7 | 1 | 73.3 | 0.9±0.5 | 1 | 56.8 |
| | $0.970 < \nu_N < 0.987$ | 1.8±0.7 | 2 | 51.3 | 0.9±0.5 | 0 | 40.7 |
| | $0.987 < \nu_N < 0.995$ | 0.3±0.3 | 0 | 74.1 | 0.3±0.3 | 0 | 74.1 |
| | $0.995 < \nu_N < 1.000$ | 0.3±0.3 | 0 | 74.1 | 0.3±0.3 | 1 | 30.0 |
| SS− | $0.700 < \nu_N < 0.760$ | 7.8±1.5 | 10 | 27.8 | 6.0±1.3 | 4 | 80.9 |
| | $0.760 < \nu_N < 0.850$ | 10.5±1.8 | 11 | 47.2 | 7.2±1.5 | 7 | 55.8 |
| | $0.850 < \nu_N < 0.900$ | 4.2±1.1 | 7 | 15.9 | 3.0±0.9 | 2 | 75.8 |
| | $0.900 < \nu_N < 0.940$ | 3.6±1.0 | 4 | 47.2 | 0.9±0.5 | 7 | 0.1 |
| | $0.940 < \nu_N < 0.970$ | 3.3±1.0 | 6 | 14.3 | 3.6±1.0 | 2 | 83.4 |
| | $0.970 < \nu_N < 0.987$ | 3.0±0.9 | 3 | 55.0 | 2.4±0.8 | 5 | 12.2 |
| | $0.987 < \nu_N < 0.995$ | 2.1±0.8 | 0 | 12.2 | 1.2±0.6 | 0 | 30.1 |
| | $0.995 < \nu_N < 1.000$ | 1.2±0.6 | 1 | 65.9 | 1.8±0.7 | 0 | 16.5 |
| FM+ | $0.700 < \nu_N < 0.760$ | 152.2±9.9 | 161 | 29.6 | 66.5±6.1 | 88 | 2.5 |
| | $0.760 < \nu_N < 0.850$ | 140.9±9.5 | 157 | 15.3 | 81.7±6.9 | 76 | 70.0 |
| | $0.850 < \nu_N < 0.900$ | 65.2±6.1 | 50 | 94.4 | 44.7±5.0 | 34 | 91.6 |
| | $0.900 < \nu_N < 0.940$ | 48.7±5.2 | 40 | 85.8 | 24.4±3.6 | 38 | 2.3 |
| | $0.940 < \nu_N < 0.970$ | 27.7±3.8 | 24 | 73.1 | 12.7±2.6 | 20 | 7.1 |
| | $0.970 < \nu_N < 0.987$ | 10.9±2.3 | 12 | 41.4 | 7.7±2.0 | 13 | 8.8 |
| | $0.987 < \nu_N < 0.995$ | 11.0±2.3 | 4 | 98.3 | 2.7±1.1 | 3 | 48.3 |
| | $0.995 < \nu_N < 1.000$ | 28.3±4.1 | 32 | 30.6 | 4.4±1.6 | 8 | 13.0 |

An additional systematic uncertainty, the $B_s^0(B^0)$ mass-shape uncertainty, is assigned based on the probability for a $B_s^0(B^0)$ candidate to populate the dimuon-mass signal-search window. This uncertainty is based on the world average $B_s^0(B^0)$-mass uncertainty and the COT momentum scale and resolution. The final uncertainty ranges between 1% and 9% depending on $M_{\mu^+\mu^-}$ bin.

The leading systematic uncertainty for the peaking background is the 35% uncertainty assigned to the double-track fake rate as discussed in Sec. VI B. In addition, an uncertainty associated with the branching fraction of each $B \to h^+h'^-$ decay is taken from Ref. [36]. Branching fractions for which only upper limits are known are assigned a 100% relative uncertainty. These decays, however, contribute a small fraction to the total $B \to h^+h'^-$ background. The leading contribution to the $B \to h^+h'^-$ background in the $B^0$ signal-search mass-window comes from $B^0 \to K^+\pi^-$, $B^0 \to \pi^+\pi^-$, and $B_s^0 \to K^+K^-$. The branching fractions of these decays have relative uncertainties of 3%, 4%, and 16%, respectively. The $B \to h^+h'^-$ background estimates that require $f_d/f_s$ are treated as correlated with the $B^0 \to \mu^+\mu^-$ normalization (cf. Eq. (1)).

The leading systematic uncertainty, up to 43%, in the combinatorial background estimates is due to the dimuon-mass-sideband shape-uncertainty assigned to the three highest $\nu_N$ bins. The dimuon-mass-sideband shape-uncertainty is discussed in VI A. Another large source of systematic uncertainty, up to 42% in the highest $\nu_N$ bins, is due to finite sample-size in the $M_{\mu^+\mu^-}$ sidebands. A relatively small contribution ($\approx$ 6%) to the total-

combinatorial-background systematic uncertainty arises from the uncertainty on the fixed slope used in the estimates.

In the statistical interpretation of results all the above systematic uncertainties are taken as nuisance parameters with multidimensional Gaussian constraints that include correlations. Combinatorial background yields across all ($M_{\mu^+\mu^-}$, $\nu_N$) bins in the CC or CF channel are correlated because they are fit using the same slope. Combinatorial background estimates in a given $\nu_N$ bin are additionally correlated across the five mass bins because their normalization is determined from the same set of $M_{\mu^+\mu^-}$ sideband events. Peaking backgrounds are treated as correlated across all bins and across the CC and CF channels, due to the use of common fake-rates. All the acceptances and efficiencies of Eq. (1) are treated as correlated across $\nu_N$ and $M_{\mu^+\mu^-}$ bins and across the CC and CF channels due to the use of common MC and data control samples.

## VIII. ANALYSIS OPTIMIZATION

We optimize the $\nu_N$ and $M_{\mu^+\mu^-}$ binning by minimizing the expected 95% C.L. upper limit on $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ assuming only background. The expected limit is calculated using a modified frequentist methodology, the CLs method [42], combining all the $\nu_N$ and $M_{\mu^+\mu^-}$ bins in both the CC and CF channels, while taking correlations between bins into account.

For events surviving all selection criteria, two likelihood fits are performed, one assuming a background-only hypothesis, with the likelihood $\mathcal{L}(b)$, and one assuming a signal-plus-background hypothesis, with likelihood $\mathcal{L}(s+b)$. A log-likelihood ratio LLR $= -2\ln(\mathcal{L}(\mathrm{s+b})/\mathcal{L}(\mathrm{b}))$ is calculated. The likelihoods are defined as the product of Poisson probabilities over all $(M_{\mu^+\mu^-}, \nu_N)$ bins in both the CC and CF channels. For each bin, the mean number of expected events is estimated assuming only background for $\mathcal{L}(b)$ and assuming signal plus background for $\mathcal{L}(s+b)$, while the number of observed events is taken from the number of surviving events falling into that bin. Systematic uncertainties are treated as nuisance parameters with Gaussian constraints. At a fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, the likelihoods are minimized by varying the nuisance parameters. We denote the minimum of the log-likelihood ratio as $\mathrm{LLR_{min}}$.

Expected limits are calculated using an ensemble of background-only simulated data sets corresponding in size to the actual data set used in this analysis. The number of contributing background events in each $(M_{\mu^+\mu^-}, \nu_N)$ bin is drawn from a Poisson distribution whose mean corresponds to the values in Table VI for the combinatorial background and in Table VIII for the peaking background. The mean values shown in the tables are varied by their systematic uncertainties taking into account correlations between bins. Once assembled, each simulated data set is treated just like the experimental data. The median CLs as a function of assumed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is used to determine the expected 95% confidence level (C.L.) upper limit. For alternative choices of the $(M_{\mu^+\mu^-}, \nu_N)$ bins, the methods of Sec. VI A and Sec. VI B are used to generate a mean background-expectation for each bin and to update the systematic uncertainties, while the methods of Sec. V A are used to generate the corresponding signal acceptance.

The optimization process is iterative. While fixing the $M_{\mu^+\mu^-}$ bins, it begins with many $\nu_N$ bins and we then combine neighboring bins with similar expected signal-to-background ratios. The bin boundaries of the resulting eight $\nu_N$ bins are then varied to minimize the expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ limit. Finally the $M_{\mu^+\mu^-}$ bins are varied. The resulting expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ limit is not significantly dependent on the exact choice of bin boundaries and varies by less than 5% over reasonable variations of the bin definitions. The final configuration results in eight $\nu_N$ bins and five $M_{\mu^+\mu^-}$ bins. The $\nu_N$ bins are $0.700 < \nu_N < 0.760$, $0.760 < \nu_N < 0.850$, $0.850 < \nu_N < 0.900$, $0.900 < \nu_N < 0.940$, $0.940 < \nu_N < 0.970$, $0.970 < \nu_N < 0.987$, $0.987 < \nu_N < 0.995$, and $0.995 < \nu_N < 1.000$. The five mass bins for the $B_s^0$ ($B^0$) search are $5.310 < M_{\mu^+\mu^-} < 5.334$, $5.334 < M_{\mu^+\mu^-} < 5.358$, $5.358 < M_{\mu^+\mu^-} < 5.382$, $5.382 < M_{\mu^+\mu^-} < 5.406$, and $5.406 < M_{\mu^+\mu^-} < 5.430$ GeV/$c^2$ ($5.219 < M_{\mu^+\mu^-} < 5.243$, $5.243 < M_{\mu^+\mu^-} < 5.267$, $5.267 < M_{\mu^+\mu^-} < 5.291$, $5.291 < M_{\mu^+\mu^-} < 5.315$, and $5.315 < M_{\mu^+\mu^-} < 5.339$ GeV/$c^2$). This optimization reduces the expected limit by approximately 20% compared to using a single bin with $\nu_N > 0.7$. The final expected-upper-limit for the $B_s^0$ ($B^0$) search is $1.3 \times 10^{-8}$ ($4.2 \times 10^{-9}$) at 95% C.L. and $1.0 \times 10^{-8}$ ($3.4 \times 10^{-9}$) at 90% C.L.

## IX. RESULTS

The background estimates, systematic uncertainties on the background estimates, and the observed number of events for the $B^0 \to \mu^+\mu^-$ search are given in Table XII and summarized in Fig. 23. An ensemble of simulated experiments assuming the background-only hypothesis and including the effects of systematic uncertainties is used to estimate the probability that backgrounds alone could produce a $\mathrm{LLR_{min}}$ value at least as small as the one observed in the data. The resulting $p$-value for the $B^0 \to \mu^+\mu^-$ search is 41%, indicating that the observed events are consistent with the background expectations. The observed upper limits are shown in Fig. 24 and are obtained using the CLs method to give $4.6 \times 10^{-9}$ ($3.8 \times 10^{-9}$) at 95% (90%) C.L.

The results of the $B_s^0 \to \mu^+\mu^-$ search are shown in Table XIII and summarized in Fig. 23. A small excess of events is observed in the CC channel and populates the most sensitive $\nu_N$ and $M_{\mu^+\mu^-}$ bins. The probability that background alone could yield a $\mathrm{LLR_{min}}$ value smaller than that observed is 0.94%, which corresponds to an ex-

TABLE XI: Summary of systematic uncertainties.

| Category | Quantity | CC (%) | CF (%) | Source |
|---|---|---|---|---|
| Efficiencies, acceptance, and normalization factors | $\alpha_{B^+}/\alpha_{B_s^0}$ | 6 | 7 | $b$ mass, renorm. scale, fragmentation modeling |
| | $\epsilon^{\rm COT}_{B^+}/\epsilon^{\rm COT}_{B_s^0}$ | 1 | 1 | Isolation, $p_T(B)$, detector effects |
| | $\epsilon^{\mu}_{B^+}/\epsilon^{\mu}_{B_s^0}$ | 3 | 3 | $Z^0 \to \mu^+\mu^-$ and $J/\psi \to \mu^+\mu^-$ differences |
| | $\epsilon^{\rm PID}_{B^+}/\epsilon^{\rm PID}_{B_s^0}$ | 3 | 3 | $B^+ \to J/\psi\ K^+$ and $J/\psi \to \mu^+\mu^-$ differences |
| | $\epsilon^{\rm SVXII}_{B^+}/\epsilon^{\rm SVXII}_{B_s^0}$ | 3 | 3 | $p_T(\mu)$, two-muon opening-angle, track isolation |
| | $\epsilon^{\rm COT}_{K}$ | 1.7 | 1.7 | Isolation, $p_T(B)$, detector effects |
| | $\epsilon^{\rm NN}$ | 4 | 4 | $B$ isolation, $p_T(B)$ |
| | $\epsilon^{\rm NN}$ for $\nu_N > 0.995$ | 3.4 | 7.0 | Data-MC differences |
| | $f_u/f_s$ | 13 | 13 | Ref. [36] |
| | $\mathcal{B}(B^+ \to J/\psi\ K^+ \to \mu^+\mu^- K^+)$ | 4 | 4 | Ref. [36] |
| | $B$-meson mass shape | 0.1–9 | 0.1–9 | Mass resolution, mass scale, Ref. [36] |
| $B \to h^+h'^-$ | Fake rates | 20 | 20 | Detector and luminosity effects |
| | $\mathcal{B}(B \to h^+h'^-)$ | 3–100 | 3–100 | Ref. [36] |
| Comb. Bkg | Slope | 6 | 6 | Fit uncertainty |
| | Normalization | 7–42 | 7–25 | Sideband sample-size |
| | Shape | 10–43 | 3–42 | Comparison of different fit functions |

cess greater than $2\sigma$. Using the expression $\Delta\chi^2 = \mathrm{LLR} - \mathrm{LLR_{min}}$ we perform a fit to estimate $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The fit, shown in Fig. 25, includes systematic uncertainties and gives $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.3^{+0.9}_{-0.7} \times 10^{-8}$ at 68% C.L. and $0.8 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.4 \times 10^{-8}$ at 95% C.L. A Bayesian method yields very similar results. The probability that the SM, including signal, could yield a value of $\mathrm{LLR_{min}}$ smaller than that observed in the data is 6.8% as determined using an ensemble of simulated experiments that include signal and background contributions, assuming the SM value for $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and including the effects of systematic uncertainties. The observed upper limits from the CLs methodology are $3.1 \times 10^{-8}$ ($2.7 \times 10^{-8}$) at 95% (90%) C.L. and are shown in Fig. 26. Relative to the previous analysis [18], which also reported a small excess using 30% less data, the significance of the excess has diminished and the estimate of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is closer to the SM value.

An excess is observed in the two highest NN bins of the CC channel, the most sensitive bins. The total background expectations for the $0.987 < \nu_N < 0.995$ and $0.995 < \nu_N < 1.000$ bins are 4.56 and 1.29 events while the SM expected signal yields are 0.75 and 0.20 events, respectively. We observe a total of 6 and 4 events, respectively, for these bins. As a check of consistency we redo the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ fit using only the two highest NN bins. This yields a central value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.0^{+0.8}_{-0.6} \times 10^{-8}$, consistent with the full fit. When considering only the two highest NN bins, the probability that background-only (background plus SM signal) could yield a log-likelihood ratio smaller than that observed in data is 2% (22%) including the effects of systematic uncertainties.

We also observe a data excess in the $0.970 < \nu_N < 0.987$ bin of the CC channel in the $B_s^0 \to \mu^+\mu^-$ search, where no significant signal contribution is expected. Note that such an excess does not appear in the $B^0 \to \mu^+\mu^-$ search. This excess originated in the previous $B_s^0 \to \mu^+\mu^-$ analysis and was thoroughly investigated [18]. It was concluded that the excess in this bin was not caused by a problem with the background estimates, a NN bias, or any mis-modeling of the data and was likely due to a statistical upward fluctuation. This conclusion is supported by Fig. 27, which compares the observed data to the background expectations for the $B_s^0 \to \mu^+\mu^-$ search for the 3 fb$^{-1}$ of data added for this analysis. No evidence of an excess in the $0.970 < \nu_N < 0.987$ bin is found in the data added since the analysis of Ref. [18].

Our $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ results are consistent with the bounds set by other experiments and with the SM expectations. This is demonstrated in Fig. 28 for the $B_s^0 \to \mu^+\mu^-$ result, where the small corrections ($< 10\%$) suggested by the recent work in Refs. [43] and [44] have not been considered.

## X. CONCLUSION

We report on the search for $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays using the full CDF Run II data set. These

TABLE XII: The results for the $B^0 \to \mu^+\mu^-$ search comparing the expected total backgrounds and their uncertainty (Exp.) to the number of observed events (Obs.) in each $(\nu_N, M_{\mu^+\mu^-})$ bin. For each $\nu_N$ bin, the sum over mass bins is also shown. The CC and CF channels are given separately.

CC channel results:

| | | $M_{\mu^+\mu^-}$ bin (GeV/$c^2$) | | | | | |
|---|---|---|---|---|---|---|---|
| $\nu_N$ bin | | 5.310–5.334 | 5.334–5.358 | 5.358–5.382 | 5.382–5.406 | 5.406–5.430 | Sum |
| | Exp. | 10.80±0.74 | 10.70±0.74 | 10.61±0.73 | 10.51±0.72 | 10.41±0.72 | 53.02 |
| 0.700–0.760 | Obs. | 15 | 14 | 10 | 7 | 11 | 57 |
| | Exp. | 11.43±0.77 | 11.33±0.76 | 11.23±0.75 | 11.12±0.75 | 11.01±0.74 | 56.12 |
| 0.760–0.850 | Obs. | 12 | 10 | 7 | 8 | 9 | 46 |
| | Exp. | 4.88±0.47 | 4.83±0.47 | 4.78±0.46 | 4.73±0.46 | 4.68±0.46 | 23.90 |
| 0.850–0.900 | Obs. | 10 | 3 | 6 | 6 | 5 | 30 |
| | Exp. | 4.68±0.46 | 4.63±0.46 | 4.59±0.45 | 4.54±0.45 | 4.49±0.44 | 22.92 |
| 0.900–0.940 | Obs. | 6 | 10 | 6 | 8 | 6 | 36 |
| | Exp. | 4.03±0.42 | 3.99±0.42 | 3.94±0.42 | 3.89±0.41 | 3.85±0.41 | 19.70 |
| 0.940–0.970 | Obs. | 2 | 3 | 4 | 4 | 5 | 18 |
| | Exp. | 2.34±0.39 | 2.31±0.39 | 2.28±0.38 | 2.24±0.38 | 2.21±0.38 | 11.39 |
| 0.970–0.987 | Obs. | 2 | 2 | 3 | 1 | 3 | 11 |
| | Exp. | 1.04±0.27 | 1.03±0.27 | 1.00±0.27 | 0.96±0.26 | 0.93±0.26 | 4.96 |
| 0.987–0.995 | Obs. | 4 | 2 | 2 | 1 | 1 | 10 |
| | Exp. | 0.57±0.19 | 0.54±0.17 | 0.47±0.16 | 0.37±0.15 | 0.29±0.14 | 2.24 |
| 0.995–1.000 | Obs. | 2 | 1 | 1 | 0 | 1 | 5 |

CF channel results:

| | | $M_{\mu^+\mu^-}$ bin (GeV/$c^2$) | | | | | |
|---|---|---|---|---|---|---|---|
| $\nu_N$ bin | | 5.310–5.334 | 5.334–5.358 | 5.358–5.382 | 5.382–5.406 | 5.406–5.430 | Sum |
| | Exp. | 10.65±0.75 | 10.53±0.74 | 10.40±0.73 | 10.28±0.73 | 10.15±0.72 | 52.01 |
| 0.700–0.760 | Obs. | 8 | 13 | 12 | 16 | 10 | 59 |
| | Exp. | 11.74±0.80 | 11.61±0.79 | 11.47±0.78 | 11.33±0.77 | 11.19±0.76 | 57.33 |
| 0.760–0.850 | Obs. | 9 | 13 | 13 | 13 | 12 | 60 |
| | Exp. | 6.40±0.56 | 6.32±0.55 | 6.24±0.55 | 6.17±0.54 | 6.09±0.53 | 31.22 |
| 0.850–0.900 | Obs. | 3 | 4 | 3 | 2 | 1 | 13 |
| | Exp. | 4.88±0.48 | 4.82±0.48 | 4.76±0.47 | 4.70±0.47 | 4.64±0.46 | 23.80 |
| 0.900–0.940 | Obs. | 3 | 8 | 7 | 8 | 5 | 31 |
| | Exp. | 4.14±0.44 | 4.09±0.44 | 4.04±0.43 | 3.99±0.43 | 3.94±0.42 | 20.20 |
| 0.940–0.970 | Obs. | 5 | 7 | 2 | 1 | 2 | 17 |
| | Exp. | 2.89±0.46 | 2.85±0.46 | 2.82±0.45 | 2.78±0.45 | 2.74±0.44 | 14.07 |
| 0.970–0.987 | Obs. | 2 | 1 | 3 | 1 | 4 | 11 |
| | Exp. | 0.88±0.26 | 0.87±0.25 | 0.86±0.25 | 0.85±0.25 | 0.83±0.24 | 4.30 |
| 0.987–0.995 | Obs. | 4 | 0 | 1 | 0 | 1 | 6 |
| | Exp. | 0.82±0.37 | 0.81±0.36 | 0.79±0.36 | 0.75±0.35 | 0.72±0.35 | 3.89 |
| 0.995–1.000 | Obs. | 1 | 0 | 0 | 0 | 1 | 2 |

are the most sensitive searches for these decays at the Tevatron. For the $B^0 \to \mu^+\mu^-$ search, the observed data are in agreement with background-only expectations and an upper limit of $\mathcal{B}(B^0 \to \mu^+\mu^-) < 4.6 \times 10^{-9}$ ($3.8 \times 10^{-9}$) at 95% (90%) C.L. is set. For the $B_s^0 \to \mu^+\mu^-$ search, a small excess of events is observed relative to expectations from background-only sources with a $p$-value of 0.94% (6.8%) assuming only background (background plus SM signal). Using a fit to the data we measure $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.3^{+0.9}_{-0.7} \times 10^{-8}$ and the following bounds are set, $2.2\times10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.0\times10^{-8}$ and $0.8 \times 10^{-9} < \mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.4 \times 10^{-8}$ at 90% and 95% C.L., respectively. These measurements are consistent with our previous result, the recent results from other experiments, and the SM expectations.

These are the final CDF results for searches for these rare FCNC decays and are the culmination of a program spanning nearly two decades. The sensitivity of the $B_s^0 \to \mu^+\mu^-$ analysis reported here is better than the pioneering measurement by CDF [45] by a factor of over 800, which exceeds by a factor of 35 the gain in sensitivity expected by just increasing the sample size. The gains in search sensitivity originated from continual improvements to the analysis techniques employed. Those techniques are described in detail to afford future experiments performing similar searches or measurements the opportunity to benefit from this research. These results form the most sensitive search for $B_s^0(B^0) \to \mu^+\mu^-$ de-

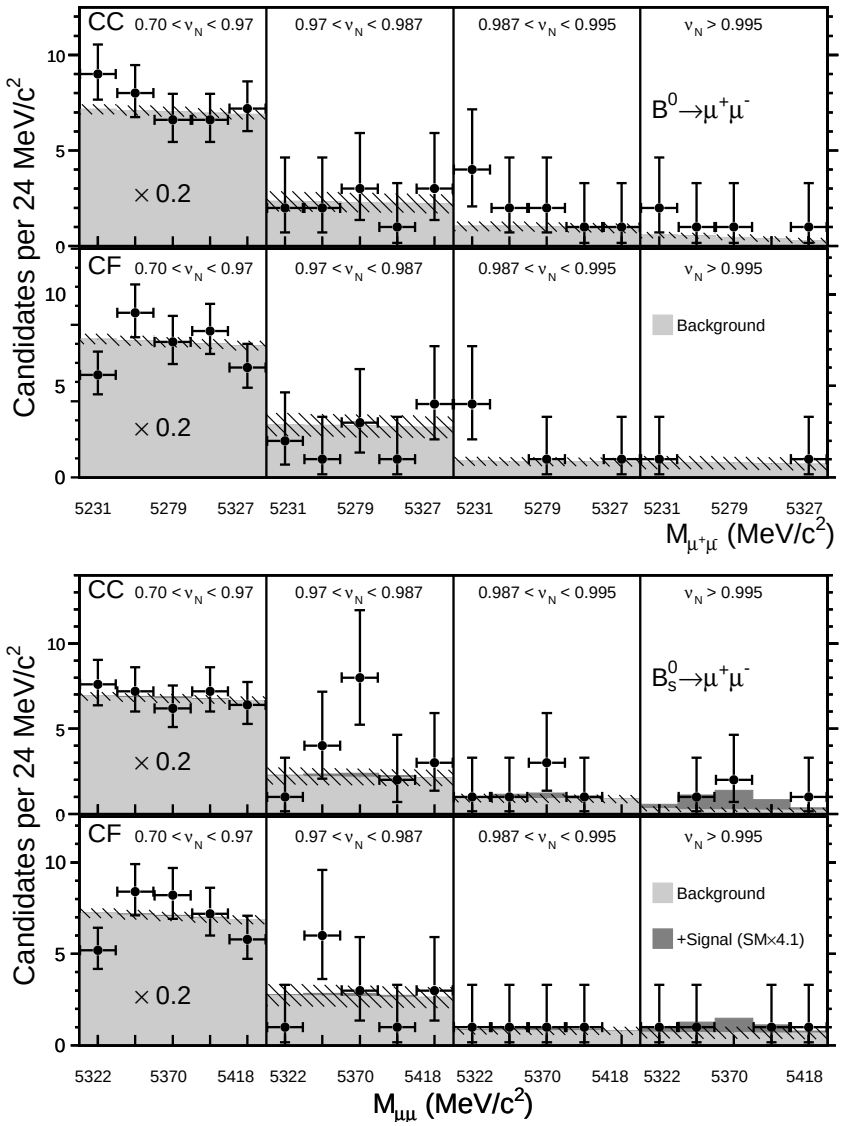


FIG. 23: The background estimate (light gray) and its systematic uncertainty (hashed area) is compared to the data (points), and their Poisson uncertainties (error bars on points) for the CC and CF channels for the $B^0 \to \mu^+\mu^-$ (top) and $B_s^0 \to \mu^+\mu^-$ (bottom) searches. Expectations that include signal at 4.1 times the SM rate (dark gray), corresponding to the fitted value from Fig. 25, are also shown for the $B_s^0$ results. The lowest five NN bins have been combined because the signal sensitivity is concentrated in the highest three NN bins.

cays performed previous to the LHC operational period and remain competitive with the most recent LHC results. The $B_s^0$ and $B^0$ results from all experiments are compatible with one another, indicate that there is no strong enhancement to the $B_s^0 \to \mu^+\mu^-$ decay rate, and strongly constrain new physics models that predict significant deviations from the standard model [14].

TABLE XIII: The results for the $B_s^0 \to \mu^+\mu^-$ search comparing the expected total backgrounds and their uncertainty (Exp.) to the number of observed events (Obs.) in each $(\nu_N, M_{\mu^+\mu^-})$ bin. For each $\nu_N$ bin, the sum over mass bins is also shown. The CC and CF channels are given separately.

CC channel results:

| | | $M_{\mu^+\mu^-}$ bin (GeV/$c^2$) | | | | | |
|---|---|---|---|---|---|---|---|
| $\nu_N$ bin | | 5.310–5.334 | 5.334–5.358 | 5.358–5.382 | 5.382–5.406 | 5.406–5.430 | Sum |
| | Exp. | 10.43±0.72 | 10.33±0.71 | 10.23±0.71 | 10.14±0.70 | 10.04±0.69 | 51.17 |
| 0.700–0.760 | Obs. | 13 | 8 | 7 | 6 | 7 | 41 |
| | Exp. | 11.03±0.74 | 10.93±0.74 | 10.82±0.73 | 10.72±0.72 | 10.62±0.72 | 54.13 |
| 0.760–0.850 | Obs. | 9 | 8 | 12 | 15 | 8 | 52 |
| | Exp. | 4.70±0.46 | 4.65±0.45 | 4.61±0.45 | 4.56±0.44 | 4.52±0.44 | 23.03 |
| 0.850–0.900 | Obs. | 6 | 8 | 5 | 6 | 5 | 30 |
| | Exp. | 4.50±0.45 | 4.45±0.44 | 4.41±0.44 | 4.37±0.43 | 4.33±0.43 | 22.05 |
| 0.900–0.940 | Obs. | 5 | 5 | 5 | 6 | 8 | 29 |
| | Exp. | 3.86±0.41 | 3.82±0.41 | 3.78±0.40 | 3.74±0.40 | 3.71±0.39 | 18.91 |
| 0.940–0.970 | Obs. | 5 | 7 | 2 | 3 | 4 | 21 |
| | Exp. | 2.22±0.38 | 2.19±0.37 | 2.17±0.37 | 2.15±0.37 | 2.12±0.36 | 10.84 |
| 0.970–0.987 | Obs. | 1 | 4 | 8 | 2 | 3 | 18 |
| | Exp. | 0.94±0.26 | 0.92±0.26 | 0.91±0.26 | 0.90±0.25 | 0.89±0.25 | 4.56 |
| 0.987–0.995 | Obs. | 1 | 1 | 3 | 1 | 0 | 6 |
| | Exp. | 0.31±0.14 | 0.26±0.14 | 0.25±0.14 | 0.24±0.14 | 0.23±0.14 | 1.29 |
| 0.995–1.000 | Obs. | 0 | 1 | 2 | 0 | 1 | 4 |

CF channel results:

| | | $M_{\mu^+\mu^-}$ bin (GeV/$c^2$) | | | | | |
|---|---|---|---|---|---|---|---|
| $\nu_N$ bin | | 5.310–5.334 | 5.334–5.358 | 5.358–5.382 | 5.382–5.406 | 5.406–5.430 | Sum |
| | Exp. | 10.18±0.72 | 10.05±0.71 | 9.93±0.7 | 9.80±0.69 | 9.68±0.68 | 49.64 |
| 0.700–0.760 | Obs. | 10 | 16 | 12 | 11 | 10 | 59 |
| | Exp. | 11.22±0.76 | 11.08±0.75 | 10.94±0.74 | 10.8±0.73 | 10.67±0.72 | 54.71 |
| 0.760–0.850 | Obs. | 8 | 13 | 9 | 13 | 4 | 47 |
| | Exp. | 6.11±0.54 | 6.03±0.53 | 5.96±0.52 | 5.88±0.52 | 5.81±0.51 | 29.79 |
| 0.850–0.90 | Obs. | 1 | 5 | 9 | 3 | 6 | 24 |
| | Exp. | 4.65±0.46 | 4.60±0.46 | 4.54±0.45 | 4.48±0.44 | 4.42±0.44 | 22.69 |
| 0.900–0.940 | Obs. | 6 | 2 | 8 | 5 | 4 | 25 |
| | Exp. | 3.95±0.42 | 3.90±0.42 | 3.85±0.41 | 3.80±0.41 | 3.75±0.40 | 19.25 |
| 0.940–0.970 | Obs. | 1 | 6 | 3 | 4 | 5 | 19 |
| | Exp. | 2.75±0.44 | 2.71±0.44 | 2.68±0.43 | 2.64±0.43 | 2.61±0.42 | 13.38 |
| 0.970–0.987 | Obs. | 1 | 6 | 3 | 1 | 3 | 14 |
| | Exp. | 0.83±0.25 | 0.82±0.24 | 0.81±0.24 | 0.80±0.24 | 0.79±0.23 | 4.06 |
| 0.987–0.995 | Obs. | 1 | 1 | 1 | 1 | 0 | 4 |
| | Exp. | 0.73±0.35 | 0.71±0.34 | 0.69±0.34 | 0.68±0.34 | 0.67±0.33 | 3.48 |
| 0.995–1.000 | Obs. | 1 | 1 | 0 | 1 | 1 | 4 |

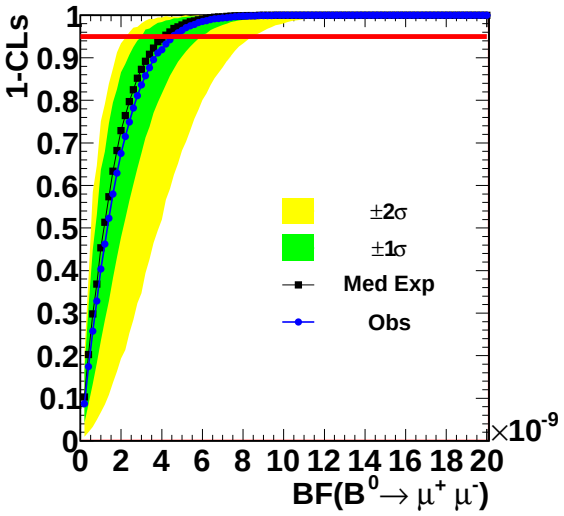


FIG. 24: Distribution of 1–CLs as a function of $\mathcal{B}(B^0 \to \mu^+\mu^-)$. The expected (observed) limit at 95% C.L. is determined by the intersection of the black (blue) points with the line at 1–CLs = 0.05. The shaded regions indicate the spread of limits obtained from simulated experiments due to fluctuations in the background alone.

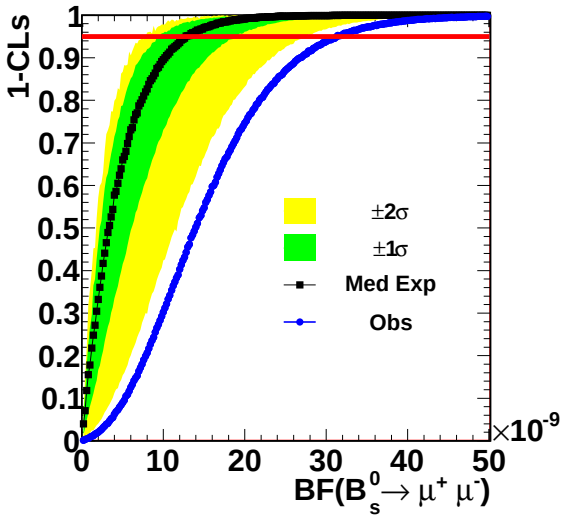


FIG. 26: Distribution of 1–CLs as a function of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The expected (observed) limit at 95% C.L. is determined by the intersection of the black (blue) points with the line at 1–CLs = 0.05. The shaded regions indicate the spread of limits obtained from simulated experiments due to fluctuations in the background alone.

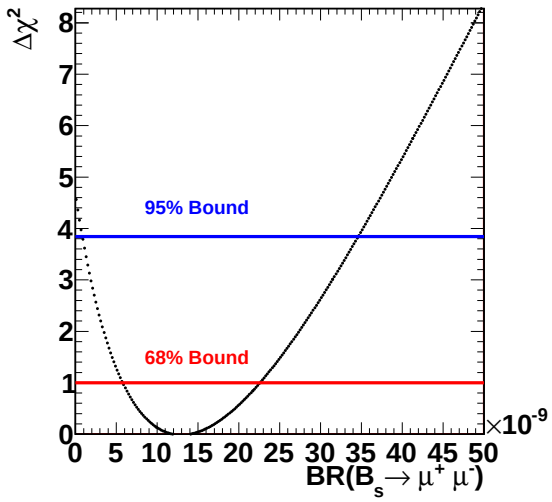


FIG. 25: $\Delta\chi^2$ distribution as a function of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$.

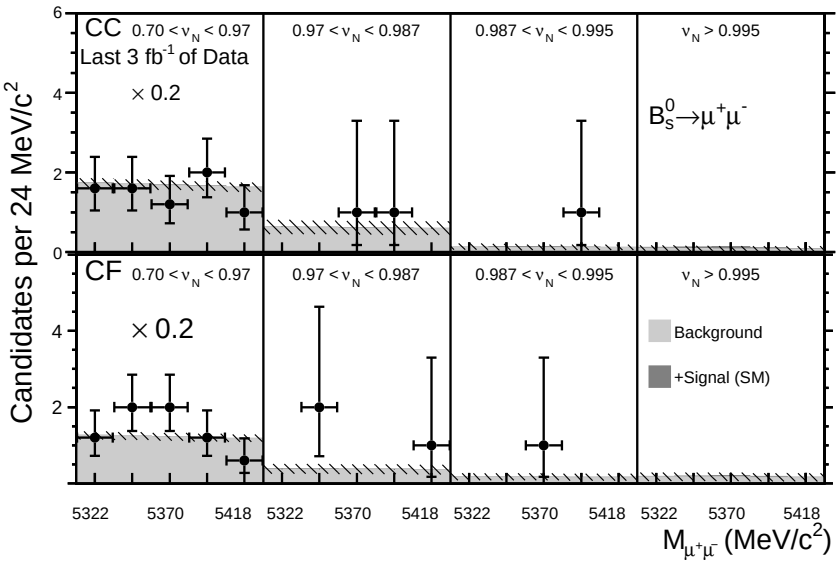


FIG. 27: For the $B_s^0 \to \mu^+\mu^-$ search, the background estimate (light gray) and its systematic uncertainty (hashed area) is compared to the data (points), and its Poisson uncertainty (error bars on points) for the CC and CF channels using only the last 3 fb$^{-1}$ of data added for this analysis. In these plots the lowest five NN bins have been combined because the signal sensitivity is concentrated in the highest three NN bins.

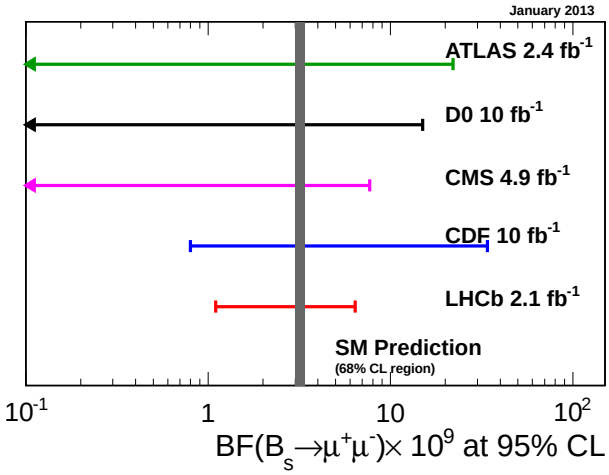


FIG. 28: A comparison of current limits from ATLAS [15], CMS [16], D0 [17], LHCb [19], and CDF (this paper) at 95% C.L. The SM prediction [2], at 68% C.L., is shown as a thin grey vertical band.

## XI. 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 World Class University Program, the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; the Academy of Finland; the Australian Research Council (ARC); and the EU community Marie Curie Fellowship contract 302103.

[1] S. L. Glashow, J. Iliopoulos, and L. Maiani, Phys. Rev. D **2**, 1285 (1970).
[2] A. J. Buras, J. Girrbach, D. Guadagnoli, and G. Isidori, Eur. Phys. J. C **72**, 2172 (2012).
[3] K.S. Babu and C. Kolda, Phys. Rev. Lett. **84**, 228 (2000).
[4] A. Dedes, H.K. Dreiner, and U. Nierste, Phys. Rev. Lett. **87**, 251804 (2001).
[5] S. Baek, P. Ko, and W. Y. Song, Phys. Rev. Lett. **89**, 271801 (2002).
[6] R. Arnowitt, B. Dutta, T. Kamon, and M. Tanaka, Phys. Lett. B **538**, 121 (2002).
[7] S. Baek, Y.G. Kim, and P. Ko, J. High Energy Phys. 02 (2005) 067.
[8] J. Ellis, K. A. Olive, Y. Santoso, and V. C. Spanos, J. High Energy Phys. 05 (2006) 063.
[9] I. Gogoladze, R. Khalid, Y. Mimura, and Q. Shafi, Phys. Rev. D **83**, 095007 (2011).
[10] M. Blanke, A. J. Buras, A. Poschenrieder, S. Recksiegel, C. Tarantino, *et al.*, J. High Energy Phys. 01 (2007) 066.
[11] M. Blanke, A. J. Buras, A. Poschenrieder, S. Recksiegel, C. Tarantino, S. Uhlig, A. Weiler, J. High Energy Phys. 03 (2009) 108.
[12] M. Bauer, S. Casagrande, U. Haisch, and M. Neubert, J. High Energy Phys. 09 (2010) 017.
[13] C. Beskidt, W. de Boer, D. I. Kazakov, F. Ratnikov, E. Ziebarth, and V. Zhukov, Phys. Lett. B **705**, 493 (2011).
[14] D. Straub, arXiv:1205.6094v1[hep-ph] (2012).
[15] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B **713**, 387 (2012).
[16] S. Chatrchyan *et al.* (CMS Collaboration), J. High Energy Phys. 04 (2012) 033.
[17] V. Abazov *et al.* (D0 Collaboration), arXiv:1301.4507v1[hep-ex] (2013).
[18] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett **107**, 191801 (2011).
[19] R. Aaij *et al.* (LHCb Collaboration), Phys. Rev. Lett. **110**, 021801 (2013).
[20] D. Acosta *et al.*, Phys. Rev. D **71**, 032001 (2005).
[21] F. Abe *et al.*, Nucl. Instrum. Methods A **271**, 387 (1988).
[22] F. Abe *et al.* (CDF Collaboration), Phys. Rev. D **50**, 2966 (1994).
[23] C. S. Hill *et al.* (CDF Collaboration), Nucl. Instrum. Methods A **530**, 1 (2004).
[24] A. Sill *et al.* (CDF Collaboration), Nucl. Instrum. Methods A **447**, 1 (2000).
[25] A. A. Affolder *et al.* (CDF Collaboration), Nucl. Instrum. Methods A **526**, 249 (2004).
[26] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **96**, 202001 (2006).
[27] G. Ascoli *et al.*, Nucl. Instrum. Methods A **268**, 33 (1988).
[28] E. J. Thomson *et al.*, IEEE Trans. Nucl. Sci. **49**, 1063 (2002).
[29] R. Downing, N. Eddy, L. Holloway, M. Kasten, H. Kim, J. Kraus, C. Marino, and K. Pitts *et al.*, Nucl. Instrum. Meth. A **570**, 36 (2007).
[30] T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna, and E. Norrbin, Comput. Phys. Commun. **135**, 238 (2001).
[31] D. Lange, Nucl. Instrum. and Methods A **462**, 152 (2001).
[32] R. Field, Acta Phys. Polon. B **36**, 167 (2005).
[33] E. Gerchtein and M. Paulini, eConf C **0303241**, TUMT005 (2003) [physics/0306031].
[34] G. Giurgiu, Ph. D. thesis, Carnegie Mellon University, 2005, FERMILAB-THESIS-2005-41.
[35] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **106**, 181802 (2011).
[36] K. Nakamura *et al.* (Particle Data Group), J. Phys. G **37**, 075021 (2010).
[37] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **106**, 161801 (2011).
[38] M. Feindt, arXiv:physics/0402093v1[physics.data-an].
[39] M. Feindt and U. Kerzel, Nucl. Instrum. and Methods A **559**, 190 (2006).
[40] D. Allspach *et al.*, IEEE Trans. Nucl. Sci. **52**, 2956 (2005)
[41] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. D **85**, 012009 (2012).
[42] T. Junk, Nucl. Instrum. Methods A **434**, 435 (1999); A.L. Read, J. Phys. G **28**, 2693 (2002).
[43] K. De Bruyn, R. Fleischer, R. Knegjens, P. Koppenburg, M. Merk, and N. Tuning, Phys. Rev. D **86**, 014027 (2012).
[44] Y.G. Aditya, K.J. Healey, and A. Petrov, arXiv:1212.4166v1[hep-ph].
[45] F. Abe *et al.* (CDF Collaboration), Phys. Rev. Lett. **76**, 4675 (1996).