Study of the $X^{\pm}(5568)$ state with semileptonic decays of the B_s^0 meson

V. M. Abazov,[31] B. Abbott,[67] B. S. Acharya,[25] M. Adams,[46] T. Adams,[44] J. P. Agnew,[41] G. D. Alexeev,[31] G. Alkhazov,[35] A. Alton,[56,a] A. Askew,[44] S. Atkins,[54] K. Augsten,[7] V. Aushev,[38] Y. Aushev,[38] C. Avila,[5] F. Badaud,[10] L. Bagby,[45] B. Baldin,[45] D. V. Bandurin,[74] S. Banerjee,[25] E. Barberis,[55] P. Baringer,[53] J. F. Bartlett,[45] U. Bassler,[15] V. Bazterra,[46] A. Bean,[53] M. Begalli,[2] L. Bellantoni,[45] S. B. Beri,[23] G. Bernardi,[14] R. Bernhard,[19] I. Bertram,[39] M. Besançon,[15] R. Beuselinck,[40] P. C. Bhat,[45] S. Bhatia,[58] V. Bhatnagar,[23] G. Blazey,[47] S. Blessing,[44] K. Bloom,[59] A. Boehnlein,[45] D. Boline,[64] E. E. Boos,[33] G. Borissov,[39] M. Borysova,[38,l] A. Brandt,[71] O. Brandt,[20] M. Brochmann,[75] R. Brock,[57] A. Bross,[45] D. Brown,[14] X. B. Bu,[45] M. Buehler,[45] V. Buescher,[21] V. Bunichev,[33] S. Burdin,[39,b] C. P. Buszello,[37] E. Camacho-Pérez,[28] B. C. K. Casey,[45] H. Castilla-Valdez,[28] S. Caughron,[57] S. Chakrabarti,[64] K. M. Chan,[51] A. Chandra,[73] E. Chapon,[15] G. Chen,[53] S. W. Cho,[27] S. Choi,[27] B. Choudhary,[24] S. Cihangir,[45,*] D. Claes,[59] J. Clutter,[53] M. Cooke,[45,k] W. E. Cooper,[45] M. Corcoran,[73,*] F. Couderc,[15] M.-C. Cousinou,[12] J. Cuth,[21] D. Cutts,[70] A. Das,[72] G. Davies,[40] S. J. de Jong,[29,30] E. De La Cruz-Burelo,[28] F. Déliot,[15] R. Demina,[63] D. Denisov,[45] S. P. Denisov,[34] S. Desai,[45] C. Deterre,[41,c] K. DeVaughan,[59] H. T. Diehl,[45] M. Diesburg,[45] P. F. Ding,[41] A. Dominguez,[59] A. Drutskoy,[32,q] A. Dubey,[24] L. V. Dudko,[33] A. Duperrin,[12] S. Dutt,[23] M. Eads,[47] D. Edmunds,[57] J. Ellison,[43] V. D. Elvira,[45] Y. Enari,[14] H. Evans,[49] A. Evdokimov,[46] V. N. Evdokimov,[34] A. Fauré,[15] L. Feng,[47] T. Ferbel,[63] F. Fiedler,[21] F. Filthaut,[29,30] W. Fisher,[57] H. E. Fisk,[45] M. Fortner,[47] H. Fox,[39] J. Franc,[7] S. Fuess,[45] P. H. Garbincius,[45] A. Garcia-Bellido,[63] J. A. García-González,[28] V. Gavrilov,[32] W. Geng,[12,57] C. E. Gerber,[46] Y. Gershtein,[60] G. Ginther,[45] O. Gogota,[38] G. Golovanov,[31] P. D. Grannis,[64] S. Greder,[16] H. Greenlee,[45] G. Grenier,[17] Ph. Gris,[10] J.-F. Grivaz,[13] A. Grohsjean,[15,c] S. Grünendahl,[45] M. W. Grünewald,[26] T. Guillemin,[13] G. Gutierrez,[45] P. Gutierrez,[67] J. Haley,[68] L. Han,[4] K. Harder,[41] A. Harel,[63] J. M. Hauptman,[52] J. Hays,[40] T. Head,[41] T. Hebbeker,[18] D. Hedin,[47] H. Hegab,[68] A. P. Heinson,[43] U. Heintz,[70] C. Hensel,[1] I. Heredia-De La Cruz,[28,d] K. Herner,[45] G. Hesketh,[41,f] M. D. Hildreth,[51] R. Hirosky,[74] T. Hoang,[44] J. D. Hobbs,[64] B. Hoeneisen,[9] J. Hogan,[73] M. Hohlfeld,[21] J. L. Holzbauer,[58] I. Howley,[71] Z. Hubacek,[7,15] V. Hynek,[7] I. Iashvili,[62] Y. Ilchenko,[72] R. Illingworth,[45] A. S. Ito,[45] S. Jabeen,[45,m] M. Jaffré,[13] A. Jayasinghe,[67] M. S. Jeong,[27] R. Jesik,[40] P. Jiang,[4,*] K. Johns,[42] E. Johnson,[57] M. Johnson,[45] A. Jonckheere,[45] P. Jonsson,[40] J. Joshi,[43] A. W. Jung,[45,o] A. Juste,[36] E. Kajfasz,[12] D. Karmanov,[33] I. Katsanos,[59] M. Kaur,[23] R. Kehoe,[72] S. Kermiche,[12] N. Khalatyan,[45] A. Khanov,[68] A. Kharchilava,[62] Y. N. Kharzheev,[31] I. Kiselevich,[32] J. M. Kohli,[23] A. V. Kozelov,[34] J. Kraus,[58] A. Kumar,[62] A. Kupco,[8] T. Kurča,[17] V. A. Kuzmin,[33] S. Lammers,[49] P. Lebrun,[17] H. S. Lee,[27] S. W. Lee,[52] W. M. Lee,[45,*] X. Lei,[42] J. Lellouch,[14] D. Li,[14] H. Li,[74] L. Li,[43] Q. Z. Li,[45] J. K. Lim,[27] D. Lincoln,[45] J. Linnemann,[57] V. V. Lipaev,[34,*] R. Lipton,[45] H. Liu,[72] Y. Liu,[4] A. Lobodenko,[35] M. Lokajicek,[8] R. Lopes de Sa,[45] R. Luna-Garcia,[28,g] A. L. Lyon,[45] A. K. A. Maciel,[1] R. Madar,[19] R. Magaña-Villalba,[28] S. Malik,[59] V. L. Malyshev,[31] J. Mansour,[20] J. Martínez-Ortega,[28] R. McCarthy,[64] C. L. McGivern,[41] M. M. Meijer,[29,30] A. Melnitchouk,[45] D. Menezes,[47] P. G. Mercadante,[3] M. Merkin,[33] A. Meyer,[18] J. Meyer,[20,i] F. Miconi,[16] N. K. Mondal,[25] M. Mulhearn,[74] E. Nagy,[12] M. Narain,[70] R. Nayyar,[42] H. A. Neal,[56] J. P. Negret,[5] P. Neustroev,[35] H. T. Nguyen,[74] T. Nunnemann,[22] J. Orduna,[70] N. Osman,[12] A. Pal,[71] N. Parashar,[50] V. Parihar,[70] S. K. Park,[27] R. Partridge,[70,e] N. Parua,[49] A. Patwa,[65,j] B. Penning,[40] M. Perfilov,[33] Y. Peters,[41] K. Petridis,[41] G. Petrillo,[63] P. Pétroff,[13] M.-A. Pleier,[65] V. M. Podstavkov,[45] A. V. Popov,[34] M. Prewitt,[73] D. Price,[41] N. Prokopenko,[34] J. Qian,[56] A. Quadt,[20] B. Quinn,[58] P. N. Ratoff,[39] I. Razumov,[34] I. Ripp-Baudot,[16] F. Rizatdinova,[68] M. Rominsky,[45] A. Ross,[39] C. Royon,[8] P. Rubinov,[45] R. Ruchti,[51] G. Sajot,[11] A. Sánchez-Hernández,[28] M. P. Sanders,[22] A. S. Santos,[1,h] G. Savage,[45] M. Savitskyi,[38] L. Sawyer,[54] T. Scanlon,[40] R. D. Schamberger,[64] Y. Scheglov,[35,*] H. Schellman,[69,48] M. Schott,[21] C. Schwanenberger,[41] R. Schwienhorst,[57] J. Sekaric,[53] H. Severini,[67] E. Shabalina,[20] V. Shary,[15] S. Shaw,[41] A. A. Shchukin,[34] O. Shkola,[38] V. Simak,[7] P. Skubic,[67] P. Slattery,[63] G. R. Snow,[59] J. Snow,[66] S. Snyder,[65] S. Söldner-Rembold,[41] L. Sonnenschein,[18] K. Soustruznik,[6] J. Stark,[11] N. Stefaniuk,[38] D. A. Stoyanova,[34] M. Strauss,[67] L. Suter,[41] P. Svoisky,[74] M. Titov,[15] V. V. Tokmenin,[31] Y.-T. Tsai,[63] D. Tsybychev,[64] B. Tuchming,[15] C. Tully,[61] L. Uvarov,[35] S. Uvarov,[35] S. Uzunyan,[47] R. Van Kooten,[49] W. M. van Leeuwen,[29] N. Varelas,[46] E. W. Varnes,[42] I. A. Vasilyev,[34] A. Y. Verkheev,[31] L. S. Vertogradov,[31] M. Verzocchi,[45] M. Vesterinen,[41] D. Vilanova,[15] P. Vokac,[7] H. D. Wahl,[44] M. H. L. S. Wang,[45] J. Warchol,[51,*] G. Watts,[75] M. Wayne,[51] J. Weichert,[21] L. Welty-Rieger,[48] M. R. J. Williams,[49,n] G. W. Wilson,[53] M. Wobisch,[54] D. R. Wood,[55] T. R. Wyatt,[41] Y. Xie,[45] R. Yamada,[45] S. Yang,[4] T. Yasuda,[45] Y. A. Yatsunenko,[31] W. Ye,[64] Z. Ye,[45] H. Yin,[45] K. Yip,[65] S. W. Youn,[45] J. M. Yu,[56] J. Zennamo,[62] T. G. Zhao,[41] B. Zhou,[56] J. Zhu,[56] M. Zielinski,[63] D. Zieminska,[49] and L. Zivkovic[14,p]

(D0 Collaboration)

[1]LAFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, RJ 22290, Brazil
[2]Universidade do Estado do Rio de Janeiro, Rio de Janeiro, RJ 20550, Brazil
[3]Universidade Federal do ABC, Santo André, SP 09210, Brazil
[4]University of Science and Technology of China, Hefei 230026, People's Republic of China
[5]Universidad de los Andes, Bogotá 111711, Colombia
[6]Charles University, Faculty of Mathematics and Physics, Center for Particle Physics, 116 36 Prague 1, Czech Republic
[7]Czech Technical University in Prague, 116 36 Prague 6, Czech Republic
[8]Institute of Physics, Academy of Sciences of the Czech Republic, 182 21 Prague, Czech Republic
[9]Universidad San Francisco de Quito, Quito 170157, Ecuador
[10]LPC, Université Blaise Pascal, CNRS/IN2P3, Clermont, F-63178 Aubière Cedex, France
[11]LPSC, Université Joseph Fourier Grenoble 1, CNRS/IN2P3, Institut National Polytechnique de Grenoble, F-38026 Grenoble Cedex, France
[12]CPPM, Aix-Marseille Université, CNRS/IN2P3, F-13288 Marseille Cedex 09, France
[13]LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, F-91898 Orsay Cedex, France
[14]LPNHE, Universités Paris VI and VII, CNRS/IN2P3, F-75005 Paris, France
[15]CEA Saclay, Irfu, SPP, F-91191 Gif-Sur-Yvette Cedex, France
[16]IPHC, Université de Strasbourg, CNRS/IN2P3, F-67037 Strasbourg, France
[17]IPNL, Université Lyon 1, CNRS/IN2P3, F-69622 Villeurbanne Cedex, France and Université de Lyon, F-69361 Lyon CEDEX 07, France
[18]III. Physikalisches Institut A, RWTH Aachen University, 52056 Aachen, Germany
[19]Physikalisches Institut, Universität Freiburg, 79085 Freiburg, Germany
[20]II. Physikalisches Institut, Georg-August-Universität Göttingen, 37073 Göttingen, Germany
[21]Institut für Physik, Universität Mainz, 55099 Mainz, Germany
[22]Ludwig-Maximilians-Universität München, 80539 München, Germany
[23]Panjab University, Chandigarh 160014, India
[24]Delhi University, Delhi-110 007, India
[25]Tata Institute of Fundamental Research, Mumbai-400 005, India
[26]University College Dublin, Dublin 4, Ireland
[27]Korea Detector Laboratory, Korea University, Seoul, 02841, Korea
[28]CINVESTAV, Mexico City 07360, Mexico
[29]Nikhef, Science Park, 1098 XG Amsterdam, the Netherlands
[30]Radboud University Nijmegen, 6525 AJ Nijmegen, the Netherlands
[31]Joint Institute for Nuclear Research, Dubna 141980, Russia
[32]Institute for Theoretical and Experimental Physics, Moscow 117259, Russia
[33]Moscow State University, Moscow 119991, Russia
[34]Institute for High Energy Physics, Protvino, Moscow region 142281, Russia
[35]Petersburg Nuclear Physics Institute, St. Petersburg 188300, Russia
[36]Institució Catalana de Recerca i Estudis Avançats (ICREA) and Institut de Física d'Altes Energies (IFAE), 08193 Bellaterra (Barcelona), Spain
[37]Uppsala University, 751 05 Uppsala, Sweden
[38]Taras Shevchenko National University of Kyiv, Kiev, 01601, Ukraine
[39]Lancaster University, Lancaster LA1 4YB, United Kingdom
[40]Imperial College London, London SW7 2AZ, United Kingdom
[41]The University of Manchester, Manchester M13 9PL, United Kingdom
[42]University of Arizona, Tucson, Arizona 85721, USA
[43]University of California Riverside, Riverside, California 92521, USA
[44]Florida State University, Tallahassee, Florida 32306, USA
[45]Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA
[46]University of Illinois at Chicago, Chicago, Illinois 60607, USA
[47]Northern Illinois University, DeKalb, Illinois 60115, USA
[48]Northwestern University, Evanston, Illinois 60208, USA
[49]Indiana University, Bloomington, Indiana 47405, USA
[50]Purdue University Calumet, Hammond, Indiana 46323, USA
[51]University of Notre Dame, Notre Dame, Indiana 46556, USA
[52]Iowa State University, Ames, Iowa 50011, USA
[53]University of Kansas, Lawrence, Kansas 66045, USA
[54]Louisiana Tech University, Ruston, Louisiana 71272, USA
[55]Northeastern University, Boston, Massachusetts 02115, USA
[56]University of Michigan, Ann Arbor, Michigan 48109, USA

[57]Michigan State University, East Lansing, Michigan 48824, USA
[58]University of Mississippi, University, Mississippi 38677, USA
[59]University of Nebraska, Lincoln, Nebraska 68588, USA
[60]Rutgers University, Piscataway, New Jersey 08855, USA
[61]Princeton University, Princeton, New Jersey 08544, USA
[62]State University of New York, Buffalo, New York 14260, USA
[63]University of Rochester, Rochester, New York 14627, USA
[64]State University of New York, Stony Brook, New York 11794, USA
[65]Brookhaven National Laboratory, Upton, New York 11973, USA
[66]Langston University, Langston, Oklahoma 73050, USA
[67]University of Oklahoma, Norman, Oklahoma 73019, USA
[68]Oklahoma State University, Stillwater, Oklahoma 74078, USA
[69]Oregon State University, Corvallis, Oregon 97331, USA
[70]Brown University, Providence, Rhode Island 02912, USA
[71]University of Texas, Arlington, Texas 76019, USA
[72]Southern Methodist University, Dallas, Texas 75275, USA
[73]Rice University, Houston, Texas 77005, USA
[74]University of Virginia, Charlottesville, Virginia 22904, USA
[75]University of Washington, Seattle, Washington 98195, USA

(Received 29 December 2017; published 18 May 2018)

We present a study of the $X^{\pm}(5568)$ using semileptonic decays of the B_s^0 meson using the full run II integrated luminosity of 10.4 fb^{-1} in proton-antiproton collisions at a center of mass energy of 1.96 TeV collected with the D0 detector at the Fermilab Tevatron Collider. We report evidence for a narrow structure, $X^{\pm}(5568)$, in the decay sequence $X^{\pm}(5568) \to B_s^0\pi^{\pm}$ where $B_s^0 \to \mu^{\mp}D_s^{\pm}$X, $D_s^{\pm} \to \phi\pi^{\pm}$ which is consistent with the previous measurement by the D0 Collaboration in the hadronic decay mode, $X^{\pm}(5568) \to B_s^0\pi^{\pm}$ where $B_s^0 \to J/\psi\phi$. The mass and width of this state are measured using a combined fit of the hadronic and semileptonic data, yielding $m = 5566.9^{+3.2}_{-3.1}(\text{stat})^{+0.6}_{-1.2}(\text{syst})$ MeV$/c^2$, $\Gamma = 18.6^{+7.9}_{-6.1}(\text{stat})^{+3.5}_{-3.8}(\text{syst})$ MeV$/c^2$ with a significance of 6.7σ.

DOI: 10.1103/PhysRevD.97.092004

I. INTRODUCTION

Since the creation of the quark model [1,2] it was understood that exotic mesons containing more than one quark-antiquark pair are possible. However, for exotic mesons containing only the up, down and strange quarks it has been difficult to make a definitive experimental case for such exotic states, although some persuasive arguments have been made (for recent comprehensive discussions of

[*]Deceased.
[a]Visitor from Augustana College, Sioux Falls, SD 57197, USA.
[b]Visitor from University of Liverpool, Liverpool L69 3BX, United Kingdom.
[c]Visitor from Deutshes Elektronen-Synchrotron (DESY), Notkestrasse 85, Germany.
[d]Visitor from CONACyT, M-03940 Mexico City, Mexico.
[e]Visitor from SLAC, Menlo Park, CA 94025, USA.
[f]Visitor from University College London, London WC1E 6BT, United Kingdom.
[g]Visitor from Centro de Investigacion en Computacion—IPN, CP 07738 Mexico City, Mexico.
[h]Visitor from Universidade Estadual Paulista, São Paulo, SP 01140, Brazil.
[i]Visitor from Karlsruher Institut für Technologie (KIT)—Steinbuch Centre for Computing (SCC), D-76128 Karlsruhe, Germany.
[j]Visitor from Office of Science, U.S. Department of Energy, Washington, D.C. 20585, USA.
[k]Visitor from American Association for the Advancement of Science, Washington, D.C. 20005, USA.
[l]Visitor from Kiev Institute for Nuclear Research (KINR), Kyiv 03680, Ukraine.
[m]Visitor from University of Maryland, College Park, MD 20742, USA.
[n]Visitor from European Organization for Nuclear Research (CERN), CH-1211 Geneva, Switzerland.
[o]Visitor from Purdue University, West Lafayette, IN 47907, USA.
[p]Visitor from Institute of Physics, Belgrade, Belgrade, Serbia.
[q]Visitor from P.N. Lebedev Physical Institute of the Russian Academy of Sciences, 119991, Moscow, Russia.

exotic hadrons containing both light and heavy quarks, see Refs. [3–6]). Multiquark states that contain heavy quarks can be more recognizable owing to the distinctive decay structure of heavy quark hadrons. The 2003 discovery by the Belle experiment [7] of the $X(3872)$ in the channel $B^\pm \to K^\pm X(\to \pi^+\pi^- J/\psi)$ was the first candidate exotic meson in which heavy flavor quarks participate. This state was subsequently confirmed in several production and decay modes by ATLAS [8], *BABAR* [9], BES III [10], CDF [11], CMS [12], D0 [13] and LHCb [14] Collaborations. Several additional four-quark candidate exotic mesons have since been found, though in many cases not all experiments have been able to confirm their existence.

Four-quark mesons can be generically categorized as either "molecular states" or tetraquark states of a diquark and an anti-diquark. In the example of the $X(3872)$, a molecular state interpretation would be a colorless D^0 $(c\bar{u})$ and a colorless $\bar{D}^{*0}$ $(u\bar{c})$ in a loosely bound state. Such a state would be expected to lie close in mass to the $D^0\bar{D}^{*0}$ threshold. The tetraquark mode of a colored diquark (cu) and colored anti-diquark $(\bar{c}\,\bar{u})$ is more strongly bound by the exchange of gluons and would be expected to have a mass somewhat below the $D^0\bar{D}^{*0}$ threshold. In many cases, interpretations of four-quark mesons as pure molecular or tetraquark states are difficult, and more complex mechanisms may be required [4–6]. The firm identification of multiquark mesons and baryons and the study of their properties are of importance for further understanding of nonperturbative QCD.

Recently the D0 Collaboration presented evidence for a new four-quark candidate that decays to $B_s^0\pi^\pm$ where B_s^0 decays to $J/\psi\phi$ [15]. This system would be composed of two quarks and two antiquarks of four different flavors: b, s, u, d, with either a molecular constitution as a loosely bound B_d^0 and $K^\pm$ system or a tightly bound tetraquark such as $(bd) - (\bar{s}\,\bar{u})$, $(bu) - (\bar{s}\,\bar{d})$, $(su) - (\bar{b}\,\bar{d})$, or $(sd) - (\bar{b}\,\bar{u})$ (because the B_s^0 meson is fully mixed, the exact quark antiquark composition cannot be determined). The mass of $X^\pm(5568)$ is about 200 MeV$/c^2$ below the $B_d^0K^\pm$ threshold, thus disfavoring a $B_d^0 - K^\pm$ molecular interpretation.

The $X^\pm(5568)$ was previously reported [15] with a significance of 5.1σ (including systematic uncertainties and the look-elsewhere effect [16]) in the decay $X^\pm(5568) \to B_s^0(J/\psi\phi)\pi^\pm$ in proton-antiproton collisions at a center of mass energy of 1.96 TeV. The ratio of the number of B_s^0 that are from the decay of the $X^\pm(5568)$ to all B_s^0 produced was measured to be $[8.6 \pm 1.9(\text{stat}) \pm 1.4(\text{syst})]\%$. In order to reduce the background, a selection was imposed on the angle between the B_s^0 and $\pi^\pm$ (the "cone cut", $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ [17]). Without the cone cut the significance was found to be 3.9σ. In addition to increasing the signal-to-background ratio this cone cut limits backgrounds, such as possible excited states of the B_c meson, that are not included in the available simulations. Multiple checks were carried out to ensure that the cone cut did not create an anomalous signal [15]. Varying the cone cut from $\Delta R_{\text{max}} = 0.2$ to 0.5 gave stable fitted masses and resulted in no unexpected changes in the result. The invariant mass spectra of the B_s^0 candidates and charged tracks with kaon or proton mass hypotheses were checked, and no resonant enhancements in these distributions were found. The invariant mass distribution of $B_d^0\pi^\pm$ was also examined with no unexpected resonances or reflections found. Subsequent analyses by the LHCb Collaboration [18] and by the CMS Collaboration [19] have not found evidence for the $X^\pm(5568)$ in proton-proton interactions at $\sqrt{s} = 7$ and 8 TeV. The CDF Collaboration has recently reported no evidence for $X^\pm(5568)$ in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV [20] with different kinematic coverage than that of Ref. [15].

In this article, we present a study of the $X^\pm(5568)$ in the decay to $B_s^0\pi^\pm$ using semileptonic B_s^0 decays, $B_s^0 \to \mu^+D_s^-\text{X}$, where $D_s^- \to \phi\pi^-$, $\phi \to K^+K^-$, using the full run II integrated luminosity of 10.4 fb^{-1} in proton-antiproton collisions at a center of mass energy of 1.96 TeV collected with the D0 detector at the Fermilab Tevatron Collider. Charge conjugate states are assumed. Here X includes the unseen neutrino and possibly a photon or π^0 from a D_s^* decay or other hadrons from the B_s^0 decay. The decay process is illustrated in Fig. 1. The semileptonic decay channel has a higher branching fraction than the hadronic channel ($B_s^0 \to J/\psi\phi$). However the presence of the unmeasured neutrino in the final state deteriorates the mass resolution of the signal. Still, a good mass resolution for the $X^\pm(5568)$ can be obtained in the semileptonic channel for events with a large invariant mass of the $\mu^+D_s^-$ system, yielding a comparable number of selected B_s^0 candidates in the two channels. The backgrounds in the semileptonic channel are independent of, but somewhat larger than, those in the hadronic channel. The character of possible reflections of other resonant structures is quite different in the semileptonic and hadronic channels.

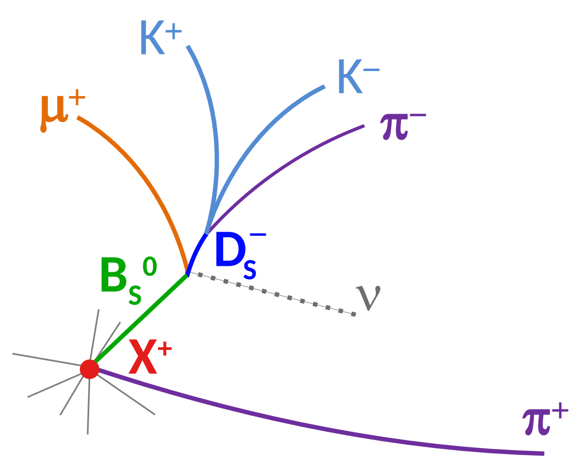


FIG. 1. An illustration of the decay $X^+(5568) \to B_s^0\pi^+$ where $B_s^0 \to \mu^+D_s^-\text{X}$ in the plane perpendicular to the beam.

Thus observation of the $X^{\pm}(5568)$ in the semileptonic decay channel enables an independent confirmation of its existence. We report here the results of the search for the $X^{\pm}(5568)$ in the semileptonic channel, as well as a combination of the results in the hadronic and semileptonic channels.

II. D0 DETECTOR

The detector components most relevant to this analysis are the central tracking and the muon systems. The D0 detector has a central tracking system consisting of a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located within a 2 T superconducting solenoidal magnet [21,22]. The SMT has a design optimized for tracking and vertexing for pseudorapidity of $|\eta| < 3$. For charged particles, the resolution on the distance of closest approach as provided by the tracking system is approximately 50 μm for tracks with $p_T \approx 1$ GeV/c, where p_T is the component of the momentum perpendicular to the beam axis. It improves asymptotically to 15 μm for tracks with $p_T > 10$ GeV/c. Preshower detectors and electromagnetic and hadronic calorimeters surround the tracker. A muon system, positioned outside the calorimeter, covering $|\eta| < 2$ consists of a layer of tracking detectors and scintillation trigger counters in front of 1.8 T iron toroidal magnets, followed by two similar layers after the toroids [23].

III. EVENT RECONSTRUCTION AND SELECTION

The $B_s^0 \to \mu^+ D_s^- X$ selection requirements have been chosen to optimize the mass resolution of the $B_s^0\pi^+$ system and to minimize background from random combinations of tracks from muons and charged hadrons. The selection criteria are based on those used in Ref. [24] with the cut on the B_s^0 isolation removed and have been selected by maximizing the significance of the signal.

The data were collected with a suite of single and dimuon triggers (approximately 95% of the sample is recorded using single muon triggers). The selection and reconstruction of $\mu^+ D_s^-$ decays requires tracks with at least two hits in both the CFT and SMT.

The muon is required to have hits in at least two layers of the muon system, with segments reconstructed both inside and outside the toroid. The muon track segment is required to be matched to a track found in the central tracking system that has transverse momentum $3 < p_T < 25$ GeV/c.

The $D_s^- \to \phi\pi^-$; $\phi \to K^+K^-$ decay is selected as follows. The two particles from the ϕ decay are assumed to be kaons and are required to have $p_T > 1.0$ GeV/c, opposite charge and an invariant mass $1.012 < m(K^+K^-) < 1.03$ GeV/c^2. The charge of the third particle, assumed to be a pion, has to be opposite to that of the muon. This particle is required to have transverse momentum $0.5 < p_T < 25$ GeV/c. The mass of the three particles must satisfy $1.91 < m(K^+K^-\pi^-) < 2.03$ MeV/c^2. The three tracks are combined to form a common D_s^- decay vertex using the algorithm described in Ref. [25]. The D_s^- vertex is required to be displaced from the $p\bar{p}$ primary interaction vertex (PV) in the transverse plane with a significance of at least three standard deviations. The cosine of the angle between the D_s^- momentum and the vector from the PV to the D_s^- decay vertex is required to be greater than 0.9.

The trajectories of the muon and D_s^- candidate are required to be consistent with originating from a common vertex (assumed to be the B_s^0 semileptonic decay vertex). The cosine of the angle between the combined $\mu^+ D_s^-$ transverse momentum, an approximation of the B_s^0 direction, and the direction from the PV to the B_s^0 decay vertex has to be greater than 0.95. The B_s^0 decay vertex has to be displaced from the PV in the transverse plane with a significance of at least four standard deviations. The transverse momentum of the $\mu^+ D_s^-$ system is required to satisfy the condition $p_T > 10$ GeV/c to suppress backgrounds. To minimize the effect of the neutrino in the final state the effective mass is limited to $4.5\ \mathrm{GeV}/c^2 < m(\mu^+ D_s^-) < m(B_s^0)$.

The impact parameters (IP) [26] with respect to the PV of the four tracks from the B_s^0 decay are required to satisfy the following criteria: the two-dimensional (2D) IPs of the tracks of the muon and the pion from the D_s^- decay are required to be at least 50 μm to reject tracks emerging promptly from the PV (this requirement is not applied to the tracks associated with the charged kaons since the mass requirements provide satisfactory background suppression). The three-dimensional (3D) IPs of all four tracks are required to be less than 2 cm to suppress combinations with tracks emerging from different $p\bar{p}$ vertices reconstructed in the same beam crossing.

The $m(K^+K^-\pi^{\pm})$ distribution of the candidates that pass these cuts [except $1.91 < m(K^+K^-\pi^-) < 2.03$ MeV/c^2] is shown in Fig. 2, where the invariant mass distribution in data is compared to a fit using a function which includes three terms: a second order polynomial used to describe combinatorial background, a Gaussian used to model the D^- peak, and a double Gaussian with similar, but different masses and widths used to model the D_s^- peak.

The selection criteria for the pion in the $B_s^0\pi^{\pm}$ combination have been chosen to match those used in the hadronic analysis. The track representing the pion is required to have transverse momentum $0.5 < p_T < 25$ GeV/c (the upper limit is applied to reduce background). The pion and the B_s^0 candidate are combined to form a vertex that is consistent with the PV. The pion is required to be associated with the PV and have a 2D IP of at most 200 μm and a 3D IP that is less than 0.12 cm. Events with more than 20 $B_s^0\pi^{\pm}$ candidates are rejected. The most likely number of candidates per event is 5.1, and only about 0.1% of the events have more than 20 candidates

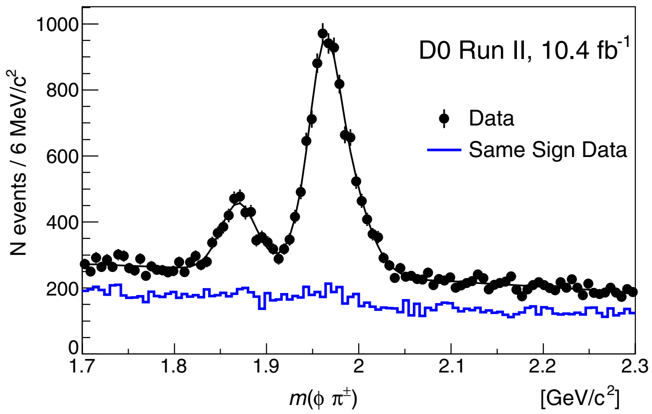


FIG. 2. The $K^+K^-\pi^\pm$ invariant mass distribution for the $\mu^\pm\phi\pi^\mp$ sample (right sign) with the solid curve representing the fit. The lower mass peak is due to the decay $D^\pm \to \phi\pi^\pm$ and the second peak is due to the $D_s^\pm$ decay. The blue histogram below the data points is the invariant mass distribution for the same-sign sample, $\mu^\pm\phi\pi^\pm$.

per event. To improve the resolution of the invariant mass of the $B_s^0\pi^\pm$ system we define the invariant mass as $m(B_s^0\pi^\pm) = m(\mu^+D_s^-\pi^\pm) - m(\mu^+D_s^-) + m(B_s^0)$ where $m(B_s^0) = 5.3667\ \text{GeV}/c^2$ [27]. We study the mass distribution in the range $5.506 < m(B_s^0\pi^\pm) < 5.906\ \text{GeV}/c^2$. When using the hadronic data from Ref. [15] in this paper we use the same mass range as the semileptonic data instead of the slightly shifted mass range used in the original analysis [$5.5 < m(B_s^0\pi^\pm) < 5.9\ \text{GeV}/c^2$]. The semileptonic data are studied with and without a cone cut which is used to suppress background, in which the angle between the $\mu^+D_s^-$ system and $\pi^\pm$ is required to satisfy $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$. The resulting invariant mass distributions for the semileptonic channel are shown in Fig. 3.

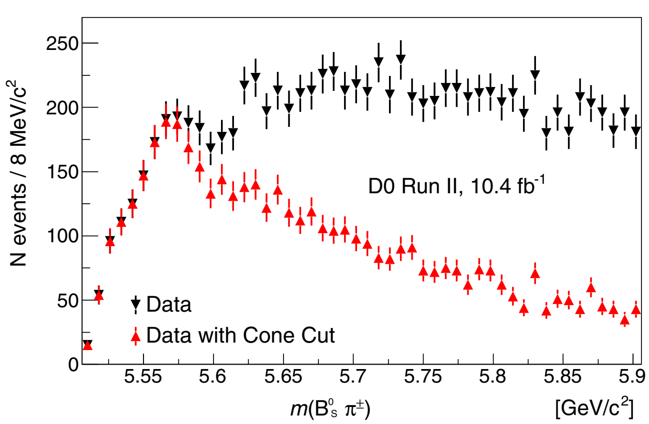


FIG. 3. The $m(B_s^0\pi^\pm)$ distribution for the semileptonic data with (red upward triangles) and without (black downward triangles) the cone cut. Below 5.56 GeV/c^2 the red and black points have the same values.

The selection cuts and resulting kinematics for the hadronic and semileptonic channels are quite similar. The requirement that muons be seen outside the toroids means that the minimum p_T for the J/ψ in the hadronic channel is about 4 GeV and about 3 GeV for the single muon in the semileptonic channel. The minimum p_T for the additional pion is 0.5 GeV for both the hadronic and semileptonic channels. For both channels, we require the minimum $p_T(B_s^0\pi)$ to be greater than 10 GeV and the average $p_T(B_s^0\pi)$ for events with $m(B_s\pi) \approx 5.5$ GeV is ≈ 17 GeV. For both channels the $B_s^0\pi$ candidates are in the range of $-2 < \eta < 2$, and more than half of the events have a muon with $|\eta| > 1$.

IV. MONTE CARLO SIMULATION, BACKGROUND MODELING AND PARAMETRIZATION

Monte Carlo (MC) samples are generated using the PYTHIA [28] event generator, modified to use EVTGEN [29] for the decay of hadrons containing b or c quarks. The generated events are processed by the full detector simulation chain. Data events recorded in random beam crossings are overlaid on the MC events to simulate the effect of additional collisions in the same or nearby bunch crossings. The resulting events are then processed with the same reconstruction and selection algorithms as used for data events.

The MC sample for $X^\pm(5568)$ signal is generated by modifying the mass of the $B^\pm$ meson and forcing it to decay to $B_s^0\pi^\pm$ using an isotropic S-wave decay model. The $X^\pm(5568)$ is simulated with zero width and zero lifetime. The resulting $K^+K^-\pi^-$ and $B_s^0\pi^\pm$ invariant mass distributions are shown in Fig. 4 with all selection requirements.

The signal component of the $K^+K^-\pi^\pm$ invariant mass distribution (Fig. 4a) is modeled by two Gaussian functions and the background by a second-order polynomial. The signal of the $m(B_s^0\pi^\pm)$ distribution (Fig. 4b) is well modeled with a single Gaussian and the background with a third-order polynomial times an exponential. Using the results of these fits the reconstruction efficiency of the charged pion in the decay $X^\pm(5568) \to B_s^0\pi^\pm$ is $[32.0 \pm 1.8(\text{stat}) \pm 1.6(\text{syst})]\%$ for $p_T(\mu^+D_s^-) > 10\ \text{GeV}/c$ where the systematic uncertainty represents the expected differences between the reconstruction efficiencies for low-momentum tracks in the data and MC simulation.

It is not possible to create a model of the background that is based only on data. Since the $X^\pm(5568)$ decays to B_s^0 mesons, any data sample that includes B_s^0 decays will also include the signal and is unsuitable for modeling the background. Hence, we use MC-generated B_s^0 events that result from known particles that have decays that include a B_s^0 in the decay chain, combined with data events where the muon has the same sign as the D_s^- candidate (SS events). MC event generators do not include all possible states as in

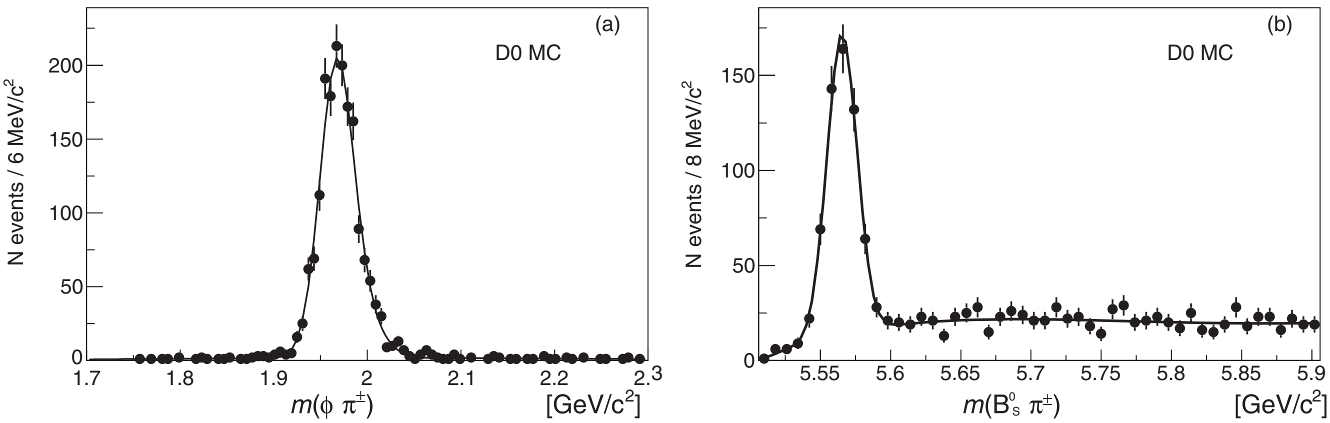


FIG. 4. MC simulation of $X^{\pm}(5568) \to B^0_s\pi^{\pm}$ where $B^0_s \to \mu^+ D^-_s X$ and the width of the $X^{\pm}(5568)$ is zero. The invariant mass distributions (a) $m(K^+K^-\pi^-)$ and (b) $m(B^0_s\pi^+)$ are shown. The background in the $m(B^0_s\pi^+)$ distribution is produced by the combination of a random charged track with the B^0_s meson.

many cases they have not been experimentally observed. For example, $b\bar{c}$ resonances decaying to B^0_s mesons could contribute to our sample.

There are two distinct sources of background in this analysis. The first occurs when an $X^{\pm}(5568)$ candidate is reconstructed from a real μ^+ and D^-_s together with a random charged track. This background is modeled using MC samples.

The background MC sample is generated using the PYTHIA inclusive heavy flavor production model, and events are selected that contain at least one muon and a $D^{\mp}_s \to \phi\pi^{\mp}$ decay where $\phi \to K^+K^-$. To correct for the difference in lifetimes in the MC simulation and data, a weighting is applied to all nonprompt events in the simulation, based on the generated lifetime of the B candidate, to give the world-average B hadron lifetimes [27]. To correct for the effects of the trigger selection and the reconstruction in data, we also weight each MC event so that the transverse momenta of the reconstructed muon and the $\mu^+D^-_s$ system agree with those in the data. The p_T distribution of the $B^0_s\pi^{\pm}$ system is altered significantly by the weighting as shown in Fig. 5(a). However, the effect is relatively small for the $B^0_s\pi^{\pm}$ mass distribution as seen in Fig. 5(b).

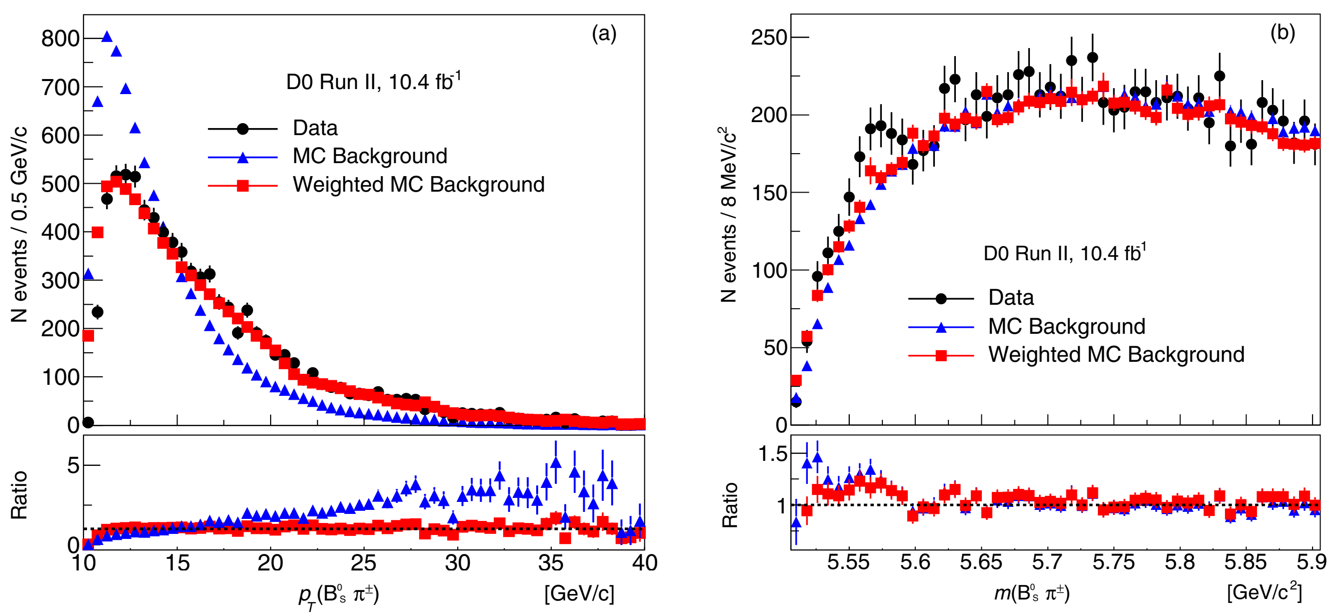


FIG. 5. The MC background distribution, without the cone cut, before and after weighting is compared with data (black points). The unweighted MC simulation is in blue, and the weighted is in red. The (a) $p_T(B^0_s\pi^{\pm})$ and (b) invariant mass distributions $m(B^0_s\pi^{\pm})$ are shown. The excess in the data around $m(B^0_s\pi^{\pm}) = 5565$ MeV/c is the $X^{\pm}(5568)$ signal. The lower panels show the ratio between the data and corresponding MC simulation.

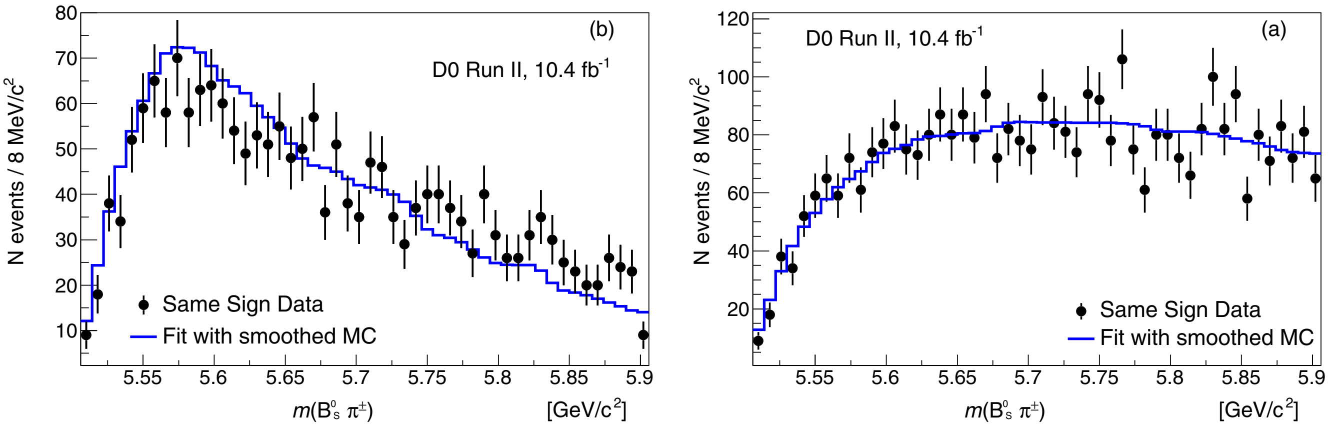


FIG. 6. The comparison of the $m(B_s^0\pi^\pm)$ background only distributions (a) with the cone cut and (b) without the cone cut, obtained using the weighted MC (histogram) and from the same sign data samples (points with error bars). The fluctuations in the number of MC events with the cone cut are due to the weighting procedure and the size of the sample.

The second source of background is the combinatorial background that occurs when a $X^\pm(5568)$ candidate is reconstructed from a spurious D_s^- candidate formed from three random charged tracks that form a vertex. This background is modeled using data events where the muon has the same sign as the D_s^- candidate (SS events).

In Fig. 6(b) we compare the reweighted MC background simulation, smoothed using one iteration of the 353QH algorithm [30], with the SS data for the no cone cut case. These two backgrounds are in good agreement since the χ^2 between them is 50 for 50 bins. We therefore choose to use the MC background shape only, for the data without the cone cut. In Fig. 6(a) we make the same comparison for the data with the cone cut. In this case, $\chi^2 = 77$ for the 50 bins, and we therefore need to model the background shape with a combination of the MC and SS backgrounds.

To construct the background sample for the data with the cone cut the fraction of MC and SS backgrounds need to be determined. This is found by fitting the data with a combination of the MC and SS with the fraction of MC events as a free parameter in the sideband mass range $5.506 < m(B_s^0\pi^\pm) < 5.55$ and $5.650 < m(B_s^0\pi^\pm) < 5.906\ \mathrm{GeV}/c^2$. The best agreement is found when the MC fraction is $(62 \pm 2)\%$.

We choose the background parametrization for the invariant mass distribution, both with and without the cone cut, to be

$$F_{\mathrm{bgr}}(m) = (C_1 m_0 + C_2 m_0^2 + C_3 m_0^3 + C_4 m_0^4) \times \exp(C_5 m_0 + C_7 m_0^2), \quad (1)$$

where $m = m(B_s^0\pi^\pm)$, $m_0 = m - m_{\mathrm{th}}$ and $m_{\mathrm{th}} = 5.5063\ \mathrm{GeV}/c^2$ is the mass threshold. Our baseline choice of Eq. (1) gives an equivalently good description of the background as that used in Ref. [15] [Eq. (2)]. It has the advantages of having one fewer parameter and being zero at the mass threshold.

Three alternative parametrizations are used to model the background. The first is that used in Ref. [15],

$$F_{\mathrm{bgr}}(m) = (C_1 + C_2 m_\Delta^2 + C_3 m_\Delta^3 + C_4 m_\Delta^4) \times \exp(C_5 + C_6 m_\Delta + C_7 m_\Delta^2), \quad (2)$$

where $m_\Delta = m - \Delta$ and $\Delta = 5.500\ \mathrm{GeV}/c^2$. The second is the ARGUS function [31] which is specifically constructed to describe background near a threshold,

$$F_{\mathrm{bgr}}(m) = m\left(\frac{m^2}{m_{\mathrm{th}}^2} - 1\right)^{C_1} \exp(C_2 m). \quad (3)$$

The third alternative model used to fit the background is the MC histogram (or combined MC and SS data) smoothed using one iteration of the 353QH algorithm [30].

The ARGUS function is not used as an alternate parametrization in the semileptonic data with the cone cut, because the fit to background is strongly disfavored (the χ^2 of the fit to the MC background is 145 compared with approximately 50 for the alternate functions). The χ^2 per number of degrees of freedom (ndf) for the four representations of the background are shown in Table I.

TABLE I. Fit results for different parametrizations to the background model.

Background function	χ^2/ndf Cone cut	χ^2/ndf No cone cut
Eq. (1)	$51.0/(50-6) = 1.2$	$48.1/(50-6) = 1.1$
Eq. (2)	$42.9/(50-7) = 1.0$	$48.1/(50-7) = 1.1$
Eq. (3)	$145/(50-2) = 3.0$	$38.3/(50-2) = 0.8$
Smoothed background	$33.8/(50-1) = 0.7$	$30.9/(50-1) = 0.6$

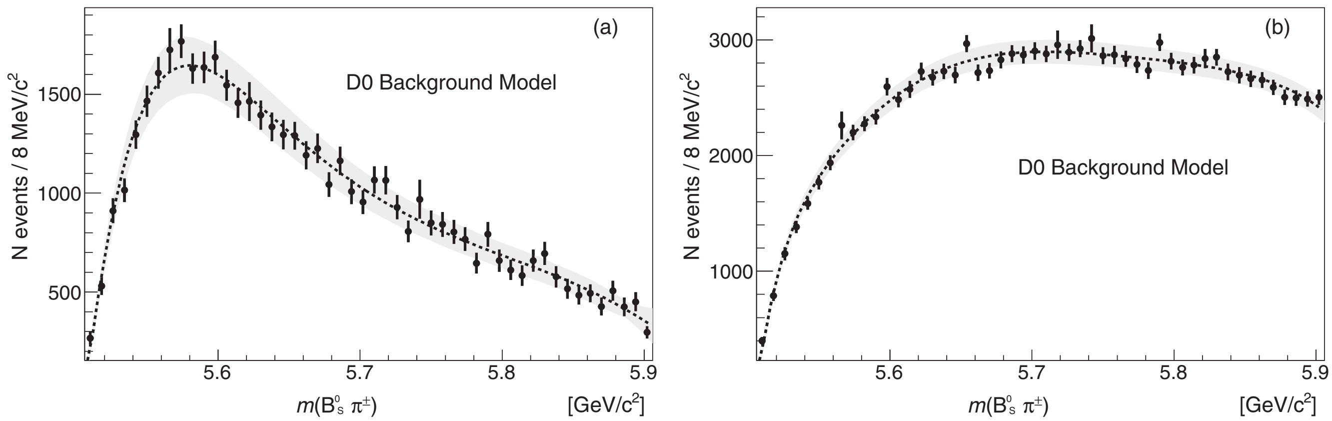


FIG. 7. The background model produced according to the procedure described in the text is shown along with background function (1) (dotted line) (a) with and (b) without the cone cut. The gray band shows the systematic uncertainties on the background model (see Sec. VII D).

We choose the background description of Eq. (1) as the baseline. The alternative functions and the smoothed MC are used to estimate the systematic uncertainty on the background shape. The $m(B_s^0\pi^{\pm})$ background model distribution along with the fit using Eq. (1) is presented in Fig. 7.

V. SIGNAL MASS RESOLUTION

We calculate the mass of the $B_s^0\pi^{\pm}$ system using the quantity,

$$m(B_s^0\pi^{\pm}) = m(\mu^{\pm}D_s^{\mp}\pi^{\pm}) - m(\mu^{\pm}D_s^{\mp}) + m(B_s^0). \quad (4)$$

Before carrying out the search for the $X^{\pm}(5568)$ in the semileptonic channel we ensure that it is an unbiased and precise estimator of the mass of the $B_s^0\pi^{\pm}$ system. This is studied by simulating the two body decay $X(5568)^{\pm} \to B_s^0\pi^{\pm}$ where $B_s^0 \to \mu^{\pm}D_s^{\mp}X$, starting with a range of input masses $\tilde{m}(B_s^0\pi^{\pm})$. Following the decay chain $B_s^0 \to \mu^{\pm}D_s^{\mp}X$ and forming the invariant masses $m(\mu^{\pm}D_s^{\mp}\pi^{\pm})$ and $m(\mu^{\pm}D_s^{\mp})$ are found. Then $m(B_s^0\pi^{\pm})$ is calculated and compared to the input mass $\tilde{m}(B_s^0\pi^{\pm})$.

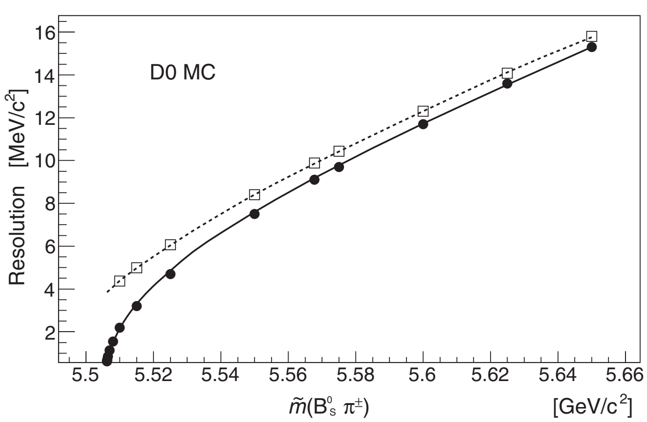


FIG. 8. The resolution for a zero width resonance as a function of $\tilde{m}(B_s^0\pi^{\pm})$. The solid circles and the solid line show the effect of the missing neutrino, and the open squares and dashed line show the convolution of the resolution due to the missing neutrino convolved with the 3.85 MeV$/c^2$ detector mass resolution.

To evaluate how well the mass approximation works to compensate for the missing neutrino, we model the decay with a toy MC that simulates the virtual W in $B_s^0 \to D_s^{\mp} + W^*$ with an isotropic distribution of μ and ν in the W boson rest frame. The resulting resolution of a zero width resonance due to the presence of the neutrino is modeled by a Gaussian. The width varies according to $\tilde{m}(B_s^0\pi^{\pm})$ as illustrated by the solid line in Fig. 8.

The mass resolution for the D0 detector of a state decaying into five reconstructed charged particles with a similar kinematic range as in this study is measured using the MC simulation and is given by a Gaussian function of width 3.85 MeV$/c^2$. The $m(B_s^0\pi^{\pm})$ resolution function is obtained by convoluting the Gaussian tracking resolution and the smearing resolution resulting from the missing neutrino. The resulting combined resolution, the dashed line in Fig. 8, can be approximated by

$$\sigma_{\rm SL} = [3.85 + 60.93(m_0^{0.85})]\ \mathrm{MeV}/c^2, \quad (5)$$

where m_0 has the same definition as in Eq. (1). These studies show that the difference between $m(B_s^0\pi^{\pm})$ and $\tilde{m}(B_s^0\pi^{\pm})$ is less than 1 MeV$/c^2$ in the search region. This is confirmed with the signal MC sample.

VI. SIGNAL FIT FUNCTION

The $X^{\pm}(5568)$ resonance is modeled by a relativistic Breit-Wigner function convolved with a Gaussian detector resolution function given in Eq. (5), $F_{\rm sig}(m, m_X, \Gamma_X)$,

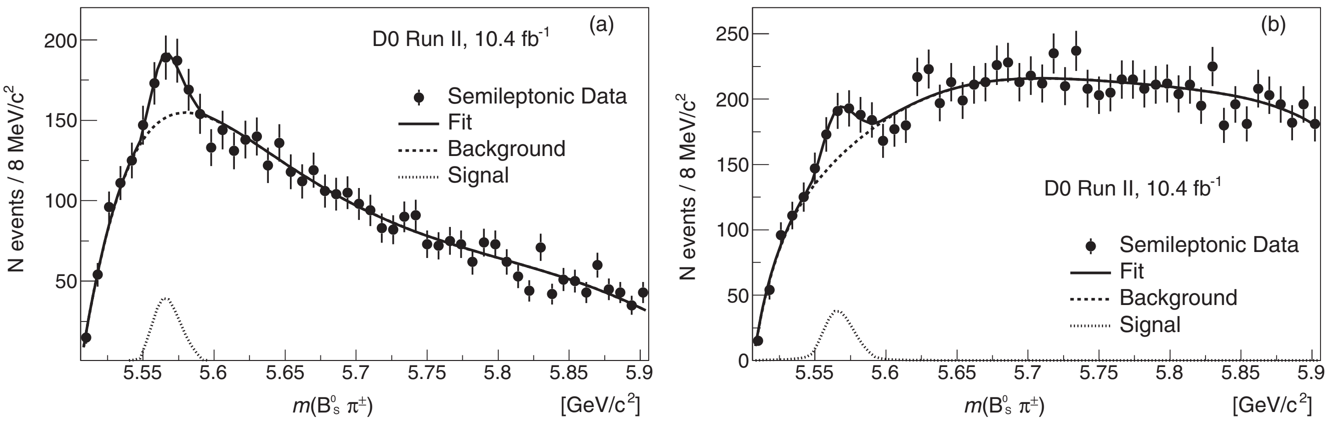


FIG. 9. The $m(B_s^0\pi^\pm)$ distribution (a) with and (b) without the cone cut. The fitting function is superimposed (see text for details).

where m_X and Γ_X are the mass and the width of the resonance.

The fit function has the form

$$F = f_{\rm sig}F_{\rm sig}(m, m_X, \Gamma_X) + f_{\rm bgr}F_{\rm bgr}(m), \tag{6}$$

where $f_{\rm sig}$ and $f_{\rm bgr}$ are normalization factors. The shape parameters in the background term $F_{\rm bgr}$ are fixed to the values obtained from fitting the MC background distribution (see Fig. 7).

We use the Breit-Wigner parametrization appropriate for an S-wave two-body decay near threshold,

$$\mathrm{BW}(m) \propto \frac{m_X^2\Gamma(m)}{(m_X^2 - m^2)^2 + m_X^2\Gamma^2(m)}. \tag{7}$$

The mass-dependent width $\Gamma(m) = \Gamma_X \cdot (q_1/q_0)$, where q_1 and q_0 are the magnitudes of momenta of the B_s^0 meson in the rest frame of the $B_s^0\pi^\pm$ system at the invariant mass equal to m and m_X, respectively.

VII. $X^\pm(5568)$ SEMILEPTONIC FIT RESULTS

In the fit to the semileptonic data with the cone cut shown in Fig. 9(a), the normalization parameters $f_{\rm sig}$ and $f_{\rm bgr}$ and the Breit-Wigner parameters m_X and Γ_X are allowed to vary. The fit yields the mass and width of $m_X = 5566.4^{+3.4}_{-2.8}$ MeV$/c^2$, $\Gamma_X = 2.0^{+9.5}_{-2.0}$ MeV$/c^2$, the number of signal events, $N = 121^{+51}_{-34}$, and a $\chi^2 = 34.9$ for 46 degrees of freedom. The local statistical significance of the signal is defined as $\sqrt{-2\ln(\mathcal{L}_0/\mathcal{L}_{\rm max})}$, where $\mathcal{L}_{\rm max}$ and $\mathcal{L}_0$ are likelihood values at the best-fit signal yield and the signal yield fixed to zero obtained from a binned maximum-likelihood fit. The p-value of the background only fit is 2.1×10^{-5}, and the local statistical significance is 4.3σ.

In the fit to the semileptonic data without the cone cut shown in Fig. 9(b), the mass and width of $m_X = 5566.7^{+3.6}_{-3.4}$ MeV$/c^2$, $\Gamma_X = 6.0^{+9.5}_{-6.0}$ MeV$/c^2$, the number of signal events, $N = 139^{+51}_{-63}$, and a $\chi^2 = 30.4$ for 46 degrees of freedom. The p-value of the background only fit is 7.7×10^{-6}, and the local statistical significance is 4.5σ. The fit results, both for the cone cut and no cone cut cases, are given in Table II and for various background parametrizations in Table III. The $X^\pm(5568)$ parameters for the cone cut and no cone cut cases are consistent.

A. Systematic uncertainties

Systematic uncertainties (Table IV) are obtained for the measured values of the mass, width and event yield of the $X^\pm(5568)$ signal. The dominant uncertainty is due to

TABLE II. Results for the fit to the semileptonic data sets (see Fig. 9).

	Cone cut	No cone cut
Fitted mass, MeV$/c^2$	$5566.4^{+3.4}_{-2.8}(\mathrm{stat})^{+1.5}_{-0.6}(\mathrm{syst})$	$5566.7^{+3.6}_{-3.4}(\mathrm{stat})^{+1.0}_{-1.0}(\mathrm{syst})$
Fitted width, MeV$/c^2$	$2.0^{+9.5}_{-2.0}(\mathrm{stat})^{+2.8}_{-2.0}(\mathrm{syst})$	$6.0^{+9.5}_{-6.0}(\mathrm{stat})^{+1.9}_{-4.6}(\mathrm{syst})$
Fitted number of signal events	$121^{+51}_{-34}(\mathrm{stat})^{+9}_{-28}(\mathrm{syst})$	$139^{+51}_{-63}(\mathrm{stat})^{+11}_{-32}(\mathrm{syst})$
χ^2/ndf	$34.9/(50-4)$	$30.4/(50-4)$
p-value	2.1×10^{-5}	7.7×10^{-6}
Local significance	4.3σ	4.5σ
Significance including systematic uncertainties	3.2σ	3.4σ

TABLE III. Semileptonic data fits for the different background parametrizations.

	Equation (1)	Equation (2)	Equation (3)	Smoothed MC simulation
Cone cut				
Fitted mass, MeV/c^2	$5566.4^{+3.4}_{-2.8}$	$5566.1^{+3.7}_{-3.2}$	···	$5567.1^{+4.4}_{-3.3}$
Fitted width, MeV/c^2	$2.0^{+9.5}_{-2.0}$	$1.0^{+12.8}_{-1.0}$	···	$1.2^{+12.9}_{-1.2}$
Fitted number of signal events	121^{+51}_{-34}	98^{+52}_{-29}	···	95^{+51}_{-30}
χ^2/ndf	$34.9/(50-4)$	$43.2/(50-4)$	···	$50.5/(50-4)$
Local significance	4.3σ	3.6σ	···	3.5σ
No cone cut				
Fitted mass, MeV/c^2	$5566.7^{+3.6}_{-3.4}$	$5566.2^{+4.2}_{-4.1}$	$5566.0^{+3.6}_{-3.4}$	$5566.1^{+4.5}_{-4.5}$
Fitted width, MeV/c^2	$6.0^{+9.5}_{-6.0}$	$6.0^{+12.0}_{-6.0}$	$6.5^{+8.9}_{-6.5}$	10^{+13}_{-10}
Fitted number of signal events	139^{+51}_{-63}	116^{+52}_{-48}	146^{+51}_{-54}	130^{+56}_{-48}
χ^2/ndf	$30.4/(50-4)$	$50.3/(50-4)$	$43.8/(50-4)$	$44.8/(50-4)$
Local significance	4.5σ	3.7σ	4.7σ	3.8σ

(i) the description of the background shape. We evaluate this systematic uncertainty by using the alternative parametrizations of the background, Eqs. (2) and (3) and the smoothed MC histogram and finding the maximal deviations from the nominal fit.

The effect of (ii) the MC weighting is estimated by creating 1000 background samples where the weights have been randomly varied based on the uncertainties in the weighting procedure.

Other sources of systematic uncertainty are evaluated by (iii) varying the energy scale in the MC sample relative to the data by ± 1 MeV/c^2, (iv) varying the mass resolution of the $X^{\pm}(5568)$ either by ± 1 MeV/c^2 around the mean value, or by using a constant resolution of 11.1 MeV/c^2 obtained from the MC simulation of the $X^{\pm}(5568)$ signal, (v) using a P-wave relativistic Breit-Wigner function, and (vi) estimating the shift of the fitted mass peak due to the missing neutrino.

Systematic uncertainties are summarized in Table IV. The uncertainties are added in quadrature separately for positive and negative values to obtain the total systematic uncertainties for each measured parameter. The results including systematic uncertainties are given in Table II.

B. Significance

Since we are seeking to confirm the result presented in Ref. [15] we do not apply a correction for a look elsewhere effect. The systematic uncertainties are treated as nuisance parameters to construct a prior predictive model [27,32] of our test statistic. When the systematic uncertainties are included, the significance of the observed semileptonic signal with the cone cut is 3.2σ (p-value $= 1.4 \times 10^{-3}$). The significance of the semileptonic signal without the cone is 3.4σ (p-value $= 6.4 \times 10^{-4}$).

TABLE IV. Systematic uncertainties for the $X^{\pm}(5568)$ state mass, width and the event yield obtained from the semileptonic data.

Source	Mass, MeV/c^2	Width, MeV/c^2	Event yield, events
Cone cut			
(i) Background shape description	+0.7; −0.3	+0.0; −1.0	+0.0; −26.6
(ii) Background reweighting	+0.1; −0.1	+0.4; −0.4	+3.9; −4.2
(iii) B_s^0 mass scale, MC simulation and data	+0.1; −0.3	+0.8; −1.0	+5.1; −7.8
(iv) Detector resolution	+0.9; −0.0	+2.7; −1.0	+6.5; −0.0
(v) P-wave Breit-Wigner	+0.0; −0.4	+0.0; −1.0	+0.0; −3.7
(vi) Missing neutrino effect	+1.0; −0.0	···	···
Total	+1.5; −0.6	+2.8; −2.0	+9.1; −28.3
No cone cut			
(i) Background shape description	+0.0; −0.7	+0.7; −2.5	+4.8; −28.0
(ii) Background reweighting	+0.1; −0.1	+0.7; −0.7	+5.0; −5.0
(iii) B_s^0 mass scale, MC simulation and data	+0.3; −0.5	+1.0; −1.4	+7.5; −9.6
(iv) Detector resolution	+0.0; −0.5	+1.3; −2.6	+3.7; −6.4
(v) P-wave Breit-Wigner	+0.0; −0.2	+0.0; −2.4	+0.0; −7.0
(vi) Missing neutrino effect	+1.0; −0.0	···	···
Total	+1.0; −1.0	+1.9; −4.6	+10.9; −31.5

TABLE V. Mean values and uncertainties for fitted number of events, mass and width from Gaussian fits to corresponding distributions from 10,000 trials with the cone cut. Also given is the expected statistical uncertainties on the fitted number of events, $\Delta(N_{\rm fit}(\rm sl))$, and the expected uncertainties on the measurement of the width, $\Delta(\Gamma_X)$ MeV/c^2. A range of signals with 75, 100, 125, 150, 175 and 200 signal events, mass $m_x = 5568.3$ MeV/c^2 and width $\Gamma_X = 21.9$ MeV/c^2 have been simulated. Background parametrization Eq. (1) is used.

$N_{\rm in}(\rm sl)$	$N_{\rm fit}(\rm sl)$	$\Delta(N_{\rm fit}(\rm sl))$	m_X MeV/c^2	Γ_X MeV/c^2	$\Delta(\Gamma_X)$ MeV/c^2
75	80.4 ± 0.9	61	5577.9 ± 0.24	13.1	15.3
100	108.5 ± 0.7	58	5572.9 ± 0.17	15.8	15.6
125	133.3 ± 0.6	59	5570.4 ± 0.12	17.7	15.3
150	156.7 ± 0.6	58	5569.3 ± 0.08	19.3	14.6
175	181.0 ± 0.6	59	5568.9 ± 0.07	20.2	13.8
200	204.2 ± 0.6	61	5568.7 ± 0.05	20.8	12.9

C. Closure tests

We have tested the accuracy of the fitting procedure using toy MC event samples constructed with input mass and width of 5568.3 and 21.9 MeV/c^2, respectively, with the number of input signal events varied in steps of 25 between 75 and 350. At each number of input signal events, 10,000 pseudoexperiments were generated. The signals are modeled with a relativistic Breit-Wigner function convolved with a Gaussian function representing the appropriate detector resolution. The background distribution is based on Eq. (1). For each trial the fitting procedure is performed to obtain the mass and width and the number of semileptonic signal events. The results of each set of trials is fitted with a Gaussian to determine the mean and the uncertainty in the number of signal events, the mass and the width (see Table V). The number of fitted signal events vs the number of injected signal events for the semileptonic samples are plotted in Fig. 10.

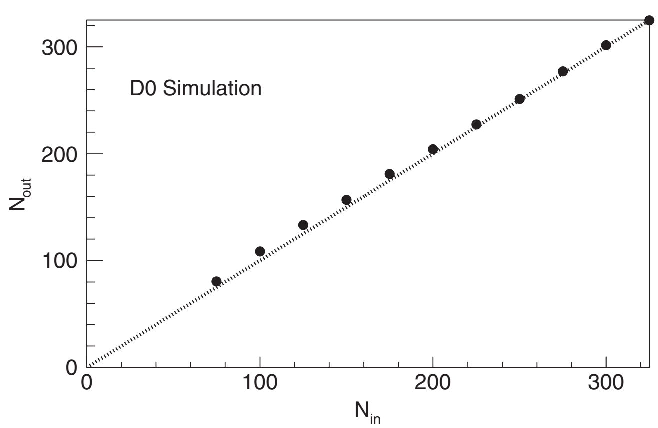


FIG. 10. Results of the toy MC tests of the fitting procedure (black circles) used in the analysis of the semileptonic data with the cone cut. The number of fitted signal events are plotted vs fitted number of injected signal events. The dotted line shows $N_{\rm in} = N_{\rm out}$.

For the ensembles with a number of input events similar to that observed in data, there is a slight overestimate of the yield and fitted mass, and the width is underestimated. This width reduction is in agreement with the results of the fits to data (Sec. VII) and indicate that the semileptonic data are not sensitive to the width. These effects are accounted for in the calculation of the significance.

D. Comparison with hadronic channel

The measured values of the mass, width, the number of signal events, and significance of the signal for the semileptonic channel and the hadronic channel [15] are given in Table VI. The mass and width of the $X^\pm(5568)$ for the semileptonic and hadronic channels are consistent taking into account the uncertainties. The observed yields are consistent with coming from a common particle given the number of B_s^0 events in the sample and the B_s^0 branching ratios.

E. Cross-checks

As a cross-check the $B_s^0\pi^\pm$ mass-bin size is set to 5 MeV/c^2 and to 10 MeV/c^2 instead of 8 MeV/c^2, and the lower edge of the fitted mass range is shifted by 2, 3, 5, and 7 MeV/c^2. This leads to maximal variations in the mass of $^{+0.1}_{-0.6}$ MeV/c^2, in the width of $^{+1.7}_{-0.9}$ MeV/c^2 and in the number of signal events $^{+0}_{-9}$ which are small compared to the statistical and systematic uncertainties.

To test the stability of the results, alternative choices are made regarding the fit parameters. In the first, the background fit parameters are allowed to float. The resulting fit is consistent with the nominal fit and the p-value of the background-only fit is 1.7×10^{-4} corresponding to a statistical significance of 3.8σ (Table VII). The second cross-check fixes the mass and width of the $X^\pm(5568)$ to the values found in Ref. [15]. Again, the resulting fit is consistent with the nominal fit with an increase in the number of signal events due to the increased width of the peak. The p-value of the background-only fit is 1.1×10^{-4} corresponding to a statistical significance of

TABLE VI. Fit results obtained in the semileptonic channel and in the hadronic channel (Ref. [15]). In the hadronic channel with no cone cut the mass and width of the $X^{\pm}(5568)$ were set to the values found with the cone cut. LEE—look elsewhere effect.

	Semileptonic		Hadronic (from Ref. [15])	
	Cone cut	No cone cut	Cone cut	No cone cut
Fitted mass, MeV$/c^2$	$5566.4^{+3.4}_{-2.8}{}^{+1.5}_{-0.6}$	$5566.7^{+3.6}_{-3.4}{}^{+1.0}_{-1.0}$	$5567.8 \pm 2.9^{+0.9}_{-1.9}$	5567.8
Fitted width, MeV$/c^2$	$2.0^{+9.5}_{-2.0}{}^{+2.8}_{-2.0}$	$6.0^{+9.5}_{-6.0}{}^{+1.9}_{-4.6}$	$21.9 \pm 6.4^{+5.0}_{-2.5}$	21.9
Fitted number of signal events	$121^{+51}_{-34}{}^{+9}_{-28}$	$139^{+51}_{-63}{}^{+11}_{-32}$	$133 \pm 31 \pm 15$	106 ± 23(stat)
Local significance	4.3σ	4.5σ	6.6σ	4.8σ
Significance with systematics	3.2σ	3.4σ	5.6σ	...
Significance with LEE + systematics	...	...	5.1σ	3.9σ

TABLE VII. Fit results for the semileptonic channel using parametrization (1) with the nominal fit, with all parameters free and the mass and width fixed to those of the hadronic channel. Statistical uncertainties only.

	Nominal fit	All parameters free	Mass and width fixed to hadronic
Cone cut			
Fitted mass, MeV$/c^2$	$5566.4^{+3.4}_{-2.8}$	5567.2 ± 2.9	5567.8
Fitted width, MeV$/c^2$	$2.0^{+9.5}_{-2.0}$	8.3 ± 11.0	21.9
Fitted number of signal events	121^{+51}_{-34}	181 ± 88	164 ± 44
χ^2/ndf	$34.9/(50-4)$	$30.9/(50-10)$	$38.0/(50-2)$
Local significance	4.3σ	3.8σ	3.9σ
No cone cut			
Fitted mass, MeV$/c^2$	$5566.7^{+3.6}_{-3.4}$	5566.6 ± 3.5	5567.8
Fitted width, MeV$/c^2$	$6.0^{+9.5}_{-6.0}$	8.4 ± 14.5	21.9
Fitted number of signal events	139^{+51}_{-63}	144 ± 101	168 ± 42
χ^2/ndf	$30.4/(50-4)$	$27.4/(50-10)$	$32.8/(50-2)$
Local significance	4.5σ	4.4σ	4.2σ

3.9σ (Table VII). These cross-checks are also repeated without the cone cut (Table VII).

VIII. PRODUCTION RATIO OF $X^{\pm}(5568)$ TO B_s^0

To calculate the production ratio of the $X^{\pm}(5568)$ to B_s^0, the number of the B_s^0-mesons needs to be estimated. The fitting of the $K^+K^-\pi^{\mp}$ mass distribution is described in Sec. IV. The number of $D_s^{\mp}$ mesons extracted from the fit and adjusted for the mass range $1.91 < m(K^+K^-\pi^{\mp}) < 2.03$ MeV$/c^2$ is $N(D_s^{\mp}) = 6648 \pm 127$ (see Fig. 2). The number of $\mu^{\pm}D_s^{\mp}$ events in the signal sample that are the result of a random combination of a promptly produced $D_s^{\mp}$ and a muon in the event is estimated using events where the muon and the $D_s^{\mp}$-meson have the same sign. The same sign data sample is analyzed using the same model as the opposite sign data with the means and widths of the Gaussians fixed to the values obtained from the opposite sign data. The number of events in the same-sign sample is $N(D_s^{\pm}) = 426 \pm 61$. The mass distributions of the $K^+K^-\pi^{\mp}$ for opposite and same-sign data are shown in Fig. 2.

The number of B_s^0-meson decays in the semileptonic data is estimated by subtracting the contribution of the promptly produced $\mu^{\pm}D_s^{\mp}$ events from the overall $\mu^{\pm}D_s^{\mp}$ sample. A study of the MC background simulations shows that the purity of the resulting sample is $99.5^{+0.5}_{-1.0}\%$. We find 6222 ± 141 B_s^0 events.

Combining these results and using the efficiency for the charged pion in the X(5568) decay (Sec. IV), we obtain a production ratio for the semileptonic data of

$$\rho = \frac{N_{\mathrm{sl}}(X^{\pm}(5568) \to B_s^0(\mathrm{sl})\pi^{\pm})}{N_{B_s^0}(\mathrm{sl})} = [7.3^{+2.8}_{-2.4}(\mathrm{stat})^{+0.6}_{-1.7}(\mathrm{syst})]\%, \quad (8)$$

for our fiducial selection [which includes the requirements $p_T(\mu^{\pm}D_s^{\mp}) > 10$ GeV$/c^2$ and 4.5 GeV$/c^2 < m(\mu^{\pm}D_s^{\mp}) < m(B_s^0)$], where $N_{\mathrm{sl}}(X^{\pm}(5568) \to B_s^0(\mathrm{sl})\pi^{\pm})$ is the number of $X^{\pm}(5568)$ decays to $B_s^0\pi^{\pm}$ and $N_{B_s^0}(\mathrm{sl})$ is the inclusive number of B_s^0 decays, both for semileptonic decays of the B_s^0. This result is similar to the ratio measured in

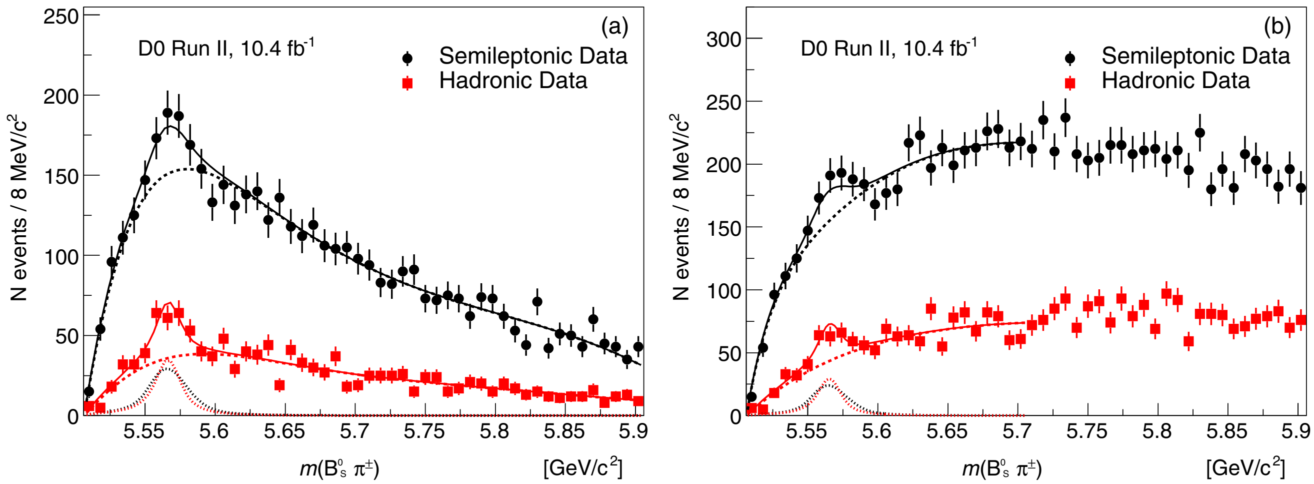


FIG. 11. The $m(B_s^0\pi^\pm)$ distribution for the hadronic (red squares) and semileptonic (black circles) data with the combined fitting function superimposed (a) with and (b) without the cone cut. (see text for details, the resulting fit parameters are given in Table VIII). The background parametrization function is taken from Eq. (1).

the hadronic channel $[8.6 \pm 1.9(\mathrm{stat}) \pm 1.4(\mathrm{syst})]\%$ for $p_T(J/\psi\phi\pi^\pm) > 10\ \mathrm{GeV}/c^2$ [15].

IX. COMBINED SIGNAL EXTRACTION

We now proceed to fit the hadronic and semileptonic data sets simultaneously. The hadronic data set is the same as used in Ref. [15] except that the data are fitted in the mass range $5.506 < m(B_s^0\pi^\pm) < 5.906\ \mathrm{GeV}/c^2$ instead of $5.500 < m(B_s^0\pi^\pm) < 5.900\ \mathrm{GeV}/c^2$. The data selection and background modeling for the hadronic data set are described in detail in Ref. [15].

The fit function has the form

$$F_{\mathrm{h}} = f_{\mathrm{h,sig}} F_{\mathrm{h,sig}}(m, m_X, \Gamma_X) + f_{\mathrm{h,bgr}} F_{\mathrm{h,bgr}}(m), \tag{9}$$

$$F_{\mathrm{sl}} = f_{\mathrm{sl,sig}} F_{\mathrm{sl,sig}}(m, m_X, \Gamma_X) + f_{\mathrm{sl,bgr}} F_{\mathrm{sl,bgr}}(m), \tag{10}$$

where $f_{\mathrm{h(sl),sig}}$ and $f_{\mathrm{h(sl),bgr}}$ are normalization factors. The shape parameters in the background terms $F_{\mathrm{h(sl),bgr}}$ are fixed to the values obtained from fitting the respective background models for the hadronic (h) and semileptonic (sl) samples to Eq. (1). The signal shape $F_{\mathrm{h(sl),sig}}$ is modeled by relativistic Breit-Wigner function convolved with a Gaussian detector resolution function that depends on the data sample. For the semileptonic sample the detector resolution is given by Eq. (5), and for the hadronic channel it is $3.85\ \mathrm{MeV}/c^2$. For the data without the cone cut the combined data are fitted in the range $5.506 < m(B_s^0\pi^\pm) < 5.706\ \mathrm{GeV}/c^2$ as the hadronic background is not well modeled for $m(B_s^0\pi^\pm) > 5.706\ \mathrm{GeV}/c^2$ [15]. The same Breit-Wigner parameters m_X and Γ_X are used for the hadronic and semileptonic samples. In the fits shown in Fig. 11, the normalization parameters $f_{\mathrm{h(sl),sig}}$ and $f_{\mathrm{h(sl),bgr}}$ and the Breit-Wigner parameters m_X and Γ_X are allowed to vary. Since the fraction of B_s^0 events produced by the decay of the $X^\pm(5568)$ should be essentially the same in the hadronic and semileptonic channels the $X^\pm(5568)$ event yields (N_{h} and N_{sl}) are constrained using the parameter

$$A_{\mathrm{sl,h}} = \frac{N_{\mathrm{sl}} - N_{\mathrm{h}}}{N_{\mathrm{sl}} + N_{\mathrm{h}}}, \tag{11}$$

which is required to be consistent with the B_s^0-meson production rate in the hadronic and semileptonic channels

$$A_{\mathrm{sl,h}}(B_s^0) = \frac{N_{B_s^0}(\mathrm{sl}) - N_{B_s^0}(\mathrm{h})}{N_{B_s^0}(\mathrm{sl}) + N_{B_s^0}(\mathrm{h})} = 0.054 \pm 0.020, \tag{12}$$

where $N_{B_s^0}(\mathrm{sl}) = 6222 \pm 144$, $N_{B_s^0}(\mathrm{h}) = 5582 \pm 100$ are the number of semileptonic and hadronic B_s^0 decays in the sample. A likelihood penalty of $0.5[(A_{\mathrm{sl,h}} - A_{\mathrm{sl,h}}(B_s^0))/\Delta A_{\mathrm{sl,h}}(B_s^0)]^2$ is applied where $\Delta A_{\mathrm{sl,h}}(B_s^0) = 0.020$ is the uncertainty. This uncertainty includes the statistical uncertainty in the number of B_s^0 events and the uncertainties in the relative reconstruction efficiencies and acceptances between the hadronic and semileptonic data. A ratio has been chosen for the constraint as it is well behaved if either of the event yields (N_{h} and N_{sl}) approaches zero.

The fit results are summarized in Table VIII, and the correlations between the fit parameters are given in Table IX. The correlation of nearly one between $N_X(\mathrm{sl})$ and $N_X(\mathrm{had})$ is a result of the constraint on the event yields [Eq. (11)]. The local statistical significance of the signal is defined as $\sqrt{-2\ln(\mathcal{L}_0/\mathcal{L}_{\max})}$, where $\mathcal{L}_{\max}$ and $\mathcal{L}_0$ are likelihood values at the best-fit signal yield and the signal yield fixed to zero obtained from a binned maximum-likelihood fit. For the cone cut the p-value of the fit to the data with the cone cut is 2.2×10^{-14} and the local statistical

TABLE VIII. Results for the combined fit to the hadronic and semileptonic data sets (see Fig. 11).

	Cone cut	No cone cut
Fitted mass, MeV$/c^2$	$5566.9^{+3.2}_{-3.1}(\text{stat})^{+0.6}_{-1.2}(\text{syst})$	$5565.8^{+4.2}_{-4.0}(\text{stat})^{+1.3}_{-2.0}(\text{syst})$
Fitted width, MeV$/c^2$	$18.6^{+7.9}_{-6.1}(\text{stat})^{+3.5}_{-3.8}(\text{syst})$	$16.3^{+9.8}_{-7.6}(\text{stat})^{+4.2}_{-6.5}(\text{syst})$
Fitted number of hadronic signal events	$131^{+37}_{-33}(\text{stat})^{+15}_{-14}(\text{syst})$	$99^{+40}_{-34}(\text{stat})^{+18}_{-33}(\text{syst})$
Fitted number of semileptonic signal events	$147^{+42}_{-37}(\text{stat})^{+17}_{-16}(\text{syst})$	$111.7^{+46}_{-39}(\text{stat})^{+20}_{-38}(\text{syst})$
χ^2/ndf	$94.7/(100-6)$	$54.2/(50-6)$
p-value	2.2×10^{-14}	1.9×10^{-8}
Local significance	7.6σ	5.6σ
Significance with LEE	6.9σ	5.0σ
Significance with LEE + systematics	6.7σ	4.7σ

significance is 7.6σ. The p-value without the cone cut is 8.2×10^{-9}, and the local statistical significance is 5.8σ.

A. Systematic uncertainties

The systematic uncertainties of the combined fit are given in Table X. The uncertainty on (i) the background shape descriptions is evaluated by using the alternative parametrizations of the background, Eqs. (2) and (3) and the smoothed MC histogram independently for the semileptonic and the hadronic channels (16 different fits) and finding the maximal deviations from the nominal fit.

The effect of (ii) the MC weighting for the semileptonic background is estimated by creating 1000 background samples where the weights have been randomly varied based on the uncertainties in the weighting procedure and measuring the standard deviation and bias of the measured values.

The (iii) MC component of the background for the hadronic sample is made up of a mixture of two independent MC samples with different production properties (see Ref. [15]), and the systematic uncertainties due to this are found by varying the composition of this mixture and measuring the standard deviation and bias of the measured values. The (iv) size of the hadronic sidebands is varied using the maximal deviations from the nominal fit to estimate the systematic uncertainty.

The systematic uncertainty due to the (v) fraction of MC and SS data in the semileptonic sample, (vi) the MC and sideband data in the case of the hadronic, is varied independently between the two samples measuring the standard deviation and bias of the measured values. Since the background model for the semileptonic sample without the cone cut only uses the MC background simulation this uncertainty (v) does not apply.

All of the uncertainties due to the modeling of the background are assumed to be independent for the hadronic and semileptonic data samples.

The remaining uncertainties are measured by finding the maximal deviations from the nominal fit for (vii) varying the energy scale in the semileptonic and hadronic MC data samples by ±1 MeV$/c^2$ in both samples simultaneously; (viii) varying the nominal mass resolution of 3.85 MeV$/c^2$ for the D0 detector by ±1 MeV$/c^2$ and $+2$ MeV$/c^2$ in both the hadronic and semileptonic data samples simultaneously; (ix) varying the resolution of the $X^{\pm}(5568)$ peak in the semileptonic channel either by ±1 MeV$/c^2$ around the mean value given by Eq. (5) or by using a constant resolution of 11.1 MeV$/c^2$ for the semileptonic data while the mass resolution in the hadronic channel remains at 3.85 MeV$/c^2$; (x) using a P-wave relativistic Breit-Wigner function for both data sets; (xi) setting the shift of the fitted mass peak in the semileptonic data with respect to the hadronic data due to the missing neutrino to ±1 MeV$/c^2$; and (xii) varying the constraint on the relative number of signal events in hadronic and semileptonic channels [Eq. (11)] between 0.034 and 0.074. The correlation of each of the sources of systematic uncertainty between the hadronic and semileptonic data sets is indicated in Table X. The uncertainties are added in quadrature separately for positive and negative values to obtain the total systematic

TABLE IX. Correlations between the parameters of the combined fit to the hadronic and semileptonic data sets (see Fig. 11). The yield in the semileptonic channel is $N_X(\text{sl})$, the hadronic channel $N_X(\text{h})$, while the fraction of background events is $f_{\text{sl,bgr}}$ and $f_{\text{h,bgr}}$, respectively.

	Mass	Width	$N_X(\text{sl})$	$N_X(\text{h})$	$f_{\text{sl,bgr}}$	$f_{\text{h,bgr}}$
Cone cut						
Mass	1	0.22	0.37	0.37	−0.06	−0.11
Width	0.22	1	0.58	0.59	−0.16	−0.29
$N_X(\text{sl})$	0.37	0.58	1	0.98	−0.31	−0.44
$N_X(\text{h})$	0.37	0.59	0.98	1	−0.30	−0.45
$f_{\text{sl,bgr}}$	−0.06	−0.16	−0.31	−0.30	1	0.14
$f_{\text{h,bgr}}$	−0.11	−0.29	−0.44	−0.45	0.14	1
No cone cut						
Mass	1	0.38	0.49	0.49	−0.11	−0.17
Width	0.38	1	0.64	0.64	−0.18	−0.31
$N_X(\text{sl})$	0.49	0.64	1	0.99	−0.33	−0.45
$N_X(\text{h})$	0.49	0.64	0.99	1	−0.33	−0.46
$f_{\text{sl,bgr}}$	−0.11	−0.18	−0.33	−0.33	1	0.15
$f_{\text{h,bgr}}$	−0.17	−0.31	−0.45	−0.46	0.15	1

TABLE X. Systematic uncertainties of the combined fit for the $X^{\pm}(5568)$ state mass, width and the event yields. Each uncertainty is either correlated or uncorrelated between the hadronic and semileptonic data sets.

				Event yields, events	
Source	Sample	Mass, MeV/c^2	Width, MeV/c^2	Hadronic	Semileptonic
Cone cut					
(i) Background shape description	Both	+0.3; −0.6	+1.9; −0.0	+0.0; −6.6	+0.0; −7.8
(ii) SL background reweighting	Semileptonic	+0.1; −0.2	+0.2; −0.2	+2.5; −3.3	+2.9; −3.9
(iii) Hadronic MC samples	Hadronic	+0.3; −0.2	+1.2; −0.4	+7.0; −2.5	+7.8; −2.8
(iv) Hadronic sidebands	Hadronic	+0.1; −0.1	+0.5; −1.3	+2.3; −9.3	+2.5; −10.2
(v) SL MC simulation/data ratio	Semileptonic	+0.0; −0.1	+0.1; −0.1	+1.0; −1.2	+1.1; −1.4
(vi) Hadronic MC simulation/data ratio	Hadronic	+0.0; −0.0	+0.2; −0.2	+1.0; −1.1	+1.1; −1.2
(vii) B_s^0 mass scale, MC simulation and data	Both	+0.2; −0.2	+0.8; −0.8	+3.7; −4.3	+4.1; −4.7
(viii) Detector resolution	Both	+0.1; −0.3	+1.3; −3.4	+1.4; −3.8	+1.6; −4.2
(ix) Missing neutrino effect	Semileptonic	+0.1; −0.1	+0.1; −0.0	+0.5; −0.1	+0.0; −0.4
(x) P-wave Breit-Wigner	Both	+0.0; −0.0	+2.1; −0.0	+11.7; −0.0	+13.0; −0.0
(xi) Mass offset	Both	+0.3; −0.3	+0.1; −0.0	+0.2; −0.4	+0.3; −0.4
(xii) Production fraction	Both	+0.0; −0.0	+0.1; −0.1	+1.4; −1.6	+4.2; −4.2
Total		+0.6; −1.2	+3.5; −3.8	+14.7; −13.6	+16.9; −15.8
No cone cut					
(i) Background shape description	Both	+1.1; −1.9	+1.4; −5.1	+7.6; −32.8	+8.4; −37.1
(ii) SL background reweighting	Semileptonic	+0.1; −0.0	+0.1; −0.3	+1.8; −1.1	+2.0; −1.4
(iii) Hadronic MC samples	Hadronic	+0.3; −0.0	+1.1; −0.0	+7.2; −0.0	+7.9; −0.0
(iv) Hadronic sidebands	Hadronic	+0.3; −0.1	+0.2; −0.6	+4.5; −3.7	+4.9; −4.2
(v) SL MC simulation/data ratio	Not applicable	…; …	…; …	…; …	…; …
(v) Hadronic MC simulation/data ratio	Hadronic	+0.1; −0.0	+0.5; −0.0	+7.4; −0.1	+8.1; −0.2
(vii) B_s^0 mass scale, MC simulation and data	Both	+0.1; −0.1	+0.9; −0.2	+5.1; −0.0	+5.6; −0.0
(viii) Detector resolution	Both	+0.1; −0.2	+1.6; −3.9	+1.5; −3.5	+1.6; −4.0
(ix) Missing neutrino effect	Semileptonic	+0.2; −0.1	+0.1; −0.1	+0.4; −0.0	+0.1; −0.3
(x) P-wave Breit-Wigner	Both	+0.0; −0.6	+3.3; −0.0	+10.7; −0.0	+11.8; −0.0
(xi) Mass offset	Both	+0.4; −0.4	+0.2; −0.2	+0.0; −0.0	+0.0; −0.1
(xii) Production fraction	Both	+0.0; −0.0	+0.1; −0.1	+0.8; −0.8	+3.5; −3.6
Total		+1.3; −2.0	+4.2; −6.5	+18.2; −33.2	+20.3; −37.8

uncertainties for each measured parameter. The results including systematic uncertainties are given in Table VIII.

B. Significance

The look-elsewhere effect (LEE) is determined using the approach proposed in Ref. [33]. We have generated 250,000 simulated background distributions with no signal, both with and without the cone cut. These distributions are fit using the same procedure as the data. The mass parameter of the relativistic Breit-Wigner is constrained to be between 5506 to 5675 MeV/c^2 (the sum of the mass of the B_d^0 and $K^{\pm}$) with a starting value of $m_X = 5600$ MeV/c^2. The width of the signal is allowed to vary between 0.1 and 60 MeV/c^2 with a starting value of $\Gamma_X = 21$ MeV/c^2. The maximum local statistical significance for each distribution is calculated. The resulting distribution of the local significance is fitted with the function

$$f_{\rm loc} = N_{\rm trials}[\chi^2(2) + P_1\chi^2(3)], \qquad (13)$$

where $N_{\rm trials}$ is the number of generated distributions, P_1 is a free parameter and $\chi^2(n)$ is the χ^2 cumulative distribution function for n degrees of freedom. We have used $n = 2$ and 3 as we are fitting two spectra simultaneously. The resulting function is integrated above the measured local significance to determine the global significance (Table VIII). The significance, not including the systematic uncertainty, of the observed signal accounting for the LEE and with the cone cut applied is 6.9σ (p-value $= 4.1 \times 10^{-12}$). The significance of the signal without the cone cut is 5.0σ (p-value $= 4.1 \times 10^{-7}$). The effect of choosing the function in Eq. (13) is studied by modifying it to $f_{\rm loc} = N_{\rm trials}[\chi^2(2) + P_1\chi^2(4)]$ and $f_{\rm loc} = N_{\rm trials}[\chi^2(2) + P_1\chi^2(3) + P_2\chi^2(4)]$ with no significant change to the significance being observed. The look-elsewhere effect on the signal significance is checked with a method described in Ref. [33] that relates the tail probability with the number of "upcrossing" at a small reference level. Five hundred simulated background spectra are generated. Each of these 500 distributions is fitted with the background plus signal function with different initial masses from 5506 to 5675 MeV/c^2 in 5 MeV/c^2 steps along with a background-only fit. The significance is plotted for each of the mass points and the number of upcrossings (each time the significance crosses a

TABLE XI. Mean values and uncertainties for fitted number of events, mass and width from Gaussian fits to corresponding distributions from 10,000 trials with the cone cut. Also given is the expected statistical uncertainties on the fitted number of events, $\Delta(N_{\rm fit})$, and the expected uncertainties on the measurement of the width, $\Delta(\Gamma_X)$ MeV/c^2. A range of signals with 75, 100, 125, 150, 175 and 200 signal events, mass $m_x = 5568.3$ MeV/c^2 and width $\Gamma_X = 21.9$ MeV/c^2 have been simulated. Background parametrization Eq. (1) is used.

Semileptonic channel			Hadronic channel			m_X	Γ_X	$\Delta(\Gamma_X)$
$N_{\rm in}$(sl)	$N_{\rm fit}$(sl)	$\Delta(N_{\rm fit}(\rm sl))$	$N_{\rm in}$(h)	$N_{\rm fit}$(h)	$\Delta(N_{\rm fit}(\rm h))$	MeV/c^2	MeV/c^2	MeV/c^2
75	73.8 ± 0.3	25.7	67.3	66.0 ± 0.2	23.0	5569.0 ± 0.076	19.3	10.9
100	99.1 ± 0.3	26.3	89.8	88.7 ± 0.2	23.6	5568.4 ± 0.042	20.8	9.2
125	124.9 ± 0.3	26.8	112.2	111.7 ± 0.2	24.0	5568.4 ± 0.032	21.5	7.8
150	149.6 ± 0.3	26.5	134.6	133.8 ± 0.2	23.6	5568.4 ± 0.027	21.9	6.8
175	175.9 ± 0.3	27.2	157.1	157.3 ± 0.2	24.3	5568.4 ± 0.023	22.3	6.0
200	200.8 ± 0.3	27.2	179.5	179.6 ± 0.2	24.2	5568.4 ± 0.021	22.4	5.4

small reference value) is measured. The mean number of upcrossings for a reference level of 0.5 is determined, and the global significance is calculated. The resulting significance is consistent with the method described above.

The systematic uncertainties are treated as nuisance parameters to construct a prior predictive model [27,32] of our test statistic. When the systematic uncertainties are included, the significance of the observed signal with the cone cut applied for the combined fit is reduced to 6.7σ (p-value = 1.5×10^{-11}), and the significance of the signal without the cone cut is 4.7σ (p-value $= 2.0 \times 10^{-6}$).

C. Closure tests

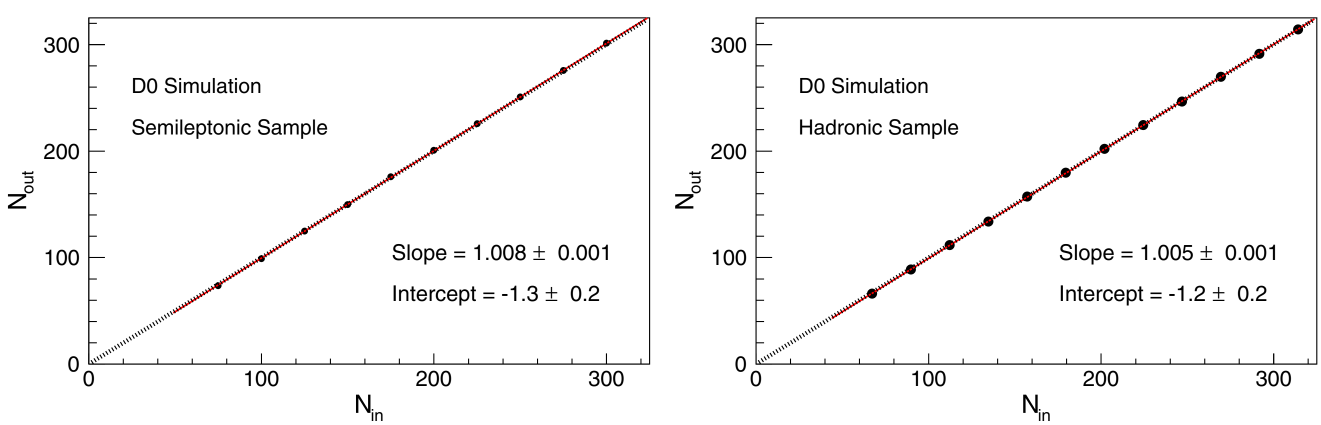
To test the sensitivity and accuracy of the fitting procedure for the combined signal extraction we repeat the closure tests carried out in Sec. VII C with the following modifications. The size of the associated hadronic signal is set using Eqs. (11) and (12). The appropriate detector resolution is used, Eq. (5) for the semileptonic sample and 3.85 MeV/c^2 for the hadronic sample. For each trial the fitting procedure is performed to obtain the mass and width and the number of semileptonic and hadronic signal events. The results of each set of trials is fitted with a Gaussian to determine the mean and the uncertainty in the number of signal events, the mass and the width (see Table XI). The number of fitted signal events vs the number of injected signal events for the semileptonic and hadronic samples is plotted in Fig. 12. These results show excellent agreement between the input and fit parameters.

D. Cross-checks

To test the stability of the results, alternative choices are made regarding the fit parameters (see Table XII).

When no constraint is placed on the ratio of the event yields in the hadronic and semileptonic channels, Eq. (11), the results are entirely consistent with the fit with the constraint.

We have also carried out a fit in which two of the systematic effects are treated as nuisance parameters in the fit. We allow a mass shift, Δm, between the hadronic and



(a) Semileptonic Sample (b) Hadronic Sample

FIG. 12. Results of the toy MC tests of the combined sample fitting procedure (black circles) used in the analysis with the cone cut. The number of fitted signal events are plotted vs fitted number of injected signal events for the (a) semileptonic and (b) hadronic samples. The dotted line shows $N_{\rm in} = N_{\rm out}$ and the red line shows the fit to a line.

TABLE XII. Various cross-checks for the combined fit of the hadronic and semileptonic data sets.

	Default fit	No production constraint	Nuisance parameter	Zero width
Cone cut				
Fitted mass, MeV/c^2	$5566.9^{+3.2}_{-3.1}$	$5566.8^{+3.2}_{-3.1}$	$5567.4^{+3.2}_{-3.4}$	$5569.9^{+1.3}_{-1.3}$
Fitted width, MeV/c^2	$18.6^{+7.9}_{-6.1}$	$18.3^{+8.0}_{-6.2}$	$21.7^{+7.3}_{-5.5}$	0
Fitted number of hadronic signal events	131^{+37}_{-33}	127^{+34}_{-29}	134^{+37}_{-33}	60^{+17}_{-16}
Fitted number of semileptonic signal events	147^{+42}_{-37}	159^{+66}_{-59}	151^{+41}_{-37}	68^{+19}_{-18}
χ^2/ndf	$94.7/(100-6)$	$94.5/(100-6)$	$94.8/(100-8)$	$115.4/(100-7)$
p-value	2.2×10^{-14}	2.0×10^{-14}	2.4×10^{-14}	8.5×10^{-10}
Local significance	7.6σ	7.7σ	7.6σ	6.1σ
No cone cut				
Fitted mass, MeV/c^2	$5565.8^{+4.2}_{-4.0}$	$5565.8^{+4.1}_{-3.9}$	$5566.3^{+4.4}_{-4.6}$	$5569.7^{+1.6}_{-1.9}$
Fitted width, MeV/c^2	$16.3^{+9.8}_{-7.6}$	$15.0^{+9.6}_{-7.8}$	$20.0^{+9.1}_{-9.4}$	0
Fitted number of hadronic signal events	99^{+40}_{-34}	84^{+43}_{-35}	103^{+40}_{-37}	48^{+17}_{-16}
Fitted number of semileptonic signal events	112^{+46}_{-39}	151^{+72}_{-61}	115^{+45}_{-42}	54^{+20}_{-18}
χ^2/ndf	$54.2/(50-6)$	$52.5/(50-6)$	$54.8/(50-8)$	$101.3/(50-7)$
p-value	1.9×10^{-8}	8.2×10^{-9}	2.7×10^{-8}	5.1×10^{-6}
Local significance	5.6σ	5.8σ	5.6σ	4.6σ

semileptonic data with a likelihood penalty of $0.5(\Delta m/1\ \mathrm{MeV}/c^2)^2$. We also allow the overall resolution of the semileptonic signal to vary by $\Delta\sigma_{\mathrm{SL}}$ with a likelihood penalty of $0.5(\Delta\sigma_{\mathrm{SL}}/1\ \mathrm{MeV}/c^2)^2$. The resultant fit produces a mass, width and event yields that are consistent with the default fit and shifts of $\Delta m = (0.0 \pm 1.4)\ \mathrm{MeV}/c^2$ and $\Delta\sigma_{\mathrm{SL}} = (-0.1 \pm 1.4)\ \mathrm{MeV}/c^2$.

The significance of a nonzero width is determined by fitting the data with the width set to zero and comparing it with the fit with no constraint on the width (Table XII). Using the data with the cone cut the p-value of the width being consistent with zero is 5.4×10^{-6}, and the statistical significance is 4.5σ. The significance without the cone cut is 3.3σ (p-value = 1.1×10^{-3}).

X. CONCLUSIONS

We have presented the results of a search for the $X^\pm(5568) \to B_s^0\pi^\pm$ with semileptonic decays of the B_s^0 meson. The $X^\pm(5568) \to B_s^0\pi^\pm$ state reported in the case that $B_s^0 \to J/\psi\phi$ [15] is confirmed for the case that $B_s^0 \to \mu^\mp D_s^\pm X$, $D_s^\pm \to \phi\pi^\pm$. The analyses of the hadronic and semileptonic data give similar measurements of the mass, width and production ratio of $X^\pm(5568)$ to a B_s^0 meson. The mass and width of this state are measured using a combined fit of both data sets with the cone cut, yielding $m = 5566.9^{+3.2}_{-3.1}(\mathrm{stat})^{+0.6}_{-1.2}(\mathrm{syst})\ \mathrm{MeV}/c^2$, $\Gamma = 18.6^{+7.9}_{-6.1}(\mathrm{stat})^{+3.5}_{-3.8}(\mathrm{syst})\ \mathrm{MeV}/c^2$. The p-value for the null signal hypothesis to represent the data is 1.5×10^{-11} (6.7σ).

ACKNOWLEDGMENTS

This document was prepared by the D0 Collaboration using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359. We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the Department of Energy and National Science Foundation (United States of America); Alternative Energies and Atomic Energy Commission and National Center for Scientific Research/National Institute of Nuclear and Particle Physics (France); Ministry of Education and Science of the Russian Federation, National Research Center "Kurchatov Institute" of the Russian Federation, and Russian Foundation for Basic Research (Russia); National Council for the Development of Science and Technology and Carlos Chagas Filho Foundation for the Support of Research in the State of Rio de Janeiro (Brazil); Department of Atomic Energy and Department of Science and Technology (India); Administrative Department of Science, Technology and Innovation (Colombia); National Council of Science and Technology (Mexico); National Research Foundation of Korea (Korea); Foundation for Fundamental Research on Matter (Netherlands); Science and Technology Facilities Council and The Royal Society (United Kingdom); Ministry of Education, Youth and Sports (Czech Republic); Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research) and Deutsche Forschungsgemeinschaft (German Research Foundation) (Germany); Science Foundation Ireland (Ireland); Swedish Research Council (Sweden); China Academy of Sciences and National Natural Science Foundation of China (China); and Ministry of Education and Science of Ukraine (Ukraine).

[1] M. Gell-Mann, A schematic model of baryons and mesons, Phys. Lett. **8**, 214 (1964).
[2] G. Zweig, An SU(3) model for strong interaction symmetry and its breaking. Version 1, Report No. CERN-TH-401, 1964, https://cds.cern.ch/record/352337.
[3] L. Maiani, F. Piccinini, A. D. Polosa, and V. Riquer, A New Look at Scalar Mesons, Phys. Rev. Lett. **93**, 212002 (2004).
[4] A. Esposito, A. Pilloni, and A. D. Polosa, Multiquark resonances, Phys. Rep. **668**, 1 (2017).
[5] H. Chen, W. Chen, X. Liu, and S. Zhu, The hidden-charm pentaquark and tetraquark states, Phys. Rep. **639**, 1 (2016).
[6] S. L. Olsen, T. Skwarnicki, and D. Zieminska, Nonstandard heavy mesons and baryons: Experimental evidence, Rev. Mod. Phys. **90**, 015003 (2018).
[7] S. K. Choi *et al.* (Belle Collaboration), Observation of a Narrow Charmonium - like State in Exclusive $B^{\pm} \to K^{\pm}\pi^{+}\pi^{-}J/\psi$ Decays, Phys. Rev. Lett. **91**, 262001 (2003).
[8] M. Aaboud *et al.* (ATLAS Collaboration), Measurements of $\psi(2S)$ and $X(3872) \to J/\psi\pi^{+}\pi^{-}$ production in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, J. High Energy Phys. 01 (2017) 117.
[9] B. Aubert *et al.* (*BABAR* Collaboration), Study of the $B \to J/\psi K^{-}\pi^{+}\pi^{-}$ decay and measurement of the $B \to X(3872)K^{-}$ branching fraction, Phys. Rev. D **71**, 071103 (2005).
[10] M. Ablikim *et al.* (BESIII Collaboration), Observation of $e^{+}e^{-} \to \gamma X(3872)$ at BES III, Phys. Rev. Lett. **112**, 092001 (2014).
[11] D. Acosta *et al.* (CDF Collaboration), Observation of the Narrow State $X(3872) \to J/\psi\pi^{+}\pi^{-}$ in $\bar{p}p$ Collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. Lett. **93**, 072001 (2004).
[12] S. Chatrchyan *et al.* (CMS Collaboration), Measurement of the X(3872) production cross section via decays to $J/\psi\pi\pi$ in pp collisions at $\sqrt{s} = 7$ TeV, J. High Energy Phys. 04 (2013) 154.
[13] V. M. Abazov *et al.* (D0 Collaboration), Observation and Properties of the $X(3872)$ Decaying to $J/\psi\pi^{+}\pi^{-}$ in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV, Phys. Rev. Lett. **93**, 162002 (2004).
[14] R. Aaij *et al.* (LHCb Collaboration), Determination of the X (3872) Meson Quantum Numbers, Phys. Rev. Lett. **110**, 222001 (2013).
[15] V. M. Abazov *et al.* (D0 Collaboration), Evidence for a $B_s^0\pi^{\pm}$ State, Phys. Rev. Lett. **117**, 022003 (2016) (and the Supplemental Material at https://journals.aps.org/prl/supplemental/10.1103/PhysRevLett.117.022003 for additional figures).
[16] For a description of the look-elsewhere effect see L. Lyons, Open statistical issues in particle physics, Ann. Appl. Stat. **2**, 887 (2008).
[17] $\eta = -\ln[\tan(\theta/2)]$ is the pseudorapidity, and θ is the polar angle between the track momentum and the proton beam direction. ϕ is the azimuthal angle of the track.
[18] R. Aaij *et al.* (LHCb Collaboration), Search for Structure in the $B_s^0\pi^{\pm}$ Invariant Mass Spectrum, Phys. Rev. Lett. **117**, 152003 (2016); Publisher's Note **118**, 109904(A) (2017).
[19] A. M. Sirunyan *et al.* (CMS Collaboration), Search for the X (5568) state decaying into $B_s^0\pi^{\pm}$ in proton-proton collisions at $\sqrt{s} = 8$ TeV, arXiv:1712.06144 [Phys. Rev. Lett. (to be published)].
[20] T. Aaltonen *et al.* (CDF Collaboration), A search for the exotic meson $X(5568)$ with the collider detector at Fermilab, arXiv:1712.09620 [Phys. Rev. Lett. (to be published)].
[21] V. M. Abazov *et al.* (D0 Collaboration), The upgraded D0 detector, Nucl. Instrum. Methods Phys. Res., Sect. A **565**, 463 (2006).
[22] R. Angstadt *et al.* (D0 Collaboration), The layer 0 inner silicon detector of the D0 experiment, Nucl. Instrum. Methods Phys. Res., Sect. A **622**, 298 (2010).
[23] V. M. Abazov *et al.*, The muon system of the run II D0 detector, Nucl. Instrum. Methods Phys. Res., Sect. A **552**, 372 (2005).
[24] V. M. Abazov *et al.* (D0 Collaboration), Measurement of the Semileptonic Charge Asymmetry Using $B_s^0 \to D_s\mu X$ Decays, Phys. Rev. Lett. **110**, 011801 (2013).
[25] J. Abdallah *et al.* (DELPHI Collaboration), b tagging in DELPHI at LEP, Eur. Phys. J. C **32**, 185 (2004).
[26] The three-dimensional impact parameter (IP) is defined as the distance of closest approach of the track to the $p\bar{p}$ collision point. The two-dimensional IP is the distance of closest approach projected onto the plane transverse to the $p\bar{p}$ beams.
[27] C. Patrignani *et al.* (Particle Data Group), Review of particle physics, Chin. Phys. C **40**, 100001 (2016).
[28] T. Sjostrand, S. Mrenna, and P. Z. Skands, PYTHIA 6.4 physics and manual, J. High Energy Phys. 05 (2006) 026.
[29] D. J. Lange, The EvtGen particle decay simulation package, Proceedings, 7th International Conference on B physics at hadron machines (BEAUTY 2000): Maagan, Israel, September 13–18, 2000, Nucl. Instrum. Methods Phys. Res., Sect. A **462**, 152 (2001) for details see https://evtgen.hepforge.org.
[30] J. Friedman, *1974 CERN School of Computing, Godoysund, Norway, 11–24 Aug 1974: Proceedings* (1974) where for n data points the smoothed i^{th} point $y_s(i)$ is given by $y_s(i) = 0.25y(i-1) + 0.5y(i) + 0.25y(i+1)$ for $i = 2, n-1$ and $y_s(1) = y(1)$ and $y_s(n) = y(n)$, DOI: 10.5170/CERN-1974-023.
[31] H. Albrecht *et al.* (ARGUS Collaboration), Search for hadronic $b \to u$ decays, Phys. Lett. B **241**, 278 (1990).
[32] C. Giunti, Treatment of the background error in the statistical analysis of Poisson processes, Phys. Rev. D **59**, 113009 (1999).
[33] E. Gross and O. Vitells, Trial factors or the look elsewhere effect in high energy physics, Eur. Phys. J. C **70**, 525 (2010).