Measurement of the CP-violating phase $\phi\_s^{J/\psi\phi}$ using the flavor-tagged decay $B\_s^0 \to J/\psi\phi$ in 8 fb^{-1} of $p\overline{p}$ collisions

V.M. Abazov, F. Badaud, U. Bassler, G. Bernardi, M. Besancon, D. Brown, J. Brown, B. Calpas, E. Chapon, S. Chevalier-Thery, et al.

FERMILAB-PUB-11-463-E

Measurement of the CP-violating phase $\phi_s^{J/\psi\phi}$ using the flavor-tagged decay $B_s^0 \to J/\psi\phi$ in 8 fb^{-1} of $p\overline{p}$ collisions

V.M. Abazov,[34] B. Abbott,[72] B.S. Acharya,[28] M. Adams,[48] T. Adams,[46] G.D. Alexeev,[34] G. Alkhazov,[38] A. Alton[a],[60] G. Alverson,[59] G.A. Alves,[2] M. Aoki,[47] M. Arov,[57] A. Askew,[46] B. Åsman,[40] S. Atkins,[57] O. Atramentov,[64] K. Augsten,[9] C. Avila,[7] J. BackusMayes,[79] F. Badaud,[12] L. Bagby,[47] B. Baldin,[47] D.V. Bandurin,[46] S. Banerjee,[28] E. Barberis,[59] P. Baringer,[55] J. Barreto,[3] J.F. Bartlett,[47] U. Bassler,[17] V. Bazterra,[48] A. Bean,[55] M. Begalli,[3] M. Begel,[70] C. Belanger-Champagne,[40] L. Bellantoni,[47] S.B. Beri,[26] G. Bernardi,[16] R. Bernhard,[21] I. Bertram,[41] M. Besançon,[17] R. Beuselinck,[42] V.A. Bezzubov,[37] P.C. Bhat,[47] V. Bhatnagar,[26] G. Blazey,[49] S. Blessing,[46] K. Bloom,[63] A. Boehnlein,[47] D. Boline,[69] E.E. Boos,[36] G. Borissov,[41] T. Bose,[58] A. Brandt,[75] O. Brandt,[22] R. Brock,[61] G. Brooijmans,[67] A. Bross,[47] D. Brown,[16] J. Brown,[16] X.B. Bu,[47] M. Buehler,[47] V. Buescher,[23] V. Bunichev,[36] S. Burdin[b],[41] T.H. Burnett,[79] C.P. Buszello,[40] B. Calpas,[14] E. Camacho-Pérez,[31] M.A. Carrasco-Lizarraga,[55] B.C.K. Casey,[47] H. Castilla-Valdez,[31] S. Chakrabarti,[69] D. Chakraborty,[49] K.M. Chan,[53] A. Chandra,[77] E. Chapon,[17] G. Chen,[55] S. Chevalier-Théry,[17] D.K. Cho,[74] S.W. Cho,[30] S. Choi,[30] B. Choudhary,[27] S. Cihangir,[47] D. Claes,[63] J. Clutter,[55] M. Cooke,[47] W.E. Cooper,[47] M. Corcoran,[77] F. Couderc,[17] M.-C. Cousinou,[14] A. Croc,[17] D. Cutts,[74] A. Das,[44] G. Davies,[42] K. De,[75] S.J. de Jong,[33] E. De La Cruz-Burelo,[31] F. Déliot,[17] M. Demarteau,[47] R. Demina,[68] D. Denisov,[47] S.P. Denisov,[37] S. Desai,[47] C. Deterre,[17] K. DeVaughan,[63] H.T. Diehl,[47] M. Diesburg,[47] P.F. Ding,[43] A. Dominguez,[63] T. Dorland,[79] A. Dubey,[27] L.V. Dudko,[36] D. Duggan,[64] A. Duperrin,[14] S. Dutt,[26] A. Dyshkant,[49] M. Eads,[63] D. Edmunds,[61] J. Ellison,[45] V.D. Elvira,[47] Y. Enari,[16] H. Evans,[51] A. Evdokimov,[70] V.N. Evdokimov,[37] G. Facini,[59] T. Ferbel,[68] F. Fiedler,[23] F. Filthaut,[33] W. Fisher,[61] H.E. Fisk,[47] M. Fortner,[49] H. Fox,[41] S. Fuess,[47] A. Garcia-Bellido,[68] G.A García-Guerra[c],[31] V. Gavrilov,[35] P. Gay,[12] W. Geng,[14,61] D. Gerbaudo,[65] C.E. Gerber,[48] Y. Gershtein,[64] G. Ginther,[47,68] G. Golovanov,[34] A. Goussiou,[79] P.D. Grannis,[69] S. Greder,[18] H. Greenlee,[47] Z.D. Greenwood,[57] E.M. Gregores,[4] G. Grenier,[19] Ph. Gris,[12] J.-F. Grivaz,[15] A. Grohsjean,[17] S. Grünendahl,[47] M.W. Grünewald,[29] T. Guillemin,[15] G. Gutierrez,[47] P. Gutierrez,[72] A. Haas[d],[67] S. Hagopian,[46] J. Haley,[59] L. Han,[6] K. Harder,[43] A. Harel,[68] J.M. Hauptman,[54] J. Hays,[42] T. Head,[43] T. Hebbeker,[20] D. Hedin,[49] H. Hegab,[73] A.P. Heinson,[45] U. Heintz,[74] C. Hensel,[22] I. Heredia-De La Cruz,[31] K. Herner,[60] G. Hesketh[e],[43] M.D. Hildreth,[53] R. Hirosky,[78] T. Hoang,[46] J.D. Hobbs,[69] B. Hoeneisen,[11] M. Hohlfeld,[23] Z. Hubacek,[9,17] N. Huske,[16] V. Hynek,[9] I. Iashvili,[66] Y. Ilchenko,[76] R. Illingworth,[47] A.S. Ito,[47] S. Jabeen,[74] M. Jaffré,[15] D. Jamin,[14] A. Jayasinghe,[72] R. Jesik,[42] K. Johns,[44] M. Johnson,[47] A. Jonckheere,[47] P. Jonsson,[42] J. Joshi,[26] A.W. Jung,[47] A. Juste,[39] K. Kaadze,[56] E. Kajfasz,[14] D. Karmanov,[36] P.A. Kasper,[47] I. Katsanos,[63] R. Kehoe,[76] S. Kermiche,[14] N. Khalatyan,[47] A. Khanov,[73] A. Kharchilava,[66] Y.N. Kharzheev,[34] J.M. Kohli,[26] A.V. Kozelov,[37] J. Kraus,[61] S. Kulikov,[37] A. Kumar,[66] A. Kupco,[10] T. Kurča,[19] V.A. Kuzmin,[36] J. Kvita,[8] S. Lammers,[51] G. Landsberg,[74] P. Lebrun,[19] H.S. Lee,[30] S.W. Lee,[54] W.M. Lee,[47] J. Lellouch,[16] L. Li,[45] Q.Z. Li,[47] S.M. Lietti,[5] J.K. Lim,[30] D. Lincoln,[47] J. Linnemann,[61] V.V. Lipaev,[37] R. Lipton,[47] Y. Liu,[6] A. Lobodenko,[38] M. Lokajicek,[10] R. Lopes de Sa,[69] H.J. Lubatti,[79] R. Luna-Garcia[f],[31] A.L. Lyon,[47] A.K.A. Maciel,[2] D. Mackin,[77] R. Madar,[17] R. Magaña-Villalba,[31] S. Malik,[63] V.L. Malyshev,[34] Y. Maravin,[56] J. Martínez-Ortega,[31] R. McCarthy,[69] C.L. McGivern,[55] M.M. Meijer,[33] A. Melnitchouk,[62] D. Menezes,[49] P.G. Mercadante,[4] M. Merkin,[36] A. Meyer,[20] J. Meyer,[22] F. Miconi,[18] N.K. Mondal,[28] G.S. Muanza,[14] M. Mulhearn,[78] E. Nagy,[14] M. Naimuddin,[27] M. Narain,[74] R. Nayyar,[27] H.A. Neal,[60] J.P. Negret,[7] P. Neustroev,[38] S.F. Novaes,[5] T. Nunnemann,[24] G. Obrant[‡],[38] J. Orduna,[77] N. Osman,[14] J. Osta,[53] G.J. Otero y Garzón,[1] M. Padilla,[45] A. Pal,[75] N. Parashar,[52] V. Parihar,[74] S.K. Park,[30] J. Parsons,[67] R. Partridge[d],[74] N. Parua,[51] A. Patwa,[70] B. Penning,[47] M. Perfilov,[36] K. Peters,[43] Y. Peters,[43] K. Petridis,[43] G. Petrillo,[68] P. Pétroff,[15] R. Piegaia,[1] M.-A. Pleier,[70] P.L.M. Podesta-Lerma[g],[31] V.M. Podstavkov,[47] P. Polozov,[35] A.V. Popov,[37] M. Prewitt,[77] D. Price,[51] N. Prokopenko,[37] S. Protopopescu,[70] J. Qian,[60] A. Quadt,[22] B. Quinn,[62] M.S. Rangel,[2] K. Ranjan,[27] P.N. Ratoff,[41] I. Razumov,[37] P. Renkel,[76] M. Rijssenbeek,[69] I. Ripp-Baudot,[18] F. Rizatdinova,[73] M. Rominsky,[47] A. Ross,[41] C. Royon,[17] P. Rubinov,[47] R. Ruchti,[53] G. Safronov,[35] G. Sajot,[13] P. Salcido,[49] A. Sánchez-Hernández,[31] M.P. Sanders,[24] B. Sanghi,[47] A.S. Santos,[5] G. Savage,[47] L. Sawyer,[57] T. Scanlon,[42] R.D. Schamberger,[69] Y. Scheglov,[38] H. Schellman,[50] T. Schliephake,[25] S. Schlobohm,[79] C. Schwanenberger,[43] R. Schwienhorst,[61] J. Sekaric,[55] H. Severini,[72] E. Shabalina,[22] V. Shary,[17] A.A. Shchukin,[37] R.K. Shivpuri,[27] V. Simak,[9] V. Sirotenko,[47] P. Skubic,[72] P. Slattery,[68] D. Smirnov,[53] K.J. Smith,[66] G.R. Snow,[63] J. Snow,[71] S. Snyder,[70] S. Söldner-Rembold,[43] L. Sonnenschein,[20] K. Soustruznik,[8] J. Stark,[13] V. Stolin,[35]

D.A. Stoyanova,[37] M. Strauss,[72] D. Strom,[48] L. Stutte,[47] L. Suter,[43] P. Svoisky,[72] M. Takahashi,[43] A. Tanasijczuk,[1] M. Titov,[17] V.V. Tokmenin,[34] Y.-T. Tsai,[68] K. Tschann-Grimm,[69] D. Tsybychev,[69] B. Tuchming,[17] C. Tully,[65] L. Uvarov,[38] S. Uvarov,[38] S. Uzunyan,[49] R. Van Kooten,[51] W.M. van Leeuwen,[32] N. Varelas,[48] E.W. Varnes,[44] I.A. Vasilyev,[37] P. Verdier,[19] L.S. Vertogradov,[34] M. Verzocchi,[47] M. Vesterinen,[43] D. Vilanova,[17] P. Vokac,[9] H.D. Wahl,[46] M.H.L.S. Wang,[47] J. Warchol,[53] G. Watts,[79] M. Wayne,[53] M. Weber^{h},[47] L. Welty-Rieger,[50] A. White,[75] D. Wicke,[25] M.R.J. Williams,[41] G.W. Wilson,[55] M. Wobisch,[57] D.R. Wood,[59] T.R. Wyatt,[43] Y. Xie,[47] C. Xu,[60] S. Yacoob,[50] R. Yamada,[47] W.-C. Yang,[43] T. Yasuda,[47] Y.A. Yatsunenko,[34] Z. Ye,[47] H. Yin,[47] K. Yip,[70] S.W. Youn,[47] J. Yu,[75] S. Zelitch,[78] T. Zhao,[79] B. Zhou,[60] J. Zhu,[60] M. Zielinski,[68] D. Zieminska,[51] and L. Zivkovic[74]

(The D0 Collaboration*)

[1] *Universidad de Buenos Aires, Buenos Aires, Argentina*
[2] *LAFEX, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, Brazil*
[3] *Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil*
[4] *Universidade Federal do ABC, Santo André, Brazil*
[5] *Instituto de Física Teórica, Universidade Estadual Paulista, São Paulo, Brazil*
[6] *University of Science and Technology of China, Hefei, People's Republic of China*
[7] *Universidad de los Andes, Bogotá, Colombia*
[8] *Charles University, Faculty of Mathematics and Physics, Center for Particle Physics, Prague, Czech Republic*
[9] *Czech Technical University in Prague, Prague, Czech Republic*
[10] *Center for Particle Physics, Institute of Physics, Academy of Sciences of the Czech Republic, Prague, Czech Republic*
[11] *Universidad San Francisco de Quito, Quito, Ecuador*
[12] *LPC, Université Blaise Pascal, CNRS/IN2P3, Clermont, France*
[13] *LPSC, Université Joseph Fourier Grenoble 1, CNRS/IN2P3, Institut National Polytechnique de Grenoble, Grenoble, France*
[14] *CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France*
[15] *LAL, Université Paris-Sud, CNRS/IN2P3, Orsay, France*
[16] *LPNHE, Universités Paris VI and VII, CNRS/IN2P3, Paris, France*
[17] *CEA, Irfu, SPP, Saclay, France*
[18] *IPHC, Université de Strasbourg, CNRS/IN2P3, Strasbourg, France*
[19] *IPNL, Université Lyon 1, CNRS/IN2P3, Villeurbanne, France and Université de Lyon, Lyon, France*
[20] *III. Physikalisches Institut A, RWTH Aachen University, Aachen, Germany*
[21] *Physikalisches Institut, Universität Freiburg, Freiburg, Germany*
[22] *II. Physikalisches Institut, Georg-August-Universität Göttingen, Göttingen, Germany*
[23] *Institut für Physik, Universität Mainz, Mainz, Germany*
[24] *Ludwig-Maximilians-Universität München, München, Germany*
[25] *Fachbereich Physik, Bergische Universität Wuppertal, Wuppertal, Germany*
[26] *Panjab University, Chandigarh, India*
[27] *Delhi University, Delhi, India*
[28] *Tata Institute of Fundamental Research, Mumbai, India*
[29] *University College Dublin, Dublin, Ireland*
[30] *Korea Detector Laboratory, Korea University, Seoul, Korea*
[31] *CINVESTAV, Mexico City, Mexico*
[32] *Nikhef, Science Park, Amsterdam, the Netherlands*
[33] *Radboud University Nijmegen, Nijmegen, the Netherlands and Nikhef, Science Park, Amsterdam, the Netherlands*
[34] *Joint Institute for Nuclear Research, Dubna, Russia*
[35] *Institute for Theoretical and Experimental Physics, Moscow, Russia*
[36] *Moscow State University, Moscow, Russia*
[37] *Institute for High Energy Physics, Protvino, Russia*
[38] *Petersburg Nuclear Physics Institute, St. Petersburg, Russia*
[39] *Institució Catalana de Recerca i Estudis Avançats (ICREA) and Institut de Física d'Altes Energies (IFAE), Barcelona, Spain*
[40] *Stockholm University, Stockholm and Uppsala University, Uppsala, Sweden*
[41] *Lancaster University, Lancaster LA1 4YB, United Kingdom*
[42] *Imperial College London, London SW7 2AZ, United Kingdom*
[43] *The University of Manchester, Manchester M13 9PL, United Kingdom*
[44] *University of Arizona, Tucson, Arizona 85721, USA*
[45] *University of California Riverside, Riverside, California 92521, USA*
[46] *Florida State University, Tallahassee, Florida 32306, USA*
[47] *Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA*
[48] *University of Illinois at Chicago, Chicago, Illinois 60607, USA*
[49] *Northern Illinois University, DeKalb, Illinois 60115, USA*
[50] *Northwestern University, Evanston, Illinois 60208, USA*

[51]Indiana University, Bloomington, Indiana 47405, USA
[52]Purdue University Calumet, Hammond, Indiana 46323, USA
[53]University of Notre Dame, Notre Dame, Indiana 46556, USA
[54]Iowa State University, Ames, Iowa 50011, USA
[55]University of Kansas, Lawrence, Kansas 66045, USA
[56]Kansas State University, Manhattan, Kansas 66506, USA
[57]Louisiana Tech University, Ruston, Louisiana 71272, USA
[58]Boston University, Boston, Massachusetts 02215, USA
[59]Northeastern University, Boston, Massachusetts 02115, USA
[60]University of Michigan, Ann Arbor, Michigan 48109, USA
[61]Michigan State University, East Lansing, Michigan 48824, USA
[62]University of Mississippi, University, Mississippi 38677, USA
[63]University of Nebraska, Lincoln, Nebraska 68588, USA
[64]Rutgers University, Piscataway, New Jersey 08855, USA
[65]Princeton University, Princeton, New Jersey 08544, USA
[66]State University of New York, Buffalo, New York 14260, USA
[67]Columbia University, New York, New York 10027, USA
[68]University of Rochester, Rochester, New York 14627, USA
[69]State University of New York, Stony Brook, New York 11794, USA
[70]Brookhaven National Laboratory, Upton, New York 11973, USA
[71]Langston University, Langston, Oklahoma 73050, USA
[72]University of Oklahoma, Norman, Oklahoma 73019, USA
[73]Oklahoma State University, Stillwater, Oklahoma 74078, USA
[74]Brown University, Providence, Rhode Island 02912, USA
[75]University of Texas, Arlington, Texas 76019, USA
[76]Southern Methodist University, Dallas, Texas 75275, USA
[77]Rice University, Houston, Texas 77005, USA
[78]University of Virginia, Charlottesville, Virginia 22901, USA
[79]University of Washington, Seattle, Washington 98195, USA

(Dated: September 14, 2011)

We report an updated measurement of the *CP*-violating phase, $\phi_s^{J/\psi\phi}$, and the decay-width difference for the two mass eigenstates, $\Delta\Gamma_s$, from the flavor-tagged decay $B_s^0 \to J/\psi\phi$. The data sample corresponds to an integrated luminosity of 8.0 fb^{-1} accumulated with the D0 detector using $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV produced at the Fermilab Tevatron collider. The 68% bayesian credibility intervals, including systematic uncertainties, are $\Delta\Gamma_s = 0.163^{+0.065}_{-0.064}$ ps^{-1} and $\phi_s^{J/\psi\phi} = -0.55^{+0.38}_{-0.36}$. The p-value for the Standard Model point is 29.8%.

PACS numbers: 13.25.Hw, 11.30.Er

I. INTRODUCTION

The meson-antimeson mixing and the phenomenon of charge-conjugation-parity *(CP)* violation in neutral mesons systems are key problems of particle physics. In the standard model (SM), the light (L) and heavy (H) mass eigenstates of the B_s^0 system are expected to have sizeable mass and decay width differences: $\Delta M_s \equiv M_H - M_L$ and $\Delta\Gamma_s \equiv \Gamma_L - \Gamma_H$. The two mass eigenstates are expected to be almost pure *CP* eigenstates. The *CP*-violating phase that appears in $b \to c\overline{c}s$ decays is due to the interference of the decay with and without mixing, and it is predicted [1] to be $\phi_s^{J/\psi\phi} = -2\beta_s^{SM} = 2\arg[-V_{tb}V_{ts}^*/V_{cb}V_{cs}^*] = -0.038 \pm 0.002$, where V_{ij} are elements of the Cabibbo-Kobayashi-Maskawa quark-mixing matrix [2]. New phenomena [3–23] may alter the observed phase to $\phi_s^{J/\psi\phi} \equiv -2\beta_s \equiv -2\beta_s^{SM} + \phi_s^{\Delta}$. A significant deviation of $\phi_s^{J/\psi\phi}$ from its small SM value would indicate the presence of processes beyond SM.

The analysis of the decay chain $B_s^0 \to J/\psi\phi$, $J/\psi \to \mu^+\mu^-$, $\phi \to K^+K^-$ separates the *CP*-even and *CP*-odd states using the angular distributions of the decay products. It is a unique feature of the decay $B_s^0 \to J/\psi\phi$ that because of the sizeable lifetime difference between the two mass eigenstates, there is a sensitivity to $\phi_s^{J/\psi\phi}$ even in the absence of the flavor tagging information. The first direct constraint on $\phi_s^{J/\psi\phi}$ [24, 25] was derived by analysing $B_s^0 \to J/\psi\phi$ decays where the flavor (i.e., B_s^0 or $\overline{B}_s^0$) at the time of production was not determined ("tagged"). It was followed by an improved analysis [26], based on 2.8 fb^{-1} of integrated luminosity, that included the information on the B_s^0 flavor at production. In ad-

*with visitors from [a]Augustana College, Sioux Falls, SD, USA, [b]The University of Liverpool, Liverpool, UK, [c]UPIITA-IPN, Mexico City, Mexico, [d]SLAC, Menlo Park, CA, USA, [e]University College London, London, UK, [f]Centro de Investigacion en Computacion - IPN, Mexico City, Mexico, [g]ECFM, Universidad Autonoma de Sinaloa, Culiacán, Mexico, and [h]Universität Bern, Bern, Switzerland. ‡Deceased.

dition, the CDF collaboration has performed a measurement [27] of $\phi_s^{J/\psi\phi}$ using 1.35 fb^{-1} of data. After the submission of this Article, new measurements of the CP violation parameters in the $B_s^0 \to J/\psi\phi$ decay have been published by the CDF [28] and the LHCb [29] Collaborations.

In this Article, we present new results from the time-dependent amplitude analysis of the decay $B_s^0 \to J/\psi\phi$ using a data sample corresponding to an integrated luminosity of 8.0 fb^{-1} collected with the D0 detector [30] at the Fermilab Tevatron Collider. In addition to the increase in the size of the data sample used in the analysis, we also take into account the $\mathcal{S}$-wave K^+K^- under the ϕ peak that has been suggested [31] to contribute between 5-10%. We measure $\Delta\Gamma_s$; the average lifetime of the B_s^0 system, $\overline{\tau}_s = 1/\overline{\Gamma}_s$, where $\overline{\Gamma}_s \equiv (\Gamma_H + \Gamma_L)/2$; and the CP-violating phase $\phi_s^{J/\psi\phi}$. Section II briefly describes the D0 detector. Section III presents the event reconstruction and the data set. Sections IV and V describe the event selection requirements and the procedure of determining the flavor of the initial state of the B_s^0 candidate. In Sec. VI we describe the analysis formalism and the fitting method, present fit results, and discuss systematic uncertainties in the results. We obtain the bayesian credibility intervals for physics parameters using a procedure based on the Markov Chain Monte Carlo (MCMC) technique, presented in Sec. VII. We summarize and discuss the results in Sec. VIII.

II. DETECTOR

The D0 detector consists of a central tracking system, calorimetry system and muon detectors, as detailed in Refs. [30, 32, 33]. The central tracking system comprises a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located inside a 1.9 T superconducting solenoidal magnet. The tracking system is designed to optimize tracking and vertexing for pseudorapidities $|\eta| < 3$, where $\eta = -\ln[\tan(\Theta/2)]$, and Θ is the polar angle with respect to the proton beam direction.

The SMT can reconstruct the $p\overline{p}$ interaction vertex (PV) for interactions with at least three tracks with a precision of 40 μm in the plane transverse to the beam direction and determine the impact parameter of any track relative to the PV with a precision between 20 and 50 μm, depending on the the number of hits in the SMT.

The muon detector is positioned outside the calorimeter. It consists of a central muon system covering the pseudorapidity region $|\eta| < 1$ and a forward muon system covering the pseudorapidity region $1 < |\eta| < 2$. Both central and forward systems consist of a layer of drift tubes and scintillators inside 1.8 T toroidal magnets and two similar layers outside the toroids.

The trigger and data acquisition systems are designed to accommodate the high instantaneous luminosities of Tevatron Run II.

III. DATA SAMPLE AND EVENT RECONSTRUCTION

The analysis presented here is based on data accumulated between February 2002 and June 2010. Events are collected with a mixture of single- and dimuon triggers. Some triggers require track displacement with respect to the primary vertex (large track impact parameter). Since this condition biases the B_s^0 lifetime measurement, the events selected exclusively by these triggers are removed from our sample.

Candidate $B_s^0 \to J/\psi\phi$, $J/\psi \to \mu^+\mu^-$, $\phi \to K^+K^-$ events are required to include two opposite charge muons accompanied by two opposite charge tracks. Both muons are required to be detected in the muon chambers inside the toroid magnet, and at least one of the muons is required to be also detected outside the toroid. Each of the four final-state tracks is required to have at least one SMT hit.

To form B_s^0 candidates, muon pairs in the invariant mass range 3.096 ± 0.350 GeV, consistent with J/ψ decay, are combined with pairs of opposite charge tracks (assigned the kaon mass) consistent with production at a common vertex, and with an invariant mass in the range 1.019 ± 0.030 GeV. A kinematic fit under the B_s^0 decay hypothesis constrains the dimuon invariant mass to the world-average J/ψ mass [34] and constrains the four-track system to a common vertex.

Trajectories of the four B_s^0 decay products are adjusted according to the decay-vertex kinematic fit. The re-adjusted track parameters are used in the calculation of the B_s^0 candidate mass and decay time, and of the three angular variables characterising the decay as defined later. B_s^0 candidates are required to have an invariant mass in the range 5.37 ± 0.20 GeV. In events where multiple candidates satisfy these requirements, we select the candidate with the best decay vertex fit probability.

To reconstruct the PV, we select tracks that do not originate from the candidate B_s^0 decay, and apply a constraint to the average beam-spot position in the transverse plane. We define the signed decay length of a B_s^0 meson, L_{xy}^B, as the vector pointing from the PV to the decay vertex, projected on the B_s^0 transverse momentum p_T. The proper decay time of a B_s^0 candidate is given by $t = M_{B_s}\vec{L}_{xy}^B \cdot \vec{p}/(p_T^2)$ where M_{B_s} is the world-average B_s^0 mass [34], and $\vec{p}$ is the particle momentum. The distance in the beam direction between the PV and the B_s^0 vertex is required to be less than 5 cm. Approximately 5 million events are accepted after the selection described in this section.

IV. BACKGROUND SUPPRESSION

The selection criteria are designed to optimimize the measurement of $\phi_s^{J/\psi\phi}$ and $\Delta\Gamma_s$. Most of the background is due to directly produced J/ψ mesons accompanied by

tracks arising from hadronization. This "prompt" background is distinguished from the "non-prompt", or "inclusive $B \to J/\psi + X$" background, where the J/ψ meson is a product of a b-hadron decay while the tracks forming the ϕ candidate emanate from a multi-body decay of a b hadron or from hadronization. Two different event selection approaches are used, one based on a multi-variate technique, and one based on simple limits on kinematic and event quality parameters.

IV-A. Signal and background simulation

Three Monte Carlo simulated samples are used to study background suppression: signal, prompt background, and non-prompt background. All three are generated with PYTHIA [35]. Hadronization is also done in PYTHIA, but all hadrons carrying heavy flavors are passed on to EvtGen [36] to model their decays. The prompt background MC sample consists of $J/\psi \to \mu^+\mu^-$ decays produced in $gg \to J/\psi g$, $gg \to J/\psi\gamma$, and $g\gamma \to J/\psi g$ processes. The signal and non-prompt background samples are generated from primary $b\bar{b}$ pair production with all b hadrons being produced inclusively and the J/ψ mesons forced into $\mu^+\mu^-$ decays. For the signal sample, events with a B_s^0 are selected, their decays to $J/\psi\phi$ are implemented without mixing and with uniform angular distributions, and the B_s^0 mean lifetime is set to $\overline{\tau}_s = 1.464$ ps. There are approximately 10^6 events in each background and the signal MC samples. All events are passed through a full GEANT-based [37] detector simulation. To take into account the effects of multiple interactions at high luminosity, hits from randomly triggered $p\bar{p}$ collisions are overlayed on the digitized hits from MC. These events are reconstructed with the same program as used for data. The three samples are corrected so that the p_T distributions of the final state particles in $B_s^0 \to J/\psi\phi$ decays match those in data (see Appendix B).

IV-B. Multivariate event selection

To discriminate the signal from background events, we use the TMVA package [38]. In preliminary studies using MC simulation, the Boosted Decision Tree (BDT) algorithm was found to demonstrate the best performance. Since prompt and non-prompt backgrounds have different kinematic behavior, we train two discriminants, one for each type of background. We use a set of 33 variables for the prompt background and 35 variables for the non-prompt background. The variables and more details of the BDT method are given in Appendix A.

The BDT training is performed using a subset of the MC samples, and the remaining events are used to test the training. The signal MC sample has about 84k events, the prompt background has 29k events, and the non-prompt background has 39k events. Figure 1 shows the BDT output discriminant for the prompt and non-prompt cases.

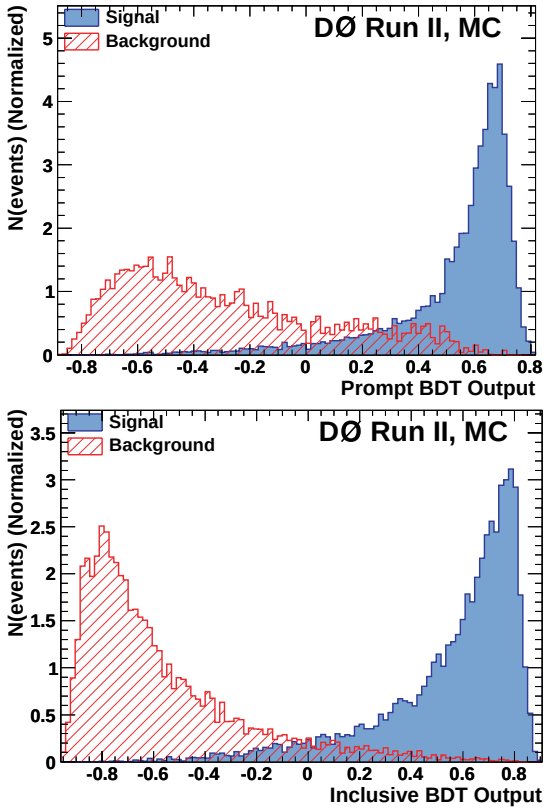


FIG. 1: (color online). BDT discriminant output for the prompt (top) and non-prompt (bottom) classifiers. The signal and background events are taken from simulation. Events used for BDT training are excluded from these samples.

IV-C. Selection Criteria

To choose the best set of criteria for the two BDT discriminants, we first step through the values of both BDT discriminants from -0.4 to 0.8 in increments of 0.01 and measure the B_s^0 signal yield for each choice of cuts. Next, we define 14 signal yield regions between 4000 and 7000 events, and for each region choose the pair of BDT cuts which gives the highest significance $S/\sqrt{S+B}$, where S (B) is the number of signal (background) events in the data sample. The 14 points, in increasing order of the signal size S, are shown in Table I. Figure 2 shows the number of signal events as a function of the total number of events for the 14 points. As the BDT criteria are loosened, the total number of events increases by a factor of ten, while the number of signal events increases by about 50%.

As a test of possible detrimental effects of training on variables with low separation power, we have repeated the above procedure using only the variables whose im-

portance (see Appendix A) exceeds 0.01, giving 18 variables for the prompt background and 13 variables for the non-prompt background. The resulting number of background events for a given number of signal events is larger by about 10%. Therefore, we proceed with the original number of variables.

Criteria Set	S	$S+B$	Non-prompt BDT	Prompt BDT
0	4550	38130	0.45	0.42
1	4699	44535	0.45	0.29
2	5008	53942	0.39	0.35
3	5213	64044	0.36	0.30
4	5364	72602	0.33	0.28
5	5558	85848	0.13	0.41
6	5767	100986	0.21	0.29
7	5988	120206	0.13	0.29
8	6097	134255	0.07	0.29
9	6399	189865	0.04	0.10
10	6489	254022	−0.05	−0.01
11	6608	294949	−0.13	0.00
12	6594	364563	−0.18	−0.14
13	6695	461744	−0.35	−0.08

TABLE I: Numbers of signal and signal-plus-background events for different sets of BDT criteria, shown in the last two columns, that give the largest value of $S/\sqrt{S+B}$ for a given S.

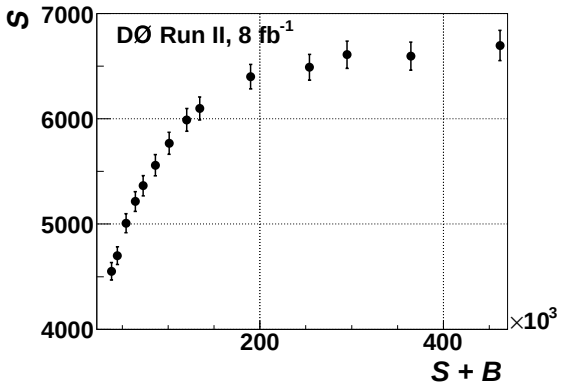


FIG. 2: Number of $B_s^0 \to J/\psi\phi$ signal events as a function of the total number of events for the 14 criteria sets considered.

The choice of the final cut on the BDT output is based on an ensemble study. For each point in Table I, we perform a maximum-likelihood fit to the event distribution in the 2-dimensional (2D) space of B_s^0 candidate mass and proper time. This 2D fit provides a parametrization of the background mass and proper time distribution. We then generate pseudo-experiments in the 5D space of B_s^0 candidate mass, proper time, and three independent angles of decay products, using as input the parameters as obtained in a preliminary study, and the background from the 2D fit. We perform a 5D maximum likelihood fit on the ensembles and compare the distributions of the statistical uncertainties of $\phi_s^{J/\psi\phi}$ ($\sigma(\phi_s^{J/\psi\phi})$) and $\Delta\Gamma_s$ ($\sigma(\Delta\Gamma_s)$) for the different sets of criteria. The dependence of the mean values of $\sigma(\phi_s^{J/\psi\phi})$ and $\sigma(\Delta\Gamma_s)$ on the number of signal events is shown in Figs. 3(a) and 3(b). The mean statistical uncertainties of both $\phi_s^{J/\psi\phi}$ and $\Delta\Gamma_s$ systematically decrease with increasing signal, favoring looser cuts. The gain in the parameter resolution is slower for the three loosest criteria, while the total number of events doubles from about 0.25×10^6 to 0.5×10^6. The fits used for these ensemble tests were simplified, therefore the magnitude of the predicted uncertainty is expected to underestimate the final measured precision. However, the general trends should be valid.

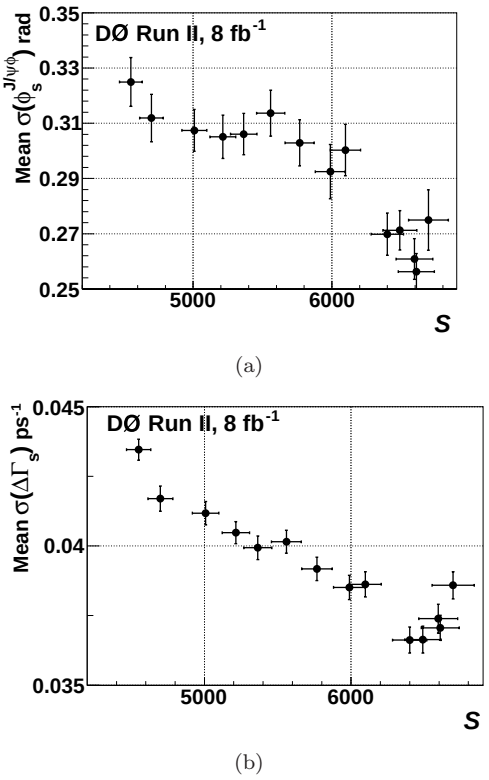


FIG. 3: Ensemble study results for (a) mean value of $\sigma(\phi_s)$ as a function of the number of signal events and (b) mean value of $\sigma(\Delta\Gamma_s)$ as a function of the number of signal events.

Based on these results, we choose the sample that contains about 6500 signal events, (labeled "Set 10" in Table I) as a final selection and refer to it as the "BDT selection". Figure 17 in Appendix A shows the ratios of the normalized distributions of the three angles (see Section VI) and the lifetime before and after the BDT selection. The ratios are consistent with unity, which means that the BDT requirements do not significantly alter these distributions.

IV-D. Simple Selection

We select a second event sample by applying criteria on event quality and kinematic quantities. We use the consistency of the results obtained for the BDT and for this sample as a measure of systematic effects related to imperfect modeling of the detector acceptance and of the selection requirements.

The criteria are the same as in Refs. [24] and [26]. Each of the four tracks is required to have at least two SMT hits and at least eight hits in SMT or CFT. We require minimum momentum in the transverse plane p_T for B_s^0, ϕ, and K meson candidates of 6.0 GeV, 1.5 GeV, and 0.7 GeV, respectively. Muons are required to have p_T above 1.5 GeV. For events in the central rapidity region (an event is considered to be central if the higher p_T muon has $|\eta(\mu_{\rm leading})| < 1$), we require the transverse momentum of the J/ψ meson to exceed 4 GeV. In addition, J/ψ candidates are accepted if the invariant mass of the muon pair is in the range 3.1 ± 0.2 GeV. Events are required to satisfy the condition $\sigma(t) < 0.2$ ps where $\sigma(t)$ is the uncertainty on the decay proper time obtained from the propagation of the uncertainties in the decay-vertex kinematic fit, the primary vertex position, and the B_s^0 candidate transverse momentum. We refer to this second sample as the "Square-cuts" sample.

V. FLAVOR TAGGING

At the Tevatron, b quarks are mostly produced in $b\overline{b}$ pairs. The flavor of the initial state of the B_s^0 candidate is determined by exploiting properties of particles produced by the other b hadron ("opposite-side tagging", or OST). The OST-discriminating variables x_i are based primarily on the presence of a muon or an electron from the semileptonic decay or the decay vertex charge of the other b hadron produced in the $p\overline{p}$ interaction.

For the initial b quark, the probability density function (PDF) for a given variable x_i is denoted as $f_i^b(x_i)$, while for the initial $\overline{b}$ quark it is denoted as $f_i^{\overline{b}}(x_i)$. The combined tagging variable y is defined as:

$$y = \prod_{i=1}^{n} y_i;\ y_i = \frac{f_i^{\overline{b}}(x_i)}{f_i^b(x_i)}. \qquad (1)$$

A given variable x_i can be undefined for some events. For example, there are events that don't contain an identified muon from the opposite side. In this case, the corresponding variable y_i is set to 1.

In this way the OST algorithm assigns to each event a value of the predicted tagging parameter $d = (1-y)/(1+y)$ in the range $[-1,1]$, with $d > 0$ tagged as an initial b quark and $d < 0$ tagged as an initial $\overline{b}$ quark. Larger $|d|$ values correspond to higher tagging confidence. In events where no tagging information is available d is set to zero. The efficiency ϵ of the OST, defined as fraction of the number of candidates with $d \neq 0$, is 18%. The OST-discriminating variables and algorithm are described in detail in Ref. [39].

The tagging dilution $\mathcal{D}$ is defined as

$$\mathcal{D} = \frac{N_{\rm cor} - N_{\rm wr}}{N_{\rm cor} + N_{\rm wr}}, \qquad (2)$$

where $N_{\rm cor}$ ($N_{\rm wr}$) is the number of events with correctly (wrongly) identified initial B-meson flavor. The dependence of the tagging dilution on the tagging parameter d is calibrated with data for which the flavor (B or $\overline{B}$) is known.

V-A. OST calibration

The dilution calibration is based on four independent $B_d^0 \to \mu\nu D^{*\pm}$ data samples corresponding to different time periods, denoted IIa, IIb1, IIb2, and IIb3, with different detector configurations and different distributions of instantaneous luminosity. The Run IIa sample was used in Ref. [39].

For each sample we perform an analysis of the $B_d^0 - \overline{B}_d^0$ oscillations described in Ref. [40]. We divide the samples in five ranges of the tagging parameter $|d|$, and for each range we obtain a mean value of the dilution $|\mathcal{D}|$. The mixing frequency ΔM_d is fitted simultaneously and is found to be stable and consistent with the world average value. The measured values of the tagging dilution $|\mathcal{D}|$ for the four data samples above, in different ranges of $|d|$, are shown in Fig. 4. The dependence of the dilution on $|d|$ is parametrized as

$$|\mathcal{D}| = \frac{p_0}{(1+\exp((p_1-|d|)/p_2))} - \frac{p_0}{(1+\exp(p_1/p_2))}. \qquad (3)$$

and the function is fitted to the data. All four measurements are in good agreement and hence a weighted average is taken.

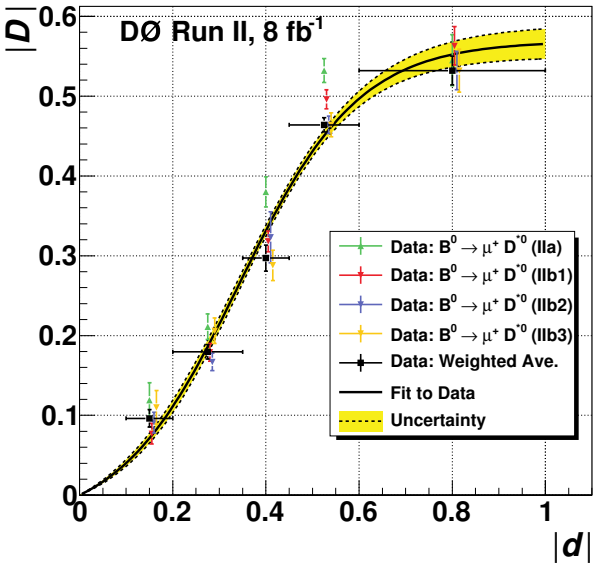


FIG. 4: (color online). Parametrization of the dilution $|\mathcal{D}|$ as a function of the tagging parameter $|d|$ for the combined opposite-side tagger. The curve is the result of the weighted fit to four self-tagging control data samples (see text).

VI. MAXIMUM LIKELIHOOD FIT

We perform a six-dimensional (6D) unbinned maximum likelihood fit to the proper decay time and its uncertainty, three decay angles characterizing the final state, and the mass of the B_s^0 candidate. We use events for which the invariant mass of the K^+K^- pair is within the range 1.01 – 1.03 GeV. There are 104683 events in the BDT-based sample and 66455 events in the Square-cuts sample. We adopt the formulae and notation of Ref. [41]. The normalized functional form of the differential decay rate includes an $\mathcal{S}$-wave KK contribution in addition to the dominant $\mathcal{P}$-wave $\phi \to K^+K^-$ decay. To model the distributions of the signal and background we use the software library RooFit [42].

VI-A. Signal model

The angular distribution of the signal is expressed in the transversity basis [43]. In the coordinate system of the J/ψ rest frame, where the ϕ meson moves in the x direction, the z axis is perpendicular to the decay plane of $\phi \to K^+K^-$, and $p_y(K^+) \geq 0$. The transversity polar and azimuthal angles θ and φ describe the direction of the positively-charged muon, while ψ is the angle between $\vec{p}(K^+)$ and $-\vec{p}(J/\psi)$ in the ϕ rest frame. The angles are shown in Fig. 5.

In this basis, the decay amplitude of the B_s^0 and $\overline{B}_s^0$ mesons is decomposed into three independent components corresponding to linear polarization states of the vector mesons J/ψ and ϕ, which are polarized either longitudinally (0) or transversely to their direction of motion, and parallel ($\|$) or perpendicular ($\perp$) to each other.

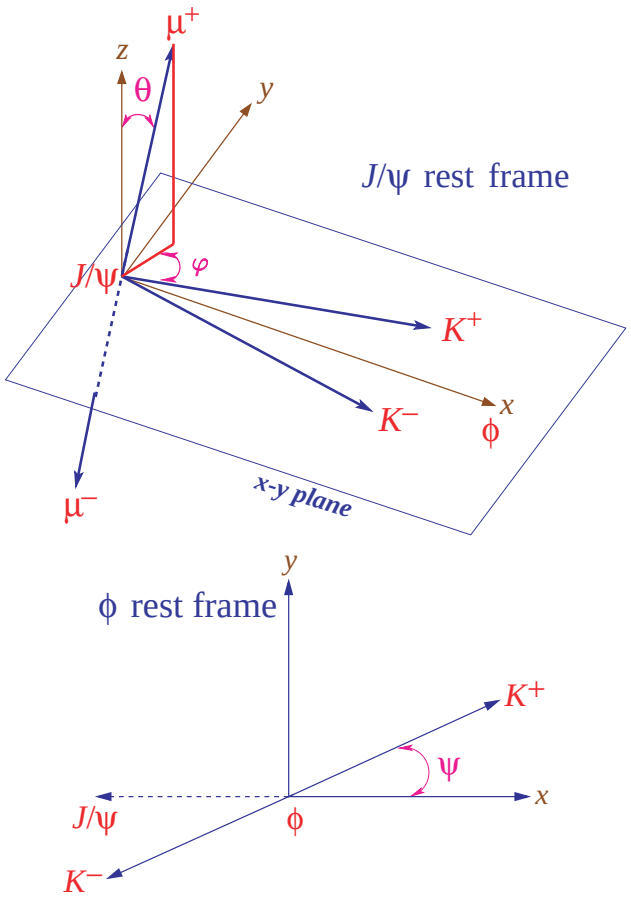


FIG. 5: (color online). Definition of the angle ψ, and the transversity angles θ and φ.

The time dependence of amplitudes $\mathcal{A}_i(t)$ and $\bar{\mathcal{A}}_i(t)$ (i denotes one of $\{\|, \perp, 0\}$), for B_s^0 and $\overline{B}_s^0$ states to reach the final state $J/\psi\ \phi$ is:

$$\begin{aligned} \mathcal{A}_i(t) &= F(t)\left[E_+(t) \pm e^{2i\beta_s} E_-(t)\right] a_i\,, \\ \bar{\mathcal{A}}_i(t) &= F(t)\left[\pm E_+(t) + e^{-2i\beta_s} E_-(t)\right] a_i\,, \end{aligned} \qquad (4)$$

where

$$F(t) = \frac{e^{-\Gamma_s t/2}}{\sqrt{\tau_H + \tau_L \pm \cos 2\beta_s\,(\tau_L - \tau_H)}}\,, \qquad (5)$$

and τ_H and τ_L are the lifetimes of the heavy and light B_s^0 eigenstates.

In the above equations the upper sign indicates a CP-even final state, the lower sign indicates a CP-odd final state,

$$E_\pm(t) \equiv \frac{1}{2}\left[e^{(\frac{-\Delta\Gamma_s}{4} + i\frac{\Delta M_s}{2})t} \pm e^{-(\frac{-\Delta\Gamma_s}{4} + i\frac{\Delta M_s}{2})t}\right], \qquad (6)$$

and the amplitude parameters a_i give the time-integrated decay rate to each of the polarization states, $|a_i|^2$, satisfying:

$$\sum_i |a_i|^2 = 1\,. \tag{7}$$

The interference terms $A_\parallel - A_\perp$ and $A_0 - A_\perp$ are proportional to $(e^{-\Gamma_H t} - e^{-\Gamma_L t})\sin\phi_s^{J/\psi\phi}$. Also, if $\cos\phi_s^{J/\psi\phi}$ is significantly different from unity, the decay rates of the CP-even and CP-odd components have two slopes each.

The normalized probability density functions P_B and $P_{\bar{B}}$ for B and $\bar{B}$ mesons in the variables t, $\cos\psi$, $\cos\theta$, and φ, are

$$\begin{aligned} P_B(\theta,\varphi,\psi,t) &= \frac{9}{16\pi}|\mathbf{A}(t)\times\hat{n}|^2, \\ P_{\bar{B}}(\theta,\varphi,\psi,t) &= \frac{9}{16\pi}|\bar{\mathbf{A}}(\mathbf{t})\times\hat{n}|^2, \end{aligned} \tag{8}$$

where $\hat{n}$ is the muon momentum direction in the J/ψ rest frame,

$$\hat{n} = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)\,, \tag{9}$$

and $\mathbf{A}(t)$ and $\bar{\mathbf{A}}(t)$ are complex vector functions of time defined as

$$\begin{aligned} \mathbf{A}(t) &= \left(\mathcal{A}_0(t)\cos\psi, -\frac{\mathcal{A}_\parallel(t)\sin\psi}{\sqrt{2}}, i\frac{\mathcal{A}_\perp(t)\sin\psi}{\sqrt{2}}\right), \\ \bar{\mathbf{A}}(t) &= \left(\bar{\mathcal{A}}_0(t)\cos\psi, -\frac{\bar{\mathcal{A}}_\parallel(t)\sin\psi}{\sqrt{2}}, i\frac{\bar{\mathcal{A}}_\perp(t)\sin\psi}{\sqrt{2}}\right). \end{aligned} \tag{10}$$

The values of $\mathcal{A}_i(t)$ at $t=0$ are denoted as A_i. They are related to the parameters a_i by

$$\begin{aligned} |A_\perp|^2 &= \frac{|a_\perp|^2 y}{1+(y-1)|a_\perp|^2}, \\ |A_{||}|^2 &= \frac{|a_{||}|^2}{1+(y-1)|a_\perp|^2}, \\ |A_0|^2 &= \frac{|a_0|^2}{1+(y-1)|a_\perp|^2}, \end{aligned} \tag{11}$$

where $y \equiv (1-z)/(1+z)$ and $z \equiv \cos 2\beta_s \Delta\Gamma_s/(2\overline{\Gamma}_s)$. By convention, the phase of A_0 is set to zero and the phases of the other two amplitudes are denoted by $\delta_{||}$ and $\delta_\perp$.

For a given event, the decay rate is the sum of the functions P_B and $P_{\bar{B}}$ weighted by the flavor tagging dilution factors $(1+\mathcal{D})/2$ and $(1-\mathcal{D})/2$, respectively.

The contribution from the decay to $J/\psi K^+K^-$ with the kaons in an $\mathcal{S}$-wave is expressed in terms of the $\mathcal{S}$-wave fraction F_S and a phase δ_s. The squared sum of the $\mathcal{P}$ and $\mathcal{S}$ waves is integrated over the KK mass. For the $\mathcal{P}$-wave, we assume the non-relativistic Breit-Wigner model

$$g(M(KK)) = \sqrt{\frac{\Gamma_\phi/2}{\Delta M(KK)}}\cdot\frac{1}{M(KK)-M_\phi+i\Gamma_\phi/2} \tag{12}$$

with the ϕ meson mass $M_\phi = 1.019$ GeV and width $\Gamma_\phi = 4.26$ MeV [34], and with $\Delta M(KK) = 1.03 - 1.01 = 0.02$ GeV.

For the $\mathcal{S}$-wave component, we assume a uniform distribution in the range $1.01 < M(KK) < 1.03$ GeV. We constrain the oscillation frequency to $\Delta M_s = 17.77 \pm 0.12$ ps^{-1}, as measured in Ref. [45]. Table II lists all physics parameters used in the fit.

Parameter	Definition
$\lvert A_0\rvert^2$	$\mathcal{P}$-wave longitudinal amplitude squared, at $t=0$
$A1$	$\lvert A_\parallel\rvert^2/(1-\lvert A_0\rvert^2)$
$\overline{\tau}_s$ (ps)	B_s^0 mean lifetime
$\Delta\Gamma_s$ (ps^{-1})	Heavy-light decay width difference
F_S	K^+K^- $\mathcal{S}$-wave fraction
β_s	CP-violating phase ($\equiv -\phi_s^{J/\psi\phi}/2$)
$\delta_\parallel$	$\arg(A_\parallel/A_0)$
$\delta_\perp$	$\arg(A_\perp/A_0)$
δ_s	$\arg(A_s/A_0)$

TABLE II: Definition of nine real measurables for the decay $B_s^0 \to J/\psi\phi$ used in the Maximum Likelihood fitting.

For the signal mass distribution we use a Gaussian function with a free mean value, width, and normalization. The function describing the signal rate in the 6D space is invariant under the combined transformation $\beta_s \to \pi/2 - \beta_s$, $\Delta\Gamma_s \to -\Delta\Gamma_s$, $\delta_\parallel \to 2\pi - \delta_\parallel$, $\delta_\perp \to \pi - \delta_\perp$, and $\delta_s \to \pi - \delta_s$. In addition, with a limited flavor-tagging power, there is an approximate symmetry around $\beta_s = 0$ for a given sign of $\Delta\Gamma_s$.

We correct the signal decay rate by a detector acceptance factor $\epsilon(\psi,\theta,\varphi)$ parametrized by coefficients of expansion in Legendre polynomials $P_k(\psi)$ and real harmonics $Y_{lm}(\theta,\varphi)$. The coefficients are obtained from Monte Carlo simulated samples, as described in Appendix B.

The signal decay time resolution is given by a Gaussian centered at zero and width given by the product of a global scale factor and the event-by-event uncertainty in the decay time measurement. The distribution of the uncertainty in the decay time measurement in the MC simulation is modeled by a superposition of five Gaussian functions. The background-subtracted signal distribution agrees well with the MC model, as seen in Fig. 6. Variations of the parameters within one sigma of the best fit are used to define two additional functions, also shown in the figure, that are used in alternative fits to estimate the systematic effect due to time resolution.

VI-B. Background model

The proper decay time distribution of the background is described by a sum of a prompt component, modeled

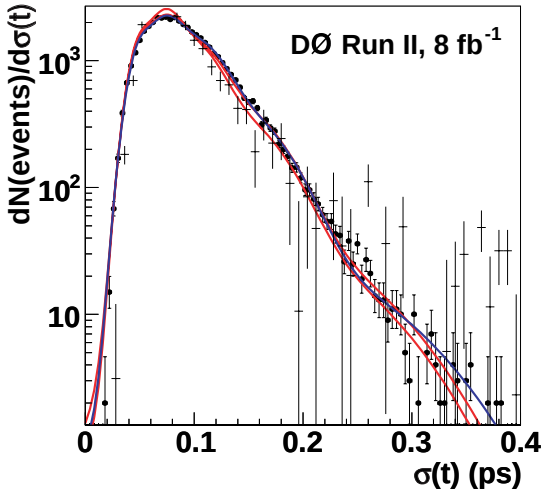


FIG. 6: (color online). The distribution of the uncertainty in the decay time for the signal, MC (squares) and background-subtracted data (crosses). The blue curve is the sum of five Gaussian functions fitted to the MC distribution. The two red lines are variations of the default function used in the studies of systematic effects.

as the same resolution function used in the signal decay time, and a non-prompt component. The non-prompt component is modeled as a superposition of one exponential decay for $t < 0$ and two exponential decays for $t > 0$, with free slopes and normalizations. The lifetime resolution is modeled by an exponential convoluted with a Gaussian function, with two separate parameters for prompt and non-prompt background. To allow for the possibility of the lifetime uncertainty to be systematically underestimated, we introduce a free scale factor.

The mass distributions of the two components of background are parametrized by low-order polynomials: a linear function for the prompt background and a quadratic function for the non-prompt background. The angular distribution of background is parametrized by Legendre and real harmonics expansion coefficients. A separate set of expansion coefficients c^k_{lm} and c^k_{lm}, with $k = 0$ or 2 and $l = 0, 1, 2$, is used for the prompt and non-prompt background. A preliminary fit is first performed with all 17 parameters ${c^{(P)}}^k_{lm}$ for prompt and 17 parameters ${c^{(NP)}}^k_{lm}$ for non-prompt allowed to vary. In subsequent fits those that converge at values within two standard deviations of zero are set to zero. Nine free parameters remain, five for non-prompt background: ${c^{(NP)}}^0_{1-1}$, ${c^{(NP)}}^0_{20}$, ${c^{(NP)}}^0_{22}$, ${c^{(NP)}}^2_{00}$, and ${c^{(NP)}}^2_{22}$, and four for prompt background: ${c^{(P)}}^0_{1-1}$, ${c^{(P)}}^0_{20}$, ${c^{(P)}}^0_{22}$, and ${c^{(P)}}^2_{2-1}$. All background parameters described above are varied simultaneously with physics parameters. In total, there are 36 parameters used in the fit. In addition to the nine physics parameters defined in Table II, they are: signal yield, mean mass and width, non-prompt background contribution, six non-prompt background lifetime parameters, four background time resolution parameters, one time resolution scale factor, three background mass distribution parameters, and nine parameters describing background angular distributions.

VI-C. Fit results

The maximum likelihood fit results for the nominal fit (Default), for two alternative time resolution functions, $\sigma_A(t)$ and $\sigma_B(t)$ shown in Fig. 6, and for an alternative $M(KK)$ dependence of the $\phi(1020) \to K^+K^-$ decay with the decay width increased by a factor of two are shown in Table III and Table IV. These alternative fits are used to estimate the systematic uncertainties. The fit assigns 5598 ± 113 (5050 ± 105) events to the signal for the BDT (Square-cuts) sample. Only the parameters whose values do not suffer from multi-modal effects are shown. A single fit does not provide meaningful point estimates and uncertainties for the four phase parameters. Their estimates are obtained using the MCMC technique. Figures 7 – 10 illustrate the quality of the fit for the background, for all data, and for the signal-enhanced sub-samples.

An independent measurement of the $\mathcal{S}$-wave fraction is described in Appendix C and the result is in agreement with F_S determined from the maximum likelihood fit.

Parameter	Default	$\sigma_A(t)$	$\sigma_B(t)$	$\Gamma_\phi = 8.52$ MeV
$\lvert A_0\rvert^2$	0.553 ± 0.016	0.553 ± 0.016	0.552 ± 0.016	0.553 ± 0.016
$\lvert A_\parallel\rvert^2/(1-\lvert A_0\rvert^2)$	0.487 ± 0.043	0.483 ± 0.043	0.485 ± 0.043	0.487 ± 0.043
$\overline{\tau}_s$ (ps)	1.417 ± 0.038	1.420 ± 0.037	1.417 ± 0.037	1.408 ± 0.434
$\Delta\Gamma_s$ (ps^{-1})	0.151 ± 0.058	0.136 ± 0.056	0.145 ± 0.057	0.170 ± 0.067
F_S	0.147 ± 0.035	0.149 ± 0.034	0.147 ± 0.035	0.147 ± 0.035

TABLE III: Maximum likelihood fit results for the BDT selection. The uncertainties are statistical.

Parameter	Default	$\sigma_A(t)$	$\sigma_B(t)$	$\Gamma_\phi = 8.52$ MeV
$\lvert A_0\rvert^2$	0.566 ± 0.017	0.564 ± 0.017	0.567 ± 0.017	0.566 ± 0.017
$\lvert A_\parallel\rvert^2/(1-\lvert A_0\rvert^2)$	0.579 ± 0.048	0.579 ± 0.048	0.577 ± 0.048	0.579 ± 0.048
$\overline{\tau}_s$ (ps)	1.439 ± 0.039	1.450 ± 0.038	1.457 ± 0.037	1.438 ± 0.042
$\Delta\Gamma_s$ (ps^{-1})	0.199 ± 0.058	0.194 ± 0.057	0.185 ± 0.056	0.202 ± 0.060
F_S	0.175 ± 0.035	0.169 ± 0.035	0.171 ± 0.035	0.175 ± 0.035

TABLE IV: Maximum likelihood fit results for the 'Square-cuts' sample.

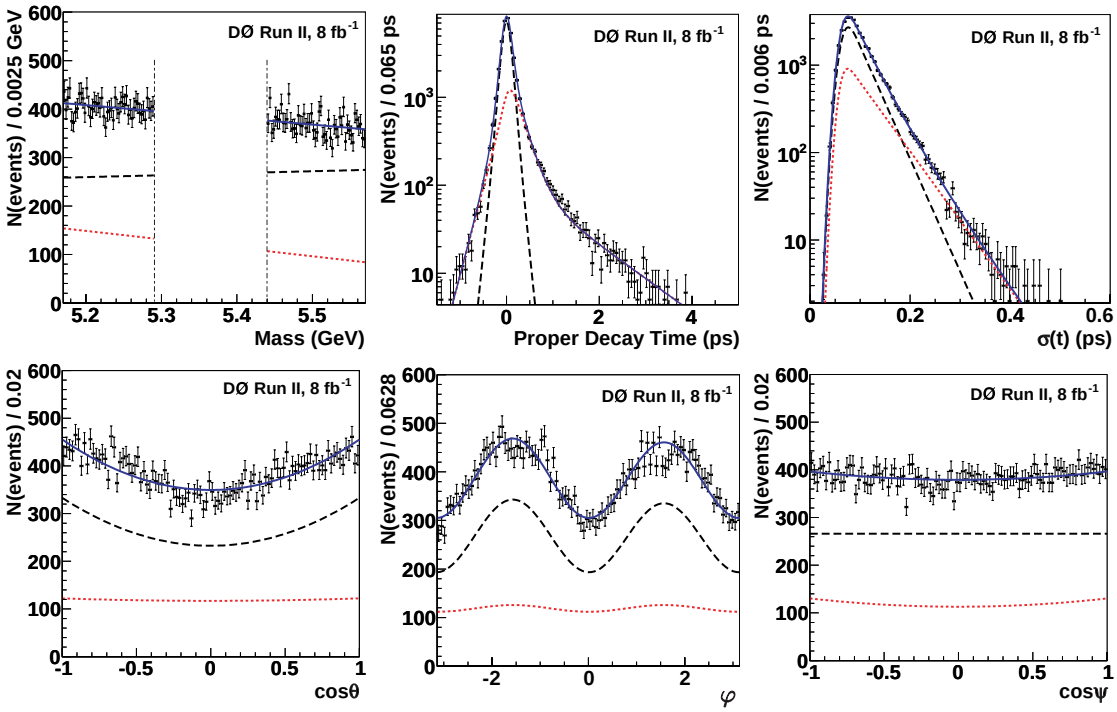


FIG. 7: (color online). The distributions in the background (B_s^0 mass sidebands) region of candidate mass, proper decay time, decay time uncertainty, transversity polar and azimuthal angles, and $\cos\psi$ for the BDT sample. The curves show the prompt (black dashed) and non-prompt (red dotted) components, and their sum (blue solid).

VI-D. Systematic uncertainties

There are several possible sources of systematic uncertainty in the measurements. These uncertainties are estimated as described below.

- **Flavor tagging**: The measured flavor mistag fraction suffers from uncertainties due to the limited number of events in the data samples for the decay $B_d^0 \to \mu\nu D^{(*)\pm}$. The nominal calibration of the flavor tagging dilution is determined as a weighted average of four samples separated by the running period. As an alternative, we use two separate calibration parameters, one for the data collected in running periods IIa and IIb1, and one for the IIb2 and IIb3 data. We also alter the nominal parameters by their uncertainties. We find the effects of the changes to the flavor mistag variation negligible.

- **Proper decay time resolution**: Fit results can be affected by the uncertainty of the assumed proper decay time resolution function. To assess the effect, we have used two alternative parametrizations obtained by random sampling of the resolution function.

- **Detector acceptance**: The effects of imperfect modeling of the detector acceptance and of the selection requirements are estimated by investigating the consistency of the fit results for the sample based on the BDT selection and on the Square-cuts selection. Although the overlap between the two samples is 70%, and some statistical differences are expected, we interpret the differences in the results as a measure of systematic effects.

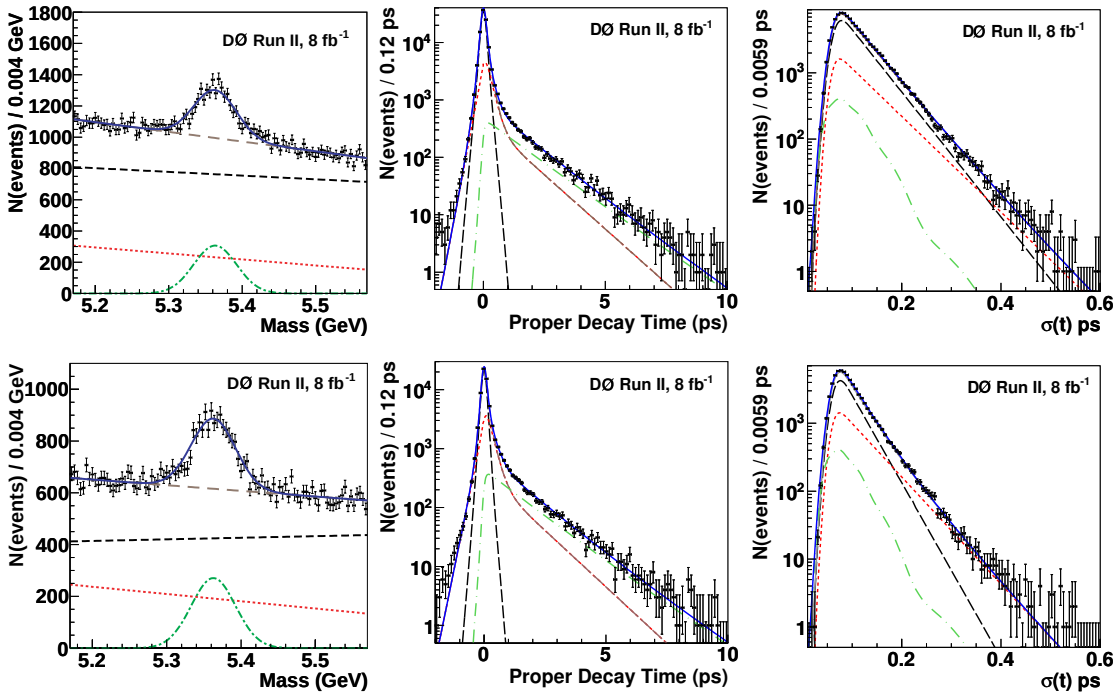


FIG. 8: (color online). Invariant mass, proper decay time, and proper decay time uncertainty distributions for B_s^0 candidates in the (top) BDT sample and (bottom) Square-cuts sample. The curves are projections of the maximum likelihood fit. Shown are the signal (green dashed-dotted curve), prompt background (black dashed curve), non-prompt background (red dotted curve), total background (brown long-dashed curve), and the sum of signal and total background (solid blue curve).

The two event selection approaches have different merits. The BDT-based approach uses more information on each event, and hence it allows a higher signal yield at lower background. However, it accepts signal events of lower quality (large vertex χ^2 or proper decay time uncertainty) that are rejected by the square cuts. Also, the BDT-based approach uses the $M(KK)$ distribution as a discriminant in the event selection, affecting the results for the parameters entering the $\mathcal{S}-\mathcal{P}$ interference term, particularly the $\mathcal{S}$-wave fraction F_S and the phase parameters.

The main difference between the two samples is in the kinematic ranges of final-state kaons, and so the angular acceptance functions and MC weights (see Appendix B) are different for the two samples. Imperfections in the modelling of the B_s^0 decay kinematics and estimated acceptances, and in the treatment of the MC weighting, are reflected in differences between fit results. The differences are used as an estimate of this class of systematic uncertainty.

- ***M*(*KK*) resolution:** The limited $M(KK)$ resolution may affect the results of the analysis, especially the phases and the $\mathcal{S}$-wave fraction F_S, through the dependence of the $\mathcal{S}-\mathcal{P}$ interference term on the $\mathcal{P}$-wave mass model. In principle, the function of Eq. (12) should be replaced by a Breit-Wigner function convoluted with a Gaussian. We avoid this complication by approximating the smeared $\mathcal{P}$-wave amplitude by a Breit-Wigner function where the width Γ_ϕ of Eq. (12) is set to twice the world average value to account for the detector resolution effects. A MC simulation-based estimate of the scale factor for the event selection criteria used in this analysis yields a value in the range 1.5 – 1.7. The resulting complex integral of the $\mathcal{S}-\mathcal{P}$ interference has an absolute value behavior closer to the data, but a distorted ratio of the real and imaginary parts compared to Eq. (12). We repeat the fits using this altered $\phi(1020)$ propagator as a measure of the sensitivity to the $M(KK)$ resolution.

Tables III and IV compare results for the default fit and the alternative fits discussed above. The differences between the best-fit values provide a measure of systematic effects. For the best estimate of the credible intervals for all the measured physics quantities, we conduct MCMC studies described in the next section.

Other sources of systematic uncertainties like the functional model of the background mass, lifetime and angle distributions were studied and give a negligible contribution.

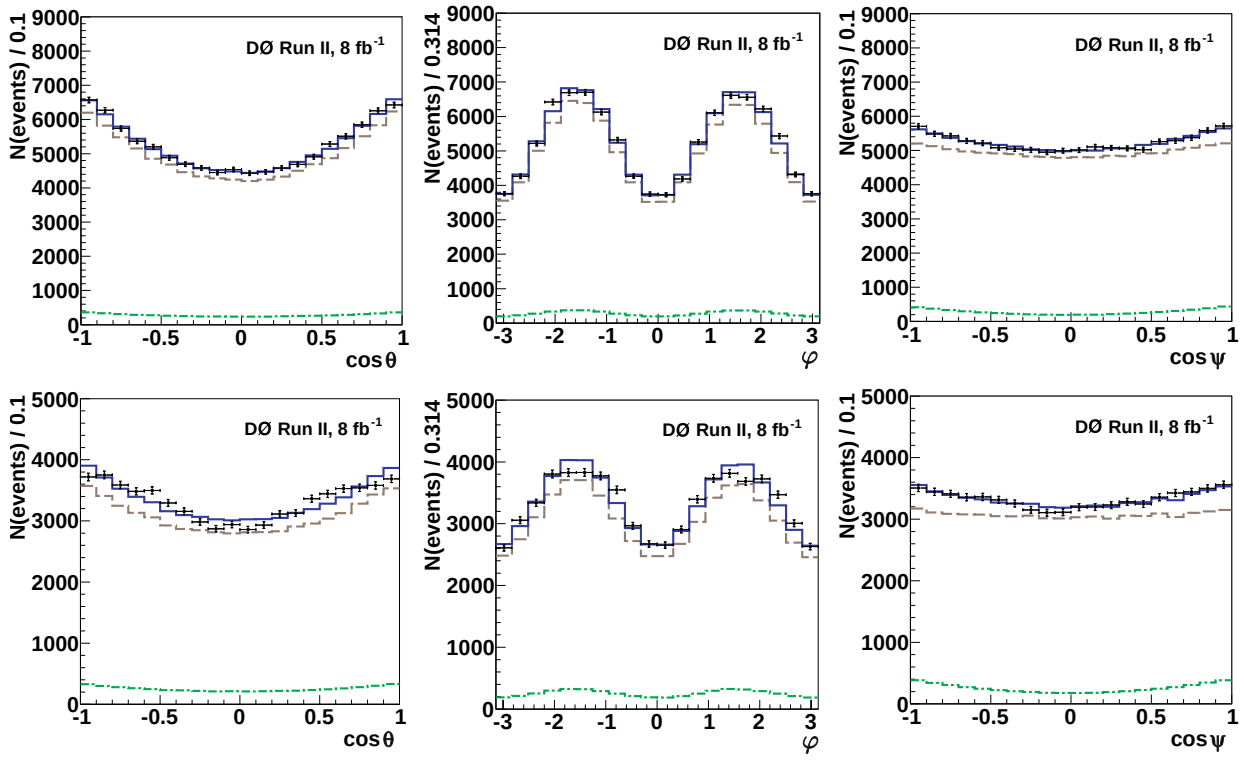


FIG. 9: (color online). Distributions of transversity polar and azimuthal angles and $\cos\psi$ for B_s^0 candidates in the BDT sample (top) and Square-cuts sample (bottom). The curves are projections of the maximum likelihood fit. Shown are the signal (green dashed-dotted curve), total background (brown long-dashed curve) and the sum of signal and total background (blue solid curve).

VII. BAYESIAN CREDIBILITY INTERVALS FROM MCMC STUDIES

The maximum likelihood fit provides the best values of all free parameters, including the signal observables and background model parameters, their statistical uncertainties and their full correlation matrix.

In addition to the free parameters determined in the fit, the model depends on a number of external constants whose inherent uncertainties are not taken into account in a given fit. Ideally, effects of uncertainties of external constants, such as time resolution parameters, flavor tagging dilution calibration, or detector acceptance, should be included in the model by introducing the appropriate parametrized probability density functions and allowing the parameters to vary. Such a procedure of maximizing the likelihood function over the external parameter space would greatly increase the number of free parameters and would be prohibitive. Therefore, as a trade-off, we apply a random sampling of external parameter values within their uncertainties, we perform the analysis for thus created "alternative universes", and we average the results. To do the averaging in the multidimensional space, taking into account non-Gaussian parameter distributions and correlations, we use the MCMC technique.

VII-A. The method

The MCMC technique uses the Metropolis-Hastings algorithm [44] to generate a sample representative to a given probability distribution. The algorithm generates a sequence of "states", a Markov chain, in which each state depends only on the previous state.

To generate a Markov chain for a given data sample, we start from the best-fit point $\vec{x}$. We randomly generate a point $\vec{x}'$ in the parameter space according to the multivariate normal distribution $\exp(-(\vec{x}'-\vec{x})\cdot\Sigma\cdot(\vec{x}'-\vec{x})/2)$, where Σ is the covariance matrix between the best fit current point $\vec{x}$ in the chain and next random point $\vec{x}'$. The best-fit point and the covariance matrix are obtained from a maximum likelihood fit over the same data sample. The new point is accepted if $\mathcal{L}(x')/\mathcal{L}(x) > 1$, otherwise it is accepted with the probability $\mathcal{L}(x')/\mathcal{L}(x)$. The process is continued until a desired number of states is achieved. To avoid a bias due to the choice of the initial state, we discard the early states which may "remember" the initial state. Our studies show that the initial state is "forgotten" after approximately 50 steps. We discard the first 100 states in each chain.

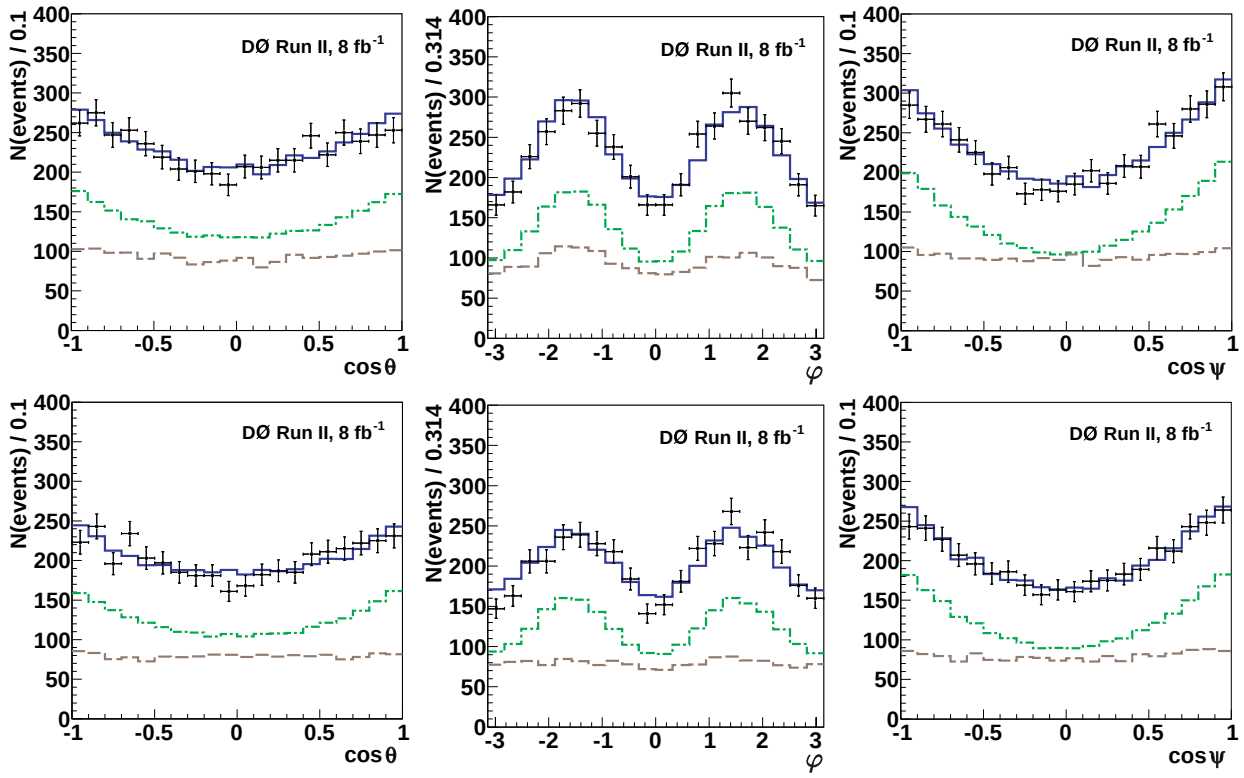


FIG. 10: (color online). Distributions of transversity polar and azimuthal angles and $\cos\psi$ for B_s^0 candidates in the BDT sample (top) and Square-cuts sample (bottom). The signal contribution is enhanced, relative to the distributions shown in Fig. 9, by additional requirements on the reconstructed mass of the B_s^0 candidates ($5.31 < M(B_s^0) < 5.43$ GeV) and on the proper time $t > 1.0$ ps. The curves are projections of the maximum likelihood fit. Shown are the signal (green dashed-dotted curve), total background (brown long-dashed curve) and the sum of signal and total background (blue solid curve).

VII-B. General properties of MCMC chains for the BDT-selection and Square-cuts samples

We generate 8 MCMC chains, each containing one million states: a nominal and three alternative chains each for the BDT-selection and Square-cuts samples, according to the fit results presented in Tables III and IV.

Figures 11 and 12 illustrate the dependence of $\phi_s^{J/\psi\phi}$ on other physics parameters, in particular on $\cos\delta_\perp$ and $\cos\delta_s$. Each point shows the Markov Chain representation of the likelihood function integrated over all parameters except the parameter of interest in a slice of $\phi_s^{J/\psi\phi}$. For clarity, the profiles are shown for $\Delta\Gamma_s > 0$ and $\Delta\Gamma_s < 0$ separately. The distributions for the Square-cuts sample are similar. We note the following salient features of these correlations for $\Delta\Gamma_s > 0$:

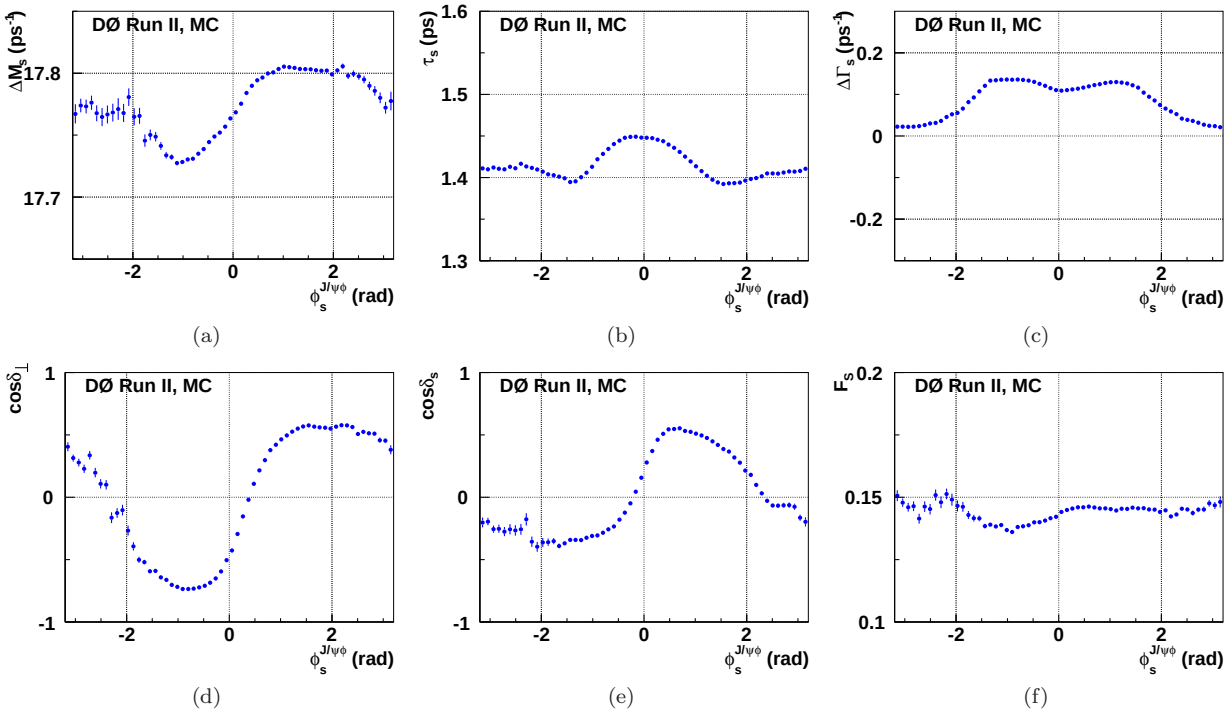


FIG. 11: Profiles of ΔM_s, $\overline{\tau}_s$, $\Delta\Gamma_s$, $\cos\delta_\perp$, $\cos\delta_s$, and F_S, for $\Delta\Gamma_s > 0$, versus $\phi_s^{J/\psi\phi}$ from the MCMC simulation for the BDT selection data sample.

a) A positive correlation between $\phi_s^{J/\psi\phi}$ and ΔM_s, with the best fit of $\phi_s^{J/\psi\phi}$ changing sign as ΔM_s increases (see also Fig. 26 in Appendix D).

b) A correlation between $|\phi_s^{J/\psi\phi}|$ and $\overline{\tau}_s$, with the highest $\overline{\tau}_s$ occuring at $\phi_s^{J/\psi\phi} = 0$.

c) For $\phi_s^{J/\psi\phi}$ near zero, $|\Delta\Gamma_s|$ increases with $|\phi_s^{J/\psi\phi}|$.

d) A strong positive correlation between $\phi_s^{J/\psi\phi}$ and $\cos\delta_\perp$ near $\phi_s^{J/\psi\phi} = 0$, with $\phi_s^{J/\psi\phi}$ changing sign as the average $\cos\delta_\perp$ increases between -0.8 and $+0.8$. For the related decay $B_d^0 \to J/\psi K^*$ the measured value is $\cos\delta_\perp = -0.97$. This indicates that a constraint of $\cos\delta_\perp$ to the $B_d^0 \to J/\psi K^*$ value would result in $\phi_s^{J/\psi\phi} < 0$ with a smaller uncertainty.

e) A strong positive correlation between $\phi_s^{J/\psi\phi}$ and $\cos\delta_s$ near $\phi_s^{J/\psi\phi} = 0$, with $\phi_s^{J/\psi\phi}$ changing sign as the average $\cos\delta_s$ increases between -0.4 and $+0.4$.

f) A weak correlation between $\phi_s^{J/\psi\phi}$ and F_S, with F_S a few percent lower for $\phi_s^{J/\psi\phi} < 0$.

While we do not use any external numerical constraints on the polarization amplitudes, we note that the best-fit values of their magnitudes and phases are consistent with those measured in the $U(3)$-flavor related decay $B_d^0 \to J/\psi K^*$ [34], up to the sign ambiguities. Ref. [46] predicts that the phases of the polarization amplitudes in the two decay processes should agree within approximately 0.17 radians. For $\delta_\perp$, our measurement gives equivalent solutions near π and near zero, with only the former being in agreement with the value of 2.91 ± 0.06 measured for $B_d^0 \to J/\psi K^*$ by B factories. Therefore, in the following we limit the range of $\delta_\perp$ to $\cos\delta_\perp < 0$.

To obtain the credible intervals for physics parameters, taking into account non-Gaussian tails and systematic effects, we combine the MCMC chains for the nominal and alternative fits. This is equivalent to an effective averaging of the resulting probability density functions from the fits. First, we combine the four MCMC chains for each sample. We then combine all eight chains, to produce the final result.

H]

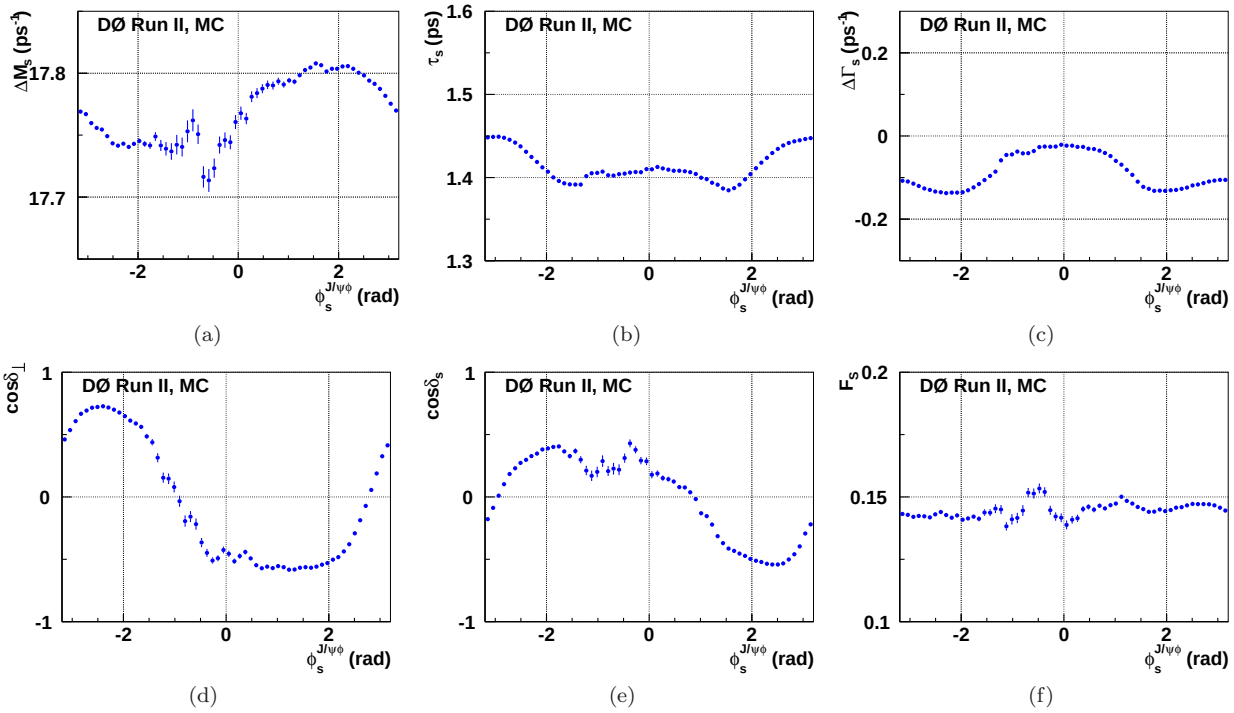


FIG. 12: Profiles of ΔM_s, $\overline{\tau}_s$, $\Delta\Gamma_s$, $\cos\delta_\perp$, $\cos\delta_s$, and F_S, for $\Delta\Gamma_s < 0$, versus $\phi_s^{J/\psi\phi}$ from the MCMC simulation for the BDT selection data sample.

VII-C. Results

Figure 13 shows 68%, 90% and 95% credible regions in the ($\phi_s^{J/\psi\phi}$,$\Delta\Gamma_s$) plane for the BDT-based and for the Square-cuts samples. The point estimates of physics parameters are obtained from one-dimensional projections. The minimal range containing 68% of the area of the probability density function defines the one standard deviation credible interval for each parameter, while the most probable value defines the central value.

The large correlation coefficient (0.85) between the two phases, $\delta_\perp$ and δ_s, prevents us from making separate point estimates. Their individual errors are much larger than the uncertainty on their difference. For the BDT selection, the measured $\mathcal{S}$-wave fraction F_S(eff) is an effective fraction of the K^+K^- $\mathcal{S}$-wave in the accepted sample, in the mass range $1.01 < M(K^+K^-) < 1.03$ GeV. It includes the effect of the diminished acceptance for the $\mathcal{S}$-wave with respect to the $\mathcal{P}$-wave in the event selection.

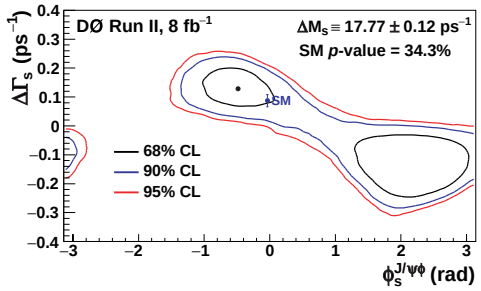
This procedure gives the following results for the BDT-based sample:

$$
\begin{aligned}
\overline{\tau}_s &= 1.426^{+0.035}_{-0.032}\ \text{ps}, \\
\Delta\Gamma_s &= 0.129^{+0.076}_{-0.053}\ \text{ps}^{-1}, \\
\phi_s^{J/\psi\phi} &= -0.49^{+0.48}_{-0.40}, \\
|A_0|^2 &= 0.552^{+0.016}_{-0.017}, \\
|A_\parallel|^2 &= 0.219^{+0.020}_{-0.021}, \\
\delta_\parallel &= 3.15 \pm 0.27, \\
\cos(\delta_\perp - \delta_s) &= -0.06 \pm 0.24, \\
F_S(\text{eff}) &= 0.146 \pm 0.035.
\end{aligned}
$$

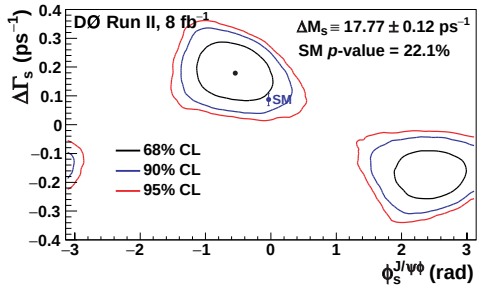
F_S(eff) in this case refers to the "effective" F_S since it is not a physical parameter: the BDT cut on the phi mass leads to the measurement of F_S in this case to depend on the efficiency of the selection to non-resonant $B_s^0 \to J/\psi K^+K^-$.

The one-dimensional estimates of physics parameters for the Square-cuts sample are:

$$
\begin{aligned}
\overline{\tau}_s &= 1.444^{+0.041}_{-0.033}\ \text{ps}, \\
\Delta\Gamma_s &= 0.179^{+0.059}_{-0.060}\ \text{ps}^{-1}, \\
\phi_s^{J/\psi\phi} &= -0.56^{+0.36}_{-0.32}, \\
|A_0|^2 &= 0.565 \pm 0.017, \\
|A_\parallel|^2 &= 0.249^{+0.021}_{-0.022}, \\
\delta_\parallel &= 3.15 \pm 0.19, \\
\cos(\delta_\perp - \delta_s) &= -0.20^{+0.26}_{-0.27}, \\
F_S &= 0.173 \pm 0.036.
\end{aligned}
$$



(a)BDT selection



(b)Square-cuts selection

FIG. 13: (color online). Two-dimensional 68%, 90% and and 95% credible regions for (a) the BDT selection and (b) the Square-cuts sample. The standard model expectation is indicated as a point with an error.

To obtain the final credible intervals for physics parameters, we combine all eight MCMC chains, effectively averaging the probability density functions of the results of the fits to the BDT- and Square-cuts samples. Figure 14 shows 68%, 90% and 95% credible regions in the $(\phi_s^{J/\psi\phi},\Delta\Gamma_s)$ plane. The p-value for the SM point [47] $(\phi_s^{J/\psi\phi},\Delta\Gamma_s) = (-0.038, 0.087\ \text{ps}^{-1})$ is 29.8%. The one-dimensional 68% credible intervals are listed in Section VIII below.

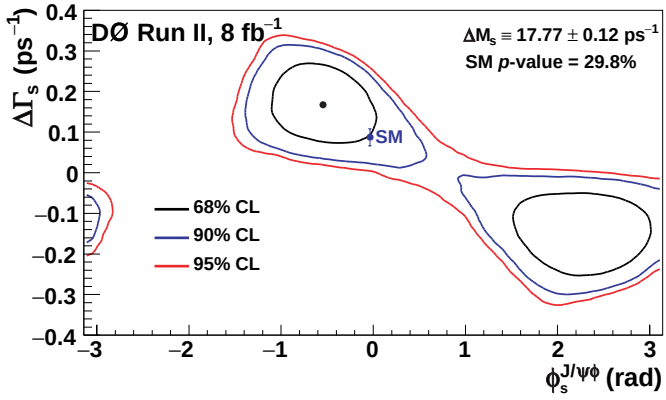


FIG. 14: (color online). Two-dimensional 68%, 90% and 95% credible regions including systematic uncertainties. The standard model expectation is indicated as a point with an error.

VIII. SUMMARY AND DISCUSSION

We have presented a time-dependent angular analysis of the decay process $B_s^0 \to J/\psi\phi$. We measure B_s^0 mixing parameters, average lifetime, and decay amplitudes. In addition, we measure the amplitudes and phases of the polarization amplitudes. We also measure the level of the KK $\mathcal{S}$-wave contamination in the mass range (1.01 – 1.03) GeV, F_S. The measured values and the 68% credible intervals, including systematic uncertainties, with the oscillation frequency constrained to $\Delta M_s = 17.77 \pm 0.12$ ps^{-1}, are:

$$
\begin{aligned}
\overline{\tau}_s &= 1.443^{+0.038}_{-0.035}\ \text{ps}, \\
\Delta\Gamma_s &= 0.163^{+0.065}_{-0.064}\ \text{ps}^{-1}, \\
\phi_s^{J/\psi\phi} &= -0.55^{+0.38}_{-0.36}, \\
|A_0|^2 &= 0.558^{+0.017}_{-0.019}, \\
|A_\parallel|^2 &= 0.231^{+0.024}_{-0.030}, \\
\delta_\parallel &= 3.15 \pm 0.22, \\
\cos(\delta_\perp - \delta_s) &= -0.11^{+0.27}_{-0.25}. \\
F_S &= 0.173 \pm 0.036,
\end{aligned}
\tag{13}
$$

The p-value for the SM point $(\phi_s^{J/\psi\phi},\Delta\Gamma_s) = (-0.038, 0.087\ \text{ps}^{-1})$ is 29.8%.

In the previous publication [26], which was based on a subset of this data sample, we constrained the strong phases to those of $B_d^0 \to J/\psi K^*$ whereas this analysis has a large enough data sample to reliably let them float. Also, the previous publication did not have a large enough data sample to allow for the measurement of a significant level of KK $\mathcal{S}$-wave, whereas it is measured together with its relative phase in the current analysis. The results supersede our previous measurements.

Independently of the Maximum Likelihood analysis, we make an estimate of the non-resonant K^+K^- in the final state based on the $M(KK)$ distribution of the B_s^0 signal yield. The result of this study (Appendix C) is consistent with the result of the Maximum Likelihood fit shown above.

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC and the Royal Society (United Kingdom); MSMT and GACR (Czech Republic); CRC Program and NSERC (Canada); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); and CAS and CNSF (China). We thank J. Boudreau who has suggested and developed the use of the MCMC method for this study.

[1] M. Bona *et al.*, J. High Energy Phys. **10**, 081 (2006).
[2] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
[3] M. Kreps *et al.*, arXiv:1103.4962 [hep-ph].
[4] J. Drobnak *et al.*, Phys. Lett. B **701**, 234 (2011).
[5] R. M. Wang *et al.*, Phys. Rev. D **83**, 0950109 (2011).
[6] A. K. Alok *et al.*, arXiv:1103.5344 [hep-ph].
[7] J. Shelton and K. M.Zurek, Phys. Rev. D **83**, 091701 (2011).
[8] Z. J. Ajaltouni and E. Di Salvo, Jour. Phys. G, **37**, 125001 (2010).
[9] S. Nandi and A. Soni, Phys. Rev. D **83**, 114510 (2011).
[10] A. Datta *et al.*, Phys. Rev. D **83**, 094501 (2011).
[11] J. Girrbach *et al.*, J. High Energy Phys. **06**, 044 (2011).
[12] A. J. Buras *et al.*, J. High Energy Phys. **05**, 005 (2011).
[13] Z. Ligeti *et al.*, Phys. Rev. Lett. **105**, 131601 (2010).
[14] A. J. Buras *et al.*, Phys. Lett. B **694**, 402, (2010).
[15] B. A. Dobrescu, P. J. Fox and A. Martin, Phys. Rev. Lett. **105**, 041801 (2010).
[16] J. P. Saha, B. Misra and A. Kundu, Phys. Rev. D **81**, 095011 (2010).
[17] A. J. Buras *et al.*, J. High Energy Phys. **10**, 106 (2010).
[18] Y. Bai and A. E. Nelson, Phys. Rev. D **82**, 114027 (2010).
[19] G. Isidori, Y. Nir, G. Perez, arXiv:1002.0900 [hep-ph].
[20] A. Soni *et al.*, Phys. Rev. D **82**, 033009 (2010).
[21] L L. Everett *et al.*, Phys. Rev. D **82**, 094924 (2010).
[22] F. J. Botella, G. C. Branco and M. N.Rebelo, Phys. Lett. B **687**, 2 (2010).
[23] C. W. Chiang *et al.*, J. High Energy Phys. **04**, 031 (2010).
[24] D0 Collaboration, V. M. Abazov *et al.*, Phys. Rev. Lett. **98**, 121801 (2007).
[25] CDF Collaboration, T. Aaltonen *et al.*, Phys. Rev. Lett. **100**, 121803 (2008).
[26] D0 Collaboration, V. M. Abazov *et al.*, Phys. Rev. Lett. **101**, 241801 (2008).
[27] CDF Collaboration, T. Aaltonen *et al.*, Phys. Rev. Lett. **100**, 161802 (2008).
[28] CDF Collaboration, T. Aaltonen *et. al.*, arXiv:1112.1726 [hep-ex], submitted to Phys. Rev. D.
[29] LHCb Collaboration, R. Aaji *et. al.*, arXiv:1112.3183 [hep-ex], submitted to Phys. Rev. Lett.
[30] D0 Collaboration, V. M. Abazov *et al.*, Nucl. Instrum. Methods Phys. Res. A **565**, 463 (2006).
[31] S. Stone and L. Zhang, Phys. Rev. D **79**, 074024 (2009).
[32] R. Angstadt *et al.* Nucl. Instrum. Methods Phys. Res. A **622**, 298 (2010).
[33] V. M. Abazov *et al.*, Nucl. Instrum. Methods Phys. Res. A **522**, 372 (2005).
[34] K. Nakamura *et al.* (Particle Data Group), J. Phys. G **37**, 075021 (2010).
[35] H. U. Bengtsson and T. Sjöstrand, J. High Energy Phys. **05**, 026 (2006).
[36] D.J. Lange, Nucl. Instrum. Meth. A **462**, 152 (2001).
[37] R. Brun and F. Carminati, CERN Program Library Long Writeup No. W5013, 1993 (unpublished).
[38] http://root.cern.ch.
[39] D0 Collaboration, V. M. Abazov *et al.*, Phys. Rev. D **74**, 112002 (2006).
[40] D0 Collaboration, V. M. Abazov *et al.* Phys. Rev. Lett. **97**, 021802 (2006).
[41] F. Azfar *et al.*, J. High Energy Phys. **11**, 158 (2010).
[42] W. Verkerke and D. Kirkby, "The RooFit Toolkit for Data Modeling", http://roofit.sourceforge.net/.
[43] A.S. Dighe, I. Dunietz, and R. Fleischer, Eur. Phys. J. C**6**, 647 (1999).
[44] W.K. Hastings, "Monte Carlo Sampling Methods Using Markov Chains and Their Applications", Biometrika **57**(1), 97 (1970).
[45] CDF Collaboration, A. Abulencia *et al.*, Phys. Rev. Lett. **97**, 242003 (2006).
[46] M. Gronau, J. L. Rosner, Phys. Lett. B **669**, 321 (2008).
[47] A. Lenz, and U. Nierste, arXiv:1102.4274 [hep-ph].
[48] H.G. Moser and A. Roussarie, Nucl. Instrum. Methods Phys. Res. Sect. A **384**, 491 (1997).

Appendix A: BDT Discriminants

Two BDT discriminants are used to reject background. One is trained to remove the prompt background (the "prompt BDT"), and the other is trained to remove inclusive B decays (the "inclusive BDT"). The prompt BDT uses 33 variables, listed in Table V. The inclusive BDT uses 35 variables, listed in Table VI. In these tables, ΔR is defined as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, where η is the pseudorapidity and ϕ is the azimuthal angle. The term "uncorrected" refers to the correction due to the J/ψ mass constraint. "Leading" ("trailing") muon or kaon refers to the particle with larger (smaller) p_T, and dE/dx is the energy loss per unit path length of a charged particle as it traverses the silicon detector. Isolation is defined as $p(B)/\sum_{<\Delta R} p$ where $p(B)$ is the sum of the momenta of the four daughter particles of the B_s^0 candidate, and the sum is over all particles within a cone defined by ΔR, including the decay products of the B_s^0 candidate. The

tables also show the importance and separation for each variable. The separation $\langle S^2 \rangle$ of a classifier y is defined as

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy, \qquad \text{(A1)}$$

where y_S is the output of the discriminant function for signal events and y_B is the discriminant function for background. The importance of each BDT input variable is derived by counting in the training how often the variable is used to split decision tree nodes and by weighting each split occurrence by its separation gain squared and by the number of events in the node.

The distributions for the six most important variables in training on prompt J/ψ decays are shown in Fig. 15. The distributions for the six most important variables in the training on inclusive $B \to J/\psi X$ decays are shown in Fig. 16.

Figure 17 compares the shapes of the distributions of the three angular variables and the lifetime, before and after the BDT requirements. The figures show that the BDT requirements do not affect these differential distributions significantly.

Rank	Variable	Importance	Separation
1	KK invariant mass	0.3655	0.3540
2	Maximum ΔR between either K meson and the B_s^0 candidate	0.1346	0.4863
3	Isolation using the maximum ΔR between either K and the B_s^0	0.0390	0.1784
4	Uncorrected p_T of the B_s^0	0.0346	0.3626
5	Minimum ΔR between either K and the B_s^0	0.0335	0.4278
6	p_T of the trailing K meson	0.0331	0.4854
7	p_T of the ϕ meson	0.0314	0.4998
8	p_T of the leading K meson	0.0283	0.4884
9	Trailing muon momentum	0.0252	0.0809
10	p_T of the leading muon	0.0240	0.1601
11	Maximum ΔR between either muon and the B_s^0	0.0223	0.1109
12	Maximum χ^2 of either K meson with the J/ψ vertex	0.0217	0.0162
13	Dimuon invariant mass	0.0215	0.0145
14	Maximum χ^2 of either of the K candidate track	0.0213	0.021
15	B_s^0 isolation using the larger K/B_s ΔR and tracks from the PV	0.0207	0.1739
16	p_T of the J/ψ meson	0.0205	0.1809
17	Minimum ΔR between either muon and the B_s^0 candidate	0.0188	0.1023
18	Trailing K momentum	0.0105	0.3159
19	χ^2 of the B_s^0 candidate vertex	0.0093	0.0119
20	B_s^0 isolation using $\Delta R < 0.75$	0.0084	0.0241
21	Minimum χ^2 of the J/ψ vertex with either K	0.0081	0.0069
22	p_T of the trailing muon	0.0079	0.0922
23	Minimum of the χ^2 of the J/ψ and ϕ vertices	0.0073	0.0057
24	Isolation using $\Delta R < 0.5$	0.0070	0.0405
25	Uncorrected B_s^0 total momentum	0.0068	0.2103
26	Minimum χ^2 of either K track fit	0.0065	0.0266
27	Isolation using $\Delta R < 0.5$ and particles from the PV	0.0057	0.0401
28	Leading K meson momentum	0.0051	0.3217
29	Leading muon momentum	0.0048	0.0908
30	ϕ meson momentum	0.0048	0.3233
31	Maximum χ^2 of the J/ψ or ϕ vertices	0.0044	0.0061
32	Isolation using $\Delta R < 0.75$ and particles from the PV	0.0037	0.0259
33	J/ψ meson momentum	0.0037	0.1004

TABLE V: Variables used to train the prompt BDT, ranked by their importance in the training.

Rank	Variable	Importance	Separation
1	KK invariant mass	0.2863	0.3603
2	B_s^0 isolation using the larger K/B_s ΔR and tracks from the PV	0.1742	0.4511
3	Minimum dE/dx of either K	0.0778	0.1076
4	χ^2 of B_s^0	0.0757	0.2123
5	p_T of the ϕ meson	0.0559	0.4856
6	p_T of the leading K meson	0.0504	0.4745
7	Isolation using the maximum ΔR between either K and the B_s^0	0.0429	0.4468
8	p_T of the trailing K meson	0.0350	0.4774
9	Maximum χ^2 of either K meson with the J/ψ vertex	0.0260	0.2051
10	Isolation using $\Delta R < 0.5$ and particles from the PV	0.0229	0.1703
11	Isolation using $\Delta R < 0.75$ and tracks from the PV	0.0154	0.2238
12	Minimum χ^2 of of either K with the J/ψ vertex	0.0151	0.1308
13	Minimum ΔR between either K meson and the B_s^0 candidate	0.0115	0.3104
14	Dimuon invariant mass	0.0099	0.0190
15	Total momentum of the ϕ meson	0.0091	0.3307
16	p_T of the J/ψ meson	0.0089	0.1198
17	Trailing muon momentum	0.0082	0.0594
18	Isolation using $\Delta R < 0.5$	0.0073	0.1695
19	Maximum ΔR between either K meson and the B_s^0 candidate	0.0070	0.3794
20	Maximum dE/dx of either K meson	0.0069	0.0528
21	Trailing K meson momentum	0.0068	0.3253
22	J/ψ vertex χ^2	0.0063	0.0057
23	Leading K meson momentum	0.0058	0.3277
24	Maximum χ^2 of either K candidate track	0.0054	0.0267
25	Isolation using $\Delta R < 0.75$	0.0046	0.2203
26	Minimum ΔR between either muon and the B_s^0 candidate	0.0041	0.0729
27	Minimum χ^2 of either K candidate track	0.0039	0.0284
28	uncorrected p_T of B_s^0 candidate	0.0036	0.2485
29	p_T of the trailing muon	0.0029	0.0702
30	J/ψ momentum	0.0027	0.0645
31	Maximum ΔR between either muon and the B_s^0 candidate	0.0026	0.0872
32	Vertex χ^2 of the ϕ meson	0.0017	0.0098
33	Uncorrected B_s^0 momentum	0.0014	0.1675
34	p_T of the leading muon	0.0011	0.1008
35	Leading muon momentum	0.0009	0.0547

TABLE VI: Variables used to train the non-prompt BDT, ranked by their importance in the training.

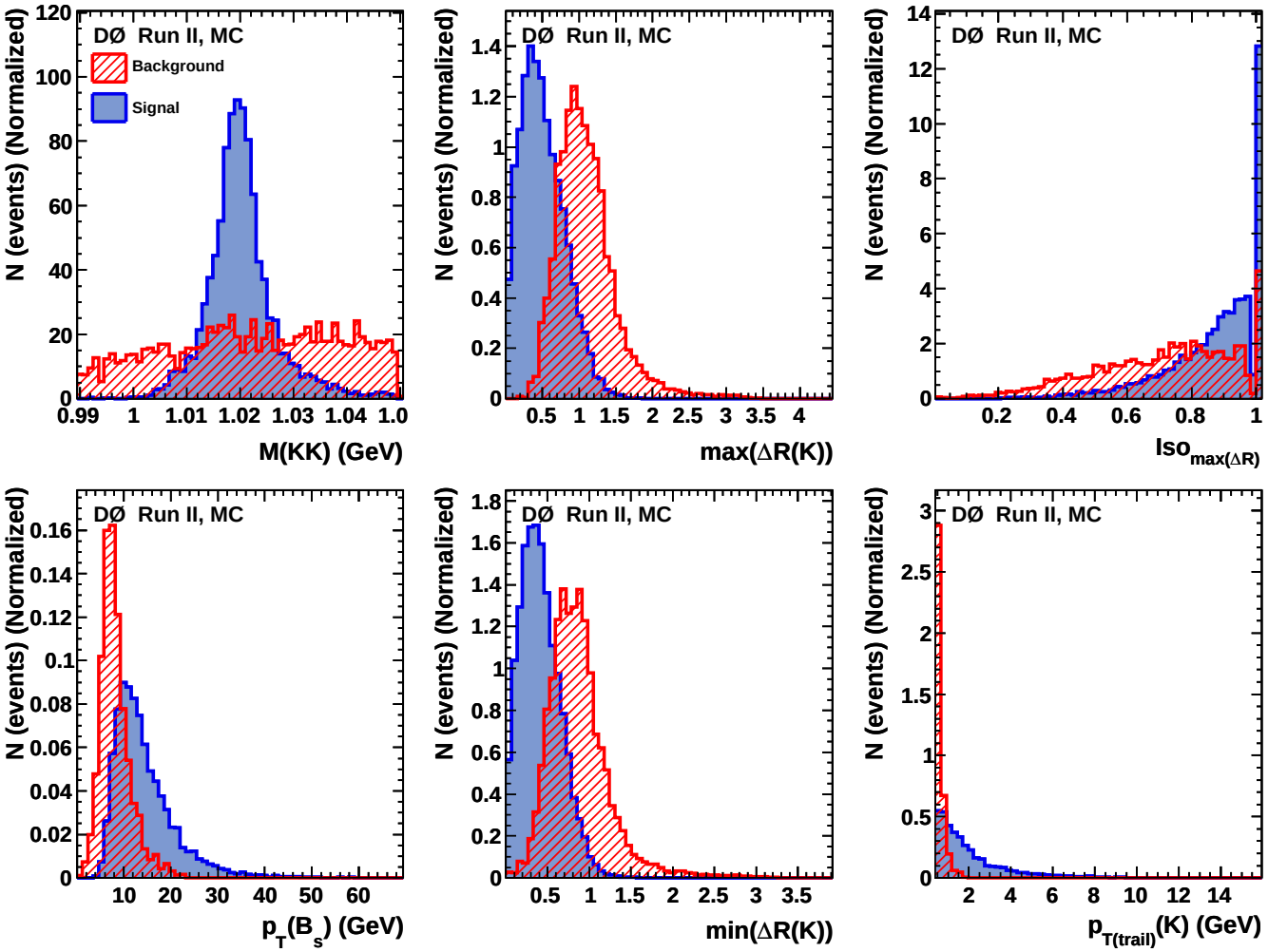


FIG. 15: (color online) The distributions of the six most important variables used in the BDT trained on prompt J/ψ production for the $B_s^0 \to J/\psi\phi$ signal (solid blue) and prompt J/ψ events (red dashed) histograms.

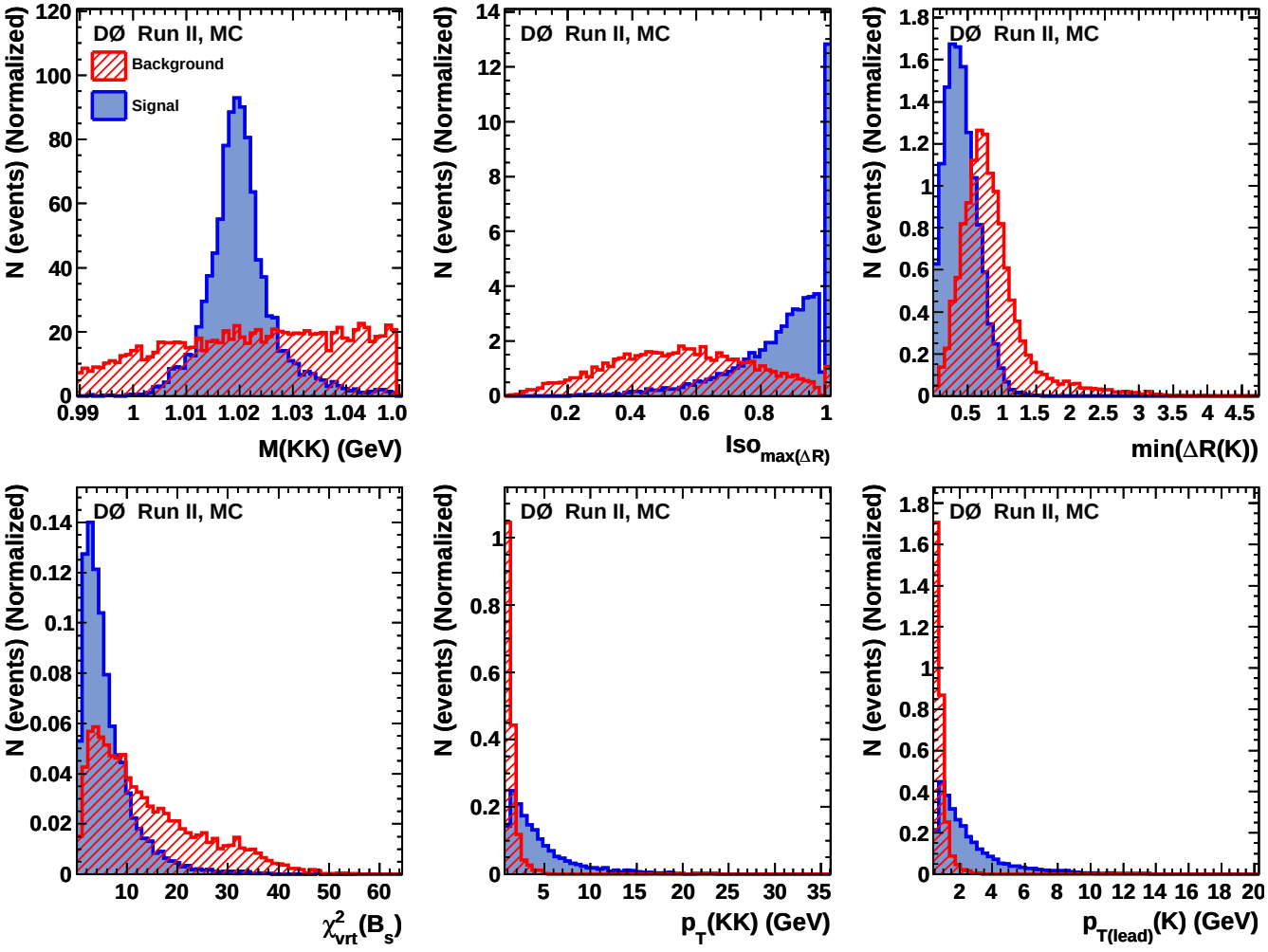


FIG. 16: (color online) The distributions of the six most important variables used in the BDT trained on inclusive $B \to J/\psi X$ decays for the $B_s^0 \to J/\psi\phi$ signal (solid blue) and inclusive $B \to J/\psi X$ decays (red dashed) histograms.

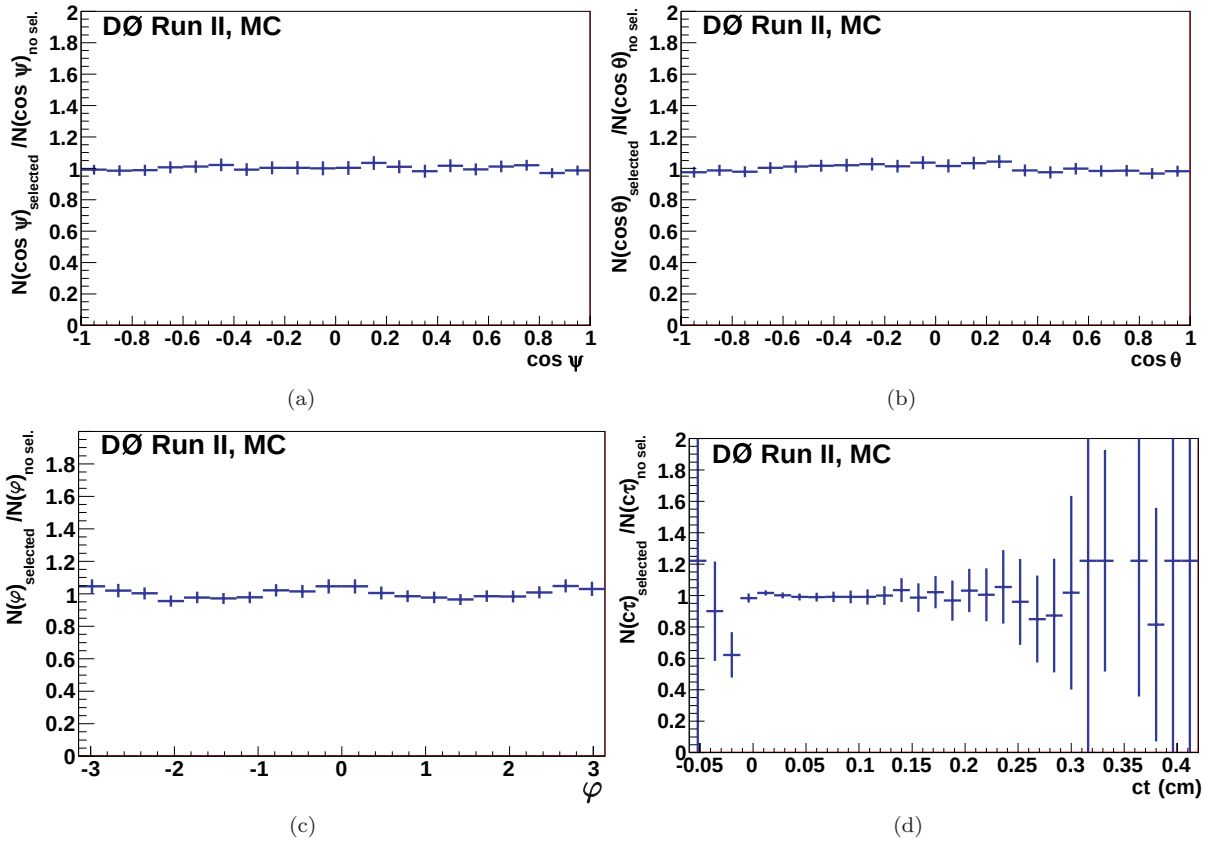
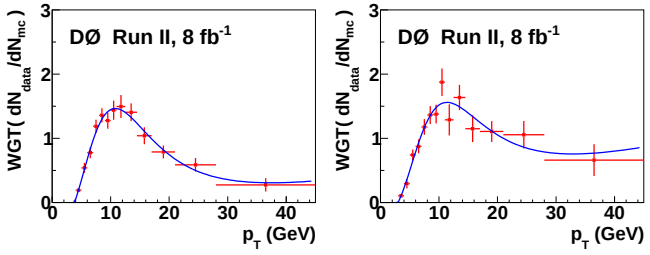


FIG. 17: Test of uniformity of the efficiencies of the BDT selection using a MC sample with $\phi_s = -0.5$. The figure shows the ratios of the normalized distributions of (a – c) the three angles and (d) the proper decay length, before and after the BDT selection.

Appendix B: Detector acceptance

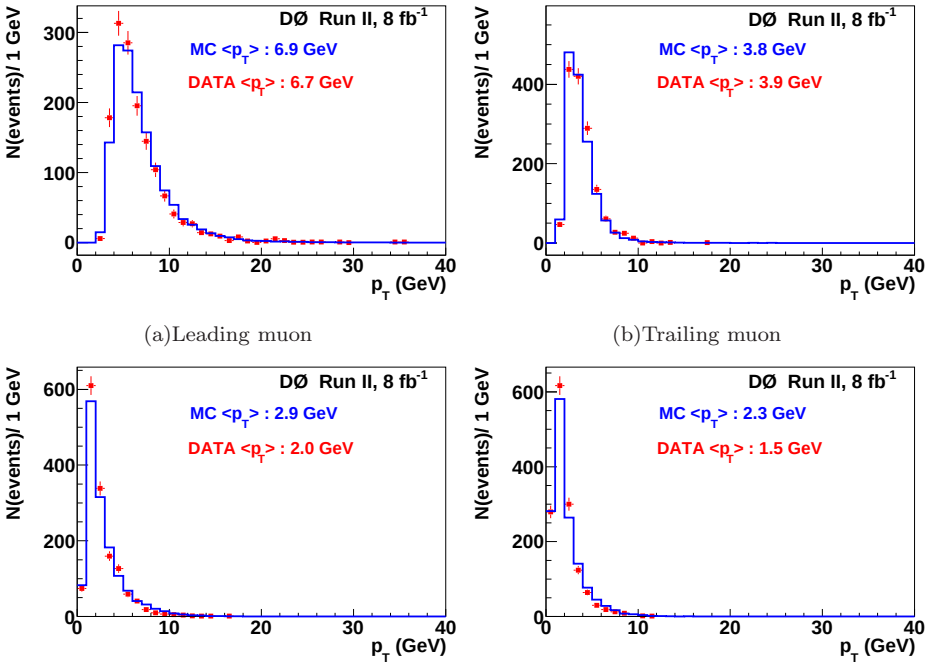
We take into account the shaping of the signal distribution by the detector acceptance and kinematic selection by introducing acceptance functions in the three angles of the transversity basis . The acceptance functions are derived from Monte Carlo simulation. Due to the event triggering effects, the momentum spectra of final-state objects in data are harder than in MC. We take into account the difference in the p_T distribution of the final-state objects in data and MC by introducing a weight factor as a function of $p_T(J/\psi)$, separately for the central ($|\eta(\mu_{\rm leading})| < 1$) and forward regions. The weight factor is derived by forcing an agreement between the J/ψ transverse momentum spectra in data and MC. The behavior of the weight factor as a function of $p_T(J/\psi)$ for the BDT-based selection, for the central and forward regions, is shown in Fig. 18.

Figure 19 shows the background-subtracted p_T distributions of the leading and trailing muon and leading and trailing kaon, in the central region. There is a good agreement between data and MC for all final-state particles after applying the weight factor. The acceptance in φ and θ is shown in Fig. 20. The acceptance in ψ is shown in Fig. 21.



(a) (b)

FIG. 18: Weight factor as a function of $p_T(J/\psi)$ used to correct MC p_T distribution of B_s^0 and B_d^0 decay objects for (a) central region, and (b) forward region. The curves are empirical fits to a sum of a Landau function and a polynomial.



(a)Leading muon (b)Trailing muon

(c)Leading kaon (d)Trailing kaon

FIG. 19: Transverse momentum distributions of the four final-state particles in data (points) and weighted MC (solid histogram), for the BDT-based event selection.

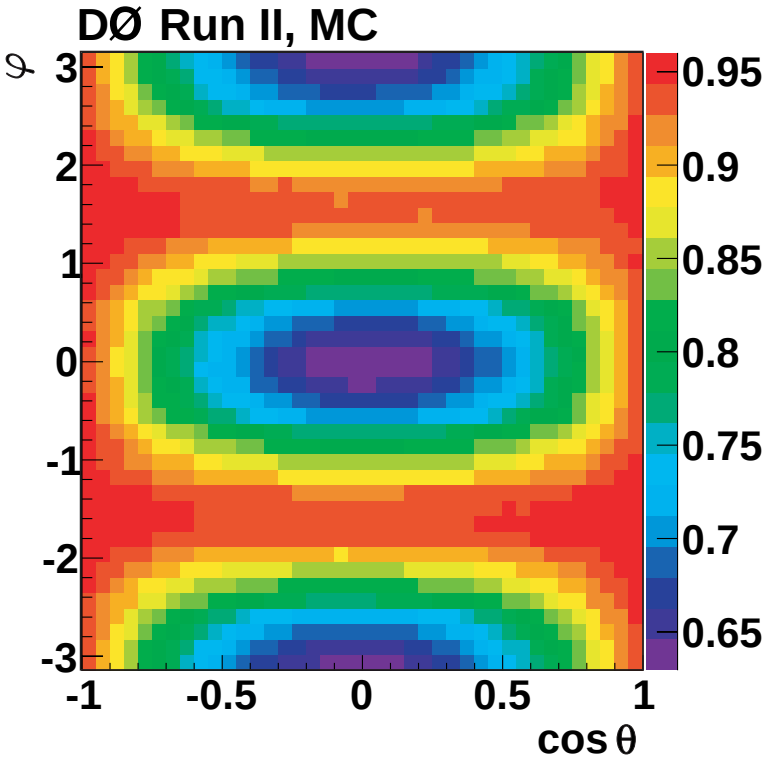


FIG. 20: (color online). Map of the detector acceptance on the plane φ – $\cos\theta$.

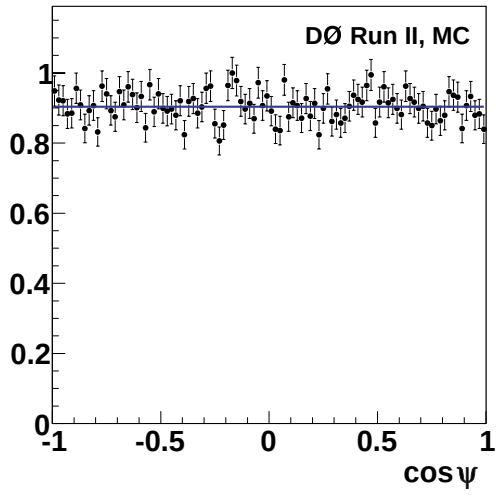


FIG. 21: Detector acceptance as a function of $\cos\psi$. The acceptance is uniform in $\cos\psi$.

Appendix C: Independent estimate of F_S

In the Maximum Likelihood fit, the invariant mass of the K^+K^- pair is not used. To do so would require a good model of the $M(K^+K^-)$ dependence of background, including a small $\phi(1020)$ component, as a function of the B_s^0 candidate mass and proper time. However, we can use the $M(K^+K^-)$ mass information to make an independent estimate of the non-resonant K^+K^- contribution in the final state.

For this study, we use the "Square-cuts" sample, for

which the event selection is not biased in $M(K^+K^-)$. Using events with decay length $ct > 0.02$ cm to suppress background, we extract the B_s^0 signal in two ranges of $M(K^+K^-)$: $1.01 < M(KK) < 1.03$ GeV and $1.03 < M(KK) < 1.05$ GeV. The first range is that used by both selections, and contains the bulk of the $\phi \to K^+K^-$ signal. The second range will still contain a small Breit-Wigner tail of $\phi \to K^+K^-$. From the simulated $M(K^+K^-)$ distribution of the $B_s^0 \to J/\psi\phi$ decay, shown in Fig. 22, we obtain the fraction of the K^+K^- decay products in the upper mass range to be 0.061 ± 0.001 of the total range $1.01 < M(KK) < 1.05$ GeV. The $\mathcal{S}$-wave component is assumed to be a flat distribution in $M(KK)$ across this range. Given that the widths of the ranges are the same, the number of candidates due to the $\mathcal{S}$-wave contribution should be the same for both.

The B_s^0 signal in each mass range is extracted by fitting the B_s^0 candidate mass distribution to a Gaussian function representing the signal, a linear function for the background, and MC simulation-based templates for the $B^0 \to J/\psi K^*$ reflection where the pion from the K^* decay is assumed to be a kaon. The two shape templates used, one for each mass range, are shown in Fig. 23. The mass distributions, with fits using the above templates, are shown in Fig. 24. The fits result in the B_s^0 yield of 3027 ± 93 events for $1.01 < M(KK) < 1.03$ GeV and 547 ± 94 events for $1.03 < M(KK) < 1.05$ GeV. In the mass range $1.01 < M(KK) < 1.03$ GeV, we extract the fraction of B_s^0 candidates decaying into non-resonant KK to be 0.12 ± 0.03. The error includes the uncertainties in the signal and background modelling. This excess may be due to an $\mathcal{S}$-wave, or a non-resonant $\mathcal{P}$-wave, or a combination of both. If we assign it entirely to the $\mathcal{S}$-wave, and assume it to be independent of $M(KK)$, we obtain the measured $\mathcal{S}$-wave fraction in the range $1.01 < M(K^+K^-) < 1.03$ GeV to be $F_S = 0.12 \pm 0.03$.

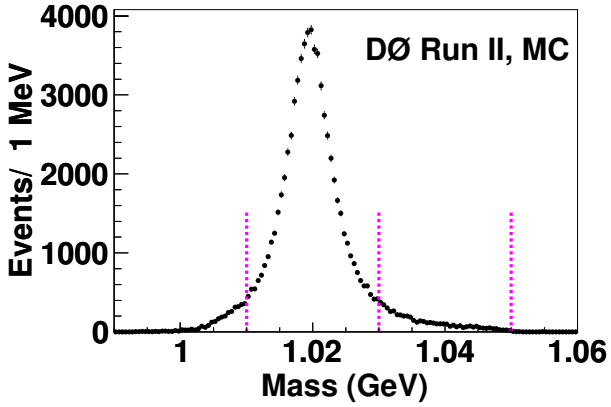


FIG. 22: Invariant mass distribution of kaon pairs from the full simulation of the decay $\phi \to K^+K^-$. Vertical dashed lines delineate the two $M(KK)$ invariant mass bins considered.

Appendix D: $B_s^0 - \overline{B}_s^0$ oscillation

Under the hypothesis of CP conservation in the B_s^0 decay, and a possible mixing-induced CP violation, the non-vanishing CP-violating mixing angle should manifest itself as a $B_s^0 - \overline{B}_s^0$ oscillation with the amplitude proportional to $\sin(\phi_s^{J/\psi\phi})$. The observed time-dependent asymmetry $\Delta N \equiv N(B_s^0) - N(\overline{B}_s^0) = N_S \cdot C \cdot \sin(\phi_s^{J/\psi\phi})$, is diluted by a product C of several factors: (i) a factor of $(1-2|A_\perp|^2)\cdot(1-2F_s) \approx 0.6 \cdot 0.7$ due to the presence of the CP-odd decay, (ii) a factor of $\epsilon \cdot \mathcal{D}^2 \approx 0.03$ due to the flavor tagging efficiency and accuracy, and (iii) a factor of $\exp(-(\Delta M_s \sigma)^2/2) \approx 0.2$ due to the limited time resolution. Thus, with $N_S \approx 6000$ events, and $C \approx 0.0025$, we expect $N_S \cdot C \approx 15$.

In Fig. 25 we show the proper decay length evolution of ΔN in the first 90 μm, corresponding to approximately twice the mean B_s^0 lifetime. The curve represents a fit to the function $N_0 \cdot \sin(\Delta M_s t) \cdot \exp(-t/\tau_s)$, with N_0 unconstrained and with $\Delta M_s \equiv 17.77$ ps^{-1}. The fit gives $N_0 = -6$ for the BDT-based sample and -8 for the Square-cuts sample, with a statistical uncertainty of ± 4, corresponding to $\sin(\phi_s^{J/\psi\phi}) = N_0/N_S \cdot C \approx -0.4 \pm 0.3$. This one-dimensional analysis gives a result for $\phi_s^{J/\psi\phi}$ that is consistent with the result of the full analysis.

Following the Amplitude Method described in Ref. [48], we fit the above distributions at discrete values of ΔM_s, and plot the fitted value of N_0 as a function of the probe frequency. The results are shown in Fig. 26. There is an undulating structure, with no significantly large deviations from zero. At ΔM_s near 17.77 ps^{-1} the data prefer a negative oscillation amplitude (and hence a

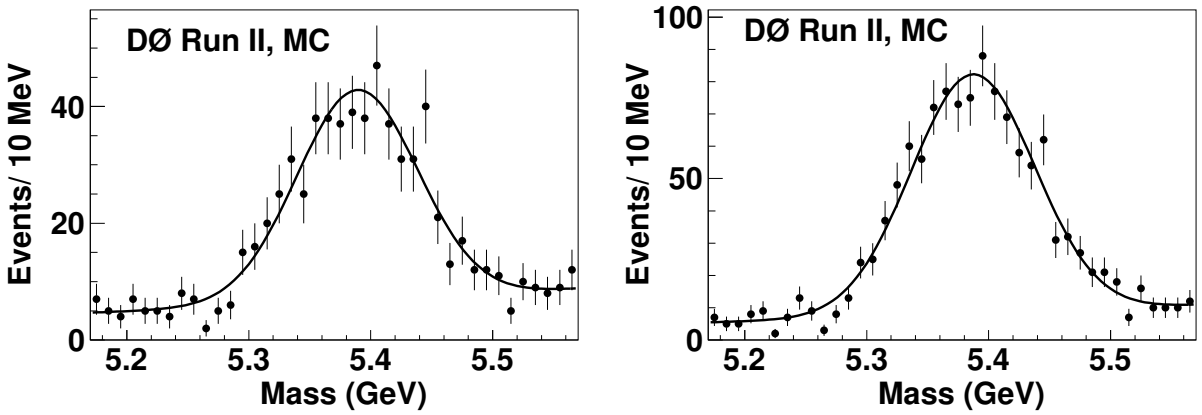


FIG. 23: The simulated distributions of the invariant mass of the $B_d^0 \to J/\psi K^*$ decay products reconstructed under the $B_s^0 \to J/\psi\phi$ hypothesis for $1.01 < M(KK) < 1.03$ GeV (left) and $1.03 < M(KK) < 1.05$ GeV (right). The curves are results of fits assuming a sum of two Gaussian functions.

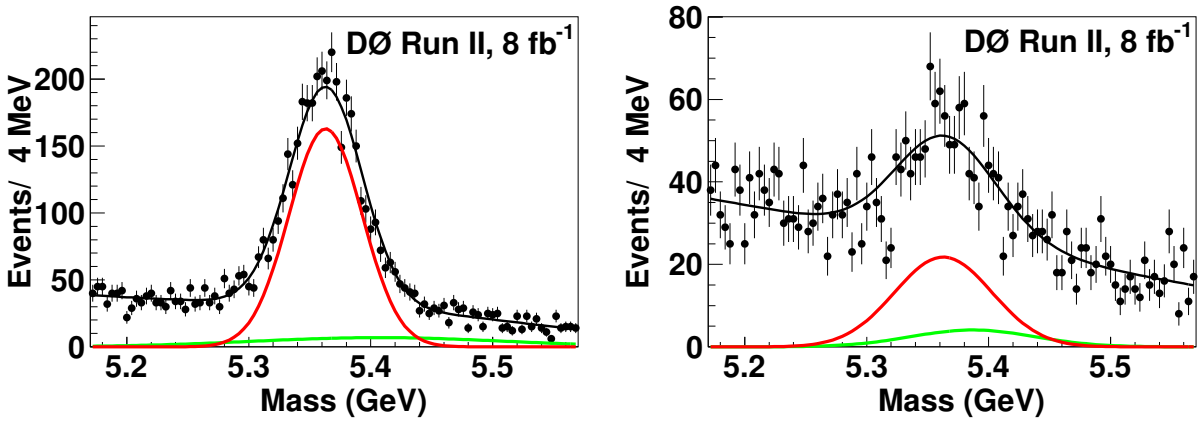


FIG. 24: (color online). Invariant mass distributions of B_s^0 candidates with decay length $ct > 0.02$ cm for $1.01 < M(KK) < 1.03$ GeV (left) and $1.03 < M(KK) < 1.05$ GeV (right). Fits to a sum (black line) of a Gaussian function representing the signal (red), an MC simulation-based template for the $B^0 \to J/\psi K^*$ reflection (green line), and a linear function representing the background are used to extract the B_s^0 yield.

negative value of $\sin\phi_s^{J/\psi\phi}$). The statistical uncertainty of the result of this simple approach does not take into account uncertainties of the dilution factors, related to the time resolution, CP-odd fraction, and the $\mathcal{S}$-wave fraction.

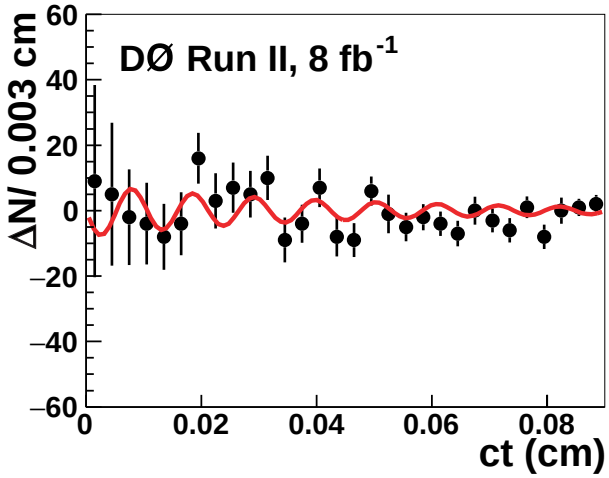
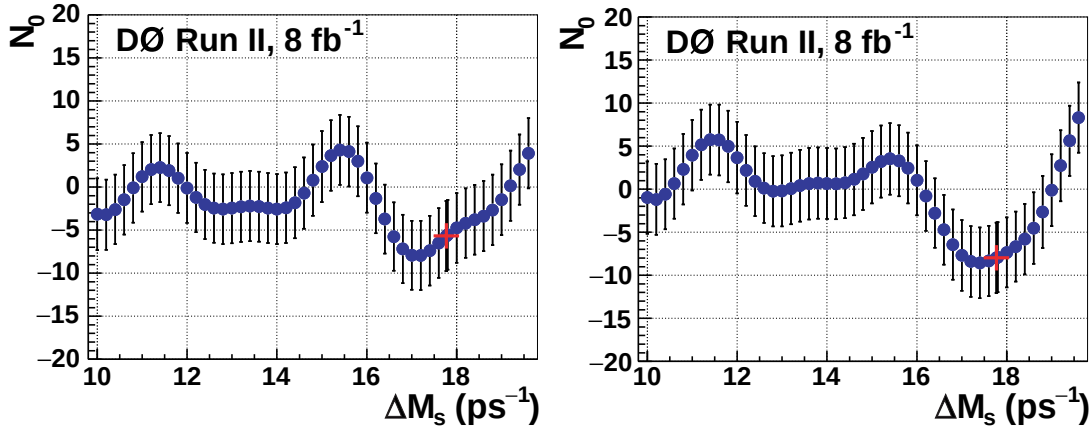


FIG. 25: Proper decay length evolution of the difference $\Delta N = N(B_s^0) - N(\overline{B}_s^0)$ in the first 0.09 cm (3 ps) for the Square-cuts sample. The curve represents the best fit to the oscillation with the frequency of $\Delta M_s = 17.77$ ps^{-1}.



(a) (b)

FIG. 26: (color online). The fitted magnitude of the $B_s^0 - \overline{B}_s^0$ oscillation as a function of ΔM_s for (a) BDT selection and (b) Square cuts. The red crosses correspond to $\Delta M_s = 17.77$ ps^{-1}.