Measurement of the B^- lifetime using a simulation free approach for trigger bias correction

T. Aaltonen,[24] J. Adelman,[14] B. Álvarez González[w],[12] S. Amerio[ee],[44] D. Amidei,[35] A. Anastassov,[39] A. Annovi,[20] J. Antos,[15] G. Apollinari,[18] J. Appel,[18] A. Apresyan,[49] T. Arisawa,[58] A. Artikov,[16] J. Asaadi,[54] W. Ashmanskas,[18] A. Attal,[4] A. Aurisano,[54] F. Azfar,[43] W. Badgett,[18] A. Barbaro-Galtieri,[29] V.E. Barnes,[49] B.A. Barnett,[26] P. Barria[gg],[47] P. Bartos,[15] G. Bauer,[33] P.-H. Beauchemin,[34] F. Bedeschi,[47] D. Beecher,[31] S. Behari,[26] G. Bellettini[ff],[47] J. Bellinger,[60] D. Benjamin,[17] A. Beretvas,[18] A. Bhatti,[51] M. Binkley*,[18] D. Bisello[ee],[44] I. Bizjak[kk],[31] R.E. Blair,[2] C. Blocker,[7] B. Blumenfeld,[26] A. Bocci,[17] A. Bodek,[50] V. Boisvert,[50] D. Bortoletto,[49] J. Boudreau,[48] A. Boveia,[11] B. Brau[a],[11] A. Bridgeman,[25] L. Brigliadori[dd],[6] C. Bromberg,[36] E. Brubaker,[14] J. Budagov,[16] H.S. Budd,[50] S. Budd,[25] K. Burkett,[18] G. Busetto[ee],[44] P. Bussey,[22] A. Buzatu,[34] K. L. Byrum,[2] S. Cabrera[y],[17] C. Calancha,[32] S. Camarda,[4] M. Campanelli,[31] M. Campbell,[35] F. Canelli[14],[18] A. Canepa,[46] B. Carls,[25] D. Carlsmith,[60] R. Carosi,[47] S. Carrillo[n],[19] S. Carron,[18] B. Casal,[12] M. Casarsa,[18] A. Castro[dd],[6] P. Catastini[gg],[47] D. Cauz,[55] V. Cavaliere[gg],[47] M. Cavalli-Sforza,[4] A. Cerri,[29] L. Cerrito[q],[31] S.H. Chang,[28] Y.C. Chen,[1] M. Chertok,[8] G. Chiarelli,[47] G. Chlachidze,[18] F. Chlebana,[18] K. Cho,[28] D. Chokheli,[16] J.P. Chou,[23] K. Chung[o],[18] W.H. Chung,[60] Y.S. Chung,[50] T. Chwalek,[27] C.I. Ciobanu,[45] M.A. Ciocci[gg],[47] A. Clark,[21] D. Clark,[7] G. Compostella,[44] M.E. Convery,[18] J. Conway,[8] M.Corbo,[45] M. Cordelli,[20] C.A. Cox,[8] D.J. Cox,[8] F. Crescioli[ff],[47] C. Cuenca Almenar,[61] J. Cuevas[w],[12] R. Culbertson,[18] J.C. Cully,[35] D. Dagenhart,[18] N. d'Ascenzo[v],[45] M. Datta,[18] T. Davies,[22] P. de Barbaro,[50] S. De Cecco,[52] A. Deisher,[29] G. De Lorenzo,[4] M. Dell'Orso[ff],[47] C. Deluca,[4] L. Demortier,[51] J. Deng[f],[17] M. Deninno,[6] M. d'Errico[ee],[44] A. Di Canto[ff],[47] B. Di Ruzza,[47] J.R. Dittmann,[5] M. D'Onofrio,[4] S. Donati[ff],[47] P. Dong,[18] T. Dorigo,[44] S. Dube,[53] K. Ebina,[58] A. Elagin,[54] R. Erbacher,[8] D. Errede,[25] S. Errede,[25] N. Ershaidat[cc],[45] R. Eusebi,[54] H.C. Fang,[29] S. Farrington,[43] W.T. Fedorko,[14] R.G. Feild,[61] M. Feindt,[27] J.P. Fernandez,[32] C. Ferrazza[hh],[47] R. Field,[19] G. Flanagan[s],[49] R. Forrest,[8] M.J. Frank,[5] M. Franklin,[23] J.C. Freeman,[18] I. Furic,[19] M. Gallinaro,[51] J. Galyardt,[13] F. Garberson,[11] J.E. Garcia,[21] A.F. Garfinkel,[49] P. Garosi[gg],[47] H. Gerberich,[25] D. Gerdes,[35] A. Gessler,[27] S. Giagu[ii],[52] V. Giakoumopoulou,[3] P. Giannetti,[47] K. Gibson,[48] J.L. Gimmell,[50] C.M. Ginsburg,[18] N. Giokaris,[3] M. Giordani[jj],[55] P. Giromini,[20] M. Giunta,[47] G. Giurgiu,[26] V. Glagolev,[16] D. Glenzinski,[18] M. Gold,[38] N. Goldschmidt,[19] A. Golossanov,[18] G. Gomez,[12] G. Gomez-Ceballos,[33] M. Goncharov,[33] O. González,[32] I. Gorelov,[38] A.T. Goshaw,[17] K. Goulianos,[51] A. Gresele[ee],[44] S. Grinstein,[4] C. Grosso-Pilcher,[14] R.C. Group,[18] U. Grundler,[25] J. Guimaraes da Costa,[23] Z. Gunay-Unalan,[36] C. Haber,[29] S.R. Hahn,[18] E. Halkiadakis,[53] B.-Y. Han,[50] J.Y. Han,[50] F. Happacher,[20] K. Hara,[56] D. Hare,[53] M. Hare,[57] R.F. Harr,[59] M. Hartz,[48] K. Hatakeyama,[5] C. Hays,[43] M. Heck,[27] J. Heinrich,[46] M. Herndon,[60] J. Heuser,[27] S. Hewamanage,[5] D. Hidas,[53] C.S. Hill[c],[11] D. Hirschbuehl,[27] A. Hocker,[18] S. Hou,[1] M. Houlden,[30] S.-C. Hsu,[29] R.E. Hughes,[40] B.T. Huffman,[43] M. Hurwitz,[14] U. Husemann,[61] M. Hussein,[36] J. Huston,[36] J. Incandela,[11] G. Introzzi,[47] M. Iori[ii],[52] A. Ivanov[p],[8] E. James,[18] D. Jang,[13] B. Jayatilaka,[17] E.J. Jeon,[28] M.K. Jha,[6] S. Jindariani,[18] W. Johnson,[8] M. Jones,[49] K.K. Joo,[28] S.Y. Jun,[13] J.E. Jung,[28] T.R. Junk,[18] T. Kamon,[54] D. Kar,[19] P.E. Karchin,[59] Y. Kato[m],[42] R. Kephart,[18] W. Ketchum,[14] J. Keung,[46] V. Khotilovich,[54] B. Kilminster,[18] D.H. Kim,[28] H.S. Kim,[28] H.W. Kim,[28] J.E. Kim,[28] M.J. Kim,[20] S.B. Kim,[28] S.H. Kim,[56] Y.K. Kim,[14] N. Kimura,[58] L. Kirsch,[7] S. Klimenko,[19] K. Kondo,[58] D.J. Kong,[28] J. Konigsberg,[19] A. Korytov,[19] A.V. Kotwal,[17] M. Kreps,[27] J. Kroll,[46] D. Krop,[14] N. Krumnack,[5] M. Kruse,[17] V. Krutelyov,[11] T. Kuhr,[27] N.P. Kulkarni,[59] M. Kurata,[56] S. Kwang,[14] A.T. Laasanen,[49] S. Lami,[47] S. Lammel,[18] M. Lancaster,[31] R.L. Lander,[8] K. Lannon[u],[40] A. Lath,[53] G. Latino[gg],[47] I. Lazzizzera[ee],[44] T. LeCompte,[2] E. Lee,[54] H.S. Lee,[14] J.S. Lee,[28] S.W. Lee[x],[54] S. Leone,[47] J.D. Lewis,[18] C.-J. Lin,[29] J. Linacre,[43] M. Lindgren,[18] E. Lipeles,[46] A. Lister,[21] D.O. Litvintsev,[18] C. Liu,[48] T. Liu,[18] N.S. Lockyer,[46] A. Loginov,[61] L. Lovas,[15] D. Lucchesi[ee],[44] J. Lueck,[27] P. Lujan,[29] P. Lukens,[18] G. Lungu,[51] L. Lyons,[43] J. Lys,[29] R. Lysak,[15] D. MacQueen,[34] R. Madrak,[18] K. Maeshima,[18] K. Makhoul,[33] P. Maksimovic,[26] S. Malde,[43] S. Malik,[31] G. Manca[e],[30] A. Manousakis-Katsikakis,[3] F. Margaroli,[49] C. Marino,[27] C.P. Marino,[25] A. Martin,[61]

V. Martin[k],[22] M. Martínez,[4] R. Martínez-Ballarín,[32] P. Mastrandrea,[52] M. Mathis,[26] M.E. Mattson,[59] P. Mazzanti,[6] K.S. McFarland,[50] P. McIntyre,[54] R. McNulty[j],[30] A. Mehta,[30] P. Mehtala,[24] A. Menzione,[47] C. Mesropian,[51] T. Miao,[18] D. Mietlicki,[35] N. Miladinovic,[7] R. Miller,[36] C. Mills,[23] M. Milnik,[27] A. Mitra,[1] G. Mitselmakher,[19] H. Miyake,[56] S. Moed,[23] N. Moggi,[6] M.N. Mondragon[n],[18] C.S. Moon,[28] R. Moore,[18] M.J. Morello,[47] J. Morlock,[27] P. Movilla Fernandez,[18] J. Mülmenstädt,[29] A. Mukherjee,[18] Th. Muller,[27] P. Murat,[18] M. Mussini[dd],[6] J. Nachtman[o],[18] Y. Nagai,[56] J. Naganoma,[56] K. Nakamura,[56] I. Nakano,[41] A. Napier,[57] J. Nett,[60] C. Neu[aa],[46] M.S. Neubauer,[25] S. Neubauer,[27] J. Nielsen[g],[29] L. Nodulman,[2] M. Norman,[10] O. Norniella,[25] E. Nurse,[31] L. Oakes,[43] S.H. Oh,[17] Y.D. Oh,[28] I. Oksuzian,[19] T. Okusawa,[42] R. Orava,[24] K. Osterberg,[24] S. Pagan Griso[ee],[44] C. Pagliarone,[55] E. Palencia,[18] V. Papadimitriou,[18] A. Papaikonomou,[27] A.A. Paramanov,[2] B. Parks,[40] S. Pashapour,[34] J. Patrick,[18] G. Pauletta[jj],[55] M. Paulini,[13] C. Paus,[33] T. Peiffer,[27] D.E. Pellett,[8] A. Penzo,[55] T.J. Phillips,[17] G. Piacentino,[47] E. Pianori,[46] L. Pinera,[19] K. Pitts,[25] C. Plager,[9] L. Pondrom,[60] K. Potamianos,[49] O. Poukhov*,[16] N.L. Pounder,[43] F. Prokoshin[z],[16] A. Pronko,[18] F. Ptohos[i],[18] E. Pueschel,[13] G. Punzi[ff],[47] J. Pursley,[60] J. Rademacker[c],[43] A. Rahaman,[48] V. Ramakrishnan,[60] N. Ranjan,[49] I. Redondo,[32] P. Renton,[43] M. Renz,[27] M. Rescigno,[52] S. Richter,[27] F. Rimondi[dd],[6] L. Ristori,[47] A. Robson,[22] T. Rodrigo,[12] T. Rodriguez,[46] E. Rogers,[25] S. Rolli,[57] R. Roser,[18] M. Rossi,[55] R. Rossin,[11] P. Roy,[34] A. Ruiz,[12] J. Russ,[13] V. Rusu,[18] B. Rutherford,[18] H. Saarikko,[24] A. Safonov,[54] W.K. Sakumoto,[50] L. Santi[jj],[55] L. Sartori,[47] K. Sato,[56] V. Saveliev[v],[45] A. Savoy-Navarro,[45] P. Schlabach,[18] A. Schmidt,[27] E.E. Schmidt,[18] M.A. Schmidt,[14] M.P. Schmidt*,[61] M. Schmitt,[39] T. Schwarz,[8] L. Scodellaro,[12] A. Scribano[gg],[47] F. Scuri,[47] A. Sedov,[49] S. Seidel,[38] Y. Seiya,[42] A. Semenov,[16] L. Sexton-Kennedy,[18] F. Sforza[ff],[47] A. Sfyrla,[25] S.Z. Shalhout,[59] T. Shears,[30] P.F. Shepard,[48] M. Shimojima[t],[56] S. Shiraishi,[14] M. Shochet,[14] Y. Shon,[60] I. Shreyber,[37] A. Simonenko,[16] P. Sinervo,[34] A. Sisakyan,[16] A.J. Slaughter,[18] J. Slaunwhite,[40] K. Sliwa,[57] J.R. Smith,[8] F.D. Snider,[18] R. Snihur,[34] A. Soha,[18] S. Somalwar,[53] V. Sorin,[4] P. Squillacioti[gg],[47] M. Stanitzki,[61] R. St. Denis,[22] B. Stelzer,[34] O. Stelzer-Chilton,[34] D. Stentz,[39] J. Strologas,[38] G.L. Strycker,[35] J.S. Suh,[28] A. Sukhanov,[19] I. Suslov,[16] A. Taffard[f],[25] R. Takashima,[41] Y. Takeuchi,[56] R. Tanaka,[41] J. Tang,[14] M. Tecchio,[35] P.K. Teng,[1] J. Thom[h],[18] J. Thome,[13] G.A. Thompson,[25] E. Thomson,[46] P. Tipton,[61] P. Ttito-Guzmán,[32] S. Tkaczyk,[18] D. Toback,[54] S. Tokar,[15] K. Tollefson,[36] T. Tomura,[56] D. Tonelli,[18] S. Torre,[20] D. Torretta,[18] P. Totaro[jj],[55] M. Trovato[hh],[47] S.-Y. Tsai,[1] Y. Tu,[46] N. Turini[gg],[47] F. Ukegawa,[56] S. Uozumi,[28] N. van Remortel[b],[24] A. Varganov,[35] E. Vataga[hh],[47] F. Vázquez[n],[19] G. Velev,[18] C. Vellidis,[3] M. Vidal,[32] I. Vila,[12] R. Vilar,[12] M. Vogel,[38] I. Volobouev[x],[29] G. Volpi[ff],[47] P. Wagner,[46] R.G. Wagner,[2] R.L. Wagner,[18] W. Wagner[bb],[27] J. Wagner-Kuhr,[27] T. Wakisaka,[42] R. Wallny,[9] S.M. Wang,[1] A. Warburton,[34] D. Waters,[31] M. Weinberger,[54] J. Weinelt,[27] W.C. Wester III,[18] B. Whitehouse,[57] D. Whiteson[f],[46] A.B. Wicklund,[2] E. Wicklund,[18] S. Wilbur,[14] G. Williams,[34] H.H. Williams,[46] P. Wilson,[18] B.L. Winer,[40] P. Wittich[h],[18] S. Wolbers,[18] C. Wolfe,[14] H. Wolfe,[40] T. Wright,[35] X. Wu,[21] F. Würthwein,[10] A. Yagil,[10] K. Yamamoto,[42] J. Yamaoka,[17] U.K. Yang[r],[14] Y.C. Yang,[28] W.M. Yao,[29] G.P. Yeh,[18] K. Yi[o],[18] J. Yoh,[18] K. Yorita,[58] T. Yoshida[l],[42] G.B. Yu,[17] I. Yu,[28] S.S. Yu,[18] J.C. Yun,[18] A. Zanetti,[55] Y. Zeng,[17] X. Zhang,[25] Y. Zheng[d],[9] and S. Zucchelli[dd6]

(CDF Collaboration[†])

[1]*Institute of Physics, Academia Sinica, Taipei, Taiwan 11529, Republic of China*
[2]*Argonne National Laboratory, Argonne, Illinois 60439, USA*
[3]*University of Athens, 157 71 Athens, Greece*
[4]*Institut de Fisica d'Altes Energies, Universitat Autonoma de Barcelona, E-08193, Bellaterra (Barcelona), Spain*
[5]*Baylor University, Waco, Texas 76798, USA*
[6]*Istituto Nazionale di Fisica Nucleare Bologna,*
[dd]*University of Bologna, I-40127 Bologna, Italy*
[7]*Brandeis University, Waltham, Massachusetts 02254, USA*
[8]*University of California, Davis, Davis, California 95616, USA*
[9]*University of California, Los Angeles, Los Angeles, California 90024, USA*
[10]*University of California, San Diego, La Jolla, California 92093, USA*
[11]*University of California, Santa Barbara, Santa Barbara, California 93106, USA*
[12]*Instituto de Fisica de Cantabria, CSIC-University of Cantabria, 39005 Santander, Spain*

[13]Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA
[14]Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA
[15]Comenius University, 842 48 Bratislava,
Slovakia; Institute of Experimental Physics, 040 01 Kosice, Slovakia
[16]Joint Institute for Nuclear Research, RU-141980 Dubna, Russia
[17]Duke University, Durham, North Carolina 27708, USA
[18]Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA
[19]University of Florida, Gainesville, Florida 32611, USA
[20]Laboratori Nazionali di Frascati, Istituto Nazionale di Fisica Nucleare, I-00044 Frascati, Italy
[21]University of Geneva, CH-1211 Geneva 4, Switzerland
[22]Glasgow University, Glasgow G12 8QQ, United Kingdom
[23]Harvard University, Cambridge, Massachusetts 02138, USA
[24]Division of High Energy Physics, Department of Physics,
University of Helsinki and Helsinki Institute of Physics, FIN-00014, Helsinki, Finland
[25]University of Illinois, Urbana, Illinois 61801, USA
[26]The Johns Hopkins University, Baltimore, Maryland 21218, USA
[27]Institut für Experimentelle Kernphysik, Karlsruhe Institute of Technology, D-76131 Karlsruhe, Germany
[28]Center for High Energy Physics: Kyungpook National University,
Daegu 702-701, Korea; Seoul National University, Seoul 151-742,
Korea; Sungkyunkwan University, Suwon 440-746,
Korea; Korea Institute of Science and Technology Information,
Daejeon 305-806, Korea; Chonnam National University, Gwangju 500-757,
Korea; Chonbuk National University, Jeonju 561-756, Korea
[29]Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA
[30]University of Liverpool, Liverpool L69 7ZE, United Kingdom
[31]University College London, London WC1E 6BT, United Kingdom
[32]Centro de Investigaciones Energeticas Medioambientales y Tecnologicas, E-28040 Madrid, Spain
[33]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
[34]Institute of Particle Physics: McGill University, Montréal,
Québec, Canada H3A 2T8; Simon Fraser University, Burnaby,
British Columbia, Canada V5A 1S6; University of Toronto,
Toronto, Ontario, Canada M5S 1A7; and TRIUMF,
Vancouver, British Columbia, Canada V6T 2A3
[35]University of Michigan, Ann Arbor, Michigan 48109, USA
[36]Michigan State University, East Lansing, Michigan 48824, USA
[37]Institution for Theoretical and Experimental Physics, ITEP, Moscow 117259, Russia
[38]University of New Mexico, Albuquerque, New Mexico 87131, USA
[39]Northwestern University, Evanston, Illinois 60208, USA
[40]The Ohio State University, Columbus, Ohio 43210, USA
[41]Okayama University, Okayama 700-8530, Japan
[42]Osaka City University, Osaka 588, Japan
[43]University of Oxford, Oxford OX1 3RH, United Kingdom
[44]Istituto Nazionale di Fisica Nucleare, Sezione di Padova-Trento,
[ee]University of Padova, I-35131 Padova, Italy
[45]LPNHE, Universite Pierre et Marie Curie/IN2P3-CNRS, UMR7585, Paris, F-75252 France
[46]University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
[47]Istituto Nazionale di Fisica Nucleare Pisa, [ff]University of Pisa,
[gg]University of Siena and [hh]Scuola Normale Superiore, I-56127 Pisa, Italy
[48]University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA
[49]Purdue University, West Lafayette, Indiana 47907, USA
[50]University of Rochester, Rochester, New York 14627, USA
[51]The Rockefeller University, New York, New York 10021, USA
[52]Istituto Nazionale di Fisica Nucleare, Sezione di Roma 1,
[ii]Sapienza Università di Roma, I-00185 Roma, Italy
[53]Rutgers University, Piscataway, New Jersey 08855, USA
[54]Texas A&M University, College Station, Texas 77843, USA
[55]Istituto Nazionale di Fisica Nucleare Trieste/Udine,
I-34100 Trieste, [jj]University of Trieste/Udine, I-33100 Udine, Italy
[56]University of Tsukuba, Tsukuba, Ibaraki 305, Japan
[57]Tufts University, Medford, Massachusetts 02155, USA
[58]Waseda University, Tokyo 169, Japan
[59]Wayne State University, Detroit, Michigan 48201, USA
[60]University of Wisconsin, Madison, Wisconsin 53706, USA
[61]Yale University, New Haven, Connecticut 06520, USA

(Dated: April 27, 2010)

The collection of a large number of B hadron decays to hadronic final states at the CDF II detector is possible due to the presence of a trigger that selects events based on track impact parameters. However, the nature of the selection requirements of the trigger introduces a large bias in the observed proper decay time distribution. A lifetime measurement must correct for this bias and the conventional approach has been to use a Monte Carlo simulation. The leading sources of systematic uncertainty in the conventional approach are due to differences between the data and the Monte Carlo simulation. In this paper we present an analytic method for bias correction without using simulation, thereby removing any uncertainty between data and simulation. This method is presented in the form of a measurement of the lifetime of the B^- using the mode $B^- \to D^0\pi^-$. The B^- lifetime is measured as $\tau_{B^-} = 1.663 \pm 0.023 \pm 0.015$ ps, where the first uncertainty is statistical and the second systematic. This new method results in a smaller systematic uncertainty in comparison to methods that use simulation to correct for the trigger bias.

PACS numbers: 14.40Nd;13.25.Hw;29.85.Fj

I. INTRODUCTION

The weak decay of quarks depends on fundamental parameters of the standard model, including the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which describes mixing between quark families [1, 2]. Extraction of these parameters from weak decays is complicated because the quarks are confined within color-singlet hadrons as described by quantum chromodynamics (QCD). An essential tool used in this extraction is the heavy quark expansion (HQE) technique [3]. In HQE the total decay width of a heavy hadron is expressed as an expansion in inverse powers of the heavy quark mass m_q. At $\mathcal{O}(1/m_b)$ the lifetimes of all B hadrons are identical. Corrections to this simplification are given by $\mathcal{O}(1/m_b^2)$ and $\mathcal{O}(1/m_b^3)$ calculations leading to the predicted lifetime hierarchy: $\tau(B^\pm) > \tau(B^0) \approx \tau(B_s^0) > \tau(\Lambda_b) >> \tau(B_c)$ and quantitative predictions of the lifetime ratios with respect to the B^0 meson [4–9].

The Tevatron $p\overline{p}$ collider at $\sqrt{s} = 1.96$ TeV has the energy to produce all B hadron species. The decays of these hadrons are selected by a variety of successive trigger selection criteria applied at the three trigger levels. Unique to the CDF II detector is the silicon vertex trigger (SVT), which selects events based on pairs of tracks displaced from the primary interaction point. This exploits the long lived nature of B hadrons and results in yields of B hadrons in several decay modes, targeting in particular the fully hadronic B decays. Many different measurements of the properties of B hadrons have been made using samples selected by this trigger, examples of which are given in Ref. [10–14].

However, this trigger preferentially selects those events in which the decay time of the B hadron is longer. This leads to a biased proper decay time distribution. The conventional approach to correct this bias has been through the use of a full detector and trigger simulation. An important source of systematic uncertainty, inherent in this conventional approach, is how well the simulation represents the data. A full and accurate simulation of data collected by this trigger is particularly difficult due to the dependence on many variables including particle kinematics, beam-interaction positions, and the instantaneous luminosity. The difference between data and simulation is the dominant systematic uncertainty in the recent CDF measurement of the Λ_b lifetime [15]. These systematic uncertainties

*Deceased
†With visitors from [a]University of Massachusetts Amherst, Amherst, Massachusetts 01003, [b]Universiteit Antwerpen, B-2610 Antwerp, Belgium, [c]University of Bristol, Bristol BS8 1TL, United Kingdom, [d]Chinese Academy of Sciences, Beijing 100864, China, [e]Istituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09042 Monserrato (Cagliari), Italy, [f]University of California Irvine, Irvine, CA 92697, [g]University of California Santa Cruz, Santa Cruz, CA 95064, [h]Cornell University, Ithaca, NY 14853, [i]University of Cyprus, Nicosia CY-1678, Cyprus, [j]University College Dublin, Dublin 4, Ireland, [k]University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom, [l]University of Fukui, Fukui City, Fukui Prefecture, Japan 910-0017, [m]Kinki University, Higashi-Osaka City, Japan 577-8502, [n]Universidad Iberoamericana, Mexico D.F., Mexico, [o]University of Iowa, Iowa City, IA 52242, [p]Kansas State University, Manhattan, KS 66506, [q]Queen Mary, University of London, London, E1 4NS, England, [r]University of Manchester, Manchester M13 9PL, England, [s]Muons, Inc., Batavia, IL 60510, [t]Nagasaki Institute of Applied Science, Nagasaki, Japan, [u]University of Notre Dame, Notre Dame, IN 46556, [v]Obninsk State University, Obninsk, Russia, [w]University de Oviedo, E-33007 Oviedo, Spain, [x]Texas Tech University, Lubbock, TX 79609, [y]IFIC(CSIC-Universitat de Valencia), 56071 Valencia, Spain, [z]Universidad Tecnica Federico Santa Maria, 110v Valparaiso, Chile, [aa]University of Virginia, Charlottesville, VA 22906, [bb]Bergische Universität Wuppertal, 42097 Wuppertal, Germany, [cc]Yarmouk University, Irbid 211-63, Jordan, [kk]On leave from J. Stefan Institute, Ljubljana, Slovenia,

will be the limiting factor in obtaining precision measurements of b hadron lifetimes in data samples collected by such triggers. In this paper we present a new analytical technique for correction of the bias induced by this trigger. This technique uses no information from simulations of the detector or physics processes, and thus incurs none of the uncertainties intrinsic to the simulation based method.

The technique is presented in a measurement of the B^- meson lifetime using the decay mode $B^- \to D^0\pi^-$ (charge conjugate decays are implied throughout). This decay channel is chosen as the lifetime of the B^- is already well known and the high yield available in this channel allows a good comparison to the world average. This measurement demonstrates the ability of this method to reduce the overall systematic uncertainty on a lifetime measurement. A displaced track trigger is expected to operate at the LHCb detector, and the technique of lifetime measurement presented here is applicable to any data where the method of collection induces a bias in the proper decay time distribution.

II. OVERVIEW

The simulation-independent method for removing the trigger-induced lifetime bias, presented here, is based on using an candidate-by-candidate efficiency function for each B meson candidate. This efficiency function is calculated from the event data, without recourse to simulation. This approach is based on the observation that for a given set of decay kinematics of the decay $B^- \to D^0\pi^-$ (four momenta of the final state particles and the flight distance of the D) the decay time dependent efficiency function has a simple shape that can easily be calculated from the measured decay kinematics and the known decay time dependent cuts. This provides a simple and robust method of taking into account the effect of the trigger by calculating a different efficiency function for each candidate and applying it candidate-by-candidate in the likelihood fit. The details of this calculation are presented in Sec. V.

As discussed in Ref. [16], if a candidate-by-candidate quantity (here, the efficiency function) enters a fit with a signal and background component, the probability density function (PDF) for this quantity needs to be included in the fit, unless it happens to be identical for both components. In our case, this constitutes a significant complication as it requires fitting a distribution of efficiency functions rather than just numbers. How this is accomplished, with an unusual application of the Fisher discriminant method to translate the efficiency functions into numbers, is described in Sec. VII.

While we take care that in extracting the B lifetime from the data we do not use any input from simulation, we do use simulated events to test our analysis method and also to evaluate systematic uncertainties. We use a full GEANT3-based detector simulation [17], (which includes a trigger simulation), as well as a detailed fast simulation for high-statistics studies. The results of the simulation studies are presented in Sec. VI, and Sec. IX. Finally, in Sec. VIII we show the results of applying the method to our data, and in Sec. X we summarize our conclusions. We start the remainder of the paper with a brief description of the relevant components of the CDF detector, in particular, the trigger (Sec. III), followed by the description of the event reconstruction, data selection and sample composition (Sec. IV).

III. THE CDF II DETECTOR AND TRIGGER SELECTION

This analysis uses data corresponding to 1 fb^{-1} of integrated luminosity collected by the CDF II detector at the Fermilab Tevatron using $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. The data were collected during the first four years (2002–2006) of the ongoing Run-II data taking period. The CDF II detector is described in detail elsewhere [18]. A brief description of the most relevant components and aspects for this analysis follows.

A. CDF II Detector

The CDF II detector has a cylindrical geometry with forward-backward symmetry and a tracking system in a 1.4 T magnetic field, coaxial with the beam. The tracking system is surrounded by the calorimeters and muon detection chambers. A cylindrical coordinate system, (r, ϕ, z) is used with origin at the geometric center of the detector, where r is the perpendicular distance from the beam, ϕ is the azimuthal angle and the $\hat{z}$ direction is in the direction of the proton beam. The polar angle θ with respect to the proton beam defines the pseudorapidity η which is given by $\eta = -\ln(\tan\frac{\theta}{2})$.

The CDF II detector tracking system consists of an open cell argon-ethane gas drift chamber called the central outer tracker (COT) [19] and a silicon vertex microstrip detector (SVX-II) [20]. The SVX-II is 96 cm long, with three sub-sections in z and has five concentric layers of double

sided silicon microstrip detectors from r=2.45 to r=10.60 cm segmented into 12 wedges in ϕ. The COT is 310 cm long, consisting of 96 sense wire layers grouped into eight alternating axial and 2° stereo superlayers. The three barrel intermediate silicon layer detector (ISL) [21] lies between a radius of 20.0 and 29.0 cm and helps in extending the η coverage of the SVX-II and COT. Together the SVX-II, ISL and COT provide r-ϕ and z measurements in the pseudorapidity range $|\eta|<2$ ($|\eta|<1$ for tracks traversing all eight superlayers).

B. Track Parametrization

A charged particle has a helical trajectory in a constant magnetic field. A description of the five parameters used to describe tracks at the CDF experiment follows. In the transverse plane, which is the plane perpendicular to the beam direction, the helix is parametrized with track curvature C, impact parameter d_0, and azimuthal angle ϕ_0. The projection of the track helix onto the transverse plane is a circle of radius R, and the absolute value of the track curvature is $|C|=\frac{1}{2R}$. The curvature is related to the magnitude of the track's transverse momentum, p_T, by $|C|=\frac{1.49898\cdot 10^{-3}\cdot B}{p_T}$, where C is in cm^{-1}, B is in T and p_T is in GeV/c, where c is the speed of light in vacuum. The sign of the curvature matches the sign of the track charge. The d_0 of a track is another signed variable; its absolute value corresponds to the distance of closest approach of the track to the beam line. The sign of d_0 is taken to be that of $(\hat{p}\times\hat{d})\cdot\hat{z}$, where $\hat{p}$ is the unit vector in the direction of the particle trajectory, $\hat{d}$ is the direction of the vector from the primary interaction point to the point of closest approach to the beam, and $\hat{z}$ is the unit vector in the direction of increasing z. The angle ϕ_0 is the azimuthal angle between $\hat{x}$ and the particle momentum at closest approach. The two remaining parameters that uniquely define the helix in three dimensions are the cotangent of the angle θ between the z axis and the momentum of the particle and z_0, the position along the z axis at the point of closest approach to the beam.

C. Trigger Selection

The CDF II detector hadronic B trigger is at the heart of this analysis. It collects large quantities of hadronic B decays, but through its impact-parameter-based selection also biases the measured proper decay time distribution. The CDF II detector has a three level trigger system. The first two levels, level 1 (L1) and level 2 (L2), are implemented in hardware and the third, level 3 (L3), is implemented in software on a cluster of computers using reconstruction algorithms that are similar to those used offline. The CDF trigger has many different selections. In this paper we refer to the family of triggers aimed at collecting samples of multi-body hadronic B decays as the "two track trigger". The silicon vertex trigger (SVT) single track finding efficiency as a function of d_0, $\varepsilon(d_0)$, is an important factor in this analysis. There have been three improvements in the SVT efficiency over the course of the data taking time period used by this analysis due to changes in the pattern recognition algorithm. These have led to three consecutive time periods in which $\varepsilon(d_0)$ has improved. These three periods and different resulting efficiencies are incorporated into the analysis as described in Sec. VI. A determination of the beam collision point or primary vertex is continuously made by the SVT during each data taking period (defining a run) and is used by all relevant triggers. After data taking is complete, the offline algorithm uses full detector information and fully reconstructed three dimensional tracks for a more accurate determination.

At L1 the trigger uses information from the extremely fast tracker (XFT) [22] finds two tracks in the COT and imposes criteria on track p_T and opening angle; at L2 the (SVT) [23] which uses silicon hits and fast pattern recognition reapplies the p_T criteria, associates silicon hits with each XFT track and requires that the absolute value of each track's d_0 lies between 120 and 1000 μm. Due to the different algorithms, the impact parameter for any given track measured by the L2 (SVT) is, in general, different from the impact parameter calculated by the L3 or offline reconstruction algorithms for the same track. These different measurements of impact parameter are referred to in this paper as $d_0^{\rm L2}$, $d_0^{\rm L3}$, and $d_0^{\rm off}$ from L2 (SVT), L3, and offline algorithms, respectively.

At L2 additional criteria are imposed on variables calculated from each track pair found by the SVT. The variables are: the product of the track charges (opposite or same sign), the projection of the vector from the primary vertex to the two track vertex in the transverse plane along the vectorial sum of the transverse momentum of the tracks, referred to as L_{xy}, a track fit χ^2 quantity, the opening angle of the two tracks in the transverse plane and finally the scalar sum of the p_T of the two tracks. In addition, one of the trigger paths requires a higher p_T for each individual track. The L3 trigger uses a full reconstruction of the event with all detector information, but uses

a slightly simpler tracking algorithm than the one used offline and reconfirms the criteria imposed by L2. In addition, the difference in z_0 of the two tracks is required to be less than 5 cm removing events where the pair of tracks originate from different collisions within the same crossing of p and $\overline{p}$ bunches.

The criteria for the three different two-track trigger configurations, used in this analysis, are summarized in Table I in terms of the quantities described above. It is clear that the impact parameter and L_{xy} requirements will preferentially select long-lived B hadron decays over prompt background. The three selections are referred to as the low-p_T, medium-p_T and high-p_T selections, this is a reference to their single track p_T (> 2.0, 2.0, 2.5 GeV/c, respectively) and track pair p_T scalar sum (> 4.0, 5.5, 6.5 GeV/c, respectively) cuts.

The requirements of the three trigger selections mean that any event that passes the high-p_T selection, simultaneously satisfies the requirements of the low and medium p_T selections. The three separate selection criteria exist because of the need to control the high trigger acceptance rates that occur at high instantaneous luminosity due to high track multiplicity. The rates are controlled by the application of prescaling, which is the random rejection of a predefined fraction of events accepted by each trigger selection dependent on the instantaneous luminosity. Therefore only the higher purity, but less efficient high-p_T selection is available to accept events at higher luminosities.

IV. DATA SELECTION AND EVENT RECONSTRUCTION

A. Reconstruction of the decay $B^- \to D^0\pi^-$

The reconstruction of the decay $B^- \to D^0\pi^-$ uses data collected by the two track trigger described in Sec. III C. Standard track quality selection criteria are applied to all individual tracks; each track is required to have $p_T > 0.4$ GeV/c, $\mid \eta \mid < 2$, a minimum of five hits in at least two axial COT super layers, a minimum of five hits in at least two stereo COT super layers and a minimum of three silicon hits in the SVX-II r-ϕ layers.

Candidate $D^0 \to K^-\pi^+$ decays are searched for first. As no particle identification is used in this analysis, the search for D^0 candidates considers all pairs of oppositely charged tracks which are then assumed to be K^- and π^+ and assigned the kaon and pion mass, respectively. The two tracks are then constrained to come from a common vertex and the invariant mass (m_{D^0}) and $p_T(D^0)$ is calculated. Candidates are required to have a mass within 0.06 GeV/c^2 of the world average D^0 mass, 1.8645 GeV/c^2 [24], and $p_T(D^0) > 2.4$ GeV/c. The $K^-\pi^+$ pair is required not to exceed a certain geometric separation in the detector. Defining the separation in the η-ϕ plane, in terms of the differences in η and ϕ of the two tracks, as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, we require $\Delta R < 2$. The separation in z_0 of the two tracks is required to be $\Delta z_0 < 5$ cm. The candidate D^0 is then combined with each remaining negatively charged track with $p_T > 1$ GeV/c in the event. These are assumed to be pions from the decay $B^- \to D^0\pi^-$. The D^0 and the π^- are constrained to a common vertex assumed to be the decay point of the B^- with the D^0 mass constrained to the world average. The three tracks can be combined to measure the invariant mass of the candidate B^-, m_B. It is possible to have more than one candidate event from the same bunch crossing.

Proper decay time calculations in this paper are made using distances measured in the plane transverse to the beam. The proper decay time of the B^-, t, is given by

$$t = \frac{L_{xy}}{c\,(\beta\gamma)_T} = L_{xy} \cdot \frac{m_B}{cp_T}, \qquad (1)$$

where L_{xy} is the projection of the distance from the primary vertex to the B^- vertex along the direction of the transverse momentum of the B^- and $(\beta\gamma)_T = \frac{p_T}{m_B}$ is the transverse Lorentz factor. The statistical uncertainty on L_{xy}, $\sigma_{L_{xy}}$, is calculated from the full covariance matrix of the vertex constrained fit and is dominated by the primary vertex resolution which is approximately 33 μm. We have used the average beam position per run, which is calculated offline for each run, as an estimate of the primary vertex position. The uncertainty on the proper decay time is calculated by transforming $\sigma_{L_{xy}}$ into the B rest frame.

To reduce background we require that the B^- candidate must have: $5.23 < m_B < 5.5$ GeV/c^2, $0 < t < 10$ ps, $p_T > 5.5$ GeV/c, $L_{xy} > 350$ μm, $\sigma_t < 0.333$ ps, and that the impact parameter of the B with respect to the beam spot is smaller than 80 μm . We also require that the χ^2 of the vertex constrained fit is less than 15, that all tracks have z_0 within 5 cm of each other, and that $\Delta R(D^0, \pi^-) < 2$.

It is possible to reconstruct candidates where no pair of tracks in the final state meet the trigger criteria. The lifetime measurement method presented here cannot be used on these candidates, and these candidates are removed by reconfirm-

TABLE I: Trigger selection criteria for the three two-track trigger paths. We use n/a where no criterion is applied. † - The trigger requirements on the χ^2 were altered during the data taking period. The quantity in brackets refers to the first 0.21 fb$^-$1 collected.

Trigger criteria L1	Units	`Low` p_T	`Medium` p_T	`High` p_T
Minimum track p_T	GeV/c	2.0	2.0	2.5
Two track charge product	-	n/a	-1	-1
Two track max $\Delta\phi$	degrees	90°	135°	135°
Two track p_T scalar sum	GeV/c	4.0	5.5	6.5
Trigger criteria L2				
Minimum $\lvert d_0^{\mathrm{L2}}\rvert$	μm	120	120	120
Maximum $\lvert d_0^{\mathrm{L2}}\rvert$	μm	1000	1000	1000
Minimum track p_T	GeV/c	2.0	2.0	2.5
Maximum track χ^2	-	15(25)†	15(25)†	15(25)†
Track pair charge product	-	n/a	-1	-1
Maximum pair $\Delta\phi$	degrees	90°	90°	90°
Minimum pair $\Delta\phi$	degrees	2°	2°	2°
Minimum pair p_T scalar sum	GeV/c	0	5.5	6.5
Minimum pair L_{xy}	μm	200	200	200
Trigger criteria L3				
Minimum $\lvert d_0^{\mathrm{L3}}\rvert$	μm	80	80	80
Maximum $\lvert d_0^{\mathrm{L3}}\rvert$	μm	1000	1000	1000
Minimum track p_T	GeV/c	2.0	2.0	2.5
Maximum track η	-	1.2	1.2	1.2
Track pair charge product	-	n/a	-1	-1
Maximum pair $\Delta\phi$	degrees	90°	90°	90°
Minimum pair $\Delta\phi$	degrees	2°	2°	2°
Maximum pair Δz_0	cm	5.0	5.0	5.0
Minimum pair p_T scalar sum	GeV/c	4.0	5.5	6.5
Minimum pair L_{xy}	μm	200	200	200

ing the trigger. We require that at least one track pair from each candidate decay passes the L2 and L3 trigger selection requirements. The particular L2 and L3 selection that the decay must pass depends on which trigger selection accepted the event during data taking. In the case where more than one trigger selection was satisfied during data taking, we require that the candidate satisfies the least stringent selection. Reconfirmation of the trigger requires that the offline reconstructed tracks are associated to L2 and L3 tracks in the event. To match an offline track to a L2 or L3 track we calculate the $\chi^2 = \left(\frac{\Delta C}{\sigma_C}\right)^2 + \left(\frac{\Delta\phi}{\sigma_\phi}\right)^2$ between an offline track and each L2 or L3 track in the candidate, where ΔC and $\Delta\phi$ are the differences between the offline and L2 or L3 track C (curvature) and ϕ, respectively, and σ_C and σ_ϕ are the mean uncertainties on the offline track C and ϕ, respectively. The L2 or L3 track that has the lowest χ^2 is associated with the corresponding offline track. We apply $\chi^2 < 95(25)$ to remove matches in the case where there is no L2(L3) track that is similar to the offline track. These requirements are loose due to the significant differences between the impact parameter calculation algorithms.

Collectively, the trigger selection requirements and the cuts made on offline or derived variables are referred to as the selection criteria. The kinematics of each track is used to calculate the efficiency function central to this method. We use the following nomenclature to refer to each individual track. The pion originating from the B^- vertex is referred to as π_B and the pion and kaon originating from the D vertex are referred to as π_D and K_D, respectively.

B. Sample composition and signal yield

The resulting mass distribution after the selection criteria have been applied is shown in Fig. 1.

The lower background sideband and a small part of the signal peak have been removed by the requirement that $m_B > 5.23$ GeV/c^2. This cut has been applied to remove partially reconstructed $B^- \to D^{*0}\pi^-/\rho^-$ and $B^0 \to D^{(*)-}\pi^+/\rho^+$ decays, where only three tracks of the final state are used leading to a lower B mass. If left in the sample, these partially reconstructed B mesons would bias the proper decay time distribution, since they resemble signal candidates, but, due to the missing momentum, their proper decay time has been mis-measured (see Eq. (1)). Detailed Monte Carlo has shown that the applied mass removes the contamination from these partially reconstructed B decays. No other partially reconstructed B hadron decays are expected to populate this mass range. The Cabibbo suppressed decay $B^- \to D^0K^-$ is also present in this sample where the kaon from the B is reconstructed as a pion. The lower mass cut does not remove all of these candidates, but a tighter cut would remove too many $B^- \to D^0\pi^-$ candidates. For simplicity, the $B^- \to D^0K^-$ candidates are not fit separately and are treated as $B^- \to D^0\pi^-$ candidates for the lifetime determination. This simplification is motivated by the small size of the contamination (3%), and the small difference in reconstructed proper decay time between the K and the π mass assignment of the kaon track which is of order 1%. The resulting systematic uncertainty was evaluated and found to be negligible (Sec. IX). The mass distribution of the remaining signal candidates, including both $B^- \to D^0\pi^-$ and $B^- \to D^0K^-$, is modeled by the sum of two Gaussians each with an independent mean and width. The background candidates are due to track combinations that mimic the signature of signal decays. The mass distribution of background candidates is modeled by a linear function. An alternative description which allows for a second order polynomial to model the background was found to be degenerate with the linear function.

To determine the signal yield the mass distribution is fit by maximizing an unbinned log likelihood $\mathcal{L}$ which is calculated using the mass, m_i, for each candidate. The letters s and b define whether the PDF describes signal or background candidates. The likelihood is given by

$$\log(\mathcal{L}) = \log\left\{\prod_i^N \left[f_s\mathcal{P}(m_i|s) + (1-f_s)\,\mathcal{P}(m_i|b)\right]\right\}, \tag{2}$$

where $\mathcal{P}(m_i|s)$ is given by

$$\mathcal{P}(m_i|s) = \left[\frac{f_1}{\sigma_1\sqrt{2\pi}}e^{-\frac{(m_i-m_1)^2}{2\sigma_1^2}} + \frac{(1-f_1)}{\sigma_2\sqrt{2\pi}}e^{-\frac{(m_i-m_2)^2}{2\sigma_2^2}}\right]\cdot\mathcal{A}, \tag{3}$$

where the factor $\mathcal{A}$ is required to satisfy the normalization condition

$$\int_{m_{\rm low}}^{m_{\rm high}} \mathcal{P}(m_i|s)dm_i = 1. \tag{4}$$

$\mathcal{P}(m_i|b)$ is described by a first order polynomial and is given by:

$$\mathcal{P}(m_i|b) = \frac{1-\alpha m_i}{\left[m_{\rm high} - m_{\rm low} - \frac{\alpha}{2}\left(m_{\rm high}^2 - m_{\rm low}^2\right)\right]}, \tag{5}$$

where $m_{\rm low}$ and $m_{\rm high}$ are the lower and upper mass limits, 5.23 and 5.5 GeV/c^2, respectively.

The free parameters in the mass fit are m_1, m_2, σ_1, σ_2, α, f_1, and f_s. The data are fit and the mass fit projection is shown in Fig. 1. From the results of the mass fit a yield of 23900±200 signal candidates is determined. We define the upper sideband to be the candidates with $5.38 < m_B < 5.5$ GeV/c^2. These candidates are retained to constrain the parameters of the background component of the lifetime fit. The best fit parameters are given in Appendix D.

The results of the mass fit are also used to extract the signal distribution of various parameters using background subtraction. We use this technique in several places for cross checks, but not as a method to extract the lifetime or any other fit parameter. For the purpose of background subtraction, we define a signal window by $5.25 < m_B < 5.31$ GeV/c^2. The results of the mass fit are used to calculate the fraction of background candidates in the signal region. For any given parameter, we subtract an appropriately scaled upper sideband distribution from the distribution found in the signal region to obtain the signal distribution in data.

V. REMOVING THE SELECTION-INDUCED BIAS FOR SIGNAL EVENTS

A. Introduction

In this section we derive the PDF that takes into account lifetime bias due to the trigger and other selection criteria without input from simu-

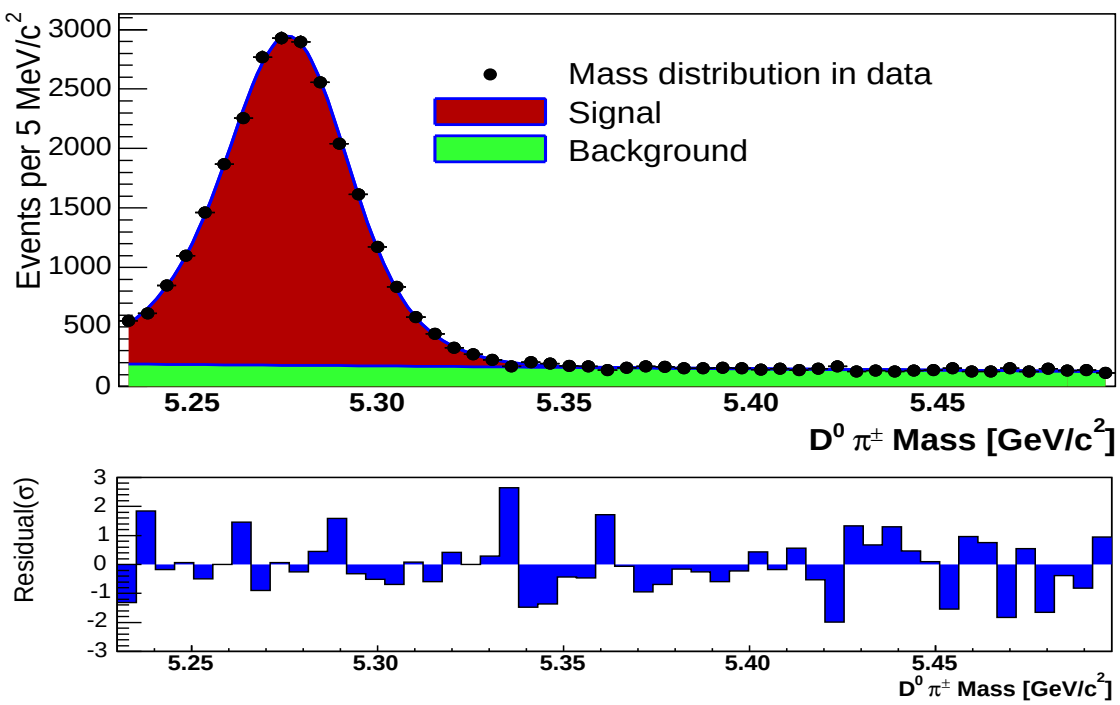


FIG. 1: The top plot shows the mass fit projection (line) on the data (points). The bottom plot shows the residual for each bin: $(N_{\rm fit} - N_{\rm data})/\sqrt{N_{\rm data}}$.

lation. Only the case of pure signal is considered in this section, whereas the complications introduced by the presence of background events are discussed in Sec. VII.

We start by considering a proper decay time distribution that is not subject to any bias, which is given by an exponential. To incorporate detector effects, the exponential is then convolved with a resolution function. For the purpose of this measurement, the proper decay time resolution function at the CDF detector is adequately described by a single Gaussian of fixed width. For a decay with mean lifetime τ and Gaussian proper decay time resolution of width σ_t, the probability density to observe a signal candidate decaying with proper time t_i, where the subscript i labels the candidate, is given by

$$\mathcal{P}\left(t_i;\tau|s\right) = \frac{1}{\tau} e^{\frac{-t_i}{\tau}+\frac{\sigma_t^2}{2\tau^2}} F\left(\frac{t_i}{\sigma_t} - \frac{\sigma_t}{\tau}\right), \quad \text{where } F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-y^2}{2}}\, dy. \tag{6}$$

The effect of the trigger and offline selection on the observed proper decay time distribution is taken into account by introducing a time-dependent efficiency function $E(t)$. This efficiency function depends on a number of variables including the *single* track finding efficiency ε_s. The crux of the analysis presented here is that we use a different efficiency function $E_i(t,\varepsilon_s)$ for each candidate i, and that we calculate each $E_i(t,\varepsilon_s)$ analytically from the candidate's decay kinematics and the selection criteria. The selection criteria are referred to by the symbol T. With this, the PDF for a candidate with decay time t_i is

$$\mathcal{P}\left(t_i;\tau|\mathrm{T}, E_i(t,\varepsilon_s), s\right) = \frac{E_i(t_i,\varepsilon_s) \times \frac{1}{\tau} e^{\frac{-t_i}{\tau}+\frac{\sigma_t^2}{2\tau^2}} F\left(\frac{t_i}{\sigma_t} - \frac{\sigma_t}{\tau}\right)}{\int_{-\infty}^{\infty} E_i(t,\varepsilon_s) \times \frac{1}{\tau} e^{\frac{-t}{\tau}+\frac{\sigma_t^2}{2\tau^2}} F\left(\frac{t}{\sigma_t} - \frac{\sigma_t}{\tau}\right) dt}. \tag{7}$$

B. Calculation of $E_i(t,\varepsilon_s)$

1. *Scanning through different potential proper decay times*

In order to find the function $E_i(t,\varepsilon_s)$ for a given candidate i, we need to find the trigger efficiency for that candidate for all possible B proper decay times. We scan through different B decay times by translating the B decay vertex along the B flight direction, defined by the reconstructed B momentum. At each point in the scan, we re-

calculate all decay-time dependent properties of the candidate, in particular the impact parameters and decay distance. Properties that are independent of proper decay time (before cuts are applied), such as the four momenta of all particles or the flight distance of the intermediate D meson, remain constant. We re-apply the trigger and other selection criteria to the translated candidate. If the translated candidate fails the selection criteria, $E_i(t, \varepsilon_s)$ is zero for that candidate at the corresponding decay time. Otherwise $E_i(t, \varepsilon_s)$ is non-zero at time t and its exact value depends on the SVT (L2) track-finding efficiency, ε_s. This process is illustrated and described in detail in Sec. V B 4. Prior to this, we discuss two complications to the basic idea presented above. Firstly, not all tracks found offline are also found by the SVT and furthermore, the probability that the SVT finds a track depends on its impact parameter. Secondly, at different stages in the event reconstruction and selection, different algorithms are used to calculate the track parameters - very fast algorithms at L2, more detailed ones at L3, and finally the full tracking and vertexing in the final offline reconstruction. The measured values of track parameters such as impact parameters differ slightly depending on the algorithm used for the calculation. Below we describe these and other important aspects of the acceptance function calculation in greater detail.

2. *The value of $E_i(t, \varepsilon_s)$ and its dependence on the SVT track finding efficiency*

a. The need to include the dependence on ε_s If the track-finding efficiency is proper decay time independent, one can base the fit on the PDF *given* that a certain track combination has been reconstructed and seen by the trigger. In practice, this means that the track finding efficiency is constant as a function of the impact parameter since the decay time and the impact parameter are correlated. In the case where the track-finding efficiency is proper decay time independent, the set of tracks seen by the trigger would be treated exactly in the same way as the decay kinematics, something that can be kept constant as the decay distance is changed for the efficiency function evaluation. Given that a certain track combination has been found, the trigger efficiency at a certain decay time is either 1 (passes cuts) or 0 (fails), independent of ε_s. This PDF would ignore one factor; the probability that exactly this track combination has been found. If this factor is proper decay time independent it does not affect the maximum of the likelihood and hence the result of the fit.

However, the SVT track finding efficiency of the CDF II detector is impact parameter, and therefore time dependent as can be seen in Fig. 2. The figure shows the SVT track finding efficiency for tracks found in the offline reconstruction in data as a function of the track's offline impact parameter ($|d_0^{\rm off}|$). The SVT track finding efficiency is approximately constant for $0 < |d_0| < 1000\,\mu$m and falls rapidly for $|d_0| > 1000\,\mu$m. The correct efficiency distribution is obtained from the signal region of the data sample used in the fit, using the following method: The efficiency prior to triggering is obtained by considering the sub-sample of candidates where two particular tracks can pass the trigger requirements. For these candidates, the remaining third track is used to obtain the SVT track finding efficiency.

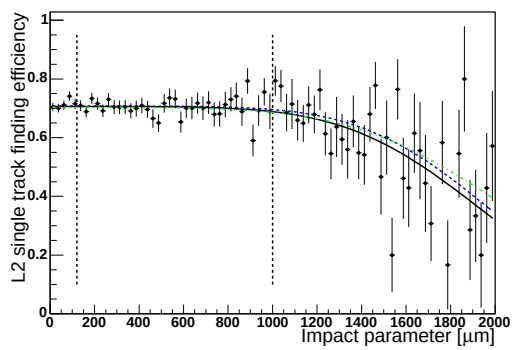


FIG. 2: The L2 single track finding efficiency, relative to the offline efficiency, as a function of $|d_0^{\rm off}|$. The points represent the data. The vertical dashed lines represent the trigger selection cuts. The fitted curves represent possible descriptions of the efficiency which are described in the discussion of systematic uncertainties in Sec. IX.

Even though ε_s is approximately constant within the trigger acceptance requirements, the rapid drop after $|d_0| > 1000\,\mu$m, introduces a particular problem. The trigger efficiency is calculated depending on which tracks are found by the SVT. If ε_s is constant for all impact parameters, then the tracks which *were actually* found by the SVT can be used calculate the trigger efficiency, and we can assume that the *same* tracks would be found as the decay vertex is slid along the direction of the B momentum. However, since ε_s does vary with $|d_0^{\rm off}|$, the probability of the SVT finding tracks is dependent on the decay position. This means that at a given decay time, for a particular candidate, the efficiency depends on how likely the SVT is to actually find those track combinations that could be accepted by the trigger. The more tracks combinations there are that pass the selection criteria,

the higher the probability that at least one is found by the SVT. Under these circumstances, the proper decay time dependence of the SVT trigger efficiency has to be taken into account, which requires some parametrization of the single track finding efficiency as a function of $d_0^{\rm off}$.

b. Parametrizing $\varepsilon_s(|d_0|)$ While the inclusion of the single track finding efficiency in the PDF increases the complexity of the measurement, we can take, as a good approximation, the following simple model. We model the SVT track finding efficiency as constant for $0 < |d_0| < 1000\,\mu$m. We treat all tracks with $|d_0| > 1000\,\mu$m as not-found by the SVT (which does not affect the trigger decision as it requires $120\,\mu\text{m} < |d_0| < 1000\,\mu\text{m}$), so that we can describe the SVT efficiency by the following simple description:

$$\varepsilon_s(d_0) = \left\{ \begin{array}{c} \varepsilon_s \;\; \text{if } |d_0| < 1\,\text{mm} \\ 0 \;\; \text{otherwise} \end{array} \right\}. \qquad (8)$$

The consequence on the lifetime measurement of the small deviations of the real SVT efficiency from this simple model are discussed Sec. IX A. We also assume that there is no variation in track finding efficiency as a function of track p_T or η. Such variations can alter the probability of finding a particular track combination. However as this is time independent, the effect on the lifetime measurement is expected to be small. This is also discussed in Sec. IX A, where we show that the effects of these simplifications on the lifetime measurements are indeed, sufficiently small. There is an alternative, simpler approach, that does not depend on ε_s, which is suitable in situations where the track-finding efficiency is constant over a larger range than for the SVT at the CDF II detector. This is discussed in Appendix C.

c. Calculating $E_i(t, \varepsilon_s)$ The value of $E_i(t, \varepsilon_s)$ for a given decay time is the probability that at least one of the possible track combinations that pass the trigger criteria is in fact found by the L2 tracking algorithms. For example, if there is only one track pair in the candidate that can pass the selection requirements, then the probability of finding both those tracks is ε_s^2, where we simply take the product of two single track finding efficiencies. For a three body final state, where there are two possible track pairs that pass the trigger, the probability is given by $2\varepsilon_s^2 - \varepsilon_s^3$. In cases where there are three possible track pairs (only possible for the `low`-p_T selection that makes no requirement on track charge), the probability to find sufficient tracks to pass the trigger is $3\varepsilon_s^2 - 2\varepsilon_s^3$.

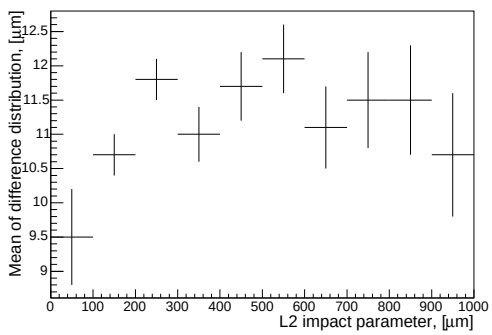
3. *Translating Online and Offline quantities*

To calculate the trigger efficiency for all possible B proper decay times we scan through different B decay points along the B flight path and determine the probability that the trigger was passed at that point. As we re-apply the trigger cuts, we always base the decision on the quantities accessible to the relevant trigger level i.e., L2 criteria to SVT tracks, L3 criteria to L3 tracks, and offline criteria to the fully reconstructed offline tracks. Certain quantities such as the track momentum or the opening angle between two tracks will remain constant as the vertex is translated along the B flight path. Other quantities such as the impact parameter will change. Therefore, as we translate the B decay along its flight direction, we need to re-calculate the decay time-dependent quantities for each level: L2, L3, and offline.

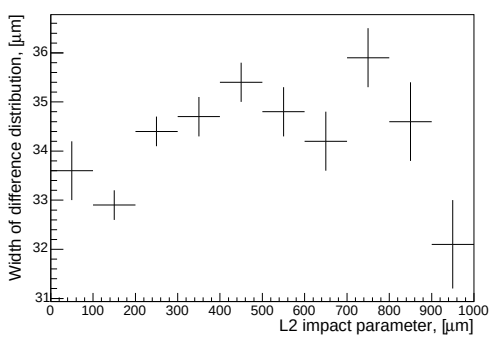
The calculation of the offline impact parameters or reconstructed proper decay time as the candidate is translated along its flight path is trivial. Furthermore, as the candidate-by-candidate efficiency is a function of the *offline-reconstructed* proper decay time, rather than the true decay time, it is not necessary to reconsider the effects of detector resolution. This means that there is a simple, one-to-one relationship between the offline-reconstructed decay time of the translated candidates and the other time-dependent offline quantities such as impact parameters and L_{xy}, without the need to take into account further resolution effects. We aim to retain a similarly simple direct relationship between proper decay time and trigger cuts for online quantities as well. The complication arises only for decay time dependent quantities, and since all L2 and L3 decay time dependent quantities (d_0, L_{xy}) are calculated from the impact parameters of the tracks, d_0 is the only parameter we need to consider.

As we translate the candidate along, we recalculate each track's online d_0 at L2 and L3 by assuming that the differences between online and offline quantities are not decay time dependent. This way, we can treat this difference in exactly the same way as the other proper decay time independent quantities in the candidate, such as track p_T. We measure them in each candidate and keep them constant as we translate the candidate along. It is possible that the difference between the L2 and offline impact parameter, $(\Delta d_0)_{L2} = d_0^{L2} - d_0$, could vary as a function of impact parameter due to the finite hit recognition patterns used to measure the L2 impact parameter. We verify in data that these differences are time independent. To check this, we

calculate $(\Delta d_0)_{L2}$ and bin it according to track $|d_0^{\mathrm{L2}}|$. In each bin, the $(\Delta d_0)_{L2}$ distribution is fitted with a Gaussian, and the mean and width of the fitted Gaussian for different impact parameter ranges is shown in Fig. 3. There are some deviations from a straight line, but there is no systematic dependence on impact parameter, and hence on impact parameter resolution as a function of decay time. In Sec. IX we evaluate the systematic uncertainty on the fitted lifetime that variations of the impact parameter resolution as a function of impact parameter such as those observed in the data have and find it to be very small (0.002 ps).



(a)



(b)

FIG. 3: The difference, $(\Delta d_0)_{L2}$ is binned as a function of $|d_0^{\mathrm{L2}}\|$ and fitted to a Gaussian. The mean of the fitted Gaussian is shown in (a) while the width is given in (b). The variation is of order a few microns.

The $(\Delta d_0)_{L2}$ for a given track is measured at the actual point of decay by accessing the information of the L2 track matched to the offline track. This is then used to calculate the translated L2 impact parameter $d_0^{L2}(t)$ from the translated offline $d_0(t)$ at each point: $d_0^{L2}(t) = d_0(t) + (\Delta d_0)_{L2}$. A complication arises for those tracks not found by the SVT (such as those with $|d_0| \gg 1$ mm). In this case, a value of $(\Delta d_0)_{L2}$ is assigned by drawing a value at random from the distribution of $(\Delta d_0)_{L2}$ from tracks where it is possible to calculate $(\Delta d_0)_{L2}$. One further issue to consider is that the L2 algorithm measures impact parameters to the closest 10 μm. To emulate this feature of the L2 tracking algorithm the calculated $d_0^{L2}(t)$L2 is rounded to the closest multiple of 10 μm. The same procedure is applied to estimate d_0^{L3}except that no discretization is necessary. The online L_{xy} values at L2 and L3 for each track pair are then re-calculated from the translated d_0 of each track.

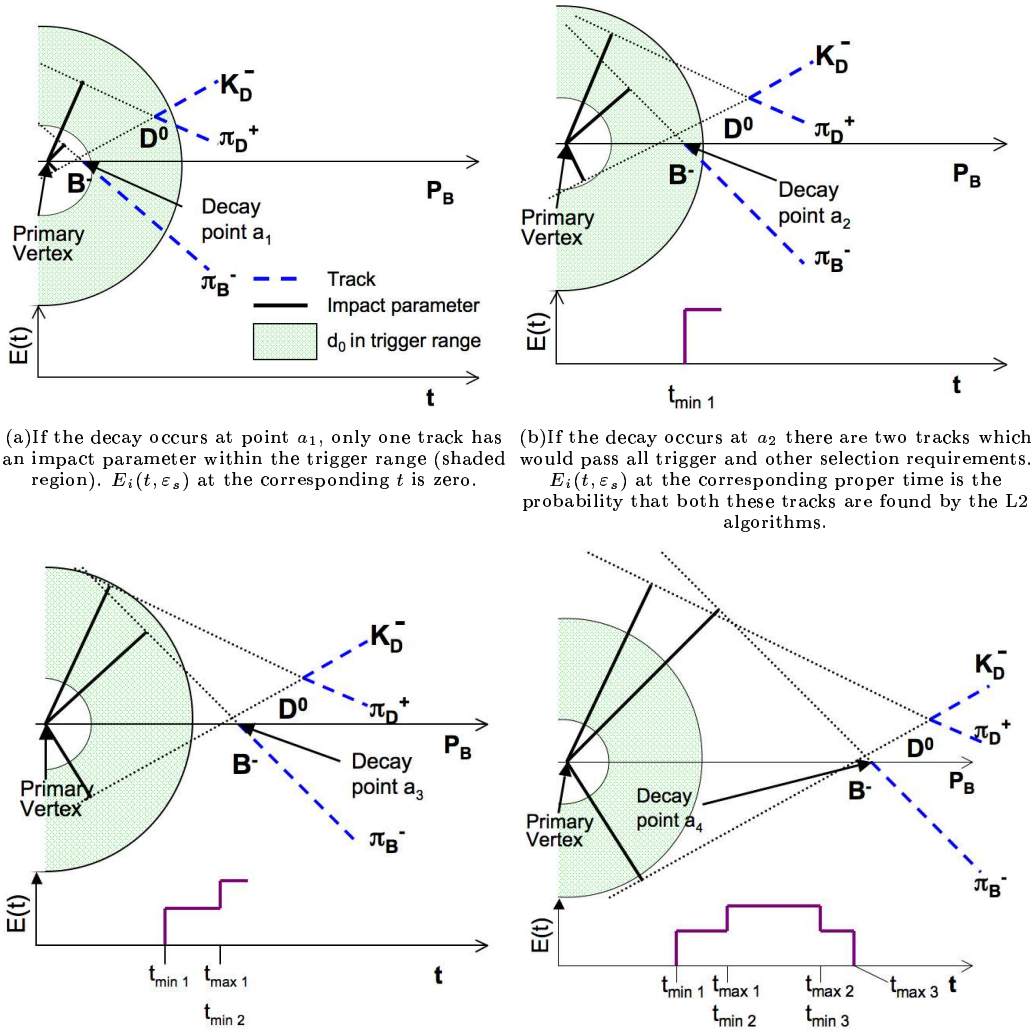
4. *Example*

To illustrate the entire process, we describe in detail a specific example shown in Fig. 4. Figure 4 shows the same decay at four different decay times. For the purposes of this illustration we assume this decay has been accepted by the medium-p_T trigger selection. First, we consider the decay vertex translated to point a_1 as shown in Fig. 4(a). The decay vertex is close to the primary interaction point and only one track has $|d_0^{L2}| > 120\,\mu$m, therefore the selection requirements are not met and the value of $E_i(t, \varepsilon_s)$ at the proper decay time corresponding to a_1 is $H_1 = 0$.

In Fig. 4(b) the decay vertex has been translated further along the B momentum direction and is at the point where one track pair satisfies the trigger selection and the B decay satisfies all other selection requirements listed in Sec. IV A. At this point, a_2, the value of $E_i(t, \varepsilon_s)$ is given by probability of finding both the π_B and the K_D track, which is $H_2 = \varepsilon_s^2$.

As this candidate is further translated along its B momentum direction it moves into the region where all three tracks can participate in the trigger decision. In Fig. 4(c), two track combinations fulfill the trigger requirements, (π_B, π_D) and (π_D, K_D). The remaining combination, (π_B, K_D), does not pass the trigger in this case as it does not satisfy the opposite charge requirement of the medium-p_T trigger. The value of $E_i(t, \varepsilon_s)$ at the decay point a_3 is the probability that at least one of the two possible track combinations is found by the SVT, $H_3 = 2\varepsilon_s^2 - \varepsilon_s^3$.

In Fig. 4(d) the decay vertex has been translated to the point a_4 where the track impact parameter requirements are not satisfied. The value of $E_i(t, \varepsilon_s)$ returns to zero at the point where the trigger requirements are not met. Hence $E_i(t, \varepsilon_s)$ can be described by a series of intervals limited by $t_{\min}$ and $t_{\max}$ and within an interval the value of $E_i(t, \varepsilon_s)$ is given by a polynomial in terms of ε_s, $H(\varepsilon_s)$. The efficiency function can be written



(a)If the decay occurs at point a_1, only one track has an impact parameter within the trigger range (shaded region). $E_i(t, \varepsilon_s)$ at the corresponding t is zero.

(b)If the decay occurs at a_2 there are two tracks which would pass all trigger and other selection requirements. $E_i(t, \varepsilon_s)$ at the corresponding proper time is the probability that both these tracks are found by the L2 algorithms.

(c)If the decay occurs at a_3 there are two tracks pairs which would pass all trigger requirements. $E_i(t, \varepsilon_s)$ rises at the point $t_{min\ 2}$ as the probability to find at least one of the two available track pairs is greater than the probability to find a particular track pair.

(d)If the decay occurs at a_4 all the track impact parameters are above the trigger threshold. $E_i(t, \varepsilon_s)$ at $t_{max\ 3}$ returns to zero.

FIG. 4: The decay vertex is translated along the direction of the B momentum while the decay kinematics are held fixed. At each decay point it is determined whether or not the selection criteria could be satisfied and $E_i(t, \varepsilon_s)$ is calculated. Diagrams not to scale.

in terms of the Heaviside step function θ as

$$E_i(t, \varepsilon_s) = \sum_{\substack{k_i = \text{all} \\ \text{intervals} \\ \text{in event } i}} \Big\{ H_{k_i}(\varepsilon_s) \big[\theta\left(t - t_{\min\ k_i}\right) - \theta\left(t - t_{\max\ k_i}\right) \big] \Big\}. \tag{9}$$

C. The signal PDF and its parameters

Substitution of the form of $E_i(t, \varepsilon_s)$ into Eq. (7) leads to the following PDF for observing a decay at time t_i:

$$\mathcal{P}(t_i;\tau|\mathrm{T},E_i(t,\varepsilon_s),s) = \frac{E_i(t_i,\varepsilon_s)\frac{1}{\tau}e^{\frac{-t_i}{\tau}+\frac{1}{2}\frac{\sigma_t^2}{\tau^2}}F\left(\frac{t_i}{\sigma_t}-\frac{\sigma_t}{\tau}\right)}{\sum\limits_{\substack{k_i=\text{all}\\ \text{intervals}\\ \text{in event } i}} \mathrm{H}_{k_i}(\varepsilon_s)\left[-e^{\frac{-t}{\tau}+\frac{1}{2}\frac{\sigma_t^2}{\tau^2}}F\left(\frac{t}{\sigma_t}-\frac{\sigma_t}{\tau}\right)+F\left(\frac{t}{\sigma_t}\right)\right]_{t=t_{\mathrm{min\ k_i}}}^{t=t_{\mathrm{max\ k_i}}}}. \quad (10)$$

We describe the detector resolution as a Gaussian with width $\sigma_t = 0.087$ ps. This is the average of the calculated candidate-by-candidate σ_{t_i} of the background subtracted signal region in data. Using a single Gaussian based on a single global σ_t instead of a candidate-by-candidate value significantly simplifies the analysis and is justified since the PDF is not very sensitive to the exact value of σ_t. This is the case for two reasons: the lifetime to be measured, $\mathcal{O}(1.6\,\mathrm{ps})$, is much larger than $\sigma_t = 0.087\,\mathrm{ps}$; and the "turn-on" of the curve, for un-biased data at $t \approx 0$, is cut off by our L_{xy} cut.

In terms of the PDF in Eq. (10), this implies that all terms with σ_t only play a minor role because $t/\tau \gg \frac{1}{2}\sigma_t^2/\tau^2$ and $F\left(\frac{t}{\sigma_t}-\frac{\sigma_t}{\tau}\right) \approx F\left(\frac{t}{\sigma_t}\right) \approx 1$. These approximations are not made in the PDF, but they illustrate why the dependence on σ_t is small. In Sec. IX we confirm that the systematic uncertainty due to the resolution parametrization is small.

To use this PDF to extract the lifetime, knowledge of ε_s is also required. Although Eq. (10) could be used to simultaneously fit τ and ε_s, there is extra information available in the data that can be used to help determine ε_s with greater precision. The extra information used is simply the knowledge of exactly which tracks do and which do not have L2 information. To add this information to the PDF, we introduce a candidate observable called track configuration, C_i. This observable is defined by n, the number of tracks that are within the reach of the SVT ($p_T > 2.0\,\mathrm{GeV}/c$, $|d_0| \in [0,1]\,\mathrm{mm}$), and by r, the number of those that have L2 information. The configuration also distinguishes which specific tracks have L2 information. The probability of observing a particular C_i, i.e., that of n tracks within the reach of the SVT, a specific set of r tracks have matches, while the remaining $n-r$ tracks do not, is given by

$$\mathcal{P}(C_i|\mathrm{T},E_i(t,\varepsilon_s),t_i,s) = \frac{\varepsilon_s^r(1-\varepsilon_s)^{(n-r)}}{E_i(t,\varepsilon_s)|_{t=t_i}}, \quad (11)$$

where the factor $E_i(t,\varepsilon_s)|_{t=t_i}$ provides the correct normalisation as it is the sum of all possible configurations that could have passed the trigger.

We multiply the probabilities defined in Eqs. (10) and (11) to obtain the PDF which is used to simultaneously fit the proper decay time and ε_s. It is given by

$$\mathcal{P}(t_i;\tau|\mathrm{T},E_i(t,\varepsilon_s),s)\cdot\mathcal{P}(C_i|\mathrm{T},E_i(t,\varepsilon_s),t_i,s) = \frac{\varepsilon_s^r(1-\varepsilon_s)^{(n-r)}\frac{1}{\tau}e^{\frac{-t_i}{\tau}+\frac{1}{2}\frac{\sigma_t^2}{\tau^2}}F\left(\frac{t_i}{\sigma_t}-\frac{\sigma_t}{\tau}\right)}{\sum\limits_{\substack{k_i=\text{all}\\ \text{intervals}\\ \text{in event } i}} \mathrm{H}_{k_i}(\varepsilon_s)\left[-e^{\frac{-t}{\tau}+\frac{1}{2}\frac{\sigma_t^2}{\tau^2}}F\left(\frac{t}{\sigma_t}-\frac{\sigma_t}{\tau}\right)+F\left(\frac{t}{\sigma_t}\right)\right]_{t=t_{\mathrm{min\ k_i}}}^{t=t_{\mathrm{max\ k_i}}}}. \quad (12)$$

In the case of a two body decay, we would always find, in both the numerator and denominator of the expression, that $\mathrm{H}_{k_i}(\varepsilon_s) = \varepsilon_s^r(1-\varepsilon_s)^{n-r} = \varepsilon_s^2$; all factors containing ε_s would cancel and we would recover the expression for two-body decays derived in Ref. [25]. If there is no upper impact parameter cut or equivalent ($t_{\max} = \infty$), and the lower cut that is hard enough so that for each event $t_{\min} \gg \sigma_t$, Eq. (12) reduces to $\frac{1}{\tau}e^{-(t-t_{\min})/\tau}$, equivalent to a re-definition of $t = 0$, as used by DELPHI in Ref. [26]. Other special cases leading to some simplifications are discussed in Appendix C. However, none of these apply here and we use the full expression given in Eq. (12).

VI. VALIDATION OF THE METHOD

We test the signal PDF derived in Sec. V, and the full PDF with both signal and background component that will be derived in Sec. VII, on simulated events. We use two kinds of simulations: a full GEANT 3-based [17] detector simulation and a fast parametric simulation for high statistics studies.

A. The Full Detector Simulation

We use the full CDF II detector simulation to test whether the signal PDF constructed in Sec. V can correctly remove the selection bias. The simulated data samples used for this test consist of single B hadrons generated with p_T spectra consistent with NLO QCD [27, 28] and decayed with EvtGen [29]. A detailed GEANT 3-based detector and trigger simulation is used to produce the detector response, which is processed using the same reconstruction algorithms as data. In addition to a $B^- \to D^0\pi^-$ sample, we also use samples of three other decay modes; $B^0 \to D^+\pi^-$, $B_s \to \phi\phi$ and $B_s \to K^+K^-$, where the offline selection criteria applied are broadly similar to that of the $B^- \to D^0\pi^-$ candidates. These distinct samples, with differing topologies, allow for further crosschecks of the basis of the method to correct the selection biases . The calculation of the efficiency function is easily extended to include four track decays using the same principle of scanning through all possible proper decay times as described in Sec. V.

As these samples contain only signal events, we use the PDF described in Eq. (12) to simultaneously extract the lifetime and the L2 single track finding efficiency. The fitted lifetimes, along with the input truth lifetimes and size of each sample are given in Table II. The fitted lifetime is consistent with the input lifetime for each Monte Carlo sample. These results indicate that the method of calculating the event efficiency can be used to correct the selection biases.

B. The Fast Simulation

In addition to the full CDF II detector simulation we use a custom fast simulation which is several orders of magnitude faster than the detailed simulation. It allows production of many thousands of independent samples, each approximately the size of our data yield (24,000 signal events), which are used for the extensive validation and systematics studies. The fast simulation is used for validating the technique with simulated signal and background events, and for evaluating systematic uncertainties. Neither the fast simulation nor the full simulation described earlier is used to determine or constrain any of the parameters that enter the likelihood fit to data from which we extract the B^- lifetime. Below we describe the fast simulation with its default settings. These form the basis of the validation studies presented later. How the default behavior is altered to estimate systematic uncertainties is discussed in Sec. IX.

In order to reproduce the data as well as possible with a relatively simple simulation, we generate many of the kinematic variables in each event based on distributions observed in data, in particular when generating background. The most important ones are summarized in Table III. For every event i we generate the B^- proper decay time, t_i, the reconstructed mass, m_i, the measured momentum, $\boldsymbol{P_i}$, and the D^0 meson proper decay time. The B^- mass is generated from the PDF described in Eq. (3) using the best fit parameters from the mass fit to the data sample. For signal events, the B^- and D^0 proper decay times are generated as exponentials using the 2008 world average values of the lifetimes, which are 1.637 ps and 0.41 ps for the B^- and D^0 mesons, respectively [24]. The generated proper decay times are smeared by a Gaussian of width 0.087 ps to simulate the detector resolution. The generation of the reconstructed proper B^- decay time in background events is based on the PDF described in Sec. VII A. Its parameters are determined from data, by fitting the lifetime distribution of the events in the upper mass sideband. The background D^0 proper decay time is taken from the D^0 decay time distribution observed in the same sample. The direction of the B^- momentum is generated uniformly in ϕ and η. As transverse quantities are used to determine the measured proper decay time in data, it is important to match the p_T distribution in the simulation, to that observed in data. The magnitude of the B momentum is generated such that, after the selection criteria are applied, the distribution of p_T of the remaining simulated signal events matches the p_T distribution observed in the background subtracted signal region. Similarly, we generate the magnitude of the momentum for background events so that after selection there is agreement between the p_T of simulated events and the upper sideband in data.

We calculate the remaining kinematic variables as follows. In the rest frame of the B^- particle, the magnitudes of the reconstructed D^0 and π_B momenta are defined by the generated mass of the B^- meson and the world average values for

TABLE II: The fit results on full detector simulated B decay samples. The table also gives the true input lifetime and the size of the sample after selection cuts had been applied.

Decay	Sample size	Input lifetime	Measured lifetime
$B^- \to D^0\pi^-$	75000	496 μm	493.3 $\pm$ 3.2 μm
$B^0 \to D^+\pi^-$	71000	464 μm	467.8 $\pm$ 2.8 μm
$B_s \to \phi\phi$	35000	438 μm	443 $\pm$5 μm
$B^0 \to K^+K^-$	75000	438 μm	441.5$\pm$ 2.9 μm

TABLE III: Kinematic parameters of the fast simulation and the parent distribution used for generation. Details are given in the text.

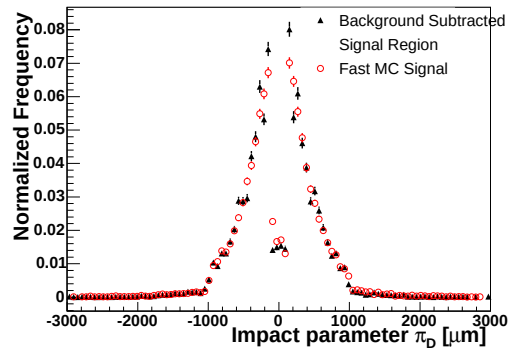
Randomly generated parameter			Parent distribution
t	reconstructed decay time of B^-	(signal)	$\frac{1}{\tau_B}e^{-t/\tau_B} \otimes \frac{1}{\sqrt{2\pi}\sigma_t}e^{-\frac{t^2}{2\sigma_t{}^2}}$
t_D	reconstructed decay time of D^0	(signal)	$\frac{1}{\tau_D}e^{-t_D/\tau_D}$
$\|P\|$	magnitude of B^- momentum	(signal)	background-subtracted data
m_B	B^- mass	(signal)	PDF given in Eq. (3), fitted to data
t	reconstructed decay time of B^-	(bkg)	PDF given in Eq. (16), fitted to data
t_D	reconstructed decay time of D^0	(bkg)	sideband data
$\|P\|$	magnitude of B^- momentum	(bkg)	sideband data
m_B	B^- mass	(bkg)	PDF given in Eq. (3), fitted to data
ϕ	azimuth angle of B^- momentum		uniform
η	pseudorapidity of B^-		uniform with $\|\eta\| < 1.5$
Δd_0^{L2}	$d_0^{\rm L2} - d_0^{\rm off}$		Gaussian, then round $d_0^{\rm L2}$ to nearest 10 μm.

the D^0 and π masses [24]. The reconstructed D^0 mass is kept fixed because in data, the mass-constrained vertex forces the reconstructed D^0 mass to the world average value. We pick a direction for the π_B momentum isotropic in the B^- rest frame; the D^0 momentum is in the opposite direction. These momenta are then transformed into the laboratory frame to calculate the simulated D^0 and π_B momenta. The equivalent procedure is carried out to calculate the π_D and K_D momenta in the laboratory frame. The B^- and D^0 decay vertex positions are calculated from the generated proper decay time and momentum; knowledge of these allows for track impact parameter calculation. These impact parameters are defined as the offline impact parameters.

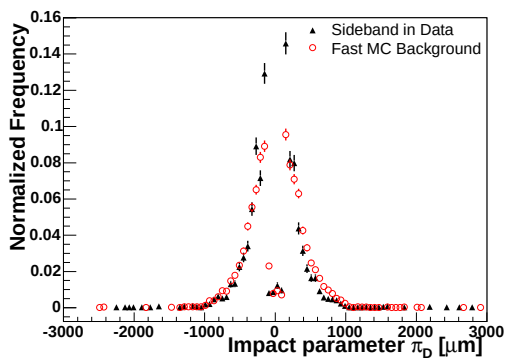
We simulate the SVT with a single track finding efficiency of $\varepsilon_s^{\rm sig} = 65\%$ for signal events and $\varepsilon_s^{\rm bkg} = 55\%$ for background events. The efficiency is different for signal and background because in general, we find in our data that background tracks have fewer hits in the silicon layers and hence a lower track finding efficiency. The values for the track finding efficiency we use for the simulation are approximately those found in data for tracks with $|d_0| < 1000\,\mu$m obtained from the simultaneous proper decay time, mass, and efficiency fit (the fit results for all parameters can be found in Appendix IV). Tracks with $|d_0| > 1000\,\mu$m do not participate in the trigger decision, and are treated in the fit as not found by the SVT, so there is no need to model the behavior of the SVT efficiency for tracks with $|d_0| > 1000\,\mu$m. For those tracks that are found, the SVT-measured impact parameter $d_0^{\rm L2}$ is obtained by adding a Gaussian-distributed random number to the offline d_0. The Gaussian is centered at 0 and has a width of 35 μm, which is consistent with the width observed in data (Fig. 3(b)). The result is then rounded to the nearest 10 μm, as in the real SVT. The difference between the L3 impact parameter $d_0^{\rm L3}$ and $d_0^{\rm off}$ is not simulated. Although the mean $d_0^{\rm L2}$ in data is shifted from zero, further tests detailed below confirm that the central value of the d_0 distribution does not affect the results. Therefore, these

differences between the fast simulation and the data will have a negligible effect on the interpretation of the results.

After all kinematic quantities have been obtained in the way described above, the selection criteria are applied to replicate the biases observed in data. All decay products are required to lie in the fiducial volume of the CDF II detector. The three different trigger requirements represent three different sets of selection criteria. Events are generated with each set of cuts separately and then combined in the fractions observed in the data. In data we observe very few events with tracks that have $|\eta| > 1.5$. Therefore events that have simulated tracks with $|\eta| > 1.5$ are removed from the sample. For background, prior to applying selection cuts, we further reject events so that the p_T spectrum of the candidate π_B after the cuts are applied matches that observed in the data upper sideband. This further rejection for background events effectively changes all kinematic distributions observed after the selection criteria are applied and forces the simulated background to have the characteristics of background observed in data. Overall there is broad agreement between the distributions of impact parameters, momenta, $\Delta\phi$ of track pairs, and η in the simulated and real data. As impact parameters are particularly important in this analysis, we show the π_D impact parameter distribution from the fast simulation and in real data in Fig. 5. Given the simple nature of the fast simulation, the agreement with data is remarkably good, although of course not perfect. Since the simulation is not used to determine any parameters in the final fit to data, but only to test the robustness of the method and to estimate systematic uncertainties, we do not rely on a perfect match between the simulation and the data, and the agreement we observe is sufficient.



(a)



(b)

FIG. 5: A comparison of the impact parameter distribution of the π_D track in data (triangle points) and fast simulation (circular points). The comparison between generated signal events and the background subtracted signal region in data is shown in (a) while (b) shows the comparison between generated background events and the upper sideband in data. All distributions are normalized to one event.

C. Validation of the method on signal events

We use the custom fast simulation for high-statistics tests of the signal PDF given in Eq. (12). We generate 1000 samples of 24,000 signal events each, similar to the yield observed in data. The proper decay time distribution for each sample has been sculpted by the same decay time dependent selection cuts as in real data, applied to the simulated data as described in the previous section. We maximize the likelihood function for signal events to extract a best fit lifetime for each sample. Fitting the resulting pull distribution with a Gaussian we find a mean of $\mu = -0.026 \pm 0.034$ and a standard deviation $\sigma = 1.027 \pm 0.024$. This demonstrates that $E_i(t, \varepsilon_s)$ is correctly calculated, and that the likelihood formed from the PDF in Eq. (12) can correct for the selection biases. It also shows that assigning the value of Δd_0 to tracks that did not have an SVT match, from the distribution of Δd_0 of tracks that did have an SVT match, does not cause any bias. In addition to this single test to validate the method itself we performed further tests to cross check our assumptions.

There are some differences between the value of the single track finding efficiency applied in the fast simulation and that observed in data. To test that the results were not sensitive to the default values of the efficiency chosen for the fast simulation, we varied the input efficiencies around the default values and saw no bias due to the value of input efficiency or the difference between signal

and background efficiencies. The fitted efficiency was always consistent with the input value. In Sec. III C we note that there have been three changes to ε_s over the course of the period of data taking used for this analysis. To determine whether in such a situation parameterizing the SVT track finding efficiency with one single value representing the average ε_s over these three data taking periods is sufficient, the value of ε_s in the simulation is altered so that events in each sample have been generated using three separate values of ε_s, in the proportions expected from each data period. These samples are fit using only one average ε_s parameter, which is allowed to float in the fit. The resulting pull distribution from all the fits is still described by a Gaussian of mean 0 μm, however the width is 1.19±0.03. This can be understood as follows: Each $E_i(t, \varepsilon_s)$ is a measure of the statistical power of each event [25]. By using an average ε_s, the statistical power of each event has been incorrectly assumed in the fit leading to an incorrect estimate of the statistical uncertainty. If instead, we allow for three floating efficiency parameters where each parameter is only sensitive to the events in one of the data taking periods, the resulting pull distribution once again has unit width. In the fit to data, we therefore use three parameters to describe ε_s, each floating in the fit, one for each data taking period.

In the default simulation the Δd_0 distribution is generated by a Gaussian with mean $\mu = 0\ \mu m$ and width $\sigma = 35\,\mu$m. Since the fit method takes all its information about Δd_0 from data, and makes no assumptions about the shape of the Δd_0 distribution, we expect it to perform equally well for any Δd_0 distribution, including asymmetric and biased distributions. We test this by generating data with two alternative models for Δd_0: For the first model we used a biased impact parameter resolution function described by a Gaussian with mean $\mu = 35\,\mu$m and width $\sigma = 35\,\mu$m. To truly stress-test the sensitivity of the method to the Δd_0 distribution, the second alternative model is a somewhat unrealistic, biased and asymmetric resolution function described by an exponential decay distribution with mean 35 μm, so that all $d_0^{\mathrm{L}2}$ are larger than the d_0^{off}. For both models we perform pull studies with the same sample size as observed in data, and observe no bias in the fitted lifetime. This confirms that the fit method is robust with respect to the shape and mean of the Δd_0 resolution function, and that the observed shift from zero in data of the mean of the distribution in Fig. 3(a) does not affect the fit result.

Other assumptions, including the dependence of the SVT efficiency on impact parameter, p_T, η and the effect of small differences in the Δd_0 resolution depending on impact parameter, are discussed as sources of systematic uncertainties in Sec. IX.

VII. THE COMBINED PDF FOR SIGNAL AND BACKGROUND EVENTS

In this section we derive the PDF for a sample containing signal and background events. We remind the reader that we use four measured observables in the fit; the measured proper decay time, t_i, the efficiency function, $E_i(t, \varepsilon_s)$, the mass, m_i, and the track-configuration observed, C_i. An unbinned maximum likelihood fit is used to determine the lifetime of the B meson and other parameters. Candidates in the data sample have passed the selection criteria, T, which means that we must consider the conditional probability that a candidate has a particular t_i , m_i, $E_i(t, \varepsilon_s)$, and C_i, given that the selection criteria have been satisfied. There are only two classes of candidates in the data sample: signal or background, therefore the likelihood function is defined as

$$\mathcal{L} = \prod_i \left[\mathcal{P}\left(s, t_i, m_i, C_i, E_i(t, \varepsilon_s); \tau|\mathrm{T}\right) + \mathcal{P}\left(b, t_i, m_i, C_i, E_i(t, \varepsilon_s^{\mathrm{bkg}})|\mathrm{T}\right)\right], \tag{13}$$

where the first term represents the likelihood for signal candidates and the second term is the likelihood for background candidates. For readability, the dependence on other fit parameters such as those related the parametrization of the mass distribution, is suppressed and only the dependence on the fit parameter τ is explicitly written.

The PDF for signal candidates can be factorized into the following form,

$$\begin{aligned}\mathcal{P}(s, t_i, m_i, C_i, E_i(t, \varepsilon_s); \tau|\mathrm{T}) = \\ \mathcal{P}(t_i; \tau|\mathrm{T}, E_i(t, \varepsilon_s), s) \\ \times \mathcal{P}(C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s) \\ \times \mathcal{P}(E_i(t, \varepsilon_s)|\mathrm{T}) \times \mathcal{P}(m_i|\mathrm{T}, s) \\ \times \mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})\end{aligned} \tag{14}$$

where a detailed derivation of this factorization is given in Appendix A. There is also an entirely analogous factorization for background candidates. The combined factor $\mathcal{P}(t_i; \tau|\mathrm{T}, E_i(t, \varepsilon_s), s)\mathcal{P}(C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s)$ describes the proper decay time distribution and includes the track configuration information which determines ε_s. Note that $\mathcal{P}(E_i(t, \varepsilon_s)|\mathrm{T})$ and similar expressions refer to the probability to find a

given efficiency *function* $E_i(t,\varepsilon_s)$. It does not refer to the function as evaluated for a given t or ε_s, but to the function as a whole. $\mathcal{P}(E_i(t,\varepsilon_s)|\mathrm{T})$ therefore does not depend on the value of t_i or ε_s. The factor $\mathcal{P}(E_i(t,\varepsilon_s)|\mathrm{T})$ is independent of τ and whether a candidate is signal or background. Hence, it can be ignored in the likelihood. The factors $\mathcal{P}(m_i|\mathrm{T},s)$ and $\mathcal{P}(m_i|\mathrm{T},b)$ (from the background part of the PDF) describe the mass distribution and are described earlier in Sec. IV B. The final factor $\mathcal{P}(s|E_i(t,\varepsilon_s),\mathrm{T})$ is the probability that a candidate is signal given its efficiency function. Each factor of the likelihood is normalised to one candidate.

A. The parametrization of the proper decay time PDF

This section considers the proper decay time term in the PDF in Eq. (14), the analogous terms for the background candidates and describes the parametrizations used for the fit. For the signal component a physics model is used, for the background contribution it is sufficient to provide an empirical description of the data. The first two factors on the right hand side of Eq. (14) are identical to the left hand side of Eq. (12), and this is the PDF used to fit the proper decay time and single track finding efficiency for signal candidates. Three different values of $\varepsilon_s^{\mathrm{sig}}$ are fit, one for each time period as described in Sec. VI.

For background candidates

$$\mathcal{P}(t_i;\tau|\mathrm{T},E_i(t,\varepsilon_s),b)\mathcal{P}(C_i|\mathrm{T},E_i(t,\varepsilon_s),t_i,b) = \frac{{\varepsilon_s^{\mathrm{bkg}}}^n(1-\varepsilon_s^{\mathrm{bkg}})^{n-r}y(t_i)}{\sum\limits_{\substack{k=all\\intervals}} H_k(\varepsilon_s^{\mathrm{bkg}})\int\limits_{t_{\mathrm{min\ k}}}^{t_{\mathrm{max\ k}}} y(t)dt}, \tag{15}$$

where, similarly to signal, there are three values of $\varepsilon_s^{\mathrm{bkg}}$ to fit. The function $y(t)$ can be determined empirically from the data. Simple forms of $y(t)$, such as a sum of exponentials convoluted by a Gaussian, were found to provide an unsatisfactory description of the data. Therefore, the function $y(t)$ is empirically determined using an interpolation of exponentials given by

$$y(t) = e^{a_j+\left(\frac{a_{j+1}-a_j}{t_{j+1}-t_j}\right)(t-t_j)} \text{ for } t_j \le t \le t_{j+1}. \tag{16}$$

We use ten fit points (t_j), which are spaced more closely at lower t where the proper decay time distribution of background candidates is concentrated. The values of the corresponding a_j are determined alongside the other fit parameters in the unbinned maximum likelihood fit. This parametrization was tested on data from the upper sideband to ensure that it is a good model for the data. The tests on the upper sideband were only used to distinguish the performance of different parametrizations, and no fit parameters are fixed from this test.

B. The complication in combining the signal PDF and the background PDF when using a candidate-by-candidate efficiency function

Combining the signal and the background PDF while using a candidate-by-candidate efficiency function introduces a significant complication into the analysis. The rest of the section describes this problem, and its solution, in detail. As discussed in [16], when a candidate-by-candidate quantity enters a fit with a signal and background component, the PDF for this quantity needs to be included in the fit. In our PDF this effect is taken into account by a term that describes the candidate-by-candidate signal probability depending on the $E_i(t,\varepsilon_s)$. So, instead of an overall signal fraction $P(s)$, there is a signal weighting for each candidate which depends on the efficiency function. This is described by the factor $\mathcal{P}(s|E_i(t,\varepsilon_s),\mathrm{T})$ where the corresponding term for background is simply $\mathcal{P}(b|E_i(t,\varepsilon_s),\mathrm{T}) = 1-\mathcal{P}(s|E_i(t,\varepsilon_s),\mathrm{T})$. Alternative ways of factorizing the PDF would lead to different ways to take this effect into account, but regardless of the choice of factorization, the underlying need to include a PDF for the efficiency function remains.

A simplification to an overall signal fraction, $\mathcal{P}(s)$, is only possible in the case where the efficiency function distributions are the same for signal and background. Figure 6 shows the mean efficiency function, $\overline{E_i(t,\varepsilon_s)}$ for candidates in the upper sideband (background) and the background subtracted signal region. The mean is determined simply by summing all efficiency functions in a sample and dividing by the number of candidates. The two $\overline{E_i(t,\varepsilon_s)}$ are clearly different which shows that the distribution of efficiency functions in signal and background must be different. We can estimate the bias on the lifetime measurement if we were to ignore the differences in the efficiency function by simplifying $\mathcal{P}(s|E_i(t,\varepsilon_s),\mathrm{T})$ to $\mathcal{P}(s)$. Using the custom fast simulation described in Sec. VI B, we find a bias of approximately -0.018 ± 0.001 ps. Any advantage gained in precision by using a simulation independent method would be negated by a bias of this size. Therefore, a successful simulation-

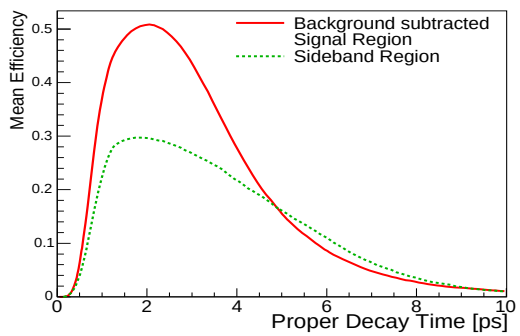


FIG. 6: The mean $E(t, \varepsilon_s)$ function for signal and background candidate. Signal (solid line) and background (dashed line) candidates have different $E_i(t, \varepsilon_s)$.

independent method of correcting the trigger bias must also include a proper description of the term $\mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$.

C. Calculating the term $\mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$

1. *Overview*

To correctly represent the PDF in the fit we require a parametrization of the signal fraction dependent on the candidate-by-candidate efficiency function, $E_i(t, \varepsilon_s)$. It is, however, difficult to parametrize a distribution of functions to derive a signal probability as a function of each individual function. The problem is simplified if we represent $E_i(t, \varepsilon_s)$ by a number, x_i, as it is considerably easier to parametrize the distribution of the scalar variable x rather than a distribution of functions, i.e., we aim to find a variable x such that we can replace $\mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$ by $\mathcal{P}(s|x, \mathrm{T}) \approx \mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$. For this approach to succeed, x must be chosen in such a way that the loss of information regarding the signal probability of $E_i(t, \varepsilon_s)$ is minimized as we transform from $E_i(t, \varepsilon_s)$ to x. Note that the transformation of $E_i(t, \varepsilon_s)$ to x is only used for determining the signal probability of each candidate. The proper decay time probabilities are unchanged and continue to use $E_i(t, \varepsilon_s)$ as the trigger bias cannot be corrected without the full description. To summarize, the parametrization of the term $\mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$ involves two steps:

- Transforming the efficiency function $E_i(t, \varepsilon_s)$ into a representative number x.
- Describing the signal fraction as a function of x, $P(s|x, \mathrm{T})$ with a suitable function whose parameters will be determined in the fit.

These are discussed below.

2. *Representing the Efficiency function by a Scalar*

In order to translate the efficiency function $E_i(t, \varepsilon_s)$ into a scalar variable, we make use of the Fisher Linear Discriminants method [30]. This method transforms a vector of variables into a single scalar variable. Therefore, we represent each $E_i(t, \varepsilon_s)$ as a vector that contains all the relevant information about $E_i(t, \varepsilon_s)$ and then use the Fisher discriminant method to translate this vector into a number, the Fisher scalar, x. Note that we do not use the Fisher discriminant method to select candidates. The scalar resulting from the Fisher discriminant method is optimized for distinguishing signal from background, and therefore fulfills the requirement of minimizing the loss of information about the signal probability as we translate $E_i(t, \varepsilon_s)$ to x_i, so that $P(s|x_i, \mathrm{T}) \approx P(s|E_i(t, \varepsilon_s), \mathrm{T})$ to a very good approximation. How good this approximation is, is quantified below. Here, we summarize the method rather briefly. Further details can be found in Appendix B.

3. *Finding the Fisher discriminant in a simulation independent way*

The Fisher scalar is the scalar product of a vector $\boldsymbol{v}_i$ of parameters, which in our case represents one candidate's efficiency function, with another fixed vector $\boldsymbol{w}$ so $x_i = \boldsymbol{w} \cdot \boldsymbol{v}_i$ and is a single variable for each (i^{th}) candidate. How we represent $E_i(t, \varepsilon_s)$ by a vector is described in Appendix II. It is done in a way so that the value of $\varepsilon_s^{\mathrm{sig}}$ or $\varepsilon_s^{\mathrm{bkg}}$ is not required and hence can be done before the fit determines the values for the efficiencies. The Fisher Linear Discriminant method provides a way to a vector $\boldsymbol{w}$ such that it maximizes the separation of signal and background candidates in the variable x.

In typical uses of the Fisher Linear Discriminant method, the calculation of $\boldsymbol{w}$ requires not only the knowledge of all the $\boldsymbol{v}_i$ but also of $\boldsymbol{v}_s$ and $\boldsymbol{v}_b$ which are the mean $\boldsymbol{v}_i$ for signal and background candidates, respectively. Traditionally, $\boldsymbol{v}_s$ and $\boldsymbol{v}_b$ are determined from independent training samples such as detailed Monte Carlo data. Since this analysis uses no input from simulation we use the data itself to calculate $\boldsymbol{v}_s$ and $\boldsymbol{v}_b$. For this measurement, we use candidates in the upper sideband to determine $\boldsymbol{v}_b$. We

perform a background subtraction on candidates with $5.25 < m_B < 5.32$ GeV/c^2 to determine $\boldsymbol{v}_s$.

4. Testing the assumption that $P(s|x_i, T) \approx P(s|E_i(t, \varepsilon_s), T)$

Before proceeding further, it is important to test the assumption that the Fisher scalar variable x_i is representative of $E_i(t, \varepsilon_s)$. We use a custom fast simulation and fit the lifetime of the 1000 independent samples of signal and background candidates, using the Fisher scalar x_i to determine a signal probability per candidate. As these are simulated data, where the truth information is known, $\mathcal{P}(s|x_i, \mathrm{T})$ can be calculated by finely binning the distribution of x into 100 bins and simply counting the number of signal and background candidates in any particular bin of the variable x. So, for each x_i, we determine $\mathcal{P}(s|x_i, \mathrm{T})$ by reading its value off a histogram generated from the truth information. Note that this is not the way we calculate $\mathcal{P}(s|x_i, \mathrm{T})$ in our fit to data, where we do not have access to truth information, or in fact any simulation information. The purpose of this step is solely to quantify how the assumption that $P(s|E_i(t, \varepsilon_s), T) \approx P(s|x_i, T)$ affects the fit result, in a way that is independent of any particular parametrization of $P(s|x_i, T)$, which is investigated separately. We find that the mean lifetime shift in those 1000 fits is only 0.0013 ps, which is significantly smaller than -0.018 ps found when the distribution of efficiency functions is ignored. This demonstrates that the variable x_i is a satisfactory substitute for $E_i(t, \varepsilon_s)$ for the purposes of calculating the probability that a candidate is signal given its efficiency function. This mean shift of 0.0013 ps is small in comparison to the statistical uncertainty from the data sample size and is taken as a systematic uncertainty due to assuming the scalar variable is entirely equivalent to using the full efficiency function.

5. Parametrizing the signal fraction as a function of the Fisher scalar variable

In order to apply this method to real data in a simulation-independent way, we need to find a function that parametrizes $P(s|x_i, \mathrm{T})$, whose parameters can then be determined in the fit to data. We use Lagrange interpolating polynomials as they provide a very general parametrization that makes minimal assumptions about the shape of the distribution to be fitted. This parametrization has, as its parameters, the signal fractions p_j at certain discrete values of the Fisher discriminant x_j, so $P(s|x_j, \mathrm{T}) = p_j$. The value of $P(s|x, \mathrm{T})$ for general x is calculated by a smooth interpolation between those points. The p_j are determined in the fit.

Our default choice for the x_j is the following: We divide the x axis into $N = 15$ equal bins. Since the number of candidates at the edges of the distribution is small, we merge the first two bins, as well as the last two bins. We place our x_j at the center of each resulting bins. This results in 13 fit parameters p_j, representing the signal fraction at the 13 x_j. We tested the robustness of this choice by trying out different numbers of bins N, and found that there is negligible difference in performance for any value of N from 10 to 20.

This parametrization is tested using the fast simulation. Figure 7 shows the projection of the fitted Lagrange interpolating polynomial, $f(x)$, where the truth information has been superimposed for one sample of simulated data. In contrast to the test in Sec. VII C 2 where we evaluated the assumption $P(s|x_i, \mathrm{T}) \approx P(s|E_i(t, \varepsilon_s), \mathrm{T})$, here the fit is performed in the same way as in our final fit to real data: at no point is truth information or any external simulation input used in the fit, and the p_j parameters of $P(s|x_i, \mathrm{T})$ are determined in the fit at the same time as all other fit parameters, such as the lifetime or ε_s. The projection of $P(s|x, \mathrm{T})$ obtained in this fit matches closely the histogram obtained from truth information, giving us confidence that this parametrization provides a good description. We tested this parametrization using 1000 simulation samples and observed no shift in the mean residual position relative to the small shift resulting from the assumption $P(s|x_i, \mathrm{T}) \approx P(s|E_i(t, \varepsilon_s), \mathrm{T})$ observed in Sec. VII C 2. The lifetime pull distribution is described by a Gaussian with mean 0.039 ± 0.036 and width 1.097 ± 0.029.

6. Summary: The full signal & background PDF with the factor $\mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$

In summary, we find that the PDF in Eq. (13) with the factor $P(s|E_i(t, \varepsilon_s), \mathrm{T})$ parametrized as described in this section successfully corrects for the selection bias in data samples where both a signal and a background component is present. The 0.0013 ps residual is taken as a systematic uncertainty due to the method of parametrizing the term $\mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$. The width of the pull indicates that the method underestimates the statistical uncertainty by 10 ± 3 %. To be conservative we increase the statistical uncertainty of the fit to data accordingly.

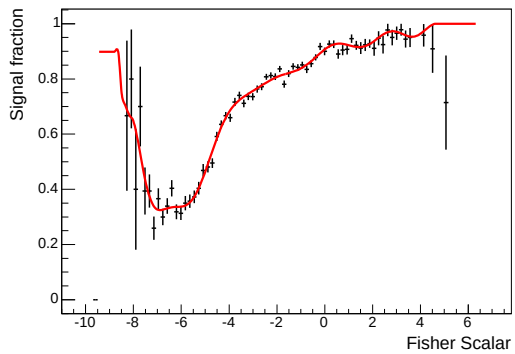


FIG. 7: The data points show the signal fraction as a function of Fisher scalar for a sample of simulated data. The line shows the projection of the Lagrange interpolating polynomial determined by the simultaneous fit to proper decay time, signal fraction and other parameters.

VIII. FIT RESULTS

This section describes the fit to data selected by applying the selection criteria listed in Sec. IV A. An initial mass fit is performed as described in Sec. IV B with seven free parameters, where the best fit results are given in Appendix D. The results of the mass fit are used to perform the background subtraction to calculate the $\boldsymbol{v}_s$ for signal events required for the Fisher Discriminant Analysis.

The lifetime is determined in a second fit. The probability density function used in this unbinned maximum likelihood fit is given in Eq. (13). The parameters that determine the mass shapes for signal and background are held fixed at the values determined in the initial mass fit. However, the signal fraction is not taken from the mass fit, as this is now redefined in terms on the Fisher scalar variable. In total there are 30 free parameters in the lifetime fit. These are the following: one for the signal lifetime, ten to describe the background proper decay time distribution as described in Sec. VII A, 13 parameters to determine the signal fraction as a function of the Fisher scalar, $f(x)$, defined in Sec. VII C 5 and six parameters to describe the single track finding efficiency as described in Sec. VII A.

The proper decay time fit projection for all events in the fit is shown in Fig. 8. The function $f(x)$ determined by the fit is shown in Fig. 9, and the distribution of the variable x itself is shown in Fig. 10. To assess how well $f(x)$ determines the signal fraction, the data sample with $-7 < x < 2$ is divided into nine bins. A mass fit is performed separately for the events in each bin to obtain an independent measure of the signal fraction in that bin. For x outside the range $-7 < x < 2$, there is insufficient data to perform a mass fit. The signal fractions, as determined by the series of mass fits, are overlaid on the function $f(x)$ in Fig. 9 and there is good agreement between the two determinations of signal fraction.

The fit result for the B^- lifetime is $\tau(B^-) = 1.663 \pm 0.023(\text{stat})$ ps, where the uncertainty has already been scaled by the factor 1.1 as discussed in Sec. VII C 5. The fit results for all other parameters can be found in Appendix D. The lifetime is only weakly correlated to the other fit parameters; the correlation coefficient between the lifetime and any other fit parameter is always less than 10%. The statistical uncertainty on $\tau(B^-)$ is about twice as large as one would naively expect from dividing the fit result by the square-root of the number of signal events, $\sigma_{\tau\,\text{naive}} \approx \tau/\sqrt{N_{\text{sg}}} = 0.011$ ps, which normally gives a reasonable estimate for data with good proper decay time resolution and small background contamination as we have here. As shown in Ref. [25], the cause for the increased uncertainty is the trigger bias, specifically the *upper* impact parameter cut in the trigger, leading to a significantly reduced statistical precision per event. The size of the effect is consistent with that calculated in Ref. [25].

IX. SYSTEMATIC UNCERTAINTIES

In this section we evaluate the systematic uncertainty on our measurement from a variety of possible causes. The two dominant uncertainties are due to the dependence of the single track finding efficiency of the SVT on impact parameter (Sec. IX A) and the correlation between the measured t_i and m_i that we observe in background data from the upper sideband.

We evaluate each uncertainty as follows: for each source of uncertainty, 1000 samples of simulated data are generated using the fast simulation. Each sample contains approximately the same number of signal and background candidates as found in data. The samples are generated using a non-standard configuration that simulates the effect under consideration; we then extract the B^- lifetime from each sample in the same way as we do for data, using the standard PDF described in Sec. VII. For each source of systematic uncertainty, the mean residual, (fitted lifetime − input lifetime), averaged over the 1000 samples, is taken as the systematic uncertainty. The statistical uncertainty on the mean residual from 1000 generated samples of simulated data is approximately 0.0007 ps and sys-

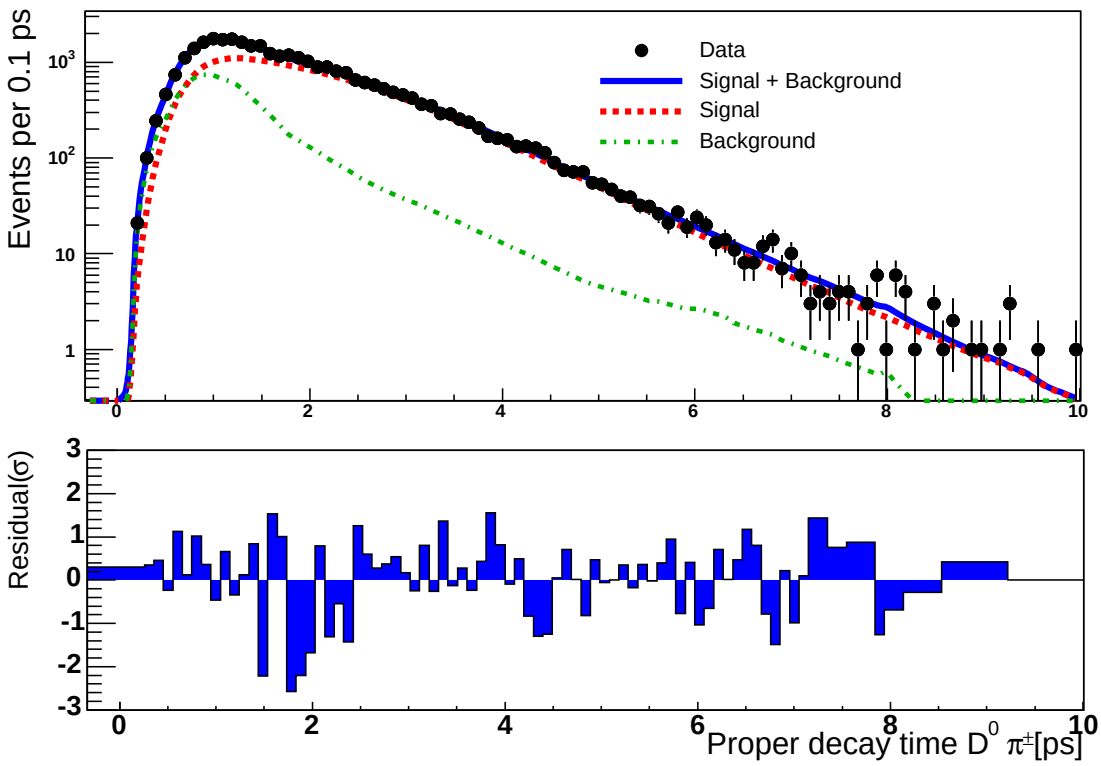


FIG. 8: This figure shows the projection of the lifetime fit onto the data. The signal and background components are shown separately (dotted lines) and in addition (solid line). The points are data. The lower plot shows the residual/error for each bin.

tematic uncertainties of this size or smaller are deemed negligible.

A. The dependence of the single track finding efficiency on impact parameter

The leading source of systematic uncertainty is the parametrization of the L2 single track finding efficiency as a function of track impact parameter. As described in Sec. V, we assume in the fit that $\varepsilon_s(|d_0|)$ is constant for $|d_0| < 1000\,\mu$m. Figure 2 in Sec. V B 3 shows the efficiency as a function of $|d_0|$ in data, and indicates that ε_s starts dropping slightly before $|d_0| = 1000\,\mu$m. To obtain a model for the track finding efficiency to be used in the simulation, we fit the SVT single track finding efficiency as a function of $|d_0|$ found in data, using the function $\varepsilon(|d_0|) = p_0 \times G\left(\frac{|d_0|-p_1}{p_2}\right)$, where p_0, p_1, and p_2 are free parameters and $G(x)$ is the complementary error function defined as $G(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty \exp(-t^2)dt$. This fit results in one particular determination of the single track finding efficiency shape. We create other SVT single-track efficiency functions, consistent with the data, by varying p_0, p_1 and p_2 by $\pm\sigma$ of their fitted values. We consider further the three where we expect the largest effects, which are the ones where difference in efficiency between $|d_0| = 0$ and $|d_0| = 1000\,\mu$m is largest. These three SVT single track efficiency functions, one of which is the original fit result itself, are represented by the three lines in Fig. 2. The different single track efficiency functions are implemented in the simulation by assigning SVT matches with the probability determined by the given function. For each of the three functions considered, a set of 1000 simulated samples is generated and fit with the standard PDF that assumes a flat SVT single track efficiency for $|d_0| < 1000\,\mu$m. The mean lifetime residual from the fits to these samples varies from -0.0060 to -0.010 ps, depending on the values of p_0, p_1 and p_2 used. To be conservative we assign a 0.010 ps systematic uncertainty due to assuming that $\varepsilon_s(d_0)$ is constant for tracks with impact parameter less than $1000\,\mu$m.

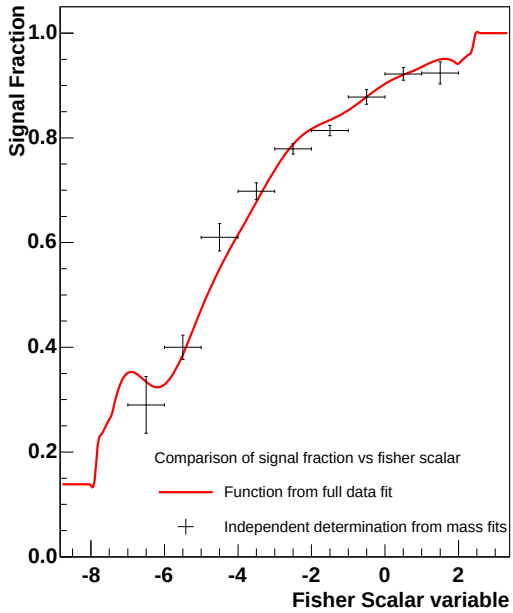


FIG. 9: Projection of the signal fraction as a function of the Fisher scalar, $f(x)$, determined from fit (line). The data points are the signal fraction determined from mass fits using events only lying in that particular bin of x.

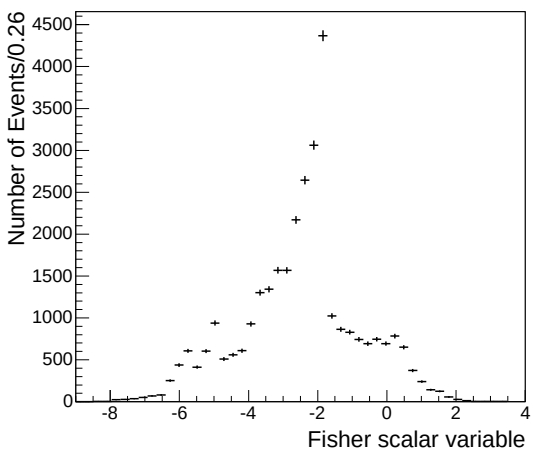


FIG. 10: The distribution of the Fisher Scalar variable x in data.

B. Single track finding efficiency dependence on p_T and η

The fit also neglects the dependence of ε_s on p_T and η for tracks that pass the trigger criteria, i.e., with $p_T > 2$ GeV/c. Figure 11 shows $\varepsilon_s(p_T)$ in data; the line through the data represents a fit using third order polynomial. The efficiency is obtained in a similar manner to Figure 2, where the third track, in the sub-sample of candidates where the other two tracks are sufficient to pass the trigger, is used to determine the efficiency. The p_T dependence is incorporated into the fast simulation, by assigning SVT matches based on the probability given by the polynomial function and we determine a systematic uncertainty of 0.006 ps. Similarly, we evaluate the effect of the dependence of the SVT single

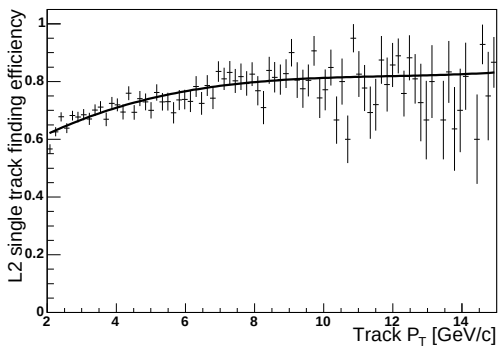


FIG. 11: SVT single track finding efficiency as a function of track transverse momentum. The fit function (line) is a third order polynomial to the data (points)

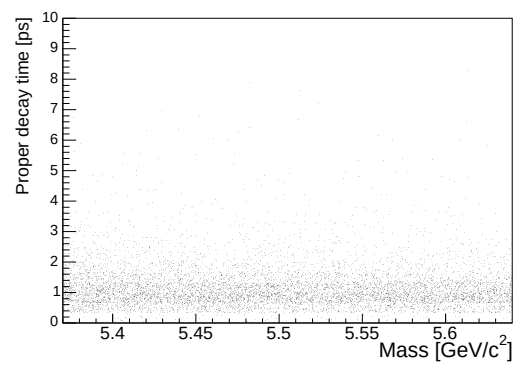
track finding efficiency on the track's pseudorapidity which results in a systematic uncertainty of 0.001 ps. The dependence on track p_T and η are not large sources of uncertainty since they are not directly related to the proper decay time while the impact parameter is.

C. Dependence on the impact parameter resolution shape

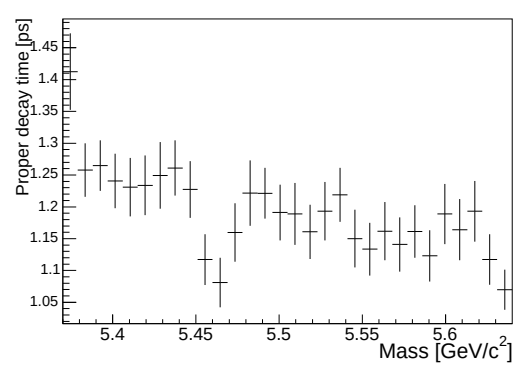
In the fit, we assume that the impact parameter resolution between the offline and online algorithms remains constant as a function of impact parameter. As discussed in Sec. VI B, it has been shown that the technique of sliding the decay vertex is insensitive to the actual shape of the resolution as long as the shape remains constant. In the data we do, however, see subtle differences, at the level of a few microns, in the mean and width of the resolution as a function of impact parameter (Fig. 3, Sec. V). To test this effect we incorporate such differences, as found in data, into the fast simulation. The bias observed due to this is 0.002 ps.

D. Dependence of background observables on mass

In data we observe a correlation between the measured t_i and m_i for background candidates in the upper sideband, which is shown in the scatter and profile plot in Fig. 12. We assume that this correlation is described well by a linear relationship and determine that the mean reconstructed proper decay time of background varies by approximately 40 μm over a mass range of 0.27 GeV/c^2 which is the mass range used in the lifetime fit. However, the derivation of the PDF assumes that there is no dependence on mass for the proper decay time of a background candidate. To test the effect of neglecting this correlation in the PDF we extrapolate the same linear correlation for background candidates underneath the peak as observed in the sideband. We generate simulated data where background candidates are rejected in such a way as to introduce a correlation between the mass and proper decay time of the candidate, similar to that observed in data. We determine the systematic uncertainty of 0.0083 ps using samples of fast simulation signal and background candidates. This is one of the leading sources of systematic uncertainty. It could be reduced in future measurements by defining a proper decay time parametrization for background that includes dependence on the mass. One possible way to do this would be to assume that $\mathcal{P}(t_i|\mathrm{T}, E_i(t,\varepsilon_s), b, m_i) = \mathcal{P}(t_i^\dagger|\mathrm{T}, E_i(t,\varepsilon_s), b)$ where $t_i^\dagger = t_i + \beta(m_i - m_0)$ where m_0 is a central mass value, and care is taken to ensure proper normalization.



(a)



(b)

FIG. 12: A scatter and profile plot shows the correlation between the mass and proper decay time of candidates in the upper sideband.

In the derivation of the PDF, we also assumed

that there was no relation between $E_i(t, \varepsilon_s)$ and m_i for background candidates. Candidates in the upper sideband are used to calculate $\boldsymbol{v}_b$ which is necessary to determine the Fisher discriminant. We assume that the calculated $\boldsymbol{v}_b$ is representative of all background candidates. To test the sensitivity of the lifetime result to the particular background sample, we repeat the lifetime fit to data where candidates with reconstructed mass between $5.5 < M_B < 5.7$ GeV/c^2 are used as the background sample. There is no change in the fitted lifetime for data which demonstrates that there is no significant relation between $E_i(t, \varepsilon_s)$ and m_i for background candidates.

E. Background proper decay time parametrization

To test the reliability of the $y(t)$ parametrization described in Sec. VII A, we seek an alternate parametrization of the data. We use the sum of two exponentials convoluted with the detector resolution. This parametrization of the background is not used in the main fit, as the quality of fit is poor. Nonetheless, we can generate simulated data where the background proper decay times are generated using the sum of two exponential functions with mean lifetimes of 0.787 ps and 0.0282 ps in the ratio 1:7.3 as found from this fit of the sideband. This results in a background proper decay length distribution that has similar characteristics to the distribution observed in the upper sideband. We fit these simulated data samples with the standard PDF. The mean lifetime residual is 0.0027 ps and we take this as an estimate of the systematic uncertainty due to the background proper decay time parametrization.

F. Silicon Alignment

To determine the uncertainty due to a possible misalignment of the silicon (SVX-II) detector we consider radial shifts in the silicon layers towards and away from the beam pipe by 50 μm as in other lifetime measurements [31] at the CDF experiment. The shifts in the silicon layers change the measured hit positions of the tracks. To first order, the mis-measurement in track impact parameters are related to a 50 μm shift in the silicon layers by the relation $\delta(d_0) = 50 \cdot \sin(\alpha)$, where the angle α is that between the track and perpendicular to the silicon layer in the transverse plane. We recalculate the measured impact parameters in the fast simulation due to the misalignment model. The proper decay time of the candidate is recalculated using the shifted impact parameter values. Fitting 1000 samples generated in this way we find a systematic uncertainty in lifetime measurements due to silicon layer misalignment of 0.0013 ps.

G. Detector Resolution Model

In the fit to data, we describe the detector time-resolution with a Gaussian of width 0.0087 ps. We estimate the systematic uncertainty by generating data sets with an alternative resolution function and fitting it using the standard PDF. The alternative resolution function is described by a sum of three Gaussians with widths 0.0067, 0.0124, 0.0249 ps and relative fractions 1:0.92:0.04. This resolution function derives from a study of prompt D mesons combined with an extra track from the primary vertex. We test the effect of this alternate resolution using the fast simulation. From the 1000 samples of fast simulated data, we find that the mean residual is 0.0010 ps and we take this as the systematic uncertainty.

H. Signal composition

We also consider the contamination of the signal peak by the decay $B^- \to D^0 K^-$. This decay can appear in the sample if the kaon track is reconstructed as a pion and the resulting decay passes the selection criteria. Although this is the decay of a charged B meson, the proper decay time distribution of this decay mode will be altered as the mass has been miscalculated. We use the fast simulation to estimate the fraction of $B^- \to D^0 K^-$ candidates that pass the lower mass cut. This information, in conjunction with the relative branching fractions of the $B^- \to D^0 \pi^-$ and $B^- \to D^0 K^-$ decay modes [24], results in the estimate that 3% of the candidates in the signal peak are actually misreconstructed $B^- \to D^0 K^-$ decays. This fraction is introduced into the fast simulation and the effect on the best fit lifetime is negligible.

I. Summary

A list of systematic uncertainties is given in Table IV. We combine the uncertainties in quadrature to find a total systematic uncertainty of 0.015 ps. This is to be compared to a statistical uncertainty of 0.023 ps. The leading sources of systematic uncertainty are related to the details of the SVT single track finding efficiency, and the correlation in background between the

TABLE IV: Summary of systematic uncertainties.

Source of systematic uncertainty	Uncertainty(ps)
Track finding efficiency dependence on d_0	0.0103
Track finding efficiency dependence on p_T	0.0060
Variation in impact parameter resolution	0.0020
Track finding efficiency dependence on η	0.0010
Mass-proper decay time correlation in background	0.0083
Background proper decay time parametrization	0.0027
Silicon alignment	0.0013
Transformation of $E_i(t)$ to scalar variable	0.0013
Detector resolution model	0.0010
Signal composition	negligible
Total systematic uncertainty	0.015

reconstructed proper decay time and mass. Neither of them are irreducible, and should the systematic uncertainty become a limiting factor in future measurements, it should be possible to improve it significantly. A more detailed description of the SVT track finding efficiency, which can be obtained from data, can be incorporated into the fit to reduce the leading systematic error. Similarly, the correlation between the mass and proper decay time in background candidates can be incorporated into a future version of this technique to reduce the second largest contribution.

X. RESULT AND CONCLUSION

We introduce a simulation-independent method for measuring lifetimes in event samples where the selection criteria bias the proper decay time distribution. We apply it to measure the B^- lifetime in data collected by the hadronic B trigger at CDF, which selects events with displaced tracks and thus biases the measured proper decay time distribution.

In previous analyses of data biased in this way, the trigger bias has usually been corrected for with an efficiency function that has been obtained from Monte Carlo simulation. This simulation dependence can be a significant source of systematic uncertainty. A recent example is the measurement of the Λ_b lifetime in the hadronic decay channel $\Lambda_b \to \Lambda_c \pi$ at CDF, which found $\tau(\Lambda_b) = 1.401 \pm 0.046$ (stat) ± 0.035 (syst) ps [15]. The systematic uncertainty in this measurement is almost entirely due to the simulation dependence. While adequate for the available statistics, in future high-statistics measurements, this could limit the experimental precision.

The technique introduced here removes the simulation dependence by replacing the global efficiency function with candidate-by-candidate efficiency functions that can be calculated analytically from the event data, without recourse to simulation. We test the method extensively with simulated data, and finally apply it to measure the lifetime of the B^- meson, $\tau(B^-)$, using 23900 ± 200 $B^- \to D^0\pi^-$ candidates, where $D^0 \to K^-\pi^+$, collected by CDF's hadronic B trigger in $1\,\mathrm{fb}^{-1}$ of data. We extract $\tau(B^-)$ from the data without input from simulation. We measure $\tau(B^-) = 1.663 \pm 0.023$ (stat) ± 0.015 (syst) ps. This result is in good agreement with the world average of 1.638 ± 0.011 ps [24]. This technique generalizes easily to other decay channels, as we have demonstrated in Sec. VI A. It can be applied to any situation where the trigger or other selection criteria bias the proper decay time distribution of the reconstructed data.

XI. ACKNOWLEDGMENTS

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the World Class University Program,

the National Research Foundation of Korea; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

APPENDIX A: FACTORIZING THE PDF

This appendix details the factorization of the PDF term $\mathcal{P}(s, t_i, m_i, C_i, E_i(t, \varepsilon_s); \tau|\mathrm{T})$ which describes the probability that we observe an event with its particular values of t_i, m_i, $E_i(t, \varepsilon_s)$, and C_i. Although there are a number of ways to factorize the expression in Eq. (13) we aim to find a final form that includes the factor $\mathcal{P}(t_i; \tau|\mathrm{T}, E_i(t, \varepsilon_s), s) \times \mathcal{P}(C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s)$ as a parametrization for this factor is well understood and given in Eq. (12). We make use of the following relation

$$\mathcal{P}(A, B) = \mathcal{P}(A)\mathcal{P}(B|A) = \mathcal{P}(B)\mathcal{P}(A|B). \quad \text{(A1)}$$

We only explicitly write the dependence of the PDF on parameters where such relations are used. Using Eq. (A1), $\mathcal{P}(s, t_i, m_i, C_i, E_i(t, \varepsilon_s); \tau|\mathrm{T})$ can be split into two factors:

$$\begin{aligned} &\mathcal{P}(s, t_i, m_i, C_i, E_i(t, \varepsilon_s); \tau|\mathrm{T}) = \\ &\quad \mathcal{P}(s, t_i, E_i(t, \varepsilon_s); \tau|\mathrm{T}) \\ &\quad \times \mathcal{P}(m_i, C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s; \tau). \end{aligned} \quad \text{(A2)}$$

The last factor, $\mathcal{P}(m_i, C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s; \tau)$ is concerned with the probability of observing a particular mass and track configuration. This can be factorized further, again using Eq. (A1):

$$\begin{aligned} &\mathcal{P}(m_i, C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s; \tau) = \\ &\quad \mathcal{P}(m_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s; \tau) \\ &\quad \times \mathcal{P}(C_i|m_i, \mathrm{T}, E_i(t, \varepsilon_s), t_i, s; \tau). \end{aligned} \quad \text{(A3)}$$

The measured mass is independent of the mean lifetime and the efficiency function and we make the assumption that it is independent of the measured proper decay time. These simplifications result in

$$\mathcal{P}(m_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s; \tau) = \mathcal{P}(m_i|\mathrm{T}, s). \quad \text{(A4)}$$

The track configuration is independent of the mean lifetime and the mass and therefore we can make the simplification $\mathcal{P}(C_i|m_i, \mathrm{T}, E_i(t, \varepsilon_s), t_i, s; \tau) = \mathcal{P}(C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s)$. Substituting this and Eq. (A4) into Eq. (A3) leads to

$$\begin{aligned} &\mathcal{P}(m_i, C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s; \tau) = \\ &\quad \mathcal{P}(m_i|\mathrm{T}, s)\, \mathcal{P}(C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s). \end{aligned} \quad \text{(A5)}$$

The other factor in Eq. (A2) is factorized further, using Eq. (A1) once more:

$$\begin{aligned} &\mathcal{P}(s, t_i, E_i(t, \varepsilon_s); \tau|\mathrm{T}) = \\ &\quad \mathcal{P}(t_i; \tau|\mathrm{T}, E_i(t, \varepsilon_s), s)\, \mathcal{P}(s, E_i(t, \varepsilon_s)|\mathrm{T}). \end{aligned} \quad \text{(A6)}$$

The first factor on the right hand side is the one we desire as it corresponds to that in Eq. (12). Applying Eq. (A1) one more time we find $\mathcal{P}(s, E_i(t, \varepsilon_s)|\mathrm{T}) = \mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})\mathcal{P}(E_i(t, \varepsilon_s)|\mathrm{T})$. Therefore

$$\begin{aligned} &\mathcal{P}(s, t_i, E_i(t, \varepsilon_s); \tau|\mathrm{T}) = \\ &\quad \mathcal{P}(t_i; \tau|\mathrm{T}, E_i(t, \varepsilon_s), s) \\ &\quad \times \mathcal{P}(E_i(t, \varepsilon_s)|\mathrm{T})\, \mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T}). \end{aligned} \quad \text{(A7)}$$

Substitution of Eq. (A7) and Eq. (A5) into Eq. (A2) leads to

$$\begin{aligned} &\mathcal{P}\,(s, t_i, m_i, C_i, E_i(t, \varepsilon_s); \tau|\mathrm{T}) = \\ &\quad \mathcal{P}(t_i; \tau|\mathrm{T}, E_i(t, \varepsilon_s), s) \\ &\quad \times \mathcal{P}(C_i|\mathrm{T}, E_i(t, \varepsilon_s), t_i, s)\, \mathcal{P}(E_i(t, \varepsilon_s)|\mathrm{T}) \\ &\quad \times \mathcal{P}(m_i|\mathrm{T}, s)\, \mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T}), \end{aligned} \quad \text{(A8)}$$

which is the same as the expression given in Eq. (14) in Sec. VII.

There are other valid factorizations of the PDF. One that might be of particular interest is a parameterization that depends on the overall signal fraction rather than the event-by-event signal probability used here. This can be obtained by replacing $\mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$ with $\mathcal{P}(s)\mathcal{P}(E_i(t, \varepsilon_s)|s, \mathrm{T})$, (and equivalently for the background terms). $P(s)$, often written as f_s, is the overall signal fraction, and $P(b) = 1 - P(s)$ the background fraction. This PDF differs from the one we use by an overall factor $P(E_i(t, \varepsilon_s)|\mathrm{T})$ which does not affect the maximum of the likelihood function. We find it easier to parameterize than $\mathcal{P}(E_i(t, \varepsilon_s)|s, \mathrm{T})$ and $\mathcal{P}(E_i(t, \varepsilon_s)|b, \mathrm{T})$ (compare Fig. 9 and Fig. 10).

APPENDIX B: CHARACTERIZING THE EFFICIENCY FUNCTION BY A VECTOR OF VARIABLES

In order to parametrize the term $\mathcal{P}(s|E_i(t, \varepsilon_s), \mathrm{T})$ which arises in the PDF in Eq. (14) we have chosen to use Fisher Discriminants to characterize each $E_i(t, \varepsilon_s)$ by a

scalar variable x_i as described in Sec. VII C 2. To use the Fisher Discriminant method we need to construct the vector $\boldsymbol{v}_i$ whose components describe the efficiency function $E_i(t, \varepsilon_s)$. How this vector is obtained is described here.

The vector $\boldsymbol{v}_i$ contains a series of variables: v_1, v_2..., v_n . Each variable should describe a property of the efficiency function in a way that allows comparison of one candidate's efficiency function with another. In Sec. V B we showed that the efficiency function can be written as

$$E_i(t, \varepsilon_s) = \sum_{\substack{k_i = \text{all} \\ \text{intervals} \\ \text{in event } i}} \left[\theta(t_i - t_{\min\, k}) - \theta(t_i - t_{\max\, k})\right] H_k(\varepsilon_s), \tag{B1}$$

where $H(\varepsilon_s)$ is either ε_s^2, $2\varepsilon_s^2 - \varepsilon_s^3$ or $3\varepsilon_s^2 - 2\varepsilon_s^3$, depending on whether there were one, two or three track pairs that could have passed the trigger.

From inspection of Eq.(B1), a single efficiency function can be uniquely defined by a series of variables which are $t_{\min\, k}$, $t_{\max\, k}$ and the value of $H_k(\varepsilon_s)$. However, this is not a useful description for comparing one efficiency function to the next simply because the number of intervals and hence the number of variables required to describe the efficiency function varies from one candidate to the next.

The efficiency function is defined for all candidates between a measured proper decay time of 0 and 10 ps. Another trivial way to construct $\boldsymbol{v}_i$ would be to bin the efficiency into n equal bins of time and take the mean value of $H_k(\varepsilon_s)$ in each bin as the elements v_1, v_2..., v_n. In this way, the value of any particular element of $\boldsymbol{v}_i$ for one candidate can be compared to the same element of $\boldsymbol{v}_i$ for another candidate. However, this method is also problematic as the ε_s is a floating parameter and the Fisher Discriminant and the observable, x_i, cannot be recalculated at each iteration of the likelihood minimization.

We take the following approach which allows construction of $\boldsymbol{v_i}$ so that its elements can be compared across candidates and knowledge of the value of ε_s is not required. The efficiency function can be re-written in the form

$$\begin{aligned} E_i(t, \varepsilon_s) = & \mathcal{A}_a \cdot \varepsilon_s^2 + \mathcal{A}_b \cdot (2\varepsilon_s^2 - \varepsilon_s^3) \\ & + \mathcal{A}_c \cdot (3\varepsilon_s^2 - 2\varepsilon_s^3), \end{aligned} \tag{B2}$$

where,

$$\mathcal{A}_a = \sum_{\substack{j = \text{all intervals} \\ \text{with} \\ H(\varepsilon_s) = \varepsilon_s^2}} \left[\theta(t_i - t_{\min\, j}) - \theta(t_i - t_{\max\, j})\right], \tag{B3}$$

and corresponding terms for $\mathcal{A}_b$ and $\mathcal{A}_c$. The value of the functions $\mathcal{A}$ at any time is either 0 or 1. Writing the efficiency function this way splits it into three sections, dependent on the value of $H(\varepsilon_s)$. Comparing $\mathcal{A}_a$ from one candidate to the next allows comparison of the efficiency function arising from the parts where there was only one track pair available to pass the trigger. To construct $\boldsymbol{v_i}$ we bin each of the $\mathcal{A}$ functions into 20 bins as a function of proper decay time. The value of v_1...v_{20} are the values of $\mathcal{A}_a$ in each bin. Nominally the value in any given bin is either 0 or 1, however, where the efficiency turns on or off within the bin an intermediate value is taken to represent the mean efficiency in that bin. Similarly the values of v_{21}...v_{40} are the values of $\mathcal{A}_b$ in each bin and v_{41}...v_{60} are the values of $\mathcal{A}_c$. By splitting the effect function into three, dependent on the form of $H_k(\varepsilon_s)$, we have removed the need to know the absolute value of ε_s and determining the v_i from bins allows for a vectorial representation of the efficiency function that is consistent and allows comparison of $E_i(t, \varepsilon_s)$ between different candidates. We now have a prescription for converting $E_i(t, \varepsilon_s)$ into $\boldsymbol{v_i}$ for each candidate and use this and the mass distribution to determine $\boldsymbol{v_s}$ and $\boldsymbol{v_b}$ as defined in Sec. VII C 2 and use these vectors to determine the direction $\boldsymbol{w}$, and therefore x_i using Fisher Discriminant Analysis [30].

APPENDIX C: A SIMPLER PDF

A lot of the complexity of the method presented here results directly or indirectly from the tight upper impact parameter cut applied by the two track trigger. In situations where this upper impact parameter cut is significantly looser, or ideally where no such cut is applied at all, one would not only benefit from a higher statistical precision for each candidate [25], but would also be able to employ a significantly simpler version of the method that will be outlined below. In this simpler version

- the dependence on ε_s can be removed,
- under many circumstances, there is no need to use the Fisher discriminant.

While we did not choose this approach for reasons specific to the CDF II detector trigger that will become apparent below, it is summarized here for the benefit of potential users of this method at other experiments.

Removing the dependence on ε_s

If the track-finding efficiency is decay time independent one can base the fit on the PDF *given* that a certain track combination has been reconstructed and seen by the trigger. The set of tracks seen by the trigger would be treated exactly in the same way as the decay kinematics; something that can be kept constant as the decay distance is changed for the efficiency function calculation.

Given that a certain track combination has been found, the trigger efficiency at a certain decay time is either 1 (passes cuts) or 0 (fails), independent of ε_s. This PDF would ignore one factor, the probability that exactly this track combination has been found. If this factor is decay time independent it does not affect the maximum of the likelihood and hence the result of the fit. With this, the signal PDF given in Eq. (10) reduces to:

$$\mathcal{P}(t_i;\tau|\mathrm{T},E_i(t),s) = \frac{\frac{1}{\tau}e^{\frac{-t_i}{\tau}+\frac{1}{2}\frac{\sigma_t^2}{\tau^2}}F\left(\frac{t_i}{\sigma_t}-\frac{\sigma_t}{\tau}\right)}{\sum\limits_{\substack{k_i=\text{all}\\ \text{intervals}\\ \text{in event } i}}\left[-e^{\frac{-t}{\tau}+\frac{1}{2}\frac{\sigma_t^2}{\tau^2}}F\left(\frac{t}{\sigma_t}-\frac{\sigma_t}{\tau}\right)+F\left(\frac{t}{\sigma_t}\right)\right]_{t=t_{\min\,\mathrm{k_i}}}^{t=t_{\max\,\mathrm{k_i}}}}. \tag{C1}$$

This approach, which is independent of ε_s, is valid whenever the track-finding efficiency is independent of the decay time for all tracks in the candidate. Despite the drop of the SVT track finding efficiency beyond $|d_0| > 1\,\mathrm{mm}$, this approach could in principle be used in the data analyzed in this paper if we applied a fiducial cut $|d_0| < 1\,\mathrm{mm}$ (where the SVT efficiency is effectively constant) for *all* tracks in the decay (this cut would of course need to be reflected in the efficiency function calculation). This is a significantly harsher requirement than that of the trigger, which requires two out of three tracks to have $0.12 < |d_0| < 1\,\mathrm{mm}$, allowing one of the tracks to be have $|d_0| > 1\,\mathrm{mm}$. We studied this option and found that the loss in statistical precision due to the additional cut is too large, mainly because of the effects discussed in Ref. [25]. This simpler approach would however be suitable in a situation where the track-finding efficiency is constant over a larger range than for the SVT.

Removing the need for a Fisher Discriminant

If the dependence on ε_s has been removed as described above, and in addition there is no variable upper proper decay time cut (no upper impact parameter cut), the candidate-by-candidate $E_i(t)$ is fully determined by one single parameter, the decay time $t_{\min}$ where the acceptance "turns on", i.e., above which the decay is accepted. Remembering that the motivation for introducing the Fisher discriminant was to translate the efficiency function into a simple number, this would clearly be unnecessary, as $E_i(t)$ is already fully described by a single number, $t_{\min}$. The factor $P(s|E_i(t,\varepsilon_s))$ can then be replaced by $P(s|t_{\min})$, with $P(b|E_i(t,\varepsilon_s)) = P(b|t_{\min}) = 1 - P(s|t_{\min})$. There is now no need for the Fisher scalar variable although the PDF term still requires a description of the signal fraction as a function of t_{min}.

Even simpler: Redefining t=0

Finally, in the case where there is no upper lifetime cut (i.e. $t_{\max} = \infty$), and the lower lifetime cut is hard enough to satisfy $t_{\min} \gg \sigma_t$, all the "F-terms" in Eq. (C1) are 1, and the equation reduces to

$$P(t) = \frac{1}{\tau}e^{-(t-t_{\min})/\tau} \tag{C2}$$

which is equivalent to an event-by-event redefinition of $t = 0$, as used by DELPHI in Ref. [26].

APPENDIX D: FULL FIT RESULTS

TABLE V: Summary of best fit mass parameters and uncertainties

Parameter	Best fit	Uncertainty
m_1 [GeV/c^2	5.2762]	$\pm$0.0004
m_2 [GeV/c^2	5.2711]	$\pm$0.0025
σ_1 [GeV/c^2	0.0247]	$\pm$0.0033
σ_2 [GeV/c^2	0.0138]	$\pm$0.0010
f_1	0.481	$\pm$0.13
f_s	0.741	$\pm$0.0050
α [(GeV/c^2)$^{-1}$]	-0.1658	$\pm$0.0035

TABLE VI: Summary of best fit efficiency parameters and uncertainties. The three periods correspond to the changes in the SVT described in Sec. III C.

Efficiency parameter	Best fit	Uncertainty
Signal Period 1	0.488	$\pm$0.033
Signal Period 2	0.656	$\pm$0.009
Signal Period 3	0.725	$\pm$0.006
Background Period 1	0.496	$\pm$0.064
Background Period 2	0.502	$\pm$0.019
Background Period 3	0.560	$\pm$0.017

TABLE VII: Summary of best fit background proper decay time parameters and uncertainties. The ct value represents the points where the background proper decay time distribution is sampled as defined by t_j in Eq. (16).

Backg. Parameter	ct_j (μm)	Best fit	Uncertainty
a_1	0	10.80	$\pm$0.39
a_2	146.9	7.08	$\pm$0.06
a_3	322.6	4.79	$\pm$0.04
a_4	532.7	2.73	$\pm$0.04
a_5	783.9	1.29	$\pm$0.07
a_6	1084.3	1.28	$\pm$0.10
a_7	1443.5	-1.19	$\pm$0.19
a_8	1873.1	-1.93	$\pm$0.29
a_9	2386.7	-2.73	$\pm$0.47
a_{10}	3000	-7.16	$\pm$2.87

[1] N. Cabibbo, Phys. Rev. Lett. **10**, 531 (1963).
[2] M. Kobayashi and T. Maskawa, Prog. Theor. Phys. **49**, 652 (1973).
[3] I. Bigi, M. Shifman, and N. Uraltsev, Annu. Rev. Nucl. Part. Sci. **47**, 591 (1997).
[4] C. Tarantino, Eur. Phys. J **C33**, S895 (2004).
[5] F. Gabbiani, A. I. Onischenko, and A. A. Petrov, Phys. Rev. D **68**, 114006 (2003).
[6] F. Gabbiani, A. I. Onischenko, and A. A. Petrov, Phys. Rev. D **70**, 094031 (2004).
[7] A. J. Lenz, AIP Conf. Proc. **1026**, 36 (2008), 0802.0977.
[8] M. Beneke, G. Buchalla, C. Greub, A. Lenz, and U. Nierste, Nucl. Phys. **B639**, 389 (2002), hep-ph/0202106.
[9] E. Franco, V. Lubicz, F. Mescia, and C. Tarantino, Nucl. Phys. **B633**, 212 (2002), hep-ph/0203089.
[10] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **97**, 242003 (2006).
[11] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **95**, 031801 (2005).
[12] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. Lett. **94**, 122001 (2005).
[13] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **97**, 211802 (2006).
[14] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **103**, 031801 (2009).
[15] T. Aaltonen *et al.* (CDF Collaboration), Phys. Rev. Lett. **104**, 102002 (2010).
[16] G. Punzi (2004), arXiv:physics/0401045.
[17] R. Brun, R. Hagelberg, M. Hansroul, and J. C. Laselle, Cern programme library, Cern-DD-78-2-REV, Cern-DD-78-2 (1994).
[18] D. Acosta *et al.* (CDF Collaboration), Phys. Rev. D **71**, 032001 (2005).
[19] T. Affolder *et al.*, Nucl. Instrum. Methods A **526**, 249 (2004).
[20] A. Sill *et al.*, Nucl. Instrum. Methods A **447**, 1 (2000).
[21] A. Affolder *et al.*, Nucl. Instrum. Methods A **453**, 84 (2000).
[22] E. J. Thomson *et al.*, IEEE Trans. Nucl. Sci. **49**, 1063 (2002).

TABLE VIII: Summary of best fit signal fraction as a function of Fisher scalar and uncertainties. The values of Fisher scalar gives the mid point of each bin used by the Lagrange interpolating polynomial function as described in Sec. VII C 5.

Fisher parameter	Fisher scalar, x	Best fit signal fraction	Uncertainty
x_1	-8.35	0.139	± 0.072
x_2	-7.05	0.273	± 0.071
x_3	-6.19	0.333	± 0.030
x_4	-5.32	0.379	± 0.014
x_5	-4.45	0.535	± 0.011
x_6	-3.59	0.657	± 0.008
x_7	-2.73	0.768	± 0.006
x_8	-1.86	0.825	± 0.005
x_9	-1.00	0.860	± 0.007
x_{10}	-0.13	0.907	± 0.007
x_{11}	0.73	0.937	± 0.011
x_{12}	1.60	0.941	± 0.034
x_{13}	2.89	1.00	± 0.045

[23] W. Ashmanskas *et al.*, Nucl. Instrum. Methods A **518**, 532 (2004).

[24] C. Amsler *et al.* (Particle Data Group), Phys. Lett. B **667**, 1 (2008).

[25] J. Rademacker, Nucl. Instrum. Methods A **570**, 525 (2007), arXiv:hep-ex/0502042.

[26] W. Adam *et al.* (DELPHI), Z. Phys. **C68**, 363 (1995).

[27] P. Nason, S. Dawson, and R. K. Ellis, Nucl. Phys. **B303**, 607 (1988).

[28] P. Nason, S. Dawson, and R. K. Ellis, Nucl. Phys. **B327**, 49 (1989).

[29] D. J. Lange, Nucl. Instrum. Methods Phys. Res., Sect. A **462**, 152 (2001).

[30] R. A. Fisher, Ann. of Eugenics **7**, 179 (1936).

[31] A. Abulencia *et al.* (CDF Collaboration), Phys. Rev. Lett. **98**, 122001 (2007).