CDF/BOTTOM/THESIS/PUBLIC/8051
V3
12/06/2005

SEARCH FOR SUPERSYMMETRY USING RARE $B^0_{s(d)} \to \mu^+\mu^-$ DECAYS AT CDF RUN II

A Dissertation

by

VYACHESLAV KRUTELYOV

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2005

Major Subject: Physics

SEARCH FOR SUPERSYMMETRY USING RARE $B^0_{s(d)} \to \mu^+\mu^-$ DECAYS AT CDF RUN II

A Dissertation

by

VYACHESLAV KRUTELYOV

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Dr. Teruki Kamon
Committee Members,	David Toback
	Richard L. Arnowitt
	Robert C. Webb
	Philip B. Yasskin
Head of Department,	Edward Fry

December 2005

Major Subject: Physics

ABSTRACT

Search for Supersymmetry Using Rare $B_s^0 \to \mu^+\mu^-$ Decays

at CDF Run II. (December 2005)

Vyacheslav Krutelyov, B.S., Belarus State University

Chair of Advisory Committee: Dr. Teruki Kamon

A search for rare $B_s^0 \to \mu^+\mu^-$ and $B_d^0 \to \mu^+\mu^-$ decays has been performed in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV using 364 pb^{-1} of data collected by the CDF II experiment at the Fermilab Tevatron Collider. The rate of each decay is sensitive to contributions from physics beyond the Standard Model (SM). No events pass the optimized selection requirements, consistent with the SM expectation. The resulting upper limits on the branching ratios are $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.5 \times 10^{-7}$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 3.8 \times 10^{-8}$ at the 90% confidence level. The limits are used to exclude some parameter space for several supersymmetric models.

F. Goya, *"The sleep of reason brings forth monsters."*

ACKNOWLEDGMENTS

I would like to acknowledge that without Teruki Kamon I would probably not have gotten the exciting experience of working in the field of the experimental high energy physics.

I would like to thank the core members of the $B_s^0 \to \mu^+\mu^-$ analysis group, Doug Glenzinski, Cheng-Ju Lin, Matt Herndon, Teruki Kamon, and those who worked with the group at various stages of this analysis, Julia Thom, Avi Yagil, Charles Plager and Nick Wisnewski. It is only with group effort that we could do the amount of work needed to perform the analysis.

I would like to thank our theoretical physics colleagues: Richard Arnowitt, Bhaskar Dutta, Uli Nierste, and many others. Special thanks to Bhaskar Dutta for preparing many of the important plots used in this dissertation.

I want to thank my CDF collaborators for their work and experience that keep everything running in CDF. Thanks to Prem Singh, Ken Bloom, Michael Schmitt, David Dagenhart and other members of the muon offline group we could make the muon reconstruction work. Thanks to Pasha Murat for his help and discussions, especially when I was entering the strange world of the CDF software. I appreciate the suggestions and comments from Jonathan Lewis that helped me understand the triggers. Thanks to many members of the muon hardware group who made the hardware work and helped me understand the many details, especially to Eric James and Phil Schlabach. Many thanks to the members of J/ψ and B-physics group at CDF, especially Thom LeCompte, Christoph Paus, Marjorie Shapiro and Steve Nahn who have helped to improve the analysis and facilitated the approval within the collaboration (aren't we blessed?).

Thanks to my friends who made my life easier and fun during the years spent in

Fermilab village.

Special thanks to David Toback for his valuable comments during the preparation of this dissertation, which made substantial improvements to the text, and for his time spent on proof-reading the text.

TABLE OF CONTENTS

CHAPTER Page

I INTRODUCTION 1

II THEORY 4

A. The Standard Model of particle physics 4
B. Beyond the Standard Model 8
C. Supersymmetry 10
D. Flavor changing neutral currents 18
E. Rare $B^0_{s(d)} \to \mu^+\mu^-$ decays 20
1. Standard Model expectations 22
2. SUSY expectations 24
a. Dominant diagrams in SUSY 25
b. Sensitivity to SUSY model parameters 29
F. B-hadron production and decay at hadron colliders 38

III EXPERIMENTAL SETUP 43

A. Accelerator chain 43
B. CDF II overview 46
C. CDF II detector elements 47
1. Silicon vertex detector 49
2. Central outer tracker 50
3. Calorimeter 52
4. Muon detectors 54
D. Data acquisition and trigger 61
1. Infrastructure 62
2. Level-1 (di)muon trigger 66
3. Level-2 (di)muon trigger 69
4. Level-3 (di)muon trigger 70
E. Offline software 71
1. Offline reconstruction 72
2. Monte Carlo simulation 76
a. Event generation 77

CHAPTER Page

b. Detector simulation 78
F. Calibrations 79

IV ANALYSIS 81

A. Overview 81
B. Methodology 82
1. Branching ratio observable 82
2. Blind analysis technique 83
3. Setting the limit 84
C. Data samples 85
1. The $B^0_{s(d)} \to \mu^+\mu^-$ sample 90
2. The $B^+ \to J/\psi K^+$ sample 91
3. $B^0_{s(d)} \to \mu^+\mu^-$ control samples 94
D. Monte Carlo samples 94
E. Discriminating variables for the $B^0_{s(d)} \to \mu^+\mu^-$ search . . . 99
1. Selection of variables 99
2. A multivariate relative likelihood discriminant 105
F. Background estimate 110
1. Estimating the mass window cut rejection and a number of sideband events 111
2. Estimating the likelihood cut rejection 112
3. Cross-checks using control samples 114
G. Estimate of the signal acceptance and efficiency 115
1. Acceptance 117
2. Trigger efficiencies 123
3. Reconstruction efficiencies 125
a. Dimuon reconstruction efficiencies 126
b. Vertex efficiencies 132
c. Kaon reconstruction efficiencies 132
4. Efficiency of the likelihood requirement 135
H. Optimization 137
1. The optimization procedure 137
2. Expectations for $B^0_s \to \mu^+\mu^-$ 140
3. Expectations for $B^0_d \to \mu^+\mu^-$ 140
I. Results 141

CHAPTER Page

J. Prospects for $B^0_{s(d)} \to \mu^+\mu^-$ at the Tevatron and beyond . 144
K. Exclusion of SUSY models parameters 149

V CONCLUSION 154

REFERENCES 155

APPENDIX A: SETTING UPPER LIMIT ON HEAVY HIGGS MASS USING $\mathcal{B}(B^0_S \to \mu^+\mu^-)$ 168

APPENDIX B: SILICON TRACKER CONFIGURATION 170

APPENDIX C: OFFLINE RECONSTRUCTION ALGORITHMS 173

A. Tracking 173
B. Muon reconstruction 178
C. Vertex reconstruction 181

APPENDIX D: ABSOLUTE VS. RELATIVE NORMALIZATION 185

APPENDIX E: SUMMARY OF THE TRIGGERS USED IN THE ANALYSIS 187

APPENDIX F: MEASUREMENT OF LEVEL-1, 2, AND 3 LOW-P_T DIMUON TRIGGER EFFICIENCIES 189

A. Trigger efficiency 189
B. Level-1 dimuon trigger efficiency (ϵ_{L1}) 192
C. Single muon Level-1 trigger efficiency 193
1. Dataset 193
2. Methodology 193
3. Level-1 CMU trigger efficiency: XFT 2-miss 197
4. Level-1 CMX trigger efficiency: XFT 2-miss 199
5. Level-1 CMU trigger efficiency: XFT 1-miss 200

Page

6. Level-1 CMX trigger efficiency: XFT 1-miss 207
7. Systematics 208

D. Level-2 dimuon trigger efficiency (ϵ_{L2}) 210
E. Level-3 dimuon trigger efficiency (ϵ_{L3}) 212
F. Cross-check using $J/\psi \to \mu^+\mu^-$ yields 214
G. Summary 217

APPENDIX G: KINEMATIC-DEPENDENT MATCHING CUT FOR LOW-P_T CMX MUONS 218

A. Methodology 219
B. Parameterization of $\sigma_{\Delta X}$ 220
C. Efficiency of $\chi^2_{r\phi,\mathrm{CMX}} < 9$ cut 222
D. Background rejection 225
E. Summary 227

APPENDIX H: RESIDUAL BACKGROUND CONTRIBUTIONS FROM B-DECAYS 228

APPENDIX I: NORMALIZATION MODE CORRECTION 232

VITA 234

LIST OF FIGURES

FIGURE Page

1 SUSY Higgs mass correction diagrams. 12

2 A schematic of a sample MSSM spectrum. 16

3 Effective FCNC vertex with neutral field N. 19

4 Example FCNC $b \to s\gamma$ diagram. 19

5 Standard Model Feynman diagrams contributing to $B_s^0 \to \mu^+\mu^-$. 23

6 The leading tree diagram in RPV mSUGRA. 26

7 Dominant diagram for $B_{s(d)}^0 \to \mu^+\mu^-$ in the MSSM. 29

8 RPV mSUGRA allowed λ dependence in $(m_0, m_{1/2})$ plane for a fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.6 \times 10^{-7}$. 31

9 RPV mSUGRA $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ dependence in $(m_0, m_{1/2})$ plane for a fixed λ. 31

10 Fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ contours in mSUGRA in $(m_0, m_{1/2})$ plane. 33

11 Fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ contours in mSUGRA in $(m_0, m_{1/2})$ plane with experimental constraints included. 33

12 Fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ contours in mSUGRA in $(\tan\beta, m_0)$ plane. 34

13 Fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ contours and Tevatron trilepton reach in mSUGRA in $(\tan\beta, m_0)$ plane. 34

14 Correlation between $\sigma_{\tilde{\chi}_1^0 p}$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in mSUGRA (left) and SUGRA with non-universal Higgs masses (right). 36

15 Allowed region in $\mathrm{MSO_{10}SM}$ in $(\mu, m_{1/2})$ plane. 37

16 Allowed range of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ for the given m_A in $\mathrm{MSO_{10}SM}$. 38

FIGURE Page

17 Production cross section of b-hadrons. 40

18 Drawing and aerial view of the Fermilab's accelerator chain. Protons (anti-protons) inside the Tevatron are moving clock-wise (counter clock-wise). 44

19 An isometric view of the CDF II detector showing the main components. Most name component abbreviations are also shown. 48

20 An r-z view of the silicon detector. Note different radial and longitudinal scales. 50

21 An r-ϕ view of the silicon detector. 50

22 One sixth of the COT endplate. For each superlayer is given the total number of supercells, the wire orientation (axial or stereo), and the average radius. The enlargement shows the sense and field sheet slot geometry in detail. Dimensions are in cm. 51

23 Coverage provided by CDF II muon systems. 55

24 Number of pion interaction lengths for muon systems. 56

25 Cut view of one CMU module. The three modules in a wedge are slightly different. Shown is the central one. 58

26 Cut view of a part of CMP. 58

27 Cut view of a part of CMX arch. An illustration of a passing track and the resulting hit pattern is also shown. 61

28 Cut view of a part of BMU. 61

29 Data flow diagram for deadtimeless trigger and DAQ system. 63

30 Data flow for L1 muon trigger. 67

31 Impact parameter resolution of tracks with silicon hits (that with L00 is shown separately). 74

FIGURE Page

32 Event display r-ϕ view of a (simulated) muon with $p_T = 2$ GeV/c, $\eta = 0.1$, and $\phi_0 = 35°$. 76

33 Di-muon invariant mass distributions for events in the $B_s^0 \to \mu^+\mu^-$ CMU-CMU (left) and CMU-CMX (right) search samples. The results of a fit to a straight line ($p_0 + p_1 M_{\mu^+\mu^-}$) are shown. 91

34 The $\mu^+\mu^-K^+$ invariant mass distributions for CMU-CMU (left) and CMU-CMX (right) events in the $B^+ \to J/\psi K^+$ sample. 93

35 The $p_T(B)$ spectrum from $B^+ \to J/\psi K^+$ events for CMU-CMU and CMU-CMX combined. 96

36 The invariant mass distributions for $B^+ \to J/\psi K^+$ (left) and $B_s^0 \to \mu^+\mu^-$ (right) MC events. The results of a fit to a Gaussian are also shown. 97

37 The invariant mass distribution for $J/\psi \to \mu^+\mu^-$ from $B^+ \to J/\psi K^+$ candidates in the data (left) and the PYTHIA MC (right). Each is fit to a Gaussian distribution. 98

38 Distributions of discriminating variables for MC signal (dashed histograms) and a $B_s^0 \to \mu^+\mu^-$ data sample (solid histograms) events passing the baseline and vertex requirements. The histograms are all normalized to unit area. For the $M_{\mu^+\mu^-}$ plots, different binning is used for the data and MC. 101

39 Profile plots showing the correlations among the discriminating variables for CMU-CMU $\mu^+\mu^-$ pairs. 102

40 Profile plots showing the correlations among the discriminating variables for CMU-CMX $\mu^+\mu^-$ pairs. 103

41 The λ distribution for $B_s^0 \to \mu^+\mu^-$ MC events with $p_T > 4$ GeV/c and $|\eta(B)| < 1$ (top), additionally satisfying the trigger acceptance criteria described in Section IV.G.1 (middle), and also satisfying the remaining baseline and vertex requirements (bottom). 106

FIGURE Page

42 The I (top), $\Delta\alpha$ (middle), and λ (bottom) probability distributions used to construct the LH variable for the CMU-CMU events. 107

43 The I (top), $\Delta\alpha$ (middle), and λ (bottom) probability distributions used to construct the LH variable for the CMU-CMX events. 108

44 Likelihood curves for signal ($B_s^0 \to \mu^+\mu^-$ MC) and background (data sidebands) in the CMU-CMU channel. 109

45 A comparison of the LH distribution from data sideband events (histogram) and from the toy-MC (points): CMU-CMU (left) and CMU-CMX (right) channels. The KS-probability is 2% (3%) for the left (right). 113

46 Rapidity and transverse momentum distributions for B_s^0 mesons at the generator level. 119

47 Total (▲) and geometric (histogram) acceptance for B_s^0 vs. $p_T^{B_s^0}$. 121

48 Total (▲) and geometric (histogram) acceptance for B_s^0 vs. $\eta^{B_s^0}$. 121

49 Total (▲) and geometric (histogram) acceptance for B^+ vs. $p_T^{B^+}$. 122

50 Total (▲) and geometric (histogram) acceptance for B^+ vs. η^{B^+}. 122

51 Efficiency of $\chi^2_{r\phi} < 9$ for CMU muons vs p_T. 129

52 $\epsilon(\chi^2_{r\phi,\mathrm{CMX}} < 9)$ vs. p_T: μ^+ (▲), μ^- (▼), and charge averaged (•). 129

53 Silicon reconstruction efficiency per muon vs. run number. 131

54 Silicon reconstruction efficiency per muon vs. dimuon track isolation. . . 131

55 The COT reconstruction efficiency for kaons from $B^+ \to J/\psi K^+$ decays: default (central curve) and $\pm 1\sigma_{KN}$ (envelope). 133

56 The relative COT reconstruction efficiency in bins of B-hadron isolation (left) and instantaneous luminosity (right) for kaons from $B^+ \to J/\psi K^+$ decays. 134

FIGURE Page

57 A comparison of the input distributions and the resulting LH distribution from sideband subtracted $B^+ \to J/\psi K^+$ data to that from the PYTHIA MC sample for the CMU-CMU channel. 136

58 A comparison of the input distributions and the resulting LH distribution from sideband subtracted $B^+ \to J/\psi K^+$ data to that from the PYTHIA MC sample for the CMU-CMX channel. 137

59 Distribution of $p_T^{\mu\mu}$ for data sidebands and $B_s^0 \to \mu^+\mu^-$MC events passing the baseline and vertex requirements. Both distributions are normalized to the unit area. 139

60 The invariant mass distribution for events satisfying all selection criteria in the CMU-CMU (left) and CMU-CMX (right) channels. 142

61 The invariant mass distribution versus the event likelihood for events satisfying baseline in the CMU-CMU (left) and CMU-CMX (right) channels. 143

62 The likelihood distribution for events in the B_s search window for the CMU-CMU (left) and CMU-CMX (right) channel. 143

63 Expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ upper limit based on the presented analysis. . 145

64 Allowed region in $\mathrm{MSO_{10}SM}$ in $(\mu, m_{1/2})$ plane with current experimental limits included. 150

65 Allowed range of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ for the given m_A in $\mathrm{MSO_{10}SM}$: $m_A < 500$ GeV$/c^2$ is excluded. 151

66 Parameter space in $(m_0, m_{1/2})$ plane of RPV mSUGRA excluded by $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2.1 \times 10^{-7}$. 152

67 Parameter space in $(m_0, m_{1/2})$ plane of mSUGRA excluded by $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2.1 \times 10^{-7}$ (orange filled). 153

68 Setting upper limit on m_A in MFV MSSM at $\tan\beta = 50$. 169

FIGURE Page

69 Dimuon invariant mass distributions for events in the CMU unbiased trigger path. 195

70 L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ combined for ranges #1 and #2. 198

71 L1 CMU source of inefficiency for run ranges #1 and #2 combined: $\epsilon^{\mu}_{\mathrm{CMU}}$ (■), $\epsilon^{\mu}_{\mathrm{stub,CMU}}$ (▲), and $\epsilon^{\mu}_{\mathrm{XTRP,CMU}}$ (○). 199

72 L1 $\epsilon^{\mu}_{\mathrm{CMX}}$ (XFT 2-miss). 200

73 L1 CMX source of inefficiency for XFT 2-miss period: $\epsilon^{\mu}_{\mathrm{stub,CMX}}$ (▲), $\epsilon^{\mu}_{\mathrm{XTRP,CMX}}$ (○), and $\epsilon^{\mu}_{\mathrm{CMX}}$ (■). 201

74 L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ (XFT 1-miss) vs. muon $|\eta|$ for two low p_T bins for the top ϕ bin for ranges #3 (■), #5 (▲), #6 (○), and #8 (□). 202

75 L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ (XFT 1-miss) vs. muon $|\eta|$ for two low-p_T bins for the bottom ϕ bin for ranges #3 (■), #5 (▲), #6 (○), and #8 (□). 203

76 L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ (XFT 1-miss) vs. muon ϕ_{SL6} for ranges #3 (■), #5 (▲), #6 (○), and #8 (□). 204

77 L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ (XFT 1-miss) vs. muon p_T^{-1} in slices of $|\eta|$ for run range #3. . . 206

78 L1 CMU source of inefficiency in XFT 1-miss period for run range #4 averaged over ϕ: $\epsilon^{\mu}_{\mathrm{CMU}}$ (■), $\epsilon^{\mu}_{\mathrm{stub,CMU}}$ (▲), and $\epsilon^{\mu}_{\mathrm{XTRP,CMU}}$ (○). 207

79 L1 CMX source of inefficiency for XFT 1-miss period (ranges #3-#5 combined only): $\epsilon^{\mu}_{\mathrm{stub,CMX}}$ (▲), $\epsilon^{\mu}_{\mathrm{XTRP,CMX}}$ (○), $\epsilon^{\mu}_{\mathrm{CSX,CMX}}$ (□), and $\epsilon^{\mu}_{\mathrm{CMX}}$ (■). 209

80 Raw $J/\psi \to \mu^+\mu^-$ yield vs. run number in binned by run ranges given in Table XVII as collected on `JPSI_CMUCMU` and `JPSI_CMUCMX` trigger paths. The magenta error bars represent the statistical uncertainty of the $J/\psi \to \mu^+\mu^-$ sample. 215

FIGURE Page

81 Yield of $J/\psi \to \mu^+\mu^-$ scaled by $1/\epsilon_{\text{trig}}$ vs run number as collected on JPSI_CMUCMU and JPSI_CMUCMX trigger paths. The magenta (black) error bars represent the statistical uncertainty of the $J/\psi \to \mu^+\mu^-$ sample (uncertainty of $1/\epsilon_{\text{trig}}$). 216

82 Distribution of χ_{CMX} for $p_T > 3$ GeV/c. 221

83 p_T dependence of the pull width $\sigma_{\chi_{r\phi,\text{CMX}}}$: μ^+ (▲), μ^- (▼), and charge averaged (•). 222

84 η dependence of the pull width $\sigma_{\chi_{r\phi,\text{CMX}}}$: μ^+ (▲), μ^- (▼), and charge averaged (•). 222

85 $\epsilon(\chi^2_{r\phi,\text{CMX}} < 9)$ vs. p_T: μ^+ (▲), μ^- (▼), and charge averaged (•). 223

86 $R(\chi^2_{r\phi,\text{CMX}} < 9)$ vs. p_T for the $J/\psi \to \mu^+\mu^-$ sidebands: μ^+ (▲), μ^- (▼), and charge averaged (•). 225

87 χ_{CMX} distribution for $p_T > 3$ GeV/c for the right $J/\psi \to \mu^+\mu^-$ sideband. The Gaussian component is large. 227

88 Muon fake rate for kaons vs. p_T for CMU. 229

89 Muon fake rate for pions vs. p_T for CMU. 229

90 Invariant mass distribution for $\mu^+\mu^-$candidates in generic QCD heavy flavor MC, including double-fake-muon (top), muon-fake-muon (middle), and dimuon contributions (lower) for MC events passing baseline and vertex selections applied on generator level. The distributions with hadrons are not rescaled to account for fake rates. To increase the statistics of the test both electrons and muons are counted as muons (denoted as l in the plot), this is possible due to approximate $e - \mu$ symmetry of the dominant (semi)leptonic heavy flavor decay modes. . . . 231

LIST OF TABLES

TABLE Page

I Undiscovered particles in the MSSM. The gauge eigenstates denoted by “” are the same as the mass eigenstates. 15

II Properties of the mesons relevant to this analysis. 42

III CDF II calorimetry summary. 53

IV Operation parameters of the CDF II muon detectors. 57

V Baseline quality requirements. 88

VI B-candidate kinematic requirements. 89

VII B-candidate vertex requirements for $B^0_{s(d)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ samples. 89

VIII The linear correlation coefficients among the four discriminating variables. 104

IX The LH rejection factor calculated from the $\mu^+\mu^-$ toy-MC. Only the statistical uncertainties are included. 114

X The KS test probabilities when comparing the LH curve from toy-MC to that from data observed using the control samples. 115

XI A comparison of the predicted and observed number of events in the signal mass region as a function of likelihood cut for the various control samples. 116

XII A summary of the inputs used in to estimate the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$. The relative uncertainties are given parenthetically. The single-event-sensitivity, *ses*, is $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ corresponding to $N_{B^0_s} = 1$. 117

XIII Normalized acceptance ratio variation in the generated systematic samples. Deviations from 1 quantify the associated systematic uncertainty. . 124

TABLE Page

XIV The $\epsilon^{LH}_{B^0_s}$ for the CMU-CMU (left) and CMU-CMX (right) channels. The uncertainties are the binomial statistical uncertainties. 135

XV The ratio of the $\epsilon^{LH}_{B^+}$ determined from data to that from MC, including the statistical uncertainties. 138

XVI Silicon detector configuration. 172

XVII Summary of the run ranges for the efficiency measurement with the approximate begin-end dates specified. The luminosity per range is calculated for the runs with good silicon. 192

XVIII L2 trigger efficiencies with statistical and systematic uncertainties included. Efficiency for `L2_TWO_CMU_2G` is applicable to run range #8 only. 211

XIX L3 efficiencies with statistical and systematic uncertainties ($\epsilon \pm \delta_{\text{stat}} \pm \delta_{\text{syst}}$), given in per cent. 213

XX $\chi^2_{r\phi,\text{CMX}} < 9$ cut efficiency summary. 224

XXI $R(\chi^2_{r\phi,\text{CMX}} < 9)$ for $J/\psi \to \mu^+\mu^-$ and $B^0_{s(d)} \to \mu^+\mu^-$ sidebands. Only the statistical errors are shown. 226

CHAPTER I

INTRODUCTION

The Standard Model (SM) of particle physics has been the most successful in explaining the interactions between the elementary particles for the past two decades [1]. Nevertheless, there are certain theoretical problems and experimental observations that it can not explain [2–5]. While the recent experimental observations of non-zero neutrino masses are striking [6], perhaps the most important are those of dark matter [7–9]. Cosmological observations indicate that of all the matter in the Universe only about one sixth is the visible (interacting with light and contained in the SM) matter, the rest is not visible (dark matter), but should be there to explain the structure of the Universe [1, 7, 9]. The above issues give strong evidence that there is new physics beyond the SM.

A very theoretically attractive extension of the SM that potentially solves the above problems is supersymmetry (SUSY) [1, 2, 4, 10, 11], a theory that predicts that for each kind of elementary particle in the SM there should be a superpartner. It also indicates that these superpartners can be observed at the experimental collider facilities available now, or in the near future. The search for evidence of these predicted particles is thus of the highest priority.

In principle, one can detect the presence of these new particles not only by their direct observation (via production and observation of their decay products), but also by the effect these particles produce by appearing virtually in the interactions between other particles. Specifically, the latter can be done by searching for rare decays of some copiously produced known SM particles, which are either prohibited or highly

---

This dissertation follows the style of Physical Review D.

suppressed in the SM but might be not nearly as rare due to the presence of the new particles or interactions. For example, flavor changing neutral current (FCNC) processes [1, 12, 13], described by a transformation (flavor change) between quarks with the same electric charge, are strongly suppressed in the SM. Should there be new particles appearing virtually, they could effectively relax this suppression and raise the rates of FCNC processes above the SM expectations and give the first indications of new physics.

Recently, it was proposed that a FCNC process potentially sensitive to new physics is the rare decay of B_s^0 meson to two muons [14–16].[1] This decay is heavily suppressed in the SM with a predicted branching ratio of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.5 \pm 0.9) \times 10^{-9}$ [13, 17]. In contrast, some attractive SUSY models predict a much higher rate, with values as much as a few orders of magnitude higher than the SM expected value. In fact, this decay is currently considered as one of the golden discovery modes in the search for SUSY.

One of the most promising ways to search for evidence of $B_s^0 \to \mu^+\mu^-$ decays is at the Tevatron collider at Fermi National Accelerator Laboratory, the worlds most powerful particle accelerator. At Fermilab, anti-protons and protons are accelerated and then collide head-on at a center of mass energy, of 1.96 TeV [18]. The Tevatron produces more than a million collisions per second, with more than 20 trillion collisions, billions of them with B_s^0 mesons produced, since 2001. The products of these collisions are observed and analyzed using the Collider Detector at Fermilab (CDF II), a general purpose detector [19]. This suggests a powerful potential sensitivity to supersymmetry through the direct observation of $B_s^0 \to \mu^+\mu^-$ decays.

[1]The rare decay $B_d^0 \to \mu^+\mu^-$, which is very similar experimentally, is also concurrently searched for in this analysis, thus defining the notation $B_{s(d)}^0 \to \mu^+\mu^-$. While it is less sensitive to commonly considered new physics scenarios, in many others it has the potential to dominate the sensitivity.

Presented in this thesis is the search for $B^0_{s(d)} \to \mu^+\mu^-$ decays performed with the data collected by CDF between 2002 and 2004. The sensitivity of this analysis is more than a factor of two better than the previous world's most sensitive search [20] and a factor of twelve better than the measurement of a similar analysis performed by CDF in Run I (1992-1996) [21] which was the best published limit at the moment the present analysis began. Note that an intermediate result, based on the first part (collected by 2003) of the data used in this analysis was published in 2004 [22] and, as of this writing, is the second most cited CDF Run II analysis. The results of the full analysis presented here have also been accepted for publication [23].

A brief outline of the thesis is the following. A more detailed theoretical motivation for SUSY and the search for $B^0_{s(d)} \to \mu^+\mu^-$ decays is given in Chapter II. Chapter III describes the experimental setup and the analysis tools. A detailed description of the analysis is provided in Chapter IV along with the results, prospects, and their implications for the various SUSY models, followed by the conclusions in Chapter V.

CHAPTER II

THEORY

The rare FCNC decay, $B_s^0 \to \mu^+\mu^-$, has been suggested as a powerful tool to probe for a new physics beyond the SM. This chapter first provides a description of the SM framework related to flavor changing interactions in Section II.A. The theoretical concepts addressed by the models beyond the SM are presented in Section II.B, followed in Section II.C by a discussion of supersymmetry and the phenomenological framework of the Minimal Supersymmetric Standard Model (MSSM) together with specific model implementations relevant to this analysis. The importance of FCNC decays is discussed in Section II.D. The chapter is concluded with a detailed discussion of the expected branching ratio of the $B_s^0 \to \mu^+\mu^-$ decay in the SM and SUSY models given in Section II.E: the dominant contributions available in SUSY models are identified first to determine the model parameters sensitive to these contributions and to be able to exclude these parameters given the measurement of (the upper limit of) $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. B-meson production and dominant decay properties at hadron colliders are discussed in Section II.F.

A. The Standard Model of particle physics

The SM defines the types of interactions between elementary particles and stratifies the particles based on the types of interactions between them [1, 2]. The basics of the SM are described in this section, in particular, the weak (flavor changing) interactions that govern decay properties of B-mesons in particular. The section concludes with an overview of success and problems of the SM.

The main interaction types or forces in the SM are the strong and electroweak forces, mathematically related to gauge fields of $\mathrm{SU}(3)_C$ and $\mathrm{SU}(2)_W \times \mathrm{U}(1)_Y$ groups

for the strong (quantum chromodynamics or QCD) and electroweak interactions respectively. The gauge fields correspond to the particles mediating the corresponding interactions between the particles. Initially, all the particles in the model are introduced massless, and masses are generated as a result of interaction with a scalar (Higgs) field via a spontaneous electroweak symmetry breaking (EWSB) mechanism. This mechanism also defines the electromagnetic (massless photon field A_μ or γ) and the weak (massive charged fields $W^\pm$ and a neutral field Z) components of the original unbroken $\mathrm{SU}(2)_W \times \mathrm{U}(1)_Y$ electroweak symmetry. The gauge fields define the vector field content of the model: gluon (g for $\mathrm{SU}(3)_C$) and $W^\pm, Z$ and γ (for $\mathrm{SU}(2)_W \times \mathrm{U}(1)_Y$). The Higgs field is the only scalar elementary field in the model. The remaining particles are fermions with spin 1/2.

The fermions in the SM are grouped by their interaction properties. The fermions partaking in the strong interactions are the six quarks (quark flavors): up (u), charm (c) and top (t) quarks with electric charge[1] 2/3; and down (d), strange (s), and bottom (b) quarks with electric charge $-1/3$. The fermions not partaking in the strong interactions, are leptons of six types: electron (e^-) and its neutrino (ν_e), muon (μ^-) and its neutrino (ν_μ), and tau (τ^-) and its neutrino ν_τ, where e^-, μ^-, and τ^- have an electric charge -1, while the neutrinos are neutral. The masses of the SM

[1]In units of absolute value of the electric charge of electron.

particles are

$$m_e = 0.511 \text{ MeV}/c^2, \qquad m_d = 6 \text{ MeV}/c^2, \qquad m_u = 3 \text{ MeV}/c^2,$$
$$m_\mu = 0.1057 \text{ GeV}/c^2, \qquad m_s = 0.1 \text{ GeV}/c^2, \qquad m_c = 1.3 \text{ GeV}/c^2,$$
$$m_\tau = 1.777 \text{ GeV}/c^2, \qquad m_b = 4.26 \text{ GeV}/c^2, \qquad m_t = 175 \text{ GeV}/c^2,$$
$$m_\gamma = 0 \text{ GeV}/c^2, \qquad m_{W^\pm} = 80.4 \text{ GeV}/c^2, \qquad m_Z = 91.2 \text{ GeV}/c^2,$$
$$m_g = 0 \text{ GeV}/c^2, \qquad m_h > 114 \text{ GeV}/c^2.$$

The fermions with different chirality belong to different SU(2) representations, which defines the structure and the properties of their interactions in the SM. The left-handed fermions ($\psi_L \equiv \frac{1-\gamma_5}{2}\psi$) form SU(2) doublets:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \qquad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \qquad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \tag{2.1}$$

$$\begin{pmatrix} u \\ d' \end{pmatrix}_L \qquad \begin{pmatrix} c \\ s' \end{pmatrix}_L \qquad \begin{pmatrix} t \\ b' \end{pmatrix}_L \tag{2.2}$$

The primed quark flavors, q_i', are the gauge eigenstates which are the mix of the mass-eigenstates q_i defined by a unitary 3×3 matrix V, known as Cabbibo-Kobayashi-Maskawa (CKM) matrix $q_i' = V_{ij} q_i$. The absolute values of the CKM matrix are given by [1]

$$\hat{V}_{CKM} = \begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} \approx \begin{pmatrix} 0.975 & 0.224 & 0.004 \\ 0.224 & 0.974 & 0.042 \\ 0.01 & 0.040 & 0.999 \end{pmatrix}. \tag{2.3}$$

The right-handed fermions form SU(2) singlets. The neutrinos are left-handed only, which, by construction, makes them massless. A left-handed pair of quarks or leptons, with the corresponding right-handed components combined, define a generation:

three quark and three lepton generations, ordered in increasing constituent mass – corresponding to the pairs given in Eqs. (2.1) and (2.2).

The flavor changing interactions involve only the left-handed quark field interactions with the $W^{\pm}$ bosons, also referred to as a charged current interactions. The term of the SM interaction Lagrangian describing the flavor changing interaction is

$$\mathcal{L}_I = \frac{g}{2\sqrt{2}} \sum_i \begin{pmatrix} \bar{u} & \bar{c} & \bar{t} \end{pmatrix} \gamma^{\mu}(1-\gamma^5) W_{\mu}^{+} V \begin{pmatrix} d \\ s \\ b \end{pmatrix} + \text{h.c.}, \tag{2.4}$$

where $g \sim \frac{2}{3}$ is the weak coupling constant, V is the CKM matrix, and $h.c.$ stands for the Hermitian conjugate term.

In the SM it is assumed that quarks at low energies only exist in the form of bound states, hadrons. The hadrons corresponding to quark-antiquark states are called mesons, and those corresponding to three-quark states are called baryons. The hadrons with b and c constituent quarks are referred to as heavy flavor hadrons: B-hadrons and charm hadrons respectively.

The SM has been very successful at predicting the results of an enormous number experiments with incredible precision for the past two decades. The electroweak observables as predicted by the SM are all substantially within the experimental uncertainties [24]. The flavor mixing, as described by the CKM matrix is also well within the theory [25, 26].[2] The perturbative QCD predictions are also a success, many of those contribute to the prediction of the electroweak observables as QCD corrections, as well as to many other predictions. The non-perturbative QCD calculations, which are notoriously difficult, have even made successful mass and decay

[2]Should there be a substantial contribution to the heavy flavor processes from the particles not included in the SM, the unitarity triangle fit would not be consistent.

constant predictions recently by solving the QCD field equations numerically on the space-time lattice [27–29].

Despite the success of the SM in many areas, some of the experimental observations are not explained by the model, such as neutrino oscillations and the dark matter in the Universe. Neutrino oscillations [6], observation of neutrino flavor change with distance/time, unequivocally suggest that neutrinos have mass. The dark matter problem is related to the cosmological observations that show that of all the matter in the Universe only a fraction consists of the particles included in the SM. Namely from all the matter contributing about $\Omega_M \approx 30\%$ of the total universe energy density, only about $\Omega_B = 5\%$ can be attributed to baryons, the rest is not visible (does not interact with photons) [7–9]. Physics beyond the SM is also suggested by the measurement of the muon anomalous magnetic moment $a_\mu \equiv (g_\mu - 2)/2$ by the $(g-2)$ collaboration [30]: the measurement is close to 3σ away from the SM expectations [31].

B. Beyond the Standard Model

There is a substantial need to extend the SM. The models extending the SM are driven by a number of theoretical concepts not included in the SM, as well as by a desire to explain the experimental observations described above. The most discussed models are inspired by the combination of some of the following concepts: grand unification, Yukawa or gauge hierarchy problems, extra dimensions, formulations of particles as strings (superstring theories), dynamic symmetry breaking, quark and lepton compositeness, boson-fermion symmetry or supersymmetry [1]. A short description of these concepts as addressed in the models extending the SM is given below, and SUSY models are detailed in the next section. Note that successful SUSY models in many cases implement most of these concepts (compositeness not included).

Grand unified theories (GUTs) attempt to unify the electroweak and strong interactions. They describe the SM gauge group by a single simple gauge group in which all the particles can be combined in a single representation (one for matter and one for the gauge fields). This group symmetry is then considered to be broken below some energy scale (the GUT scale) which gives rise to the $\mathrm{SU}(3)_C \times \mathrm{SU}(2)_W \times \mathrm{U}(1)_Y$. This is mostly inspired by the fact that the three running couplings of the SM do seem to converge to roughly the same value at the scale of around 10^{16} GeV. The common consequences of the GUTs are lepton and baryon number violation and, as a result, the possibility of proton decay.

The Yukawa hierarchy problem refers to the vast mass difference between the particles of the SM – the quark masses alone vary within five orders of magnitude. The gauge hierarchy problem is usually connected to the Higgs field in the model in application to the higher energy scales. The Higgs mass gains a correction proportional to the mass-squared of the fields that the Higgs interacts with.[3] If one is now to assume that there are particles at energies above the electroweak scale of about 100 GeV, some possibly as heavy as Plank mass or GUT scale mass, this would mean that the Higgs mass corrections from those particles would have to miraculously cancel at the level of $m_{\mathrm{EW}}^2/m_X^2 \sim 10^{-30}$, which does not seem natural [3, 4].

The extra dimensions theories are formulated in $4+N$ dimensions, where the four dimensions correspond to those we live in. In most of the cases the N extra dimensions are compact [1, 32]. In these models the gauge hierarchy problem is solved naturally, since the Plank scale visible in the $4+N$ dimensions is smaller than that in 4 dimensions and can be as low as the TeV range.

[3]Note that it is common in the literature to consider the self-energy diagram loop integral with a fixed cut-off scale Λ, i.e. $\int d^4k/k^2 \sim \Lambda^2$ thus assigning a physical meaning to the mathematical procedure of regularization needed to apply renormalization.

The models with quark or lepton compositeness assume that the SM fermions are made of constituents. The large binding energy and chiral invariance of the interaction between the constituents can explain the low mass of the quarks and leptons. These models suggest the existence of excited fermions, which however have not been observed.

The models with dynamic symmetry breaking suggest a mechanism alternative to the Higgs mechanism of symmetry breaking. Instead of the Higgs field that gains a non-zero vacuum expectation value (*vev*), it is assumed that a fermion bilinear (i.e., a fermion-antifermion combination) gains a *vev*. In these models it is made possible to give rise to the masses of the W and Z bosons and all the fermions [1, 33].

The string theories describe particles as extended objects, one-dimensional strings. The consistent string theories that include both bosons and fermions are supersymmetric. They are formulated in a higher dimensional space time: the most theoretically attractive M theory is defined in 11 dimensions. It is highly anticipated that the M theory (or other superstring theory) would be a unified theory that includes gravity, electroweak, and strong forces in terms of quantized gauge field theory [34, 35].

C. Supersymmetry

Concepts of supersymmetry and the dominant phenomenological models are described in this section. Primary properties of SUSY and how it solves some theoretical and experimental problems, as well as the concept of SUSY breaking are discussed first. The minimal supersymmetric extension of the SM (MSSM) is discussed next, including the description of its particle content and interactions, giving a particular attention to the interactions defined by the Higgs sector which play an important role in defining the $B_s^0 \to \mu^+\mu^-$ decays. The problem of a large number of parameters in

the MSSM and the possible solutions to this problem are discussed next. The section is concluded with the description of the benchmark models sensitive to the present analysis, such as mSUGRA with possible RPV or Higgs non-universality extensions, and $MSO_{10}SM$ model (a minimal SUSY GUT SO(10) model).

Supersymmetry is the symmetry between bosons and fermions. A supersymmetric theory has the same number of bosonic and fermionic degrees of freedom, that is, every particle has a corresponding superpartner. The names of the superpartners to fermions have a prefix "s" (e.g., squarks), and the names of the superpartners to bosons have a suffix "ino" (e.g., gluino). Mathematically supersymmetry arises from the requirement that the theory should be invariant under the transformations that convert bosons to fermions and vice versa. These are global transformations of the fields and can be considered as an extension to the Poincare group transformations [1, 2, 4].

SUSY is theoretically attractive from various standpoints: it provides a natural solution to the gauge hierarchy problem, the consistent string theories require supersymmetry, the SM gauge coupling unification is achievable (gives rise to SUSY GUTs), it can possibly explain the neutrino masses and the presence of the dark matter. The gauge hierarchy problem is solved in the following way. The Higgs mass corrections by the virtual fermion and boson exchange at the lowest loop level is expressed by the self-energy diagrams shown in Fig. 1. The leading terms for the fermion and boson corrections are of the opposite sign, thus, the quadratic divergence cancels for the case of the same masses of the boson and fermion.

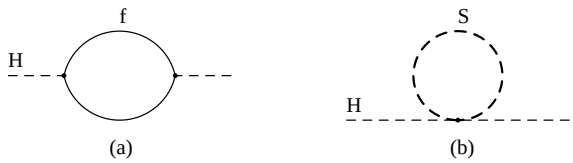

FIG. 1: SUSY Higgs mass correction diagrams.

Since the supersymmetric partners of the SM particles are not mass degenerate with their SM counterparts, a SUSY extension of the SM must have supersymmetry broken at the low energy scale. Since the nature of supersymmetry breaking is yet to be identified and there are a multitude of ways it can happen, the common approach to extend the SM with SUSY is phenomenological. The supersymmetrized SM Lagrangian prior to the electroweak symmetry breaking (EWSB) is extended by the electroweak and supersymmetry breaking terms.

The Minimal Supersymmetric Standard Model (MSSM) is the minimal phenomenological SUSY extension of the SM. In the MSSM in addition to the SM gauge bosons and fermions two Higgs doublets with hypercharge $Y = \pm 1$ are introduced, and all the particles are assigned their superpartners.[4] All renormalizeable supersymmetric interactions consistent with $B - L$ conservation (B and L are the baryon and lepton numbers) and then the most general soft-supersymmetry-breaking terms are added [1].[5] Due to $B - L$ invariance, the MSSM possesses an R-parity invariance, where $R = (-1)^{3(B-L)+2S}$ for a particle of spin S. Thus, all the SM particles have even R-parity, while the superpartners have odd R-parity. The conservation of R-parity also implies that the lightest SUSY particle (LSP) is stable and is always one of the final products of any SUSY particle decay chain. The LSP is also a prime can-

[4]The minimum of two Higgs doublets is required.

[5]*Soft* SUSY breaking implies a particular SUSY breaking scale (of the order of a few TeV). Thus, the theory should behave as unbroken SUSY above that scale.

didate for the dark matter. The MSSM can also be extended to contain the R-parity violation (RPV).

The field content of the MSSM prior to the electroweak symmetry breaking is given by the gauge and matter supermultiplets. The gauge supermultiplets contain the SM gauge bosons with their fermionic superpartners: (B_μ, λ_B), (W^i_μ, λ^i_W), and (G^a_μ, λ^a_G) for $\mathrm{U}(1)_Y$, $\mathrm{SU}(2)_W$, and $\mathrm{SU}(3)_C$ groups respectively. The matter supermultiplets are the pairs of SM matter fermions with their scalar superpartners, and Higgs scalars with their fermion superpartners. The two neutral Higgs fields couple separately to the matter (s)fermions: one Higgs (H_d) couples to charged (s)lepton fields and the down-type (s)quarks, while the other Higgs (H_u) couples to the up-type (s)quarks. After the Higgs fields gain the *vev* the particles get their masses proportional to different *vev*'s (times the Yukawa couplings). One important parameter of the model is the Higgs *vev* ratio $\tan\beta = \frac{v_u}{v_d}$, the large value ($\gtrsim 40$) of which is crucial for this analysis.

The interaction Lagrangian of the MSSM is rather complex. The part describing the interactions of gauge bosons and gauginos with themselves and with the matter supermultiplet fields does not introduce new parameters and is only dependent on the corresponding gauge group coupling constants. The supersymmetric part of the Lagrangian, describing the interactions among the matter supermultiplets is described using the superpotential, given in the MSSM by [4]

$$W_{\mathrm{MSSM}} = \bar{u}\mathbf{y_u}QH_2 - \bar{d}\mathbf{y_d}QH_1 - \bar{e}\mathbf{y_e}LH_1 + \mu H_2 H_1 \,, \tag{2.5}$$

where $\mathbf{y_{u,d}}$ are the corresponding Yukawa coupling matrices in the generation space, and μ is the Higgs mass parameter. Note that by requiring the EWSB the absolute value of the parameter μ can always be constrained and derived from other parameters

of the theory, thus, usually only sign(μ) enters as a parameters. Supersymmetry breaking terms in the Lagrangian are the gaugino and scalar mass terms, plus the trilinear couplings of the scalar fields. An extensive review of SUSY breaking terms can be found in Ref. [36]. In general both the (s)quarks and (s)leptons are allowed to mix, similar to CKM mixing in SM. After the EWSB the corresponding charge higgsinos and electroweak gauginos are allowed to mix too. The mixing in the squark and charged gaugino-higgsino sectors defines the flavor changing current properties of the model. The final particle content of the MSSM not included in SM is given in Table I [4]. A typical mass spectrum of the MSSM particles is given in Fig. 2 [4]: the first two generation squark masses are usually degenerate, as well as the slepton masses; the third generation sfermions are usually lighter, with stop (stau) being the lightest in the squark (slepton) spectrum; the neutralino is LSP, the lightest Higgs (h) is (one of) the lightest particles with a mass of 115 GeV/c^2 to 200 GeV/c^2.

The total number of independent parameters in the MSSM is 124 (including 18 from the SM). The complete Lagrangian and the full set of Feynman diagrams can be found in Ref. [37]. In its most general form, the MSSM is not a phenomenologically viable theory over most of its parameter space [1]. For a generic point there is no conservation of separate lepton numbers $L_{e,\mu.\tau}$, unsuppressed FCNCs, and new sources of CP violation inconsistent with experimental bounds. As a result, the model is viable only at very special points/regions of the full parameter space. In addition the MSSM is theoretically incomplete as it does not provide any explanation of SUSY breaking parameters and of the choice of parameters leading to EWSB.

A typical phenomenological way of reducing the number of parameters is to assume a certain universality: squark flavor mixing can be constrained to CKM (minimal flavor violation or MFV); SUSY breaking trilinear coupling and scalar mass matrices might be required to be generation independent (horizontal universality), or

TABLE I: Undiscovered particles in the MSSM. The gauge eigenstates denoted by “” are the same as the mass eigenstates.

Names	Spin	R	Mass Eigenstates	Gauge Eigenstates
Higgs bosons	0	+1	$h^0\ H^0\ A^0\ H^\pm$	$H_u^0\ H_d^0\ H_u^+\ H_d^-$
squarks	0	−1	$\widetilde{u}_L\ \widetilde{u}_R\ \widetilde{d}_L\ \widetilde{d}_R$	“ ”
			$\widetilde{s}_L\ \widetilde{s}_R\ \widetilde{c}_L\ \widetilde{c}_R$	“ ”
			$\widetilde{t}_1\ \widetilde{t}_2\ \widetilde{b}_1\ \widetilde{b}_2$	$\widetilde{t}_L\ \widetilde{t}_R\ \widetilde{b}_L\ \widetilde{b}_R$
sleptons	0	−1	$\widetilde{e}_L\ \widetilde{e}_R\ \widetilde{\nu}_e$	“ ”
			$\widetilde{\mu}_L\ \widetilde{\mu}_R\ \widetilde{\nu}_\mu$	“ ”
			$\widetilde{\tau}_1\ \widetilde{\tau}_2\ \widetilde{\nu}_\tau$	$\widetilde{\tau}_L\ \widetilde{\tau}_R\ \widetilde{\nu}_\tau$
neutralinos	1/2	−1	$\widetilde{\chi}_1^0\ \widetilde{\chi}_2^0\ \widetilde{\chi}_3^0\ \widetilde{\chi}_4^0$	$\widetilde{B}^0\ \widetilde{W}^0\ \widetilde{H}_u^0\ \widetilde{H}_d^0$
charginos	1/2	−1	$\widetilde{\chi}_1^\pm\ \widetilde{\chi}_2^\pm$	$\widetilde{W}^\pm\ \widetilde{H}_u^+\ \widetilde{H}_d^-$
gluino	1/2	−1	$\widetilde{g}$	“ ”
gravitino/goldstino	3/2	−1	$\widetilde{G}$	“ ”

better yet all these matrices can be required to be diagonal in the basis where the corresponding lepton and quark mass matrices are diagonal (flavor alignment). On the other hand one can consider the corresponding model parameters at a higher energy scale and apply constraints based on the model parameters at that scale. Inspired by the GUTs, one can require the unification of gaugino mass parameters at the GUT scale. It is most common though to derive SUSY breaking in the MSSM based on certain theoretical assumptions of the nature of the breaking. The theoretically motivated SUSY breaking scenarios available on the market are: supergravity (SUGRA), gauge mediated SUSY breaking (GMSB), anomaly-mediated (AMSB), and SUSY GUTs. In all cases a certain amount of universality at higher energy scale is either

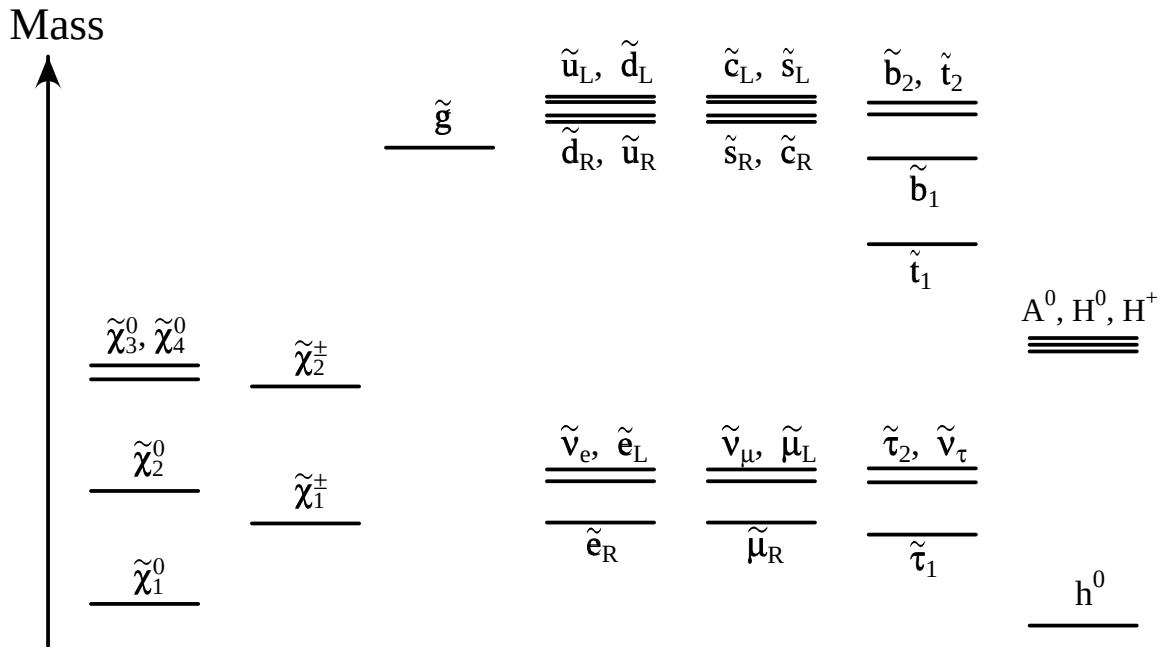

FIG. 2: A schematic of a sample MSSM spectrum.

postulated or derived from some principles [1]. The main features of mSUGRA scenario (a benchmark SUSY scenario) and SO(10) GUT sensitive to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurement are given below, including the parameters defining these models.[6]

In the minimal SUGRA (mSUGRA) scenario it is assumed that at the Plank (or GUT) scale there is a horizontal universality and a flavor alignment, namely all the scalar masses unify and are given by m_0, all the trilinear couplings are unified at A_0, and the gaugino masses are at $m_{1/2}$. After the assumption that there are no other SUSY breaking terms, the mSUGRA model is defined by 5 parameters (additional to the SM): $m_0, m_{1/2}, A_0, \mathrm{sign}(\mu), \tan\beta$. It is then possible to evolve the parameters from the unification scale to the corresponding experiment energy scale using the

[6]The dependence of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ on these parameters is discussed later in Section II.E.1.

renormalization group (RG) equations, which would define the particle spectrum and the couplings at low energy. The assumptions of EWSB and neutralino LSP usually substantially constrain the model parameter space. Further constraints come from the experimental observations [38, 39]. Although the mSUGRA scenario is rather simplistic it serves as a good benchmark model. The commonly used extensions include adding some type of non-universality to the parameters at the unification scale: gaugino mass non-universality [10, 40] or Higgs mass non-universality [41, 42]. Another extension of mSUGRA considered in this analysis is the one with RPV included. In general the RPV extension introduces all the possible RPV coupling factors, here the most relevant ones are λ'_{ijk} (λ_{ijk}) which determine the strength of RPV couplings of $\widetilde{\nu}_i$ to quarks (leptons) with flavors j and k.

SUSY GUT SO(10) model is one of the most discussed SUSY GUTs [1, 43–48]. The particular set of MSSM parameters arises from a SUSY GUT as a result of the GUT group symmetry breaking to the SM gauge group and further down to the strong interactions and electromagnetism. The SO(10) is the smallest simple Lie group for which a single irreducible representation can accommodate the entire generation, the spinor (complex) **16** representation. Thus, the whole generation can be combined in a single **16** multiplet: twelve complex quark fields (three colors, two flavors, and left and right chirality) and four complex lepton fields (charged lepton and neutrino, both left and right chirality). The corresponding squark and slepton fields are again in the same representation. Including the right-handed neutrino allows to explain the neutrino masses. Technically, the uniqueness of the SO(10) application ends here, this symmetry has to be broken to the SM group symmetry and then further (EWSB) to the $\mathrm{SO}(3)_C \times \mathrm{U}(1)_{EM}$ and there are quite a number of ways this can be done, quite often with the largest complexity in the Higgs sector [1, 43]. The typical approach is to select some kind of minimal SO(10) configuration that will have a small number of

parameters and would produce predictions consistent with experimental observations or limits [44–48]. A minimal SO(10) GUT model solution ($\mathrm{MSO_{10}SM}$) discussed in Refs. [16, 45, 46] is the most sensitive to the present analysis. In the $\mathrm{MSO_{10}SM}$ the two Higgs doublets of the MSSM reside in one **10** dimensional representation. The third generation Yukawa couplings are unified, and a hierarchical structure of the mass and mixing matrices is assumed for the first two generation. The soft SUSY breaking terms consistent with SO(10) are: a universal gaugino mass $M_{1/2}$, a universal squark and slepton mass m_{16}, a universal scalar Higgs mass m_{10}, a universal trilinear coupling parameter A_0. The requirement of Yukawa coupling unification naturally leads to $\tan\beta \approx 50$ ($m_t/m_b \sim \tan\beta$). The analysis of electroweak data in addition requires the following constraints on the parameters: $-A_0/2 \approx m_{10}/\sqrt{2} \approx m_{16}$ with $m_{16} \gg M_{1/2}, \mu$ and additionally $m_{16} \gtrsim 1$ TeV.

D. Flavor changing neutral currents

A flavor changing neutral current (FCNC) is referred to when the hadronic state changes its flavor composition without a change in charge. This can be explained by an effective coupling of the same level quarks in two generations, similar to that shown in Fig. 3. FCNCs are observed to be small, thus, any successful model must take this into account.

Historically, the suppression of the FCNC processes was an essential ingredient in the construction of the SM. Thus, the SM does not contain the FCNC processes at the tree level by construction. In the SM the unitarity (triangle) conditions[7]

$$V_{uq}V_{ub}^* + V_{cq}V_{cb}^* + V_{tq}V_{tb}^* = 0, \tag{2.6}$$

[7]This condition is specific to the b-hadron decays, a similar condition might be used for other flavors.

where $q = d$ or s are responsible for the suppression of the FCNC induced processes. For a double flavor changing process, e.g., $b \to q \to s$ as shown in Fig. 4, the amplitude is proportional to $V_{qb}V_{qs}^*$, after the contributions from $q = u, c$, and t quarks are summed up (in the limit of zero or same quark masses) the total is zero. This is referred to as GIM mechanism (Glashow-Illiopoulos-Maiani) [1, 12]. Naturally the difference in quark masses involved in the FCNC loop allows the process to happen. FCNC processes b-flavored mesons are dominated by the virtual top-quark exchange due to its large mass and large $V_{tb} \sim 1$ factor.

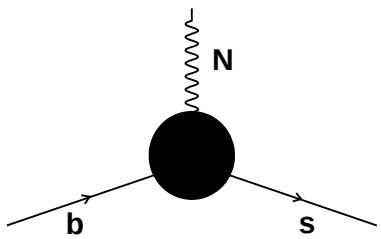

FIG. 3: Effective FCNC vertex with neutral field N.

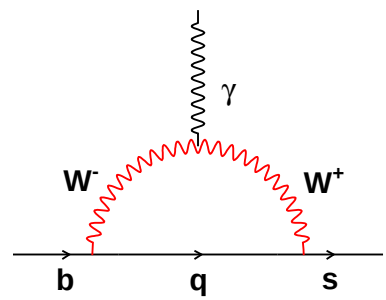

FIG. 4: Example FCNC $b \to s\gamma$ diagram.

Naturally, the suppression of FCNCs is in fact one of the most essential constraints applied to any new model beyond the SM. In fact, $b \to s\gamma$ transition observed by CLEO in 1993 [49] is one of the top golden FCNC modes. The improvements in both experimental measurements and the SM predictions of the mode allowed to constrain a lot of new physics models. Unfortunately, for the past few years both experiment and theory are dominated by the systematic uncertainties with the progress of improving those being rather slow [50, 51].[8] The $B_s^0 \to \mu^+\mu^-$ decay, on the other

[8]Theoretical predictions of FCNC induced decays (and any hadron decay in general) usually have significant uncertainties because the processes that involve hadrons

hand, has a much cleaner experimental signature and is much easier to describe theoretically [10, 13, 21, 52, 53]. Although not observed the improved upper limit on its rate becomes increasingly restrictive to a number of models, especially SUSY models.

E. Rare $B^0_{s(d)} \to \mu^+\mu^-$ decays

This section describes the general approach in calculating the branching ratio of the $B^0_{s(d)}$ decays to a dilepton final state, which is followed by a detailed discussion of the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ in the SM and SUSY. The expectation for $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ in SUSY models begins by identifying the Feynman diagrams corresponding to the processes dominating the decay. This is followed by a discussion of the sensitivity to these processes in specific SUSY models in terms of the model parameters. Thus, relations between the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ measurement and the model parameters are established, which can be used to exclude regions of SUSY parameter space once the measurement becomes available.

The branching ratio[9] for the decay of $B^0_{s(d)}$ to a dilepton final state is given by

$$\mathcal{B}(B \to l^+l^-) = \frac{1}{16\pi}\frac{\tau_B}{m_B}\sqrt{1-\left(\frac{2m_l}{m_B}\right)^2}|\mathcal{M}|^2, \tag{2.7}$$

where a dilepton (l^+l^-) final state is assumed for a meson B of mass m_B and lifetime τ_B and a lepton mass m_l. The matrix element $\mathcal{M}$ is defined by a particular model. Unless the decay amplitude is dominated by a tree diagram, the number of contributing loop diagrams is usually large. Because of this it is common to con-

in initial and/or final states are non-perturbative by nature. In fact most of the present theoretical calculations of the observables are limited by the systematic uncertainties that arise from the QCD-related effects. Lattice QCD methods would at some point be able to describe such processes from the first principles, but the current state of the theoretical methods requires a certain amount of phenomenology involved to describe processes with hadrons.

[9]Decay width divided by the total width $\Gamma_{\text{total}} \equiv 1/\tau$.

sider the decay in terms of an effective Lagrangian in order to analyze the complete decay in the first order of this effective perturbation theory [13]. Furthermore, it is standard to consider the effective Lagrangian as a sum of external field operators expanded in phenomenological terms, the operator product expansion $\mathcal{L}_{\text{eff}} = \sum C_i \mathcal{O}_i$, where C_i are the Wilson coefficients. The operators $\mathcal{O}_i$ are (esp. in the case of final state leptons) the Lorentz scalars which are the result of just a product of initial state operator times the final state operator, e.g., scalar×scalar (SS), vector×vector (VV) etc., separately for left-handed (L) and right-handed (R) combinations. For $B^0_{s(d)} \to \mu^+\mu^-$ decays the main contributions are SS and VV contributions from the possible L and R combinations.

The branching ratio for the $B^0_{s(d)} \to \mu^+\mu^-$ decay in terms of Wilson coefficients is given by [52, 53]:

$$\mathcal{B}(B^0_{s(d)} \to l^+l^-) = \frac{\tau_B m_B}{16\pi}\sqrt{1-\left(\frac{2m_l}{m_B}\right)^2}\left\{|a|^2 + |b|^2\left(1 - 4\frac{m_l^2}{m_B^2}\right)\right\}, \tag{2.8}$$

where the coefficients a and b in terms of Wilson coefficients are given by

$$\begin{aligned} a &= \frac{f_B}{4}\left\{2m_l[C^V_{LL} - C^V_{LR} + C^V_{RR} - C^V_{RL}] - \frac{m_B^2}{m_b}[C^S_{LL} - C^S_{LR} + C^S_{RR} - C^S_{RL}]\right\}, \\ b &= \frac{f_B}{4}\frac{m_B^2}{m_b}[C^S_{LL} + C^S_{LR} - C^S_{RR} - C^S_{RL}], \end{aligned} \tag{2.9}$$

where $f_B \approx 0.25$ GeV is the B-meson decay constant, C^Z_{XY} are the Wilson coefficients [52]. The decay constant is a non-perturbative value that arises from the matrix elements of the quark operators between the initial state and the vacuum: $\langle 0|\bar{b}\gamma^\mu\gamma^5 s|B^0_s(p)\rangle = -\imath f_{B_s} p^\mu$, where p is the initial meson momentum. As it can be seen from Eq. (2.9), the contributions from the vector operators is proportional to the mass of the lepton – this is the chirality suppression effect for the pseudo-scalar meson decays. In practice that means that vector exchange contributions to this

decay are proportional to lepton mass squared and are substantially suppressed for $B^0_{s(d)} \to e^+e^-$ compared to the $B^0_{s(d)} \to \mu^+\mu^-$ decays by a factor of $\sim 2.5 \times 10^{-5}$. Obviously, the VV contributions to the decays $B^0_{s(d)} \to \tau^+\tau^-$ will be enhanced by a factor of about 280 compared to $B^0_{s(d)} \to \mu^+\mu^-$, but substantial experimental limitations[10] make the final ditau state less interesting. The contribution from the scalar particle exchange does not have the inherent suppression and can potentially give a substantial enhancement depending on the model parameters [54], which in particular is the case in SUSY models.

1. *Standard Model expectations*

In the SM the $B^0_{s(d)} \to \mu^+\mu^-$ process is dominated by virtual gauge boson exchanges [13, 52]. As shown by the diagrams in Fig. 5, these include the flavor changing self-energy (a), penguin Z^0-exchange (b), and box $W^\pm$-exchange (c) diagrams. The quark entering in the loop is either u, c or t-quark. As mentioned in Section II.D, due to the CKM unitarity conditions, large m_t, and the large value of V_{tb}, only the top-quark contribution is important. The scalar Higgs exchange is suppressed by the small Yukawa coupling and is negligible in SM. Note that there is no contribution from photon exchange, which is a pure vector exchange, which makes the $B^0_{s(d)} \to \mu^+\mu^-$ decay different from the $b \to s\gamma$ related decays. The calculation of each particular diagram is similar to that performed in Ref. [55] made for $K_L \to \mu^+\mu^-$ decay.

[10] The main limitation is that $D^{\pm,0}$ decays look very much like $\tau^\pm$ decays.

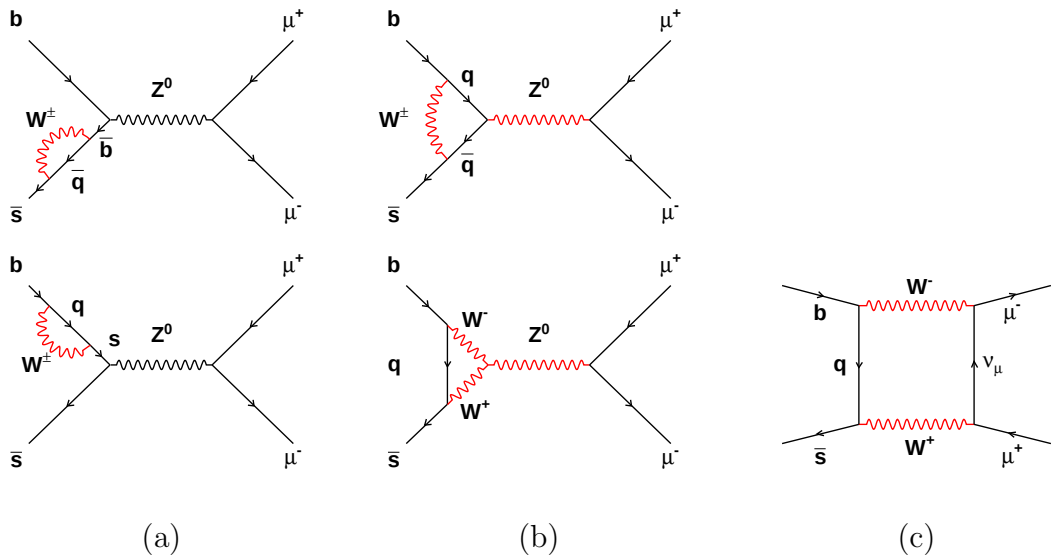

FIG. 5: Standard Model Feynman diagrams contributing to $B_s^0 \to \mu^+\mu^-$.

The total contribution from the SM in terms of observables with substantial measurement uncertainties is given by [13, 56]

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 3.5 \cdot 10^{-9} \left[\frac{\tau_{B_s}}{1.6 \text{ ps}}\right] \left[\frac{f_{B_s}}{210 \text{ MeV}}\right]^2 \left[\frac{|V_{ts}|}{0.04}\right]^2 \left[\frac{\overline{m_t}}{170 \text{ GeV}/c^2}\right]^{3.12}, \tag{2.10}$$

where $\overline{m_t}$ is the running top-quark mass in $\overline{MS}$ scheme taken at the pole. Incorporating the uncertainties on the parameters, as available in the Refs. [13, 56] the SM expectation is $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.4 \pm 1.2) \times 10^{-9}$. Using more recent values of the input parameters available in Refs. [1, 26, 28] the expected value is $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (5.1 \pm 1.1) \times 10^{-9}$, where the values $\tau_{B_s} = (1.47 \pm 0.06)$ ps, $\overline{m_t} = 167.5 \pm 4$ GeV/c^2, $|V_{ts}| = (41.6 \pm 0.6) \times 10^{-3}$, and $f_{B_s} = 260 \pm 27$ MeV are used. Nevertheless, the dominant uncertainty still comes from the value of the meson decay constant. The value for $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ can be obtained from Eq. (2.10) after replacing the lifetime and decay constant for B_s^0 with corresponding values for B_d^0

and replacing V_{ts} with V_{td}. In the case of $B_d^0 \to \mu^+\mu^-$ though it is possible to use the measured $B^0 - \overline{B^0}$ meson mass splitting to substantially reduce the uncertainty on the expected value of the branching ratio [56]. Following the Ref. [56] the expected SM branching ratio is $\mathcal{B}(B_d^0 \to \mu^+\mu^-) = (1.00 \pm 0.14) \times 10^{-10}$, which is a substantial improvement in the prediction uncertainty. Unfortunately the corresponding $B_s^0 - \overline{B_s^0}$ meson mass difference direct measurement is not available.

2. SUSY expectations

The models most sensitive to $B_s^0 \to \mu^+\mu^-$ are discussed in this section, following a brief overview of the sensitivity to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in other models. Feynman diagrams for the processes dominating the $B_s^0 \to \mu^+\mu^-$ decay are identified first in order to determine the sensitivity to the branching ratio in terms of the properties of particles entering the diagrams. The allowed parameters of SUSY models can then be related to $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ via these diagrams. The most sensitive models are those with R-parity violation and those with large $\tan\beta$ and relatively small heavy Higgs mass. In the latter case the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is proportional to $\tan^6\beta/m_A^4$, where m_A is the heavy Higgs mass, and can be as large as $\sim 1 \times 10^{-6}$.

There has been a substantial interest in the $B_s^0 \to \mu^+\mu^-$ decay in the past years mostly related to SUSY. Partly due to that interest, the $B_s^0 \to \mu^+\mu^-$ decay branching ratio was considered in a substantially extensive set of models available on the market. As it is anticipated, during Tevatron Run II it would be possible to probe the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ down to the level of 10^{-8} [10], which is about an order of magnitude above the SM expectation. Since the measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ at the SM level is not expected until about 2010 to be measured by the LHC experiments [57–60], the theoretical models that gained the most interest are those the most sensitive to the $B_s^0 \to \mu^+\mu^-$ decay at the level accessible by Tevatron. Certainly

the theoretical attraction of those models was the cause of that interest, the sensitivity to the $B_s^0 \to \mu^+\mu^-$ being just a symptom of those particular models. The models beyond the SM discussed in Section II.B have all being considered with respect to the $B_s^0 \to \mu^+\mu^-$. No significant enhancement over the SM value is suggested by the technicolor [61], universal extra dimensions [62, 63], and GMSB supersymmetry [63]. Although the enhancement from these models is usually not more than a factor of two, it will become substantial for the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurement once the experimental sensitivity improves as expected to happen at LHC experiments.

Overall, the models with vector boson exchange dominating the $B_{s(d)}^0 \to \mu^+\mu^-$ process do not exhibit a substantial enhancement in branching ratio over the SM value. In addition, due to the chirality suppression of the decay in such models the $b \to s\gamma$ and $B \to X_s\mu^+\mu^-$ provide a better sensitivity. Thus, in general one would expect that $B_s^0 \to \mu^+\mu^-$ would be important in cases where the scalar exchange dominates [54]. The models where this is the case are RPV SUSY models, where the decay proceeds at tree level, and the MSSM with large $\tan\beta$, where the decay proceeds via a loop diagram. The dominant SUSY diagrams in these models are discussed next, followed by a discussion of the sensitivity to these diagrams in the RPV mSUGRA and in a set of MSSM models discussed in the end of Section II.C.

a. Dominant diagrams in SUSY

Depending on the model, either tree or loop diagram can dominate the $B_s^0 \to \mu^+\mu^-$ decay. The case with tree level contribution in RPV SUSY models is discussed first. A contribution to $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, proportional to $\tan^4\beta$ and common to all models with two Higgs doublets is described next. An additional contribution to the branching ratio, proportional to $\tan^2\beta$ and explained by a chirality flip in the flavor-changing self-energy insert, specific to MSSM, is discussed after that. These two contributions

combined determine the dominant loop diagram in SUSY scenario, which gives a contribution to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ proportional to $\tan^6\beta$. The treatment of the $B_d^0 \to \mu^+\mu^-$ decay is similar.

In the RPV SUSY [10, 39] the $B_s^0 \to \mu^+\mu^-$ decay proceeds via sneutrino exchange in a tree diagram as shown in Fig. 6, where λ'_{ijk} (λ_{ijk}) determines the strength of RPV couplings of $\widetilde{\nu}_i$ to quarks (leptons) with flavors j and k. Assuming that the couplings λ'_{i23} and λ_{i22} are the same for all i and are equal to λ after simple calculations the expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is given by

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \approx 2 \times 10^8 \lambda^4 \left(\frac{m_{B_s}}{m_{\tilde{\nu}}}\right)^4, \tag{2.11}$$

where the sneutrino mass is considered approximately equal in all three generations. Depending on the specifics of the RPV model related to the quark mixing at the GUT scale, the present constraints on the λ' parameters can be rather strong or rather weak, with the relevant parameter range possibly as high as 0.3 [64].

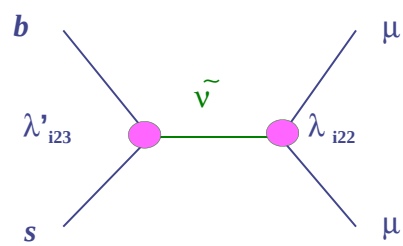

FIG. 6: The leading tree diagram in RPV mSUGRA.

Most of the models with a substantial enhancement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ at the loop level are those with more than one (usually just two) Higgs doublets. About the same time that the first experimental limits on the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ made by UA1 in the late 80's [65, 66] the first papers suggesting an enhancement to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in

the two Higgs doublet models (2HDMs) had come out [67, 68]. It is by the nature of the 2HDMs that the couplings can be enhanced by a factor $\tan\beta \equiv v_2/v_1$ [69], which was one of the reason these models were originally introduced in order to suggest an experimentally observable contribution from the Higgs sector. In a particular case of type-II 2HDM in which the Higgs content and Yukawa couplings are identical to that of the MSSM the Feynman diagram set can be easily derived from that in Figs. 5 by replacing the vector boson lines with the Higgs boson lines of the corresponding charge. The diagrams with H, A exchange are giving the dominant contribution due to the $\tan\beta$ enhancement on the corresponding vertices. The contribution to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is proportional to $\tan^4\beta$ [70]:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \approx 5 \times 10^{-9} \left[2\mathcal{X}(\mathcal{X}-1)+1\right], \tag{2.12}$$

where $\mathcal{X} \equiv \kappa \tan^2\beta \frac{\ln r_{H^+}}{r_{H^+}-1}$, $r_{H^+} \equiv m_{H^+}^2/m_t^2$, and $\kappa \tan^2\beta \equiv \frac{m_{B_s}^2}{8m_W^2}\tan^2\beta \approx 1.3\left(\frac{\tan\beta}{50}\right)^2$. Taking into account the constraint from $b \to s\gamma$, with the sensitivity of 1×10^{-8} the Tevatron reach is only for the values of $\tan\beta \gtrsim 60$. By construction of the type-II 2HDM, the same conclusion can be made about the MSSM in the case that all SUSY particles do not contribute which can easily happen in the limit of large SUSY masses.

One substantial enhancement to the rate of the $B_{s(d)}^0 \to \mu^+\mu^-$ decay in the MSSM comes from the self-energy FCNC diagrams that allow a chirality flip of a B-meson constituent quark. If the chirality flip is allowed in the self-energy diagram during the loop integration, the contribution from m_f from the internal fermion propagator $(\not{p}_f + m_f)$ is allowed in addition to that from p. The net contribution from p at best gives a value around the meson mass, compare that to m_f, which can give an enhancement to the amplitude of up to the value of m_f/m_B ratio. In the MSSM the flip is possible due to the presence of $\tilde{t}_{L,R}$ mass mixing present as a result of the

large Yukawa top coupling. The mass eigenstates are $\tilde{t}_{L,R}$ rotated by an angle θ_t. The self-energy diagram with chirality flip is the one with stop and chargino (the fermion) virtual exchange. It is proportional to $\mu \tan\beta \sin 2\theta_t$ [10], where μ expresses the (dimensional) relation to $m_{\chi_1^\pm}$. In the case of large $\tan\beta$ the enhancement can become significant even with a relatively small stop mixing. Combining this with the $\tan^2\beta$ enhancement due to the Higgs exchange, it is clear that at the large $\tan\beta$ the diagram in Fig. 7 gives rise to a branching ratio proportional to $\tan^6\beta$. With a good choice of parameters this can give as much as 3 to 4 orders of magnitude enhancement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ over the SM values [15]. Analytically the dominant contribution from the MSSM Higgs exchange-induced diagrams for $B_s^0 \to \mu^+\mu^-$ is given by [10, 52]

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) \approx 5 \times 10^{-9} \left[1 + 2\mathcal{X}(\mathcal{X}-1) + 2\mathcal{Y}\left\{\mathcal{Y} - \mathrm{sign}(\mu)(2\mathcal{X}-1)\right\}\right], \quad (2.13)$$

where

$$\mathcal{Y} \approx \kappa \tan^3\beta \frac{|\mu| m_t}{m_A^2} \frac{\sin 2\theta_t}{2} \left(\frac{r_{\tilde{t}_1}}{r_\mu - r_{\tilde{t}_1}} \ln \frac{r_{\tilde{t}_1}}{r_\mu} - \frac{r_{\tilde{t}_2}}{r_\mu - r_{\tilde{t}_2}} \ln \frac{r_{\tilde{t}_2}}{r_\mu} \right), \quad (2.14)$$

$\mathcal{X}$ and κ are defined in Eq. (2.12), $r_\mu \equiv |\mu^2|/m_t^2$, and $r_{\tilde{t}_i} \equiv m_{\tilde{t}_i}^2/m_t^2$. Note that as in case of Eq. (2.12) the numerical coefficient in front of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in Eq. (2.13) is in fact the SM expectation value. Equation (2.13) in practice gives a reasonably good approximation of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in the MSSM in case of minimal flavor violation. The sensitivity of a particular model to the $B_s^0 \to \mu^+\mu^-$ depends on the size of the parameters defining the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The relation between the model parameters and the expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ value is discussed in the next section.

The SUSY diagrams for the $B_d^0 \to \mu^+\mu^-$ decay are similar. As mentioned earlier in the section, for the models with MFV where $B_{s(d)}^0 \to \mu^+\mu^-$ proceeds via a loop diagram, the expectation for $B_d^0 \to \mu^+\mu^-$ is smaller by a factor of $|V_{ts}/V_{td}|^2 \approx 40$ than

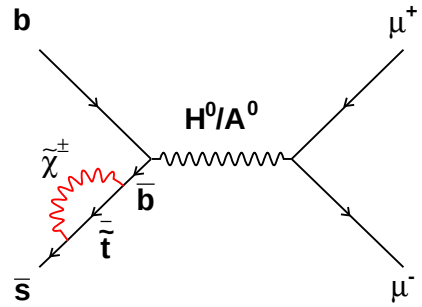

FIG. 7: Dominant diagram for $B^0_{s(d)} \to \mu^+\mu^-$ in the MSSM.

that of $B^0_s \to \mu^+\mu^-$. Since presently the experimental situation is such that the best limit on $\mathcal{B}(B^0_{s(d)} \to \mu^+\mu^-)$ comes from the hadron collider experiments, one would not expect the limit for $\mathcal{B}(B^0_d \to \mu^+\mu^-)$ to be better than by a factor of $f_d/f_s \approx 4$ [1]. Thus, the limit for $\mathcal{B}(B^0_d \to \mu^+\mu^-)$ from the hadron colliders will be less sensitive to the model parameters by about a factor of 6. The situation for $\mathcal{B}(B^0_d \to \mu^+\mu^-)$ can be different though in the case of RPV scenario.

b. Sensitivity to SUSY model parameters

SUSY models most sensitive to the $B^0_{s(d)} \to \mu^+\mu^-$ decay are discussed in this section. As discussed in the previous section, either tree or loop diagrams can give the dominant contribution to the decay rate. In case the FCNC is allowed at the tree level, the natural SUSY candidate is the model with RPV, whereas at the loop level the largest contribution to the branching ratio of $B^0_{s(d)} \to \mu^+\mu^-$ would come from SUSY models with large $\tan\beta$. The benchmark models discussed below are considered sensitive to the presented analysis: the RPV mSUGRA scenario [10, 64], the mSUGRA [10, 71], the SUGRA scenario with Higgs non-universality [42, 72], and the SUSY SO(10) model [48, 73]. Appendix A describes a potential sensitivity to

a measurement of m_A in the general MSSM scenario with large $\tan\beta$ in the case of observation of the $B_s^0 \to \mu^+\mu^-$ decay.

In the RPV mSUGRA the dominant contribution comes from the sneutrino exchange with $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ defined by the RPV trilinear couplings λ, λ' and the sneutrino mass, $m_{\tilde{\nu}}$. For simplicity, the couplings can be assumed equal. In mSUGRA $m_{\tilde{\nu}}$ is related to the model parameters by a simple relation $m_{\tilde{\nu}}^2 \approx m_0^2 + 0.54 m_{1/2}^2 + 0.5 \cos 2\beta m_Z^2$, which means that the sneutrino mass is mostly only dependent on the m_0 and $m_{1/2}$. The curves that relate the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ to λ in the $(m_0, m_{1/2})$ plane are given in Figs. 8 and 9 [74]: in both figures the red dashed line denotes the expected light Higgs mass and the red hatched region corresponds to the $m_h < 114$ GeV/c^2 as excluded by LEP [24]. The plot in Fig. 8 can be interpreted as $(m_0, m_{1/2})$ exclusion regions for a given value of λ – the excluded region being above the curve. The experimental perspective can be gained from the plot in Fig. 9: for the given value of λ the future expected upper limits on $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ would define the exclusion region in $(m_0, m_{1/2})$ space gradually moving up with improved limit measurements.

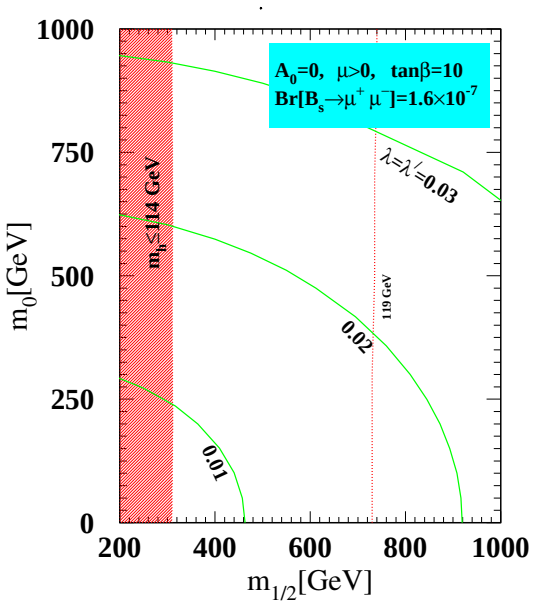

FIG. 8: RPV mSUGRA allowed λ dependence in $(m_0, m_{1/2})$ plane for a fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = 1.6 \times 10^{-7}$.

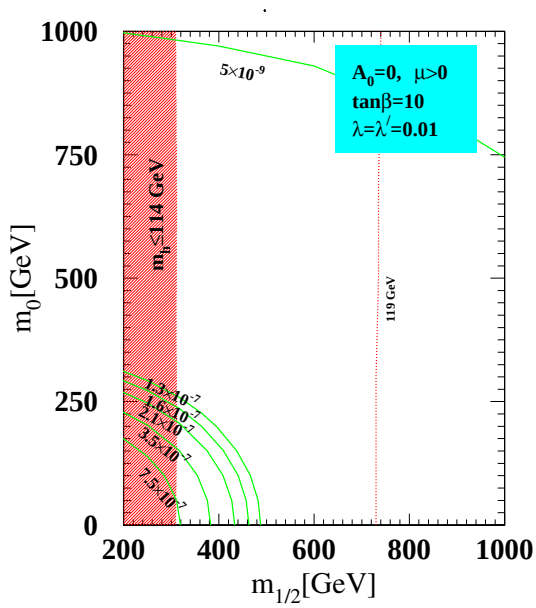

FIG. 9: RPV mSUGRA $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ dependence in $(m_0, m_{1/2})$ plane for a fixed λ.

In mSUGRA (R-parity conserving case) the loop diagrams give a $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ that is proportional to $\tan^6\beta \sin^2(2\theta_t)/m_A^4$. The relation between θ_t or m_A and the mSUGRA parameters is not so trivial (compared to the simplicity in RPV case). One has to solve the set of RG equations to extract the inputs to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ at the low energy scale based on the mSUGRA input parameters. The curves of constant $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ are shown in Figs. 10 through 13. The plot in Fig. 10 [75] shows the constant $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ (black solid), muon anomaly due to mSUGRA $\delta a_\mu \equiv a_\mu^{\rm exp} - a_\mu^{\rm SM}$ scaled by 10^{10} (green dashed), and a constant light Higgs mass (red dot-dashed) in the $(m_0, m_{1/2})$ plane for $\tan\beta = 50$. The red filled region corresponds to the theoretically unallowed parameters: either with no EWSB or neutralino being not an LSP. A similar plot only in the $(\tan\beta, m_0)$ at $m_{1/2} = 300$ GeV is given in

Fig. 12 [75]. Note that in mSUGRA the $b \to s\gamma$ measurement sets a rather strong constraint on the parameter space. Accidentally at large $\tan\beta$ it almost overlaps with the $m_h < 115$ GeV/c^2 region [10, 75]. The mSUGRA parameter space is further substantially constrained by the cosmological observation (mostly from WMAP data). The picture corresponding to the one in Fig. 10 but with the $b \to s\gamma$ excluded (green filled) and WMAP allowed (red filled) regions is given in Fig. 11 [10]. A certain amount of interest has been directed to the a_μ measurement by $(g-2)$ collaboration [30, 75]. Although the measured value differs from the SM expectations by almost 3σ the possible systematics in the SM prediction renders this difference less significant [31, 76–78]. Assuming $\delta a_\mu > 11$ which is the range starting about 1σ below it's central value, it is possible to constrain the mSUGRA space from above, as shown in Fig. 11. This makes the search for $B_s^0 \to \mu^+\mu^-$ complementary to the measurement of a_μ. An additional complementarity exists between $B_s^0 \to \mu^+\mu^-$ and the trilepton searches, now between the high and low $\tan\beta$ values. The trilepton signature is the golden supersymmetry mode of the chargino-neutralino pair production. At large $\tan\beta$ the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ increases, while the sensitivity to the trileptons decreases due to the decreasing branching ratio of $\widetilde{\chi}^{\pm,0}$ to electrons and muons. At the Tevatron the complementarity is possible at reasonably low chargino and neutralino masses, as shown in Fig. 13 [71]. Shown on this plot are the constant $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ curves (black solid), the constant $\delta a_\mu \times 10^{10}$ (dashed), and the constant m_h (dotted). The red filled region corresponds to the no-EWSB or neutralino not an LSP regions. The magenta, blue, and green filled regions represent the expected Tevatron experiments reach at 30 fb^{-1}, 10 fb^{-1}, and 2 fb^{-1} of integrated luminosity respectively.

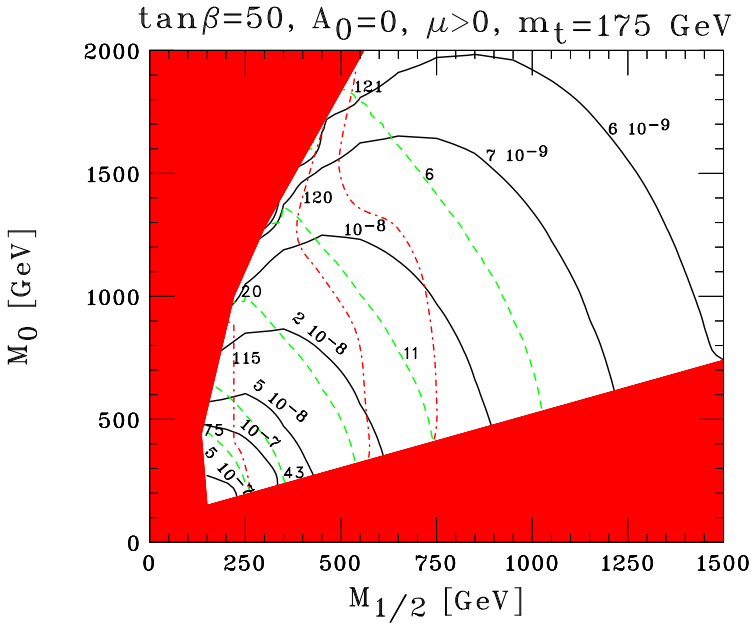

FIG. 10: Fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ contours in mSUGRA in $(m_0, m_{1/2})$ plane.

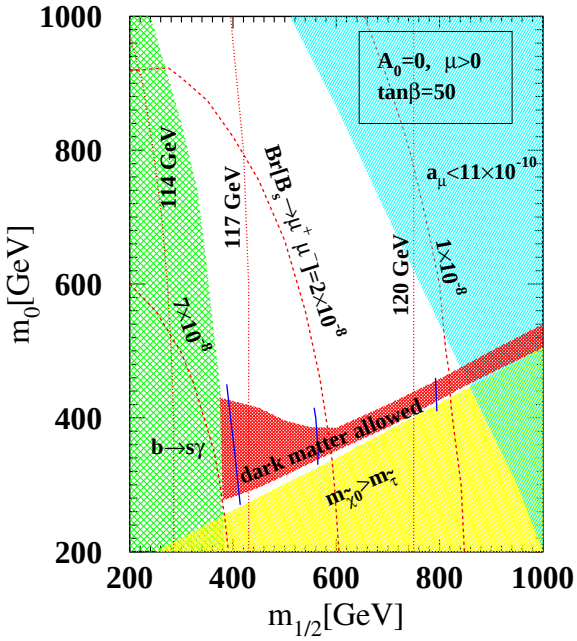

FIG. 11: Fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ contours in mSUGRA in $(m_0, m_{1/2})$ plane with experimental constraints included.

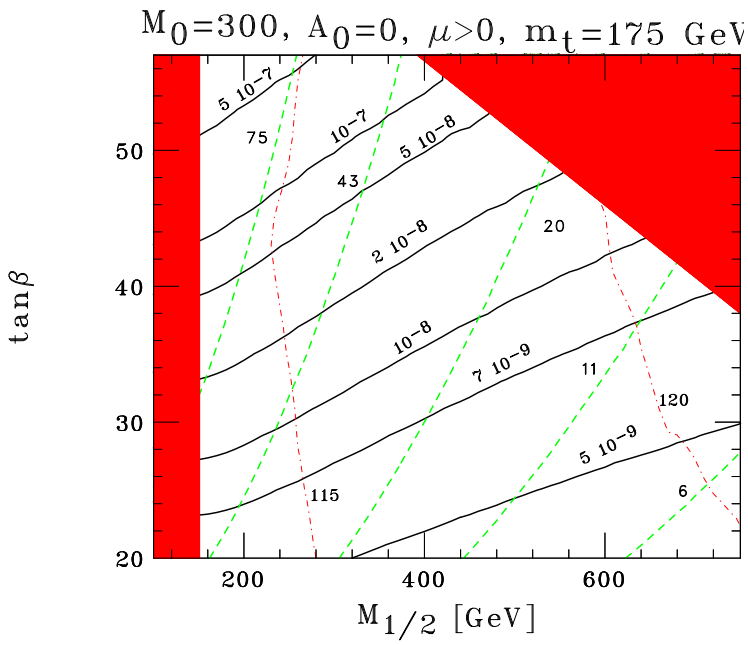

FIG. 12: Fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ contours in mSUGRA in $(\tan\beta, m_0)$ plane.

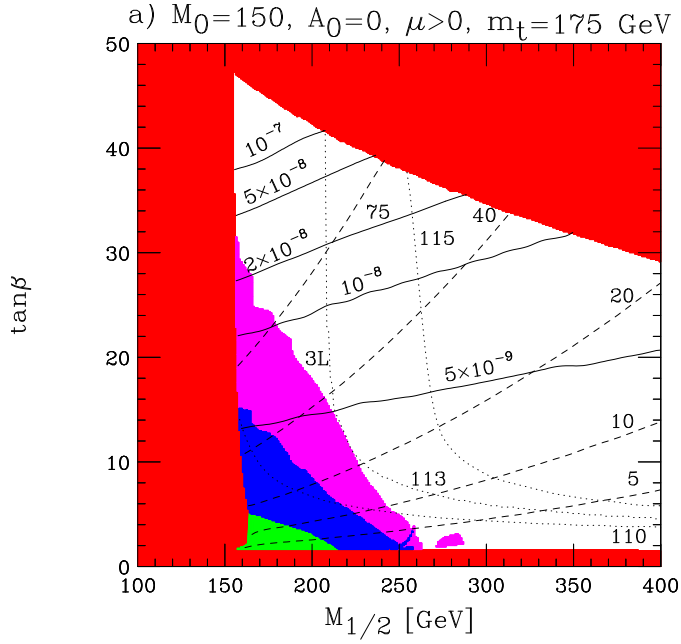

FIG. 13: Fixed $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ contours and Tevatron trilepton reach in mSUGRA in $(\tan\beta, m_0)$ plane.

In the SUGRA scenario with Higgs non-universality the same proportionality to $\tan^6\beta \sin^2(2\theta_t)/m_A^4$ holds, but it is possible in this case to have a lower heavy Higgs mass by a proper choice of parameters, which in turn enhances the value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ [42, 72]. The Higgs masses at the unification scale are selected as $m_{H_u}^2 = m_0^2(1+\delta_{H_u})$ and $m_{H_d}^2 = m_0^2(1+\delta_{H_d})$. A successful choice of parameters can give a value of the branching ratio an order of magnitude (or more) larger than that in mSUGRA, as shown in Fig. 14 [72]. The plots show the sets of points for various $\tan\beta$ values corresponding to the given $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and neutralino-dark-matter (NDM) on proton cross section $\sigma_{\tilde{\chi}_1^0 p}$: the red, green, and black dots correspond to the neutralino density within $0.095 < \Omega_{\tilde{\chi}} h^2 < 0.13$, below, and above the region allowed by the WMAP data respectively. The distribution for non-universal SUGRA is made for $\delta_{H_u} = -1$ and $\delta_{H_d} = 1$. The representation in the plane of $(\mathcal{B}(B_s^0 \to \mu^+\mu^-), \sigma_{\tilde{\chi}_1^0 p})$ shows the significant correlation between the two values. It also suggests that with respect to the model with Higgs non-universality it is possible at large $\tan\beta$ to obtain stricter constraints on the model by measuring the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ compared to the direct dark matter search provided the present limits are $\sigma_{\tilde{\chi}_1^0 p} \lesssim 1 \times 10^{-7}$ pb and $\mathcal{B}(B_s^0 \to \mu^+\mu^-) \lesssim 2 \times 10^{-7}$ [11, 72]. A similar analysis shows that $B_s^0 \to \mu^+\mu^-$ measurement is more restrictive then the present experimental sensitivity to the muon flux from the neutralino annihilation [79].

In SO(10) SUSY models, which have the same $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ dependence on $\tan\beta$ and m_A, large values of $\tan\beta$ ($\tan\beta \sim 50$) are naturally suggested by virtue of the Yukawa couplings unification requirement. The minimal SO(10) model solution with Yukawa couplings unification constraint is somewhat controversial. In particular, one group had found a solution ($\mathrm{MSO_{10}SM}$) that allows a substantial enhancement to $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ up to within 1×10^{-6} which is well within the reach of the Tevatron [16, 73], the solution being a region of parameter space allowed by the

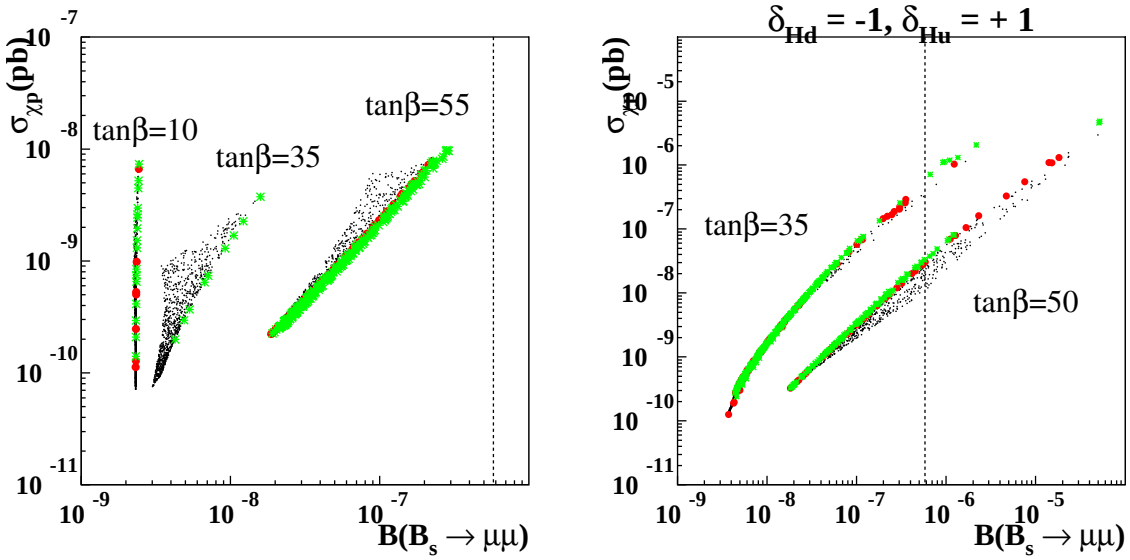

FIG. 14: Correlation between $\sigma_{\tilde{\chi}_1^0 p}$ and $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ in mSUGRA (left) and SUGRA with non-universal Higgs masses (right).

experiment and consistent with Yukawa couplings unification. This solution could not be confirmed by another group [48]. Despite the argument that the $\mathrm{MSO_{10}SM}$ solution is not confirmed independently, the fact that it is well within experimental reach makes it only more interesting as being possible to falsify. The plot in Fig. 15 [73] shows the allowed parameter space in $(\mu, m_{1/2})$ plane for $m_{16} = 3$ TeV and $m_A = 700$ GeV: the magenta filled area is excluded by the WMAP, LEP, or $b \to s\gamma$ data, the green narrow band is consistent with the WMAP data within 2σ assuming the dark matter is due to the neutralinos LSP, the steel-blue region corresponds to $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2 \times 10^{-7}$, the blue dashed lines are the constant $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ lines, and the magenta contours correspond to the constant χ^2 of the fit of the given parameter point to a set of observables (strong and electroweak couplings, gauge boson and the third generation fermion masses etc.). This particular $\mathrm{MSO_{10}SM}$ solution does not allow the m_A to be higher than about 1.5 TeV. Shown in the Fig. 16 [73] is the

allowed region of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ values for the given m_A, where the envelope[11] of the magenta open circles (dark blue crosses) limits the allowed branching ratio for $m_{16} = 5$ TeV (3 TeV). For a given branching ratio limit this can easily be converted into the limit on m_A, e.g., for $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2 \times 10^{-7}$ one gets $m_A > 500$ GeV/c^2.

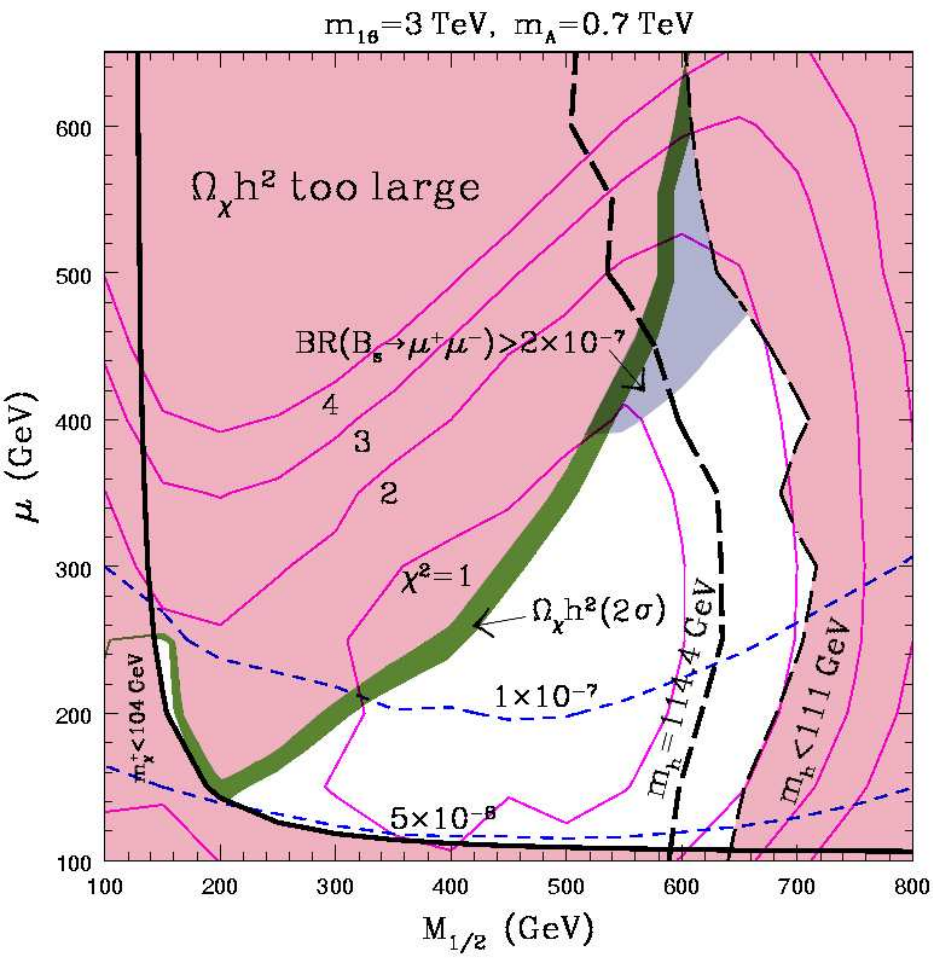

FIG. 15: Allowed region in $\mathrm{MSO_{10}SM}$ in $(\mu, m_{1/2})$ plane.

[11]The solid lines are not the part of original plot and are shown here to guide where the envelope is.

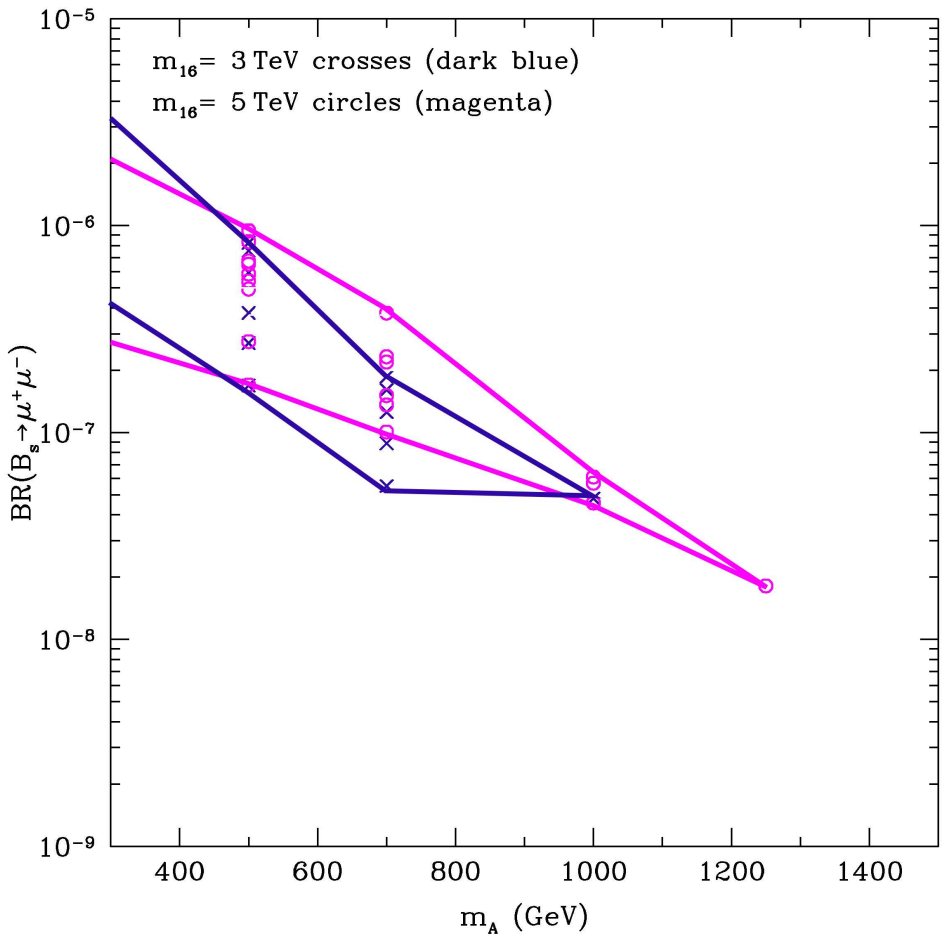

FIG. 16: Allowed range of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ for the given m_A in MSO$_{10}$SM.

F. *B*-hadron production and decay at hadron colliders

The present analysis relies on B-meson production in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV and studies the rare $B^0_{s(d)} \to \mu^+\mu^-$ decays based on the properties of B-mesons determined by their dominant decay modes. B-meson production process in high energy collision is considered as a result of b-quark production, followed by the stage of fragmentation and hadronization. Detection of B-mesons in an experiment is based on their production and decay properties.

The properties of high energy collisions can be described in terms of the parton model [80–83]. According to the parton model the nucleons (neutron and proton) in a high energy collision process can be considered as an ensemble of free point-like particles, partons. In terms of the SM, based on the concept of asymptotic freedom, the proton in a high energy collision can be viewed as consisting of free quarks and gluons, which are the partons in the parton model. The production of the final state particles can then be described as a result of a single parton-parton interaction, hard scattering. For the hadronic final states for the final state energies below about 5% of the $p\bar{p}$ collision energy the production is dominated by the $gg \rightarrow X$ processes [1].

The production of the heavy flavor quarks (b or c) can proceed via three following processes: flavor creation, flavor excitation, and gluon splitting. Flavor creation corresponds to $gg \rightarrow b\bar{b}$ or $q\bar{q} \rightarrow b\bar{b}$ processes and flavor excitation corresponds to $gb \rightarrow gb$ or $qb \rightarrow qb$ processes Gluon splitting corresponds to higher order QCD processes with a "final" state gluon splitting to a $b\bar{b}$ pair, e.g., $q\bar{q} \rightarrow g(g \rightarrow b\bar{b})$. These three processes correspond to substantially different final state topologies if one of the heavy flavor quarks is required to have a large momentum when viewed in the plane transverse to the collision axis: a flavor creation process would correspond to two heavy flavor quarks going in roughly opposite directions both having a large transverse momentum; the flavor excitation process would correspond to only one heavy flavor quark having a large transverse momentum; and the gluon excitation would correspond to the signature with both heavy flavor quarks going in roughly the same direction. The cross section of the b-hadron production at the Tevatron as measured by CDF is shown in Fig. 17, and is discussed in Refs. [84, 85]: the production cross section is maximal at transverse momentum (p_T) around 3 to 4 GeV/c with the total cross section of about 20 nb.

The stage following the "free" quark production, where the final state hadrons

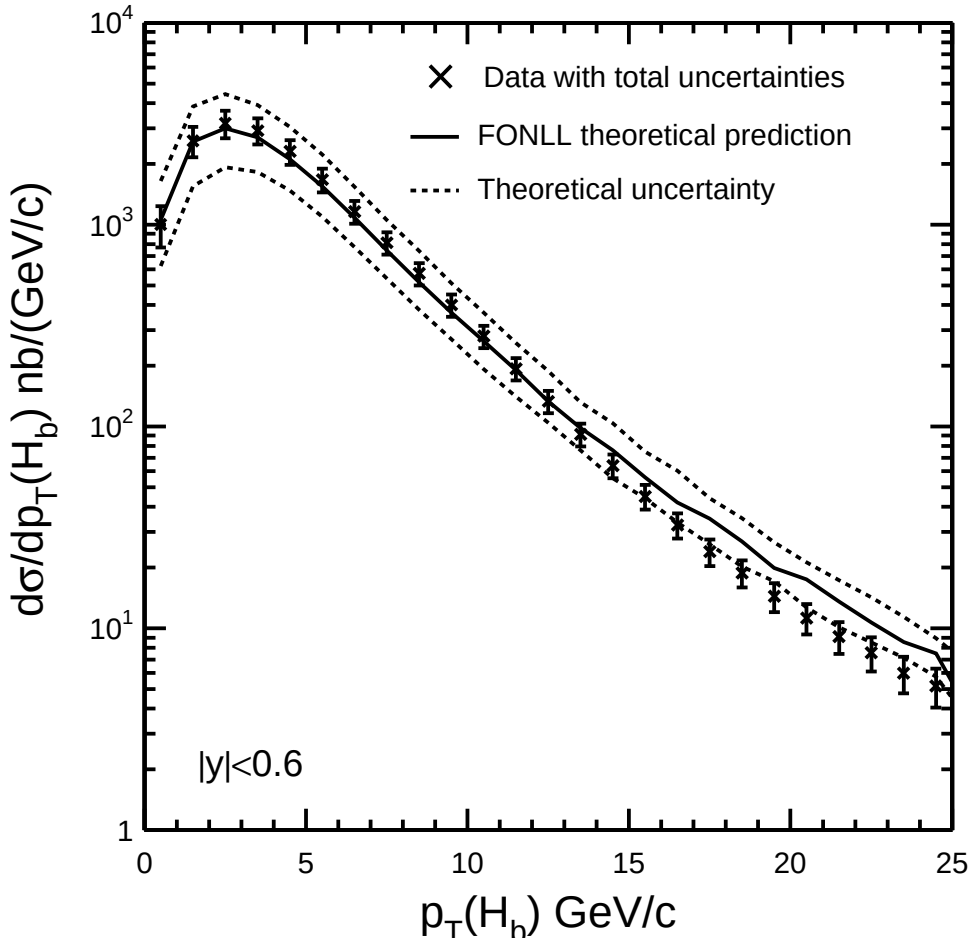

FIG. 17: Production cross section of b-hadrons.

are produced, is non-perturbative and is usually modeled as a result of two processes: fragmentation and hadronization [86, 87]. The fragmentation describes how the quark-antiquark or three-quark combinations are selected from the input combinations of the quarks and gluons. The particular hadron final states for these multiquark combinations is then defined by hadronization. For the b-quarks the final state hadron flavor combination is quantified in terms of the fragmentation fractions: the original b-quark after fragmentation and hadronization would correspond to a

meson with the constituent quark of flavor q (bq) with a fraction f_q or a baryon with a fraction f_{baryon}. Assuming $f_u = f_d$, the measured values are $f_u = 0.399 \pm 0.010$, $f_s = 0.102 \pm 0.014$, and $f_{\text{baryons}} = 0.100 \pm 0.017$, as summarized in Ref. [51].

The lifetime of B-mesons differs noticeably from that of other hadrons, which, combined with their production properties, defines their signature in experiment. Except for the proton (uud) and the neutron (udd), the remaining hadronic states are relatively short lived with lifetimes below 50 ns. The type of the interaction that describes the dominant decay mode of the given hadron is in relation to the lifetime of this hadron. The weakly decaying hadrons have lifetimes above a hundred femtoseconds.The particles decaying via electromagnetic or strong interaction are usually characterized by their decay width, Γ: the electromagnetic decay widths are generally above 1 keV, and the strong decay widths are relatively large and correspond to width to mass ratio of above 10%. The properties of the hadrons (mesons) referred to in this analysis are summarized in Table II, including masses, lifetime, constituent quark content, relevant quantum numbers (spin J, and parity P), and some of the branching ratios [1, 51, 88]. Due to the fact that the $B^0_{s(d)}$ and B^+ mesons are the lowest in mass b-flavored mesons, their decay can only proceed via a weak charged current process which makes it a weak decay. Since most of the B-mesons are produced at relatively high momenta, most of them decay after traveling a distance of a few hundred microns, which gives them a particularly spectacular experimental signature of a decay vertex with a large displacement.[12] The dominant decay mode of B-mesons proceeds via $b \to c$ transition described by the V_{cb} element of the CKM matrix. Because of the small value of V_{cb}, compared to, e.g., V_{cs} describing the c-meson decays, the life

[12]Note that except for the charm (D) mesons, all other long-lived hadrons have a lifetime at least an order of magnitude larger than that of B-mesons, while the rest of the hadrons don't have a significant lifetime at all.

time of the B-mesons is larger than that of charm mesons, which makes easier the identification of the B-mesons in the experiment.

TABLE II: Properties of the mesons relevant to this analysis.

Meson	quarks	m, GeV/c^2	$c\tau$, μm (Γ)	J^P	Mode	$\mathcal{B}$
π^+	$u\bar{d}$	0.1396	7.8×10^6	0^-	$\mu^+\nu_\mu$	99.99%
K^+	$u\bar{s}$	0.494	3.7×10^6	0^-	$\mu^+\nu_\mu(X)$	67%
B^+	$u\bar{b}$	5.279	501	0^-	μ^+X	10%
					$J/\psi K^+$	1×10^{-3}
					$J/\psi\pi^+$	4×10^{-5}
B^0	$d\bar{b}$	5.2794	460	0^-	μ^+X	10.5%
					$\mu^+\mu^-$	$< 1.6 \times 10^{-7}$
					$\pi^+\pi^-$	4.8×10^{-6}
					$K^+\pi^-$	1.8×10^{-5}
					K^+K^-	$< 3.1 \times 10^{-6}$
B_s^0	$s\bar{b}$	5.369	438	0^-	μ^+X	$\sim 8\%$
					$\mu^+\mu^-$	2×10^{-6}
					$\pi^+\pi^-$	3.4×10^{-6}
					$K^+\pi^-$	7.6×10^{-6}
					K^+K^-	3.4×10^{-5}
$J/\psi(1S)$	$c\bar{c}$	3.0969	(91 keV)	1^-	$\mu^+\mu^-$	5.88%

CHAPTER III

EXPERIMENTAL SETUP

The main components of hardware and software tools required for a measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ are described in this chapter: the accelerator chain used to produce the $p\overline{p}$ collisions described in Section III.A and the CDF II detector overview given in Section III.B, followed by a detailed description provided in Sections III.C through III.F.

A. Accelerator chain

Colliding protons and antiprotons are delivered to the collision point by the Fermilab accelerator chain in a multi-step process [18]. The protons are accelerated in stages starting from the proton source inside the pre-accelerator, then passing the Linac, Booster, and the Main Injector to be finally accelerated to the final energy of 980 TeV and collided in the Tevatron. The antiprotons produced by the antiproton source using the protons diverted from the Main Injector are accelerated by the Main Injector to be transfered to the Tevatron where they are accelerated to the energy of 980 TeV and collided with protons. High energy $p\overline{p}$ collisions are produced in the Tevatron at its interaction points. The protons and anti-protons in 36 bunches are focused and forced to cross at the interaction points of the Tevatron (centers of the CDF II and DØ detectors) every 396 ns, providing a collision spot with a spread of about 30 cm along the beam direction and 30 μm transverse to the beam direction. The accelerator chain drawing and aerial view are shown in Fig. 18 [89].

The proton beam delivery to the Tevatron proceeds through several stages. The source located inside the pre-accelerator produces negatively charged hydrogen ions (H-) which are then accelerated to 750 keV by the pre-accelerator every 66 ms in

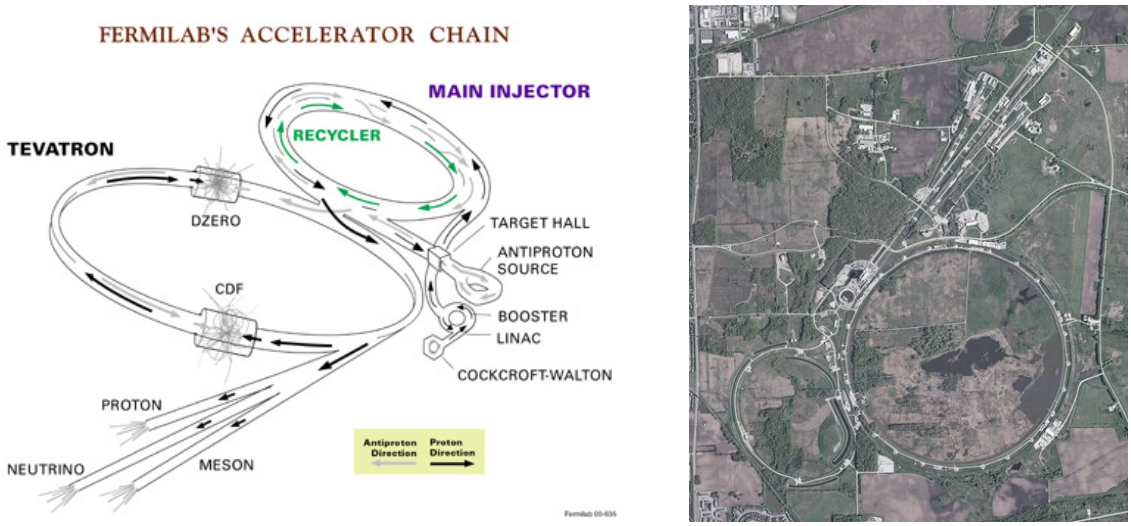

FIG. 18: Drawing and aerial view of the Fermilab's accelerator chain. Protons (antiprotons) inside the Tevatron are moving clock-wise (counter clock-wise).

electrostatic field. The H^- ions are then accelerated by Linac (linear accelerator) up to an energy of 400 MeV. The beam is then transfered to the Booster, a synchrotron with a radius of 75 m, where the electrons are stripped off the H^- ions and the protons are accelerated to 8 GeV. The proton beam is then transfered to the Main Injector (MI), a synchrotron with $r = 525$ m, where it can be accelerated to either 120 GeV (for antiproton production) or 150 GeV (to be transfered to the Tevatron). The Main Injector can also accept antiprotons from the Antiproton Source and accelerate them from 8 GeV to 150 GeV. The Main Injector can accelerate beam as fast as every 2.2 seconds. When loading the Tevatron with protons, 7 bunches are injected from the Booster, accelerated to 150 GeV and then combined into a single bunch. By repeating this process 36 times in a row, the MI loads the protons necessary for a 36x36 store. When loading antiprotons, 4 sets of 7 bunches are sent from the Antiproton Source to the MI to have the antiprotons accelerated to 150 GeV, and combined into 4 bunches to be sent into the Tevatron, which is repeated nine times to load the antiprotons for

a 36x36 store.

The stages for delivering the antiprotons to the Tevatron are the following. For antiproton production, the protons coming from the MI strike a nickel target. Using a magnet, antiprotons with energy of 8 GeV are collected from the spray of secondary particles produced on target. These antiprotons are then directed into the Debuncher, a triangular-shaped synchrotron with $\langle r \rangle = 90$ m, where the antiprotons are cooled (beam phase space reduced) and then transfered to the Accumulator at the same 8 GeV. The Accumulator is also a triangular-shaped synchrotron with $\langle r \rangle = 75$ m. It is the storage ring for the antiprotons; all of the antiprotons made are stored here at 8 GeV and cooled further until needed. The antiproton beam is sent towards the MI on demand either to be accelerated and transfered to the Tevatron or to be stored in the Recycler. The Recycler is an antiproton storage ring located along the ceiling of the MI tunnel, where the antiprotons are kept at 8 GeV. Since 2003 the Recycler is used for additional cooling and storing the antiprotons sent directly from $\overline{p}$-source, which has allowed to substantially increase the Tevatron luminosity.

The Tevatron is the largest of the Fermilab accelerators, it is a synchrotron with $r = 1$ km. The Tevatron is the only cryogenically cooled accelerator at Fermilab: the magnets used in the Tevatron are made up of a superconducting Ni/Ti alloy kept at about 4 K to remain a superconductor. The ring is divided into six sectors labeled A through F. The CDF II collision hall is located at the straight section of B sector. The primary purpose of the Tevatron is to act as a storage ring where protons and antiprotons can collide with each other. The protons and antiprotons injected from the MI are accelerated to 980 GeV and once the final energy is reached, the two counter-rotating beams pass through each other for hours at a time, defining a store. After the number of collisions per second drops too low to be useful for the experimenters, the store is ended and the Tevatron is prepared for new beam.

B. CDF II overview

The upgraded Collider Detector at Fermilab (CDF II) is a general purpose experiment for the study of $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV at the Fermilab Tevatron Collider [19]. The original CDF detector was used to collect data during Tevatron Run I (1992-1996) with $p\overline{p}$ collisions at $\sqrt{s} = 1.8$ TeV. Approximately 100 pb^{-1} of collision data were collected and used in multitude of analyses, including the discovery of the top-quark [90], b-quark cross section measurement [91], search for $B^0_{s(d)} \to \mu^+\mu^-$ decay [21], etc. In the course of 1996-2001 the CDF detector was upgraded for operations in Tevatron Run II. The amount of data collected by the Fall 2005 had surpassed that collected in Run I by an order of magnitude. The analysis presented in this work uses approximately 360 pb^{-1} of data collected by the Fall 2004.

The specific elements of the CDF II detector and the data collection and reconstruction procedures relevant to the search for $B^0_{s(d)} \to \mu^+\mu^-$ are described in the following sections. These cover the high precision charged particle kinematic measurement (tracking) used to determine mass and decay vertex position, the muon identification, and the high efficiency selection (triggering) of the collision events with two muons. The detector elements used to perform these functions are described in Section III.C, followed by the triggering described in Section III.D. The software tools used to reconstruct, analyze, and simulate the information provided by the detector elements are discussed in Section III.E. The feedback needed for proper functioning of all these elements is provided by calibrations, which are briefly discussed in Section III.F.

C. CDF II detector elements

The CDF II, as shown in Fig. 19 [19], is a mirror and (relatively) axially symmetric detector approximately $16 \times 11 \times 11$ m^3 in size. Solenoidal magnetic field of 1.41 T is produced by the superconducting solenoid in the center of the CDF II inside the radius of about 1.5 m. A spherical coordinates system is used to identify directions and positions on the detector with respect to its center. The z-axis is directed along the nominal proton beam direction, ϕ denotes the azimuthal angle with the x-axis ($\phi = 0$) being horizontal, and the y-axis going up. The pseudorapidity, $\eta = -\ln(\tan\theta/2)$, is nominally used to identify the polar direction instead of polar angle θ. The positive (negative) pseudo-rapidity side of the detector is called the east (west) side, while the positive and negative x sides are referred to as the north and south respectively, roughly corresponding to the geographical orientation of the detector. Particle momentum is denoted as p, and its component perpendicular to the z-axis, the transverse momentum, is denoted as p_T. Particle energy, as measured in calorimeter is referred to as E.

Particles are identified and measured via their interactions with various parts of the detector. Only the particles that do not decay before entering the detector and leave noticeable energy inside the CDF II detector material can be identified directly.[1] Charged particles follow a helical trajectory in a magnetic field and leave traces of ionization energy in the tracker material allowing a measurement of the momentum and the point of origin of the particles. The calorimeters are used to measure the total energy of the particles and to discriminate hadrons from electrons and photons. Muons, as minimum ionizing particles (MIPs), do not leave significant energy in

[1]The silicon tracker, which is the part of the detector closest to the beam is only 2 cm away from the nominal collision point.

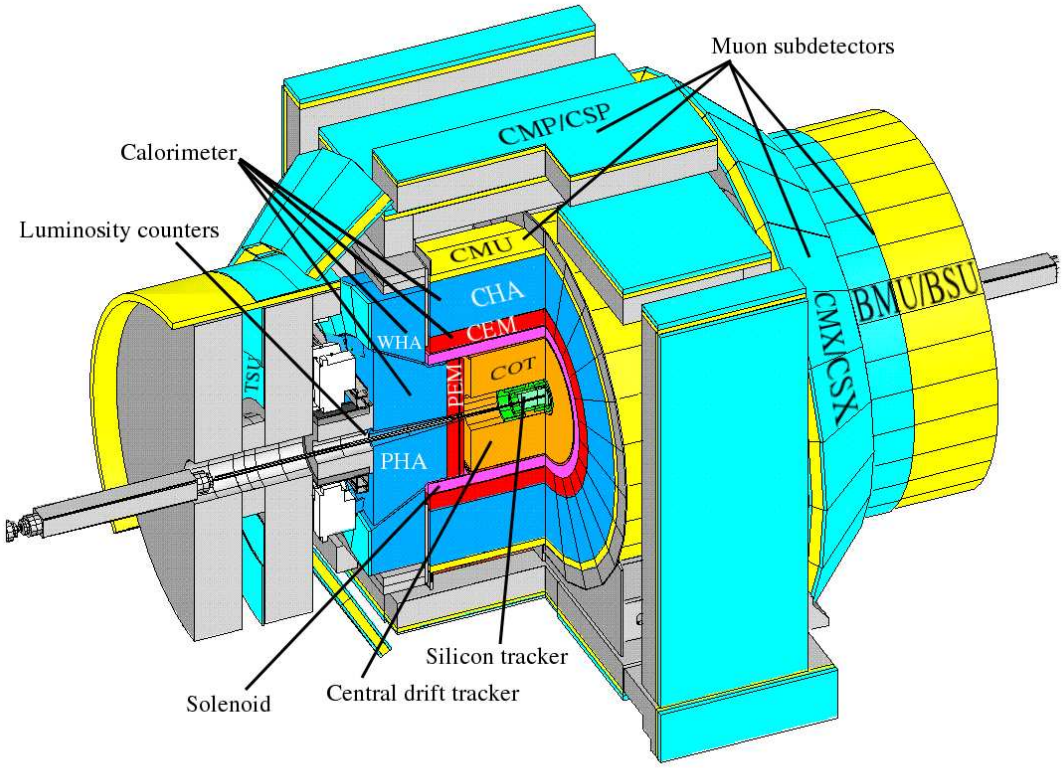

FIG. 19: An isometric view of the CDF II detector showing the main components. Most name component abbreviations are also shown.

the calorimeter and eventually exit the detector. Thus, they are identified as charged particles that traverse the tracker and the calorimeter (and possibly an additional steel absorber) and further deposit ionization energy in the muon detectors. The CDF II detector elements performing the above mentioned functions are described below in more detail: trackers at small (performed by silicon tracker) and large (performed by central outer tracker) radii are described in Sections III.C.1 and III.C.2; calorimeters are described in Section III.C.3; and muon detectors are discussed in Section III.C.4.

1. *Silicon vertex detector*

The charged particle tracking at small radii in the pseudo-rapidity region $|\eta| < 2$ is done with a high precision silicon strip detector. The Run II silicon detector is an eight layer device with 0.7M channels [92, 93]. All layers except the innermost are double sided to allow full 3D standalone tracking.

The silicon detector consists of three subdetectors: the inner silicon layer (L00) [94], the Run II silicon vertex detector (SVX II) [19], and the intermediate silicon layers detector (ISL) [95]. Figures 20 and 21 [93, 96] show side and end views of the silicon detector subsystems. SVX II is the main vertexing device, which has five double-sided layers spanning radii from 2.4 to 10.6 cm. One side is used for r-ϕ measurement; and one side for stereo angle measurement with two small angle stereo (1.2°) layers and three 90° stereo layers. L00, located at 1.4-1.6 cm in radius, is a single-sided radiation hard detector mounted on the beampipe. It significantly improves the vertexing resolution and radiation lifespan of the silicon tracker beyond SVX II alone. ISL is the outermost two layers with r-ϕ and small angle stereo strips at 20 and 28 cm. It allows for precise extrapolation from the COT inwards and increases the coverage of the silicon forward region (in η) beyond the COT into the plug calorimeter. A more detailed description of the silicon tracker configuration and operation can be found in Appendix B.

The analog performance of the detector has signal to noise ratios over 12:1 for SVX II and ISL, and 10:1 for L00. The detector can be operated at 99% single hit efficiency with virtually no noise occupancy. The data from multiple readout chips is converted to an 8-bit optical form and is then sent to the subsequent parts of the DAQ.

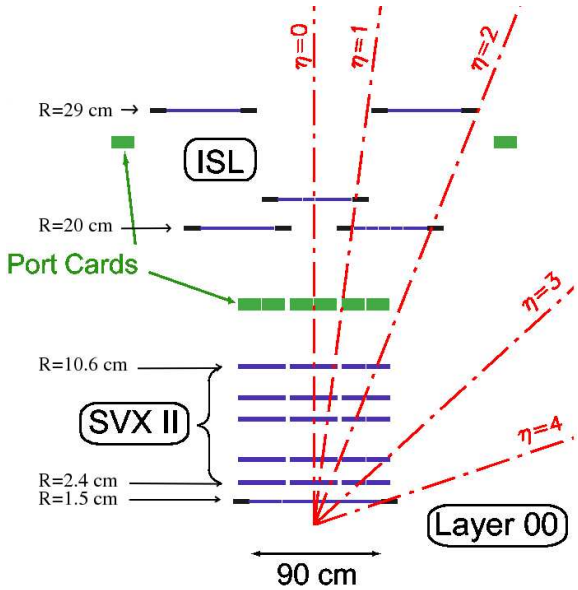

FIG. 20: An r-z view of the silicon detector. Note different radial and longitudinal scales.

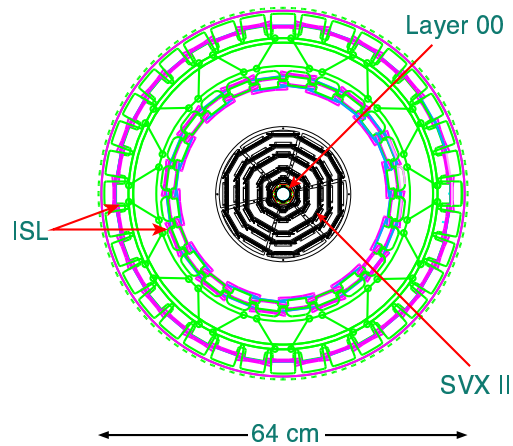

FIG. 21: An r-ϕ view of the silicon detector.

2. *Central outer tracker*

In CDF II the charged particle tracking at large radii in the central pseudo-rapidity region ($|\eta| < 1$) is done with the central outer tracker (COT). The COT is a multi-wire, large open cell, cylindrical drift chamber using a read out that can record multiple hits from each sense wire [97, 98]. The COT is located inside a 1.41 T solenoidal magnetic field in the radial region outside the silicon micro-strip detectors. The active volume of the COT spans 310 cm in the beam direction; between 43.3 cm and 132.3 cm in radius.

The COT contains 30,240 sense wires that run the length (in z) of the chamber between two endplates. Approximately half of the COT wires are axial (run along the z direction) and half are small angle (have 2° with respect to the radial plane) stereo,

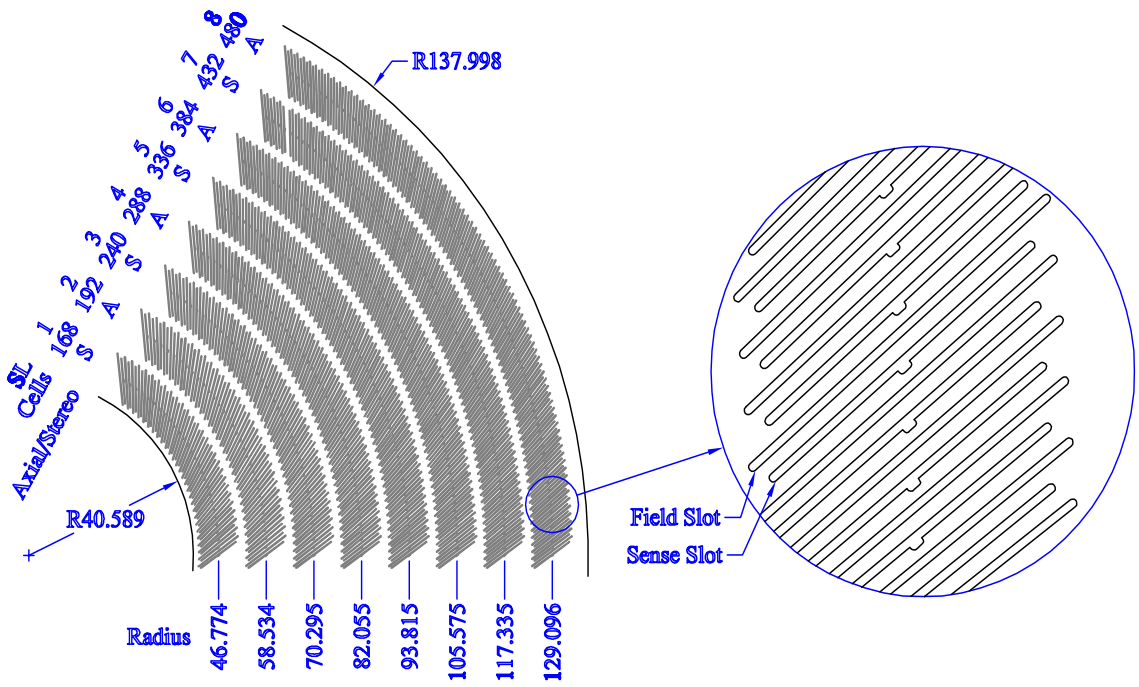

FIG. 22: One sixth of the COT endplate. For each superlayer is given the total number of supercells, the wire orientation (axial or stereo), and the average radius. The enlargement shows the sense and field sheet slot geometry in detail. Dimensions are in cm.

which allows to measure the charged particle momentum and position in 3D. The COT contains 96 sense wire layers in radius that are grouped into eight "superlayers", as shown in Fig. 22 [99]. Each superlayer is divided in ϕ into "supercells" with 12 sense wires each. The supercell consists of a wire plane containing sense and potential (or field shaping) wires and a field (or cathode) sheet on either side. Both the sense and potential wires are 40 μm diameter gold-plated tungsten. The field sheet is 6.35 μm thick Mylar with vapor-deposited gold on both sides. Each field sheet is shared with the neighboring supercell. The supercell is tilted by 35° with respect to the radial direction to compensate for the Lorentz angle of the drifting electrons in the magnetic field. Each sense wire has a potential wire on either side to form the drift cell (open

cell), and the nominal spacing between sense and potential wires is 0.3556 cm. In the middle (in z) of each wire plane there is a 0.16 cm in diameter center support glass rod that is epoxied to each wire in the plane to limit the stepping of wires out of the plane due to electrostatic forces.

To operate as a proportional drift chamber, during collision operations the COT can is filled with Ar:C_2H_6 (50:50) mixture with a small admixture of isopropyl alcohol and oxygen[2] and high voltage is applied to the sense and potential wires to generate a uniform gain and drift field for all the sense wire drift cells. The goal is to create an electric field at the surface of the wire to produce a gain (charge deposited by track to that collected on wire) of $\sim 2 \times 10^4$. With the nominal drift field of 1.9 kV/cm the maximal drift distance in a cell is 0.88 cm which corresponds to approximately 180 ns drift time.

Analog pulses from the drift electrons collected on sense wires with efficiency $\gtrsim$ 97% (on as much as 96 wires per particle traversing the COT volume) are transformed by readout chips into digital pulses with start time corresponding to the analog pulse arrival and the width proportional to the logarithm of the collected charge. These pulses are converted by common start TDCs (Time-to-Digital Converters) into digital hits with 1 ns precision. The TDC hits from all the COT channels are read out every event by the CDF II data acquisition system to be used in subsequent track reconstruction.

3. Calorimeter

Outside the tracking chambers and the solenoid, scintillator-based sampling calorimetry covers the region $|\eta| < 3$ with separate electromagnetic and hadronic measure-

[2]The original design flow rate was 20 SCFH, which has been increased in 2004 to 200 SCFH to decrease the COT wire aging [100].

ments [19]. The central electromagnetic calorimeter (CEM) [101], and the central and endwall hadron calorimeters (CHA, WHA) [102] provide the coverage in the central region ($|\eta| < 1$). The coverage in the pseudorapidity range $1 < |\eta| < 3$ is provided by the plug electromagnetic and hadronic calorimeters (PEM and PHA) [103]. The main parameters of the calorimetry in CDF II is given in Table III with the main geometric features visible in Figs. 19.

TABLE III: CDF II calorimetry summary.

Calorimeter	CEM	CHA	WHA	PEM	PHA
ϕ-segment	15°	15°	15°	7.5° ($\|\eta\|\in[1,2.1]$) 15° ($\|\eta\|>2.1$)	15° ($\|\eta\|\in[1,1.2]$) 7.5° ($\|\eta\|\in[1.2,2.1]$) 15° ($\|\eta\|>2.1$)
η-segment	0.1	0.1	0.1	0.1 ($\|\eta\|\in[1,1.8]$) $0.2-0.6$ ($\|\eta\|>1.8$)	0.1 ($\|\eta\|\in[1,1.8]$) $0.2-0.6$ ($\|\eta\|>1.8$)
$N_{\text{towers}} : \phi\times\eta$	24×20	24×9	24×6	$24(48)\times12$	$24(48)\times11$
Thickness	$18X_0, 1\lambda_I$	$4.7\lambda_I$	$4.5\lambda_I$	$21X_0, 1\lambda_I$	$7\lambda_I$
Resolution, %	$\frac{14}{\sqrt{E_T(\text{GeV})}} \oplus 2$	$\frac{80}{\sqrt{E_T(\text{GeV})}}$		$\frac{15}{\sqrt{E(GeV)}} \oplus 0.7$	$\frac{68}{\sqrt{E(GeV)}} \oplus 4$

The calorimeter is arranged in the form of projective towers pointing to the center of the detector. Each tower is a set of plastic scintillator tiles interleaved with sampling material (Pb for EM calorimeters and steel for hadronic calorimeters) with $20-30$ layers. The towers are arranged into 24 (48 in plug) wedges in r-ϕ view and into η-towers in r-z view (10 each on both sides for plug and central calorimeters). The depth of the electromagnetic calorimeter is around 20 radiation lengths (X_0) or roughly one nuclear interaction lengths (λ_I). The light produced in the scintillator in response to the energy deposited is collected using wave-length-shifting fiber (WLS) spliced

to clear fibers. The fibers carry the light to the photomultiplier tubes (PMTs) and produce a light yield of 40 (CHA) to 300 (PEM) photoelectrons per GeV. The PMTs are running with roughly $(1-10)\times 10^5$ gain with the gain adjusted per PMT to have approximately the same output signal response per GeV "input" energy. The average energy loss per MIP in the calorimeter is around 0.5(1.6) GeV for electromagnetic (hadronic) parts. The energy resolution and signal to noise ratio of the system allows to use the calorimeter information for muon identification with efficiency above 95% in the calorimeter fiducial region.

The signal from the calorimeter PMTs is integrated and digitized to produce an energy measurement using a custom charge integrating circuit. Additionally, the signal from the PMTs is used for arrival time measurement. The calorimeter energy and time signals are further used by the CDF II trigger and DAQ systems.

4. Muon detectors

CDF II has a set of systems of scintillators and proportional chambers to detect muons over the region $|\eta| < 2$. The absorbers for these systems are the calorimeter steel, the magnet return yoke, and additional steel absorbers [19, 104]. The geometric and engineering problems of covering the full η region using these absorbers leads to the four logical systems, but as seen in Table IV [19], they are all functionally similar. The systems are: the central muon chambers (CMU), the central muon upgrade chambers (CMP) with scintillator counters (CSP), the central muon extension (CMX) with scintillator counters (CSX), and the intermediate muon system (IMU). The latter consists of barrel muon chambers (BMU) and scintillator counters (BSU), and toroid scintillator counters (TSU). Rapidity coverage of the systems is given in Fig. 23 [19]. The amount of material, measured in pion absorption lengths (λ_I, for $p_T^{\pi^+} = 5$ GeV), averaged over the azimuth is shown in Fig. 24, which also defines the amount of

multiple scattering the muons undergo while traversing this material (the spread due to multiple scattering, σ_{muls} in r-ϕ view is given in Table IV). The analysis presented here relies heavily on the CMU and CMX detectors.

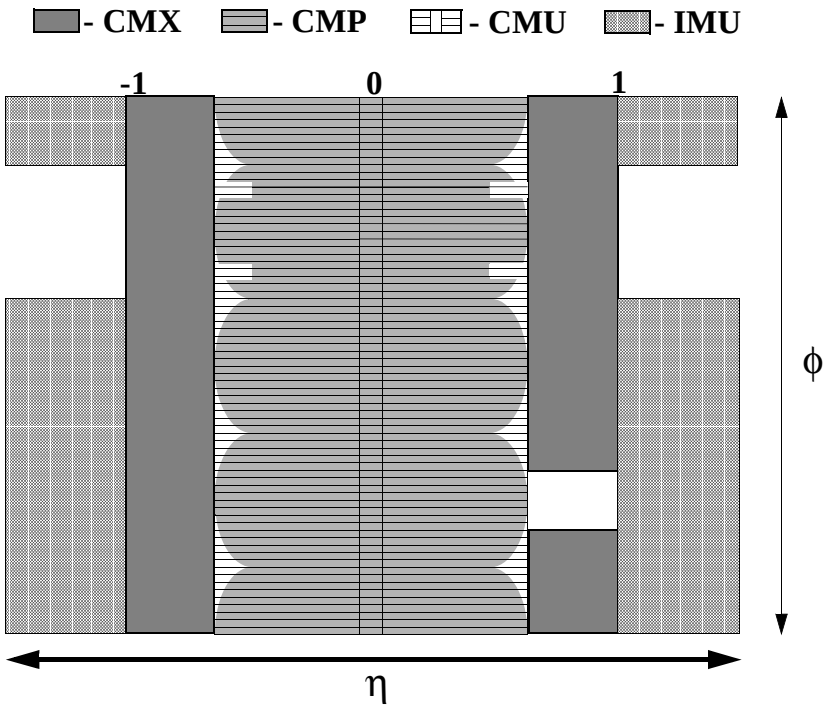

FIG. 23: Coverage provided by CDF II muon systems.

The CMU is located around the outside of the central hadron calorimeter at a radial distance 347 cm from the beam axis [105], see also Fig. 19, and provides the coverage within $|\eta| < 0.6$. The detector is segmented in ϕ into 24 (on each η-sides) 12.6° wedges which fit into the top of each central calorimeter wedge. This leaves a gap in the CMU coverage of 2.4° between each wedge. There is an 18 cm gap between east and west chambers ($z = 0$) due to the central calorimeter arch support structures and the high-voltage fanout modules located at this end of the chambers. The chambers form a cylindrical structure and provide a 2π coverage in azimuth and

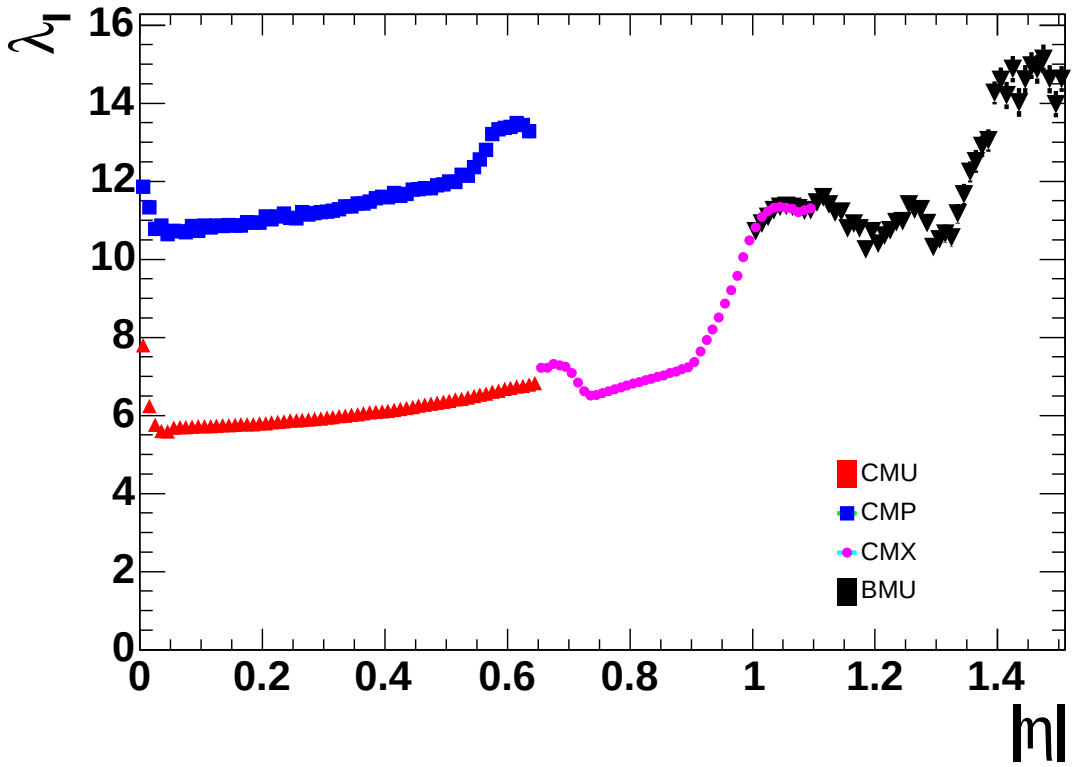

FIG. 24: Number of pion interaction lengths for muon systems.

extend to $|\eta| < 0.6$ (less the gaps). In each wedge the detector is segmented into three 4.2° modules. Each module consists of four layers of four rectangular drift cells, as shown in Fig. 25.[3] A stainless steel resistive 50 μm sense wire is located at the center of the cell. Four chambers, one from each layer, make up a stack. Two of the four sense wires in a stack from alternating layers lie on a radial line which passes through the interaction point. The wires in the neighboring layers are offset by 2 mm to resolve the ambiguity as to which side of the wires (in ϕ) a track passes (drift

[3]The Figs. 25-28 are produced in interactive GEANT3 session using the detector geometry definitions in GEANT3. The sense wires are shown out of proportion for illustrative purposes.

TABLE IV: Operation parameters of the CDF II muon detectors.

Detector	CMU	CMP/CSP	CMX/CSX	IMU
Coverage	$\|\eta\| < 0.6$	$\|\eta\| < 0.6$	$0.6 < \|\eta\| < 1$	$1 < \|\eta\| < 1.5$
Chamber size, cm^3	$2.7\times6.4\times226$	$2.5\times15\times640$	$2.5\times15\times180$	$2.5\times8.4\times363$
$\tau_{\rm drift}^{\rm max}$, ns	800	1400	1400	800
Chamber layers	4	4	8	4
Stacks/wedge	12	~ 12	6	12
Total chambers	2304	1076	2208	1728
Counter size, cm^3		$2.5\times30\times320$	$1.5\times(30\text{-}40)\times180$	$2.5\times17\times180$
Counter layers		1	2	1
Total counters		269	324	864
λ_I	5.5	7.8	6.2	6.2-20
$p_T^{\mu,\rm min}$, GeV/c	1.4	2.2	1.6	1.6-2.2
$\sigma_{\rm muls}\times p_T$, cm·GeV/$c$	15	40	40	40

sign ambiguity). The wires on neighboring stacks are ganged into pairs on the $z = 0$ side. This is done to allow hit position measurement along the wire while keeping the readout on one side of the wire (outer side).

The CMP drift chambers are placed parallel to the beam axis on the horizontal/vertical planar structures (parts) thus forming a box around the central calorimeter at a distance 5-6 m from the beam line and cover the region within $|\eta| < 0.6$. The chambers run along z and extend up to the central calorimeter support structures (approximately 3 m from the center). Due to the geometrical and design limitations in the collision hall, each side of CMP is not a continuous horizontal/vertical plane. Overall, there are 6 parts on the top, 3 on the bottom, and 2 on south and north sides. The coverage of the CMP chambers is shown in hatched gray in Fig. 23. The amount of steel absorber in front of the chambers varies between 0.3 and 0.7 m in depth, which on average corresponds to 4 pion interaction lengths, as shown in Fig. 24. The absorber is formed by the solenoid return yoke on (most of) the top and bottom and the moving steel walls on the sides. The drift chambers in each part are arranged

in 4 layers. The chambers in the consecutive layers are offset by half chamber width (staggered) to help resolve the drift sign ambiguity, as shown in Fig. 26.

The CSP scintillator counters are placed in one layer right behind the CMP chambers on the side away from the beam. Each counter covers 2 muon chambers (in width), as shown in Fig. 26 (the counters are painted blue). The counters can help identify the collision signal (muons) and suppress the beam-related backgrounds and the signal from the neighboring bunch-crossings collision.

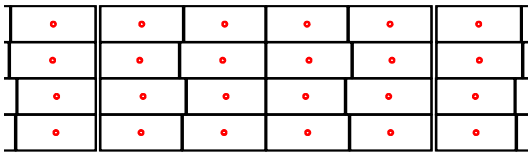

FIG. 25: Cut view of one CMU module. The three modules in a wedge are slightly different. Shown is the central one.

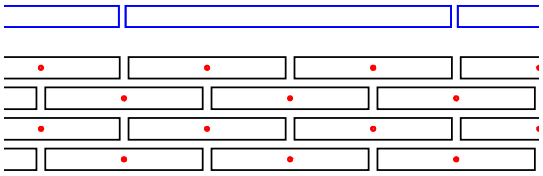

FIG. 26: Cut view of a part of CMP.

The CMX extends the coverage of the central muon detectors for $0.6 \lesssim |\eta| \lesssim 1.1$. In each instrumented ϕ-wedge there are eight layers of chambers, with 6 chambers in each layer per wedge (6 stacks). The CMX is positioned roughly between 4 and 6 m in radius and 3.5 to 5 m in $|z|$. A more complicated geometry is used in this case for chamber arrangement. The chambers covering ϕ-wedges 21-23$\oplus$0-14 are arranged on the conical surfaces: the sense wires are aligned on radial lines (in r-ϕ view) and point to the beamline at $\approx 41°$. On each side of the detector the chambers in wedges 21-23,0-4 (7-14) form north (south) arch. Each arch is a free-standing (movable) structure. Chambers in wedges 5 and 6 on the west side form a key-stone (hangs

from the ceiling).[4] The chambers in the wedges 15-20 are arranged on the planar structure: in each layer the sense wires point to the beamline and are aligned on the plane tilted at 60° with respect to beamline. The planar structures are called "miniskirts": this detector was not fully instrumented until early 2004 and was not fully operational (due to the trigger related issues) until 2005. The chambers are staggered, as shown in Fig. 27: the sense wire azimuthal position corresponds to the center of the gap between chambers on the neighboring layers. Unlike in the case of CMU, there is no gap in coverage between wedges.

The CSX scintillator counters are placed on either sides of the CMX chambers in the arches and keystone. Each counter is a trapezoid and covers roughly 1.5 CMX chamber (in width). There is only one layer of scintillator coverage for the miniskirt wedges and each counter covers one chamber (stack). Starting late 2002 the CSX was used extensively in the trigger to select the particles originating from the collision point.

The IMU system extends the muon pseudo-rapidity coverage up to $|\eta| < 1.5$. Four layers of drift chambers (BMU) with a layer of scintillator (BSU) are parallel to the beamline and placed at the radius 3.6 m, as shown in Fig. 28. The BMU/BSU covers top 270° in azimuth. Two rings (east and west) of 72 trapezoidal counters perpendicular to the beamline and placed inside the toroid steel cover 360° and form the TSU, covering the range $1.3 < |\eta| < 1.5$.

All the muon systems function similarly. In each muon system the rectangular-shaped drift chambers are filled with Ar:C_2H_6 (50:50) gas mixture at atmospheric pressure. A positive high voltage is applied to the sense wire so that the chamber performs in proportional regime. The charge pulse produced as a result of ionizing

[4]The east side wedges 5 and 6 are not instrumented due to the presence of solenoid cooling infrastructure.

radiation from the muons is then read out on the end of the wire, amplified, shaped, and eventually sent to TDC. In the case of CMU and BMU ganged wires both ends of the ganged "wire" is being read out. The amount of charge collected on each side is proportional to the distance from the end of the wire to the ionization point [105]. The charge collected on the wire is encoded into the pulse width in this case. This can later be used for z-position reconstruction using charge division. The light in the scintillators, produced in response to the ionizing radiation, is collected via WLS fibers and is then sent further to PMT. The pulses from the photomultiplier tube are eventually converted to the TDC hits which are read out by the data acquisition system and are used for trigger and further processing.

The hit position resolution (r-ϕ view) across the systems is within a few hundred microns, and that along the z-direction is slightly above 10 cm. This is enough to efficiently reconstruct the tracks in the muon system (stubs) and resolve the drift sign ambiguities (especially in CMU, as the wire offset there is small). The single hit efficiency across the systems is close to 100% for both chambers and the counters.

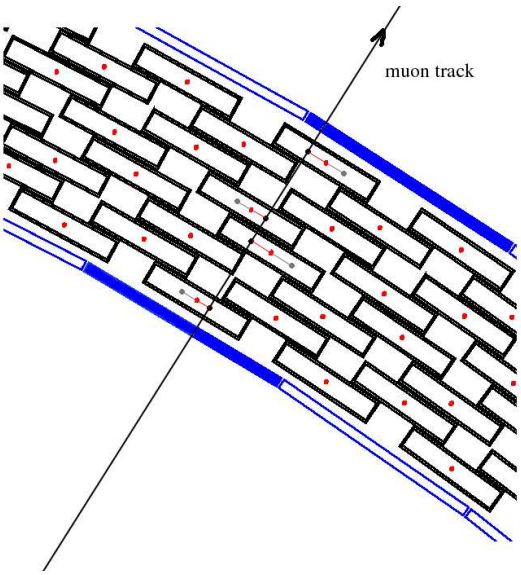

FIG. 27: Cut view of a part of CMX arch. An illustration of a passing track and the resulting hit pattern is also shown.

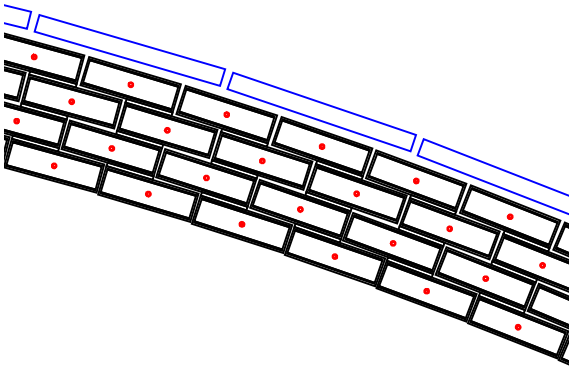

FIG. 28: Cut view of a part of BMU.

D. Data acquisition and trigger

The present analysis relies on the selection of important collision events with (di)muons in the CMU(P) and CMX systems. The primary selection stage of these events is performed by the (di)muon triggers operating within the framework of the detector data acquisition (DAQ) and trigger systems. This section provides a description of the CDF II DAQ and trigger system framework followed by the description of the trigger algorithms used to select the (di)muon events.

1. Infrastructure

The CDF II DAQ and trigger systems consist of the infrastructure, electronics and software used to collect data, calibrate detectors, monitor and configure the electronics. The CDF II is a high rate large event experiment. The detector front end electronics and data acquisition (DAQ) system was designed to handle 132 ns bunch-crossing time or 7.6 MHz input rate.[5] With roughly one million readout channels the total data input rate is huge. Thus, only a small fraction of important events can be read out and sent to mass storage for later analysis. A selection of the events is decided and controlled by the trigger system. The system is designed in such a way that the operation results in a minimal or no loss of important data (no dead time). With roughly only 30 events per million collisions being selected, the amount of raw detector data sent to the mass storage exceeds 100 TB per year or about 20 MB/s logging rate. The data flow diagram of the CDF II trigger and DAQ system is shown in Fig. 29 [19].

[5]The Tevatron upgrade to 132 ns operations was eventually canceled as discussed in Section III.A.

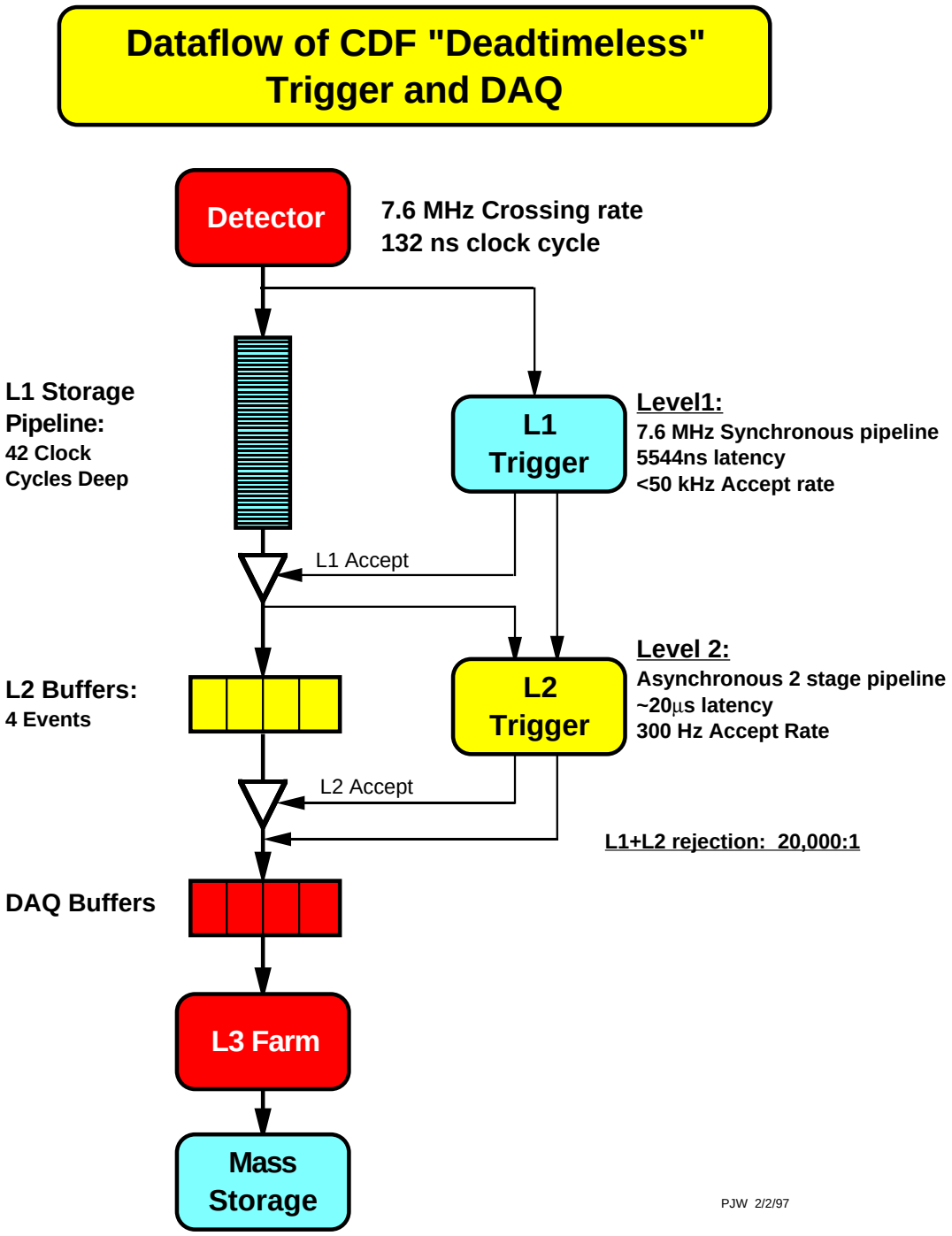

FIG. 29: Data flow diagram for deadtimeless trigger and DAQ system.

The data and control flow in the DAQ and trigger systems begins at the detector front end readout electronics, which is implemented using custom built units operated with a help of commercial processors (computers) and controlled by the global synchronization and control signals provided by the trigger electronics. The data readout, monitoring, and control outside the front end electronics is performed by general use computers (mostly Linux PCs). The flow of control/decision signals is managed by the trigger system. Three level trigger system is used in CDF II, level-1 (L1), level-2 (L2), and level-3 (L3) to select events in corresponding successive steps with increasing level of complexity of the selection requirements. This system allows to gradually reduce the rate of events from about 1.7 MHz to about 20-30 kHz on output of L1, then to 200-500 Hz on output of L2, and eventually to 50-70 Hz on output of L3. Only partial event data (trigger primitives computed by the front end electronics) is available to make decision on L1 and L2, while complete event data is available on L3.

The detector data flow controlled by the trigger decision signals proceeds from the front end readout memory buffers to the DAQ readout buffers after which it is combined into a complete event data to be eventually sent to the mass storage. Prior to the L1 decision the data is placed into a L1 pipeline, from where it is moved into the L2 buffers, where both the pipeline and the buffers are implemented using the front end readout memory buffers. After the L2 decision is made the data from the DAQ buffers specific for separate parts of the detector is combined into a single event data record by the event builder [106]. Events passing the L3 are delivered to mass storage by the data logging subsystem, consumer server/data logger (CSL) [107]. The CSL is responsible for collecting events from the L3, sorting them, and distributing to the mass storage and to the monitoring consumer processes. The data access and operations up to the input to mass storage is referred to as online, and as offline after

that.

The control flow at each trigger level proceeds in the following manner. L1 trigger operates in synchronous mode. Every clock cycle (132 ns) the data read out from the detector is put into 42 clock cycle deep pipeline. By the time the data reaches the end of the pipeline the L1 trigger electronics finishes processing of the trigger data and produces an accept/reject decision signal. The L1 processing time (latency) is thus below 5.5 μs (132 ns $\times$ 42 $\approx$ 5.5 μs). The accepted event data is then moved to one of the four available L2 buffers. The L2 decision is nominally ready within roughly 20 μs to avoid dead time, which means that the L1 accept rate can not be higher than 50 kHz. After L2 accept decision the event is moved to the DAQ readout buffers. From there the event data is transferred via network switch to the event builder [106]. There the event data is combined from the minibanks (event data pieces as produced by front end electronics) into single event data record.[6] The event record is then sent to L3 Linux PC farm [106], where it is reconstructed into physics analysis level objects and then filtered based on sophisticated selection criteria implemented within the standard CDF II software framework described in Section III.E.

During data collection operations the data is acquired in continuous periods of time (runs) characterized by a fixed configuration of trigger and front end electronics hardware and software. The three level trigger configuration is logically defined by a trigger table. In a trigger table for each trigger level a set of named conditions (triggers) is defined. As the operation of the trigger system is based on the selection of events on successive trigger levels, the L3 (named) trigger in the trigger table can be viewed as a trigger path: the L1 accept decision based on this level trigger condition defines the input to L2; the successive L2 trigger condition then defines the input

[6] A sequential integer index is assigned to an event at this point.

to L3. The purpose of the trigger table is to allow a selection of physics signals in multiple channels in an organized manner in a single run configuration.

2. Level-1 (di)muon trigger

The (di)muon L1 trigger decision is based on the muon primitives [19] created from bits of information about the track position defining the expected muon position and muon chamber hits with a potential addition of CMP or CSX hits used for confirmation.[7] The muon primitives created in 2.5° bins (muon trigger towers) are combined on 12 muon matchboxes (single hardware card, covering 30° or two wedges), as shown in Fig 30 [19]. A muon trigger summary card uses the input of the matchboxes to produce a final (di)muon trigger decision.

Tracks are reconstructed by the online hardware track processor, XFT (extremely fast tracker) [108, 109], based on the hit information from axial COT superlayers only. Thus a reconstructed XFT track has only p_T and ϕ information. The track reconstruction proceeds in two stages: first the track segments are reconstructed in each superlayer by XFT finder modules, and last the segments are linked into a single track by an XFT linker module. Depending on the finder configuration, the track segment is reconstructed with 10 to 12 hits. The 10 out of 12 hit requirement (2-miss XFT configuration) was used only in the first year of Run II. The increased luminosity and COT hit occupancy has resulted in an increased XFT fake rate, thus the XFT was moved to 1-miss configuration (11 out of 12 hits). The XFT tracks linked out of four segments each are characterized by a p_T, ϕ_{SL6}-bin pair (96 possible p_T values, 48 for each charge sign value, and 8×288 ϕ_{SL6} or $\approx 0.16°$ bins), where ϕ_{SL6} is the azimuthal position at the mid-point of superlayer six. The XFT track reconstruction threshold

[7]Functionality exists to add the hadron calorimeter timing hits, but was not used for the triggers in this analysis.

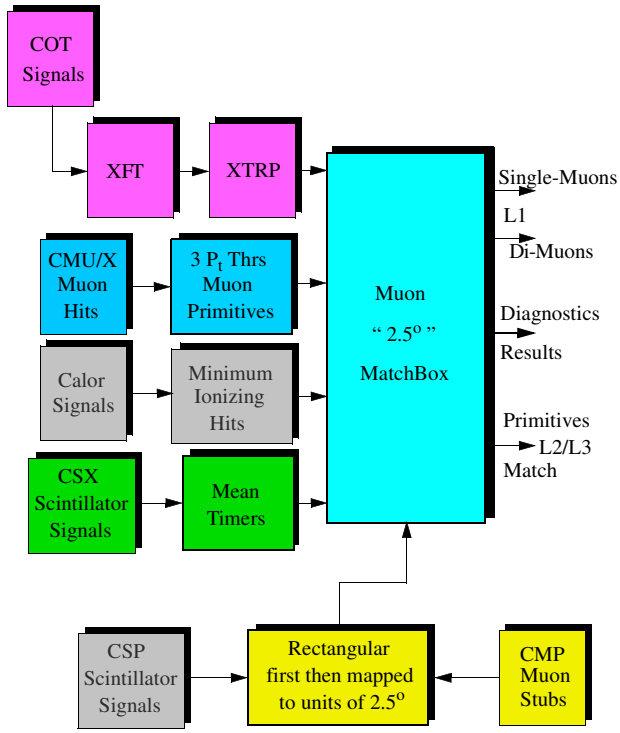

FIG. 30: Data flow for L1 muon trigger.

is 1.5 GeV/c. The XFT tracking performance is characterized by the momentum resolution $\sigma_{p_T}/p_T \approx 1.7\%/\mathrm{GeV}/c$ and ϕ resolution $\sigma_\phi \approx 5.5$ mrad $\approx 0.3°$ with typical reconstruction efficiency above 90%.[8]

The XFT tracks are then passed to the extrapolator unit, XTRP, which produces a map of expected track position at a given detector (CMU and CMX in this case) based on the track angle and momentum. This map is produced with 2.5°, or one muon tower, granularity: for a given XFT track ϕ_{SL6} bin, a number of expected hit muon towers is marked. The center tower is defined by the track deflection in

[8]A more detailed analysis of the XFT efficiency in context of the total muon trigger efficiency is given in Appendix F.

the magnetic field followed by a straight line path outside the solenoid volume. The number of the marked towers is determined from the expected multiple scattering that a muon would undergo while traversing the detector material: $\Delta_{\rm muls} = \sqrt{\xi^2/p_T^2 + \delta^2}$, where $\delta \sim 1.5^\circ$ takes into account a possible misalignment, and ξ corresponds to approximately $3\sigma_{\rm muls}$ with $\xi \approx 8^\circ(16^\circ)$ for CMU (CMX). The XTRP maps for CMU and CMX are produced for two track p_T thresholds, low and high-p_T: 1.5 GeV/c and 4 GeV/c for CMU, and 2 GeV/c and 8 GeV/c for CMX.

Independently, the muon drift chamber hits are selected/flagged into the muon trigger tower map for use in the muon primitives. Consider first the case of CMU subsystem. As mentioned in Section III.C.4, the sense wires on even (odd) layers are aligned on the radial lines. Due to the deflection by the magnetic field in the COT volume the muon track would cross the CMU layer plane at an angle $\alpha \approx 0.13/p_T(\mathrm{GeV}/c)$ away from the radial line [105]. Thus, the drift distance to sense wires on different layers will be different and for the wires laying on the radial line the absolute value of the difference will be proportional to α, $\delta t \approx 0.26/p_T(\mathrm{GeV}/c)$ μs. The preset threshold in drift time difference would then correspond to a transverse momentum cut for a muon track, with 1.5 GeV/c and 6 GeV/c cuts actually used.[9] The muon trigger electronics defines a muon stub map per each threshold with 2.5° granularity separately for east and west sides by marking the tower bin corresponding to the hit pair of radial chambers. The maps are similarly defined for CMX, corresponding to the same momentum cuts, where the wire pairs on layers 0-4, 1-5, 2-6, and 3-7 are tested for hits with matching drift distance in this case.

The CMP and CSX are potentially used as additional muon confirmation. An eight bit per 15° wedge map is created that corresponds to the CSX counter hits:

[9]The actual δt values used are larger to account for multiple scattering.

one bit per counter. Also, a map corresponding to CMP hits is created with one bit allocated per four-layer stack. An internal lookup table is used that provides a correspondence between the CSX bits and CMX trigger tower bits, and, similarly, between the CMP stack bits and CMU trigger tower bits.

All the above mentioned bit maps are combined in a matchbox to produce muon triggers. For a match between a CMU stub map and a corresponding high/low threshold XTRP map, the matchbox outputs a CMU muon match. For two such matches in the same matchbox and being either on different sides or with two unfilled towers between them, the matchbox outputs a dimuon CMU match. Once there is a CMP hit match to CMU muon, a CMUP muon match is reported. Similarly, a CMX muon match is reported if the bits in XTRP, CMX stub, and CSX maps match. The information from twelve matchboxes is combined by the muon trigger summary card. The output of the summary card is a muon trigger decision. An example using the CDF II trigger nomenclature is `L1_TWO_CMU1.5_PT1.5` which corresponds to a low-p_T L1 dimuon trigger and means that there are two CMU muon matches between a stub map with $p_T > 1.5$ GeV/c and XTRP map with 1.5 GeV/c threshold.

3. Level-2 (di)muon trigger

Two implementations of the L2 triggers were used during the data taking. For the first two thirds of this analysis data collection only a trivial L2 triggering was implemented: all the (di)muon events input from L1 were accepted and passed to L3. For the last third of the data sample the trigger hardware capable of making non-trivial (di)muon selections was implemented in early 2004. The algorithms used in this implementation are discussed below.

The data input for (di)muon L2 trigger decision is very similar to that in L1. The differences are: XFT tracks are used directly to give a one-to-one correspondence

with muon hits; tracks and CMU hits are used with 1.25° granularity. A lookup table between the XFT p_T, ϕ bins and the CMU/CMX hits similar to that of XTRP is used to select the XFT track to muon hit matches. Having a specific XFT track p_T, ϕ information for the muon match it is possible to set a larger number of cuts, and apply correlated selections in dimuon case.

The L2 dimuon triggers used in this analysis require two muon matches with oppositely charged XFT tracks and the opening angle difference $\Delta\phi_{\mathrm{SL6}} < 120°$ to reduce cosmic muon background. A specific example of a trigger used in the analysis is `L2_TWO_CMU1.5_PT1.5_DPHI120_OPPQ`, which requires two matches of XFT track with $p_T > 1.5$ GeV/c to CMU stub with $p_T > 1.5$ GeV/c with the matched XFT tracks being of opposite charge and the opening angle at superlayer six being less than 120°. Since the L2 trigger uses L1 as input and uses practically the same information as the one used by L1, the L2 efficiency is very close to 100%.

4. Level-3 (di)muon trigger

The event reconstruction, including the muon reconstruction, proceeds through practically the same set of stages as in the standard offline reconstruction (a detailed description is given in Section III.E). Only the tracking is configured to run simplified and faster algorithms. The selection of muons in the event is done by imposing requirements on the (di)muon kinematics and the set of muon systems that the muon candidate is associated with.

An example of the trigger path used in this analysis is `RAREB_CMUCMU_SUMPT`. This trigger requires two muons with CMU stubs (muon hit track) matched to tracks with $p_T > 1.5$ GeV/c. The tracks must be of opposite charge, have an opening angle at origin $\delta\phi_0 < 2.25$, be within 5 cm from each other in z at the origin, have the scalar transverse momentum sum $p_{T_1} + p_{T_2} > 5$ GeV/c, and have an invariant mass

in the range 2.7 GeV/c^2 < $M_{\mu^+\mu^-}$ < 6 GeV/c^2. A more detailed description of the requirements of all the L3 triggers used in the analysis is given in Appendix F.

E. Offline software

The data reconstruction, simulation and analysis in CDF II proceeds using the offline software. The CDF II offline software system is large, with a few million lines of code organized into over two hundred different "packages" written in four different programming languages. The simulation, reconstruction and some analysis applications must use a large fraction of this code in any one job [110, 111].

The basis of the offline software architecture is formed by a framework (AC++), event data model, and database system. The framework defines the structure of the data processing executables. The event data includes the data read out from detectors as well as calculated objects based on that readout which represent the properties of the physics event. The event data I/O is based on ROOT package I/O system [112]. The access and storage of the events on files is is managed by the data handling system. The database system with respect to offline handles the access of a user process to the calibration, run configuration (and related), and the data handling information data stored in a relational database.

By the type of output or by purpose, the offline event processing executables can be combined in three groups: simulation, reconstruction, and analysis/nTuplizing executables. The centralized effort within CDF II provides development and support of the standardized simulation (CDFSIM) and reconstruction (PRODUCTION) executables. They are briefly described in the following sections. CDFSIM produces output suitable for PRODUCTION input, thus a simulation job is usually run in a bundle: CDFSIM+PRODUCTION. Several custom-made nTuplizing executables were used at

different stages of this analysis, the details of the event processing in the analysis jobs is given in the corresponding sections of Chapter IV.

1. Offline reconstruction

This analysis relies on the reconstruction of high-level objects such as tracks, muons, and track vertices.[10] Most of the high-level objects (includes tracks and muons) are reconstructed by the standard CDF II reconstruction executable, PRODUCTION. More complex objects, peculiar to the analysis are reconstructed by the proper analysis jobs and algorithms. The primary (collision) and secondary (B-meson decay) vertex reconstruction was performed on the analysis level in this case. The track, muon, and vertex reconstruction algorithms are described in this section briefly with a more detailed description provided in Appendix C.

The reconstruction of high-level objects from the raw detector data information is performed in the order following the complexity of the objects. First the low-level objects describing the interaction of particles with the detector on the physics level (physics hits) are computed from the detector raw data: hits in drift chambers with calculated drift distances, scintillator counter hits with times relative to the collision time, calorimeter tower energies based on raw digital counts, hits in silicon tracker with calculated spacial position, etc.[11] The next stage(s) uses these physics hits to compute the quantities describing the collision and the particles partaking in it. The list high-level objects produced by the PRODUCTION includes: charged particles as tracks, charged particles crossing both the tracker and muon chambers as muons, clusters of particles leaving significant energy in the calorimeter as jets, localized large

[10]Vertex is a point common for a number of tracks.

[11]Most of the calibrations and noise/bad channel clean-up happens on this stage.

energy deposition predominantly in electromagnetic calorimeter as electromagnetic objects (photons, electrons), transverse energy non-conservation as missing E_T ($\not\!\!E_T$), a set of tracks originating from the same point as vertices, etc.

The algorithms used to reconstruct a charged particle kinematic variables use the fact that a charged particle in a magnetic field travels along a helical trajectory, which defines a track. A track helix in CDF II is defined with respect to the nominal beamline with the helix axis being parallel to the beamline. A helix is defined by five parameters: curvature ($C = q/2R$), where R is the radius of the helix projection circle in r-ϕ view and q is the sign of the particle charge; signed distance from the beamline or impact parameter (d_0),[12] z-coordinate (z_0), azimuthal and polar angle (ϕ_0 and $\lambda = \cot\theta$) of the helix direction all at the point of the helix closest approach to the beamline. The helix curvature is related to the particle transverse momentum as $|2C| = \frac{eBc}{p_T} \approx \frac{42}{p_T(\mathrm{GeV}/c)}$ cm, where $B \approx 1.41$ T is the magnetic field. Thus, the main purpose of the tracking is to reconstruct a helix based on the hits that a particle left in the silicon detector and the COT. Track reconstruction in each tracker separately is done by COT and silicon standalone algorithms, which is then used for combined COT and silicon tracking. The integrated tracking reconstruction performance is characterized by the track resolution and tracking efficiency, which largely depend on the track isolation (number of non-track hits around the track) and on the track transverse momentum with the best performance corresponding to high-p_T ($\gtrsim 5$ GeV/c) isolated tracks. For the COT algorithms the reconstruction efficiency of such tracks is close to 100%, the corresponding track parameters resolution are $\sigma_{p_T}/p_T^2 \approx 0.15\%/\mathrm{GeV}/c$, $\sigma_{z_0} \approx 0.3\ cm$, $\sigma_{d_0} \approx 250\ \mu$m, $\sigma_{\phi_0} \approx 0.6$ mrad, and $\sigma_\lambda \approx$

[12]The sign of d_0 is defined by the position of the beamline with respect to the helix: going along the helix in the particle direction if the beamline is to the right (left) the d_0 is positive (negative).

3 mrad. The efficiency of the combined COT and silicon tracking algorithms is close to 95%. While the resolution on p_T and ϕ_0 is practically the same as for the COT algorithms, the d_0 resolution improves to $\sigma_{d_0} \approx 20$ μm, and, depending on the number of attached stereo silicon hits, the stereo resolution improves down to $\sigma_{z_0} \approx 50$ μm and $\lambda \approx 1$ mrad. The σ_{d_0} decreases with lower momentum, as shown in Fig. 31 [113, 114]: the resolution distribution includes the beam size of about 30 μm, after it is subtracted the proper track resolution roughly follows the dependence $\sigma_{d_0} = \sqrt{30^2 + (30/p_T(\mathrm{GeV}/c)^2)}$ μm for the case without L00, once a track has a L00 hit the constant term in the dependence goes down by a factor of two.

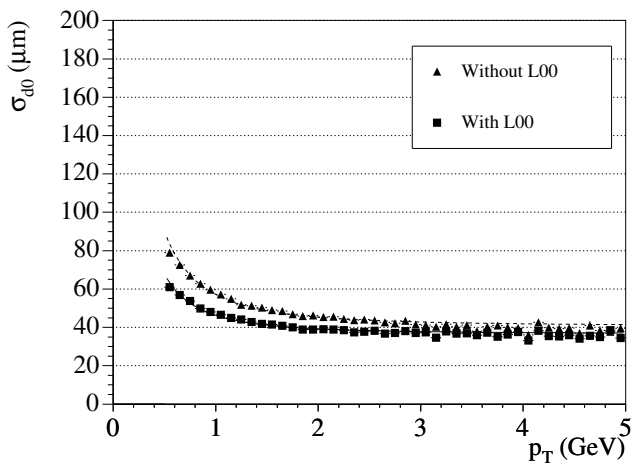

FIG. 31: Impact parameter resolution of tracks with silicon hits (that with L00 is shown separately).

Muon reconstruction is based on combining the information from the tracker, muon subdetectors, and calorimeters. A typical signature of a muon with momentum above 1-2 GeV/c the central detector region is a track, a set of muon detector hits along a line roughly pointing to the track, and a calorimeter energy deposition consistent with a MIP. An example of such signature is shown in Fig. 32: the green arc

segment corresponds to the muon track originating inside beampipe, the black line following the track represents an expected muon path neglecting the multiple scattering; the filled pie-piece corresponds to the measured calorimeter transverse energy of $E_T = 1.49$ GeV in the intersected tower (magenta for CEM energy and blue for CHA energy); the magenta points in the CMU chambers are the muon hits, and the green line segment is a reconstructed muon trace (muon stub). A combination of systems in which the muon stub is matched defines the muon type: CMU, CMP and CMX muons have a stub in the corresponding systems; a muon with matching stubs in CMU and CMP systems is a CMUP muon. The measured muon trace is different in position (and direction) from the expected muon path due to the multiple scattering: in this case the offset (Δx in chamber coordinates) is about 4 cm. The mismatch distribution is roughly Gaussian, the width of that distribution is roughly proportional to the muon p_T: the approximate dependence per muon system is summarized in Table IV as σ_{muls}. The integrated muon reconstruction performance is characterized by the efficiency close to 100% for relatively high momenta (about twice the minimum detectable muon momentum in that detector). This does not include tracking efficiency.

In the vertex reconstruction a vertex is defined by a set of tracks consistent with being originated from the common point. The important signature of a B-hadron decay is the decay vertex displaced (in the direction of the meson momentum) with respect to the primary collision vertex. Therefore it is essential for this analysis to be able to properly reconstruct the vertices. The main ingredients in this procedure are: a proper track helix parameters measurement including the correct covariance matrix, and a flexible tool that is able to fit a set of helices to a fixed point. The track helix measurement is obtained with the help of track Kalman fitter. The vertex fitting tool common for this type of analyses in CDF II is the Run I based algorithm

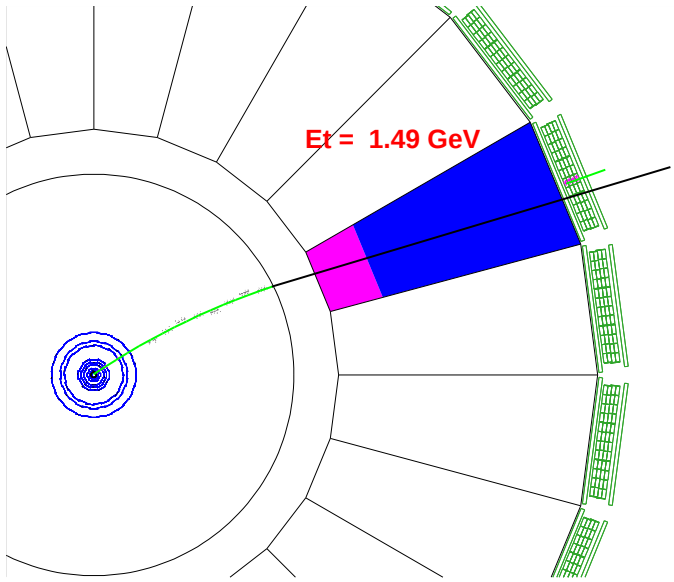

FIG. 32: Event display r-ϕ view of a (simulated) muon with $p_T = 2$ GeV/c, $\eta = 0.1$, and $\phi_0 = 35°$.

for secondary vertex fit with (optional) mass and pointing constraints, CTVMFT [115]. In addition to the decay vertex the primary collision vertex reconstruction is required in this analysis. The method used for primary vertex reconstruction (PRIMEVTX) is based on the CTVMFT.

2. *Monte Carlo simulation*

Monte Carlo (MC) simulation of high energy collisions provides an invaluable input in understanding the detector response to collision events with $B^0_{s(d)} \to \mu^+\mu^-$ and other important signal as well as the backgrounds. Its goal is to provide an adequate detector response to the known physics input. The nominal output of a simulation job is the MC event record with all the original particles, the particle material tracking information with summary results of crossing the detector volume, the detector response in the format of readout channel hits, and the MC truth information that

links the particles to the hits produced by them. Physics collision event simulation proceeds in two stages: event generation and detector simulation [116], which are briefly described below.

a. Event generation

During the event generation stage a physics collision event is generated by an event generator and a decay package(s) with an output being the list of the particles that would enter the detector volume including their kinematics and relation to the primary collision particles. The list corresponding to a single collision (primary vertex) is originally produced by an event generator: either a simplified/uncorrelated list of particles is produced following a preset kinematical distribution, or a list of particles from a modeled $p\overline{p}$ collision. The event is then processed by the decay package(s), a stage necessary if the primary event generator does not decay some particles or uses a decay model not suitable for the analysis purpose. Next, the event is optionally passed through a sequence of user-defined filters: it is much faster to generate an event and filter it out based on the list of the collision/decay products rather than passing the event through the detector simulation and reconstruction chain and filtering out based on the reconstructed objects. Finally, each primary vertex is placed in a particular space-time point following a (Gaussian) distribution with parameters corresponding to those of real collision operations. Actual beam position and collision point spread values as measured in calibrations from a set of data runs representative of the data taking period was used in this analysis. For each run a number of events proportional to the run integrated luminosity was generated.

In this analysis the following event generators were used extensively: Pythia [87, 117, 118], and Bgenerator [119]. Pythia is a general purpose generator. It includes hard scattering subprocess and underlying event generation, parton showering,

hadronization and decay. Since PYTHIA models the hadron collisions rather well, its application is abundant. In this analysis it was used for signal and background simulations, covering most of the aspects of MC simulation needed. A disadvantage of having a detailed collision simulation is that the produced number of particles is large with most of the particles being not from a physics process of interest ($B^0_{s(d)} \to \mu^+\mu^-$). The detector simulation in this case takes a large amount of time. Thus, for the studies that could be performed by considering the B-meson alone, a simplified generator was used, BGENERATOR. BGENERATOR is specifically tailored to produce B-hadrons based on the initial b-quark kinematics followed by fragmentation and hadronization stages. Since the only particles in the event in this case are the B-hadron decay products, the simulation can be performed rather fast. This generator was used extensively in the detector acceptance systematics studies, as described in Section IV.G.1.

To properly account for the decays of the B-hadrons which are not handled well by the event generator, EVTGEN decay package is used. EVTGEN is a package that includes many of the recent advances in the theory and the experimental measurements of the heavy flavor particles decays. It is developed by BaBar and CLEO [120] and is used in the CDF II environment [121]. This package was used to (re)decay all the B-hadrons produced by an event generator upstream.

b. Detector simulation

The detector simulation stage uses the generated event as input and provides the expected detector raw signal response with a flexibility to account for the detector configuration changes on a run by run basis to provide the expected detector response for the real data sample. Particle tracking through the detector medium, detector hit creation (digitization), and particle-to-hit association is performed using simulation framework [116]. In this framework each subdetector as a sensitive element is

represented as a simulation element, which includes the information about geometry, configuration, digitization functions, and MC hit data for the given subdetector.

During the detector simulation generated MC particles are passed to GEANT3 [122]. GEANT3 tracks the particles through the detector media in a step-wise manner using a detailed geometrical representation of CDF II. Physical processes related to the passage of a particle through matter are simulated in each step: energy loss, direction change due to multiple scattering, production of secondary particles, etc. After each step GEANT3 calls a user defined routine and control is passed to the digitizer of the corresponding simulation element. The digitizer creates the MC hits using true particle position and energy loss in a step. At the final stage, after all particles tracking is done the MC hit data is converted to the raw detector data format (as used by front end electronics).

To account for the changes in the configuration of the silicon tracker and to better reproduce the performance of the silicon tracking/acceptance a run-dependent realistic silicon simulation was used in this analysis. For each generated run, as described in the previous section, the silicon detector calibration data was used to configure the silicon simulation. This included the detector alignment parameters, and most importantly the set of readout chips (readout channels) that had a good data readout for that run.

F. Calibrations

To ensure stable an reliable operation the readout electronics should be regularly calibrated controlling the performance using both simple calibration signals as well as using the physics collision data. With respect to readout channels the calibration functions are: identify bad or dead channels, measure noise level and response to

fixed calibration signals. Related to physics analysis, reconstruction, and simulation algorithms the calibration functions are to provide the reliable performance of these methods.

The calibrations with signals testing the electronics alone is regularly used across the CDF II systems. The calorimeter calibration system also includes the use of radioactive sources and optical signal generators, which test both the scintillator response to the ionizing radiation and to the light (sensor calibrations). Another extensively used source of calibrations signals are high energy cosmic muons: these are used for tracker alignment and calorimeter MIP response.

The collision data is used to calibrate the detector channels by known physics signal and to measure/calibrate the beam parameters during run (luminosity, beam position) as well as to provide corrections/calibrate the reconstruction and simulation algorithms. An example of the calibration by physics signal is the mass scale calibration, which uses $J/\psi \rightarrow \mu^+\mu^-$ and $\Upsilon \rightarrow \mu^+\mu^-$ signals to calibrate the energy loss in the tracker volume and the magnetic field in the solenoid. Here the position of the dimuon mass resonance is required to be independent of kinematics, which is achieved by adjusting the amount of material in the tracker and the value of the magnetic field.

CHAPTER IV

ANALYSIS

A. Overview

As discussed in detail in Section II.E, the branching ratio of $B_s^0 \to \mu^+\mu^-$ in the SM is predicted to be $\mathcal{B}(B_s^0 \to \mu^+\mu^-) = (3.5 \pm 0.9) \times 10^{-9}$ [17, 123–125]. So far, the $B_s^0 \to \mu^+\mu^-$ final state has not been experimentally observed and the best published limit is $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 4.1 \times 10^{-7}$ at the 90% confidence level (CL), based on 240 pb^{-1} of Run II data collected by DØ [20]. The most recent CDF publication uses 171 pb^{-1} of Run II data yielding $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 5.8 \times 10^{-7}$ at the 90% CL [84]. As discussed in Section II.E.2, improving the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurement provides new sensitivity to explore various SUSY models.

The measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ using 364 pb^{-1} of Run II data is presented in this chapter [126]. The method of measuring the branching ratio observable is described in Section IV.B. Sections IV.C and IV.D describe the data and Monte Carlo (MC) samples used in this analysis. The likelihood discriminant variable used to help separating potential signal from the backgrounds is introduced in Section IV.E. The measurement of the number of expected background events is discussed in Section IV.F. The acceptance and efficiency estimates for the $B_s^0 \to \mu^+\mu^-$ signal and the $B^+ \to J/\psi K^+$ normalization decays are detailed in Section IV.G. With all this in hand, an optimization and the expected sensitivity are described in Section IV.H. The results and future prospects are discussed in Sections IV.I and IV.J. The consideration of the results with respect to SUSY models is given in Section IV.K.

B. Methodology

1. *Branching ratio observable*

The following relative normalization method is used to measure the branching ratio of the $B^0_{s(d)} \to \mu^+\mu^-$ decays. For a given decay mode X the number of events observed in the experiment is given by $N_X = \mathcal{B}(B_q \to X)\alpha_X \epsilon_X^{\text{total}} N_b f_q$, where $\mathcal{B}(B_q \to X)$ is the branching ratio of the $B_q \to X$ mode, α_X is the detector and trigger geometric and kinematic acceptance (a fraction of $B_q \to X$ events that can possibly be collected by the trigger), $\epsilon_X^{\text{total}}$ is the efficiency of the trigger, reconstruction, and analysis selections; N_b is the number of produced B-hadrons, and f_q is the fragmentation fraction for a B-hadron to fragment into B_q. The product $\alpha_X \epsilon_X^{\text{total}}$ defines the fraction of events that are observed in the analysis out of all the events produced with this decay mode. The particular choice of the separate definitions of the acceptance and efficiency is somewhat flexible, the definitions of these values for this analysis are discussed in Sections IV.G.1 and IV.G. A separate decay mode, the normalization mode, can be used to extract N_b. In this analysis, the $B^+ \to J/\psi K^+$ candidate events[1] collected on the same triggers as $B^0_{s(d)} \to \mu^+\mu^-$ events are used as a relative normalization mode to estimate the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$. Note that it is also possible to perform the measurement of the branching ratio by estimating the number of B-hadrons, N_b, from the production cross-section (σ_b) and integrated luminosity ($\mathcal{L}$) as $N_b = \sigma_b \mathcal{L}$, which defines an absolute normalization method. The absolute normalization method was used in the previous stage of the $B^0_s \to \mu^+\mu^-$ search [21, 84] and is discussed briefly in Appendix D.

Using the relative normalization method described above the branching ratio for

[1]Throughout the text, charge conjugation is implied.

$B_s^0 \to \mu^+\mu^-$ mode can be defined as:

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s}}{\alpha_{B_s}\epsilon_{B_s}^{\text{total}}} \cdot \frac{\alpha_{B^+}\epsilon_{B^+}^{\text{total}}}{N_{B^+}} \cdot \frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\psi K^+) \cdot \mathcal{B}(J/\psi \to \mu^+\mu^-), \quad (4.1)$$

where N_{B_s} is the number of candidate $B_s^0 \to \mu^+\mu^-$ events, α_{B_s} is the geometric and kinematic acceptance of the di-muon trigger for $B_s^0 \to \mu^+\mu^-$ decays, $\epsilon_{B_s}^{\text{total}}$ is the total efficiency (including trigger, reconstruction and analysis requirements) for $B_s^0 \to \mu^+\mu^-$ events in the acceptance, with N_{B^+}, α_{B^+}, and $\epsilon_{B^+}^{\text{total}}$ similarly defined for $B^+ \to J/\psi K^+$ decays; the ratio f_u/f_s accounts for the different b-quark fragmentation probabilities with $f_u/f_s = (0.398 \pm 0.010)/(0.104 \pm 0.015) = 3.83 \pm 0.57$ where the (anti)correlation between the uncertainties has been accounted for [51]; the final two terms are the relevant branching ratios $\mathcal{B}(B^+ \to J/\psi K^+) \cdot \mathcal{B}(J/\psi \to \mu^+\mu^-) = (1.00 \pm 0.04) \times 10^{-3} \cdot (5.88 \pm 0.10) \times 10^{-2} = (5.88 \pm 0.26) \times 10^{-5}$ [1]. Note that using a normalization mode with dimuon events collected on the same trigger is advantageous due to the substantial similarity between the two decay modes. This leads to a substantial cancellation of the acceptance and the efficiency ratios, as described in Sections IV.G and IV.G.1.

2. Blind analysis technique

To avoid a bias during the search, this analysis is performed using "blind" analysis technique. Simply searching for a set of events with a dimuon invariant mass consistent within detector mass resolution, with the B_d^0 or B_s^0 central mass value, is insufficient because the signal signature is expected to be overwhelmed by background events. To provide the best search sensitivity one has to make selections that keep most of the signal and remove most of the background but not to use any events from the potential final signal region during the optimization procedure ("blind" region). The optimization is performed based on the expectations or best understanding of

the signal using MC and background events outside the blind region. After the optimization is done the signal region is unblinded and the (upper limit on) branching ratio is measured based on the observed number of events.

The present analysis deviates slightly from that ideal picture of the blind analysis. Primarily this is due to the fact that the $p\overline{p}$ collision data is being collected gradually. The $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ limit measurement is done in stages, correlated with adding substantial amount of data and reoptimization of the selection criteria to produce an improved measurement. Provided no specific attention is given to the information obtained from the signal window in the previous stage it is still possible to perform mostly unbiased optimization at the given stage. In particular, the first stage of the analysis was performed using 171 pb^{-1} of data employing the absolute normalization method applied to CMU dimuon dataset and yielded a limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 5.8 \times 10^{-7}$ at 90% CL [22]. As previously discussed, the analysis presented here includes the data used in the previous stage, but uses the relative normalization method, more acceptance, and an improved discrimination method.

3. Setting the limit

Converting the observed number of events into a limit must include the proper accounting of the expected backgrounds as well as uncertainties of each input variable. The $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ definition given in Eq. (4.1) does not include the statistical treatment of the measurement and the proper accounting of the background. In reality, the observed number of events in the signal mass window is going to be a sum of background and signal contributions. Since the number of background events is only known to follow Poisson distribution with an estimated average expected value, the number of the signal events is extracted as an interval with a probability or confidence level assigned to it. Of a number of methods, the Feldman-Cousin's method [1, 127]

and the Bayesian integration method [1, 128–131] are the most commonly used ones. Similarly, the branching ratio value can be defined as a confidence interval or, depending on the number of observed signal events, as an upper limit with at given CL. Since it is straightforward with the Bayesian statistics method to incorporate the uncertainties on the input values, the Bayesian integration method is used to set the upper limit on the branching ratio in this analysis.

Due to the specifics of the dimuon trigger operations the data sample is combined from two separate subsamples: the sample composed of events with two CMU muons, CMU-CMU, and the sample of events containing one CMU and one CMX muon, CMU-CMX. Due to their different kinematics and potentially different background content, the two samples are considered separately, including the optimization of the cuts and the measurement of the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The combined branching ratio upper limit is then set based on the limit values in each sample, taking into account the correlations in the uncertainties of input values for both measurements using the Bayesian statistics method described in Ref. [130].

C. Data samples

The selection criteria used to obtain three types of samples, $B_{s(d)}^0 \to \mu^+\mu^-$ signal, $B^+ \to J/\psi K^+$ normalization, and $B_{s(d)}^0 \to \mu^+\mu^-$ control samples, are described in this section. All the data samples are collected from the `RAREB` trigger dataset. The "baseline" and B-candidate kinematic requirements common for all the samples are discussed first. The discussion of requirements specific to the samples and the results of these selections are provided in the following subsections.

The following "baseline" event requirements, summarized in Table V, are applied to the events to select the good quality candidate dimuon events.

`RAREB` **trigger:** The event should pass one of the four `RAREB` triggers.

Trigger fiducial: Each dimuon candidate should have the kinematics in the region where the trigger operation is reliable (fiducial).

XFT: The helices of each muon should cross all the axial superlayers to be able to produce an XFT track: the exit point at $r = 136$ cm should be within the active volume of COT, $|z_{\mathrm{COTexit}}| < 155$ cm. Additionally, each muon should not cross any axial superlayer in the vicinity (1.5 cm) of the sense wire support in the center of superlayer where the charge collection efficiency is low, $|z^{\mathrm{AX}}_{\mathrm{COT}}| > 1.5$ cm.

L1: In the CMU-CMU sample the two muons (muon stubs) should be separated by two muon towers (5°) when crossing the CMU detector on one (east/west) side.

L2: For approximately 99 pb^{-1} of data when L2 dimuon trigger was implemented (see Section III.D.3 or Appendix E), the two muon track positions at the center of superlayer 6 should be within $1.25° < \Delta\phi_{\mathrm{SL6}} < 120°$. In addition, these muons should satisfy the requirements imposed by the L2 muon track-to-stub matching logic. This is defined by the L2 lookup tables (LuT fiducial) and described in more detail in Appendix Section F.D and in Ref. [132].

L3: The dimuon kinematics should satisfy the L3 requirements of the trigger. Specifically, the two muons should be of the opposite charge, $q_1 + q_2 = 0$, originate close to each other, $|\Delta z_0| < 5$ cm, be not consistent with cosmic muons, $\Delta\phi < 2.25$, have an invariant mass within $2.7 < m_{\mu\mu} < 6$ GeV/c^2. Each CMU(CMX) muons is required to have $p_T > 1.6(2.2)$ GeV/c. The dimuon combination should either be a CMU-CMU(CMU-CMX) and have

$p_T^{\mu_1} + p_T^{\mu_2} > 5$ GeV/c, or one of the muons should be a CMUP type and have $p_T > 3$ GeV/c.

L1 match: The muon pair must be consistent with having fired the L1 dimuon trigger. Both muon tracks should have a matching XFT track (the matching track helices should cross the same drift cells in 3 axial superlayers). Additionally, the hits in the muon chambers associated with each muon should have a matching L1 muon trigger primitive as reported by the L1 muon trigger match hardware (TCMD).

L2 match: The muon pair must be consistent with having fired the L2 dimuon trigger. Two *distinct* XFT tracks should match the two muons. The matching XFT tracks should be of opposite charge and have an opening angle $\Delta\phi_{\mathrm{SL6}}^{\mathrm{XFT}} < 120°$.

Track quality: Each track is required to have $p_T > 0.5$ GeV/c and $|z_0| < 60$ cm, and satisfy the default track reconstruction requirement to have at least 5 hits in two axial and two stereo superlayers separately.

Muon quality: The muons must have $p_T > 2$ GeV/c. Both the CMU and CMX stubs must additionally satisfy $\chi^2_{r\phi} < 9$, where $\chi^2_{r\phi}$ is the muon matching χ^2 variable described in Ref. [133] for CMU, and Appendix G for CMX. The muons that cross CMU wedge west-17 or CMX miniskirt are excluded due to the hardware performance issues.

Silicon quality: For each muon track it is required that the number of L00+(SVX II) layers with ≥ 1 associated $r\phi$ hits be ≥ 3. Although SVX-Z and ISL hits are included, no specific requirement is made on their number.

Primary vertex: Each event should have a primary vertex reconstructed using the PrimeVtx described in Section C.C.

Good run: The data should be from the runs with validated good quality performance of the COT, SVX, CMU and CMP detector components. The runs

used correspond to the period of data taking from March 2002 to August 2004, excluding the period when COT was running in a compromised state. This corresponds to approximately 364 pb^{-1} of data. The good quality of CMX is required when events from the CMX triggers are considered. For the CMX triggers, the runs correspond to data taking from August 2002 to August 2004. There is approximately 336 pb^{-1} of data which include the CMX triggers.

TABLE V: Baseline quality requirements.

Requirement \ Sample	CMU-CMU	CMU-CMX
`RAREB` trigger	`CMUCMU_SUMPT` or `CMUPCMU`	`CMUCMX_SUMPT` or `CMUPCMX`
Trigger fiducial		
XFT	$\lvert z_{\mathrm{COTexit}}\rvert < 155$ cm, $\lvert z^{\mathrm{AX}}_{\mathrm{COT}}\rvert > 1.5$ cm	
L1	2 tower separation	
L2	$1.25^\circ < \Delta\phi_{\mathrm{SL6}} < 120^\circ$, LuT fiducial	
L3	$q_1 + q_2 = 0, \Delta\phi < 2.25, \lvert\Delta z_0\rvert < 5$ cm $2.7 < m_{\mu\mu} < 6$ GeV/c^2 $p_T^{\mu_1} + p_T^{\mu_2} > 5$ GeV/c or 1 CMUP $p_T > 3$ GeV/c	
	$p_T^{\mathrm{CMU}} > 1.6$ GeV/c	$p_T^{\mathrm{CMU(CMX)}} > 1.6(2.2)$ GeV/c
L1 match	XFT to track TCMD to stub	
L2 match	$q_1^{\mathrm{XFT}} + q_2^{\mathrm{XFT}} = 0$, $\Delta\phi^{\mathrm{XFT}}_{\mathrm{SL6}} < 120^\circ$	
	$p_T > 2$ GeV/$c, \chi^2_{r\phi} < 9$	
Muon quality	Not CMU W17 or CMX miniskirt	
Track quality	$p_T > 0.5$ GeV/$c, \lvert z_0\rvert < 60$ cm, $N^{\mathrm{hits}}_{\mathrm{COT}} \geq (2 \oplus 2) \otimes 5$	
Silicon quality	$N^{r\phi}_{\mathrm{Si}} \geq 3$	
Primary vertex	Reconstructed primary vertex exists	
Good run	Good CMU-CMU run	Good CMU-CMX run

In addition to the baseline requirements described above the B-candidates in the samples should satisfy the following kinematic requirements summarized in Table VI. The B-hadron candidates are required to have $p_T(B) > 4$ GeV/c and $\lvert y(B)\rvert < 1$,

TABLE VI: B-candidate kinematic requirements.

$B^0_{s(d)} \to \mu^+\mu^-$ sample $B^0_{s(d)} \to \mu^+\mu^-$control sample	$B^+ \to J/\psi K^+$sample
$p_T^B > 4$ GeV$/c, \lvert y^B\rvert < 1$	
$4.669 < m_{\mu\mu} < 5.969$ GeV$/c^2$	$3.017 < m_{\mu\mu} < 3.117$ GeV$/c^2$ $5.120 < m_{\mu\mu K} < 5.430$ GeV$/c^2$
	$p_T^K > 1$ GeV$/c^2$ $\lvert z^K_{\rm COTexit}\rvert < 155$ cm $\lvert z^K_0 - z_{\mu\mu}\rvert < 5$ cm

TABLE VII: B-candidate vertex requirements for $B^0_{s(d)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ samples.

$B^0_{s(d)} \to \mu^+\mu^-$ sample	$B^+ \to J/\psi K^+$sample
$\chi^2_{vtx} < 15$	$P^{B^+}_{\chi^2} > 10^{-5}, \chi^2_{\mu^+\mu^-} < 15$
$\sigma_{L_{3D}} < 150$ μm	
$L_{3D} < 1$ cm	
$\lvert\lambda\rvert < 0.3$ cm	
$\lambda/\sigma_\lambda > 2$	

where $y \equiv 0.5\ln[(E+p_z)/(E-p_z)]$ is the rapidity.[2] For the $B^0_s \to \mu^+\mu^-$ sample, the B-hadron momentum, $\vec{p}(B)$, is estimated using the vector sum of the $\mu^+\mu^-$ tracks, while for the $B^+ \to J/\psi K^+$ sample it is estimated using the vector sum of the $\mu^+\mu^-K^+$ tracks. For the $B^+ \to J/\psi K^+$ sample the dimuon is required to have an invariant mass consistent with that of J/ψ meson and to be $3.017 < m_{\mu\mu} < 3.117$ MeV$/c^2$, and the kaon track is required to satisfy the track and silicon quality requirements from Table V, have $p_T^K > 1$ GeV$/c$ and $\lvert z^K_{\rm COTexit}\rvert < 155$ cm, and to be within 5 cm from the

[2]The $p_T(B)$ requirement is included as part of the optimization described in Section IV.H.

dimuon vertex. The $B^0_{s(d)} \to \mu^+\mu^-$ sample is selected with the candidate invariant mass satisfying $4.669 < m_{\mu\mu} < 5.969$ GeV/c^2, while the $B^+ \to J/\psi K^+$ sample should have the B^+candidates with $5.120 < m_{\mu\mu K} < 5.430$ GeV/c^2.

1. The $B^0_{s(d)} \to \mu^+\mu^-$ sample

After the baseline quality and B-candidate kinematic requirements the events in the $B^0_{s(d)} \to \mu^+\mu^-$ sample are required to pass additional selection criteria aimed at selecting a well measured displaced vertex. These B-candidate vertex requirements are summarized in Table VII. For the $B^0_s \to \mu^+\mu^-$ sample, in the events passing the baseline requirements the two muons are constrained to a common 3D vertex, using CTVMFT fitter described in Section C.C, and the fit is required to have a good quality with χ^2 satisfying $\chi^2_{vtx} < 15$. The fit should correspond to a well measured vertex with 3D displacement uncertainty $\sigma_{L_{3D}} < 150$ μm, where the 3D displacement, L_{3D}, is calculated as the distance between the primary vertex and the fitted secondary vertex projected along the $\vec{p}(B)$ direction. The measured vertex should not be too far from the collision vertex, $L_{3D} < 1.0$ cm and $|\lambda| < 0.3$ cm, where λ is the proper decay length, $\lambda = c\, L_{3D}\, M_{vtx}/|\vec{p}(B)|$ calculated using the invariant mass of the tracks associated with the vertex, M_{vtx}. Finally, the vertex should be consistent with a long-lived hadron decay and have a positive significant measured proper decay length, $\lambda/\sigma_\lambda > 2$, where σ_λ is the uncertainty of the λ measurement. Except for the last, these additional cuts were chosen to minimally impact the signal efficiency, but eliminate particularly anomalous background events, like those due to misreconstructed tracks, or tracks from the particles from the secondary interactions in the detector material. The last of these requirements minimally affects $\epsilon^{\text{total}}_{B_s}$ but dramatically reduces the correlation among the discriminating variables as discussed in Section IV.E.

In the data, 22459 CMU-CMU and 14305 CMU-CMX events pass the baseline

and B-candidate kinematic and vertex cuts. The resulting $M_{\mu^+\mu^-}$ distributions are shown in Fig. 33. These events form the $B_s^0 \to \mu^+\mu^-$ search sample and, at this stage, are an entirely background dominated sample.

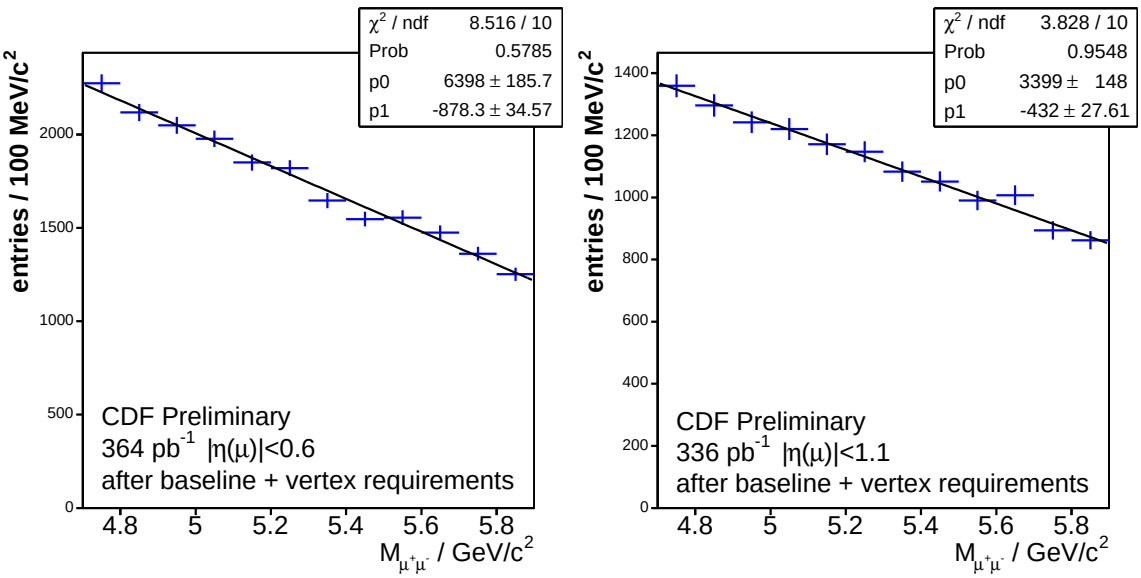

FIG. 33: Di-muon invariant mass distributions for events in the $B_s^0 \to \mu^+\mu^-$ CMU-CMU (left) and CMU-CMX (right) search samples. The results of a fit to a straight line ($p_0 + p_1 M_{\mu^+\mu^-}$) are shown.

2. The $B^+ \to J/\psi K^+$ sample

As with the $B_s^0 \to \mu^+\mu^-$ sample, the $B^+ \to J/\psi K^+$ candidates are required to pass additional selection criteria summarized in Table VII after passing the baseline and B-candidate kinematic requirements. For the two muons from the J/ψ candidate constrained to a common 3D vertex, the vertex fit should have a good quality with $\chi^2_{\mu^+\mu^-} < 15$. For each candidate kaon the $\mu^+\mu^-K^+$ tracks are constrained to a common 3D vertex and the resulting fit chi-squared probability is required to be larger than 10^{-5}. Finally, the candidates must satisfy the vertex requirements of

Section IV.C.1.

In the data, 12121 CMU-CMU and 5353 CMU-CMX events remain after these cuts. These events form the $B^+ \to J/\psi K^+$ sample and the resulting $M_{\mu^+\mu^-K^+}$ distributions are shown in Fig. 34. These are fit with a Gaussian plus first order polynomial and yield a fitted mean mass that is within 1 MeV/c^2 of the world average $M(B^+) = 5279$ MeV/c^2 [1] and a mass resolution of about $\sigma_m = 11$ MeV/c^2 for both the CMU-CMU and CMU-CMX channels. A signal region is defined as $\Delta M = |M_{\mu^+\mu^-K^+} - M(B^+)| < 35$ MeV/c^2 and the estimate of the number of B^+ candidates is made by simple sideband subtraction. A small correction accounting for the $B^+ \to J/\psi\pi^+$ contribution to the signal region is discussed in Appendix I. The number of signal $B^+ \to J/\psi K^+$ events is estimated as $N^{UU}_{B^+\to J/\psi K^+} = 1785 \pm 60$ and $N^{UX}_{B^+\to J/\psi K^+} = 696 \pm 39$, where the uncertainties are completely dominated by the statistics of the sideband subtraction. These numbers are used in Eq. (4.7) to estimate the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ value.

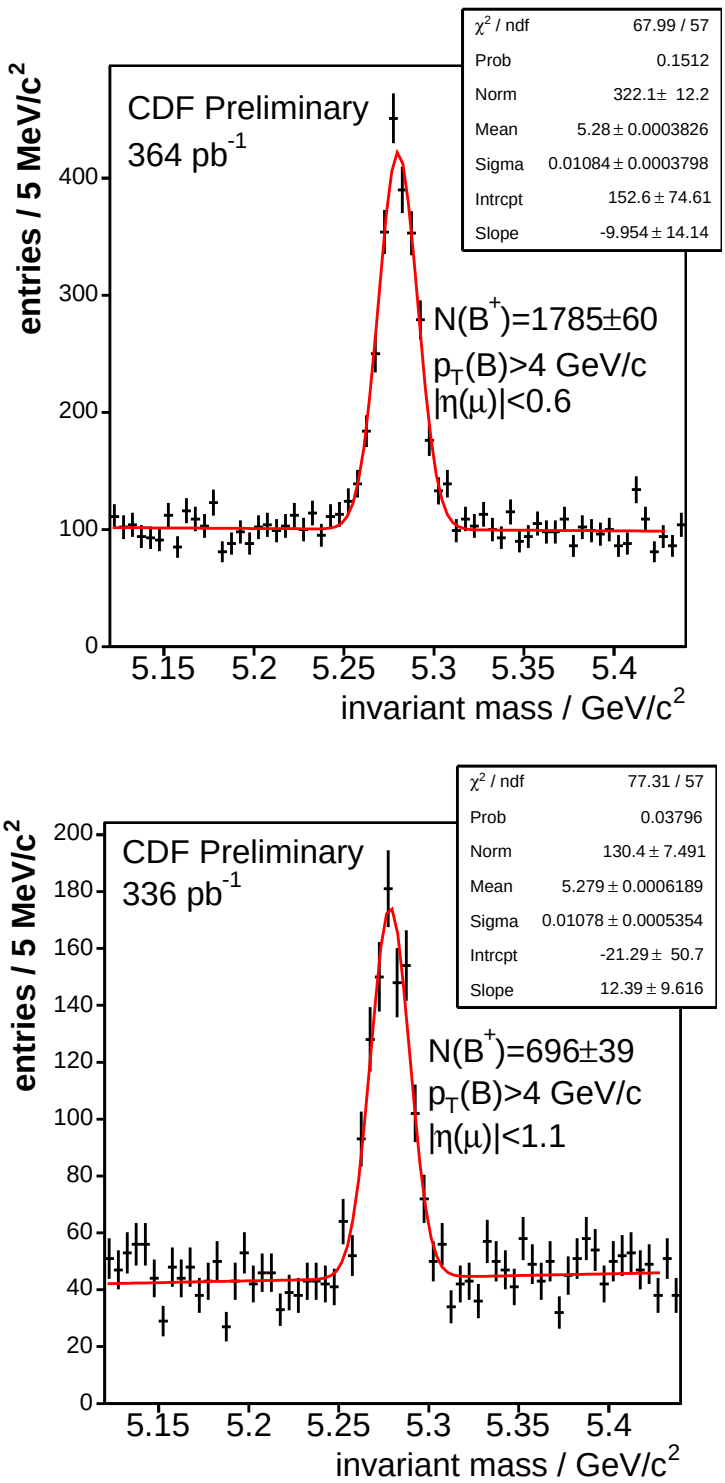

FIG. 34: The $\mu^+\mu^-K^+$ invariant mass distributions for CMU-CMU (left) and CMU-CMX (right) events in the $B^+ \to J/\psi K^+$ sample.

3. $B^0_{s(d)} \to \mu^+\mu^-$ control samples

In order to perform various cross-checks of the methods used to predict the number of background $B^0_{s(d)} \to \mu^+\mu^-$ events the following samples are defined:

OS$^+$: opposite-sign muon pairs, passing the baseline and vertex cuts and having $\lambda > 0$; this is the signal sample, and will not be used for cross-checks;

OS$^-$: opposite-sign muon pairs, passing the baseline and vertex cuts and having $\lambda < 0$;

SS$^+$: same-sign muon pairs, passing looser baseline and vertex cuts and having $\lambda > 0$;

SS$^-$: same-sign muon pairs, passing looser baseline and vertex cuts and having $\lambda < 0$.

FM$^+$: opposite-sign fake-muon pairs, at least one muon of which is required to *fail* the muon stub-track matching requirement of $\chi^2_{r\phi} < 9$, passing looser baseline and vertex cuts and having $\lambda > 0$;

For the "looser" baseline cuts, the L1 trigger matching requirements are removed and only $p^\mu_T > 1.5$ GeV$/c$ is demanded. This is necessary in order to get sufficient SS and FM statistics. It is possible to use the $\lambda < 0$ samples because of the small correlations among the discriminating variables. It should be understood that when using samples OS$^-$ or SS$^-$, the following transformations are made: $\lambda \to -\lambda$ and $\Delta\alpha \to \pi - \Delta\alpha$, where $\Delta\alpha$ is the angle between the B-candidate momentum and the direction from the primary to the secondary vertex

D. Monte Carlo samples

A set of MC samples is generated to estimate the signal efficiency and acceptance for $B^0_s \to \mu^+\mu^-$ signal and $B^+ \to J/\psi K^+$ normalization modes. The method used to simulate and reconstruct the events in these samples together with the transverse

momentum spectrum correction, and a comparison of invariant mass resolution in the samples to the data are discussed further in this section.

A sample of simulated $B_s^0 \to \mu^+\mu^-$ decays generated using PYTHIA [87, 117] is used to estimate the signal efficiency and acceptance. The sample is generated with the recommended underlying event tuning [134] and an EVTGEN [120, 121] was configured to force $B_s^0 \to \mu^+\mu^-$ decays. The events are simulated using the default CDFSIM simulation executable. A realistic silicon simulation with a realistic beamline simulation is used.[3] The sample is reconstructed using the standard PRODUCTION reconstruction executable discussed in Section III.E.1, and are required to satisfy the baseline requirements given in Tables V, VI and VII.

A $B^+ \to J/\psi K^+$ MC sample is simulated and reconstructed in exactly the same manner as the $B_s^0 \to \mu^+\mu^-$ sample, except that EVTGEN was configured to force $B^+ \to J/\psi K^+$ and $J/\psi \to \mu^+\mu^-$ decays. This sample is used to estimate the acceptance and the efficiency of some of the $B^+ \to J/\psi K^+$ selection requirements and to cross-check the MC modeling of the discriminating variables by comparing to a sample of $B^+ \to J/\psi K^+$ events in the data.

Furthermore the B-hadron p_T spectrum at generator level is reweighted to match that measured in Run II [84]. Figure 35 compares the $p_T(B^+)$ spectrum as observed in the data with that obtained from the PYTHIA MC sample for events satisfying the baseline and vertex requirements for CMU-CMU and CMU-CMX channels combined. Here the MC distribution is normalized to the number of events observed in data.

Figure 36 shows the resulting invariant mass distributions for $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ events passing the baseline and vertex requirements. Each is fit to a Gaussian distribution. The $B^+ \to J/\psi K^+$ fit yields a mean fitted mass and mass

[3]The more detailed description of event generation and detector simulation is available in Section III.E.2.

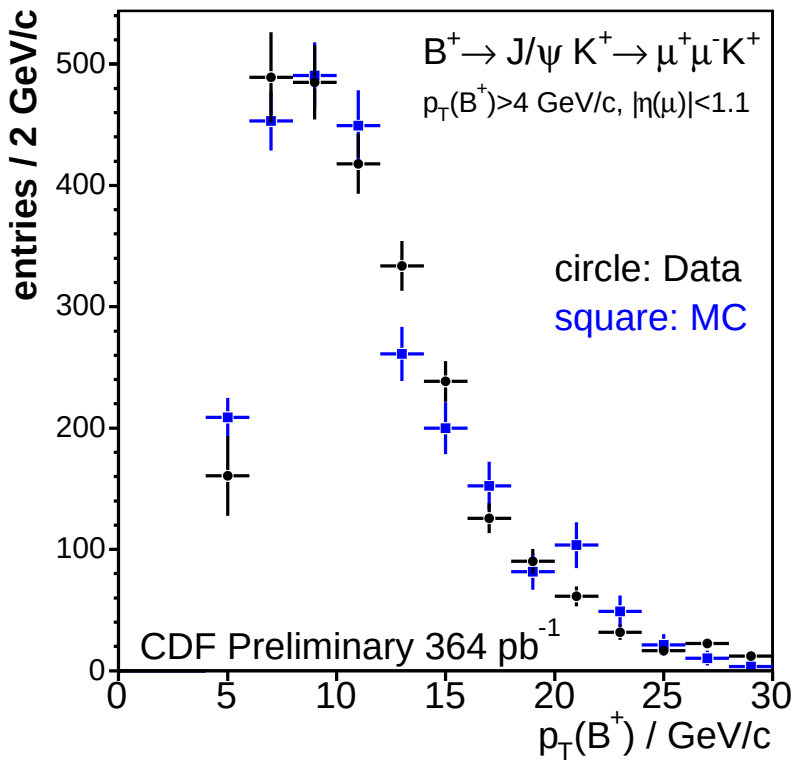

FIG. 35: The $p_T(B)$ spectrum from $B^+ \to J/\psi K^+$ events for CMU-CMU and CMU-CMX combined.

resolution consistent with that observed in the data. In Fig. 37 the di-muon invariant mass distribution for $J/\psi \to \mu^+\mu^-$ from $B^+ \to J/\psi K^+$ candidates from data and MC is fit to a Gaussian. In each case the mean fitted mass is within a few MeV/c^2 of the world average J/ψ mass (3.096 GeV/c^2 as given in Table II) and both yield a mass resolution of $\sigma_m(J/\psi \to \mu^+\mu^-) \approx 14$ MeV/c^2. The fit to the MC $B_s^0 \to \mu^+\mu^-$ sample gives an estimate for the mass resolution to be $\sigma_m(B_s^0 \to \mu^+\mu^-) = 24$ MeV/c^2.

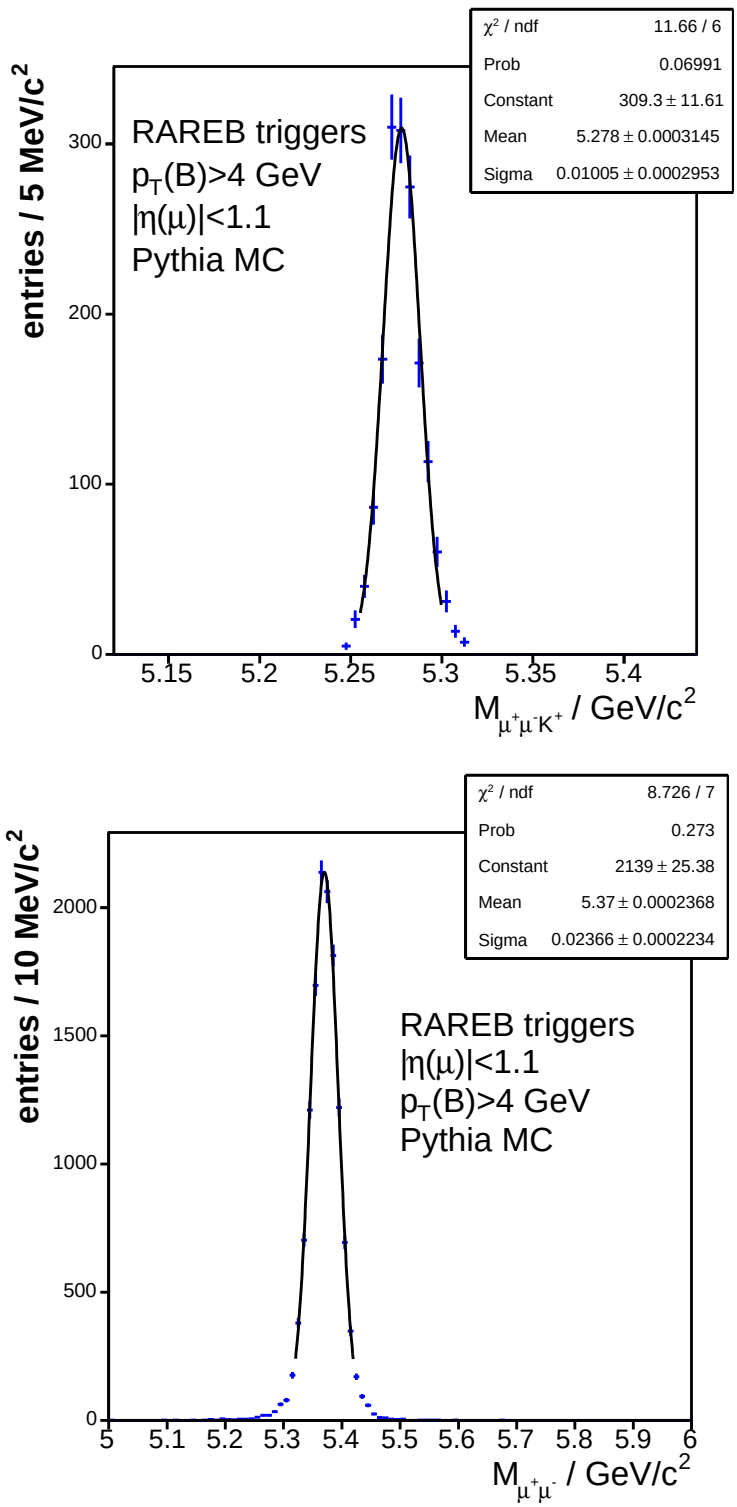

FIG. 36: The invariant mass distributions for $B^+ \to J/\psi K^+$ (left) and $B_s^0 \to \mu^+\mu^-$ (right) MC events. The results of a fit to a Gaussian are also shown.

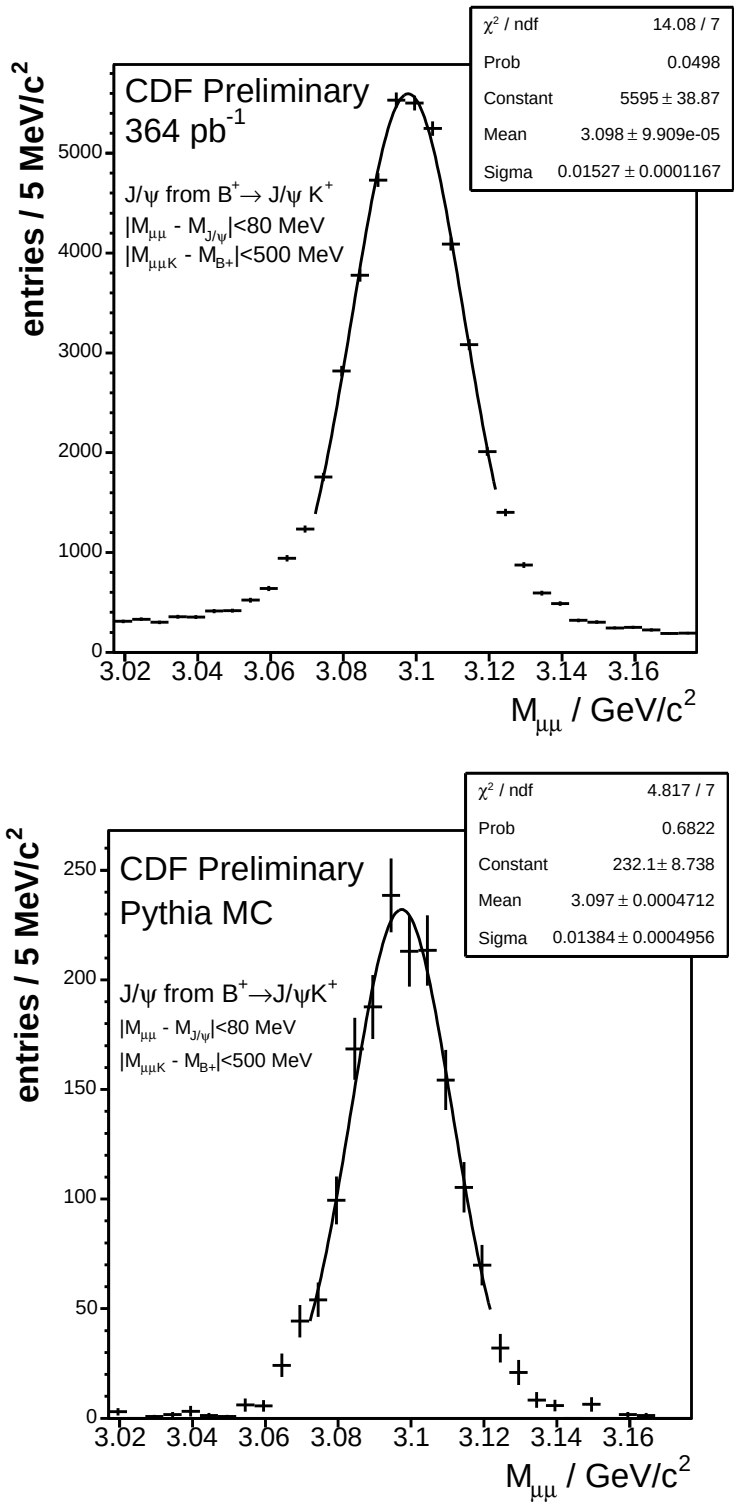

FIG. 37: The invariant mass distribution for $J/\psi \rightarrow \mu^+\mu^-$ from $B^+ \rightarrow J/\psi K^+$ candidates in the data (left) and the PYTHIA MC (right). Each is fit to a Gaussian distribution.

E. Discriminating variables for the $B^0_{s(d)} \to \mu^+\mu^-$ search

The variables used to separate B^0_s candidates from background and a likelihood discriminant using these variables as input are discussed in this section. Note that the small correlations between the input variables for the background events are important for the method used to estimate the expected background discussed in Section IV.F and to ensure the optimal power of the likelihood discriminant.

1. Selection of variables

Additional requirements have to be made to reduce backgrounds inconsistent with the properties of the B^0_s meson production and decay, because it is insufficient to select the signal as simply a pair of opposite charge muons whose invariant mass is consistent with the mass of the B^0_s. Potential sources of background include, continuum $q\bar{q} \to \mu^+\mu^-$ Drell-Yan production, sequential semi-leptonic $b \to c \to s$ decay, double semi-leptonic $g \to b\bar{b} \to \mu^+\mu^- X$ decay, $b(c) \to \mu X$+fake, and fake+fake events. A variety of discriminating variables were considered and the following were identified as among the most promising:

$\mathbf{M}_{\mu^+\mu^-}$: the invariant mass of the muon pair, identifying the B^0_s meson invariant mass

λ: the proper decay length (defined in Section IV.C.1), used to identify the displaced vertex signature of the B^0_s meson

$\Delta\alpha$: the opening angle between the B^0_s flight direction, $\vec{p}(B)$, and the direction of the decay vertex - estimated as the vector originating at the primary vertex and terminating at the muon-pair vertex. It allows to select the candidates consistent with a two-body decay of a long-lived hadron

Isolation: the track isolation of the candidate B^0_s meson defined as, $I = p_T^{B^0_s}/[p_T^{B^0_s} + \sum_i p_T^i(\Delta R < 1)]$, where the sum is over all tracks within an η-ϕ cone of radius

$R = 1$, centered on $\vec{p}(B)$; the tracks must satisfy the COT quality requirements described in Section IV.C and have a z_0 within 1 cm of the mean z_0 of the two muons. The isolation allows to additionally remove the backgrounds coming from gluon splitting heavy flavor production with semileptonic decays, QCD jets, and other multitrack environment.

Figure 38 compares the distributions of these variables in the signal MC and the data for events satisfying the baseline and vertex requirements given in Tables V, VI and VII. The shapes of the signal and background-dominated data distributions are clearly very different. To further reduce the background, very loose requirements are made on the isolation and $\Delta\alpha$ variables, $I > 0.50$ and $\Delta\alpha < 0.70$ rads. These requirements leave 6242 and 4908 events in the CMU-CMU and CMU-CMX $B_s^0 \to \mu^+\mu^-$ search samples, respectively. Using the $B_s^0 \to \mu^+\mu^-$ MC sample, the efficiency for these requirements is estimated to be 92%. For the remainder of the text it should be understood that these loose I and $\Delta\alpha$ requirements are made for the $B_s^0 \to \mu^+\mu^-$ samples.

Having minimal correlations between the four variables is useful for optimal power of the likelihood method and for improving the estimate of the background (more detail in Section IV.F). The correlations among the four discriminating variables are displayed as profile histograms in Figs. 39 and 40 for CMU-CMU and CMU-CMX data passing the baseline and vertex cuts. The vertical error bars are calculated as the uncertainty on the mean y value, $\langle y \rangle$, in each bin of x. The correlations are quantified by calculating the linear correlation coefficient between each pair of variables using:

$$\rho_{xy} = \frac{1}{N-1} \frac{\sum_i^N (x_i - \overline{x})(y_i - \overline{y})}{\sigma_x \sigma_y} \tag{4.2}$$

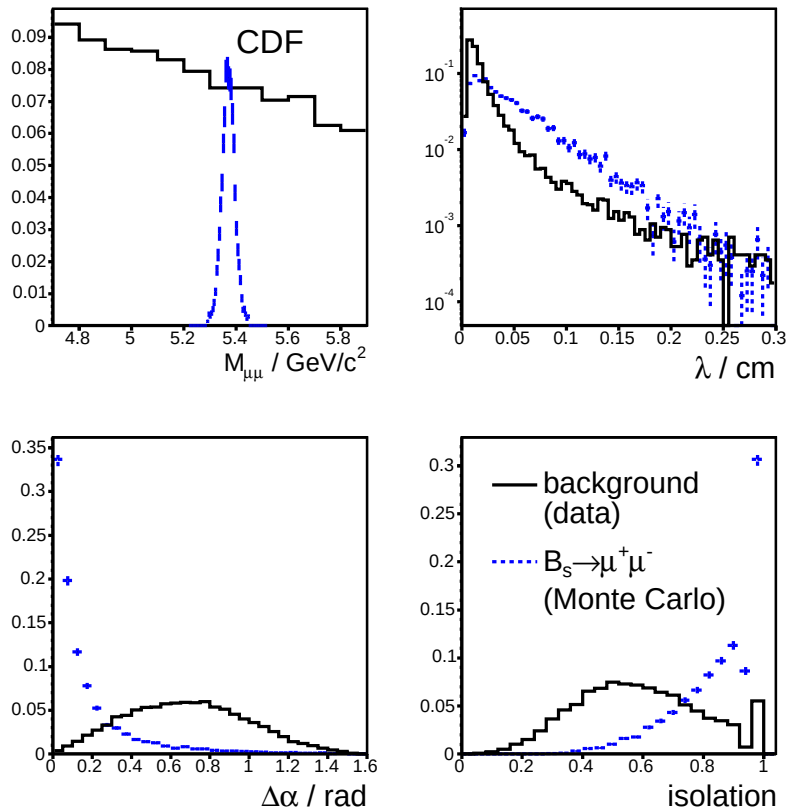

FIG. 38: Distributions of discriminating variables for MC signal (dashed histograms) and a $B_s^0 \to \mu^+\mu^-$ data sample (solid histograms) events passing the baseline and vertex requirements. The histograms are all normalized to unit area. For the $M_{\mu^+\mu^-}$ plots, different binning is used for the data and MC.

where N is the total number of events, $\overline{x}$ and σ_x^2 are the mean and variance of the variable x, respectively, and similarly for y. The statistical uncertainty of ρ_{xy} is 0.013 (0.014) for each CMU-CMU (CMU-CMX) coefficient. The resulting values are given in Table VIII. None of the variables are significantly correlated (all are with $\rho_{xy} < 0.1$). In the absence of significant correlations among a set a variables, a simple multi-variate relative likelihood can be constructed which provides approximately maximal discriminating power.

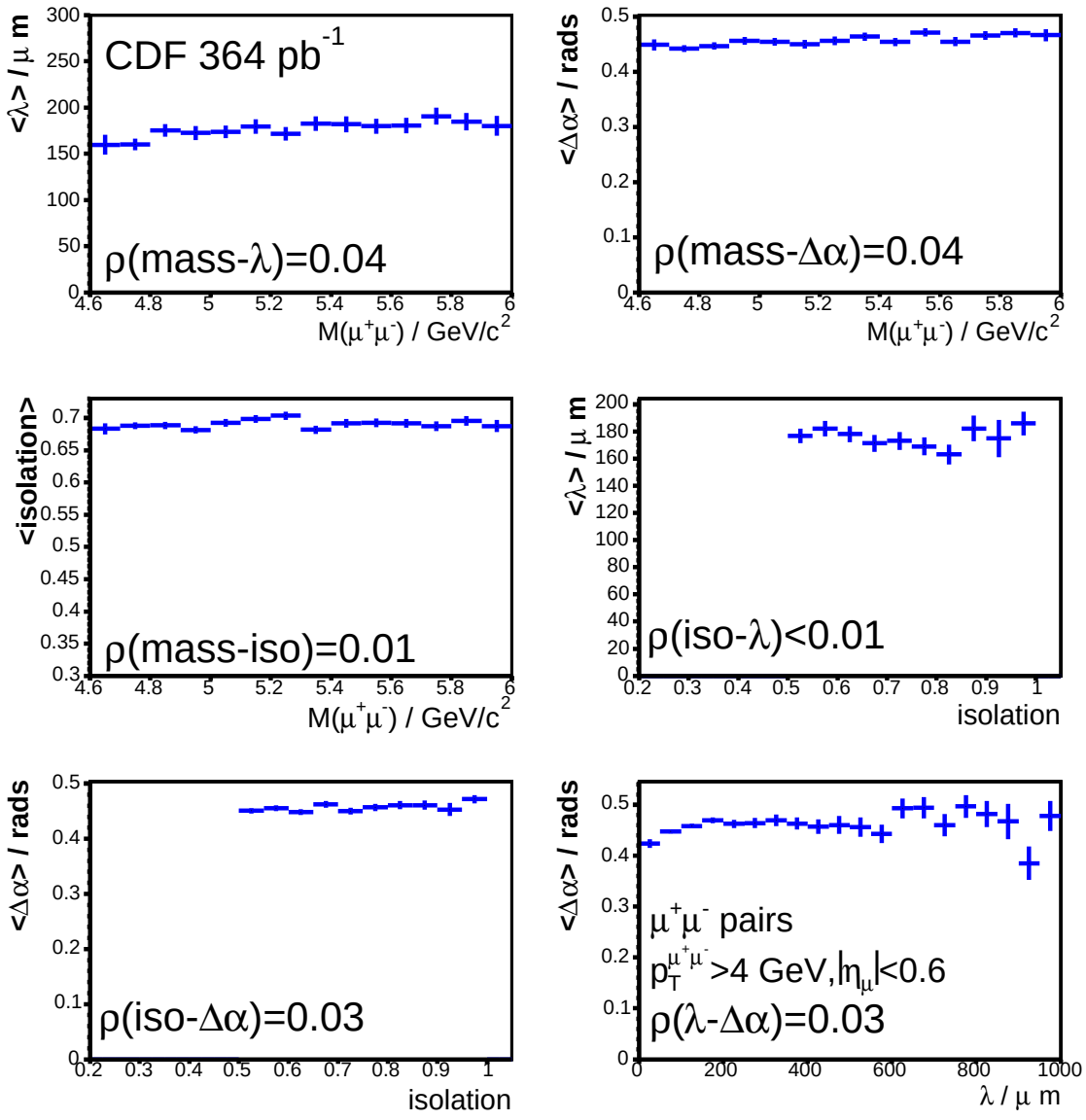

FIG. 39: Profile plots showing the correlations among the discriminating variables for CMU-CMU $\mu^+\mu^-$ pairs.

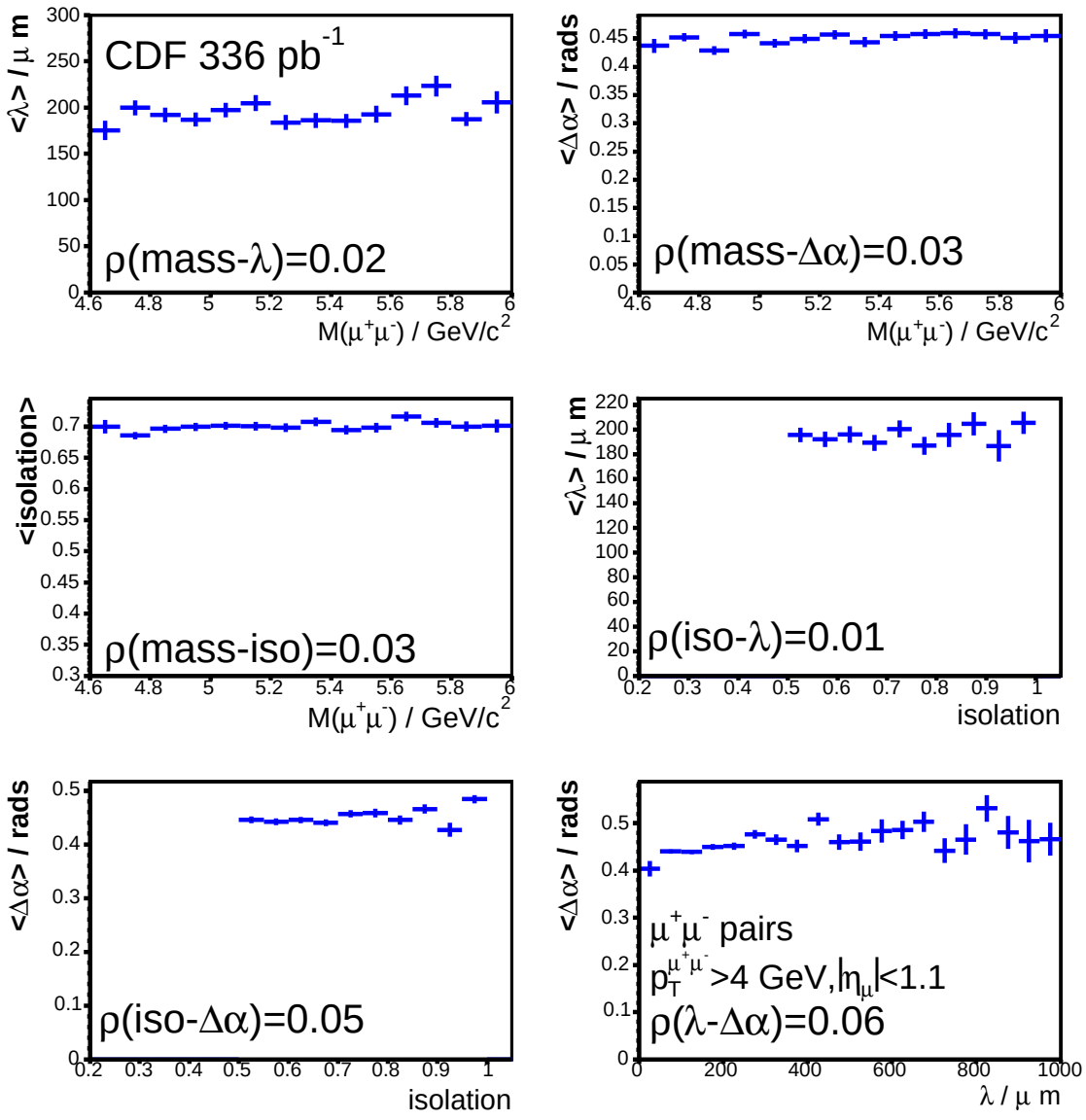

FIG. 40: Profile plots showing the correlations among the discriminating variables for CMU-CMX $\mu^+\mu^-$ pairs.

TABLE VIII: The linear correlation coefficients among the four discriminating variables.

	CMU-CMU				CMU-CMX			
	mass	I	λ	$\Delta\alpha$	mass	I	λ	$\Delta\alpha$
mass	1	0.012	0.037	0.045	1	0.029	0.018	0.026
I		1	0.003	0.028		1	0.009	0.049
λ			1	0.027			1	0.062
$\Delta\alpha$				1				1

2. *A multivariate relative likelihood discriminant*

The relative likelihood is simply defined as:

$$LH = \frac{\prod_i \mathbf{P}_s(x_i)}{\prod_i \mathbf{P}_s(x_i) + \prod_j \mathbf{P}_b(x_j)} \tag{4.3}$$

where $i, j \in [1, N_{var}]$, N_{var} is the number of variables used to construct the likelihood, and $\mathbf{P}_{s(b)}(x_i)$ is the probability that a signal (background) event has an observed x_i. By construction, $LH \in [0, 1]$, were large LH implies that an event is signal-like and small LH corresponds to the background-like events. Note that the LH method is more powerful then the method using a combined application of the cuts on the discriminating variables (box cut). E.g., for the same signal efficiency, compared to the box cut on the variables used used in Refs. [22, 135], the LH can reduce the background by more than a factor of two.

For this analysis a relative likelihood is constructed using $(I, \Delta\alpha, \lambda)$, where $P(\lambda) = \exp^{-\lambda/\lambda_{B_s^0}}$ ($\lambda_{B_s^0} = 438$ μm, the world average B_s^0 lifetime) and is the probability that a real $B_s^0 \to \mu^+\mu^-$ decay would yield a proper decay length at least as large as that observed. Figure 41 shows the λ distribution for $B_s^0 \to \mu^+\mu^-$ MC events at various stages of the selection. As expected, the MC shows a uniform distribution $\lambda \in [0, 1]$, unaffected by λ resolution effects. Even after the $\lambda/\sigma_\lambda > 2$ requirement, the distribution is flat up to about $\lambda > 0.8$. This lifetime variable was chosen for use over others that were considered because it reduces the LH sensitivity to the MC modeling of the λ resolution and provides the best discriminating power among the discriminating variables used here.

Binned histograms are used to estimate the probability distributions, $\mathbf{P}(x_i)$, used in Eq. (4.3). The signal and background probability distributions are compared in Figs. 42 and 43 with the binning used to construct the LH variable. Variable binning

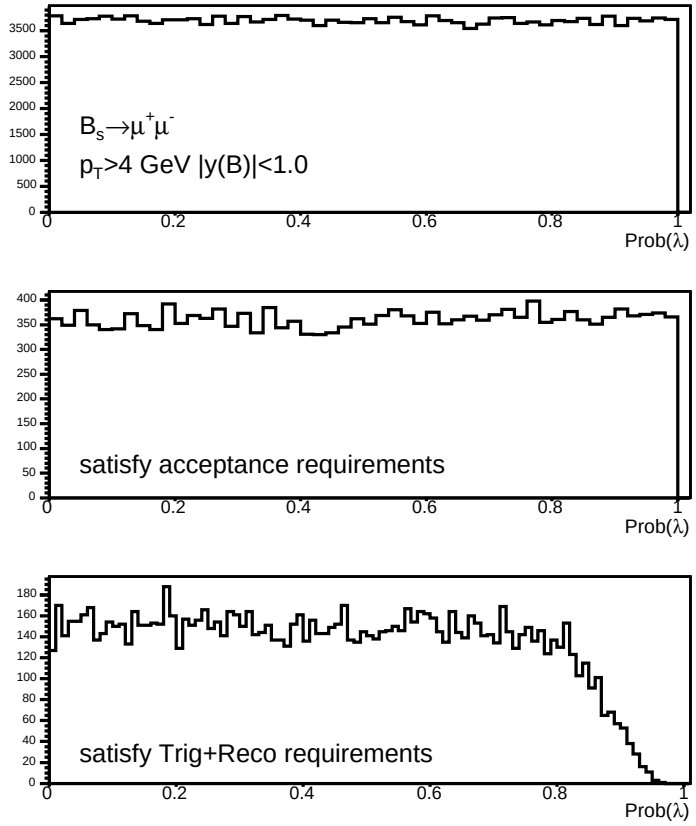

FIG. 41: The λ distribution for $B_s^0 \to \mu^+\mu^-$ MC events with $p_T > 4$ GeV/c and $|\eta(B)| < 1$ (top), additionally satisfying the trigger acceptance criteria described in Section IV.G.1 (middle), and also satisfying the remaining baseline and vertex requirements (bottom).

is used to properly account for regions with low statistics. Underflows and overflows are also properly accounted for. The probabilities are determined from events satisfying the baseline and vertex requirements and for the background are taken from the mass sidebands, $4.669 < M_{\mu^+\mu^-} < 5.169$ GeV/c^2 and $5.469 < M_{\mu^+\mu^-} < 5.969$ GeV/c^2, while for the signal are taken from even numbered $B_s^0 \to \mu^+\mu^-$ MC events. As discussed in Section IV.G, the signal efficiency of requiring $LH > LH_{cut}$ will be determined using the statistically independent odd-numbered $B_s^0 \to \mu^+\mu^-$ events. The resulting likelihood curves are shown in Fig. 44 for signal and background events.

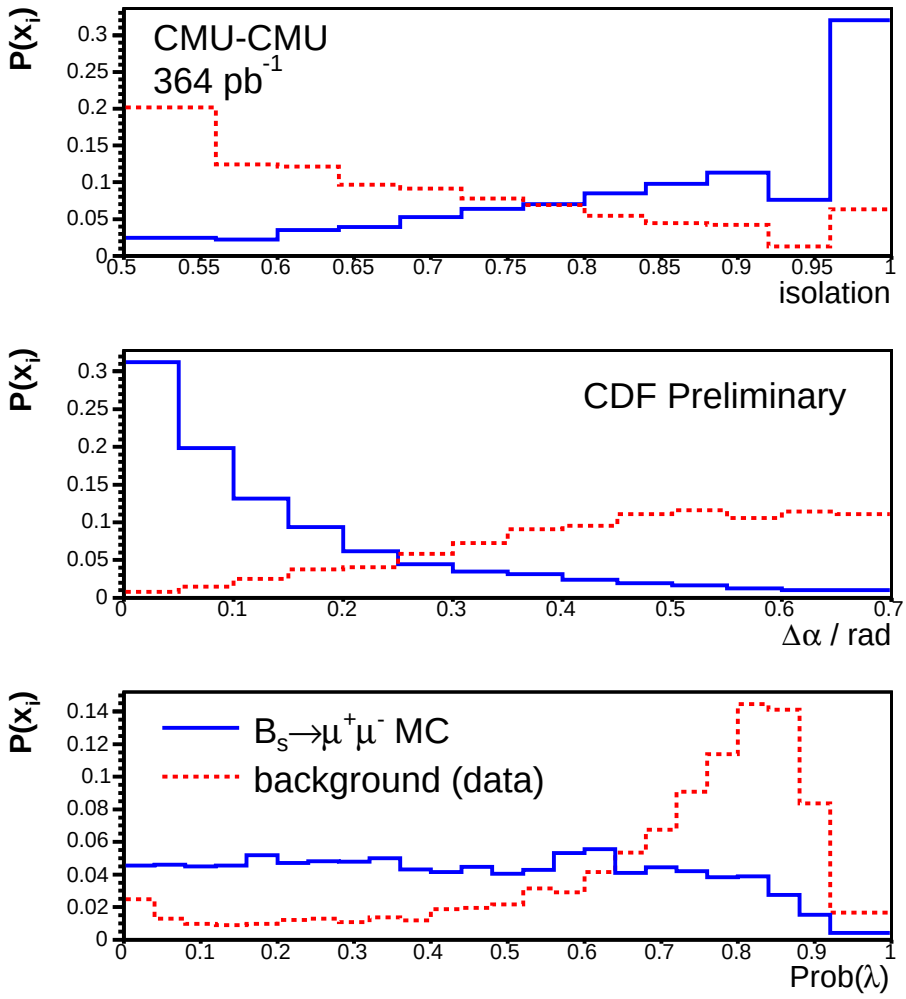

FIG. 42: The I (top), $\Delta\alpha$ (middle), and λ (bottom) probability distributions used to construct the LH variable for the CMU-CMU events.

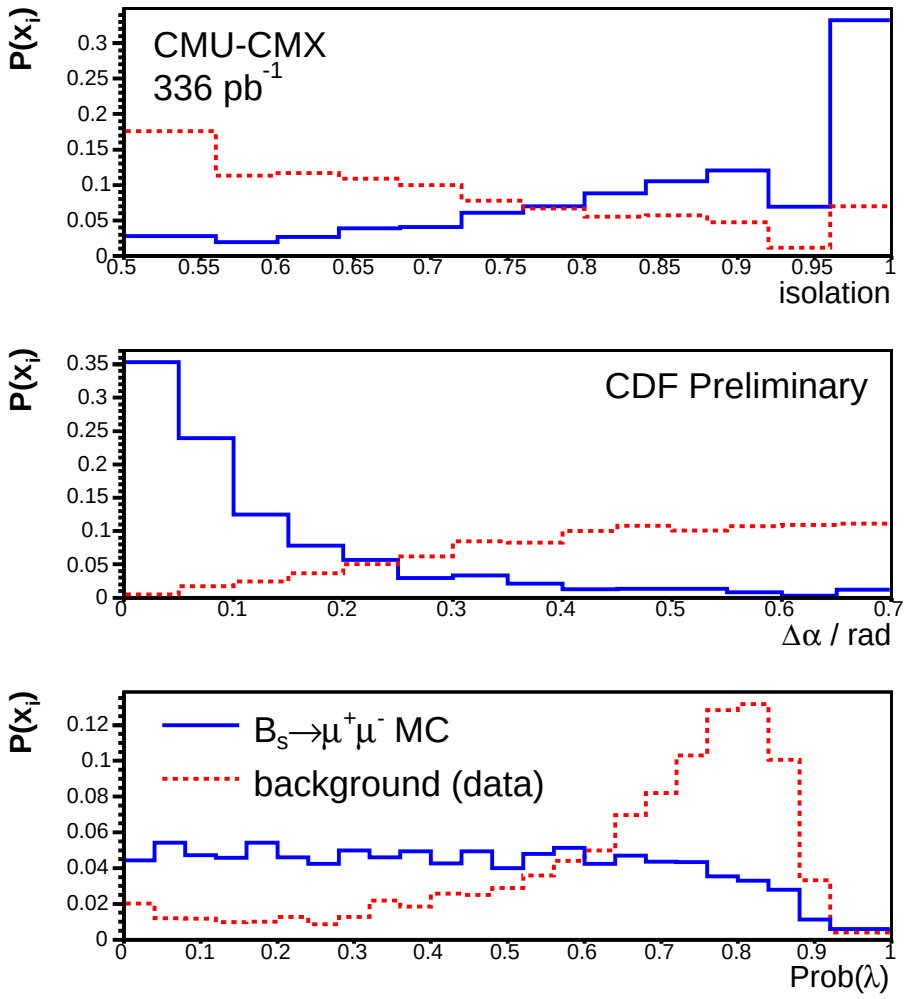

FIG. 43: The I (top), $\Delta\alpha$ (middle), and λ (bottom) probability distributions used to construct the LH variable for the CMU-CMX events.

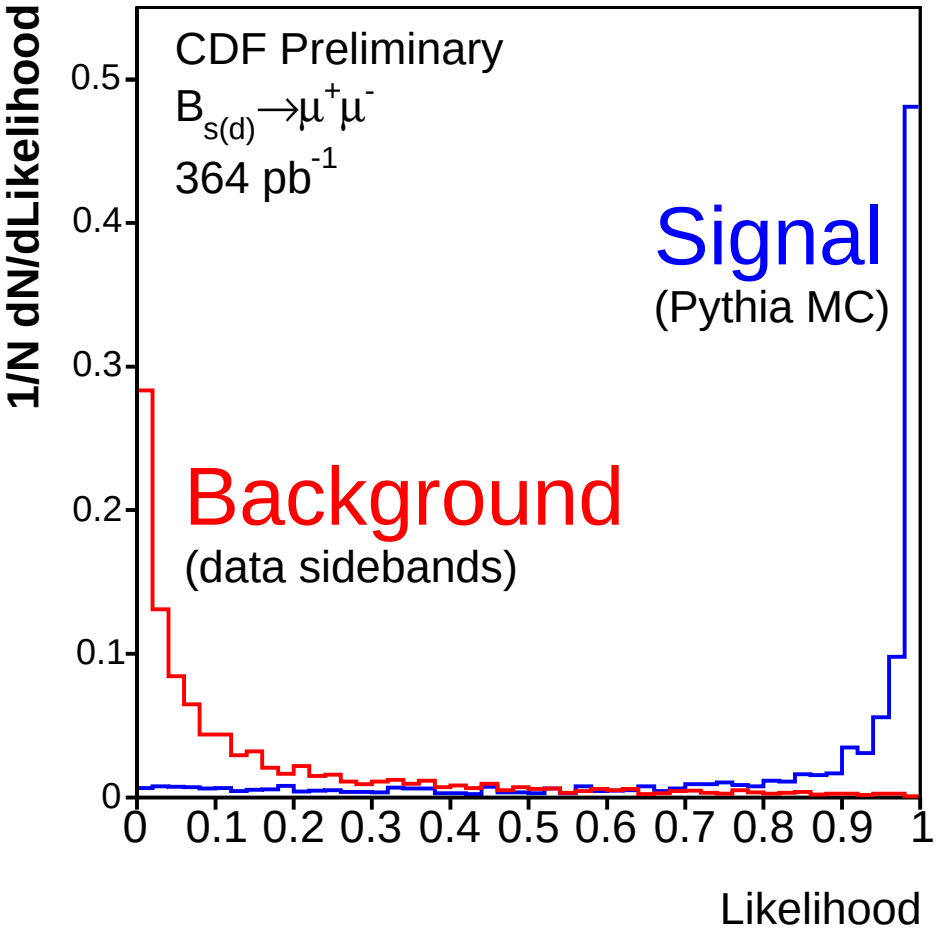

FIG. 44: Likelihood curves for signal ($B_s^0 \to \mu^+\mu^-$ MC) and background (data sidebands) in the CMU-CMU channel.

F. Background estimate

A common method for estimating backgrounds in this type of search is to apply all cuts in the mass sideband regions and scale the resulting number of events to get a background prediction in the signal mass region. The problem with this method is that very few events pass all cuts. This makes optimization difficult since the expected background is statistically consistent over a large fraction of the cut-parameter space. One way to improve the background estimate is to factorize the expected rejection for each cut separately. This will yield an accurate background expectation only when the correlations among the cut variables are small, which is the case here. Thus, the background can be estimated as

$$N_{\rm bg} = N_{SB} \cdot R_{mass} \cdot R_{LH} \tag{4.4}$$

where N_{SB} is the number of sideband events passing the baseline and vertex requirements and having $\lambda > 0$, R_{LH} is the expected background rejection for a given LH cut, and R_{mass} is the ratio of expected number of events in the signal mass-window to a known number of background events in the sideband regions. Since the mass is uncorrelated with the rest of the variables (and hence, with LH), the R_{LH} can be evaluated on samples with very loose cuts, thus reducing the associated uncertainty. This method yields background estimates whose uncertainties are significantly (2-3 times) smaller than for the common method.

The method described above relies on the assumptions that the invariant mass distribution for background events is linear and that the discriminating variables for background events are uncorrelated. Two-body hadronic $B^0_{s(d)} \to h_1^+ h_2^-$ (where $h_i^\pm = \pi^\pm$ or $K^\pm$) decays with both hadrons misidentified as muons violate the assumption of linear mass distribution and can potentially anomalously contribute to

the background. As shown in Appendix H, these decays contribute at the level more than an order of magnitude lower than the current sensitivity to $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. Thus, they can be neglected. Since it is expected that the backgrounds come primarily from the events with heavy flavor production (they naturally have displaced vertices) an additional study of generic heavy flavor production MC events was performed. This study has shown that these events do not contribute anomalously to the background and satisfy the assumptions of the background estimation method employed here, as detailed in Appendix H. Thus, the background should be properly accounted for by the proposed likelihood method.

1. Estimating the mass window cut rejection and a number of sideband events

Given the components defined in Eq. (4.4), the R_{mass} and $N_{\rm SB}$ are estimated first. A linear fit to the $M_{\mu^+\mu^-}$ distribution, for $\mu^+\mu^-$ events in the data passing the baseline and vertex cuts, is shown in Fig. 33 and yields $\chi^2/\text{ndf} = 9/10$ and $4/10$ for the CMU-CMU and CMU-CMX channels, respectively. In this case, if the sidebands are chosen to be symmetric about the signal region, R_{mass} is given by the ratio of the widths of the signal to sideband regions. A signal region is defined as ± 100 MeV/c^2 around the world average B_s^0 mass, $M_{B_s^0} = 5369$ MeV/c^2 [1]. The $M_{\mu^+\mu^-}$ resolution is estimated from the signal MC sample to be about ± 25 MeV/c^2, so this corresponds to a $\pm 4\sigma_m$ window. For now this large window is kept to avoid any bias in the cut optimization, when the signal region is considered "blind" and only sideband information is used. For the final analysis the signal window is shrunk to ± 60 MeV/c^2 (corresponding to $\pm 2.5\sigma_m$). Since the B_d^0 mass is only 90 MeV/c^2 lower, and the cuts will have similar efficiency for $B_d^0 \to \mu^+\mu^-$ decays, a ± 100 MeV/c^2 window around the B_d^0 mass is included in the signal region too. The total signal window is defined as $5.169 < M_{\mu^+\mu^-} < 5.469$ GeV/c^2. The sidebands are then defined symmetrically to

include an additional 0.500 GeV/c^2 on either side of the signal window. The ratio of the widths is then

$$R_{mass} = \frac{\Delta M^{\mathrm{Sig}}_{\mu^+\mu^-}}{\sum_{i=lo}^{hi} \Delta M^{\mathrm{SB_i}}_{\mu^+\mu^-}} = \frac{0.3}{0.5 + 0.5} = 0.3 \tag{4.5}$$

For the final analysis, the mass windows is shrunk to ± 60 MeV/c^2, $R_{mass} = 0.12$. In the data, $N^{UU}_{SB} = 4853$ and $N^{UX}_{SB} = 3768$ for the CMU-CMU and CMU-CMX channels, respectively.

2. Estimating the likelihood cut rejection

The rejection by likelihood cut R_{LH}, the fraction of (background) events satisfying the baseline and vertex requirements that pass a given $LH > x$ cut, is estimated from a likelihood curve simulated using a toy-MC method. The cumulative distribution for each of the likelihood input variables, I, $\Delta\alpha$, and λ is formed using events after the baseline and vertex requirements. For each event, the toy-MC throws a uniform random number for each of these variables and uses the cumulative distributions to extract the set of $(I, \Delta\alpha, \lambda)$. This works because the correlations among these variables are so small that the toy-MC can ignore them and still produce a good estimate of the likelihood curve. By using the toy-MC, it is possible to significantly reduce the statistical uncertainty of the R_{LH} estimate. A comparison of the likelihood curve from data sideband events and the likelihood curve from the toy-MC for $\mu^+\mu^-$ events passing the baseline and vertex requirements but with $\lambda < 0$ is shown in Fig. 45. Using 100k toy-MC events, the R_{LH} for the $B^0_s \to \mu^+\mu^-$ search sample is estimated for a variety of possible cut values. The results are summarized in Table IX.

To quantify the uncertainty associated with the statistics of the data sideband sample used to estimate the cumulative distribution, each bin of the cumulative distribution is simultaneously fluctuated by $\pm 1\sigma$(stat), where σ(stat) is the associated

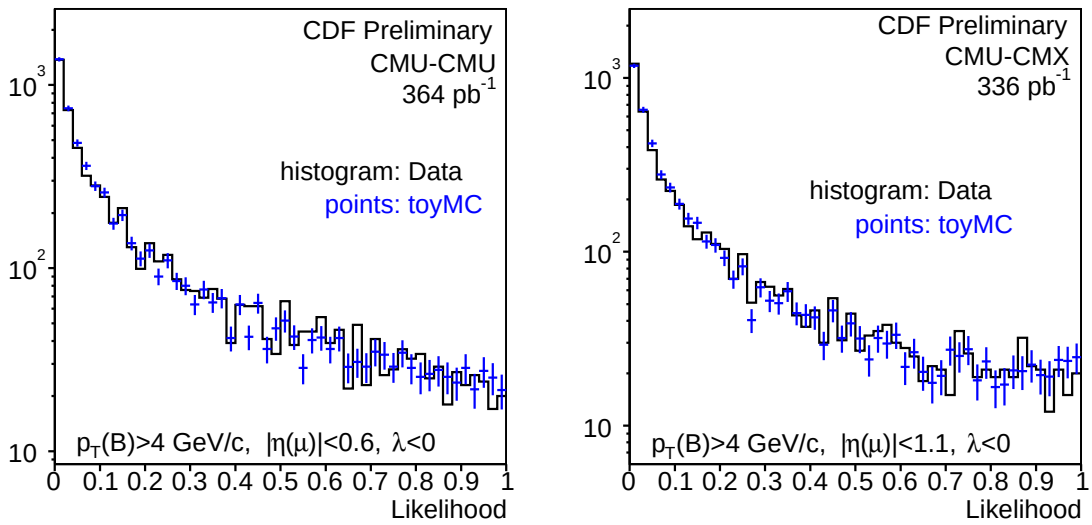

FIG. 45: A comparison of the LH distribution from data sideband events (histogram) and from the toy-MC (points): CMU-CMU (left) and CMU-CMX (right) channels. The KS-probability is 2% (3%) for the left (right).

statistical uncertainty for a particular bin, and R_{LH} is re-evaluated. The bin-to-bin correlations are properly accounted for in this case. Half the observed difference between the two fluctuated samples is assigned as a $\pm 15\%$ and $\pm 19\%$ uncertainty to the R_{LH} estimate in the CMU-CMU and CMU-CMX channel, respectively. In addition, the following variations are considered in determining the R_{LH}: the manner to determine the cumulative distributions used by the toy-MC, the total number of toy-MC events generated, and the random number seed. The R_{LH} determined is then compared to that given in Table IX. All the observed differences were smaller than the assigned statistical uncertainty and no additional systematic uncertainty is thus assigned.

TABLE IX: The LH rejection factor calculated from the $\mu^+\mu^-$ toy-MC. Only the statistical uncertainties are included.

	R_{LH}	
cut value	CMU-CMU	CMU-CMX
$LH > 0.85$	0.0245 ± 0.0005	0.0226 ± 0.0005
$LH > 0.90$	0.0165 ± 0.0004	0.0150 ± 0.0004
$LH > 0.92$	0.0130 ± 0.0004	0.0120 ± 0.0003
$LH > 0.95$	0.0082 ± 0.0003	0.0076 ± 0.0003
$LH > 0.98$	0.0031 ± 0.0002	0.0031 ± 0.0002
$LH > 0.99$	0.0014 ± 0.0001	0.0015 ± 0.0001

3. Cross-checks using control samples

To help build confidence in the method for estimating the background some cross-checks are performed in several control samples described in Section IV.C.3. For each of the control samples it is first verified that there exist no significant correlations among four discriminating variables. The likelihood curve from the toy-MC is also compared with that observed using the data sideband events in each sample. The resulting KS probabilities are given in Table X, which verify that the toy-MC should yield accurate R_{LH} estimates in each of these control samples. Thus, the background estimate discussed in Section IV.F is expected to accurately predict the number of events passing a given likelihood cut in the mass signal region for these samples. These studies are performed for the CMU-CMU and CMU-CMX channels separately. The predicted background is compared with the observed number of events in the mass signal region for $LH > 0.50$, 0.90, and 0.99, which roughly correspond to $R_{LH} = 0.10$, 0.01, and 0.001, respectively. As demonstrated in Table XI, there is no significant

discrepancy between the predicted and observed number of events in either the CMU-CMU or CMU-CMX channel.

TABLE X: The KS test probabilities when comparing the LH curve from toy-MC to that from data observed using the control samples.

	KS Probability	
sample	CMU-CMU	CMU-CMX
OS^+	11%	5%
OS^-	2%	3%
SS^+	86%	99%
SS^-	88%	92%
FM^+	44%	56%

G. Estimate of the signal acceptance and efficiency

At this stage, there is sufficient confidence in the method for estimating the background. Thus, the attention is moved toward optimizing the cuts. A necessary ingredient in the cut optimization is an estimate of the signal efficiency, as discussed below.

The total acceptance times efficiency for $B_s^0 \to \mu^+\mu^-$ decays is estimated as

$$\alpha_{B_s^0} \cdot \epsilon_{B_s^0}^{total} = \alpha_{B_s^0} \cdot \epsilon_{B_s^0}^{trig} \cdot \epsilon_{B_s^0}^{reco} \cdot \epsilon_{B_s^0}^{LH}, \tag{4.6}$$

where $\alpha_{B_s^0}$ is the geometric and kinematic acceptance of the `RAREB` triggers, $\epsilon_{B_s^0}^{trig}$ is the trigger efficiency for events within the acceptance, $\epsilon_{B_s^0}^{reco}$ is the di-muon reconstruction efficiency - including the baseline and vertex requirements - for events passing the

TABLE XI: A comparison of the predicted and observed number of events in the signal mass region as a function of likelihood cut for the various control samples.

		CMU-CMU		CMU-CMX	
sample	LH cut	predicted	observed	predicted	observed
	> 0.50	236 ± 4	235	172 ± 3	168
OS^-	> 0.90	37 ± 1	32	33 ± 1	36
	> 0.99	2.8 ± 0.2	2	3.6 ± 0.2	3
	> 0.50	2.3 ± 0.2	0	2.8 ± 0.3	3
SS^+	> 0.90	0.25 ± 0.03	0	0.44 ± 0.04	0
	> 0.99	< 0.1	0	< 0.1	0
	> 0.50	2.7 ± 0.2	1	3.7 ± 0.3	4
SS^-	> 0.90	0.35 ± 0.03	0	0.63 ± 0.06	0
	> 0.99	< 0.1	0	< 0.1	0
	> 0.50	84 ± 2	84	21 ± 1	19
FM^+	> 0.90	14.2 ± 0.4	10	3.9 ± 0.2	3
	> 0.99	1.0 ± 0.1	2	0.41 ± 0.03	0

trigger, and $\epsilon^{LH}_{B^0_s}$ is the efficiency for $B^0_s \to \mu^+\mu^-$ events to satisfy the likelihood requirement for events after the reconstruction cuts. The analogous expression for $B^+ \to J/\psi K^+$ decays is $\alpha_{B^+} \cdot \epsilon^{total}_{B^+} = \alpha_{B^+} \cdot \epsilon^{trig}_{B^+} \cdot \epsilon^{reco}_{B^+}$, where the terms are defined as for the B^0_s except that $\epsilon^{reco}_{B^+}$ includes the kaon and di-muon reconstruction efficiencies. Note that there is no $\epsilon^{LH}_{B^+}$ term since no likelihood cut is made on the $B^+ \to J/\psi K^+$

events. Using these expressions it is possible to rewrite Eq. (4.1) as

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s^0}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B_s^0}} \cdot \frac{\epsilon_{B_s^0}^{trig}}{\epsilon_{B^+}^{trig}} \cdot \frac{\epsilon_{B_s^0}^{reco}}{\epsilon_{B^+}^{reco}} \cdot \frac{1}{\epsilon_{B_s^0}^{LH}} \cdot \frac{f_u}{f_s} \cdot \mathcal{B}(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+). \tag{4.7}$$

Table XII summarizes the inputs to Eq. (4.7). These inputs are discussed in detail in the following sections.

TABLE XII: A summary of the inputs used in to estimate the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The relative uncertainties are given parenthetically. The single-event-sensitivity, *ses*, is $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ corresponding to $N_{B_s^0} = 1$.

	CMU-CMU		CMU-CMX	
$(\alpha_{B^+}/\alpha_{B_s^0})$	0.297 ± 0.020	($\pm$7%)	0.191 ± 0.013	($\pm$7%)
$(\epsilon_{B^+}^{reco}/\epsilon_{B_s^0})$	0.921 ± 0.034	($\pm$4%)	0.915 ± 0.034	($\pm$4%)
$\epsilon_{B_s^0}^{LH}$	0.348 ± 0.035	($\pm$10%)	0.360 ± 0.022	($\pm$6%)
N_{B^+}	1785 ± 60	($\pm$3%)	696 ± 39	($\pm$6%)
f_u/f_s		3.83 ± 0.57		($\pm$15%)
$\mathcal{B}(B^+ \to J/\psi K^+)\cdot$ $\mathcal{B}(J/\psi \to \mu^+\mu^-)$		$(5.88 \pm 0.26) \times 10^{-5}$		($\pm$4%)
ses	$(1.0 \pm 0.2) \times 10^{-7}$	($\pm$20%)	$(1.6 \pm 0.3) \times 10^{-7}$	($\pm$19%)

1. *Acceptance*

The acceptance is defined as a fraction of true signal production events that satisfy detector and trigger fiduciality requirements, with the contribution from the detector fiducial selections defining the geometric part of the total acceptance. The $B_s^0 \to \mu^+\mu^-$ acceptance is defined as the fraction of $B_s^0 \to \mu^+\mu^-$ events with $p_T(B_s^0) >$ 4 GeV/c and $|y(B_s^0)| < 1$ that satisfy the $\mu^+\mu^-$ fiduciality requirements described below. The $B^+ \to J/\psi K^+$ acceptance is defined as the fraction of $B^+ \to J/\psi K^+$ events with $p_T(B^+) > 4$ GeV/c and $|y(B^+)| < 1$ that satisfy the $\mu^+\mu^-$ and kaon require-

ments described below. Both are evaluated using the PYTHIA samples described in Section IV.D.

In calculating the acceptance both the numerator and the denominator is defined by the selections made using the simulation level information described in Section III.E.2. This is done in order to exclude the effects of reconstruction resolution and efficiency. The denominator of the acceptance is defined as the number of events generated with the $p_T(B_q) > 4$ GeV/c and $|y(B_q)| < 1$ requirements. The acceptance numerator is largely driven by the geometric acceptance and kinematic requirements of the `RAREB` trigger paths. The B_s^0 transverse momentum and rapidity distributions for events generated with $|y_b| < 1.5$ and $p_T^b > 2$ GeV/c, where y_b and p_T^b are the generator level b-quark rapidity and transverse momentum, are shown in Fig. 46.[4]

[4]Since the input constraints in the MC generator are made at the parton level, before fragmentation and hadronization, to avoid sculpting the final B_s^0 meson momentum and rapidity distribution the events had to be generated with the loose input selections.

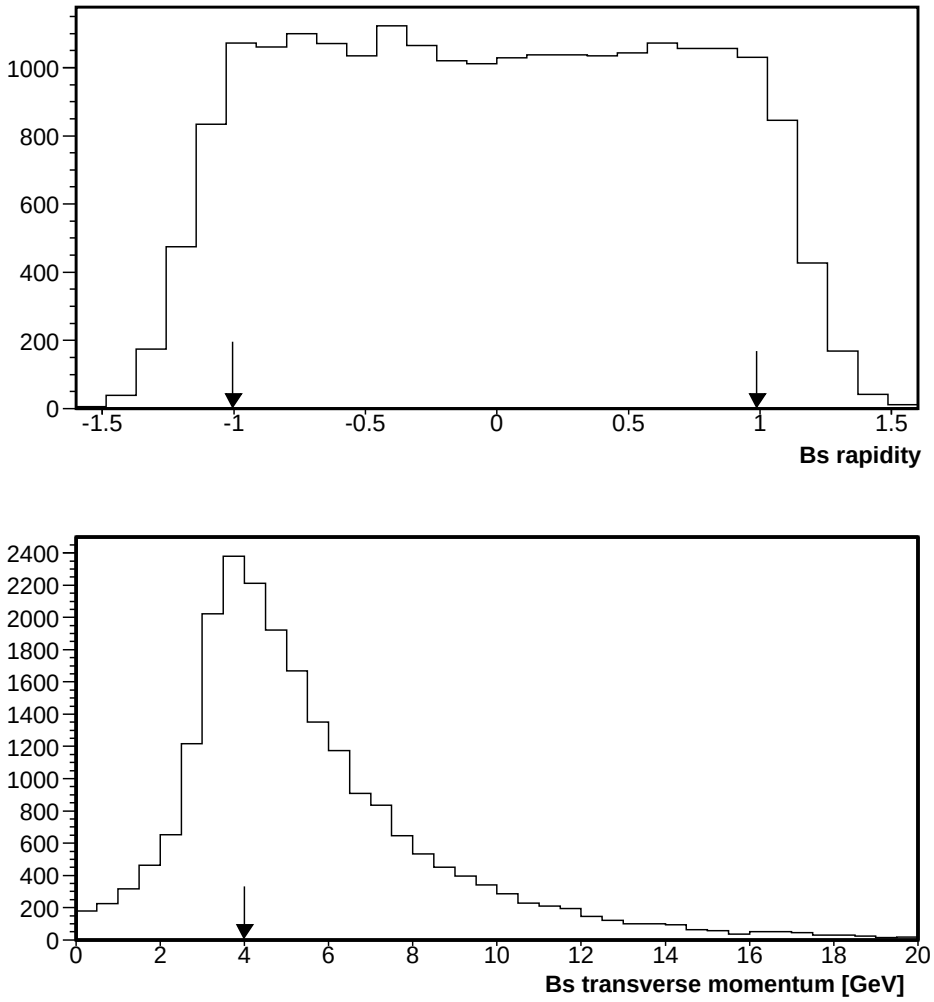

FIG. 46: Rapidity and transverse momentum distributions for B_s^0 mesons at the generator level.

The geometric portion of the acceptance (geometric acceptance) defines a fraction of events that can possibly be identified using the CMU and CMX muon systems and COT and silicon trackers, assuming no trigger requirements and an absolute reconstruction efficiency. The α_{geom} is determined by the following requirements. The muons are required to originate from the central region, that is to have $|z_0| < 60$ cm.

The muons should traverse at least 3 SVX layers, which is determined by counting the number of intersections of the muon helix with the silicon wafers using the detailed silicon geometry. In addition, each generator level muon should be expected to cross either CMU or CMX muon chambers, which is determined from the information provided by GEANT3 detector simulation as described in Section III.E.2. The geometric acceptance for the CMU and CMX muon systems combined is shown in Fig. 47 as a function of $p_T^{B_s^0}$ and in Fig. 48 as a function of $\eta^{B_s^0}$. The same z_0, and SVX requirements are made for the B^+ kaons. The corresponding acceptance distributions are given in Figs. 49 and 50. The combined geometric acceptance for $p_T(B_q) > 4$ GeV/c and $|y(B_q)| < 1$ is about 21%(9.8%) for $B_s^0(B^+)$ mesons.

The (full) acceptance, α, determines the fraction of produced events that are expected to pass the trigger and to correspond to the good quality tracks and muons after the detector reconstruction, assuming an absolute trigger and reconstruction efficiency. Thus, the numerator of the acceptance measurement is defined as a number of events passing the trigger fiducial, silicon, track and muon quality requirements specified in Table V, with the requirements on $\chi^2_{r\phi}$, and number of silicon and COT hits omitted.[5] The acceptance kinematic dependence is shown in Figs. 47 and 48 for B_s^0 and in Figs. 49 and 50 for B^+ mesons. The α_{B^+} is substantially smaller than $\alpha_{B_s^0}$ due to lower momenta of the muons from $J/\psi \to \mu^+\mu^-$ decays. The absolute values of the acceptance are $\alpha^{\rm UU}_{B_s^0} \approx 5.5\%$ and $\alpha^{\rm UX}_{B_s^0} \approx 4.5\%$ for B_s^0 mesons, and $\alpha^{\rm UU}_{B^+} \approx 1.64\%$ and $\alpha^{\rm UX}_{B^+} \approx 0.85\%$ for B_s^0 mesons. The B_s^0 and B^+ acceptance ratios are found to be $\alpha_{B^+}/\alpha_{B_s^0} = 0.297 \pm 0.008$ and $\alpha_{B^+}/\alpha_{B_s^0} = 0.191 \pm 0.006$ for the CMU-CMU and CMU-CMX channels, respectively, where only the statistical uncertainties have been included.

[5]The effect of these omitted requirements is included in the total efficiency value.

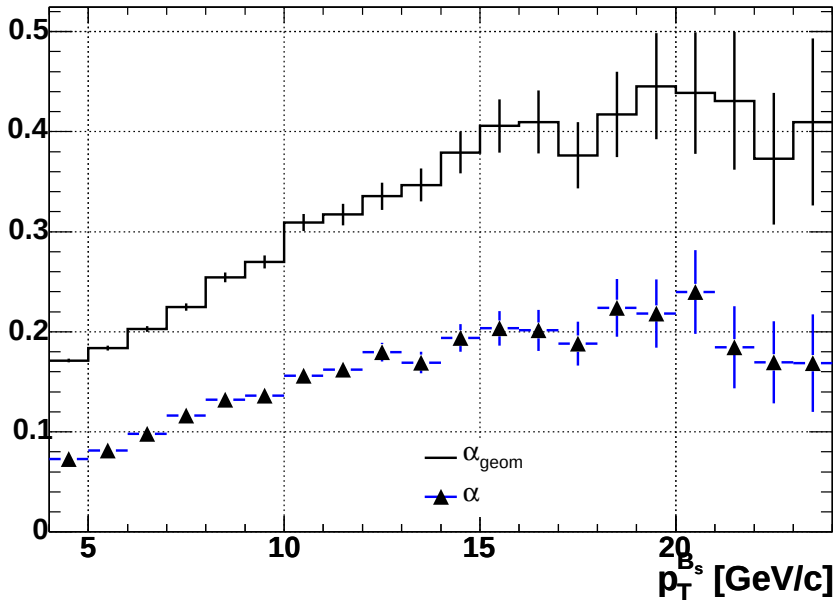

FIG. 47: Total (▲) and geometric (histogram) acceptance for B_s^0 vs. $p_T^{B_s^0}$.

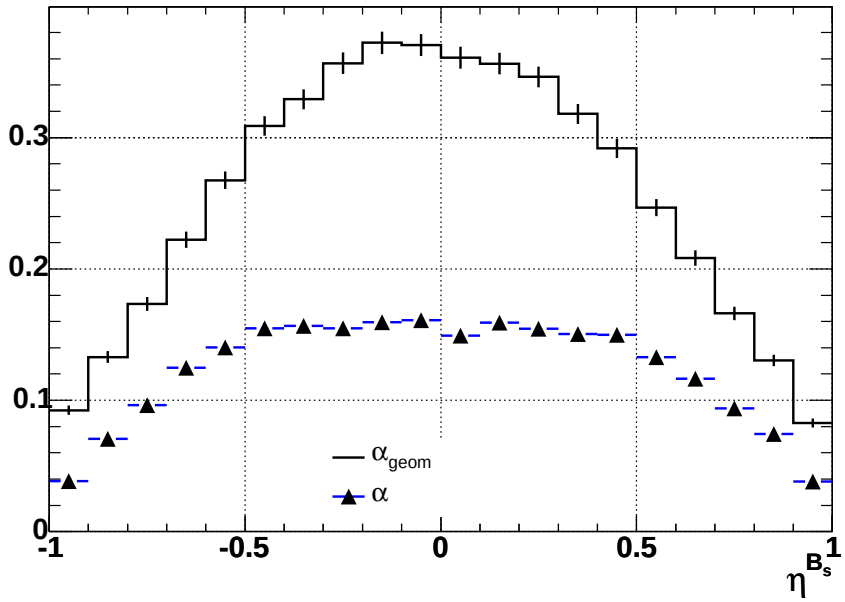

FIG. 48: Total (▲) and geometric (histogram) acceptance for B_s^0 vs. $\eta^{B_s^0}$.

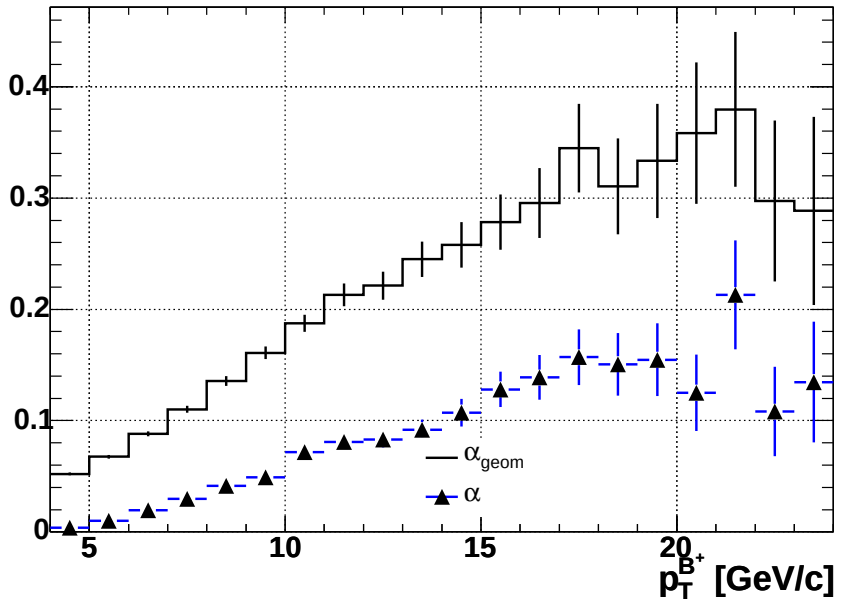

FIG. 49: Total (▲) and geometric (histogram) acceptance for B^+ vs. $p_T^{B^+}$.

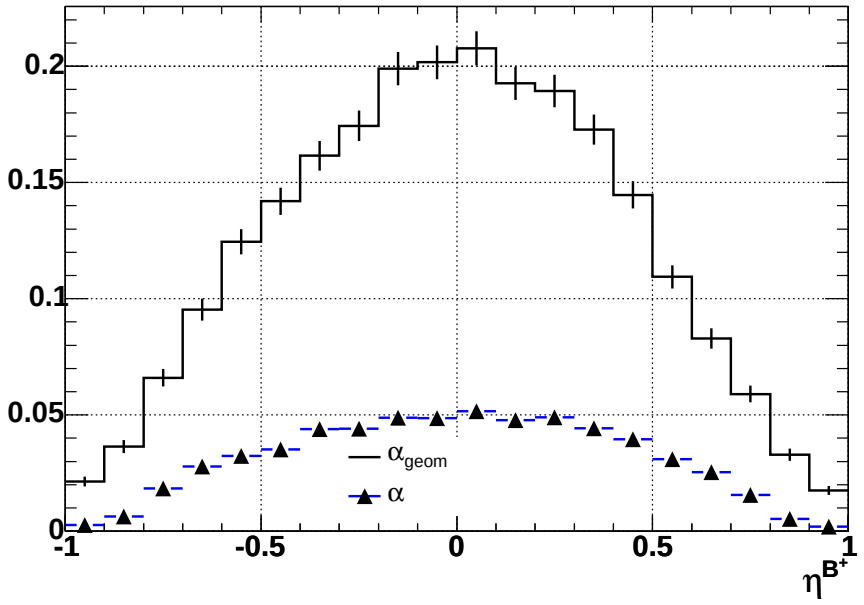

FIG. 50: Total (▲) and geometric (histogram) acceptance for B^+ vs. η^{B^+}.

As discussed in Section III.E.2, since they are much faster to generate, BGENERATOR [119] samples are used to evaluate systematic uncertainties due to variations in the b-quark mass, fragmentation modeling, and the renormalization scale. The variations are defined by the uncertainties on the corresponding values: the b-quark mass, $m_b = 4.74 \pm 0.2$ GeV/c^2; the Peterson fragmentation parameter, $\epsilon = 0.006 \pm 0.002$; and the QCD renormalization scale, $\mu_0 \to 2\mu_0$ and $\mu_0/4$. The samples are generated using the same $p_T(B)$ spectrum as used to generate the PYTHIA samples. The default PYTHIA and BGENERATOR samples yield acceptances that are consistent within 1% (5%) relative for the CMU-CMU (CMU-CMX) channel. For each systematic, the ratio of $B^+ \to J/\psi K^+$ to $B_s^0 \to \mu^+\mu^-$ acceptances for the $+1\sigma$ and -1σ samples is calculated. The resulting ratios are given in Table XIII normalized to the acceptance ratio obtained from the default samples. For each systematic sample generated, $\Delta\alpha$ is the ratio of $\alpha_{B^+}/\alpha_{B_s^0}$ for that sample normalized to the same ratio as determined from the default samples. Note that all are statistically consistent with 1. The typical statistical uncertainty of these comparisons are assigned as the associated systematic uncertainty, which is $\pm 6\%$ ($\pm 7\%$) relative, for the CMU-CMU (CMU-CMX) channel.

2. *Trigger efficiencies*

The relevant L1, L2, and L3 trigger efficiencies are all determined as described in Appendix F and summarized here. The L1 single-muon efficiency, ϵ_μ^{L1}, is measured using the $J/\psi \to \mu^+\mu^-$ events triggered at L1 by a single-muon trigger, thus giving a sample in which the second muon is not biased by the L1 trigger decision. The L1 efficiency is defined as a fraction of unbiased muons that are triggered at L1. The L1 single-muon efficiency is parameterized as a function $\epsilon_\mu^{L1}(run, \mathbf{\Omega}) \equiv \epsilon_\mu^{L1}(run, p_T^\mu, |\eta^\mu|, \phi^\mu)$ for the CMU and CMX separately. The dimuon efficiency rele-

TABLE XIII: Normalized acceptance ratio variation in the generated systematic samples. Deviations from 1 quantify the associated systematic uncertainty.

sample		CMU-CMU $\Delta\alpha$	CMU-CMX $\Delta\alpha$
mass	$+1\sigma$	0.97 ± 0.06	0.92 ± 0.07
	-1σ	1.09 ± 0.06	1.02 ± 0.08
fragmentation	$+1\sigma$	0.94 ± 0.06	0.96 ± 0.07
	-1σ	1.00 ± 0.06	0.92 ± 0.07
renormalization	$+1\sigma$	0.94 ± 0.06	0.92 ± 0.07
	-1σ	1.07 ± 0.06	0.92 ± 0.07

vant for this analysis is estimated by the convolution of the parameterization with the $(p_T^{\mu+}, |\eta^{\mu+}|, \phi^{\mu+}, p_T^{\mu-}, |\eta^{\mu-}|, \phi^{\mu-}) = (\mathbf{\Omega}_+, \mathbf{\Omega}_-)$ spectra of $B_s^0 \to \mu^+\mu^-$ or $B^+ \to J/\psi K^+$ MC events satisfying the acceptance criteria of Section IV.G.1. The final di-muon L1 efficiency is then the luminosity weighted average of these:

$$\epsilon^{L1}_{\mu^+\mu^-} = \sum_i w_i \int \epsilon^{L1}_{\mu+}(run_i, \mathbf{\Omega}_+) d\mathbf{\Omega}_+ \int \epsilon^{L1}_{\mu-}(run_i, \mathbf{\Omega}_-) d\mathbf{\Omega}_- \tag{4.8}$$

where $i \in$ [all good runs], $w_i = \frac{\mathcal{L}_i}{\sum_j \mathcal{L}_j}$, $\mathcal{L}_i$ is the integrated luminosity for the ith run, and $\epsilon^{L1}_{\mu^\pm}(run_i, \mathbf{\Omega}_\pm)$ is the parameterization of the relevant L1 single-muon efficiency for the ith run. For the CMU-CMU triggers, the parameterizations are the same for each muon and are thus correlated, while for the CMU-CMX triggers they are different and thus taken as uncorrelated. The resulting ratio of efficiencies is $\epsilon^{L1}_{\mu^+\mu^-}(B^+)/\epsilon^{L1}_{\mu^+\mu^-}(B_s^0) = 0.9954 \pm 0.0003$ (0.9889 ± 0.0003) for the CMU-CMU (CMU-CMX) channel including statistical and systematic uncertainties, which are assumed 100% correlated in the ratio.

The L2 (L3) efficiencies are measured using the sample of events collected on a trigger path that bypasses the L2 (L3) trigger decision, the details of the measurement are given in Appendix F. The L2 and L3 efficiencies depend only on run number. Luminosity and acceptance weighting is used to estimate the final L2 and L3 efficiencies, which are found to be the same for the $B_s^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ decays: $\epsilon_{UU}^{L2} = 0.9997^{+0.0003}_{-0.0016}$, $\epsilon_{UX}^{L2} = 0.9986^{+0.0014}_{-0.0017}$, $\epsilon_{UU}^{L3} = 0.989 \pm 0.022$, and $\epsilon_{UX}^{L3} = 0.980 \pm 0.015$, including statistical and systematic uncertainties. The ratio of L2 and L3 efficiencies are equal to 1.000 with total uncertainties that are $<< 1\%$ and thus negligible.

The total trigger efficiency is the product of the L1, L2, and L3 efficiencies. The final ratio of trigger efficiencies is then $\epsilon_{B^+}^{trig}/\epsilon_{B_s^0}^{trig} = 0.9954 \pm 0.0003$ and 0.9889 ± 0.0003 for the CMU-CMU and CMU-CMX channels, respectively, including statistical and systematic contributions.

3. Reconstruction efficiencies

The reconstruction efficiency accounts for the efficiency of the baseline and vertex requirements of Section IV.C. It is estimated as the product of several factors:

$$\epsilon^{reco} = \epsilon^{COT}\epsilon^{muon}\epsilon^{SVX}\epsilon^{vtx} \tag{4.9}$$

For $B_s^0 \to \mu^+\mu^-$ decays, ϵ^{COT} is the probability that the $\mu^+\mu^-$ satisfying the acceptance criteria of Section IV.G.1 are reconstructed in the COT and pass the "COT quality" requirements, $\epsilon^{muon} \equiv \epsilon^{\mu\mathrm{reco}}\epsilon^{\mu\chi^2}$ is the fraction of $\mu^+\mu^-$ satisfying the COT criteria that remain after the muon reconstruction ($\epsilon^{\mu\mathrm{reco}}$) and "muon quality" ($\epsilon^{\mu\chi^2}$) requirements, and ϵ^{SVX} is the fraction of $\mu^+\mu^-$ satisfying the COT and muon criteria and passing the "SVX quality" requirements. For $B^+ \to J/\psi K^+$ decays, the $\mu^+\mu^-$ reconstruction efficiencies are defined in exactly the same manner as described

for the $B_s^0 \to \mu^+\mu^-$ decays. Additionally ϵ_K^{COT} is defined as the fraction of kaons from $B^+ \to J/\psi K^+$ decays satisfying the acceptance criteria of Section IV.G.1 that are reconstructed in the COT and pass the COT quality requirements, and ϵ_K^{SVX} as the fraction of these that additionally satisfy the SVX quality requirements. In both cases ϵ^{vtx} is the fraction of fully reconstructed events which pass the relevant vertex requirements described in Sections IV.C.1 and IV.C.2. The resulting ratio of reconstruction efficiencies can then be expressed as:

$$\begin{aligned}\epsilon_{B^+}^{reco}/\epsilon_{B_s^0}^{reco} &= \left(\frac{\epsilon_{B^+}^{COT}}{\epsilon_{B_s^0}^{COT}}\frac{\epsilon_{B^+}^{muon}}{\epsilon_{B_s^0}^{muon}}\frac{\epsilon_{B^+}^{SVX}}{\epsilon_{B_s^0}^{SVX}}\right)\left(\frac{\epsilon_{B^+}^{vtx}}{\epsilon_{B_s^0}^{vtx}}\right)\epsilon_K^{COT}\,\epsilon_K^{SVX} \\ &= \left(\frac{\epsilon_{B^+}^{reco-\mu^+\mu^-}}{\epsilon_{B_s^0}^{reco-\mu^+\mu^-}}\right)\left(\frac{\epsilon_{B^+}^{vtx}}{\epsilon_{B_s^0}^{vtx}}\right)\epsilon_{B^+}^{reco-K}\end{aligned} \tag{4.10}$$

The first term is the ratio of $B^+ \to J/\psi K^+$ to $B_s^0 \to \mu^+\mu^-$ di-muon reconstruction efficiencies and is expected to be close to one. The second term is the ratio of $B^+ \to J/\psi K^+$ to $B_s^0 \to \mu^+\mu^-$ vertex requirement efficiencies - this will be different from one due to lifetime differences and the additional vertex requirements for the $B^+ \to J/\psi K^+$ channel. The last term accounts for the reconstruction efficiency of the kaon and is unique to the $B^+ \to J/\psi K^+$ decay. Each of these is discussed separately.

a. Dimuon reconstruction efficiencies

The dimuon reconstruction efficiency for either B_s^0 or B^+ is defined as $\epsilon^{reco-\mu^+\mu^-} = \epsilon^{COT}\epsilon^{\mu\text{reco}}\epsilon^{\mu\chi^2}\epsilon^{SVX}$. Each component is a product of single muon reconstruction efficiencies and can be defined (in general case of a measured dependence on certain variables) similar to the definition of the dimuon L1 trigger efficiency based on the single muon L1 efficiency given in Eq. (4.8). The measurement of the contribution from a single muon to each dimuon efficiency component are described below first, followed by the brief description of the results of the measurements of the ratios of

the dimuon components, conlcuded with the final value of the ratio of the dimuon reconstruction efficiencies used as input in Eq. (4.10).

The COT reconstruction efficiency is measured using a COT MC hit embedding technique. For the single muon COT efficiency measurement, $\epsilon_{\mu}^{\mathrm{COT}}$, the denominator includes the set of events from the dimuon trigger. A simulated MC muon track is added to (embedded into) these events. The number of hits and hit resolution for the embedded track is simulated distributed according to (tuned to) the distribution measured for the muon tracks unbiased by the L1 trigger decision and selected from the $J/\psi \to \mu^+\mu^-$ sample.[6] The events with the embedded tracks are then passed through the standard tracking.[7] The numerator of the efficiency measurement is the number of events with a COT track reconstructed with track helix parameters within 5σ of the embedded track parameters or with a majority of the generated hits attached to a COT track. For the muons with $p_T > 1.5$ GeV/c the efficiency is $99.61 \pm 0.02(\mathrm{stat})\ ^{+0.34}_{-0.91}(\mathrm{syst})$, where the dominant source of systematic uncertainty comes from the efficiency dependence on the track isolation. The COT reconstruction efficiency is discussed in detail in Ref. [136].

The muon reconstruction efficiency $\epsilon^{\mu\mathrm{reco}}$ is defined as a product of single muon reconstruction efficiencies, $\epsilon^{\mu 1\mathrm{reco}}$, for the two muons. The single muon reconstruction efficiency is estimated as a ratio of the raw reconstruction efficiency measured in data, $\epsilon^{\mu 1\mathrm{reco}}_{\mathrm{raw,dt}}$, to that measured in MC, $\epsilon^{\mu 1\mathrm{reco}}_{\mathrm{raw,MC}}$. Here the denominator of the raw efficiency measurement is defined by the number of tracks pointing into the given muon subdetector, where the tracks are consistent (by invariant mass) with being coming from $J/\psi \to \mu^+\mu^-$ decay [137] or $Z \to \mu^+\mu^-$ decay [138]. The numerator

[6]The events with muons unbiased by the L1 trigger are the same as those used for L1 trigger efficiency measurement described in Section F.C.1.

[7]With additional track to MC hit association methods added.

of the raw efficiency measurement is then defined by the number of events in the denominator which have a muon reconstructed as described in Section C.B. The muons that do not reach the muon detector due to the absorption in the calorimeter and the muons that fall into the gaps between chambers are already excluded by MC in the acceptance definition, they are also the natural part of the raw inefficiency value. To avoid this double-counting the raw efficiency in data should be divided by that in MC.[8] The momentum dependence of the raw efficiency is primarily due to the higher chance for a muon to be absorbed by the calorimeter. This dependence duly cancels for $\epsilon^{\mu 1\mathrm{reco}} \equiv \epsilon^{\mu 1\mathrm{reco}}_{\mathrm{raw,dt}}/\epsilon^{\mu 1\mathrm{reco}}_{\mathrm{raw,MC}}$. The muon reconstruction efficiency is thus a constant, it is approximately equal to 99%, 94%, and 97% for CMU, CMUP, and CMX muons respectively. A detailed description of the muon reconstruction efficiency can be found in Refs. [137, 138].

The efficiency of the muon $\chi^2_{r\phi}$ requirement (single muon component of $\epsilon^{\mu\chi^2}$) is defined as a fraction of muons that pass the track to stub matching requirement, $\chi^2_{r\phi} < 9$, as mentioned in Section IV.C. The $\chi^2_{r\phi}$ variable for the given muon detector subsystem is defined as $\chi^2_{r\phi} \equiv (\Delta x/\sigma_{\Delta x})^2$, where Δx is the muon track to stub matching variable discussed in Section C.B, and $\sigma_{\Delta x}$ is the width of the matching variable assuming its Gaussian distribution. Due to the multiple scattering in the material the $\sigma_{\Delta x}$ is dependent on muon kinematics. The construction of $\chi^2_{r\phi}$ variable and the measurement of the efficiency of $\chi^2_{r\phi} < 9$ requirement are detailed in Ref. [133] for CMU muons and in Appendix G for CMX muons. The corresponding efficiencies are p_T-dependent and are within $99 \pm 1\%$ range, as shown in Figs. 51 and 52 for CMU and CMX respectively.

[8]The muon simulation assumes that the chambers are 100% efficient and have an absolute resolution, the "real" muon reconstruction efficiency in MC is thus 100%.

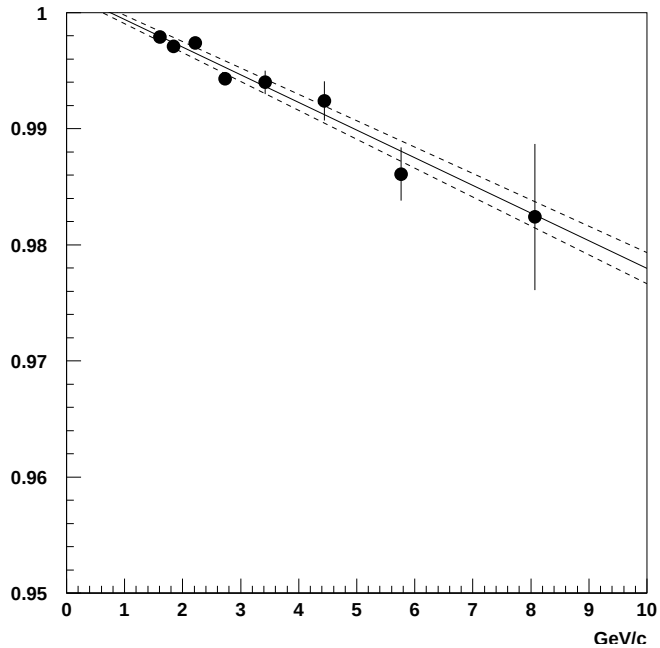

FIG. 51: Efficiency of $\chi^2_{r\phi} < 9$ for CMU muons vs p_T.

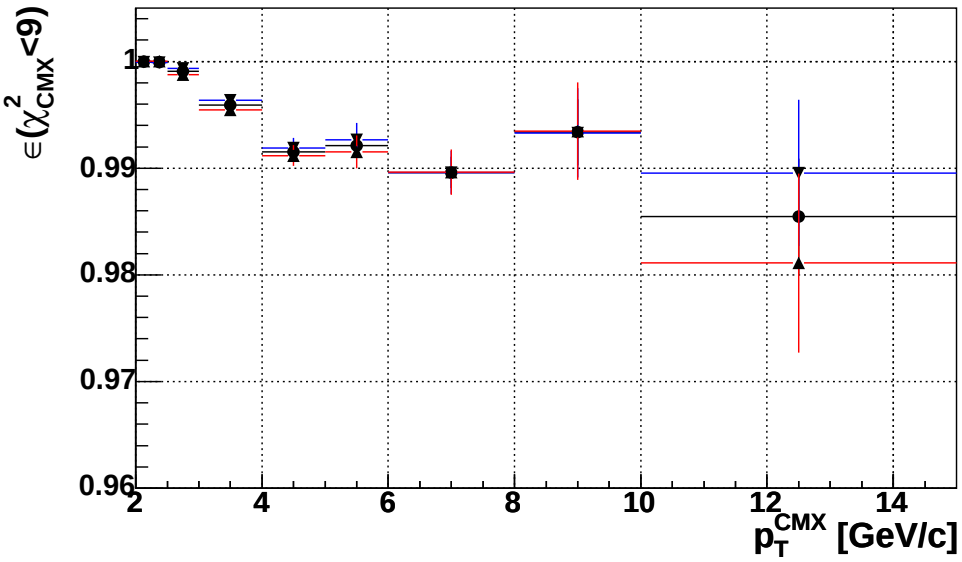

FIG. 52: $\epsilon(\chi^2_{r\phi,\mathrm{CMX}} < 9)$ vs. p_T: μ^+ (▲), μ^- (▼), and charge averaged (•).

The measurement of SVX reconstruction efficiency (the single muon component of ϵ^{SVX}) is made based on the muons from $J/\psi \rightarrow \mu^+\mu^-$ decays. The denominator is defined by the number of tracks satisfying the baseline requirements listed in Table V, less the "SVX quality" requirement (three layers crossed is still required, since it is in the definition of the acceptance). The numerator of the measurement is the number of tracks that satisfy the silicon quality requirement of Table V. The single muon track SVX efficiency, $\epsilon_\mu^{\mathrm{SVX}}$, is measured as a function of run number binned into consistent subsections, as shown in Fig. 53. The subsections are the continuous run periods with relatively small variation within a period, the systematic uncertainty due to the variations within bins is 0.7%. Changes of the $\epsilon_\mu^{\mathrm{SVX}}$ as a function of track transverse momentum, opening angle in r-ϕ view, and dimuon track isolation are considered as sources of additional systematics. Half the variation in each of those variables is considered as source of systematic uncertainty. The systematics is the largest due to the variation in isolation, as shown in Fig. 54, it is 2.5%. The systematics due to the variation in transverse momentum and the opening angle are 0.5% and 1.2% respectively. The total systematic uncertainty is estimated as a quadrature sum of the above contributions. Since the statistical uncertainty is negligible, the total uncertainty on the $\epsilon_\mu^{\mathrm{SVX}}$ is 3.4%, dominated by the systematics. The SVX reconstruction efficiency is discussed in detail in Ref. [139]. The dimuon SVX efficiency is given by $(\epsilon_\mu^{\mathrm{SVX}})^2$.

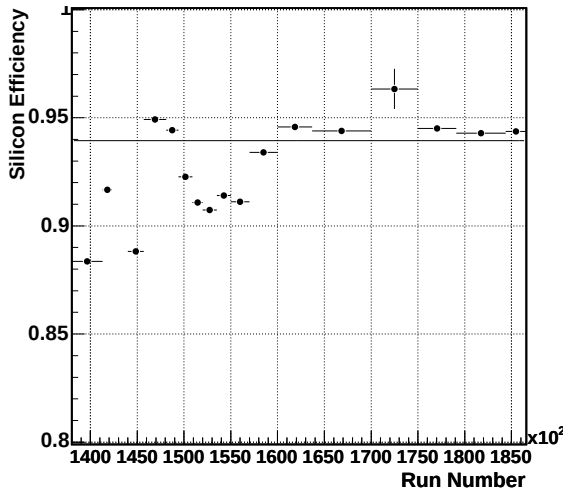

FIG. 53: Silicon reconstruction efficiency per muon vs. run number.

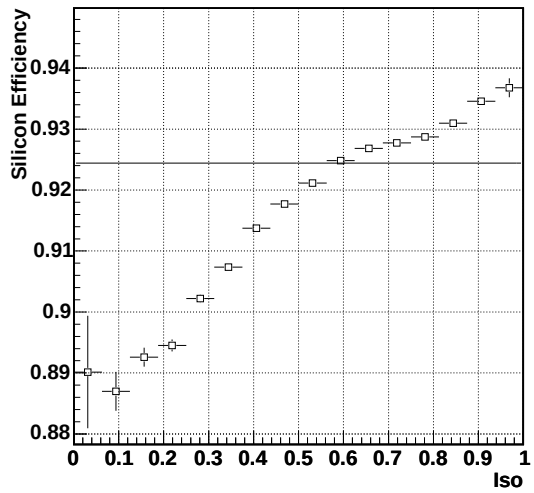

FIG. 54: Silicon reconstruction efficiency per muon vs. dimuon track isolation.

Since the COT and SVX efficiencies for $p_T > 2$ GeV/c depend only on run number they are easily calculated for dimuons. The resulting ratios are $\epsilon^{COT}_{B^+}/\epsilon^{COT}_{B^0_s} = 1.00 \pm 0.01$ and $\epsilon^{SVX}_{B^+}/\epsilon^{SVX}_{B^0_s} = 1.00 \pm 0.03$ including the statistical and systematic uncertainties for both the CMU-CMU and CMU-CMX channels.

While the (di)muon reconstruction efficiencies are constants and cancel in the ratio, the efficiency of the chi-squared track-stub matching requirements, $\epsilon^{\mu\chi^2}$, is p_T dependent. The di-muon p_T spectra are convoluted for events within the acceptance that pass the trigger and that have muons which each satisfy the COT quality criteria with the parameterizations reported in Ref. [133] for CMU muons and in Appendix G for CMX muons. The efficiencies are about 98% for the CMU and 99% for the CMX muons in both the $B^+ \to J/\psi K^+$ and $B^0_s \to \mu^+\mu^-$ decays. The resulting $B^+ \to J/\psi K^+$ to $B^0_s \to \mu^+\mu^-$ ratios are $\epsilon^{muon}_{B^+}/\epsilon^{muon}_{B^0_s} = \epsilon^{\mu\chi^2}_{B^+}/\epsilon^{\mu\chi^2}_{B^0_s} = 1.003$ and 1.002 for the CMU-CMU and CMU-CMX channels, respectively. The total associated

uncertainties are $<< 1\%$ and thus negligible.

The ratio of total di-muon reconstruction efficiencies is the product of the COT, $\mu\chi^2$, and SVX ratios reported above. It was found that $\epsilon_{B^+}^{reco-\mu^+\mu^-}/\epsilon_{B_s^0}^{reco-\mu^+\mu^-} = 1.00 \pm 0.03$ for both CMU-CMU and CMU-CMX channels, including statistical and systematic uncertainties.

b. Vertex efficiencies

The efficiencies for the vertex requirements, $\epsilon_{B^+}^{vtx}$ and $\epsilon_{B_s^0}^{vtx}$, are estimated using the default MC samples. For simplicity, the efficiency of the mass window requirements is included here too. It is found that $\epsilon_{B^+}^{vtx} = (75.7 \pm 1.0)\%$ and $\epsilon_{B_s^0}^{vtx} = (76.8 \pm 0.3)\%$ which yield a ratio of $\epsilon_{B^+}^{vtx}/\epsilon_{B_s^0}^{vtx} = 0.986 \pm 0.013$, for both the CMU-CMU and CMU-CMX channels and including the statistical uncertainties only. Since, by comparing the vertex efficiencies for the $J/\psi \to \mu^+\mu^-$ decays from $B^+ \to J/\psi K^+$ candidates between the MC and sideband subtracted data, it is observed that the MC accurately estimates these efficiencies, no additional systematic uncertainty is assigned to this ratio.

c. Kaon reconstruction efficiencies

The kaon reconstruction efficiency, $\epsilon_{B^+}^{reco-K}$, is a product of kaon COT efficiency, ϵ_K^{COT}, and SVX efficiency ϵ_K^{SVX}. The measurement of theese components is described below.

The COT reconstruction efficiency for kaons satisfying the fiducial requirements and having $p_T(K) > 1$ GeV/c, ϵ_K^{COT}, is estimated using the embedding technique of Ref. [136], the same as the one described in the previous section. The resulting efficiency curve as a function of $p_T(K)$ is shown in Fig. 55. The systematic uncertainty on ϵ_K^{COT} is estimated from the variations of the value as a function of track isolation (I), instantaneous luminosity (both in relation with the hit occupancy ef-

fects) and the kaon nuclear interaction cross-section value, σ_{KN}, used in GEANT3 MC embedding technique. The $p_T(K)$ spectrum of fiducial kaons from MC $B^+ \to J/\psi K^+$

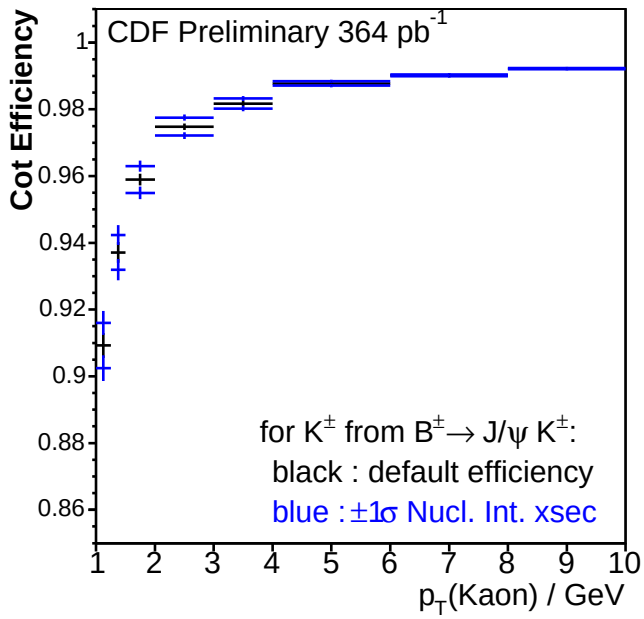

FIG. 55: The COT reconstruction efficiency for kaons from $B^+ \to J/\psi K^+$ decays: default (central curve) and $\pm 1\sigma_{KN}$ (envelope).

decays satisfying the trigger requirements is convoluted with this efficiency curve to obtain $\epsilon_K^{COT} = (96.4 \pm 1.6)\%$. The uncertainty is the quadrature sum of contributions from statistics ($\pm 0.2\%$), and systematic variations due to isolation ($\pm 1.5\%$), instantaneous luminosity ($\pm 0.5\%$), and the kaon nuclear interaction cross-section ($\pm 0.3\%$). The relative efficiency in bins of isolation and luminosity are given in Fig. 56 and the systematic uncertainties assigned cover the full observed difference. A relative uncertainty on the kaon nuclear interaction cross-section is conservatively assigned to be $\pm 25\%$. The resulting efficiency curves when varying this cross-section up and down by this amount are also included in Fig. 55 and are convoluted with the $p_T(K)$ spectrum. The difference between these resulting efficiencies and that obtained using

the default curve are assigned as the associated systematic uncertainty.

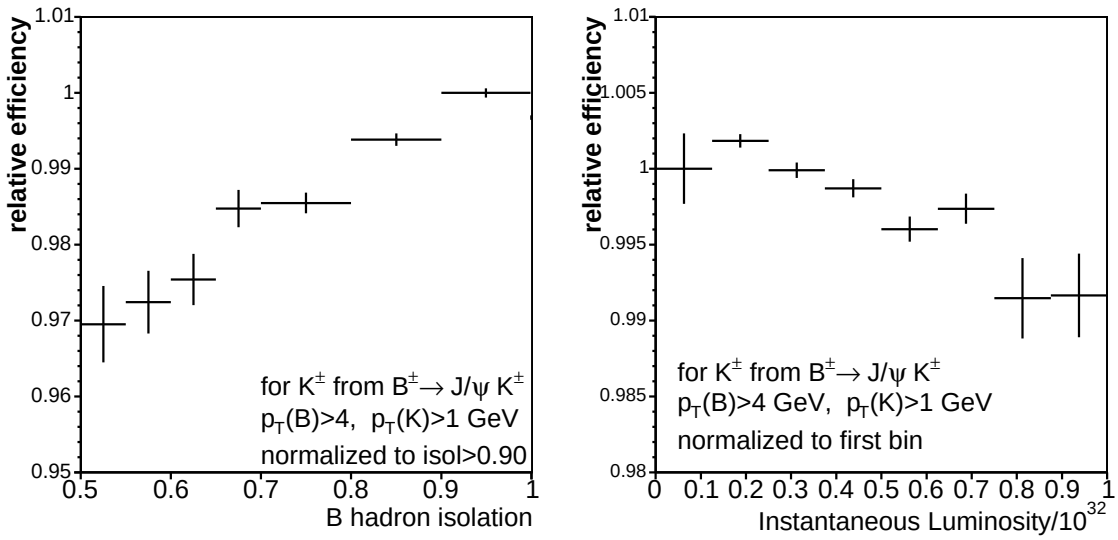

FIG. 56: The relative COT reconstruction efficiency in bins of B-hadron isolation (left) and instantaneous luminosity (right) for kaons from $B^+ \to J/\psi K^+$ decays.

The kaon specific SVX reconstruction efficiency, ϵ_K^{SVX}, is evaluated directly from the `RAREB` dataset used in the analysis. To do that the SVX quality requirements on the kaon are removed when reconstructing a sample of $B^+ \to J/\psi K^+$ candidates. The sideband subtracted efficiency for attaching > 2 $r\phi$-hits (counting one per SVX or L00 layer) to the kaon track is then obtained as a fraction of events from the sample passing these hit requirements. The resulting efficiency is $\epsilon_K^{SVX} = 0.973 \pm 0.002$, for both the CMU-CMU and CMU-CMX channels, where only the statistical uncertainty is included. Since this is evaluated on exactly the same data set with exactly the same baseline and vertex requirements as used in the analysis, the systematic uncertainties are assumed to be negligible.

The combined kaon reconstruction efficiency is $\epsilon_{B^+}^{reco-K} = 0.938 \pm 0.016$. This value is used as input for Eq. (4.10).

4. *Efficiency of the likelihood requirement*

The efficiency of the likelihood requirement, $\epsilon^{LH}_{B^0_s}$ is estimated using the odd numbered events[9] in the $B^0_s \to \mu^+\mu^-$ MC sample. The estimates are based on the sample described in Section IV.D and are given in Table XIV.

TABLE XIV: The $\epsilon^{LH}_{B^0_s}$ for the CMU-CMU (left) and CMU-CMX (right) channels. The uncertainties are the binomial statistical uncertainties.

	$\epsilon^{LH}_{B^0_s}$	
cut	CMU-CMU	CMU-CMX
$LH > 0.90$	$(70.0 \pm 0.8)\%$	$(66.3 \pm 1.0)\%$
$LH > 0.92$	$(66.5 \pm 0.8)\%$	$(64.6 \pm 1.0)\%$
$LH > 0.95$	$(61.0 \pm 0.8)\%$	$(60.1 \pm 1.0)\%$
$LH > 0.98$	$(48.1 \pm 0.9)\%$	$(48.4 \pm 1.0)\%$
$LH > 0.99$	$(37.8 \pm 0.9)\%$	$(39.1 \pm 1.0)\%$

A cross-check of the MC efficiencies is made by comparing the likelihood efficiencies between $B^+ \to J/\psi K^+$ MC and sideband subtracted $B^+ \to J/\psi K^+$ distributions from the data. The ratio of data-to-MC efficiencies for $B^+ \to J/\psi K^+$ events is given in Table XV. The distributions of the three input variables are shown in Figs. 57 and 58. Both channels are adequately modeled for this analysis purposes. Based on these comparisons systematic uncertainties of $\pm 10\%$ and $\pm 5\%$ (relative) are assigned to the CMU-CMU and CMU-CMX channels, respectively.

The efficiency for the data and MC $B^+ \to J/\psi K^+$ samples is compared for

[9] As discussed in Section IV.E.2, to avoid statistical correlations, the signal probability distributions used to determine the likelihood are made with even numbered MC events.

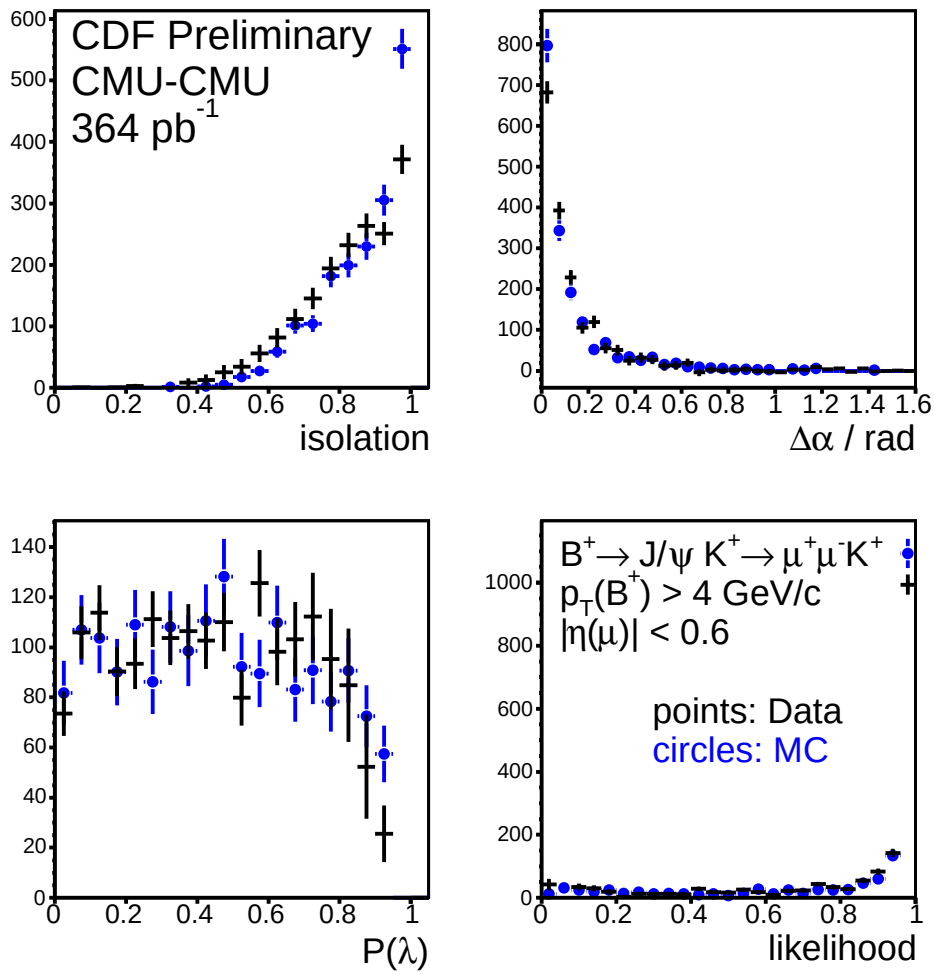

FIG. 57: A comparison of the input distributions and the resulting LH distribution from sideband subtracted $B^+ \to J/\psi K^+$ data to that from the PYTHIA MC sample for the CMU-CMU channel.

the $I > 0.50$ and $\Delta\alpha < 0.70$ rad requirements. It is found that the ratio of these efficiencies between data and MC are consistent with unity within the associated statistical uncertainty of $\pm 1.6\%$. No additional systematic uncertainty is assigned and the $B_s^0 \to \mu^+\mu^-$ MC sample is used to estimate the efficiency of these requirements to be 92%. In Eq. (4.7) this efficiency is included in the $\epsilon_{B_s^0}^{LH}$ term.

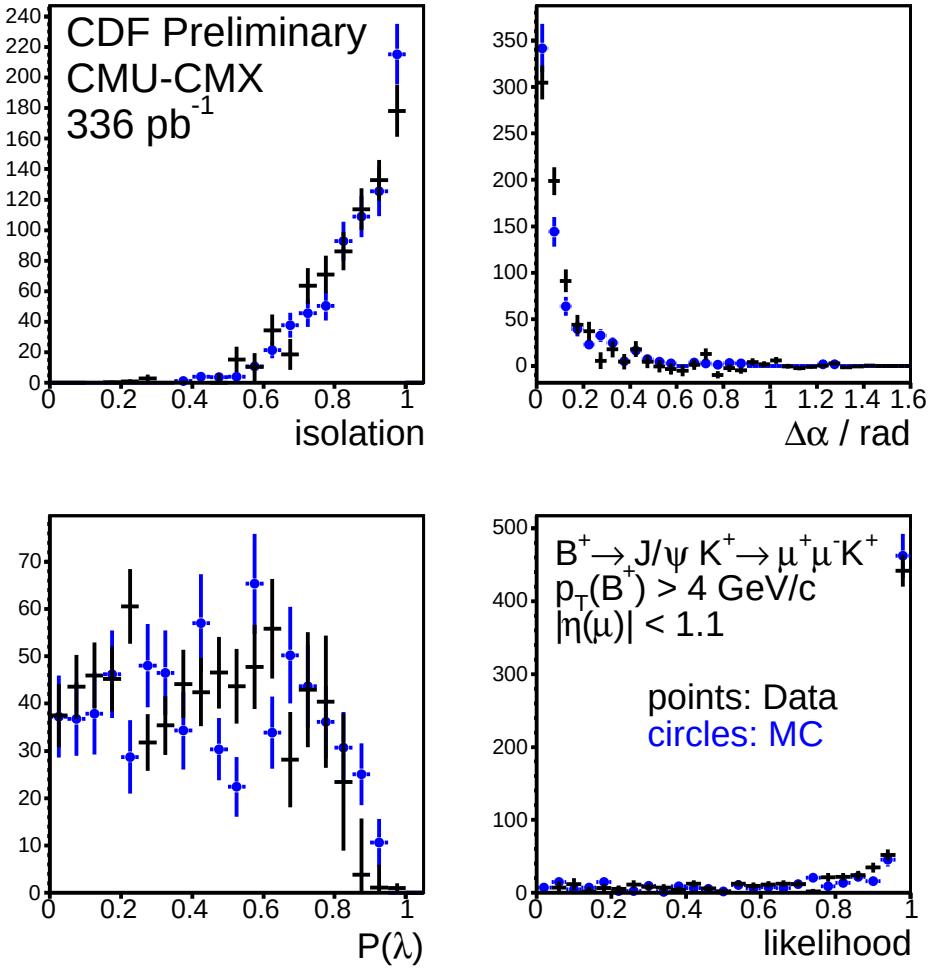

FIG. 58: A comparison of the input distributions and the resulting LH distribution from sideband subtracted $B^+ \to J/\psi K^+$ data to that from the PYTHIA MC sample for the CMU-CMX channel.

H. Optimization

1. *The optimization procedure*

For the optimization a figure-of-merit is chosen to be the *expected* 90% CL upper limit on the branching ratio, $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$, assuming no signal observation. This is a natural choice since it is statistically rigorous and optimizes the physics result itself. The effects of uncertainties are also incorporated into the optimization choice.

TABLE XV: The ratio of the $\epsilon_{B^+}^{LH}$ determined from data to that from MC, including the statistical uncertainties.

cut	$\epsilon_{B^+}^{LH}$(data)/$\epsilon_{B^+}^{LH}$(MC)	
	CMU-CMU	CMU-CMX
$LH > 0.70$	0.981 ± 0.026	1.015 ± 0.043
$LH > 0.80$	0.958 ± 0.027	1.028 ± 0.047
$LH > 0.90$	0.934 ± 0.030	0.989 ± 0.051
$LH > 0.94$	0.916 ± 0.032	0.966 ± 0.053
$LH > 0.98$	0.874 ± 0.038	0.947 ± 0.065

The methodology of setting the limit is described in Section IV.B.3. A flat prior is assumed for the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ and truncated Gaussian priors are assumed for the signal efficiencies and backgrounds [128, 129, 131]. An integrated luminosity of 1 fb^{-1} is assumed for the optimization so that the analysis can be updated with 2005 data without re-optimizing.

The optimization is performed in the range of $p_T(B)$ from 4 GeV/c to 6 GeV/c and the likelihood requirement from 0.90 to 0.99. The CMU-CMU and CMU-CMX channels are separately optimized. For the different $p_T(B)$ requirements all the relevant variables are re-evaluated: the $B^+ \to J/\psi K^+$ yields, the acceptances, any efficiency for which there exists a p_T^μ dependence, and the likelihood curves for signal and background by remaking the probability densities, $\mathbf{P}_{s(b)}(x_i)$, used in the likelihood calculation. The background is separately estimated for each likelihood cut at each $p_T(B)$ threshold using the method of Section IV.F. Note that using a $p_T(B)$ cut lower than 4 GeV/c is feasible, but that would not add substantially to the limit improvement. The reasons are that the acceptance goes down with lower p_T, as shown

in Fig. 47. Moreover, the B_s^0 p_T spectrum starts to drop below 3 GeV/c, as shown in Figs. 46 and 17. In addition, the background tends to grow faster at the lower momenta, as shown in Fig. 59.

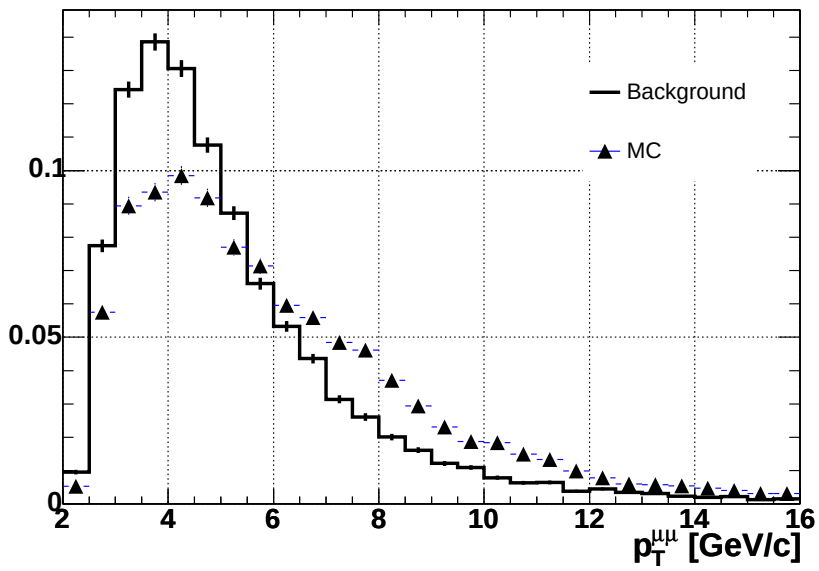

FIG. 59: Distribution of $p_T^{\mu\mu}$ for data sidebands and $B_s^0 \to \mu^+\mu^-$MC events passing the baseline and vertex requirements. Both distributions are normalized to the unit area.

Because the likelihood is so effective at suppressing the background, the optimization moves the $p_T(B)$ threshold to 4 GeV/c and chooses $LH > 0.99$ for both the CMU-CMU and CMU-CMX channels. Moving to higher $p_T(B)$ requirements makes the expected branching ratio limit $\sim 5\%$ worse in both channels (for the optimal LH requirement at that $p_T(B)$). Fixing the $p_T(B)$ requirement and loosening the LH cut makes the expected branching ratio limit $\sim 10\%$ worse for both the CMU-CMU and CMU-CMX channels.

2. *Expectations for* $B_s^0 \to \mu^+\mu^-$

For the 364 pb^{-1} of CMU-CMU data these optimal requirements correspond to an expected background of $N_{\rm bg}^{UU} = 0.81 \pm 0.12$ and a single-event-sensitivity of $ses^{UU} = (1.0 \pm 0.2) \times 10^{-7}$ and yield an expected limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 3.5 \times 10^{-7}$ at 90% CL.

For the 336 pb^{-1} of CMU-CMX data these optimal requirements correspond to an expected background of $N_{\rm bg}^{UX} = 0.66 \pm 0.13$ and a single-event-sensitivity of $ses^{UX} = (1.6 \pm 0.3) \times 10^{-7}$ and yield an expected limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 5.6 \times 10^{-7}$ at 90% CL.

The CMU-CMU and CMU-CMX channels can be combined, taking into account their correlated uncertainties [130] from the f_u/f_s and $B^+ \to J/\psi K^+$ related branching ratios, to get and expected limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2.0 \times 10^{-7}$ at 90% CL. This is significantly better than the best published result [20]. The expected combined limit for 1 fb^{-1} of data is $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.1 \times 10^{-7}$ using the same selection criteria and assuming the single-event-sensitivity and background scale only with the luminosity.

3. *Expectations for* $B_d^0 \to \mu^+\mu^-$

The expected limit on the $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ is also calculated. The acceptance, trigger, reconstruction and "final" efficiencies are estimated in the same manner as for the $B_s^0 \to \mu^+\mu^-$ decays using a MC sample of $B_d^0 \to \mu^+\mu^-$ generated in the same manner as the $B_s^0 \to \mu^+\mu^-$ sample described in Section IV.D. It is found that the ratio of $\alpha \cdot \epsilon_{total}(B_s)/\alpha \cdot \epsilon_{total}(B_d)$ is consistent with unity within the statistical uncertainties. Using the requirements optimized for the $B_s^0 \to \mu^+\mu^-$ search but for a different mass window ($[5.219, 5.339]$ GeV/c^2) yields an expected $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ limit of 4.9×10^{-8}

at the 90% CL for the CMU-CMU and CMU-CMX channels combined using 364 pb^{-1} of data and including the correlations from the $B^+ \to J/\psi K^+$ related branching ratios. This is significantly better than the present best limit from the B-factories, $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 8.3 \times 10^{-8}$ from BaBar using 111 fb^{-1} [140].

I. Results

The invariant mass distributions for events satisfying the baseline and vertex criteria given in Tables V, VI and VII and with $LH > 0.99$ are shown in Fig. 60. There are no events, in either channel, which fall within the ± 60 MeV/c^2 mass windows centered on the world average $B_{s(d)}$ mass, 5.369(5.279) GeV/c^2. The closest event has an invariant mass of 5.190 and 5.197 GeV/c^2 in the CMU-CMU and CMU-CMX channel, respectively. Taking into account the correlated systematic uncertainties from the $\mathcal{B}(B^+ \to J/\psi K^+) \cdot \mathcal{B}(J/\psi \to \mu^+\mu^-)$ and f_u/f_s a combined limit is evaluated to be $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.6 \times 10^{-7}$ (2.1×10^{-7}) at 90% (95%) CL. The resulting limit for the $B_d^0 \to \mu^+\mu^-$ decay is $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 3.9 \times 10^{-8}$ (5.1×10^{-8}) at 90% (95%) CL.

Figure 61 shows a two dimensional plot of $M_{\mu^+\mu^-}$ and LH for all events satisfying the baseline and vertex criteria and with $LH > 0.80$. The full likelihood distribution for events satisfying the baseline and vertex criteria and falling within the B_s search window, $5.309 < M_{\mu^+\mu^-} < 5.429$ GeV/c^2, is shown in Fig. 62. A comparison of the likelihood distribution for these signal events to the same distributions using events in the sideband region yields a KS probability of 66% and 76% for the CMU-CMU and CMU-CMX channels respectively.

As a cross-check, the number of observed events is compared to that predicted for looser LH requirements. In the CMU-CMU channel, 146 ± 22 events are expected

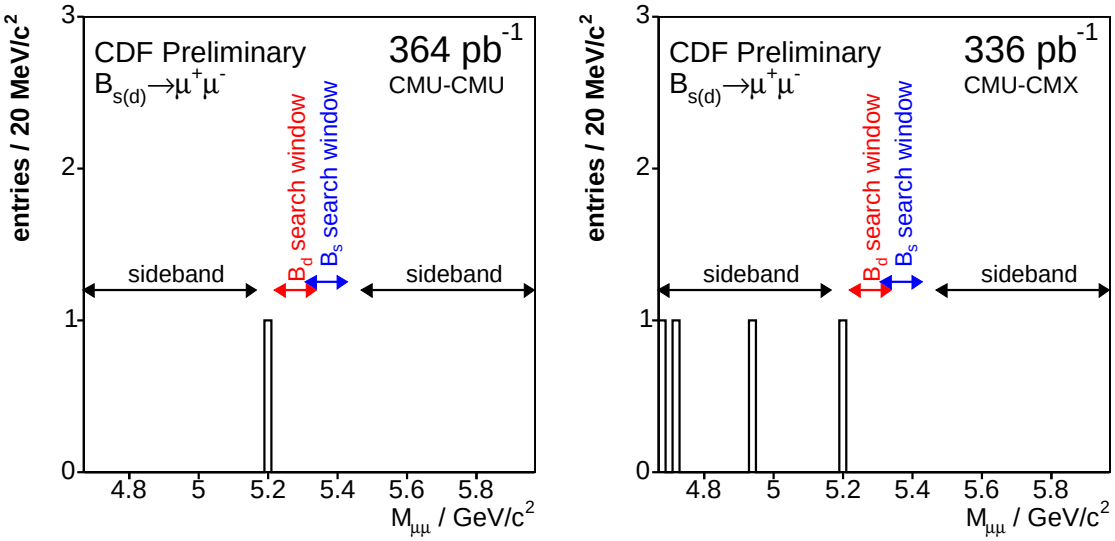

FIG. 60: The invariant mass distribution for events satisfying all selection criteria in the CMU-CMU (left) and CMU-CMX (right) channels.

and 136 events are observed for $LH > 0.50$; while 24 ± 4 events are expected and 20 are observed for $LH > 0.90$. In the CMU-CMX channel, 99 ± 20 are expected and 99 are observed for $LH > 0.50$; while 17 ± 3 events are expected and 9 are observed for $LH > 0.90$. The agreement for all of these is reasonable.[10]

[10]The Poisson probability of observing ≤ 9 while expecting 17 is approximately 2.6%, which corresponds to a fluctuation of < 2 standard deviations for a Gaussian distribution.

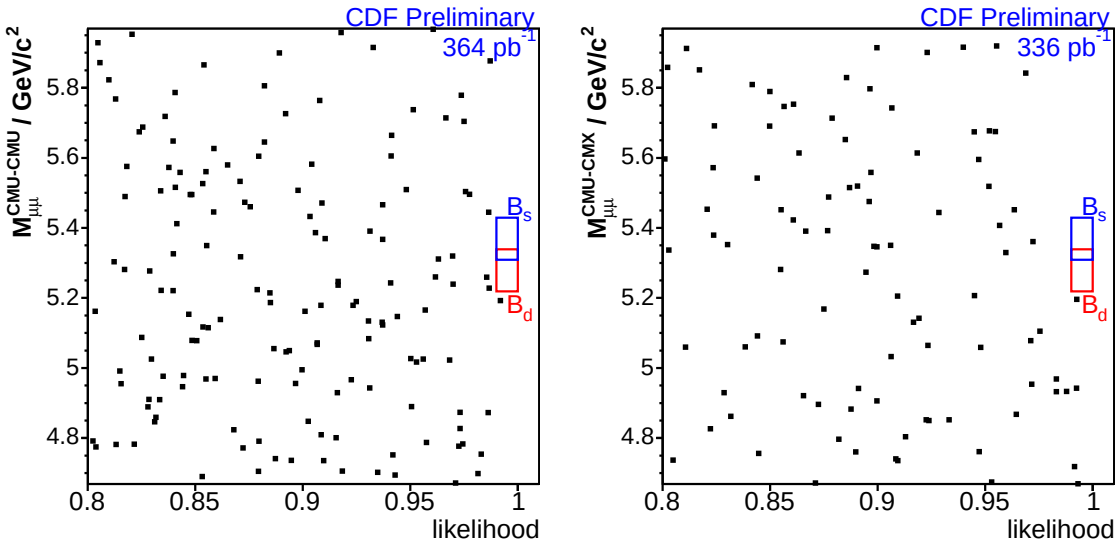

FIG. 61: The invariant mass distribution versus the event likelihood for events satisfying baseline in the CMU-CMU (left) and CMU-CMX (right) channels.

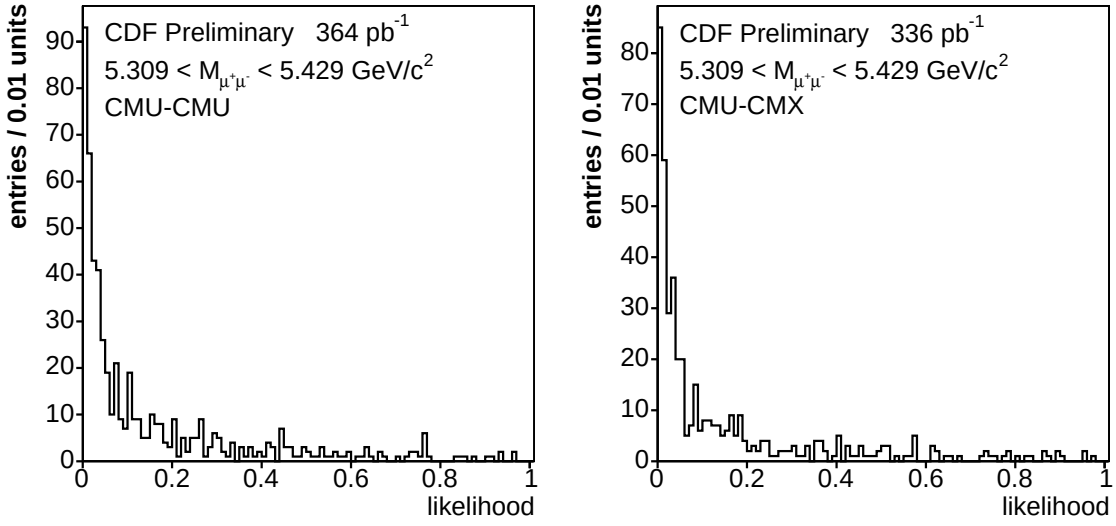

FIG. 62: The likelihood distribution for events in the B_s search window for the CMU-CMU (left) and CMU-CMX (right) channel.

J. Prospects for $B^0_{s(d)} \to \mu^+\mu^-$ at the Tevatron and beyond

Since the CDF experiment continues collecting data, it is reasonable to project the expected sensitivity for higher integrated luminosity. During the span of Run II about 5 fb^{-1} may be collected. No significant changes to the muon identification or tracking are anticipated during that time. It is also quite probable that the dimuon trigger structure relevant to the $B^0_s \to \mu^+\mu^-$ analysis will allow to continue taking data without a substantial loss in acceptance relative to the present analysis. With these facts in hand, the expected sensitivity to $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ is considered in the following scenarios when a larger dataset is considered: no improvements to the analysis; a case of an improved analysis with no increase in the expected background supported by the specific suggestions of additional improvements. Additionally, a combined measurement from CDF and DØ experiments is considered for the projected sensitivity at the Tevatron. Finally, the potential sensitivity to $B^0_s \to \mu^+\mu^-$ decays at the SM level is discussed with respect to the LHC experiments, as well as the sensitivity to $B^0_d \to \mu^+\mu^-$ decays at the SM level.

The simplest approach is to assume that no improvements are made to the analysis except for adding more data. To calculate the expected upper limit on the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ one needs only to rescale the number of events in the normalization sample as given in Table XII and the number of expected background events as detailed in Section IV.H.2. The scale factor is $\mathcal{L}/(364\ \text{pb}^{-1})$ for CMU-CMU and $\mathcal{L}/(336\ \text{pb}^{-1})$, where the same projected integrated luminosity is assumed for simplicity. These rescaled numbers are then used to calculate the expected upper limit using the method detailed in Section IV.H. The expected $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ upper limit for a set of confidence levels (90%, 3σ, and 5σ) as a function of integrated luminosity is shown in Fig. 63 [141]. Starting from about 1 fb^{-1} the upper limit scales with

luminosity as $1/\sqrt{\mathcal{L}}$, which suggests that the analysis selection requirements would have to be revisited for a larger data sample.

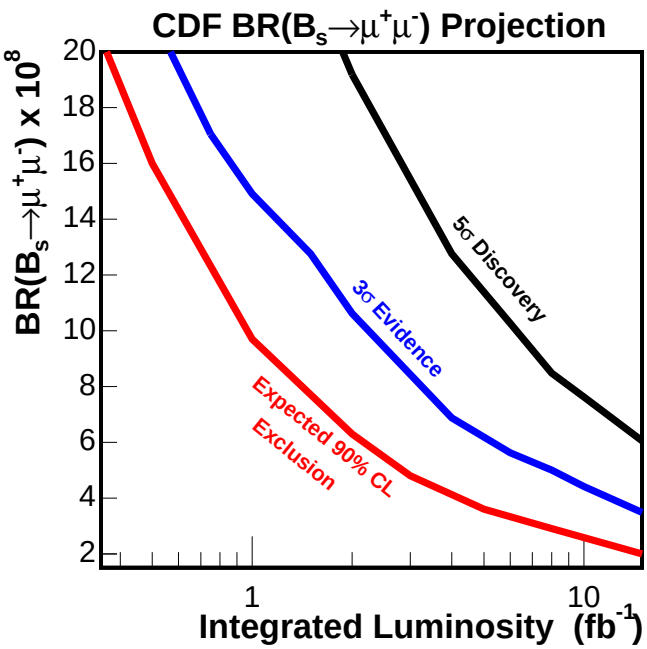

FIG. 63: Expected $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ upper limit based on the presented analysis.

A more creative approach is to suggest that the analysis can be improved when more data becomes available. As long as it is possible to reject background substantially more effectively than the corresponding loss of the signal efficiency probably the most optimal condition is to keep backgrounds at the level of about one expected event. Once the number of expected background events drops below 1 the upper limit becomes quite insensitive to the background. Thus, it is somewhat natural that the selected optimization procedure stops at the point where $N_{\rm bg} \sim 1$. A relatively optimistic scenario is that the background can be kept around 1 with increasing luminosity by the virtue of improved and tightened cuts which do not result in the loss in efficiency. In this case the upper limit is expected to scale roughly as $1/\mathcal{L}$.

Thus, at 8 fb^{-1} of integrated luminosity[11] one would expect the 90% CL upper limit of about 1×10^{-8} which would roughly correspond to one event at the SM level of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$.

The following improvements to the signal selection and background rejection are considered viable: use of calorimeter isolation energy, requirement of CMUP muons, rejection by 2D pointing angle, use of invariant mass in the likelihood. The calorimeter isolation is a variable similar to the track isolation (Iso) used in the analysis. Based on the data MC and the background samples used in the analysis it can be found that a cut[12] $\max(E_{T,R4}^{\mu^+}, E_{T,R4}^{\mu^-}) < 2$ is about 90% efficient for $B_{s(d)}^0 \to \mu^+\mu^-$ MC and rejects about half of the background. A substantial rejection of backgrounds coming from hadrons (K^+, π, or p) misidentified as muons can be achieved by a requirement that one (or two, if both are expected to cross CMP) muon is a CMUP muon. Since the CMP chambers are placed behind an additional steel absorber, the chance for a hadron to reach CMP is much lower than that for CMU. Based on the samples used in the analysis about a factor of two of background rejection is expected with a corresponding signal efficiency of roughly 90%. The 3D pointing angle is roughly a quadrature sum of the pointing polar and azimuthal angles, where both angles are roughly uncorrelated and rely heavily on the position measurement resolution in r-z and r-ϕ views. The track/vertex measurement in z direction has larger uncertainty than that in r-ϕ view. Thus, it is beneficial to consider the azimuthal ($\Delta\Phi$) and polar pointing separately, which can be achieved by using both 3D and 2D pointing angles. It can be shown from the background and MC samples that a $\Delta\Phi < 0.12$ cut is better than 90% efficient and provides a background rejection of about a factor of

[11] This is probably the most that can be integrated by CDF during Run II.

[12] $E_{T,R4}^{\mu}$ is the calorimeter energy measured in the towers within the cone $R < 0.4$ around the tower hit by the muon, where $R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

two. Assuming that the measurement of the dimuon mass is reliable, it is possible to include the measured dimuon mass into the likelihood definition. The background events would correspond to a flat distribution function, while the signal events would correspond to a Gaussian distribution function. The result of this consideration can be viewed as an effective tighter mass window. Compared to the present $\pm 2.5\sigma$ window, the mass likelihood method can provide another factor of two in background rejection. The cuts considered above can be better optimized in the upcoming analysis. But even this non-optimized set provides about a factor of 16 additional background rejection, which roughly justifies the statement that it should be possible to keep the number of background events on the order of one event throughout the length of Run II.

Another projection can be done based on the combined Tevatron results for the CDF and DØ. Since the DØ experiment also has a sensitivity to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ which is comparable to that of CDF, but slightly worse if taken for the same integrated luminosity [20], it is beneficial to combine the two independent measurements to obtain a better limit.[13] Considering the fact that the future DØ analysis can improve the background rejection it is anticipated that the parity between the CDF and DØ results will roughly hold throughout Run II. This improves the chances that a predicted $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$upper limit reach of 1×10^{-8} is achievable by the end of Run II.

In the future, the dominant sensitivity is expected to come from the LHC with a potential sensitivity at the level of the SM predictions. The collider experiments at the Tevatron, CDF and DØ, are expected to provide the best sensitivity to $B_s^0 \to \mu^+\mu^-$ for the upcoming two to three years, until about the first two years of running the

[13]In particular, after the correlations in the uncertainties in both the CDF and DØ measurements are accounted for, the combination of the presented CDF result and the DØ result yields a limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.6 \times 10^{-7}$ at 95% CL [142].

LHC experiments. After the LHC turns on (anticipated in 2007 [143]) it will be the place with the highest $b\bar{b}$ production rate in the world: around 10^6 $b\bar{b}$ pairs per second (using $\sigma_{b\bar{b}} \sim 500$ μb [58] at $\mathcal{L} \sim 2\times10^{33}$ cm^{-2}s^{-1} [143]) compared to up to 10^4 $b\bar{b}$ pairs per second at the Tevatron (assuming $\sigma_{b\bar{b}} \sim 40$ μb [85] at $\mathcal{L} \sim 2 \times 10^{32}$ cm^{-2}s^{-1}). With these production rates the experiments at LHC should have the sensitivity to the SM level $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. Despite the tighter trigger conditions both CMS [57, 58] and ATLAS [59, 60] would be able to trigger on a few tens of the SM predicted $B_s^0 \to \mu^+\mu^-$ events within a year of running (corresponding to 10 fb^{-1}). The number of actual $B_s^0 \to \mu^+\mu^-$ events selected by the analysis cuts will strongly depend on the optimization procedure and on the amount of expected background events. Due to a better mass resolution and lower detectable muon momentum thresholds the CMS experiment is is expected to have a better sensitivity compared to the ATLAS experiment. Based on the experience of the present analysis the background estimates for the $B_s^0 \to \mu^+\mu^-$ analysis in CMS [58, 59] look somewhat optimistic. Assuming that optimistic scenario with signal to background ratio of about 7 : 1 the CMS experiment should be able to measure the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ at the level of the SM prediction with a relative uncertainty about 30%.

Although it is straightforward to make projections for the $B_d^0 \to \mu^+\mu^-$ decay mode, the measurement of $\mathcal{B}(B_d^0 \to \mu^+\mu^-)$ at the level of the SM prediction does not seem easily possible at the current or anticipated (in the near future) experimental facilities. In particular, with the current state of technology, a B-factory would have a relatively clean environment but would not be able to produce enough B_d^0 mesons [144]. In contrast, the LHC would produce a large number of $B_d^0 \to \mu^+\mu^-$ events at the SM level, but the amount of backgrounds would make it really hard if possible at all to discriminate the $B_d^0 \to \mu^+\mu^-$ signal from backgrounds. Assuming substantially the same backgrounds as for $B_s^0 \to \mu^+\mu^-$, the signal to background

ratio for $B_d^0 \to \mu^+\mu^-$ at hadron collider is a factor of $|V_{td}/V_{ts}|^2 f_d/f_s \sim 0.1$ worse than that for $B_s^0 \to \mu^+\mu^-$. In an optimistic case this gives at best 1 : 1 between signal and background – not completely hopeless, but challenging.

K. Exclusion of SUSY models parameters

Based on the existing limit on $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ it is possible to exclude any model that predicts a branching ratio higher than the measured upper limit. Note that it is a common practice to report the upper limits on heavy flavor rare decays with 90% CL, e.g, Refs. [1, 51], but the standard approach to model exclusion is to apply the 95% CL (also refered to as 2σ exclusion). Thus, for the models discussed below the parameter space corresponding to a prediction of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) > 2.1 \times 10^{-7}$ is shown as excluded. Most of the exclusion plots below are derived from those described in Section II.E.2 and are considered below in the order of decreasing sensitivity to this limit.

The $\mathrm{MSO_{10}SM}$ solution is the most sensitive to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurement, based on the analysis presented in Refs. [16, 73] and further discussion in Section II.E.2. Figures 64 and 65 [73] show that the region with $m_A < 500$ GeV/c^2 is essentially excluded. Shown on these plots are the limits reported by the present analysis [23], the published DØ limit of 5×10^{-7} [20], and the preliminary DØ limit of 4.1×10^{-7} [142].

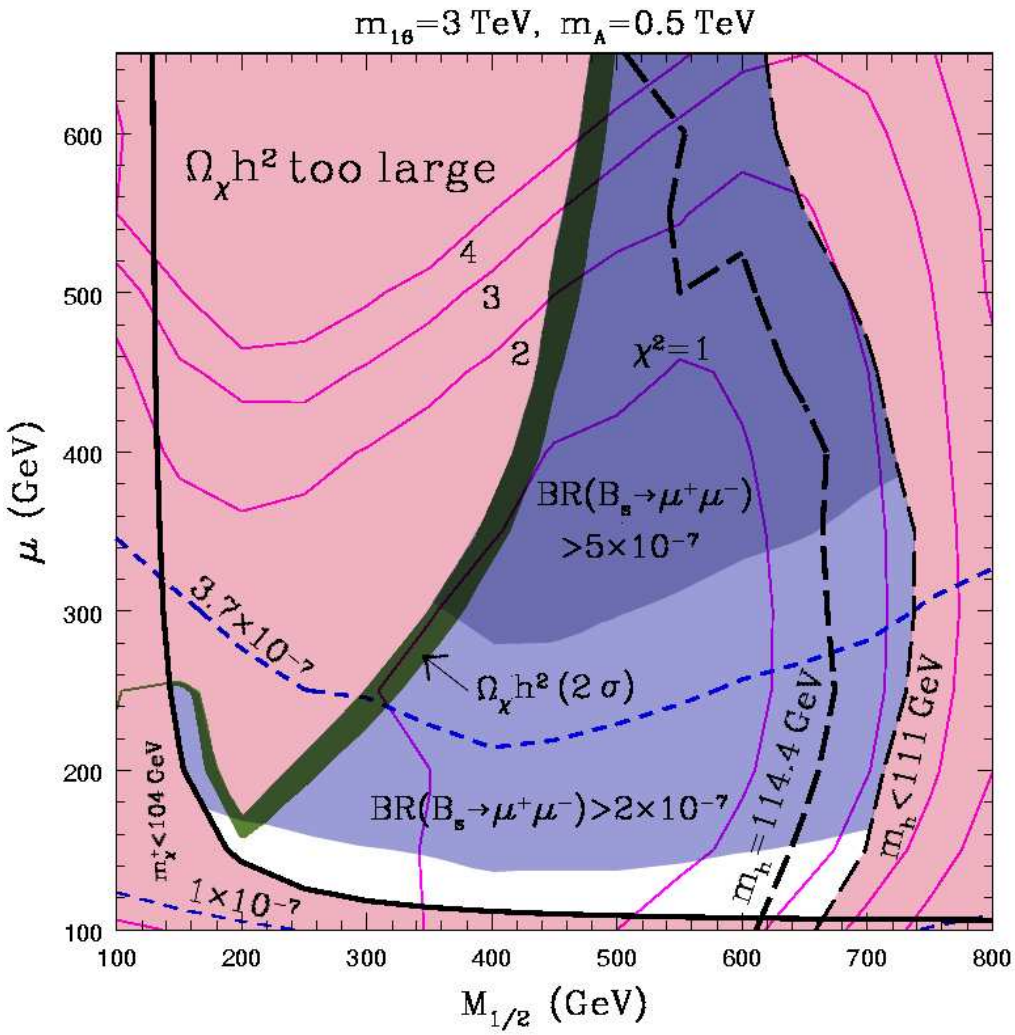

FIG. 64: Allowed region in $MSO_{10}SM$ in $(\mu, m_{1/2})$ plane with current experimental limits included.

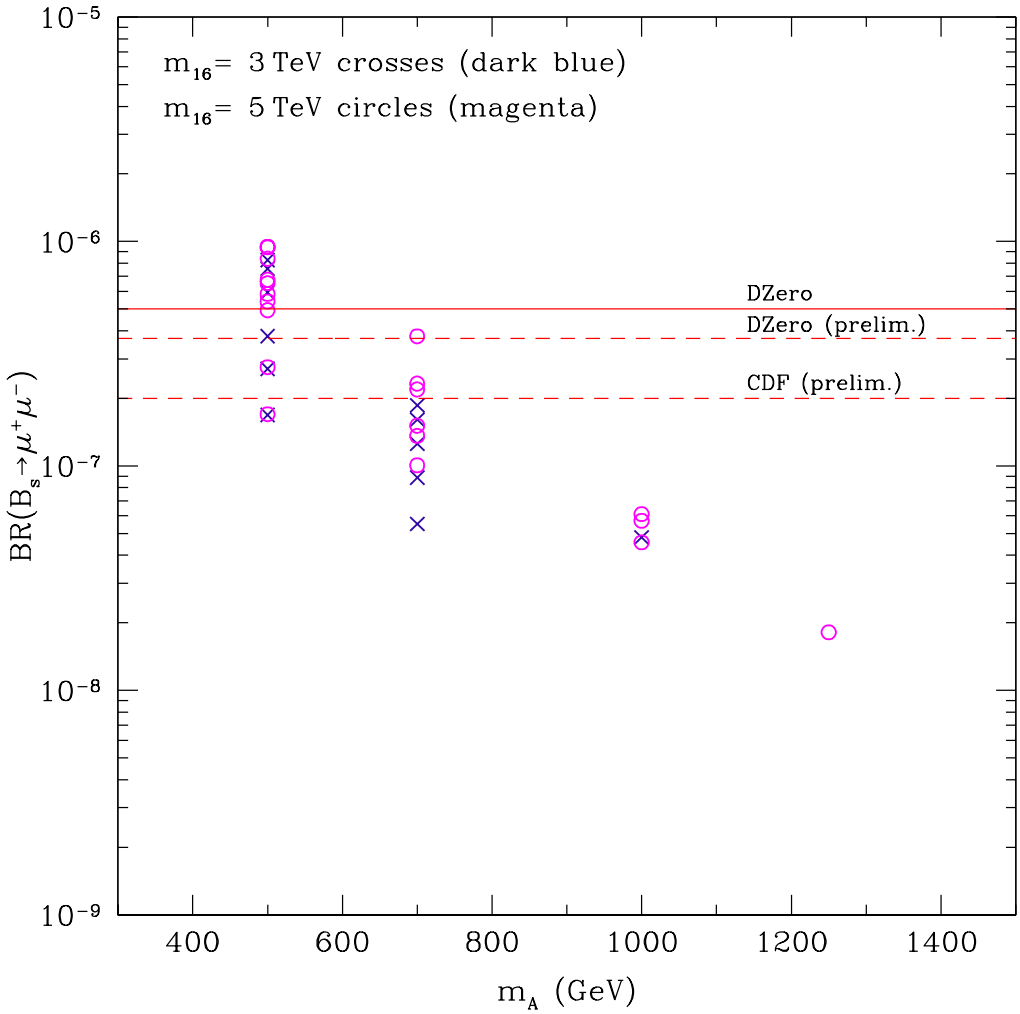

FIG. 65: Allowed range of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ for the given m_A in MSO$_{10}$SM: $m_A < 500$ GeV/c^2 is excluded.

In the RPV scenario of mSUGRA the excluded parameter space is primarily defined by the $m_0, m_{1/2}$ parameters defining the sneutrino masses and the R-parity violating couplings defined by λ. The remaining parameters of the model are practically orthogonal to the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. The parameter space of RPV mSUGRA excluded by this analysis is shown in Fig. 66 [74]: the region below the constant λ contour is excluded for the given λ.

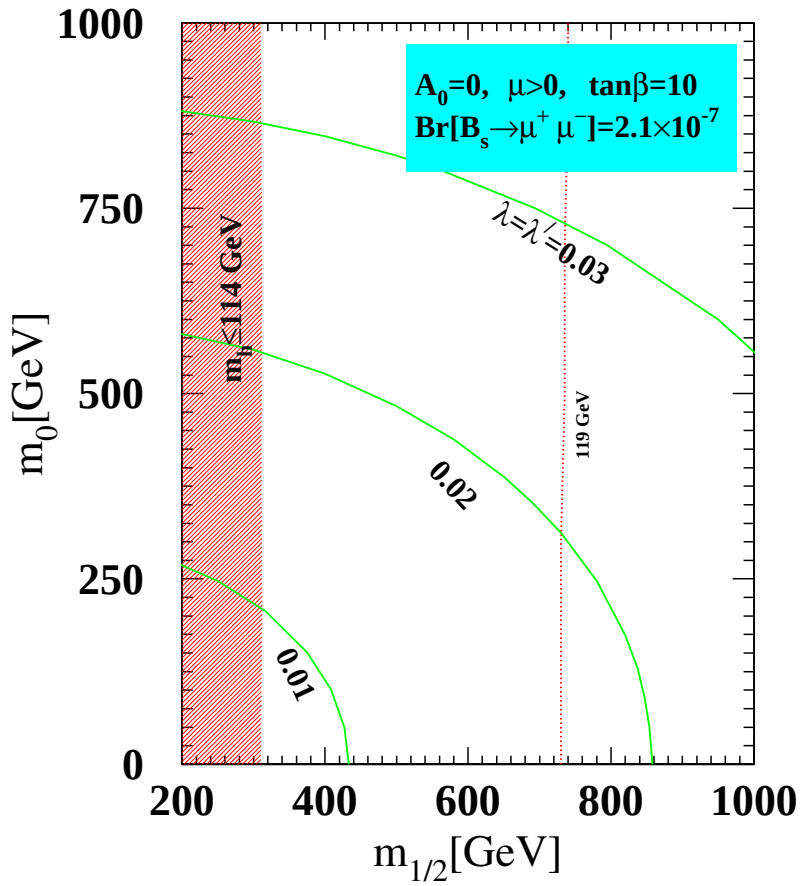

FIG. 66: Parameter space in $(m_0, m_{1/2})$ plane of RPV mSUGRA excluded by $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2.1 \times 10^{-7}$.

The amount of parameter space excluded in mSUGRA model is not significant yet, event at the large $\tan\beta \sim 50$ values. The parameter space that allows the

$\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ to be above 2×10^{-7} is already excluded by the $b \to s\gamma$ constraint, as shown in Fig. 11 [14]. The mSUGRA parameters excluded by this analysis plotted before the $b \to s\gamma$ constraint is applied are shown in Fig. 67 as gray (orange) filled region.

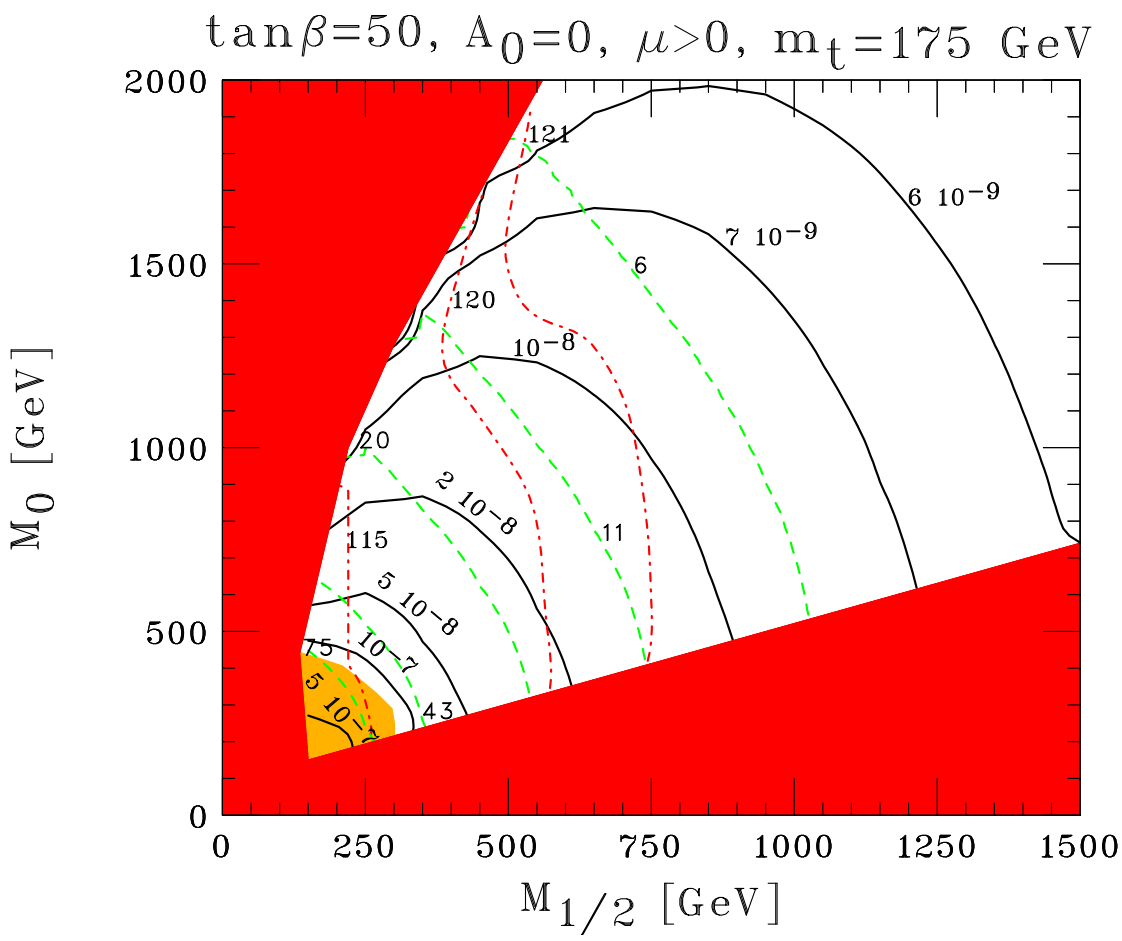

FIG. 67: Parameter space in $(m_0, m_{1/2})$ plane of mSUGRA excluded by $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 2.1 \times 10^{-7}$ (orange filled).

CHAPTER V

CONCLUSION

A search for $B_s^0 \to \mu^+\mu^-$ is performed using approximately 360 pb^{-1} of Run II data collected by the CDF II detector using $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV. This analysis greatly improves the sensitivity relative to a previous CDF analysis [84] by using a likelihood method and by extending the muon coverage. No candidate events are observed in either the CMU-CMU or CMU-CMX channels with 0.81 ± 0.12 and 0.66 ± 0.13 events expected, respectively. This corresponds to the combined limit of $\mathcal{B}(B_s^0 \to \mu^+\mu^-) < 1.6 \times 10^{-7}(2.1 \times 10^{-7})$ at the 90%(95%) confidence level.

The same analysis is also sensitive to the $B_d^0 \to \mu^+\mu^-$ decays. The combined limit is $\mathcal{B}(B_d^0 \to \mu^+\mu^-) < 4.2 \times 10^{-8}$.

These measurements provide the world's best limit, significantly extending the sensitivity of other experiments [20, 140]. Further improvements to the measurement are expected once more data is collected. With an integrated luminosity of 8 fb^{-1} collected by the CDF and DØ experiments, expected by the end of Run II, the combined sensitivity to $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is predicted to be at the level of 1×10^{-8}.

Using these results, supersymmetric models predicting a branching ratio $\mathcal{B}(B_s^0 \to \mu^+\mu^-) > 2.1 \times 10^{-7}$ are excluded at 95% CL. This includes the substantial reduction of the allowed paramer space of the MSO$_{10}$SM model solution and a significant reduction of the allowed parameter space for a R-parity violating mSUGRA model. Should the $B_s^0 \to \mu^+\mu^-$ decay be observed during Tevatron Run II, this would unequivocally point to physics beyond the Standard Model, which could be supersymmetry.

REFERENCES

[1] Particle Data Group, S. Eidelman *et al.*, Phys. Lett. B **592**, 1 (2004).

[2] S. Weinberg, *The Quantum Theory of Fields* (Cambridge University Press, Cambridge, UK, 1995-2000), Vols. 1-3.

[3] L. Susskind, Phys. Rept. **104**, 181 (1984).

[4] S. P. Martin, arXiv:hep-ph/9709356.

[5] R. N. Mohapatra, arXiv:hep-ph/9911272.

[6] M. Maltoni, T. Schwetz, M. A. Tortola and J. W. F. Valle, New J. Phys. **6**, 122 (2004); arXiv:hep-ph/0405172 and references therein.

[7] G. Bertone, D. Hooper and J. Silk, Phys. Rept. **405**, 279 (2005); arXiv:hep-ph/0404175 and references therein.

[8] WMAP Collaboration, D. N. Spergel *et al.*, Astrophys. J. Suppl. Ser. **148**, 175 (2003); arXiv:astro-ph/0302209.

[9] SDSS Collaboration, M. Tegmark *et al.*, Phys. Rev. D **69**, 103501 (2004); arXiv:astro-ph/0310723.

[10] R. Arnowitt and B. Dutta, arXiv:hep-ph/0204187.

[11] C. Balazs, M. Carena, A. Menon, D. E. Morrissey and C. E. M. Wagner, Phys. Rev. D **71**, 075002 (2005); arXiv:hep-ph/0412264.

[12] S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. D **2**, 1285 (1970).

[13] A. J. Buras, arXiv:hep-ph/9806471.

[14] R. Arnowitt, B. Dutta, T. Kamon and M. Tanaka, Phys. Lett. B **538**, 121 (2002); arXiv:hep-ph/0203069.

[15] A. Dedes and B. T. Huffman, Phys. Lett. B **600**, 261 (2004); arXiv:hep-ph/0407285.

[16] R. Dermisek, S. Raby, L. Roszkowski and R. Ruiz De Austri, J. High Energy Phys. **0304**, 037 (2003); arXiv:hep-ph/0304101.

[17] M. Misiak and J. Urban, Phys. Lett. B **451**, 161 (1999); arXiv:hep-ph/9901278.

[18] Tevatron Accelerator Division, *Operations Rookie Books*, http://www-bdnew.fnal.gov/operations/rookie_books/rbooks.html.

[19] CDF Collaboration, R. Blair *et al.*, Fermilab Report No. FERMILAB-PUB-96-390-E.

[20] DØ Collaboration, V. M. Abazov *et al.*, Phys. Rev. Lett. **94**, 071802 (2005); arXiv:hep-ex/0410039.

[21] CDF Collaboration, F. Abe *et al.*, Phys. Rev. D **57**, 3811 (1998).

[22] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. Lett. **93**, 032001 (2004); arXiv:hep-ex/0403032.

[23] CDF Collaboration, A. Abulencia *et al.*, arXiv:hep-ex/0508036.

[24] LEP Collaborations, arXiv:hep-ex/0412015; LEP Electroweak Working Group web page: http://www.cern.ch/LEPEWWG.

[25] UTfit Collaboration, M. Bona *et al.*, J. High Energy Phys. **0507**, 028 (2005); arXiv:hep-ph/0501199, http://utfit.roma1.infn.it/.

[26] CKMfitter Group, J. Charles *et al.*, Eur. Phys. J. C **41**, 1 (2005); arXiv:hep-ph/0406184, http://ckmfitter.in2p3.fr/.

[27] HPQCD Collaboration, C. T. H. Davies *et al.*, Phys. Rev. Lett. **92**, 022001 (2004); arXiv:hep-lat/0304004.

[28] HPQCD Collaboration, A. Gray *et al.*, arXiv:hep-lat/0507015.

[29] MILC Collaboration, C. Aubin *et al.*, Phys. Rev. D **70**, 114501 (2004); arXiv:hep-lat/0407028.

[30] Muon $g-2$ Collaboration, G. W. Bennett *et al.*, Phys. Rev. Lett. **92**, 161802 (2004); arXiv:hep-ex/0401008.

[31] K. Hagiwara, A. D. Martin, D. Nomura and T. Teubner, Phys. Lett. B **557**, 69 (2003); arXiv:hep-ph/0209187.

[32] T. G. Rizzo, arXiv:hep-ph/0409309.

[33] E. Farhi and L. Susskind, Phys. Rept. **74**, 277 (1981).

[34] J. H. Schwarz, arXiv:hep-ex/0008017.

[35] J. H. Schwarz, arXiv:astro-ph/0304507.

[36] D. J. H. Chung, L. L. Everett, G. L. Kane, S. F. King, J. Lykken and L. T. Wang, Phys. Rept. **407**, 1 (2005); arXiv:hep-ph/0312378.

[37] J. Rosiek, arXiv:hep-ph/9511250.

[38] H. Baer, C. Balazs, A. Belyaev, J. K. Mizukoshi, X. Tata and Y. Wang, J. High Energy Phys. **0207**, 050 (2002); arXiv:hep-ph/0205325.

[39] B. C. Allanach and C. G. Lester, arXiv:hep-ph/0507283.

[40] H. Baer, C. Balazs, A. Belyaev, R. Dermisek, A. Mafi and A. Mustafayev, J. High Energy Phys. **0205**, 061 (2002); arXiv:hep-ph/0204108.

[41] P. Nath and R. Arnowitt, Phys. Rev. D **56**, 2820 (1997); arXiv:hep-ph/9701301.

[42] H. Baer, A. Mustafayev, S. Profumo, A. Belyaev and X. Tata, J. High Energy Phys. **0507**, 065 (2005); arXiv:hep-ph/0504001 and references therein.

[43] T. Fukuyama, A. Ilakovac, T. Kikuchi, S. Meljanac and N. Okada, J. Math. Phys. **46**, 033505 (2005); arXiv:hep-ph/0405300.

[44] H. S. Goh, R. N. Mohapatra and S. P. Ng, Phys. Rev. D **68**, 115008 (2003); arXiv:hep-ph/0308197.

[45] T. Blazek, R. Dermisek and S. Raby, Phys. Rev. D **65**, 115004 (2002); arXiv:hep-ph/0201081.

[46] S. Raby, Pramana **62**, 523 (2004); arXiv:hep-ph/0304074.

[47] R. Rattazzi and U. Sarid, Phys. Rev. D **53**, 1553 (1996); arXiv:hep-ph/9505428.

[48] D. Auto, H. Baer, C. Balazs, A. Belyaev, J. Ferrandis and X. Tata, J. High Energy Phys. **0306**, 023 (2003); arXiv:hep-ph/0302155.

[49] CLEO Collaboration, R. Ammar *et al.*, Phys. Rev. Lett. **71**, 674 (1993).

[50] M. Neubert, Eur. Phys. J. C **40**, 165 (2005); arXiv:hep-ph/0408179.

[51] H. F. A. Group (HFAG), arXiv:hep-ex/0505100; http://www.slac.stanford.edu/xorg/hfag/.

[52] P. H. Chankowski and L. Slawianowska, Phys. Rev. D **63**, 054012 (2001); arXiv:hep-ph/0008046.

[53] C. Bobeth, T. Ewerth, F. Kruger and J. Urban, Phys. Rev. D **64**, 074014 (2001); arXiv:hep-ph/0104284.

[54] A. K. Alok and S. U. Sankar, arXiv:hep-ph/0502120.

[55] T. Inami and C. S. Lim, Prog. Theor. Phys. **65**, 297 (1981); Erratum-ibid. **65**, 1772 (1981).

[56] A. J. Buras, Phys. Lett. B **566**, 115 (2003); arXiv:hep-ph/0303060.

[57] A. Nikitenko, A. Starodumov and N. Stepanov, arXiv:hep-ph/9907256.

[58] F. P. Schilling, arXiv:hep-ex/0509009.

[59] N. Nikitin, AIP Conf. Proc. **619**, 665 (2002).

[60] M. Smizanska, Eur. Phys. J. C **34**, S385 (2004).

[61] Z. H. Xiong and J. M. Yang, Phys. Lett. B **546**, 221 (2002); arXiv:hep-ph/0208147.

[62] A. J. Buras, M. Spranger and A. Weiler, Nucl. Phys. B **660**, 225 (2003); arXiv:hep-ph/0212143.

[63] S. Baek, P. Ko and W. Y. Song, J. High Energy Phys. **0303**, 054 (2003); arXiv:hep-ph/0208112.

[64] B. C. Allanach, A. Dedes and H. K. Dreiner, Phys. Rev. D **69**, 115002 (2004); arXiv:hep-ph/0309196.

[65] UA1 Collaboration, C. Albajar *et al.*, Phys. Lett. B **209**, 397 (1988).

[66] UA1 Collaboration, C. Albajar *et al.*, Phys. Lett. B **262**, 163 (1991).

[67] J. L. Hewett, S. Nandi and T. G. Rizzo, Phys. Rev. D **39**, 250 (1989).

[68] X. G. He, T. D. Nguyen and R. R. Volkas, Phys. Rev. D **38**, 814 (1988).

[69] N. G. Deshpande and E. Ma, Phys. Rev. D **15**, 1950 (1977).

[70] H. E. Logan and U. Nierste, Nucl. Phys. B **586**, 39 (2000); arXiv:hep-ph/0004139.

[71] A. Dedes, H. K. Dreiner, U. Nierste and P. Richardson, arXiv:hep-ph/0207026.

[72] S. Baek, Y. G. Kim and P. Ko, J. High Energy Phys. **0502**, 067 (2005); arXiv:hep-ph/0406033.

[73] R. Dermisek, S. Raby, L. Roszkowski and R. Ruiz de Austri, arXiv:hep-ph/0507233.

[74] B. Dutta (private communication).

[75] A. Dedes, H. K. Dreiner and U. Nierste, Phys. Rev. Lett. **87**, 251804 (2001); arXiv:hep-ph/0108037.

[76] J. F. de Troconiz and F. J. Yndurain, Phys. Rev. D **71**, 073008 (2005); arXiv:hep-ph/0402285.

[77] K. Maltman, arXiv:hep-ph/0504201.

[78] M. Davier, A. Hocker and Z. Zhang, arXiv:hep-ph/0507078.

[79] S. Baek, Y. G. Kim and P. Ko, arXiv:hep-ph/0506115.

[80] R. P. Feynman, Phys. Rev. Lett. **23** (1969) 1415.

[81] R. P. Feynman, *Photon-Hadron Interactions* (Benjamin, Reading, MA, 1972).

[82] R. P. Feynman, in *Proceedings of Symposium on the Past Decade in Particle Theory, Austin, Tex., 14-17 Apr 1970,* edited by E. C. G. Sudarshan, *et al.* (Gordon and Breach, New York, 1973).

[83] J. D. Bjorken and E. A. Paschos, Phys. Rev. **185**, 1975 (1969).

[84] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. D **71**, 032001 (2005); arXiv:hep-ex/0412071.

[85] M. L. Mangano, AIP Conf. Proc. **753**, 247 (2005); arXiv:hep-ph/0411020.

[86] P. Nason, in *Proceedings of 1997 European School of High-Energy Physics, Menstrup, Denmark, 25 May - 7 Jun 1997,* edited by N. Ellis and M. Neubert (CERN, Geneva, Switzerland, 1998); CERN Yellow Report No. CERN-98-03.

[87] T. Sjostrand, P. Eden, C. Friberg, L. Lonnblad, G. Miu, S. Mrenna and E. Norrbin, Comput. Phys. Commun. **135**, 238 (2001); arXiv:hep-ph/0010017.

[88] A. Warburton for CDF Collaboration, arXiv:hep-ex/0411079.

[89] Fermilab Visual Media Services, Graphic Arts Gallery, http://www-visualmedia.fnal.gov.

[90] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **74**, 2626 (1995); arXiv:hep-ex/9503002.

[91] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. D **65**, 052005 (2002); arXiv:hep-ph/0111359.

[92] L. Miller for CDF Collaboration, Nucl. Instrum. Meth. A **518**, 281 (2004).

[93] A. Sill for CDF Collaboration, Nucl. Instrum. Meth. A **447**, 1 (2000).

[94] CDF Collaboration, F. Abe *et al.*, Fermilab Report No. Fermilab-Proposal-909, 1998 (unpublished); http://www-cdf.fnal.gov/upgrades/PAC/p909.ps.

[95] CDF Collaboration, A. Affolder *et al.*, Nucl. Instrum. Meth. A **453** (2000) 84.

[96] S. Nahn for CDF Collaboration, Fermilab Report No. FERMILAB-CONF-02-305-E.

[97] CDF Collaboration, T. Affolder *et al.*, Nucl. Instrum. Meth. A **526**, 249 (2004).

[98] D. Ambrose, Nucl. Instrum. Meth. A **518**, 117 (2004).

[99] COT group, A. Mukherjee *et al.*, CDF Report No. CDF-NOTE-6267, 2003 (unpublished).

[100] CDF collaboration, D. Allspach *et al.*, Fermilab Report No. FERMILAB-CONF-04-449-E.

[101] CDF Collaboration, L. Balka *et al.*, Nucl. Instrum. Meth. A **267**, 272 (1988).

[102] CDF Collaboration, S. Bertolucci *et al.*, Nucl. Instrum. Meth. A **267**, 301 (1988).

[103] R. Oishi for CDF Plug Upgrade Group, Nucl. Instrum. Meth. A **453**, 227 (2000).

[104] C. M. Ginsburg for CDF Collaboration, Eur. Phys. J. C **33**, S1002 (2004).

[105] G. Ascoli, L. E. Holloway, I. Karliner, U. E. Kruse, R. D. Sard *et al.*, Nucl. Instrum. Meth. A **268**, 33 (1988).

[106] G. Gomez-Ceballos, A. Belloni, A. Bolshov, B. Iyutin, I. Kravchenko *et al.*, Nucl. Instrum. Meth. A **518**, 522 (2004).

[107] M. Shimojima, B. J. Kilminster, K. S. McFarland, A. Vaiciulis and D. J. Holmgren, IEEE Trans. Nucl. Sci. **47**, 236 (2000).

[108] CDF Collaboration, S. Holm *et al.*, IEEE Trans. Nucl. Sci. **47**, 895 (2000).

[109] E. J. Thomson, C. Ciobanu, J. Y. Chung, J. Gerstenslager, J. Hoftiezer *et al.*, IEEE Trans. Nucl. Sci. **49**, 1063 (2002).

[110] E. Sexton-Kennedy, M. Shapiro, R. Snider and R. Kennedy, in Proceedings of International Conference on Computing in High-Energy Physics and Nuclear Physics (CHEP 2000), Padova, Italy, 7-11 Feb 2000, edited by M. Mazzucato and M. Michelotto (unpublished); http://chep2000.pd.infn.it/abst/abs_a245.htm.

[111] E. Sexton-Kennedy and P. Murat, arXiv:physics/0306112.

[112] ROOT Documentation Group, R. Brun *et al.*, *ROOT – An Object-oriented Data Analysis Framework*, users guide 4.04 (CERN, Geneva, Switzerland, 2005); http://root.cern.ch.

[113] B. Brau, Nucl. Instrum. Meth. A **541**, 73 (2005).

[114] CDF Run II Tracking Software web page: http://www-cdf.fnal.gov/upgrades/computing/projects/reconstruction/tracking/tracking.html.

[115] J. Marriner, CDF Report No. CDF-NOTE-1996, 1993 (unpublished); http://cdfkits.fnal.gov/CdfCode/source/VertexAlg/doc.

[116] E. Gerchtein and M. Paulini, arXiv:physics/0306031.

[117] T. Sjostrand, L. Lonnblad, S. Mrenna and P. Skands, arXiv:hep-ph/0308153.

[118] M. A. Dobbs, S. Frixione, E. Laenen, K. Tollefson, H. Baer *et al.*, arXiv:hep-ph/0403045.

[119] CDF II B-physics group, *MC Generators: BGENERATOR,* http://www-cdf.fnal.gov/cdfsim/generators/bgen.html; K. Anikeev, P. Murat, Ch. Paus, CDF Report No. CDF-NOTE-5092, 1999 (unpublished).

[120] D. J. Lange, Nucl. Instrum. Meth. A **462**, 152 (2001).

[121] A. Ryd, D. Lange, N. Kuznetsova, S. Versille, M. Rotondo *et al.*, *EvtGen: A Monte Carlo Generator for B-physics*, CDF Software Documentation Note V00-09-26 (2004), http://cdfkits.fnal.gov/CdfCode/source/EvtGen/doc/?v=5.3.4.

[122] Application Software Group, Computing and Networks Division, CERN, CERN Software Documentation Report No. W5013, 1994 (unpublished).

[123] G. Buchalla and A. J. Buras, Nucl. Phys. B **398**, 285 (1993).

[124] G. Buchalla and A. J. Buras, Nucl. Phys. B **400**, 225 (1993).

[125] G. Buchalla and A. J. Buras, Nucl. Phys. B **548**, 309 (1999); arXiv:hep-ph/9901288.

[126] D. Glenzinski *et al.*, CDF Report No. CDF-NOTE-7463, 2005 (unpublished).

[127] G. J. Feldman and R. D. Cousins, Phys. Rev. D **57**, 3873 (1998); arXiv:physics/9711021.

[128] J. Conway, CDF Report No. CDF-NOTE-6428, 2005 (unpublished).

[129] The FORTRAN code `bayes.f` provided by the CDF Statistics Committee is used. The results of `bayes.f` were cross-checked using the software provided in `bayesianlimit.tar`: the difference was shown to be within 1%. The code is available at http://www-cdf.fnal.gov/physics/statistics/statistics_software.html.

[130] J.Heinrich, CDF Report No. CDF-NOTE-7587, 2005 (unpublished).

[131] CDF Statistics Committee, J. Heinrich *et al.*, CDF Report No. CDF-NOTE-7117, 2004 (unpublished).

[132] V. Krutelyov, C.-J. Lin, D. Glenzinski, M. Herndon, T. Kamon *et al.*, CDF Report No. CDF-NOTE-7314, 2004 (unpublished).

[133] T. LeCompte, CDF Report No. CDF-NOTE-6114, 2002 (unpublished).

[134] R. Field, PYTHIA tunes, http://www.phys.ufl.edu/~rfield/cdf/tunes/rdf_tunes.html.

[135] D. Glenzinski, M. Herndon, T. Kamon, V. Krutelyov, C.-J. Lin *et al.*, CDF Report No. CDF-NOTE-6397, 2004 (unpublished).

[136] C. Chen, I. Cho, C. Hays, M. Herndon, J. Kraus *et al.*, CDF Report No. CDF-NOTE-6394, 2003 (unpublished).

[137] K. Bloom and W. D. Dagenhart, CDF Report No. CDF-NOTE-6347, 2003 (unpublished).

[138] W. D. Dagenhart, CDF Report No. CDF-NOTE-6612, 2003 (unpublished).

[139] M. Herndon, D. Glenzinski, T. Kamon, V. Krutelyov, C.-J. Lin *et al.*, CDF Report No. CDF-NOTE-6318, 2003 (unpublished).

[140] BABAR Collaboration, B. Aubert *et al.*, arXiv:hep-ex/0408096.

[141] C.-J. Lin (private communication).

[142] CDF Collaboration, R. Bernhard *et al.*, arXiv:hep-ex/0508058.

[143] LHC web page: http://lhc.web.cern.ch/lhc/.

[144] S. Hashimoto, M. Hazumi, J. Haba, J. W. Flanagan, Y. Ohnishi *et al.*, KEK Report No. KEK-REPORT-2004-4.

[145] A. Dedes (private communication).

[146] R. S. Wallny for CDF Silicon Group, Nucl. Instrum. Meth. A **546**, 56 (2005).

[147] CDF II offline code browser: http://cdfkits.fnal.gov/CdfCode/source/?v=5.3.4.

[148] C. Hays, P. Tamburello, A. Kotwal, P. Wittich, and R. Snider, CDF Report No. CDF-NOTE-6992, 2004 (unpublished).

[149] C. Hays, Y. Huang, A. V. Kotwal, H. K. Gerberich, S. Menzemer *et al.*, Nucl. Instrum. Meth. A **538** (2005) 249.

[150] W.-M. Yao and K. Bloom for CDF Tracking group, CDF Report No. CDF-NOTE-5991, 2002 (unpublished).

[151] S. Menzemer, CDF Report No. CDF-NOTE-5968, 2002 (unpublished).

[152] CDF muon offline group, J. N. Bellinger *et al.*, CDF Report No. CDF-NOTE-5870, 2002 (unpublished).

[153] M. Bishai, D. Litvintsev, T. Miao, S. Tkaczyk, and R. J. Tesarek, CDF Report No. CDF-NOTE-6144, 2002 (unpublished).

[154] V. Krutelyov, D. Glenzinski, M. Herndon, T. Kamon, C.-J.Lin *et al.* CDF Report No. CDF-NOTE-6835, 2003 (unpublished).

[155] B. Rossi and K. Greisen, Rev. Mod. Phys. **13**, 240 (1941).

[156] L. Eyges, Phys. Rev. **74**, 1534 (1948).

[157] B. Ashmanskas and R. Harr, CDF Report No. CDF-NOTE-6273, 2003 (unpublished).

[158] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. D **68**, 091101 (2003); arXiv:hep-ex/0308059.

APPENDIX A

SETTING UPPER LIMIT ON HEAVY HIGGS MASS USING $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$

It has been recently suggested [15] that a measurement of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ can be used to set an upper bound on the heavy Higgs mass m_A, as well as to set an upper bound on some of other MSSM mass parameters. Since the strongest correlation in the MSSM prediction to $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ comes from $\tan^6\beta/m_A^4$, setting the upper limit on m_A at given $\tan\beta$ is probably the most straightforward. As shown in Fig. 68 [145], for $\tan\beta = 50$, considering a large number of points in the MSSM parameter space with MFV, it is possible to set an upper bound on m_A. Here, the maximum density of the model points is denoted by the magenta curve, the green curve separates the region with 95% of the considered model points, and the green band defines the SM expectations with uncertainties included. The green curve thus defines the (reasonable) upper limit on m_A at the given $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$. On the other hand practically no limit can be set on m_A in a generic way having only the upper limit value of $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$.

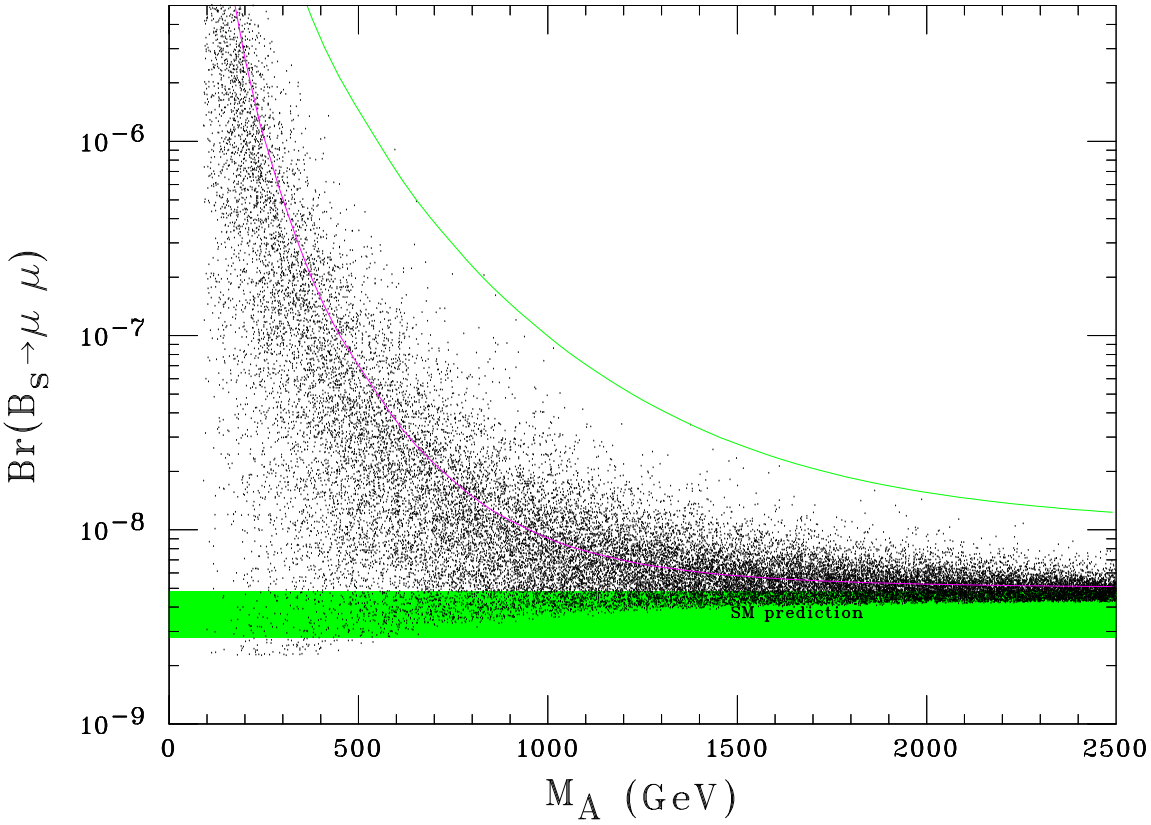

FIG. 68: Setting upper limit on m_A in MFV MSSM at $\tan\beta = 50$.

APPENDIX B

SILICON TRACKER CONFIGURATION

A description of the silicon tracker (mechanical assembly, sensors, and readout electronics) is provided in this Appendix. A summary of the silicon tracker configuration is given in Table XVI.

Silicon layers consist of a number of mechanical ladder assemblies of approximately 15 cm (22 cm for ISL) in length. Each ladder consists electronically of two half-ladders that a read out independently. The ladder assembly includes 4 (6 for ISL) silicon wafers (sensors) with each half wire-bonded to be read out on the ends of the ladder. Except for L00, where the readout is done at the end of the layer, the primary readout electronics is placed on the ends of the ladders. The half-ladders in each layer are arranged in 12 (24 to 36 for ISL) coaxial wedges (in r-ϕ view) and 6 (4 in layer 7) barrels (longitudinally). Adjacent wedges partially overlap in order to reduce gaps in r-ϕ view. As a result, the wedges in a given layer are staggered in r. For L00 due to mechanical limitations two kind of wafers are used: narrow (the closest to the beam) and wide. For ISL inner layer (6) the central ladders are placed at larger radius than the outer ladders; the number of wedges is also different.

The primary detection of the ionizing radiation produced by the charged particles in the silicon tracker is performed on the silicon sensors implemented on the thin rectangular structures, wafers. The wafers are approximately 300 μm thick n-type silicon with $p^+(n^+)$-type strips implanted on axial/top (stereo/bottom) side. A reverse bias voltage of $40 - 80$ V is applied to the sides of the wafer to achieve the depletion of the silicon volume. A charged particle crossing the silicon detector would leave a trace of electron-hole pairs: approximately 1 pair per 3.6 eV energy lost with

roughly 30 thousand pairs produced per MIP per wafer crossed. The charge collected on the sides of the wafer is then transfered via AC coupling to the readout electronics (SVX3D chip). The nominal per-strip resolution is $1/\sqrt{12}$-th of the strip pitch, but since on average the charge is shared between 2 strips, the resolution is even better. For L00 and ISL only every other strip is being read out: a charge on the floating strip is capacitively shared between the neighboring (read out) strips – this improves the resolution and allows to have a smaller number of channels. To reduce the dark current, lower the depletion voltage, and improve the signal to noise ratio the silicon is cooled to $-10°$C (6°C for ISL and portcards).[1] The analog performance of the detector is as designed, with signal to noise ratios over 12:1 for SVX II and ISL, and 10:1 for L00. The detector can be operated at 99% single hit efficiency with virtually no noise occupancy.

The primary readout of the silicon sensors is performed on 128-channel SVX3D chips. The chip includes preamplification, a multi-cell analog storage pipeline, and simultaneous analog and digital operation capability. The chip can also perform dynamic (data-driven) common-mode noise (pedestal) determination and subtraction. Each channel of the chip contains a set of charge integrators followed by 47 cell of analog storage, and an 8-bit analog-to-digital converter (only 7 bits are in fact used). The data from multiple SVX3D chips is converted to an 8-bit optical form in cards (port cards, PC). This data is then received by interface boards in the collision hall and the subsequent parts of the DAQ.

[1]The original cooling temperature of $-6°$C for L00 and SVX II was lowered to mitigate the effects of radiation damage to be able to efficiently run the detector during the whole Run II, an estimated 8 fb^{-1} of integrated luminosity by 2009. These adjustments were made shortly after Run IIb silicon project was canceled in 2003 [146].

TABLE XVI: Silicon detector configuration.

	L00	SVX II					ISL	
Layer	0	1	2	3	4	5	6	7
half-ladders	72	72	72	72	72	72	152	144
Ax. chan./HL	128/256	256	384	640	768	896	512	512
St. chan./HL	–	256	384	640	512	896	512	512
Stereo type	–	90°	90°	1.2°	90°	−1.2°	−1.2°	1.2°
HL/wedge	6	6	6	6	6	6	6	4
wedges	12	12	12	12	12	12	$\frac{28}{24}$	36
radius, mm	14/16	30/25	46/41	70/65	87/82	106/101	$\frac{231/226}{202/197}$	290/286
wafers/HL	2	2	2	2	2	2	3	3
Total $N_{\text{chan}}^{\text{axial}}$	13824	18432	27648	46080	55296	64512	77842	73728
Total $N_{\text{chan}}^{\text{axial}}$	13824			211968			151552	
Total $N_{\text{chan}}^{\text{stereo}}$	–	18432	27648	46080	36864	64512	77842	73728
Total $N_{\text{chan}}^{\text{stereo}}$	–			193536			151552	
$\text{width}_{\text{wafer}}^{\text{active}}$, mm	6.4/13	15	24	38	46	58	56	56
$\text{length}_{\text{wafer}}^{\text{active}}$, mm	76	74	74	74	74	74	72	72
wafer $N_{\text{strips}}^{\text{axial}}$	256/512	256	384	640	768	896	1024	1024
$\text{pitch}_{\text{strip}}^{\text{axial}}$, μm	25	60	62	60	60	65	55	55
$\text{pitch}_{\text{readout}}^{\text{axial}}$, μm	50	60	62	60	60	65	110	110
wafer $N_{\text{strips}}^{\text{stereo}}$	–	512	576	640	512	896	768	768
$\text{pitch}_{\text{strip}}^{\text{stereo}}$, μm	–	141	125.5	60	141	65	73	73
$\text{pitch}_{\text{readout}}^{\text{stereo}}$, μm	50	141	125.5	60	141	65	146	146

APPENDIX C

OFFLINE RECONSTRUCTION ALGORITHMS

The algorithms used to reconstruct (compute) the properties of the collision events relevant to the present analysis are detailed in this Appendix. The description of the charged particle reconstruction, tracking, is given in Section C.A, followed by the description of muon identification and reconstruction given in Section C.B. The algorithms used to calculate the properties of the multi-track vertex (B-meson candidate decay vertex and collision event vertex) are described in Section C.C. A substantial part of the documentation provided here is the result of the translation of the code used in this analysis available for browsing at [147], the references to the documentation material is provided in text whenever they are available.

A. Tracking

A charged particle in a magnetic field travels along a helical trajectory, which defines a track. A track helix in CDF II is defined with respect to the nominal beamline with the helix axis being parallel to the beamline. A helix can be defined by five parameters: curvature ($C = q/2R$), where R is the radius of the helix projection circle in r-ϕ view and q is the sign of the particle charge; signed distance from the beamline or impact parameter (d_0), z-coordinate (z_0), azimuthal and polar angle (ϕ_0 and $\lambda = \cot\theta$) of the helix direction all at the point of the helix closest approach to the beamline. The sign of the impact parameter is defined by the position of the beamline with respect to the helix: going along the helix in the particle direction if the beamline is to the right (left) the impact parameter is positive (negative). The

helix is thus defined by the following parametric equations in Cartesian coordinates:

$$\begin{aligned} x &= \frac{1}{2C}(\sin(2Cs+\phi_0) - \sin\phi_0) + d_0 \sin\phi_0, \\ y &= -\frac{1}{2C}(\cos(2Cs+\phi_0) - \cos\phi_0) + d_0 \cos\phi_0, \\ z &= z_0 + s\lambda, \end{aligned} \tag{C.1}$$

where s is the projected length along the helix. The helix curvature is related to the particle transverse momentum as $|2C| = \frac{eBc}{p_T} \approx \frac{42}{p_T(\mathrm{GeV}/c)}$ cm, where $B \approx 1.41$ T is the magnetic field. Thus for a particle of a unit charge there is one to one correspondence between the helix parameters and the particle momentum and position.

The main purpose of the tracking is to reconstruct a helix based on the hits that a particle left in the silicon detector and the COT. Track reconstruction in each tracker separately is done by COT and silicon standalone algorithms. The results of standalone algorithms are used for combined (integrated) COT and silicon tracking.

In COT the track reconstruction begins with segment reconstruction in each superlayer. The helix segments reconstruction in a superlayer begins with finding the hit triplets starting from the inner wire of a cell [148]. A line defined by a triplet should have a slope (angle with respect to radial direction $\lesssim 50°$) consistent with a trajectory of a particle with $p_T \gtrsim 0.3$ GeV/c and passing close to the beamline. The triplets are then extended iteratively by adding the hits on the neighboring wires that are roughly within 1 mm from the line defined by the set of hits from a previous iteration.

The next step after segments are searched for in all superlayers is the axial pattern recognition where tracks are reconstructed in r-ϕ view. Here two separate algorithms are run in parallel: axial segment linking (SL) and histogram linking[1] (HL). The

[1]The HL algorithm is the base of COT tracking at L3.

segments are linked successively starting from the outer two superlayers that define track parameters, the segments on the inner two axial superlayers are then linked in if consistent with the same track parameters. The histogram linking is seeded by a segment and the beam position which defines a seed circle around which the hits are counted in 200μm bins. The hits in the bin with more than 10 hits together with the hits within 1 mm around that bin define the HL axial track.

After the axial hits are reconstructed the stereo segments are linked to the axial track in r-z view. A stereo segment is considered for linking if it is in the cells close (within 4) to those intersected by the axial track. The segments are added starting from the superlayer next-in from the outermost axial superlayer of the track. All hits on the first considered stereo segment are used in r-z fit to z_0 and λ helix parameters. The outermost stereo segment r-z fit parameters are then used to select and fit the hits from segments on the next inner superlayer once that matches (with $\Delta\phi_{\mathrm{SL}} < 0.01$) with the segment in the outer segment. Once two stereo segments are linked, the segments on the remaining stereo superlayers are added if they match to the combined r-z fit with $\chi^2 < 100$. After the r-z fit for the track is defined by the linked segments, the hits from the stereo superlayers with missing segments are linked to the track.

After the segments and hits from the stereo and axial superlayers are combined in a track a five parameter fit is performed to determine the helix parameters. The final step of the COT reconstruction is to combine the results of SL and HL based tracking are combined: duplicates are removed and the best tracks are reported.

The use of silicon in tracking is based on the silicon physical hit reconstruction in each layer. To do this each strip with charge readout above noise threshold is considered on each side of all half-ladders. The strips are collected into clusters based on the collected charge. The centroids of the cluster then defines a two-dimensional

hit, depending on the readout side, as ϕ, z, or *sas*-hit. The global position of the hit is defined by a line in the half-ladder plane.

Tracks with silicon hits are reconstructed by a combination of several tracking algorithms (in the order of the call): COT track parameters are used as seeds and the silicon hits are attached using a progressive fit, outside-in tracking (OI); silicon hits alone (unused by OI algorithms) are used to reconstruct tracks (silicon standalone tracks); silicon standalone tracks are used as seeds and the matching COT hits (unused by the COT standalone tracks) are attached, inside-out tracking (IO) [149].

Two outside-in algorithms are used: a Run I-based algorithm (OI/Z) which uses a progressive χ^2-based fit [150], and an algorithm (KalOI) based on Kalman filter based fitter [151] which uses the COT tracks left unused by the OI/Z algorithm. The OI/Z algorithm proceeds in two stages. First ϕ-hits are searched for starting from the outermost intersected layer (OI stage). The best track candidates from this stage are nominated as good OI tracks. The next stage attempts to add the z and *sas* hits to the OI tracks, with any success this results in an OIZ track. In the KalOI algorithms the ϕ and stereo hits are matched together on each layer (stereo hit is considered only if a ϕ hit matches), thus producing (at best) a 3D match in each layer. All matching hit combinations are considered down to the innermost layers, producing a candidate per hit combination. Once the innermost layer is reached the best candidate is selected as KalOI output track. In both cases the multiple scattering and energy loss in the silicon and ambient material is considered, although it is more detailed in KalOI algorithm. The OI/Z algorithm is rather fast and is able to find most of the tracks. The KalOI algorithm is more efficient but it takes a larger amount of time to process, hence it is only used on the tracks leftover from OI/Z.

The silicon standalone algorithm uses practically the same algorithm as KalOI to attach the hits. The seed in this case is constructed of 2 3D silicon hits and the

beam line position. The 3D hits in this case are a result of all possible combinations of 2D silicon hits within a given half-ladder. After the seed is defined the hits on the remaining ladders are added similarly to the KalOI algorithm. In this case Kalman fitter results from the fit in both track directions are combined to better estimate the track parameters.

It should be noted that the Kalman fitter is used extensively by CDF. The ability of the fitter to produce the track helix parameters at a given point (origin being the most interesting) with proper incorporation of multiple scattering and energy loss effects is important. The tracks used in the analysis, those from $B^0_{s(d)} \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ (or other) candidates were refit using the Kalman fitter to obtain an improved mass and vertex resolution.

The integrated tracking reconstruction performance is characterized by the track resolution and tracking efficiency. In central region the performance largely depends on the track isolation (number of non-track hits around the track) and on the track transverse momentum. The best performance corresponds to high-p_T ($\gtrsim$ 5 GeV/c) isolated tracks. For the COT algorithms the reconstruction efficiency of such tracks is close to 100%, the corresponding track parameters resolution are $\sigma_{p_T}/p_T^2 \approx 0.15\%/\mathrm{GeV}/c$, $\sigma_{z_0} \approx 0.3\ cm$, $\sigma_{d_0} \approx 250\ \mu$m, $\sigma_{\phi_0} \approx 0.6$ mrad, and $\sigma_\lambda \approx 3$ mrad. The efficiency of the OI algorithms is close to 95%. In this case the resolution on momentum and azimuthal angle remains practically the same, the impact parameter resolution improves to $\sigma_{d_0} \approx 20\ \mu$m, and depending on the number of attached stereo silicon hits the stereo resolution improves down to $\sigma_{z_0} \approx 50\ \mu$m and $\lambda \approx 1$ mrad. The impact parameter resolution decreases with lower momentum, the dependence is shown in Fig. 31 [113, 114]. Here the resolution distribution includes the beam size of about 30 μm, after it is subtracted the proper track resolution roughly follows the dependence $\sigma_{d_0} = \sqrt{30^2 + (30/p_T(\mathrm{GeV}/c)^2)}\ \mu$m for the case with-

out L00, once a track has a L00 hit the constant term in the dependence goes down by a factor of two.

B. Muon reconstruction

Muon reconstruction is based on combining the information from the tracker, muon subdetectors, and calorimeters. A typical signature of a muon with momentum above 1-2 GeV/c the central detector region is a track, a set of muon detector hits along a line roughly pointing to the track, and a calorimeter energy deposition consistent with a MIP, as briefly described in Section III.E.1 and illustrated in Fig. 32.

Muon specific reconstruction steps are the following [152]: physical hits are reconstructed in all the chambers and counters with raw hit data, muon stubs are reconstructed in the drift chamber stacks with hits, finally the tracks are linked with the matching stubs.

The physical hit reconstruction in drift chambers converts the hit time measured by TDCs to the drift distance from the sense wire. In addition the position along the wire is performed in the chambers with z-position readout.

Muon stub reconstruction in both CMU and CMP uses the same pattern recognition algorithm to construct stub candidates with three or four hits. The pattern recognition is based on xy (or r-ϕ) information only: x coordinate of the hit as a rough estimate is the wire position plus/minus the drift sign, and y coordinate is the wire plane coordinate of this layer. Both drift signs are considered for all the hits. Two hits on the interleaving layers (first on 0 and 2, then on 1 and 3) are used to define a line around which the search road is defined. The line slope should be within about 35°(55°) for CMU (CMP) from normal to the wire plane to be consistent with a muon coming from inside the detector. A candidate is created if there are at least

three hits in the search road of 0.5 cm (1.5 cm) for CMU (CMP). The candidates are then fit in xy view using CMU/CMP specific drift model and a linear fitter. The drift model is used to get a predicted drift distance based on the fitted line (muon trace) location with respect to the sense wire. The model roughly follows the field lines picture: the fixed drift time points are roughly on the circle up to a certain value (1.6 cm for CMU and 0.8 cm for CMP) and effectively on the wire plane for larger drifts. An additional fit in yz view is done for the CMU candidates with valid z-position measurements on the hits. Finally the best candidates are selected based on the hit number and fit χ^2 and reported as valid muon stubs.

The details of the CMX stub pattern recognition can be better visualized with Fig. 27. First a pair of chambers with wires aligned on the radial lines (radial chambers) with hits is used to seed the algorithm. All the hit pairs in these radial chambers are considered and the one defining a line with a slope within about 12° from the normal is used as a seed pair. The drift sign of the seed pair in case of ambiguity is also constrained by the number of intersected chambers with hits. The seed pair defines the box of 14 chambers (2 on each of the six layers other than the seed layers plus the seed ones). Next another radial chamber pair is searched for hits with a slope consistent (within 3°) with the seed pair and is positioned within 4.6 to 6 cm from the line defined by seed pair. The information in the two radial hit pairs is enough to define the radius (or equivalently, z) of the candidate stub at the given layer which in turn determines the sense wire separation in azimuth between the layers. The calculated radius should correspond to the detector radial range (4 to 5.5 m). This leaves practically no ambiguity in attaching the hits in the remaining chambers using a tight search window of about 0.4 cm. In case a secondary pair is not found the box is searched for the third hit. Again, three hits is enough to define the radius and the wire separation, which allows to proceed with the tight window search. The candi-

date is then fit using its signed drift distances accounting for the particular geometry. The fit uses the drift model similar to CMP radial drift model but with additional nonlinear corrections. A candidate is accepted and made a stub if it passes the fit quality requirements. The resulting stub is three-dimensional, as in the case of CMU. In addition to the chamber hits, the scintillator hits in the counters matching with the stack, shown in blue in Fig. 27, are added to the CMX stub.

The muon linking is done in two steps. First the track are matched with stubs using a track extrapolator, and last the matched pairs are combined into a track plus matching stubs (one per detector) combination which defines a muon.

The track extrapolator is able to take a track helix parameters as an input and give an expected helix position at a particular stop point. A stop point can be one of the following: a fixed radius (used for CMU and central calorimeters), a fixed z (used for the plug calorimeters), a plane (used for CMP and CMX miniskirt), and a cone (used for CMX arches). The main function of the extrapolator is to provide a solution for a particle motion in a non-uniform magnetic field with appropriate precision. It uses a simplified detector geometry that accounts for the regions with rapid magnetic field change. An average energy loss in the calorimeter and the steel support structure is also accounted for. The energy loss is in fact used as an internal parameter tuned to produce an unbiased (independent from muon kinematics) expected track position in the muon detectors consistent with the mean measured position.[2]

In muon to stub matching the extrapolator produces an estimate of the track position and direction at the stop point defined by this stub detector geometry. The match value $\Delta x \equiv r\Delta\phi_r$ is used to select a match, where r is the radial coordinate of the stub and $\Delta\phi_r$ is the difference in the azimuthal coordinate between the stop

[2]Both GEANT3 simulation and data was used iteratively to tune this extrapolator.

point and the stub. The match cut is placed at 30 cm, 60 cm, and 50 cm for CMU, CMP, and CMX respectively. This corresponds to about $2\sigma_{\text{muls}}$ cut for the minimum detectable muon momentum in that detector. Additional loose requirements are also applied in the transverse (r-z) direction not to produce inconsistent matches: e.g., a track going to the positive z can not correspond to a stub in CMU at negative z. The result of matching is the list of stubs with all tracks matched to each stub.

In the final step the pre-sorted stub to track matches are nominated to muons starting from the best stub-track match pair. The best matching stubs from the other muon detectors associated with that track are combined into a single track plus one stub per detector match which is a base of a muon object. A combination of systems in which the muon stub is matched defines the muon type: CMU, CMP and CMX muons have a stub in the corresponding systems; a muon with matching stubs in CMU and CMP systems is a CMUP muon. The muon object is completed after the energy measured in the calorimeter towers intersected by the muon are computed and added to the muon match data.

The integrated muon reconstruction performance is characterized by the efficiency close to 100% for relatively high momenta (about twice the minimum detectable muon momentum in that detector). This does not include tracking efficiency.

C. Vertex reconstruction

A vertex is defined by a set of tracks consistent with being originated from the common point. The important signature of a B-hadron decay is the decay vertex displaced (in the direction of the meson momentum) with respect to the primary collision vertex. Therefore it is essential for this analysis to be able to properly reconstruct the vertices.

The main ingredients in the vertex reconstruction procedure are: a proper track

helix parameters measurement including the correct covariance matrix, and a flexible tool that is able to fit a set of helices to a fixed point. The track helix measurement is obtained with the help of track Kalman fitter. The vertex fitting tool common for this type of analyses in CDF is the Run I based algorithm for secondary vertex fit with (optional) mass and pointing constraints, CTVMFT [115]. In addition to the decay vertex the primary collision vertex reconstruction is required in this analysis. The method used for primary vertex reconstruction is again based on the CTVMFT, PrimeVtx.

The track helix fit to a common vertex in CTVMFT is based on the following. A point on a helix is defined by 5 parameters defining the curve plus the distance on the curve, as given in Eq. (C.1). A constraint for a helix to go through the point $\vec{r_s} = (x_s, y_s, z_s)$ would correspond to a constraint on three of the six values defining a point on the helix. In CTVMFT the d_0, z_0, and s are chosen to be constrained. They are defined by the solution of Eq. (C.1):

$$\begin{aligned} \sin 2Cs &= 2C(x_s \cos\phi_0 + y_s \sin\phi_0), \\ d_0 &= y_s \cos\phi_0 - x_s \sin\phi_0 + \frac{1}{C}\sin^2 Cs, \\ z_0 &= z_s - s\lambda. \end{aligned} \tag{C.2}$$

Assume now that the measured helix parameters are $\vec{\bar{X}} = (\bar{C}, \bar{\phi}_0, \bar{\lambda}, \bar{d}_0, \bar{z}_0)$ with a corresponding covariance matrix of the measurement denoted as G. Thus a measure of the probability that this helix can pass through the point (x_s, y_s, z_s) and have a set of parameters $\vec{X} = (C, \phi_0, \lambda, d_0, z_0)$ is given by χ^2_{trk} value defined as:

$$\chi^2_{\text{trk}} = (\vec{X} - \vec{\bar{X}})^T G^{-1} (\vec{X} - \vec{\bar{X}}) \equiv \vec{\xi}^T G^{-1} \vec{\xi}. \tag{C.3}$$

For a set of N tracks to pass through the same point $\vec{r_s}$ a combined χ^2 is defined as

$$\chi^2 = \sum_{i=1}^{N} \vec{\xi_i}^T G_i^{-1} \vec{\xi_i}. \tag{C.4}$$

With the constraints set by Eq. (C.2) applied there are three free parameters (C, ϕ_0, λ) for each track plus three defining a vertex $\vec{r_s}$. To find these the χ^2 defined by Eq. (C.4) is minimized. The minimization problem is simplified by linearization of extremum conditions and is then solved iteratively. The algorithm is seeded by the zeroth order vertex estimate given by an average (x_p, y_p) position of intersections of all the helix projection circles in the r-ϕ view and an average z of all tracks around that position.

A set of tracks constrained to a vertex can additionally be constrained by an invariant mass. Similarly the vertex can be constrained to a line (beam line constraint) or constrained to point to another vertex (vertex track vector sum parallel to the distance to another vertex). These constraints are done by adding the Lagrange multipliers corresponding to the constraint to the fit χ^2.

In the case of primary vertex the particular particle content is not as important, only the best possible vertex position measurement is. The vertex reconstruction is performed by PrimeVtx algorithm. The ctvmft is used in this case to define a quality of the primary vertex candidate and of each track in that vertex following Eqs. (C.3) and (C.4). First all OI tracks (silicon hits required) are selected from the event that are not the part of the decay vertex and that pass some quality cuts. Since the primary vertex is searched for around the expected secondary/decay vertex, only the tracks that are within 1 cm in z from the secondary candidate are selected, these tracks should also be within 1 mm from the beamline. This set is used iteratively to find the best vertex estimate. The iteration input tracks are fit constrained to the same vertex and also constrained to the beamline (position and error not worse than the beamline). The last track that causes fit errors (non-convergence and such) or

the worst from those with $\chi^2_{\rm trk} > 10$ or from vertex candidate with $\chi^2_{\rm vtx}/{\rm ndf} > 5$ is removed in each iteration. The iteration continues until no track is to be removed. On average 9 tracks are used to fit the vertex. The efficiency of the algorithm in the b-hadron production environment is above 95%. The reconstructed primary vertex allows for better displaced/secondary vertex signal selection compared to the case of secondary vertex reconstruction with respect to the beamline.

APPENDIX D

ABSOLUTE VS. RELATIVE NORMALIZATION

The relative normalization method used in the present analysis is not the only choice to perform the branching ratio measurement. An absolute normalization measurement used in the previous stages of the search for $B_s^0 \to \mu^+\mu^-$ at CDF [21, 84] can be defined as

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \frac{N_{B_s^0}}{2\mathcal{L}\sigma_{B_s^0}\alpha_{B_s^0}\epsilon_{B_s^0}}, \tag{D.1}$$

where the factor of two accounts for B_s^0 and $\overline{B_s^0}$, $\mathcal{L}$ is the integrated luminosity, and $\sigma_{B_s^0} \equiv \sigma_b f_s$ is the B_s^0 production cross-section commonly defined (and measured) as a product of b-hadron production cross-section σ_b and the B_s^0 fragmentation fraction f_s. The values $\alpha_{B_s^0}$ and $\epsilon_{B_s^0}$ are the same as in Eq. (4.1). In principle, the absolute normalization measurement would have been easier to perform should the uncertainties on the input values be insignificant. Unfortunately this is not always the case, and one should consider the measurement that provides the better statistical resolution.

The relative normalization becomes advantageous when there is a substantial similarity between the two decay modes. In particular, this is the case for the $B_{s(d)}^0 \to \mu^+\mu^-$ and $B^+ \to J/\psi K^+$ modes, which both possess the opposite sign dimuon signature. Provided that the dimuon events in both cases are selected by the same trigger, this would lead to a substantial cancellation of the acceptance and the efficiency ratios, as described in Sections IV.G and IV.G.1. Additionally, the f_u/f_s ratio has an uncertainty smaller than that of f_s due to the significant anticorrelation between the two fractions [51]. Overall, for the $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ measurement the

combined input uncertainty in the absolute normalization method is a factor of about 1.4 larger compared to the case of the absolute normalization [22]: 17% vs. 24%.

APPENDIX E

SUMMARY OF THE TRIGGERS USED IN THE ANALYSIS

The following four L3 trigger paths are used in the presented analysis[1]:

- `RAREB_CMUCMU_SUMPT` and `RAREB_CMUPCMU` which use TWO_CMU L1 and L2 triggers as input;
- `RAREB_CMUCMX_SUMPT` and `RAREB_CMUPCMX` which use CMU_CMX L1 and L2 triggers as input.

The requirements made at L3 for are listed in Table V as "L3 fiducial".

The L1 dimuon triggers used for the above mentioned trigger paths including the requirements made on each trigger are:

- L1_TWO_CMU1.5_PT1.5:
  - Two CMU low p_T stubs ($p_T > 1.5$ GeV/c),
  - CMU stubs matched to XTRP tracks with p_T >1.5 GeV/c,
  - The CMU stubs are separated by at least two L1 muon trigger towers (where the gap between the wedges is counted as a tower) or they are on separate sides (east and west) of the detector.
- L1_CMU1.5_PT1.5_&_CMX1.5_PT2_PS1 (run range #2):[2]
  - One CMU low p_T stub ($p_T > 1.5$ GeV/c) matched to an XTRP track ($p_T^{\mathrm{XFT}} > 1.5$ GeV/c),
  - One CMX low p_T stub ($p_T > 1.5$ GeV/c) matched to an XTRP track ($p_T^{\mathrm{XFT}} > 2.0$ GeV/c).

[1]The dimuon trigger algorithms are described in Sections III.D.2 through III.D.4.

[2]The run range numbers are specified in Table XVII.

- L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX_PS1 (from range #3):
  - One CMU low p_T stub ($p_T > 1.5$ GeV/c) matched to an XTRP track ($p_T^{\mathrm{XFT}} > 1.5$ GeV/c),
  - One CMX low p_T stub ($p_T > 1.5$ GeV/c) matched to an XTRP track ($p_T^{\mathrm{XFT}} > 2.0$ GeV/c),
  - CMX low p_T stub matched to CSX scintillator hits.

The L2 muon trigger was implemented for the runs starting with run range #7 on May 18, 2004. Prior to that the muons at L2 were generally auto-accepted or prescaled based on the L1 trigger information. The L2 muon logic uses the XFT track ϕ (with 1.25° granularity) and p_T information together with the muon hit information as described in detail in Section III.D.3. The following L2 dimuon triggers are used in the trigger paths considered in this analysis (the `RAREB` and `JPSI` paths):

- `L2_CMU1.5_PT1.5_&_CMX1.5_PT2_DPHI120_OPPQ`,
- `L2_TWO_CMU1.5_PT1.5_DPHI120_OPPQ`, replaced in June 2004 by `L2_TWO_CMU1.5_PT2_DPHI120_OPPQ`.

The last `TWO_CMU` trigger requires each muon to have $p_T > 2$ GeV/c (PT2). The L2 dimuon triggers in question require two XFT tracks matching to muons stubs, the tracks being of opposite charge and having $\Delta\phi_{SL6} < 120°$.

APPENDIX F

MEASUREMENT OF LEVEL-1, 2, AND 3 LOW-P_T DIMUON TRIGGER EFFICIENCIES

Measurement of the dimuon trigger efficiencies for the triggers used in the present analysis (listed in Appendix E) is given in this appendix. The description begins with Section F.A giving the definition of the trigger path efficiency in terms of the L1, L2, and L3 components together with the domain (kinematic variables and data collection periods) where the efficiency measurement is applicable. This is followed by the Section F.B which determines the L1 dimuon trigger efficiency in terms of L1 single muon trigger efficiency, which measurement is detailed in Section F.C. The L2 and L3 dimuon trigger efficiencies are detailed in Sections F.D and F.E respectively. The cross-check of the measurement is performed on the high statistics $J/\psi \rightarrow \mu^+\mu^-$ data sample as described in Section F.F. The results of the trigger efficiency measurement are summarized in Section F.G.

A. Trigger efficiency

The total trigger efficiency is the product of the L1, L2, and L3 trigger efficiencies. It can be expressed as the product of the conditional probabilities:

$$\begin{aligned}\epsilon_{\text{trig}} &= \epsilon_{\text{L1}} \times \epsilon_{\text{L2}} \times \epsilon_{\text{L3}} \\ &\equiv P(\text{L1}, \text{offl} \,|\, \text{offl}) \times P(\text{L1}, \text{L2}, \text{offl} \,|\, \text{L1}, \text{offl}) \times P(\text{L1}, \text{L2}, \text{L3}, \text{offl} \,|\, \text{L1}, \text{L2}, \text{offl}),\end{aligned} \tag{F.1}$$

where $\epsilon_{L1} \equiv P(\text{L1}, \text{offl} \,|\, \text{offl})$ is the fraction of unbiased events that pass the L1 dimuon trigger, given that both muons have been reconstructed offline. The sec-

ond probability term, $\epsilon_{L2} \equiv P(\mathrm{L1}, \mathrm{L2}, \mathrm{offl} \,|\, \mathrm{L1}, \mathrm{offl})$, is the fraction of events unbiased relative to L2 that pass the L2 trigger, given that the muons passed the L1 dimuon trigger and have been reconstructed offline. The third probability term, $\epsilon_{L3} \equiv P(\mathrm{L1}, \mathrm{L2}, \mathrm{L3}, \mathrm{offl} \,|\, \mathrm{L1}, \mathrm{L2}, \mathrm{offl})$, is the fraction of events unbiased relative to L3 that pass the L3 trigger, given that the muons passed the L1 and L2 dimuon trigger and have been reconstructed offline.

For the L1 trigger efficiency, the measured value is $\epsilon_{L1} = P(\mathrm{L1}, \mathrm{L3}, \mathrm{offl} \,|\, \mathrm{L3}, \mathrm{offl})$ instead of $P(\mathrm{L1}, \mathrm{offl} \,|\, \mathrm{offl})$. Similarly for the L2 trigger efficiency, the measured value is $\epsilon_{L2} = P(\mathrm{L1}, \mathrm{L2}, \mathrm{L3}, \mathrm{offl} \,|\, \mathrm{L1}, \mathrm{L3}, \mathrm{offl})$ instead of $P(\mathrm{L1}, \mathrm{L2}, \mathrm{offl} \,|\, \mathrm{L1}, \mathrm{offl})$. This is due to the fact that the dataset needed to measure $P(\mathrm{L1}, \mathrm{L2}, \mathrm{offl} \,|\, \mathrm{L1}, \mathrm{offl})$ and $P(\mathrm{L1}, \mathrm{offl} \,|\, \mathrm{offl})$ has very limited statistics. Since the L3 to offline tracking efficiency is nearly 100% [153], the two numbers should be nearly identical. As a cross-check, L1 trigger efficiency was measured using both methods and the results are consistent within statistical uncertainty.

Note that prior to May 2004 the L2 trigger requirements were trivial (all events triggered by dimuon L1 trigger were passed to L3). For that period of data taking (roughly 2/3 of all the data sample) ϵ_{L2} is exactly 100%.

Specified here are the selections used to define the efficiency given in Eq. (F.1):

"offl": defines the good (di)muon cuts, the same as the baseline selections listed in Table V, less the trigger matching requirements and less the silicon quality.

L1: defines the trigger match. It is the same as "L1 match" in Table V with an addition of a particular L1 trigger bit that marks that the trigger has fired.

L2: L2 trigger bit and the "L2 match" from Table V.

L3: L3 bit.

The data used for the efficiency measurement corresponds to the same period of data taking as the one of Chapter IV.

In an attempt to follow the major trigger and detector condition changes the efficiency measurement is provided in separate run ranges corresponding to the similar trigger and detector conditions, as described in Table XVII. For the CMU, in addition to the run range split, the top ($\sin\phi > -0.5$) and the bottom ($\sin\phi < -0.5$) parts are considered separately for the ranges #3-7; these are the periods that were affected by the COT aging. For the CMX the ranges #3-5 are merged; the #6 and #7 are merged too, this is done to increase the statistics assuming that the CMX is not affected by the COT aging.[1] The beginning of range #7 corresponds to the period when the COT started running in a normal configuration (after running in a compromised state with three inner superlayers turned off), and to the period when L2 muon trigger was implemented and when the L3 executable was changed to run with improved reconstruction algorithms of offline software (major) version 5.

[1]See discussion in Section F.C.6.

TABLE XVII: Summary of the run ranges for the efficiency measurement with the approximate begin-end dates specified. The luminosity per range is calculated for the runs with good silicon.

#	Dates	$\mathcal{L}$, pb^{-1}	comment
1	03/01/02-08/20/02	18.2	CMU only, 2-miss XFT, bad matchbox #8.
2	08/20/02-10/09/02	15.0	CMU (good matchbox #8), CMX: 2-miss XFT.
3	10/09/02-04/11/03	60.2	1-miss XFT, no ϕ-dependence, have η-dependence.
4	04/11/03-07/01/03	59.6	1-miss, some ϕ-dependence, have η-dependence ($\sin\phi > -0.5$ and $\sin\phi > -0.5$ considered separately).
5	07/01/03-09/06/03	50.5	significant ϕ-dependence, ditto.
6	11/24/03-02/13/04	46.4	large ϕ-dependence, minor η-dependence.
7	05/19/04-06/16/04	32.1	ditto, L2 measurement starts; L3 runs version 5.
8	06/19/04-08/23/04	84.5	good COT, no ϕ-dependence, no η-dependence

B. Level-1 dimuon trigger efficiency ($\epsilon_{\mathbf{L1}}$)

The L1 dimuon trigger efficiency is measured for the three L1 dimuon triggers specified in Appendix E:

- L1_TWO_CMU1.5_PT1.5,
- L1_CMU1.5_PT1.5_&_CMX1.5_PT2_PS1 (up to November 2002),
- L1_CMU1.5_PT1.5_&_CMX1.5_PT2_CSX_PS1 (after 10/02).

The L1 dimuon trigger efficiency can be expressed as

$$\epsilon^{\mu\mu}_{\rm L1}(x_1, x_2) = \epsilon^{\mu}_{\rm L1}(x_1) \cdot \epsilon^{\mu}_{\rm L1}(x_2) \cdot \zeta(x_1, x_2), \tag{F.2}$$

where $\epsilon_\mu(x)$ is the single muon L1 trigger efficiency and ζ corresponds to the dimuon correlation effects. The x_i denote the set of variables that the efficiency is found to depend on. It is assumed that with the event selections described in Section F.A it is reasonable to use $\zeta = 1$. Measuring the L1 dimuon trigger efficiency is thus reduced to measuring the single muon L1 trigger efficiency.

C. Single muon Level-1 trigger efficiency

1. Dataset

The dataset used to measure the single muon L1 trigger efficiency is collected from the inclusive L1 CMUP 4 GeV/c trigger. The L2 trigger is set for auto-accept. The L3 trigger requires either a CMU muon with p_T >1.5 GeV/c (L3 path `JPSI_CMUP4_CMU1.5`) or a CMX muon with $p_T > 2$ GeV/c (L3 path `JPSI_CMUP4_CMX2`).

2. Methodology

To measure the single muon L1 trigger efficiency, a sample of J/ψ events is reconstructed first by pairing the muon reconstructed offline consistent with firing the L1 CMUP4 trigger with the remaining muons reconstructed in the event. The latter muon is referred to as the unbiased muon.

The denominator of the single muon L1 trigger efficiency is the number of events which have an unbiased muon that satisfy the "offl" requirements and have a dimuon invariant mass within 3σ of the mean J/ψ mass for the given p_T. To make a uniform measurement for all the dimuons in the sample the COT-based dimuon mass is used. The 3σ window is allowed to vary based on the J/ψ mass resolution measured for each muon p_T bin. Events that have more than one offline dimuon pair that could fire the dimuon L3 CMU (CMX) triggers are excluded. This corresponds to about

0.2% (0.1%) loss of the CMU (CMX) sample.

The numerator of the efficiency measurement is given by the number of events where the unbiased muon is matched to the L1 muon primitive, following the "L1 requirement" given in Section F.A.

A side-band subtraction is made to correct for fake muons. The single muon L1 trigger efficiency is defined as:

$$\epsilon^{\mu}_{\mathrm{L1}} = \frac{N^{\mathrm{sig}}_{\mathrm{passed}} - N^{\mathrm{sb}}_{\mathrm{passed}}}{N^{\mathrm{sig}}_{\mathrm{total}} - N^{\mathrm{sb}}_{\mathrm{total}}}, \tag{F.3}$$

where $N^{\mathrm{sig}}_{\mathrm{passed}}$ is the number of events that passed the L1 trigger requirements in the J/ψ mass peak, $N^{\mathrm{sb}}_{\mathrm{passed}}$ is the number of events that passed the L1 trigger requirements in the side-band regions, $N^{\mathrm{sig}}_{\mathrm{total}}$ is the number of events in the J/ψ mass peak, and $N^{\mathrm{sb}}_{\mathrm{total}}$ is the number of events in the side-band regions. The signal region is defined to be within 3σ of the fitted J/ψ mass peak ($m_{J/\psi}$). The mass peak is modeled by a single Gaussian and the background is parameterized using a first order polynomial. The side-bands are defined within $5\sigma < |m_{J/\psi} - m_{\mu\mu}| < 8\sigma$. The number of events in the side-band and signal regions are obtained by counting the number of entries in those windows. The statistical uncertainty is calculated using the following expression:

$$\begin{aligned} \delta_{\mathrm{stat}} &= \frac{\sqrt{\sigma_c^2 + \alpha_l^2\sigma_l^2 + \alpha_r^2\sigma_r^2}}{n_s + f_s}, \\ \sigma_i^2 &\equiv \frac{n_i f_i}{n_i + f_i}, \qquad n_s = n_c - \alpha_l n_l - \alpha_r n_r, \end{aligned} \tag{F.4}$$

where $n_i(f_i)$ is the number of events that pass (fail) the trigger in the window i where i stands for either the central window (c), left (l) or right sideband (l) window, or the central window after sideband subtraction (s); $\alpha_l = \alpha_r = 2/3$ is the ratio of signal (central) window width to that of sidebands; σ_i is the variance of n_i assuming it follows the binomial distribution. Example dimuon invariant mass ($m_{\mu\mu}$) distributions from

the unbiased CMU trigger samples for run range #8 are shown in Figs. 69. The plots are shown for the run range #8 and are made for the muon p_T bins corresponding to the efficiency measurement: red histogram corresponds to the central (signal) mass window, ad the blue filled histograms correspond to the sideband mass regions defined previously in this section.

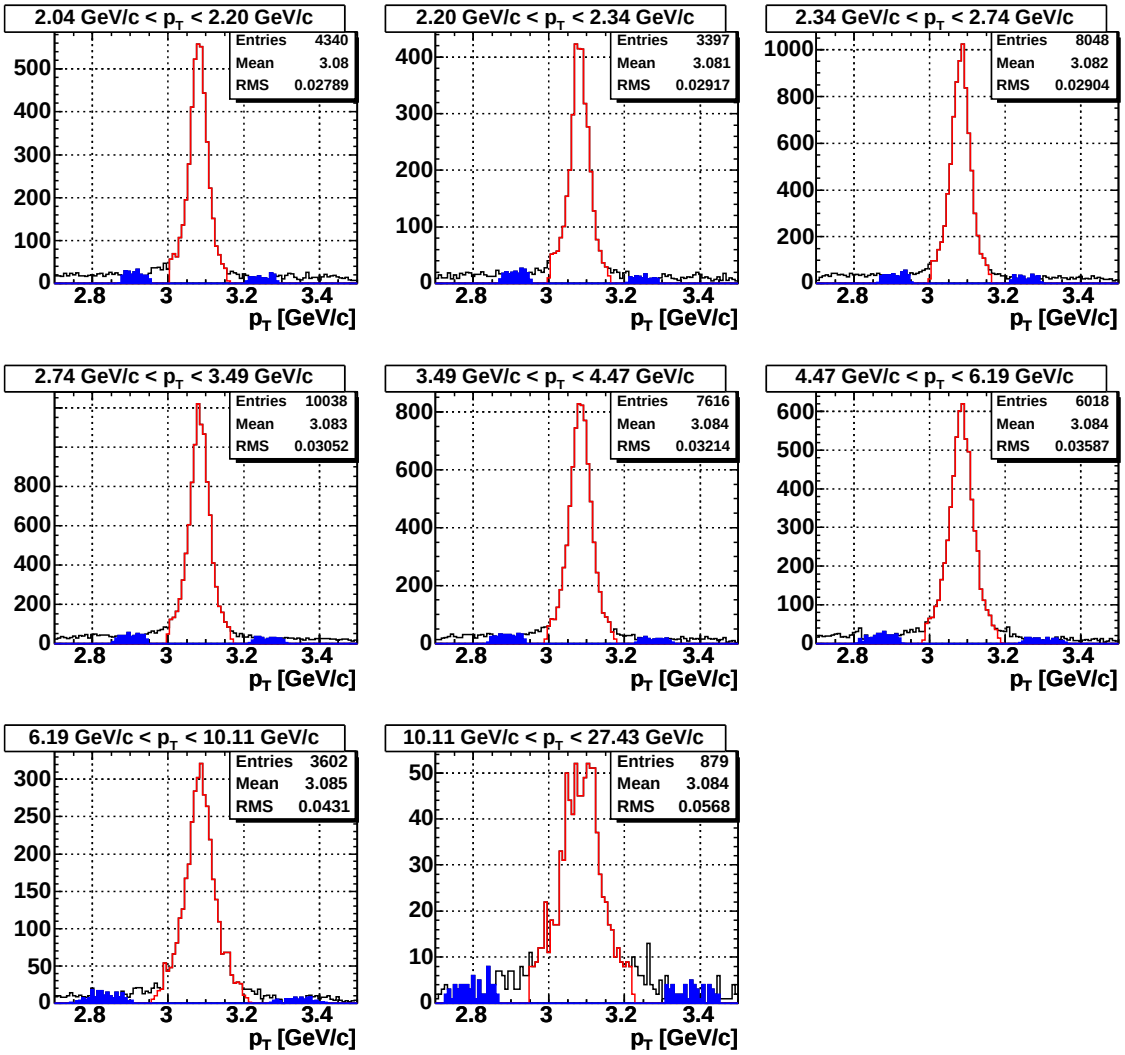

FIG. 69: Dimuon invariant mass distributions for events in the CMU unbiased trigger path.

In each case the efficiency is measured as a function of p_T the shape of the distribution is fitted using the functional form:

$$\epsilon(p_T) = A/\{1 + \exp[B * (p_T + C)]\}. \tag{F.5}$$

The fits are done to the graph: data points corresponding to the histogram binned/filled in $1/p_T$. The data point for each bin is placed at the mean value of $1/p_T$ in the bin.

The choice of the functional form is driven by the facts that this is an analytical function, it performs rather well in the fits, and it has the main features of the observed p_T dependence: monotonous growth and saturation at high p_T (see the high-p_T cross check using $Z \to \mu^+\mu^-$ below). The muon trigger efficiency can be described by slow-changing (with p_T) stub, XFT, and stub-track matching efficiencies and the fast-changing XFT trigger threshold. The selected functional form for $\epsilon(p_T)$ describes well the slow-changing component of the efficiency, but not the fast-changing one. The trigger threshold effects are only noticeable close to the threshold, with XFT resolution $\sigma_{\mathrm{XFT}}(\mathrm{GeV}/c) \approx p_T^2(\mathrm{GeV}/c) \cdot 1.7\%$. The threshold effects are excluded by selecting $p_T^{\mathrm{CMU}} > 1.6$ GeV/c and $p_T^{\mathrm{CMX}} > 2.2$ GeV/c.

The fit procedure is tailored to avoid returning the unphysical values and to improve the fit behavior. The fits are performed in the following way: $A \leq 1$ constraint is set; if $C > 0$ refit is made with $C = 0$; if $B < -10$ refit is made with $B = 10$; in case the fit probability is better when fit to *const* the fit is made with $\epsilon(p_T) = const$.

Special procedure is needed to extract the proper statistical uncertainty on the $\epsilon(p_T)$ due to the following: as the efficiency is constrained from above by 1, use of the parabolic errors and the error matrix is not appropriate, especially for cases when parameters are close to the limit. The procedure begins with the following: the fit parameter space (A, B, C) is scanned to get a set of (2000) points that satisfy $0.975 < |\chi^2(A_i, B_i, C_i) - \chi^2(A_0, B_0, C_0)| < 1.025$, where X_0 corresponds to the parameters that

minimize the χ^2 (fit result), and X_i stand for the points that give χ^2 in the vicinity of 1 from the minimum. The resulting set of points is then used to extract the $\epsilon_{\text{min,max}}(p_T)$ ($\epsilon_{\text{max}} \leq 1$ by construction) which define the 1-σ contour.

Measurement of the `CMU` L1 trigger efficiency for the run ranges described in the Table XVII is then made following the above mentioned prescription and the selections given in the Section F.A.

3. Level-1 CMU trigger efficiency: XFT 2-miss

For the 2-miss XFT configuration the only difference between the 2 run periods (#1 and #2) is the fix of the dimuon matchbox #8. Before the fix, the problem with this matchbox was that it wouldn't set a dimuon (matchbox) bit. Thus if two muons that should pass the trigger would fall into the matchbox #8 (wedges 16 and 17) the dimuon trigger would not fire. Since this effect is properly accounted for by removing such muon combinations from the denominator of the measurement, the efficiency in both ranges is expected to be the same[2] and range #1 and #2 can be combined. The efficiency for ranges #1 and #2 combined is shown in Fig. 70: the data points with statistical only error bars are overlaid by the fit result with central value represented by the blue curve and the magenta curves corresponding to the statistical uncertainty 1σ range as defined in Section F.C.2; the fit results are given in the text box. The efficiency plateau is $\epsilon^{\mu}_{\text{L1 plateau}} = 99.0 \pm 0.1\%$.

The contributions to the (in)efficiency are shown in Fig. 71: the efficiency the online stub matching the XTRP map, $\epsilon^{\mu}_{\text{stub,CMU}}$ (red triangles); the efficiency of the XTRP match, $\epsilon^{\mu}_{\text{XTRP,CMU}}$ (blue circles) and the L1 single muon efficiency itself, $\epsilon^{\mu}_{\text{CMU}}$ (black squares); the difference between the red and black points constitutes the XFT

[2]Can be shown by explicit comparison, as given in Ref. [132].

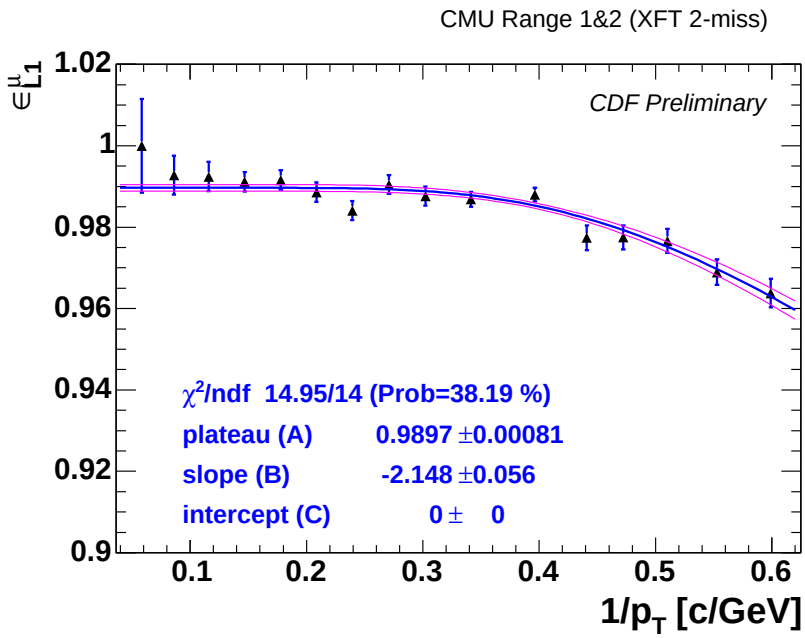

FIG. 70: L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ combined for ranges #1 and #2.

inefficiency. It can be seen from Fig. 71 that here the main source of inefficiency at the low p_T region is the L1 stub selection, the inefficiency source is split between the XFT finding and the L1 stub with roughly the same contributions for momenta $\gtrsim 5$ GeV/c. The XTRP stub to track matching is flat in p_T and is $\gtrsim 99.8\%$.

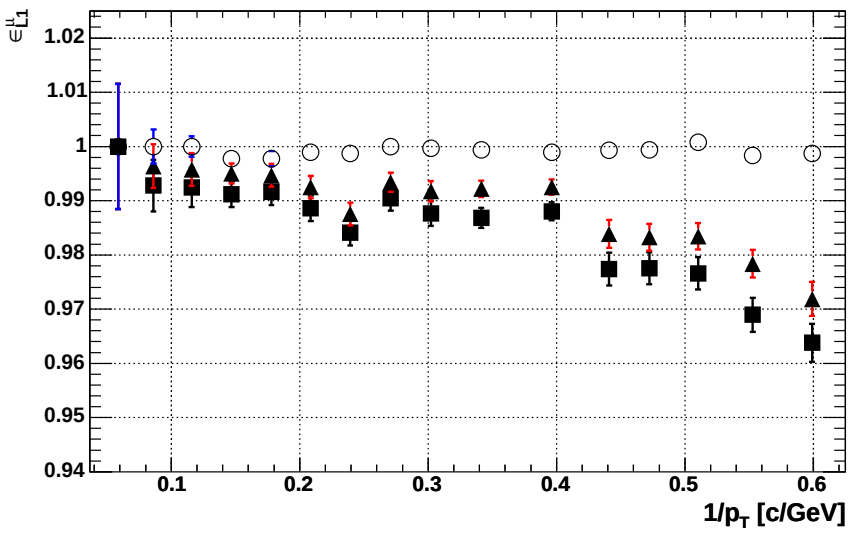

FIG. 71: L1 CMU source of inefficiency for run ranges #1 and #2 combined: $\epsilon^{\mu}_{\mathrm{CMU}}$ (■), $\epsilon^{\mu}_{\mathrm{stub,CMU}}$ (▲), and $\epsilon^{\mu}_{\mathrm{XTRP,CMU}}$ (○).

4. Level-1 CMX trigger efficiency: XFT 2-miss

The L1 CMX trigger efficiency during XFT 2-miss configuration is provided for run range #2, since the `CMX` good runs start with the run range #2 as defined in Table XVII. The single muon L1 CMX efficiency is shown in Fig. 72: the presentation format is the same as in Fig. 70 described in Section F.C.3. The efficiency plateau is $\epsilon^{\mu}_{\mathrm{L1\ plateau}} = 99.5 \pm 0.9\%$.

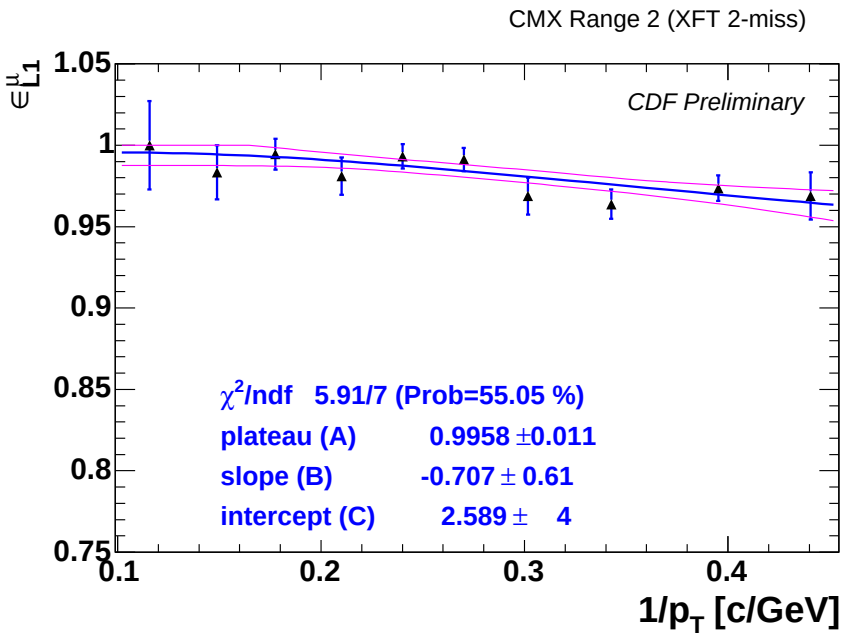

FIG. 72: L1 $\epsilon^{\mu}_{\mathrm{CMX}}$ (XFT 2-miss).

The inefficiency composition can be inferred from Fig. 73 which displays the following distributions (from bottom to top): the total L1 single muon efficiency, $\epsilon^{\mu}_{\mathrm{CMX}}$; the efficiency to have the online stub that matches the XTRP, $\epsilon^{\mu}_{\mathrm{stub,CMX}}$; the efficiency of the XTRP match, $\epsilon^{\mu}_{\mathrm{XTRP,CMX}}$. The difference between $\epsilon^{\mu}_{\mathrm{CMX}}$ and $\epsilon^{\mu}_{\mathrm{stub,CMX}}$ is due to the XFT inefficiencies. In this case the L1 stub selection dominates the inefficiency, the XFT efficiency is indistinguishable from 100%.

5. Level-1 CMU trigger efficiency: XFT 1-miss

The XFT was configured to operate using the 1-miss design from run range #3 onward. By allowing only 1 missed hit in each of the COT axial super-layers the XFT efficiency has been degraded. After changing to this configuration the trigger efficiency has developed an η-dependence, increasing with increasing $|\eta|$. Naturally the tracks with higher pseudorapidity have a longer path length in the COT thus

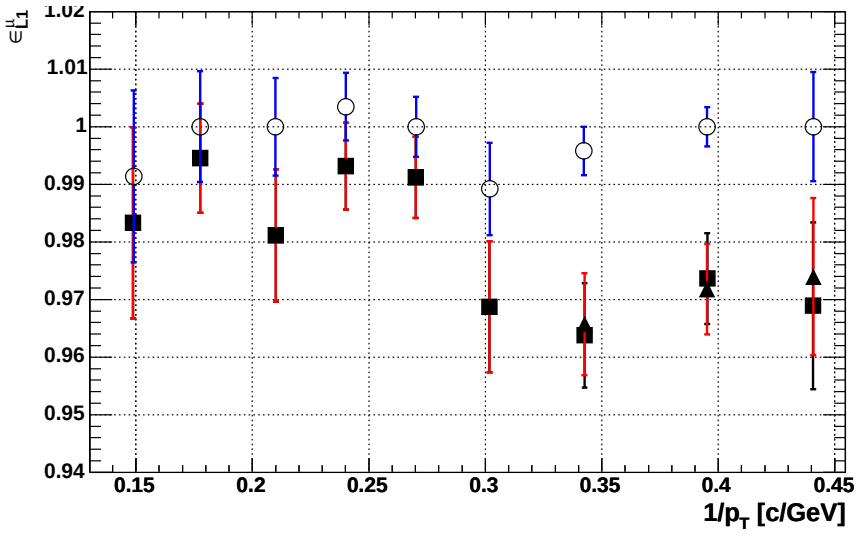

FIG. 73: L1 CMX source of inefficiency for XFT 2-miss period: $\epsilon^{\mu}_{\text{stub,CMX}}$ (▲), $\epsilon^{\mu}_{\text{XTRP,CMX}}$ (○), and $\epsilon^{\mu}_{\text{CMX}}$ (■).

producing more ionization per cell which results in a higher hit (and then XFT) efficiency. The COT thresholds were lowered during the Fall 2003 shutdown in attempt to remove this dependence, which was quite successful.

In addition to the effect introduced by switching to the 1-miss XFT configuration starting in early Summer 2003 (range #3 and later) the bottom portion of the COT had started to show an increasing hit efficiency (gain) loss (starting with range #4) which was later attributed to the drift chamber aging. The efficiency loss rate was progressing until the shutdown in September 2003 (range #5 had the worst efficiency). The COT had partially recovered during the 2003 shutdown, this and the effect of the lowered thresholds had brought the XFT efficiency back up in the Winter 2003/2004 data corresponding to range #6. For the reason that the COT had continued to age rapidly, to keep the detector alive it was decided to run the COT in a compromised state with lowered gain/turned off inner superlayers starting

February 13, 2004. Later on the solution to revive the chamber had come, first reducing (mid-April, corresponding to range #7) and then to reversing the aging until the full recovery (mid-June, corresponding to range #8). The data collected in the COT compromised state is not used in the present measurement. The data collected starting from mid-summer 2003 through mid-June has a significant ϕ-dependence.

The pseudorapidity dependence of the efficiency in bins of p_T that illustrates the above mentioned behavior related to aging and recovery while running in 1-miss configuration is shown in Figs. 74 for the top ϕ bin and in Fig. 75 for the bottom ϕ bin. It can be seen that the inefficiency is the largest for the range #5 in the bottom ϕ bin for the low-p_T low-$|\eta|$ region. In contrast, the efficiency is high and is relatively flat for the range #8.

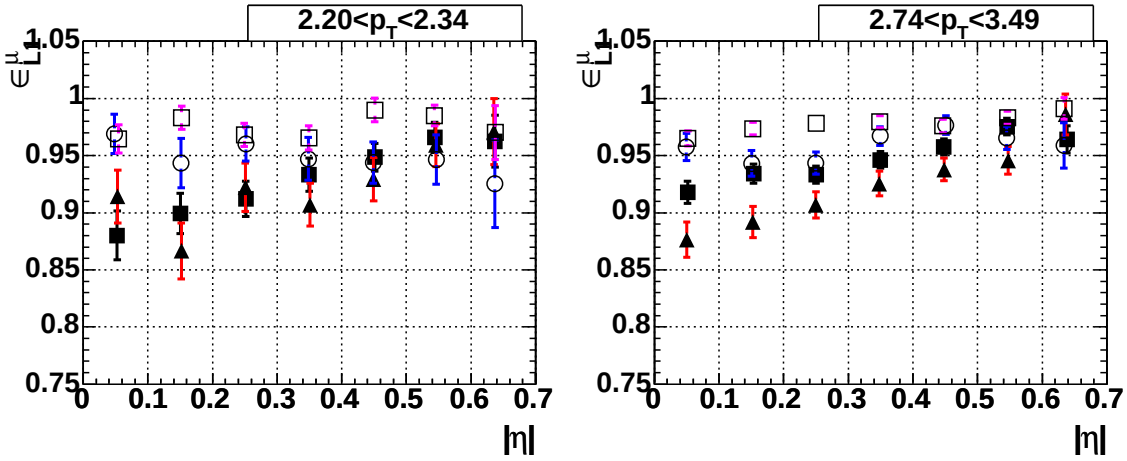

FIG. 74: L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ (XFT 1-miss) vs. muon $|\eta|$ for two low p_T bins for the top ϕ bin for ranges #3 (■), #5 (▲), #6 (○), and #8 (□).

The efficiency dependence as a function of ϕ attributed to the COT aging problem and recovery is shown in Fig. 76. The data points correspondence to the run ranges is the same as in Figs. 74 and 75, covering the momentum range of $p_T > 2$ GeV/c only.

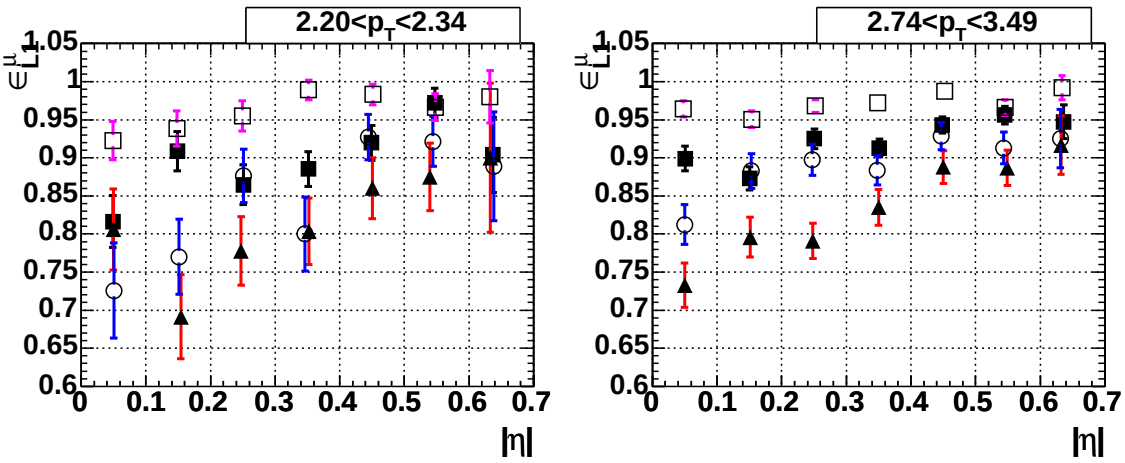

FIG. 75: L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ (XFT 1-miss) vs. muon $|\eta|$ for two low-p_T bins for the bottom ϕ bin for ranges #3 (■), #5 (▲), #6 (○), and #8 (□).

The green vertical lines mark the partitions between the top and bottom ϕ bins. It can be seen from Fig. 76 that the efficiency is relatively flat in ϕ in the top ϕ bin, which justifies the choice of binning in ϕ.

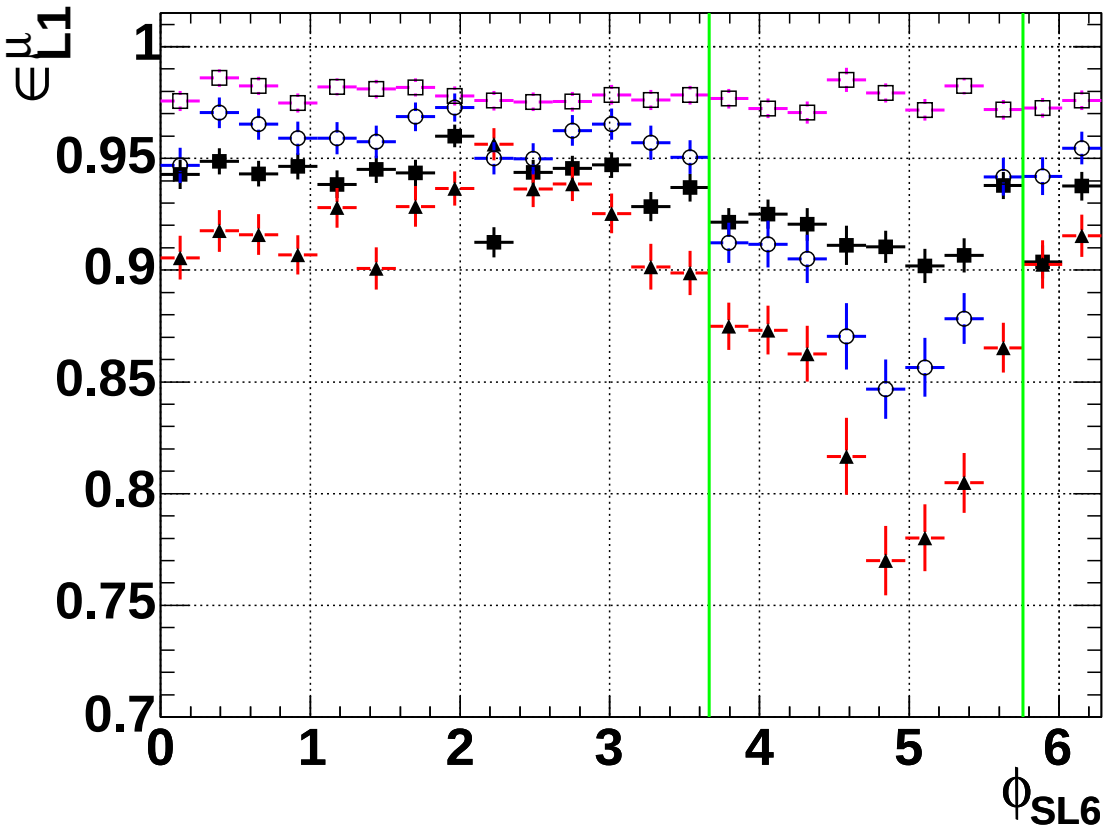

FIG. 76: L1 $\epsilon^{\mu}_{\mathrm{CMU}}$ (XFT 1-miss) vs. muon ϕ_{SL6} for ranges #3 (■), #5 (▲), #6 (○), and #8 (□).

Based on the variations observed in the efficiency behavior the efficiency is measured integrated over ϕ in ranges #3 and #8 and binned in ϕ for the ranges #4-#7 as $\sin\phi_{SL6} > -0.5$ (top), $\sin\phi_{SL6} < -0.5$ (bottom). For each run range and ϕ bin the efficiency is described as a function of p_T and $|\eta|$. The event selection follows the prescription given in Section F.A.

The efficiency is parameterized in 8 muon p_T bins and 7 $|\eta|$ bins. The efficiency value and the statistical error for the individual bins can be found in Ref. [132]. A parameterized functional dependence is used along the p_T direction and an interpolation is used along the $|\eta|$ direction. The efficiency vs. p_T^{-1} distributions for the seven $|\eta|$ slices for a sample cases of range #3 are shown in Fig. 77.

The shapes of the distributions are fitted using the prescription given in Section F.C.2. The results of the fits for each run range $|\eta|$ slice and ϕ bin are combined for easy access in the form of a C++ function. made available in public access. the user simply needs to supply the track p_T, η, ϕ_{SL6} and run number. The function will return the estimated efficiency and its statistical and systematic uncertainty. The uncertainties are obtained by propagating the errors in two adjacent $|\eta|$ slices as used in the linear interpolation formula. All the details can be found in Ref. [132].

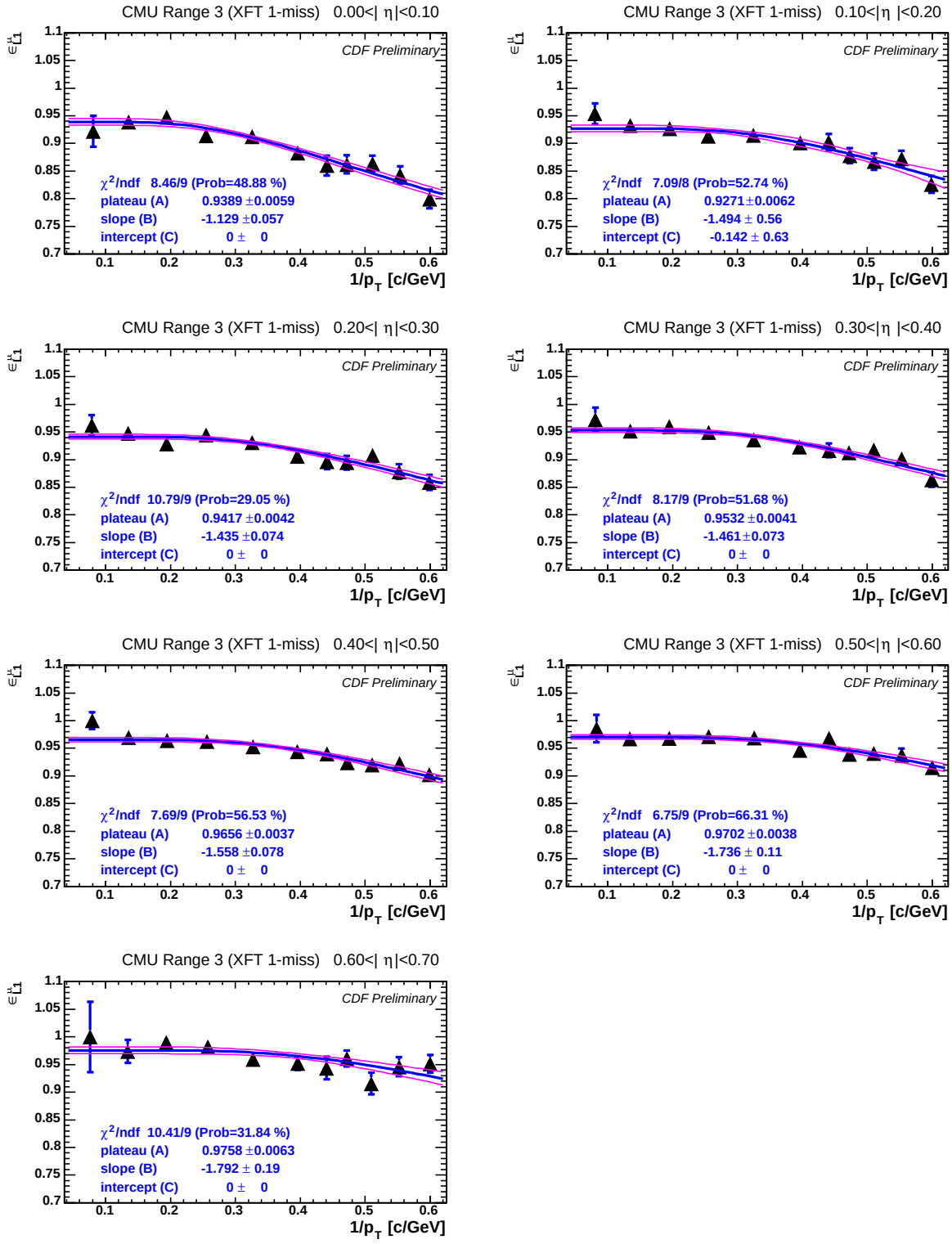

FIG. 77: L1 $\epsilon^{\mu}_{\rm CMU}$ (XFT 1-miss) vs. muon p_T^{-1} in slices of $|\eta|$ for run range #3.

The sources of inefficiency in L1 CMU trigger for a sample run range #4 are shown in Fig. 78: the difference between the $\epsilon^{\mu}_{\mathrm{CMU}}$ and the $\epsilon^{\mu}_{\mathrm{stub,CMU}}$ is described by the XFT track finding inefficiency. It is apparent that the XFT is the dominant source of inefficiency in the XFT 1-miss configuration.

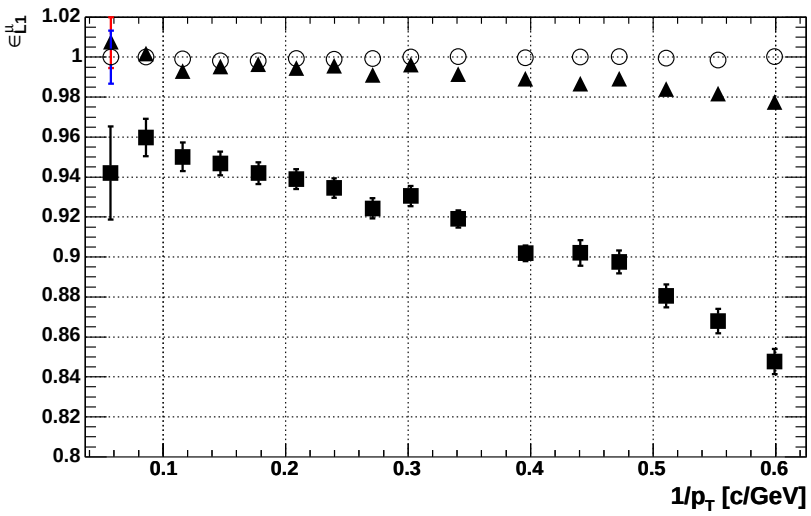

FIG. 78: L1 CMU source of inefficiency in XFT 1-miss period for run range #4 averaged over ϕ: $\epsilon^{\mu}_{\mathrm{CMU}}$ (■), $\epsilon^{\mu}_{\mathrm{stub,CMU}}$ (▲), and $\epsilon^{\mu}_{\mathrm{XTRP,CMU}}$ (○).

6. Level-1 CMX trigger efficiency: XFT 1-miss

The L1 CMX trigger efficiency measurement in XFT 1-miss configuration is made for the runs starting from range #3 and corresponding to the trigger configuration that requires a CSX hit matching to a CMX stub at L1. The miniskirt wedges (CMX wedges 15 through 20) were instrumented only during the Fall 2003 shutdown, and the trigger for the miniskirt was implemented in early May 2004 (starting with run #7) but due to the hardware reasons the trigger was not effective after that. For this reason the miniskirt wedges are not included into the efficiency measurement.

The CMX trigger efficiency was practically not affected by the COT aging prob-

lem and had no significant ϕ-dependence. This can be explained by two following factors. First, the XFT inefficiency effect is small mostly due to the larger (compared to central $|\eta| < 0.5$ region) pathlength in COT. And second, the azimuthal region that was affected the most by the COT aging was not (properly) instrumented, which is the bottom 90°-section of CMX (miniskirt). Thus, it was chosen to combine ranges #3-5 as well as the ranges #6 and #7 and to integrate over ϕ. Similar to the case with CMU, the efficiency is parameterized as a function of p_T and $|\eta|$: the sample is binned in 5 $|\eta|$ and 7 p_T bins. The resulting efficiency is implemented in the form of C++ function. The $\epsilon^{\mu}_{\mathrm{CMX}}$ values in each bin and the fit results for each run range and $|\eta|$ slice are available in Ref. [132].

The sources of CMX inefficiency for the XFT 1-miss period are shown in Fig. 79: the efficiency to have the online stub that matches the XTRP, $\epsilon^{\mu}_{\mathrm{stub,CMX}}$; the efficiency to attach a CSX hit and have the XTRP match having a stub and an XFT track, $\epsilon^{\mu}_{\mathrm{CSX,CMX}}$; and the efficiency of the XTRP match, $\epsilon^{\mu}_{\mathrm{XTRP,CMX}}$. Again, the XFT inefficiency is the difference between $\epsilon^{\mu}_{\mathrm{stub,CMX}}$ and $\epsilon^{\mu}_{\mathrm{CMX}}$. In this case the L1 trigger inefficiency is dominated by the stub and scintillator hit inefficiency and the XFT inefficiency is rather small.

7. *Systematics*

To estimate the accuracy of the L1 trigger efficiency measurement several consistency cross-checks are done first. These include the measurement of the efficiency using a sample not biased by the L3 decision, collected from the `MUON_CMUP8`; a cross-check of the efficiency value at high-p_T based on the $Z \to \mu^+\mu^-$ collected from the `MUON_CMUP18` sample; and a cross-check based on the measurements using tracks with silicon hits. The results for the first two are consistent with the main measurement within the large statistical uncertainty of the cross-check samples. The result for the

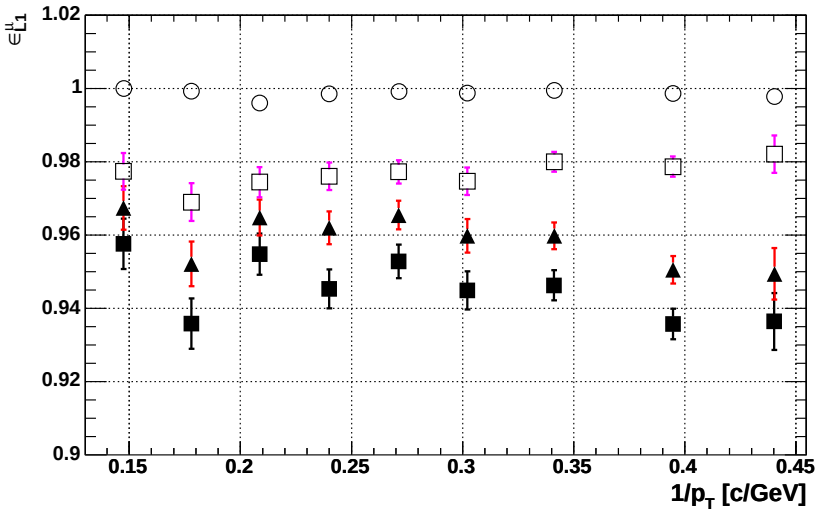

FIG. 79: L1 CMX source of inefficiency for XFT 1-miss period (ranges #3-#5 combined only): $\epsilon^{\mu}_{\text{stub,CMX}}$ (▲), $\epsilon^{\mu}_{\text{XTRP,CMX}}$ (○), $\epsilon^{\mu}_{\text{CSX,CMX}}$ (□), and $\epsilon^{\mu}_{\text{CMX}}$ (■).

latter is practically indistinguishable from the main measurement.

The systematic uncertainty of the measurement is estimated from the analysis of the efficiency change as a function of η and I. The dependence on η is considered to account for the possible differences between the efficiency for the positive and negative values of η at the same absolute value (east-west asymmetry). To account for that the estimate is made of the east-west asymmetry for each run range and ϕ bin averaged over p_T. Based on the observed asymmetry a systematic uncertainty is assigned to each of the $|\eta|$ bins. The contributions from systematics due to η and I dependence are combined and reported as a final systematic uncertainty on the measurement. Depending on the performance of the trigger, the systematic uncertainty ranges from about 0.5% up to 4%. The total statistical and systematic uncertainties can be obtained using the efficiency C++ functions.

D. Level-2 dimuon trigger efficiency ($\epsilon_{\mathbf{L2}}$)

The L2 muon trigger was implemented for the runs starting with range #7 on May 18, 2004. The efficiency is measured for the following triggers listed in Appendix E:

- `L2_CMU1.5_PT1.5_&_CMX1.5_PT2_DPHI120_OPPQ`,
- `L2_TWO_CMU1.5_PT1.5_DPHI120_OPPQ`,
- `L2_TWO_CMU1.5_PT2_DPHI120_OPPQ`.

There are a number of trigger paths that can be used to extract the L2 trigger efficiency. The several dimuon paths that are unbiased from the L2 trigger decision are

- `JPSI_CMUCMU1.5_NOL2` and `JPSI_CMU1.5_CMX2_NOL2`,
- `JPSI_CMUP4_CMU` and `JPSI_CMUP4_CMX`.

The first two are using the low-p_T dimuon L1 triggers as input, the last two are using the `CMUP4` L1 trigger as input. The first two triggers were superior in statistics (cross sections larger than those of CMUP4 by a factor of 3) for the period that corresponds to run range #7, as these were using the input dimuon L1 trigger prescaled by a factor of 10. For the run range #8 the input to these triggers had changed to have a prescale 100. This change in the prescale had made the `CMUP4` based triggers superior in statistics, the `CMUP4` trigger had been dynamically prescaled by up to a factor of ten since 2003. In addition the `CMUP4` based J/ψ's have larger momenta which make these dimuons more similar to the $B_s^0 \to \mu^+\mu^-$. The `CMUP4` based trigger paths are used to measure the L2 dimuon efficiency, the L1 dimuon based trigger paths are used for cross-checks and to infer the systematic uncertainties.

As mentioned in the Section F.A, the denominator for the L2 efficiency measurement includes the L2 muon fiducial requirement. Although in the L2 the prescription

for the XTRP matching is used, because of the finer ϕ binning the matching window is in fact tighter. Since the lookup table is a well-defined function it is easier to apply the L2 fiduciality requirement to the denominator (hence, to the acceptance), rather than trying to extract the proper kinematic dependence of the L2 matching effect from the data. Approximately 0.4(2.6)% of $J/\psi \to \mu^+\mu^-$ events are removed after applying this requirement to the `JPSI_CMUP4_CMU(X)` sample. Compare this with approximately 0.2(1.7)% increase in dimuon efficiency for two-CMU(CMU-CMX).[3]

The L2 dimuon trigger efficiency is measured averaged (as a constant) over the unbiased sample. Most of the inefficiency comes from the L2 stub to XFT matching window at low-p_T and from the XFT charge misidentification at high-p_T. The resulting L2 dimuon trigger efficiencies are summarized in Table XVIII. All efficiencies, except the one for `L2_CMU_CMX` are measured using CMUP4 trigger paths; `JPSI_CMU_CMX_NOL2` was used for the latter measurement to include the low momentum range on both muons. The systematic uncertainty is conservatively taken to be half the observed variation in p_T.

TABLE XVIII: L2 trigger efficiencies with statistical and systematic uncertainties included. Efficiency for `L2_TWO_CMU_2G` is applicable to run range #8 only.

Trigger	$\epsilon \pm stat \pm syst$, %
`L2_TWO_CMU_1G5`	$99.89 \pm 0.02 \pm 0.6$
`L2_TWO_CMU_2G`	$99.87 \pm 0.03 \pm 0.5$
`L2_CMU_CMX`	$99.51 \pm 0.07 \pm 0.6$

[3]The CMU muons here are with $p_T > 1.6$ GeV/c, the effect is even smaller for $p_T^{\mathrm{CMU}} > 2$ GeV/c.

E. Level-3 dimuon trigger efficiency ($\epsilon_{\mathbf{L3}}$)

The L3 dimuon trigger efficiency is measured for the trigger paths listed in Appendix E. The method is similar to that of the L2 efficiency measurement: a trigger sample of $J/\psi \to \mu^+\mu^-$ dimuon events collected on the trigger path bypassing the L3 requirements defines the denominator of the measurement and the events from that sample that actually pass the given L3 trigger define the numerator. The potential sources of systematic bias are determined from the trigger efficiency dependence on various kinematic variables whose distributions are different between the $J/\psi \to \mu^+\mu^-$used to estimate the trigger efficiency and the $B_s^0 \to \mu^+\mu^-$ events for which it is used.

The unbiased trigger paths used for the measurements are

- `DIMUON_L3PS200_L1_CMU1.5_PT1.5_&_CMX1.5_PT2`,
- `DIMUON_L3PS200_L1_TWO_CMU1.5_PT1.5`.

These trigger paths use the corresponding dimuon L1 triggers as input, and have events auto-accepted/prescaled at L2 and L3. For the `RAREB` trigger paths of interest in this analysis, the L3 requirements are listed in Appendix E (these are used in the denominator of the L3 efficiency measurement and are mentioned in the Section F.A as L3 fiducial requirements). The COT parent of the track is used to calculate the offline quantities since L3 tracking only utilizes COT information. In estimating the efficiency, a side-band subtraction is used following the procedure described in Section F.B.

Different reconstruction versions were used in L3 during the data collection for this analysis, which should affect the L3 trigger efficiency value: run ranges #1-6 correspond to L3 running the version 4 executable, ranges #7 and #8 correspond to running a version 5. Since the offline tracking requirements are loose compared to the

XFT requirements, the L3 efficiency is expected to be largely unaffected by the trigger and COT changes. Thus it was chosen to present the L3 efficiency measurement in 2 combined ranges: ranges #1-6 and #7,8.

The resulting L3 trigger efficiencies are summarized in Table XIX, where both statistical and systematic uncertainties are included. Note that efficiencies in range #7,8 are systematically higher. This is partly due to the better reconstruction efficiency in the version 5 executable, corresponding to approximately 1% of the difference. The values of the systematic uncertainties are determined from the analysis of the efficiency variations in four kinematic variables: dimuon opening angle ($\Delta\phi$), dimuon transverse momentum ($p_T^{\mu^+\mu^-}$), dimuon pseudorapidity ($\eta^{\mu^+\mu^-}$), and dimuon isolation ratio (I). The most significant deviations in each of the variable are considered as the possible sources of systematic uncertainty, with the largest reported as the systematic uncertainty given in Table XIX. This method is considered a conservative one.

TABLE XIX: L3 efficiencies with statistical and systematic uncertainties ($\epsilon \pm \delta_{\text{stat}} \pm \delta_{\text{syst}}$), given in per cent.

Trigger \ Run range	runs 138425-175096	runs 182843-186598
`RAREB_CMUCMU_SUMPT`	$99.1 \pm 0.4 \pm 0.8$	$99.8 \pm 0.8 \pm 0.8$
`RAREB_CMUCMX_SUMPT`	$98.0 \pm 0.6 \pm 1.7$	$100.2 \pm 1.2 \pm 1.7$
`RAREB_CMUPCMU`	$98.6 \pm 0.5 \pm 0.9$	$100.1 \pm 0.8 \pm 0.9$
`RAREB_CMUPCMX`	$97.6 \pm 0.7 \pm 1.9$	$100.2 \pm 1.1 \pm 1.9$

F. Cross-check using $J/\psi \to \mu^+\mu^-$ yields

The estimation of $J/\psi \to \mu^+\mu^-$ signal cross-section (yield per unit of luminosity) as a function of run range is used to cross-check the consistency of the measurement of the total trigger efficiency made in run ranges for the CMU-CMU and CMU-CMX dimuon trigger samples. The $J/\psi \to \mu^+\mu^-$ production cross-section is expected to be independent of the run number. The expected number of signal events in the given run range is given by

$$N_S = \mathcal{L}\sigma\epsilon_{\rm reco}\epsilon_{\rm sel}\epsilon_{\rm trig}\alpha,$$

where $\mathcal{L}$ is the integrated luminosity, σ is the signal cross-section, $\epsilon_{\rm reco}$, $\epsilon_{\rm sel}$, and $\epsilon_{\rm trig}$ are the reconstruction, event selection, and trigger efficiencies; and α is the detector acceptance. The observable $J/\psi \to \mu^+\mu^-$ cross-section is then $\sigma_{\rm obs} \equiv N_S/\mathcal{L}$. It is expected that the run dependence of $\sigma_{\rm obs}$ is dominated by the effect of $\epsilon_{\rm trig}$ changes, thus the value $\sigma_{\rm obs}/\epsilon_{\rm trig}$ should have a small run dependence. This allows to check the relative quality of measuring the trigger efficiency as a function of run range.

The events collected on the `JPSI_CMUCMU1.5` and `JPSI_CMU1.5_CMX2` trigger paths are used to check the run-dependence of the total trigger efficiency (these are high statistics $J/\psi \to \mu^+\mu^-$ trigger paths with the requirements very similar to those of `RAREB_CMUCMU(X)_SUMPT`, except for 2.7 GeV/c^2 $< m_{\mu^+\mu^-} <$ 4 GeV/c^2 cut and no $p_{T_1} + p_{T_2}$ cut applied). The events are required to pass the L1,2,3, and "offl" requirement. For each run range the number of events is counted, each event weighted by $\epsilon_{\rm trig}^{-1}(x)$, where x are the event kinematic variables and $\epsilon_{\rm trig}$ is obtained from Eq. (F.1). The weighted sum is then divided by the integrated luminosity per run range. The result is the desired $\sigma_{\rm obs}/\epsilon_{\rm trig}$. The run dependence of $\sigma_{\rm obs}$ and $\sigma_{\rm obs}/\epsilon_{\rm trig}$ is shown in Figs. 80 and 81.

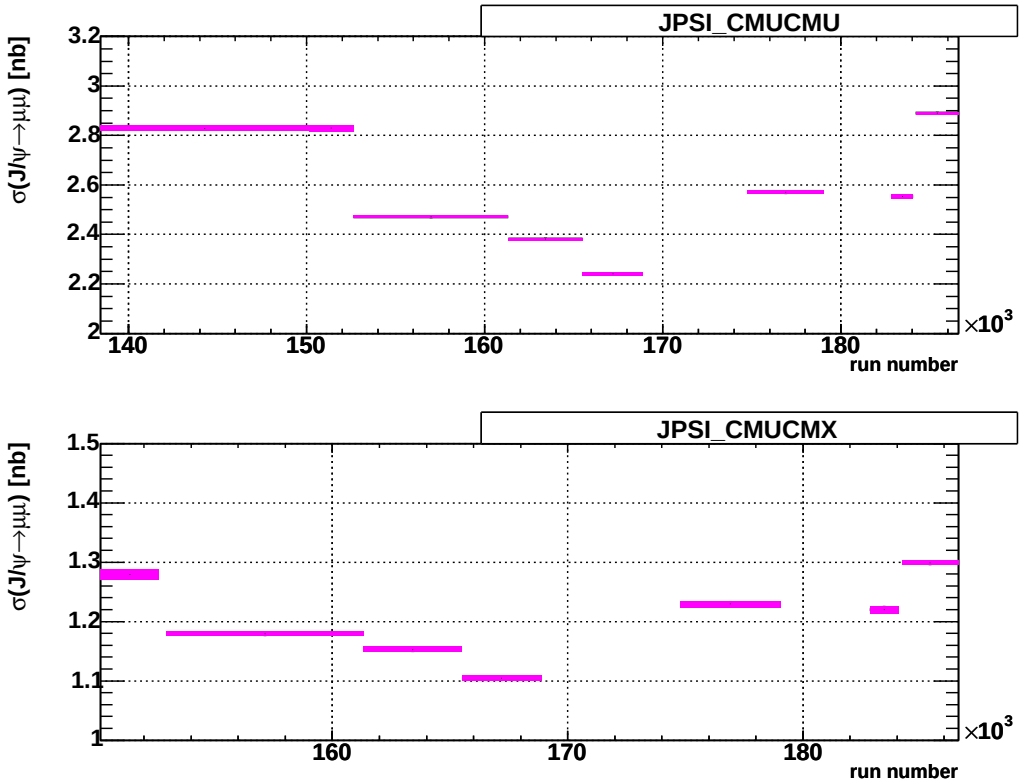

FIG. 80: Raw $J/\psi \rightarrow \mu^+\mu^-$ yield vs. run number in binned by run ranges given in Table XVII as collected on `JPSI_CMUCMU` and `JPSI_CMUCMX` trigger paths. The magenta error bars represent the statistical uncertainty of the $J/\psi \rightarrow \mu^+\mu^-$ sample.

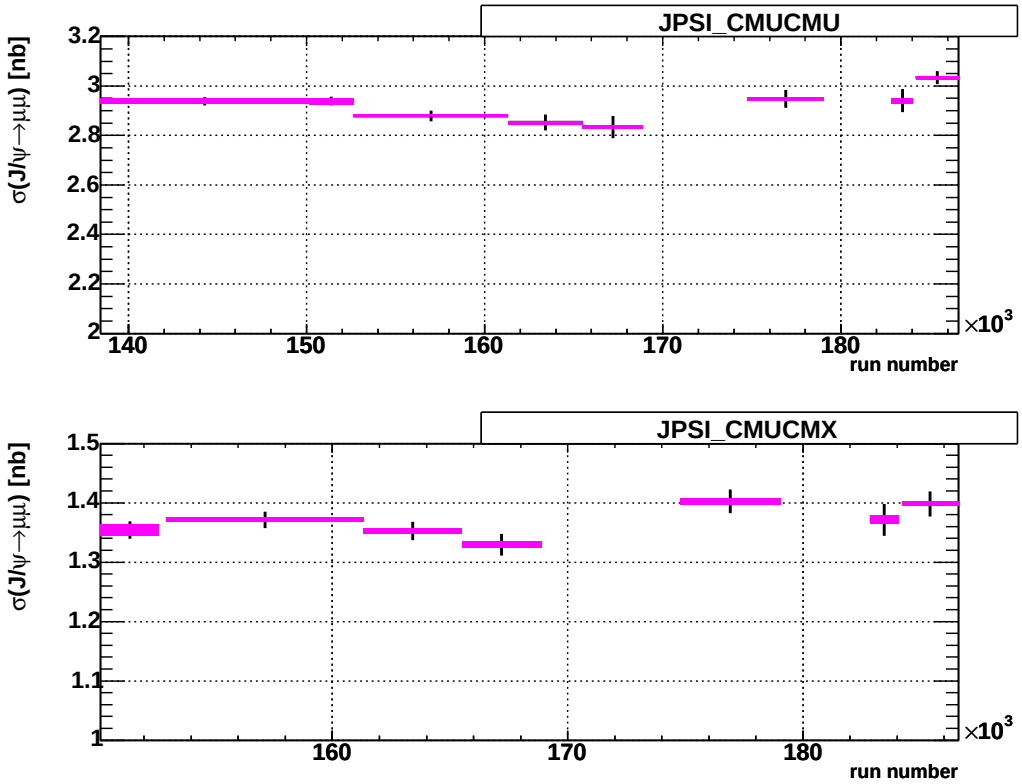

FIG. 81: Yield of $J/\psi \rightarrow \mu^+\mu^-$ scaled by $1/\epsilon_{\rm trig}$ vs run number as collected on `JPSI_CMUCMU` and `JPSI_CMUCMX` trigger paths. The magenta (black) error bars represent the statistical uncertainty of the $J/\psi \rightarrow \mu^+\mu^-$ sample (uncertainty of $1/\epsilon_{\rm trig}$).

The resulting trigger efficiency-corrected $J/\psi \rightarrow \mu^+\mu^-$ yield is within $\pm 3.5\%$ from a constant (compare this to up to 25% variations of the uncorrected yields). It is systematically lower for the period of aged COT which is expected to correspond to lower $\epsilon_{\rm reco}$ and $\epsilon_{\rm sel}$. Here no attempt is made to correct for the reconstruction and event selection efficiencies. Note that the lower $\sigma_{\rm obs}/\epsilon_{\rm trig}$ bins also correspond to the larger statistical and systematic uncertainty in the L1 trigger efficiency measurement that cover the observed deviation of $\sigma_{\rm obs}/\epsilon_{\rm trig}$ from a constant.

G. Summary

The low-p_T trigger efficiencies were measured for the dimuon trigger paths. Following the L1 trigger and detector conditions the L1 efficiency is measured in up to 8 run ranges and 2 ϕ bins and is parameterized as a function of p_T and $|\eta|$. The resulting L1 trigger efficiency is presented in the form of a C++ function.

For the runs starting with range #7 the L2 dimuon trigger efficiency is measured. Provided the L2 fiducial requirement is imposed on the muons the L2 dimuon trigger efficiency has no significant dependencies. The L2 efficiency averaged over all variables is $99.44 \pm 0.5\%$ for CMU-CMX dimuon trigger and $99.88 \pm 0.5\%$ for the CMU-CMU dimuon trigger.

The L3 dimuon trigger efficiency has been measured for 4 `RAREB` triggers. measured for the two data periods: one for the time of running the version 4 L3 executable, and one for the period of running the version 5. The efficiencies were determined to be $> 99\%$ for CMU-CMU and CMU-CMX triggers and about 98% for the CMUPCMU and CMUPCMX triggers. The L3 efficiencies corresponding to version 4 period are roughly 1% lower.

APPENDIX G

KINEMATIC-DEPENDENT MATCHING CUT FOR LOW-P_T CMX MUONS

A fixed muon matching cut applied to Δx alone has been nominally used for selecting muons during the muon reconstruction.[1] This cut is normally selected to be loose enough to maximize the acceptance for all the muons. The muons undergo multiple scattering in the detector material. Except for the very low p_T when the single scatters matter most the muon scattering is an effect of many scatters at small angles. Due to this fact the scattering distance distribution should be Gaussian to a large extent. As the multiple scattering is mostly the effect of a Coulomb scattering the deflection r.m.s. is roughly proportional to $1/p_T$. This means that the fixed muon matching cut if "loose enough" for the lowest momentum becomes unnecessarily loose for the more energetic muons. This is especially significant when the muon momentum ranges from 1-2 GeV/c to 15 GeV/c and above. Naïvely the matching cut can be smoothly tightened by an order of magnitude for the higher momenta. If made intelligently this cut can exclude a minimal amount of signal and remove a large fraction of fake muons at higher momenta. It is proposed to define a $\chi^2_{r\phi,\mathrm{CMX}} \equiv (\Delta x/\sigma_{\chi_{r\phi,\mathrm{CMX}}})^2$ matching variable for the CMX muons which can be used to define such kinematic-dependent matching (muon quality) cut.

A definition of the $\chi^2_{r\phi,\mathrm{CMX}}$ matching variable and the method of the measurement are specified in Section G.A, followed by the measurement and parameterization of $\sigma_{\chi_{r\phi,\mathrm{CMX}}}$ given in Section G.B. The efficiency and the background rejection power of

[1]The ΔX is the distance, in the r-ϕ view, between the reconstructed stub and the expected position of the muon near the stub, as detailed in Appendix C.B and in Ref. [152].

the $\chi^2_{r\phi,\mathrm{CMX}} < 9$ cut is discussed in Sections G.C and G.D respectively, followed by a short summary provided in Section G.E. A more detailed description of the analysis of the $\chi^2_{r\phi,\mathrm{CMX}}$ matching variable can be found in Ref. [154].

A. Methodology

The width of the ΔX distribution is parameterized as a function of the muon kinematic. Then the pull is constructed using this parameterization, where pull is defined as $\chi_{\mathrm{CMX}} \equiv \Delta x/\sigma_{\chi_{r\phi,\mathrm{CMX}}}$. The pull is shown to follow a Gaussian distribution of width 1. The pull squared thus represents the 1 ndf χ^2_{CMX}. The efficiency of the $\chi^2_{\mathrm{CMX}} < 9$ cut is measured. Also the fraction of background rejected by these cuts applied to the muon objects reconstructed in the $J/\psi \to \mu^+\mu^-$ and $B^0_{s(d)} \to \mu^+\mu^-$ sidebands is estimated.

The event sample consists of the events collected on the `JPSI_CMUCMX` trigger path, covering run ranges #2 to #5 only. The quality requirements are practically the same as in the case of trigger efficiency measurement described in Section F.A, only no χ^2_{CMX} requirement is made in this case.

To remove the contribution from fake muons the measurement of ΔX properties is made using sideband subtraction. To perform the sideband subtraction the following definitions are made: $2.999\ \mathrm{GeV}/c^2 < m_{\mu^+\mu^-} < 3.207\ \mathrm{GeV}/c^2$ to be the $J/\psi \to \mu^+\mu^-$ signal mass window and $2.7\ \mathrm{GeV}/c^2 < m_{\mu^+\mu^-} < 2.856\ \mathrm{GeV}/c^2$ and $3.350\ \mathrm{GeV}/c^2 < m_{\mu^+\mu^-} < 3.506\ \mathrm{GeV}/c^2$ to be the left and right sideband mass windows.

B. Parameterization of $\sigma_{\Delta X}$

The ΔX distribution is analyzed as a function of p_T and η. To do that the 2D (3D) histograms of ΔX vs one (two) of the kinematic variables are filled for both signal window and the sidebands. After that the $J/\psi \to \mu^+\mu^-$ sideband subtracted histograms are created. Gaussian fits of the ΔX distribution are then performed in each kinematic bin. The result of this fitting procedure is the binned $\sigma_{\Delta X}$ (Gaussian fit parameters) as a function of 1 or 2 kinematic variables. This $\sigma_{\Delta X}$ dependence can then be parametrized by a smooth function.

The dependence of p_T and η is significant. The η dependence can be explained by the difference in the amount of material traversed by a muon at different η.

It was chosen to parametrize the $\sigma_{\Delta X}$ as a polynomial: second order in $1/p_T$ and first order in $(|\eta| - 0.8)$. To improve the fit quality the fit is done to the binned $\sigma_{\Delta X}(p_T, \eta) \times p_T$ distribution because this greatly reduces binning effects. The following polynomial was used in the 2-dimensional fit:

$$\sigma_{\Delta X} \times p_T = p_T \left(p_0 + p_1 \frac{1}{p_T} + p_2 \frac{1}{p_T^2} + p_3 \frac{|\eta| - 0.8}{p_T} + p_4 \frac{|\eta| - 0.8}{p_T^2} \right). \tag{G.1}$$

Here the parameter p_0 represents the resolution effects and becomes essential at higher p_T, it has been fixed at 0.2 cm as it helps to improve fit quality at higher p_T. Overall, the $1/p_T$ term follows from the multiple scattering model (Rossi formula) with no energy loss [155], and the $1/p_T^2$ term is the leading term for corrections accounting for the energy loss [156]. Finally the $|\eta| - 0.8$ describes the geometrical corrections to the effective path length.

The result of the fit is ($\chi^2/\text{ndf} = 511/218$):

$$\begin{aligned} &p_0 = 0.2, && p_1 = 31.23 \pm 0.36, && \\ &p_2 = 50.55 \pm 1.13, && p_3 = -36.22 \pm 3.04, && p_4 = -69.13 \pm 9.54. \end{aligned} \tag{G.2}$$

Although the χ^2/ndf of the fit is rather high, the result is considered appropriate following the quality checks discussed below.

The parameterization quality can be checked by measuring the width of the pull distribution. The inclusive χ_{CMX} distribution for muons with $p_T > 3$ GeV/c, as seen in Fig. 82, shows that the non-Gaussian tails are small. The $p_T > 3$ GeV/c is selected here because the χ_{CMX} distribution is trimmed significantly on the edges by the reconstruction cuts for lower p_T as shown in the section below. Without this cut the large statistics at low p_T would make the distribution look more Gaussian. Figure 83 shows that the pull width is essentially flat in p_T and is close to 1. Similarly, Fig. 84 shows no significant deviation from 1 for η dependence.

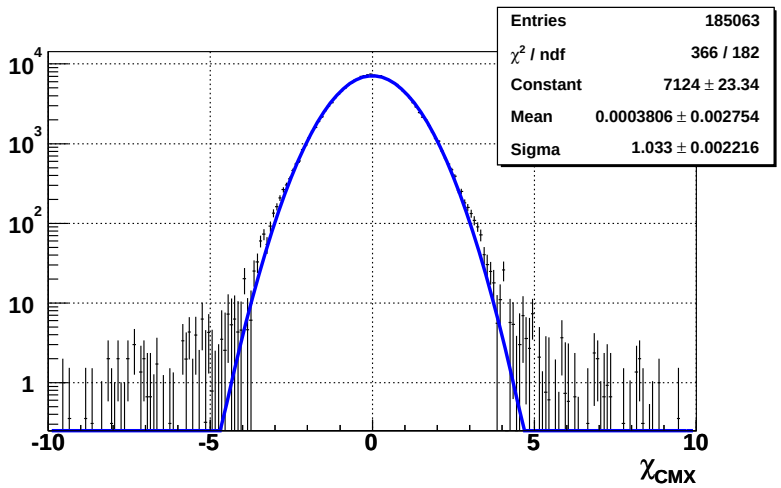

FIG. 82: Distribution of χ_{CMX} for $p_T > 3$ GeV/c.

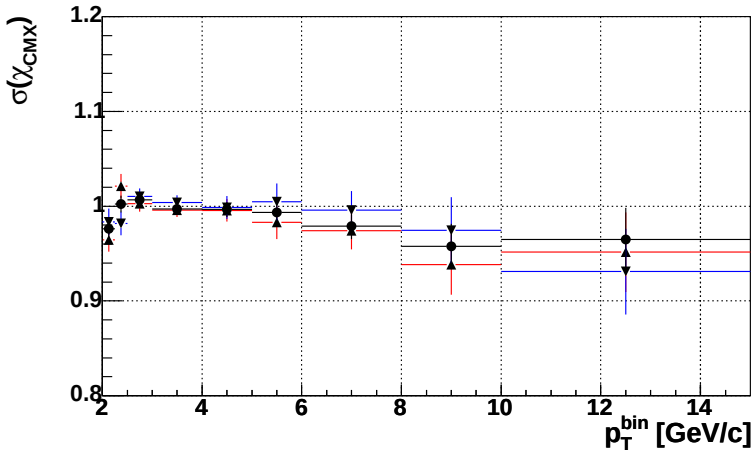

FIG. 83: p_T dependence of the pull width $\sigma_{\chi_{r\phi,\mathrm{CMX}}}$: μ^+ (▲), μ^- (▼), and charge averaged (●).

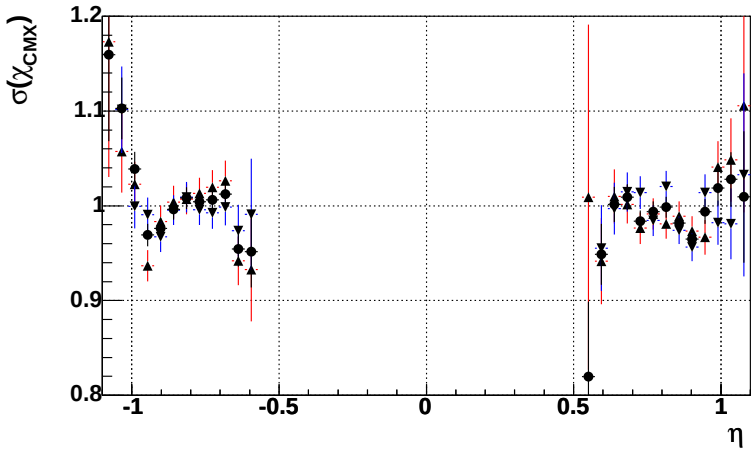

FIG. 84: η dependence of the pull width $\sigma_{\chi_{r\phi,\mathrm{CMX}}}$: μ^+ (▲), μ^- (▼), and charge averaged (●).

C. Efficiency of $\chi^2_{r\phi,\mathrm{CMX}} < 9$ cut

Efficiency of the cut applied to the sideband subtracted $J/\psi \to \mu^+\mu^-$ signal events is measured, $\epsilon(\chi^2_{r\phi,\mathrm{CMX}} < 9)$. The denominator is the number of all the events as described in Section G.A, the numerator is the number of the events passing the

$\chi^2_{r\phi,\mathrm{CMX}} < 9$ cut. The p_T dependence is shown on Fig. 85 and summarized in Table XX.

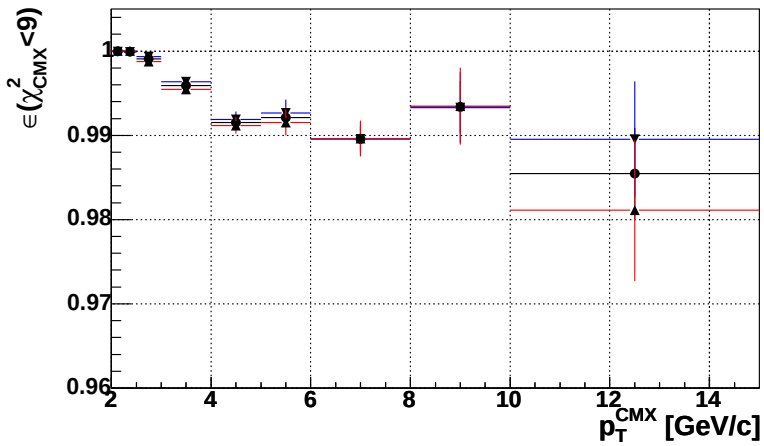

FIG. 85: $\epsilon(\chi^2_{r\phi,\mathrm{CMX}} < 9)$ vs. p_T: μ^+ (▲), μ^- (▼), and charge averaged (•).

The efficiency measurement results are fit to a smooth p_T-dependent function which can be used in the analysis. The efficiency values as measured in the p_T-plateau region are

$$\epsilon(\chi^2_{r\phi,\mathrm{CMX}} < 9 \mid p_T > 4\ \mathrm{GeV}/c) \quad = \quad 99.13 \pm 0.054(stat) \pm 0.25(syst). \qquad \text{(G.3)}$$

Here and in Table XX the systematic uncertainty is estimated from the observed η dependence of the efficiency. Half of the maximal significant difference between 2 bins is taken as a measure of the systematic uncertainty. The systematic uncertainty for bins with $p_T > 4$ GeV/c is taken to be the same for each bin and is taken from Eq. (G.3). The large uncertainties for the low-p_T bins are due to the large efficiency variations with η. The total uncertainty is dominated by the systematic component.

TABLE XX: $\chi^2_{r\phi,\mathrm{CMX}} < 9$ cut efficiency summary.

p_T range, GeV/c	$\langle p_T \rangle$, GeV/c	$\epsilon \pm \sigma_{stat} \pm \sigma_{syst}(\%)$
2-2.25	2.13	$99.9978 \pm 0.0065 \pm 0.0110$
2.25-2.5	2.37	$99.994 \pm 0.012 \pm 0.056$
2.5-3	2.73	$99.91 \pm 0.02 \pm 0.24$
3-4	3.42	$99.59 \pm 0.03 \pm 0.27$
4-5	4.43	$99.15 \pm 0.07 \pm 0.25$
5-6	5.43	$99.21 \pm 0.11 \pm 0.25$
6-8	6.78	$98.96 \pm 0.15 \pm 0.25$
8-10	8.81	$99.34 \pm 0.31 \pm 0.25$
10-15	11.6	$98.55 \pm 0.54 \pm 0.25$

It should be noted that the efficiencies plateau at the level lower than naïvely expected for the $\chi^2_{r\phi,\mathrm{CMX}}$ strictly following the χ^2 distribution with 1 ndf. This can be explained by the fact that the χ_{CMX} is not strictly Gaussian and has the non-Gaussian tails. The 0.6% difference from the strict χ^2-expected value gives the estimate of the relative size of the tails.

The efficiency dependence at low p_T can be explained by the muon matching cuts applied in the Level-1 trigger and in the offline muon reconstruction.[2] Up to the level of $p_T \sim 4$ GeV/c the reconstruction cuts are close to or even tighter than the considered $\chi^2_{r\phi,\mathrm{CMX}}$ cuts. This eventually leads to the measured efficiency increase at $p_T \lesssim 4$ GeV/c.

It is expected that the effect of the trigger requirements on the efficiency mea-

[2]Level-3 and offline CMX matching cuts are the same $\Delta X < 50$ cm.

sured here is negligible because the trigger cuts are looser than the reconstruction cuts at $p_T \lesssim 4$ GeV/c, and because the trigger cuts are looser than the considered $\chi^2_{r\phi,\mathrm{CMX}}$ cuts at the higher momenta. This statement is supported by a measurement based on the sample with CMX muons not biased by the trigger.

D. Background rejection

The fraction of background (non-muon candidates) rejected by the $\chi^2_{r\phi,\mathrm{CMX}} < 9$ cut, $R(\chi^2_{r\phi,\mathrm{CMX}} < 9)$, is estimated using $J/\psi \to \mu^+\mu^-$ signal mass sidebands.[3] The estimates using the $J/\psi \to \mu^+\mu^-$ sidebands defined in Section G.A are shown in Fig. 86. In addition the estimates are made using the $B^0_{s(d)} \to \mu^+\mu^-$ sidebands, which show a behavior similar to that of Fig. 86. Table XXI shows the results in bins of p_T for both sets of estimates.

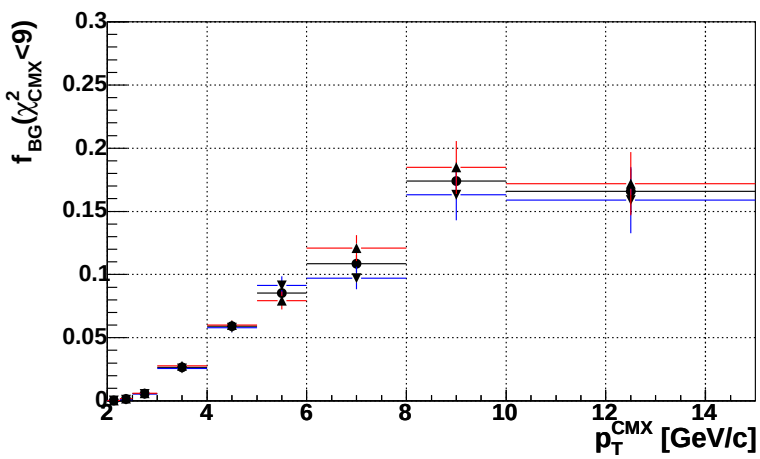

FIG. 86: $R(\chi^2_{r\phi,\mathrm{CMX}} < 9)$ vs. p_T for the $J/\psi \to \mu^+\mu^-$ sidebands: μ^+ (▲), μ^- (▼), and charge averaged (●).

[3]This sample should have a substantial fraction of fake muons, but it is not expected to be a dominant fraction. Thus, the $R(\chi^2_{r\phi,\mathrm{CMX}} < 9)$ measured here should be considered as an underestimate of the true value.

TABLE XXI: $R(\chi^2_{r\phi,\mathrm{CMX}} < 9)$ for $J/\psi \to \mu^+\mu^-$ and $B^0_{s(d)} \to \mu^+\mu^-$ sidebands. Only the statistical errors are shown.

p_T range, GeV/c	$R(\chi^2_{r\phi,\mathrm{CMX}} < 9)$, %	
	$J/\psi \to \mu^+\mu^-$ sideband	$B^0_{s(d)} \to \mu^+\mu^-$ sideband
2-2.25	0.033 ± 0.011	0.04 ± 0.04
2.25-2.5	0.14 ± 0.02	0.05 ± 0.04
2.5-3	0.58 ± 0.04	0.69 ± 0.09
3-4	2.66 ± 0.10	3.24 ± 0.14
4-5	5.9 ± 0.3	7.3 ± 0.3
5-6	8.5 ± 0.5	9.2 ± 0.6
6-8	10.9 ± 0.7	12.1 ± 0.9
8-10	17.4 ± 1.5	16.0 ± 1.9
10-15	16.6 ± 1.8	18 ± 3

The rejection power depends on the actual particle content of a selected sample, the measurement of the rejection power per particle type is outside the scope of this work. The pull distribution for the $J/\psi \to \mu^+\mu^-$ sidebands, as shown in Fig. 87, has a large Gaussian component. This suggests that most of the muon objects in the sidebands are the real muons or the muons from K or π decays.

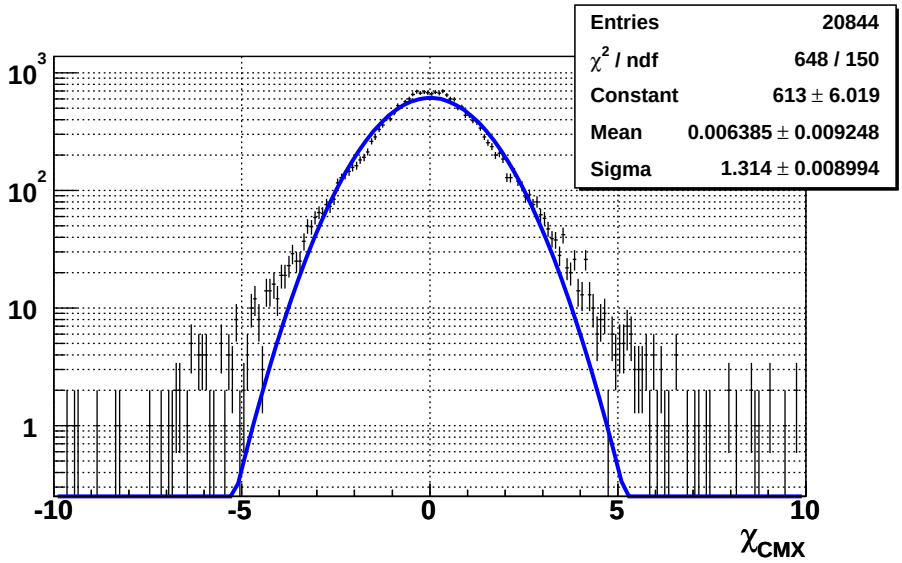

FIG. 87: $\chi_{\rm CMX}$ distribution for $p_T > 3$ GeV/c for the right $J/\psi \to \mu^+\mu^-$ sideband. The Gaussian component is large.

E. Summary

The $\chi^2_{r\phi,\rm CMX}$ variable has been constructed based on ΔX matching variable of the CMX muons. The efficiency of the $\chi^2_{r\phi,\rm CMX} < 9$ cut is measured using the $J/\psi \to \mu^+\mu^-$ signal events. The measurement is shown to be not biased by the Level-1 CMX trigger requirements. The fraction of background rejected by $\chi^2_{r\phi,\rm CMX}$ cut as applied to the sideband events is estimated for $J/\psi \to \mu^+\mu^-$ and $B^0_{s(d)} \to \mu^+\mu^-$ sidebands. It was shown that due to the stringent offline reconstruction requirements the cut starts to give the additional (background) rejection and is well-behaved at $p_T \gtrsim 4$ GeV/c. At the lower momenta the efficiency changes rather fast both in p_T and η. Here the η dependence is considered as a systematic uncertainty for the presented p_T dependent efficiency measurement. The efficiency for real muons, for $\chi^2_{r\phi,\rm CMX} < 9$ and $p_T > 4$ GeV/c, is $99.13 \pm 0.25\%$. The uncertainties are the quadrature sum of statistical and systematic contributions.

APPENDIX H

RESIDUAL BACKGROUND CONTRIBUTIONS FROM B-DECAYS

There is a potential possibility that $B \to h^+h^-$ (where $h^+ = \pi^+$ or K^+) with both hadrons misreconstructed as muons and/or generic B-hadron decays (also possibly with as much as two fake muons) might anomalously contribute to the background of $B^0_{s(d)} \to \mu^+\mu^-$ analysis. In particular if the $M_{\mu^+\mu^-}$ distribution were non-linear (as is clearly the case for the $B \to h^+h^-$ decays) or there were significant correlations among the discriminating variables, then the data-driven background estimate described in the Section IV.F may not properly account for these events.

The contributions from the potential background sources depend on the fraction of particular hadrons misidentified as muons (muon fake rate). The muon-fake rates are separately determined from the data for π^+, π^-, K^+, and K^- as a function of track p_T. The reconstructed decays $D^{\star+} \to D^0(\to K^+\pi^-)\pi+$ with the corresponding charge-conjugate mode are used to identify the particular flavor and charge of the mesons ($K^\pm$ or $\pi^\pm$) [157, 158]. For each given meson charge and flavor the muon fake rate is then defined by the fraction of the mesons reconstructed as muons passing the muon quality requirements. The dependence of the CMU muon fake rate as a function of p_T is shown in Figs. 88 and 89 for kaons and pions respectively [157].

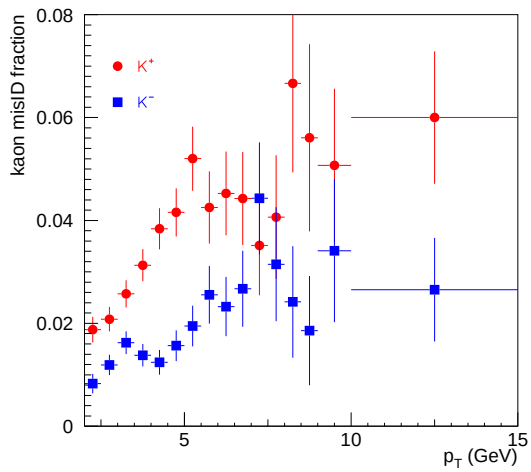

FIG. 88: Muon fake rate for kaons vs. p_T for CMU.

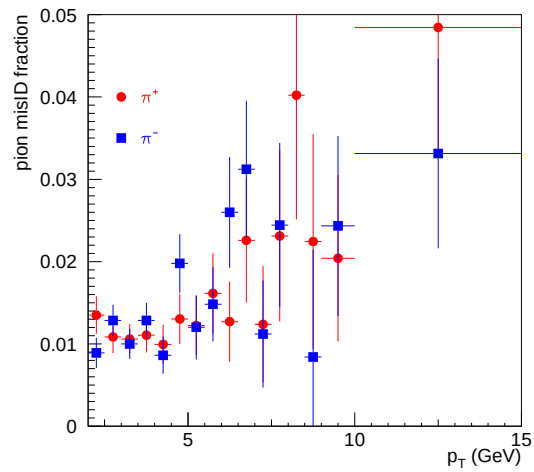

FIG. 89: Muon fake rate for pions vs. p_T for CMU.

The expected contributions for $B_d^0 \to \pi^+\pi^-$, $K^+\pi^-$, K^+K^-, and $B_s^0 \to \pi^+\pi^-$, π^+K^-, K^+K^- to the signal region are estimated using muon-fake rates. For each of the decay modes the effective observed fake $B_{s(d)}^0 \to \mu^+\mu^-$ branching ratio is defined as

$$\mathcal{B}_f(B_q^0 \to \mu^+\mu^-) = \sum_{i,j,q'} \epsilon_{m,ijq'} \frac{f_{q'}}{f_q} \kappa_i \kappa_j \mathcal{B}(B_{q'}^0 \to h_i h_j), \tag{H.1}$$

where $q, q' = \{d, s\}$ defines the meson flavor, κ_i is the muon fake rate for hadron h_i, $\epsilon_{m,ijq'}$ is the efficiency of the mass window selection for the given hadronic $B_{q'}^0 \to h_i h_j$ mode when the mass is calculated using $\mu^+\mu^-$ hypothesis, and f_q is the corresponding flavor branching fraction. Using the branching ratios for the B-meson two-body hadronic decays defined in Table II, and taking $\kappa_\pi \approx 1.4\%$ and $\kappa_{K^\pm} \approx 2.5\%$, it can be shown that the effective fake $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$ is around 5×10^{-9}. Note that even this low value is an overestimate, since the same κ_i is assumed the same for both CMU

and CMX, whereas the fake rate is smaller for CMX due to the larger amount of material in front of CMX, as shown in Fig. 24. Since the two body hadronic B-meson decays negligibly contribute to the background these are not considered further.

Using a generic $b\bar{b}$ MC sample generated by PYTHIA with direct flavor production, for generic $B \to \mu^+ X$ decays with the second muon candidate coming from either a fake or a real muon, it was found that $M_{\mu^+\mu^-}$ is linear and that no significant correlations exist among the discriminating variables, as discussed in detail in Ref. [135]. It was also demonstrated that a background estimate which exploits these assumptions (as does the method proposed in Section IV.F) accurately accounts for these events. It was found that none of the approximately 1.2×10^9 events generated were anywhere near the $(I, \Delta\alpha, \lambda)$ requirements made; this helped establish confidence that a large potential background isn't lurking near the cut boundaries. Neither of these conclusions are affected by using the likelihood discriminant.

The generic $b\bar{b}$ MC sample used in Ref. [135] omitted the gluon splitting production diagram, $g \to b\bar{b}$. An additional, but less detailed study was performed using the generic QCD PYTHIA MC sample with heavy flavor (b and c) in the final state. Due to the computational limitations in generating the generic QCD, the sample had an effective integrated luminosity of only 0.6 pb^{-1}, which in fact corresponds to about 5×10^9 generated events (an equivalent of more than 20 CPU GHz·years). This small integrated luminosity of the sample has only allowed to reliably estimate the event properties at the level of baseline selections. As shown in Fig. 90, the mass distribution is indeed linear for all the two-track combinations. This means that the corresponding true dimuon and fake (di)muon $B^0_{s(d)} \to \mu^+\mu^-$ candidates should be linear in $M_{\mu^+\mu^-}$. Given that the heavy flavor MC events are linear in $M_{\mu^+\mu^-}$ and do not have anomalous $(M_{\mu^+\mu^-}, I, \Delta\alpha, \lambda)$ correlations, they ought to be accurately accounted for in the background estimate using data.

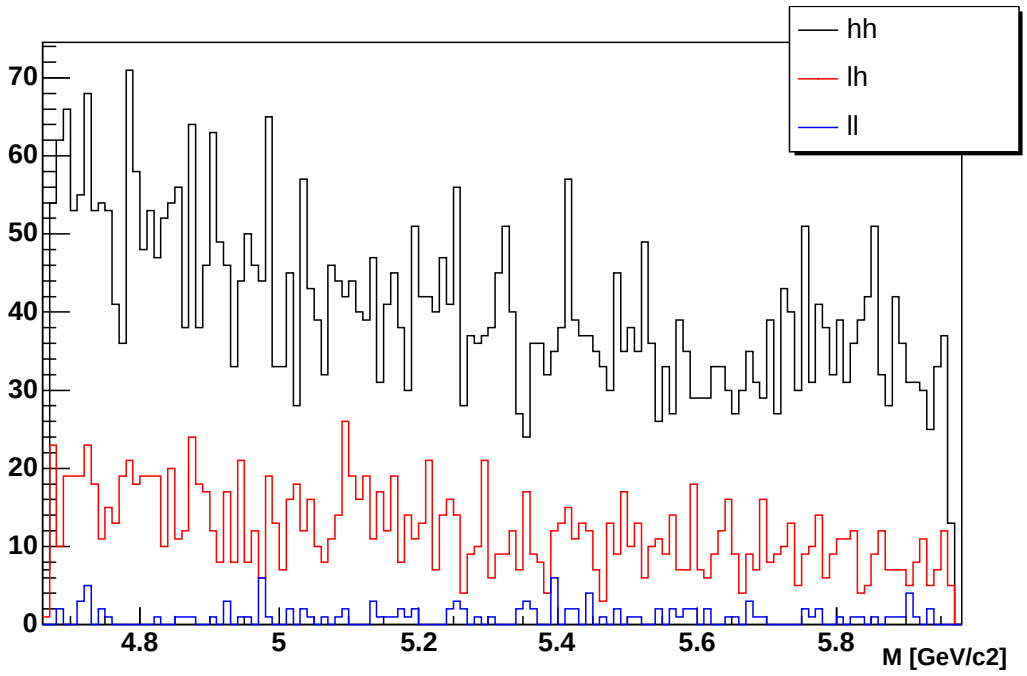

FIG. 90: Invariant mass distribution for $\mu^+\mu^-$candidates in generic QCD heavy flavor MC, including double-fake-muon (top), muon-fake-muon (middle), and dimuon contributions (lower) for MC events passing baseline and vertex selections applied on generator level. The distributions with hadrons are not rescaled to account for fake rates. To increase the statistics of the test both electrons and muons are counted as muons (denoted as l in the plot), this is possible due to approximate $e - \mu$ symmetry of the dominant (semi)leptonic heavy flavor decay modes.

APPENDIX I

NORMALIZATION MODE CORRECTION

Using the baseline and vertex requirements discussed in Section IV.C the number of $B^+ \to J/\psi K^+$ candidates, $N_{B^+\to J/\psi K^+}$, is estimated using simple sideband subtraction and correcting for the small contribution of $B^+ \to J/\psi\pi^+$ decays. The signal mass window is defined as $5.244 < M_{vtx} < 5.314$ GeV/c^2 while the sidebands are symmetrically defined to include an additional 120 MeV/c^2 on either side of the signal region. A correction is made for the number of $B^+ \to J/\psi\pi^+$ events expected to fall within this mass window using this expression:

$$N_{B^+\to J/\psi K^+} = N_{K+\pi}\left(1 + \frac{\mathcal{B}(B^+ \to J/\psi\pi^+)}{\mathcal{B}(B^+ \to J/\psi K^+)} \cdot \frac{\alpha_\pi}{\alpha_K} \cdot \frac{\epsilon_\pi^{reco}}{\epsilon_K^{reco}} \cdot \frac{\epsilon_\pi^{mass}}{\epsilon_K^{mass}}\right)^{-1} \quad \text{(I.1)}$$

$$= \frac{N_{K+\pi}}{1 + (0.0014 \pm 0.0004)} \quad \text{(I.2)}$$

where $N_{K+\pi}$ is the combined number of $B^+ \to J/\psi K^+$ and $B^+ \to J/\psi\pi^+$ decays in the $B^+ \to J/\psi K^+$ mass window determined using the sideband subtraction described above, α_π is the acceptance for $B^+ \to J/\psi\pi^+$ decays, ϵ_π^{reco} is the total reconstruction efficiency, including the trigger, COT, SVX, muon, and vertex requirements for $B^+ \to J/\psi\pi^+$ decays, and ϵ_π^{mass} is the efficiency of the $B^+ \to J/\psi K^+$ mass window requirements on the $B^+ \to J/\psi\pi^+$ sample and is equal to 0.035 ± 0.010.[1] The α_K, ϵ_K^{reco} and ϵ_K^{mass} are analogously defined for the $B^+ \to J/\psi K^+$ decays and are given in Section IV.G above. The $B^+ \to J/\psi\pi^+$ terms are determined in the same manner as the $B^+ \to J/\psi K^+$ terms. The branching ratio for $B^+ \to J/\psi\pi^+$ is $\mathcal{B}(B^+ \to J/\psi\pi^+) = (4.0 \pm 0.5) \times 10^{-5}$ [1].

[1]Note, the $B^+ \to J/\psi\pi^+$ mass will be shifted to higher values because the pion is assumed to have the kaon mass.

Using this expression the $B^+ \to J/\psi\pi^+$ corrected number of $B^+ \to J/\psi K^+$ events is estimated as $N^{UU}_{B^+\to J/\psi K^+} = 1785 \pm 60$ and $N^{UX}_{B^+\to J/\psi K^+} = 696 \pm 39$, where the uncertainties are completely dominated by the statistics of the sideband subtraction. These numbers are used in Eq. (4.7) to estimate the $\mathcal{B}(B^0_s \to \mu^+\mu^-)$ value.

VITA

Vyacheslav Krutelyov was born in 1977 in Minks, Belarus. He graduated from middle school #119 in 1992 and attended Lyceum at Belarus State University from 1992 to 1994. He attended Belarus State University where he majored in physics and specialized in theoretical physics from 1994 to 1999. Upon graduation with a B.S. in physics he joined Texas A&M University. In the winter of 1999-2000 he began working as a part of the CDF collaboration in experimental particle physics. His permanent mailing address is Department of Physics, TAMU MS4242, College Station, TX 77843.

The typist for this dissertation was Vyacheslav Krutelyov.