Cornell University
Floyd R. Newman Laboratory for
Elementary-Particle Physics

First two-sided limit of the $B_s \to \mu^+\mu^-$ decay rate

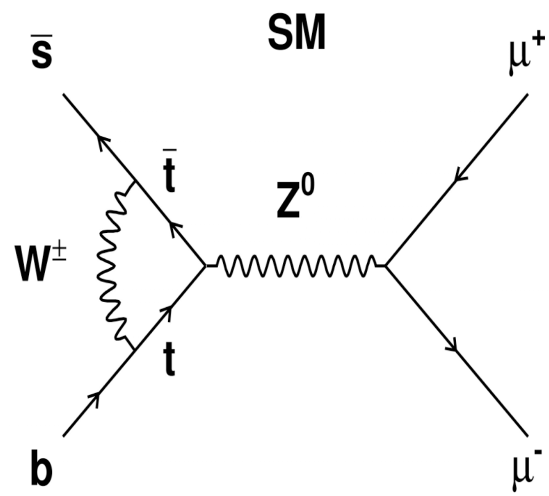
Wine and Cheese Seminar

Fermilab, 7/15/2011

Julia Thom-Levy, Cornell University

$B_s\,(B^0)\rightarrow \mu^+\mu^-$: the golden channel for FCNC searches

- In the Standard Model (SM) process is highly suppressed
  - Cabibbo and helicity suppressed
  - accessible only through higher order EWK diagrams
  - SM rate predicted with ~10% accuracy:



$$BR(B_s\rightarrow \mu^+\mu^-)=(3.2\pm0.2)\times10^{-9}$$
$$BR(B^0\rightarrow \mu^+\mu^-)=(1.0\pm0.1)\times10^{-10}$$

A. J. Buras et al., JHEP 1010:009,2010

- Search for this decay has long been of great interest:
  - robust experimental signature
  - many New Physics(NP) models predict much larger branching fraction

e.g.Choudhury, Gaur, PRB 451, 86 (1999); Babu, Kolda, PRL 84, 228 (2000).

Probing New Physics

All NP models with new scalar operators predict enhancement.
In NP models without new scalar operators,
$BR(B_s \to \mu^+\mu^-) > 10^{-8}$ are unlikely

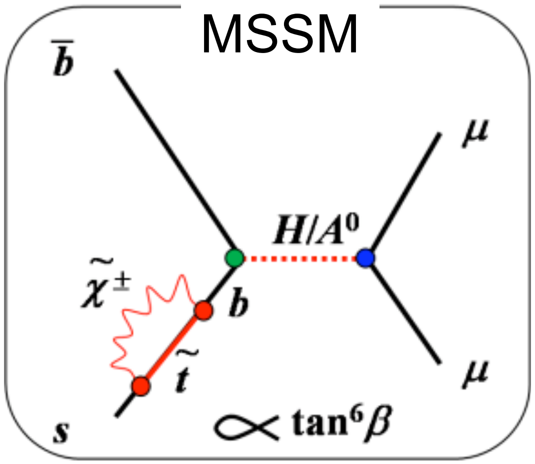
Examples of NP models:

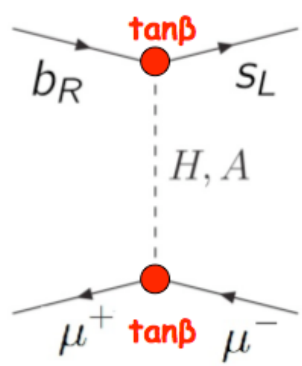
- Loop: MSSM, mSUGRA
  - Rate prop. to $\tan^6\beta$, e.g. 3 orders of magnitude enhancement

  Dedes, Dreiner, Nierste, PRL87:251804 (2001)

- Tree: Flavor violating models or R-Parity violating SUSY

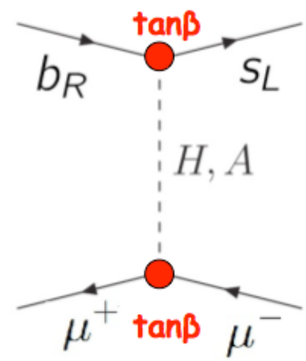
- LHT, RS, SM with 4 generations
  - modest NP contributions to $BR(B_s \to \mu^+\mu^-)$

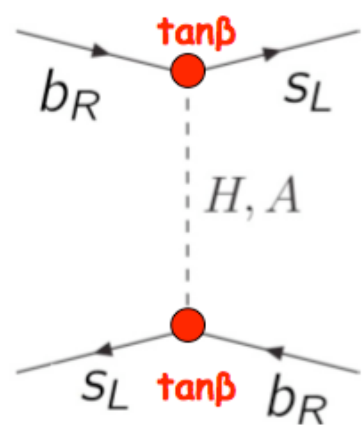




Probing New Physics

- "Smoking gun" of some Flavor Violating NP models:
  - ratio BR($B_s \rightarrow \mu^+\mu^-$)/ BR($B^0 \rightarrow \mu+\mu-$) highly informative about whether NP violates flavor significantly or not
  - clear correlation between CP violating mixing phase from $B_s \rightarrow J/\psi\phi$ and BR($B_s \rightarrow \mu^+\mu^-$)
    Altmannshofer, Buras, Gori, Paradisi, Straub, Nucl.Phys.B830:17-94,2010





- Important complementarity with direct searches at Tevatron and LHC
  - Indirect searches can access even higher mass scales than LHC COM energies

New bounds on BR($B^0 \rightarrow \mu^+\mu^-$) and BR($B_s \rightarrow \mu^+\mu^-$) are of crucial importance, and are a top priority at the Tevatron and LHC.

Probing New Physics

Plenary talk
A.Buras, Beauty 2011:

Maximal Enhancements of $S_{\psi\varphi}$, $Br(B_s \to \mu^+\mu^-)$ and $K^+ \to \pi^+\nu\bar{\nu}$

(without taking correlation between them)

Model	Upper Bound on ($S_{\psi\varphi}$)	Enhancement of $Br(B_s \to \mu^+\mu^-)$	Enhancement of $Br(K^+ \to \pi^+\nu\bar{\nu})$
CMFV	0.04	20%	20%
MFV	0.04	1000%	30%
LHT	0.30	30%	150%
RS	0.75	10%	60%
4G	0.80	400%	300%
AC	0.75	1000%	2%
RVV	0.50	1000%	10%

Large RH Currents

RS = RS with custodial protections
AC = Agashe, Carone $U(1)_F$
RVV = Ross, Velaso-Sevilla, Vives (04) $SU(3)_F$

Experimental Status

DØ 6.1 fb^{-1}:

BR($B_s \rightarrow \mu^+\mu^-$)<5.1 x$10^{-8}$

PLB 693 539 (2010)

CDF 3.7 fb^{-1}:

BR($B_s \rightarrow \mu^+\mu^-$) <4.3 x$10^{-8}$

BR($B^0 \rightarrow \mu^+\mu^-$) <7.6 x$10^{-9}$

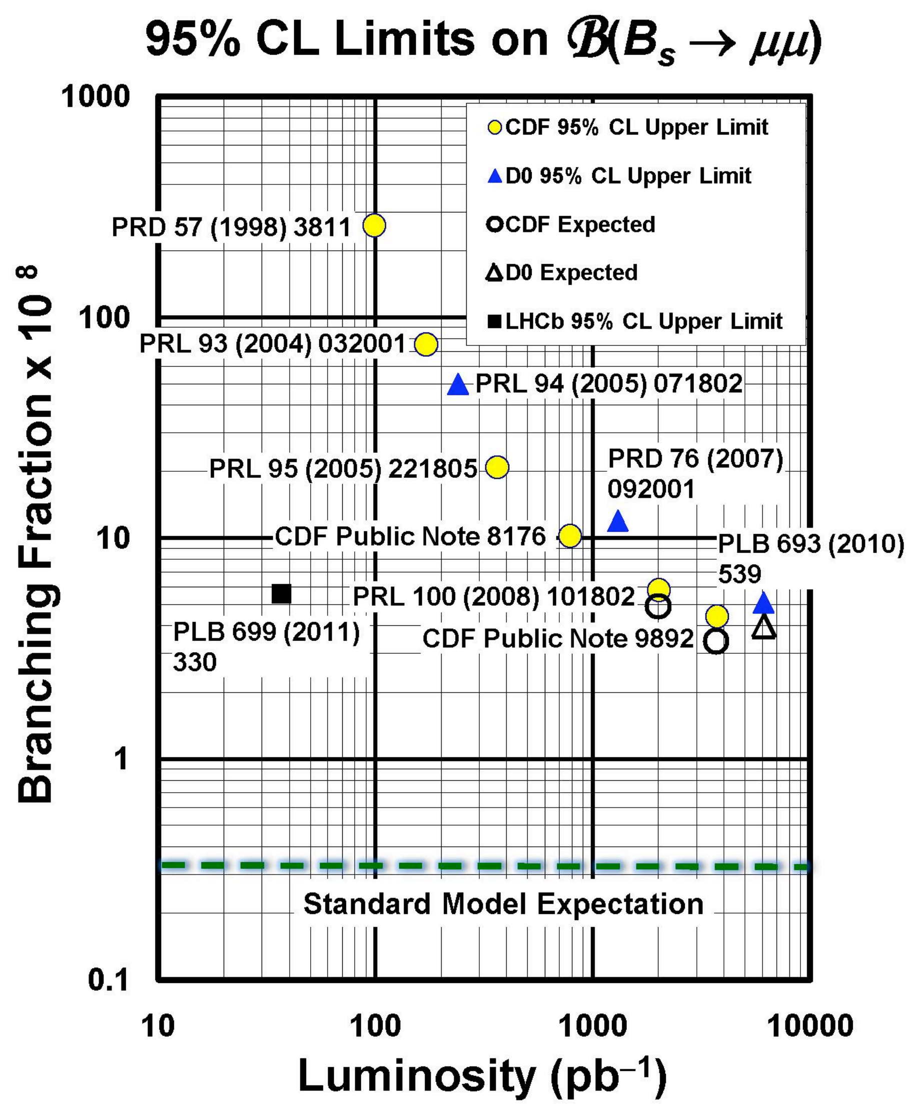
public note 9892

LHCb 37pb^{-1}:

BR($B_s \rightarrow \mu^+\mu^-$) < 5.6 x 10^{-8}

BR($B^0 \rightarrow \mu^+\mu^-$) < 15 x 10^{-9}

PLB 699, 330 (2011)



All limits quoted @95% C.L.

Analysis overview

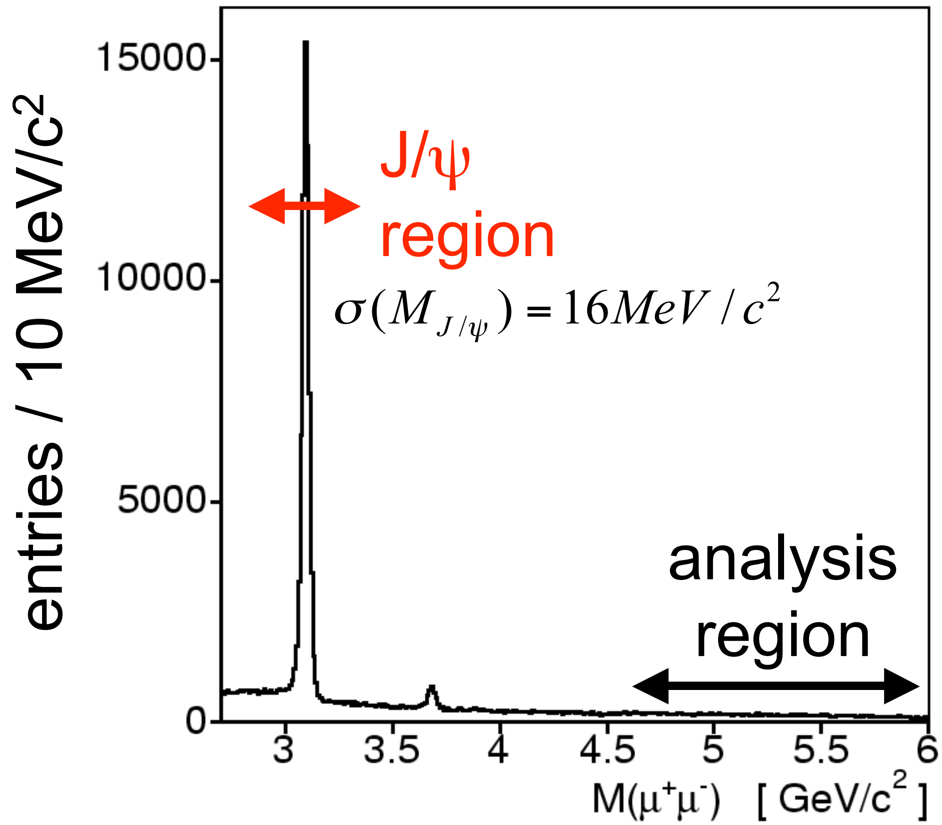
- Data collected using dimuon trigger, 7 fb^{-1}
- Loose pre-selection identifies B_s and B^+ search samples
  - $B^+ \rightarrow J/\psi K^+ \rightarrow \mu^+\mu^- K$ is used as a normalization mode to suppress common systematic uncertainties
- S/N for the $B_s(B^0)$ sample is further improved by using a Neural Net discriminant
- Signal window is blinded
- backgrounds are evaluated using sideband data and other control samples
- $BR(B_s(B^0) \rightarrow \mu^+\mu^-)$ is determined relative to the $B^+ \rightarrow J/\psi K^+$ rate after correcting for relative trigger and reconstruction efficiencies extracted from data (when possible) and simulation.
- Can search for B^0 and $B_s \rightarrow \mu^+\mu^-$ decays separately
  - dimuon mass resolution ~24 MeV < M_{Bs} - M_{B0}

Trigger

Data collected using dimuon trigger

- "CC":
  - 2 central muons "CMU", $|\eta|<0.6$,
  - $p_T>1.5$ GeV
  - $2.7<M_{\mu\mu}<6.0$ GeV
  - $p_{T(\mu)}+p_{T(\mu)}>4$ GeV

- "CF":
  - one central, one forward muon "CMX", $0.6<|\eta|<1.0$
  - $p_T>2$ GeV
  - other cuts same as above



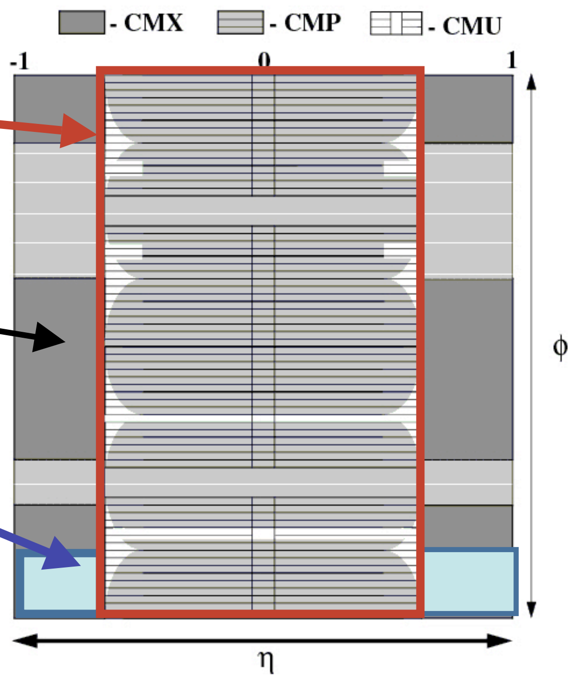
Trigger efficiency same for muons from J/ψ or B_s (for muon of a given p_T)

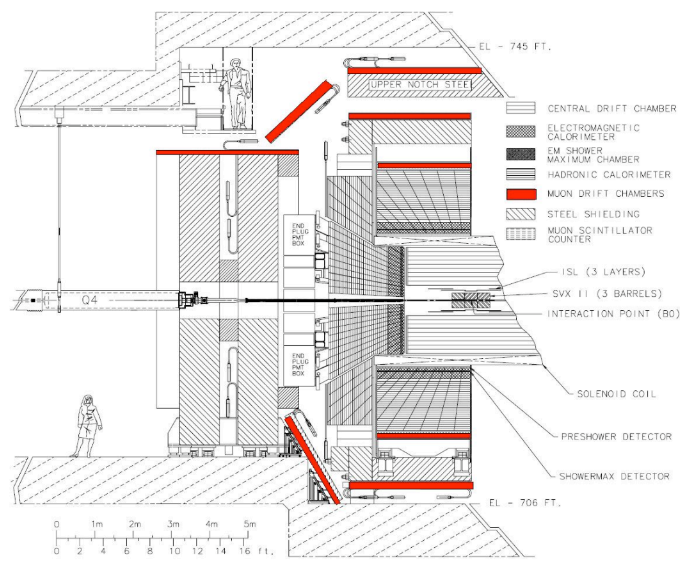
Improvements over previous $B_s(B^0) \rightarrow \mu^+\mu^-$ result from CDF

- Using twice the integrated luminosity (7 fb^{-1})
- Extended acceptance of events in the analysis by ~20%
  - muon acceptance includes forward muons detected in CMX miniskirts
  - 12% from tracking acceptance increase (using previously excluded "COT spacer region")
- Analysis improvements include an improved NN discriminant

Central muons (CMU)

Forward muons (CMX) including "miniskirts"

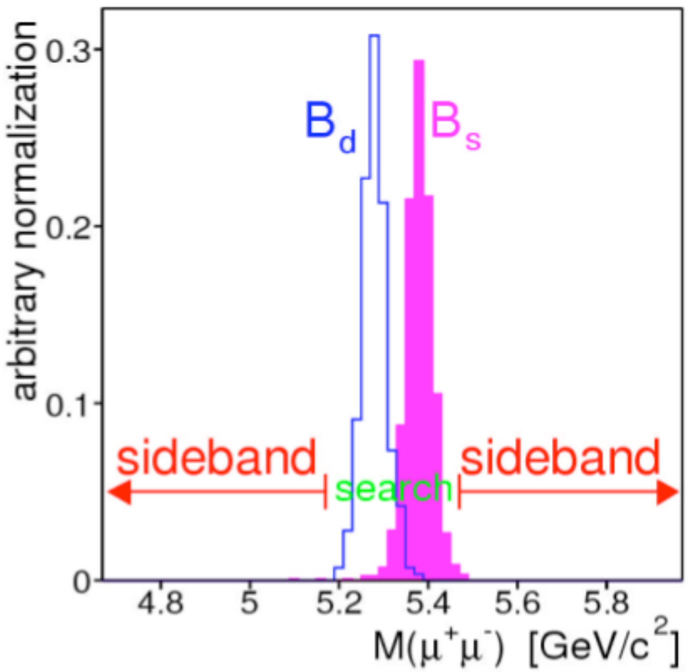




"Blind" search region

- Search region: $5.169 < M_{\mu\mu} < 5.469$ GeV
  – corresponds to $\pm 6\times\sigma_m$, where $\sigma_m \approx 24$MeV (2-track invariant mass resolution)
- Sideband regions: additional 0.5 GeV on either side
  – Used to understand background

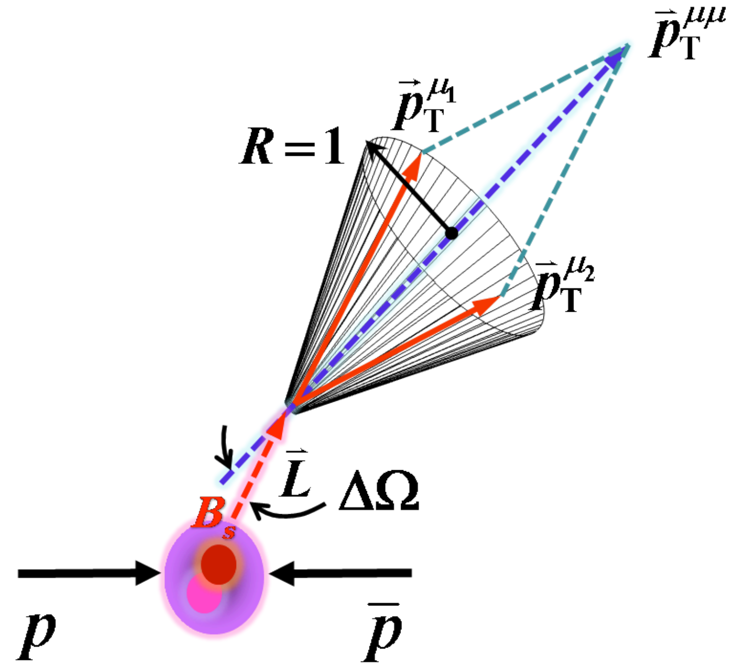
MC simulation of B_s and $B^0 \rightarrow \mu^+\mu^-$ mass peaks



Pre-selection variables

Samples of candidate B^+ and $B_s(B^0)$ decays pass track quality cuts and are constrained to a common 3D vertex. We apply loose baseline cuts on:

- isolation of B candidate and pointing angle ($\Delta\Omega$)
- transverse momentum of candidate B and muon tracks
- significance of proper decay time
- invariant mass



$$\text{Isolation} = \frac{p_T(\mu\mu)}{\sum p_T(tracks) + p_T(\mu\mu)}$$

all tracks within a cone of R=1 around $p_T(\mu\mu)$ considered

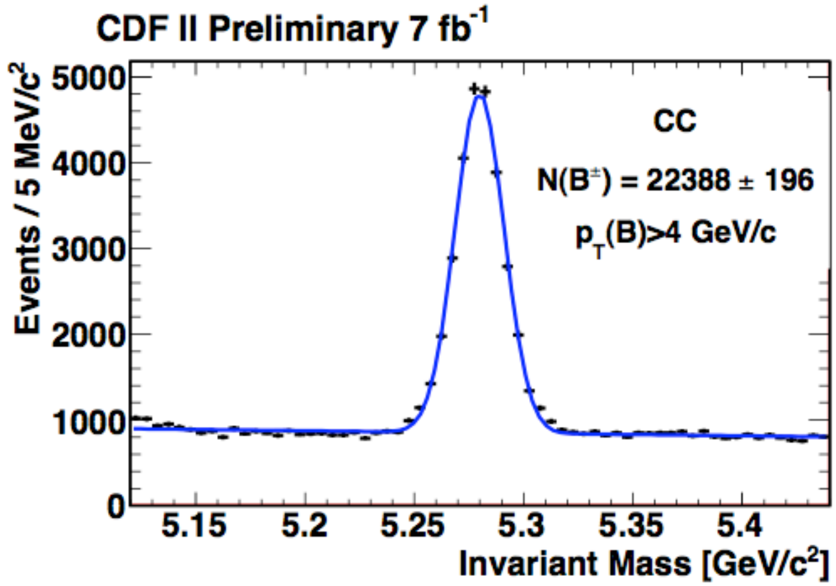
Pre-selection: B^+ normalization sample

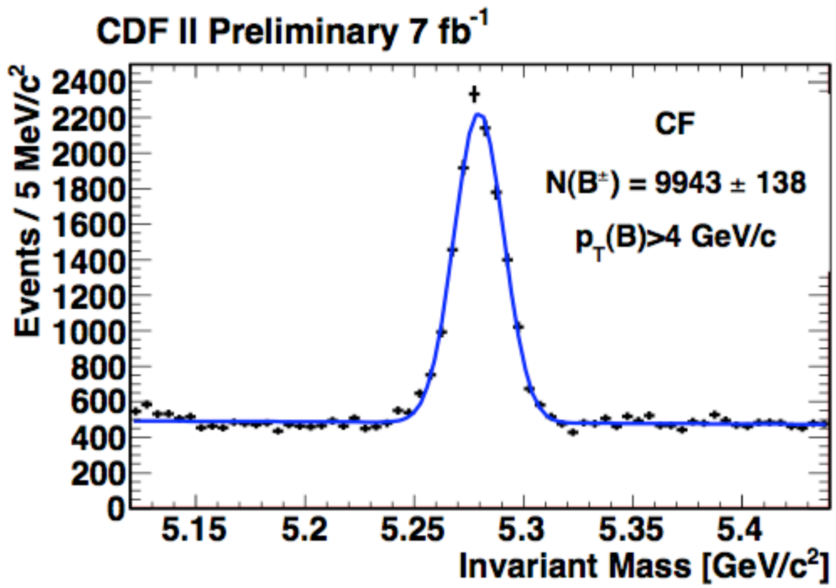
$B^+ \rightarrow J/\psi K \rightarrow \mu^+\mu^- K$,

~30k candidates.

In addition to baseline cuts, B^+ sample passes

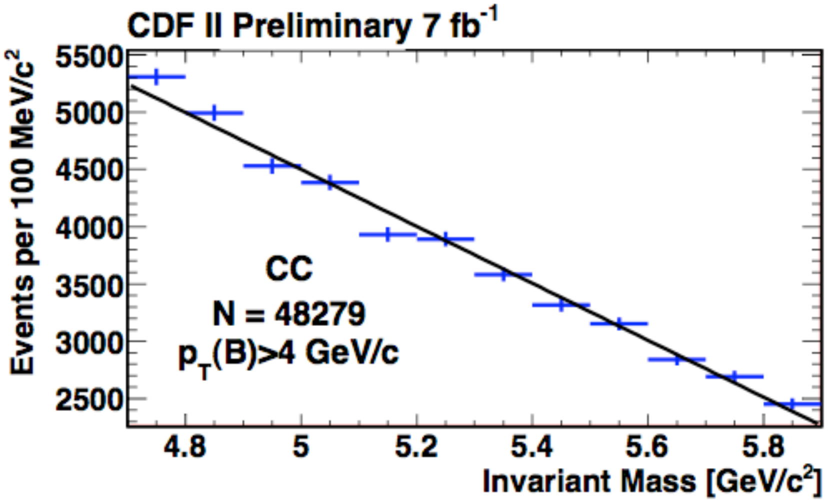
- J/ψ mass constraint for dimuons
- K quality cuts, and K and J/ψ constrained to common vertex

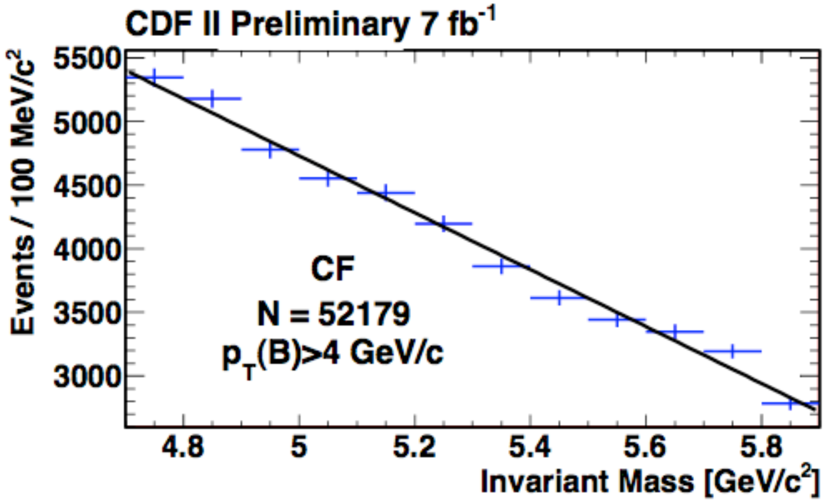




Pre-selection: $B_s(B^0)$ search samples

$B_s(B^0)$ search sample,
~100k candidates





Signal selection

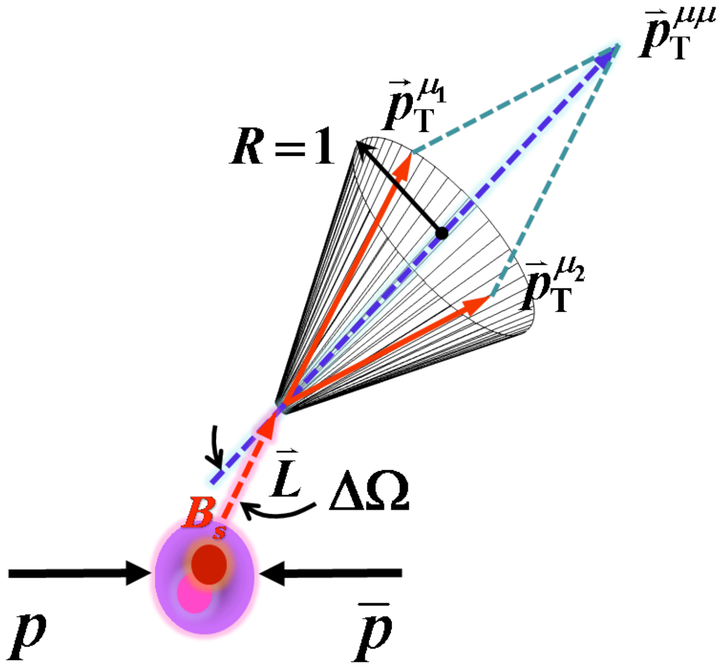
- Discriminating variables
- Neural Network

$B_s(B^0)$ Signal vs. Background

Assuming SM production, we expect ~2 events at this stage.
Need to reduce background by ~10^5!

Signal characteristics:

- final state is fully reconstructed
- B fragmentation is hard- few additional tracks, and L and $p_T(\mu\mu)$ are co-linear
- B_s has long lifetime ~1.5ps

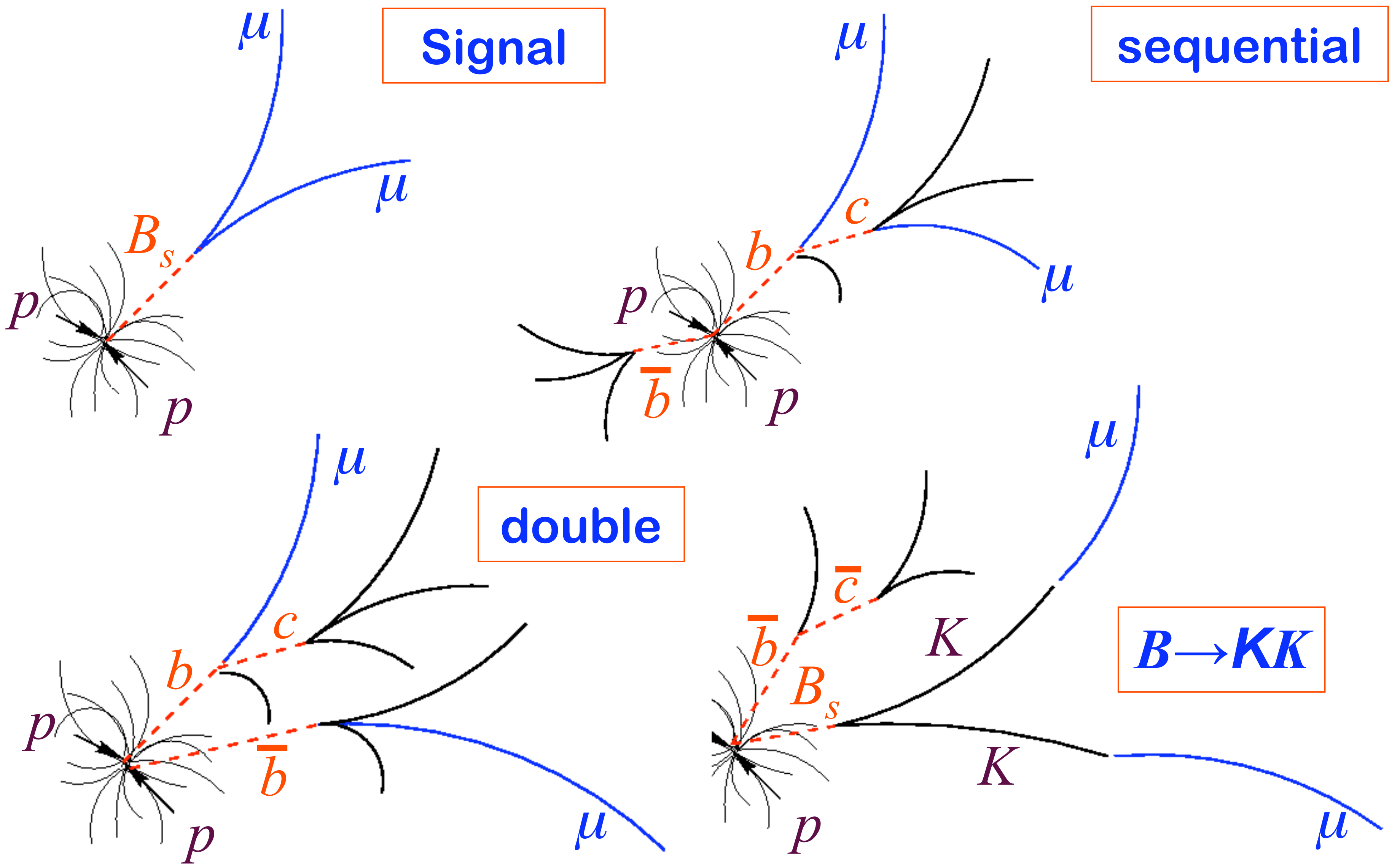


Backgrounds

- Sequential semi-leptonic decay: $b \to c\mu^- X \to \mu^+\mu^- X$
- Double semileptonic decay: $bb \to \mu^+\mu^- X$
- Continuum $\mu^+\mu^-$
- μ + fake, fake+fake

Good discriminators: isolation, mass, lifetime, p_T , how well p_T aligns with L

$B_s(B^0)$ Signal vs. Background



From B. Casey, ICHEP2010

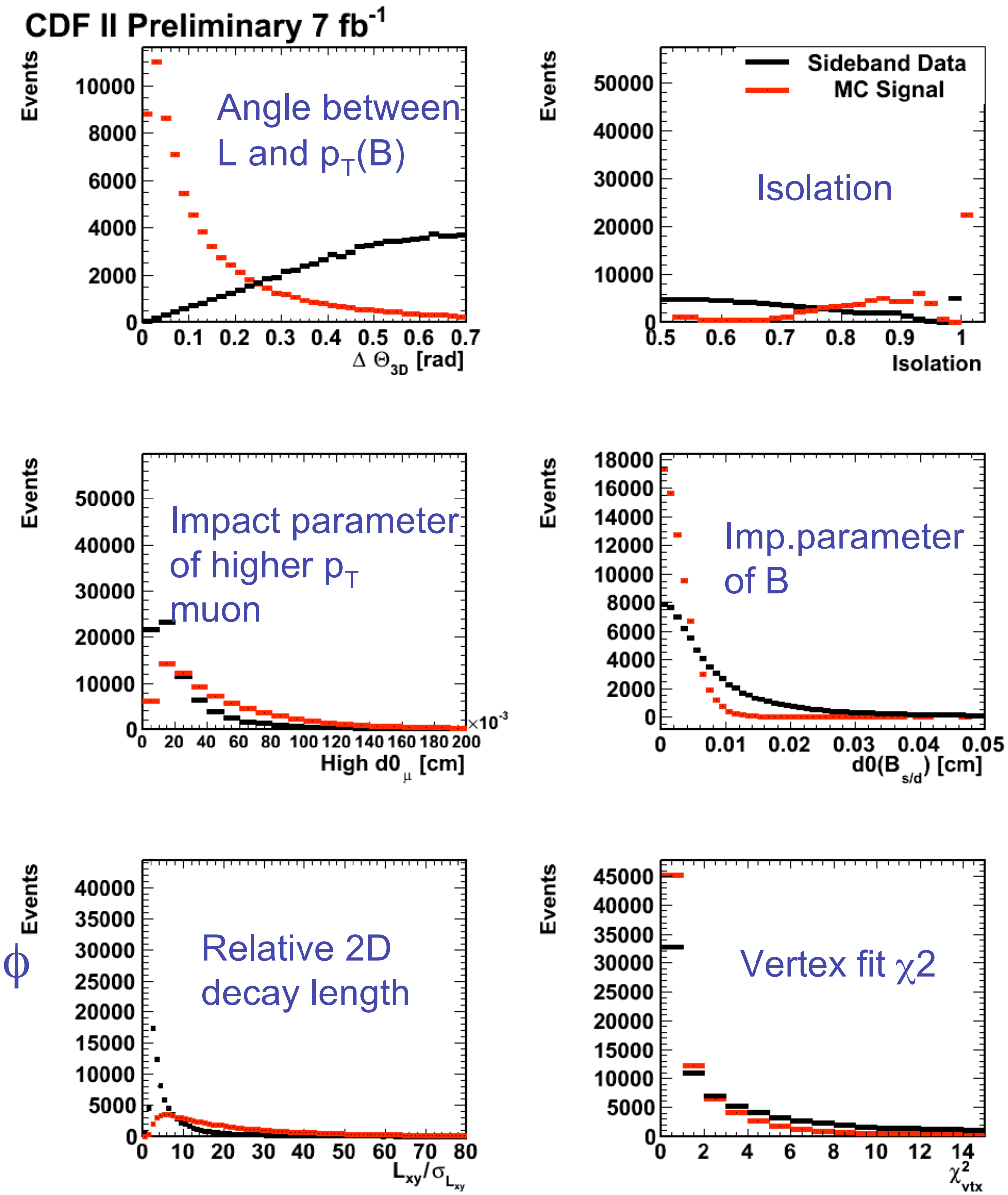
Discriminating Variables

14 variables are combined into a Neural Net (except $M_{\mu\mu}$)

6 most sensitive variables shown here:

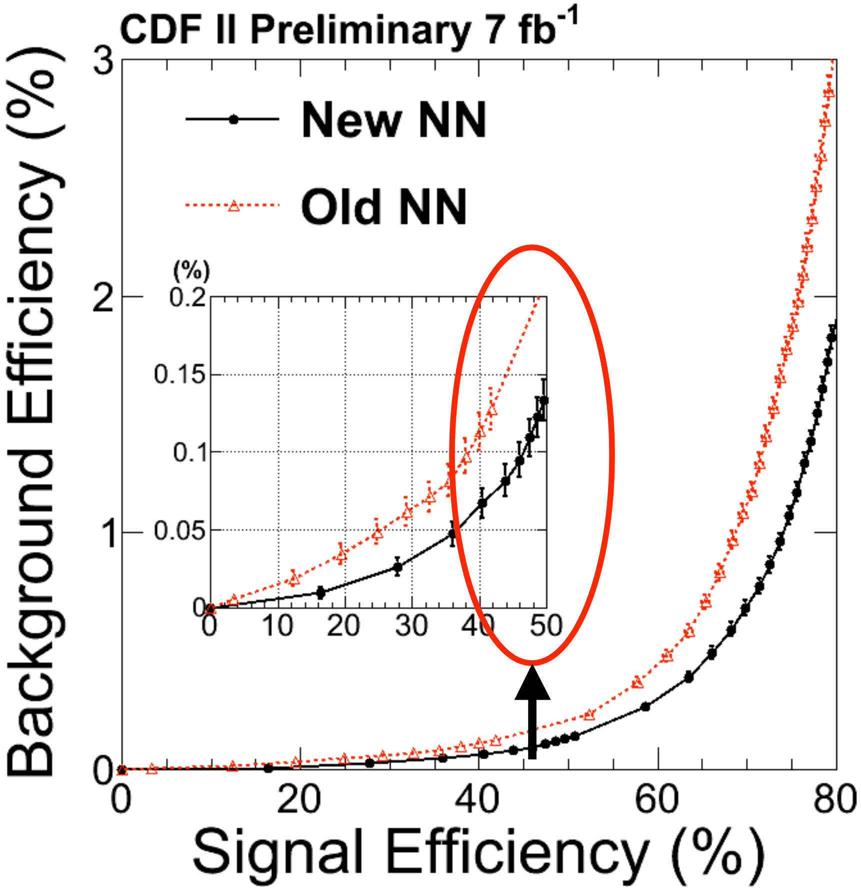
In red: MC signal (Pythia), black: sideband data

- B-hadron p_T spectrum is reweighted using $B^+ \to J/\psi K \to \mu^+\mu^- K$ data
- isolation distribution is reweighted using $B_s \to J/\psi\phi$ data



Improvements over previous $B_s(B^0) \rightarrow \mu^+\mu^-$ result

Using an improved Neural Network that achieves twice the background rejection for the same signal efficiency



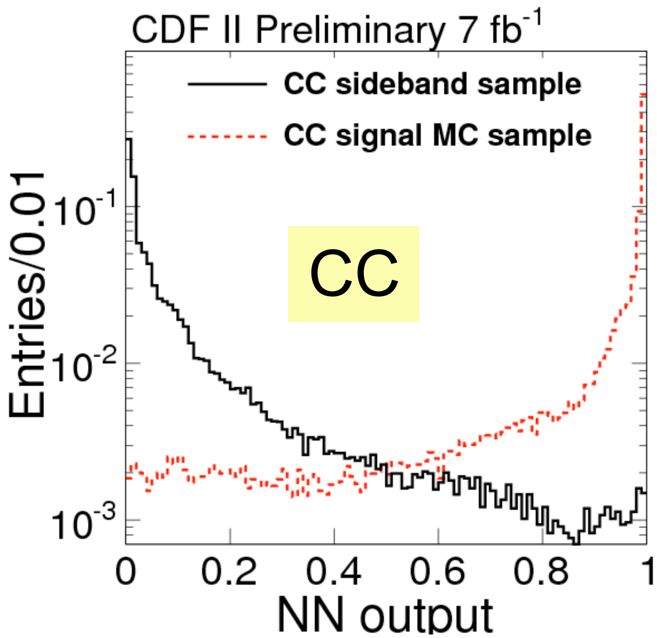
Neural Network Output

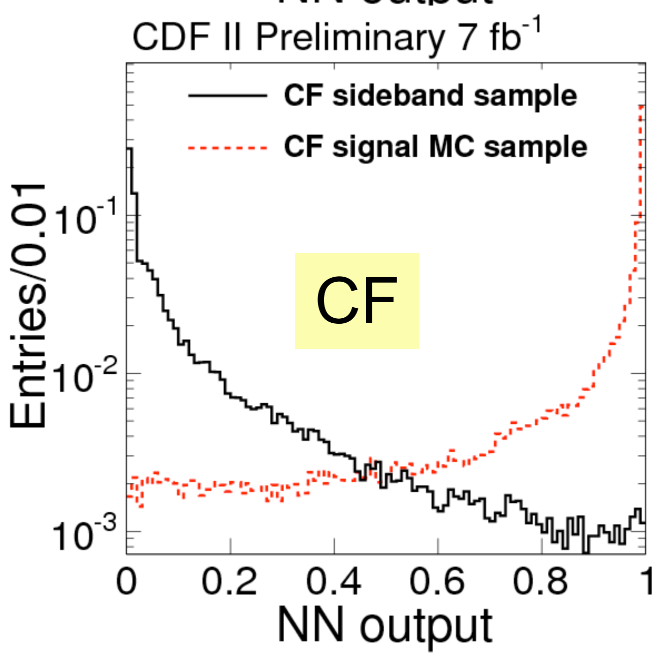
Separation between NN output for background and for signal MC

- input variables and NN signal performance has been checked in $B^+ \rightarrow J/\psi K^+ \rightarrow \mu^+\mu^- K$ data

- events with NN output >0.7 are considered candidates

- we take advantage of improved background suppression with high NN output by dividing into 8 sub-samples, using an a-priori optimization





A priori optimization

Figure of merit: expected upper limit on $BR(B_s(B^0))$, calculated using CLs method

- choice of optimal binning made using MC pseudo-exps.
  - mean expected background from data sideband
  - uncertainty (syst. and stat.) on mean included in pseudo-experiments

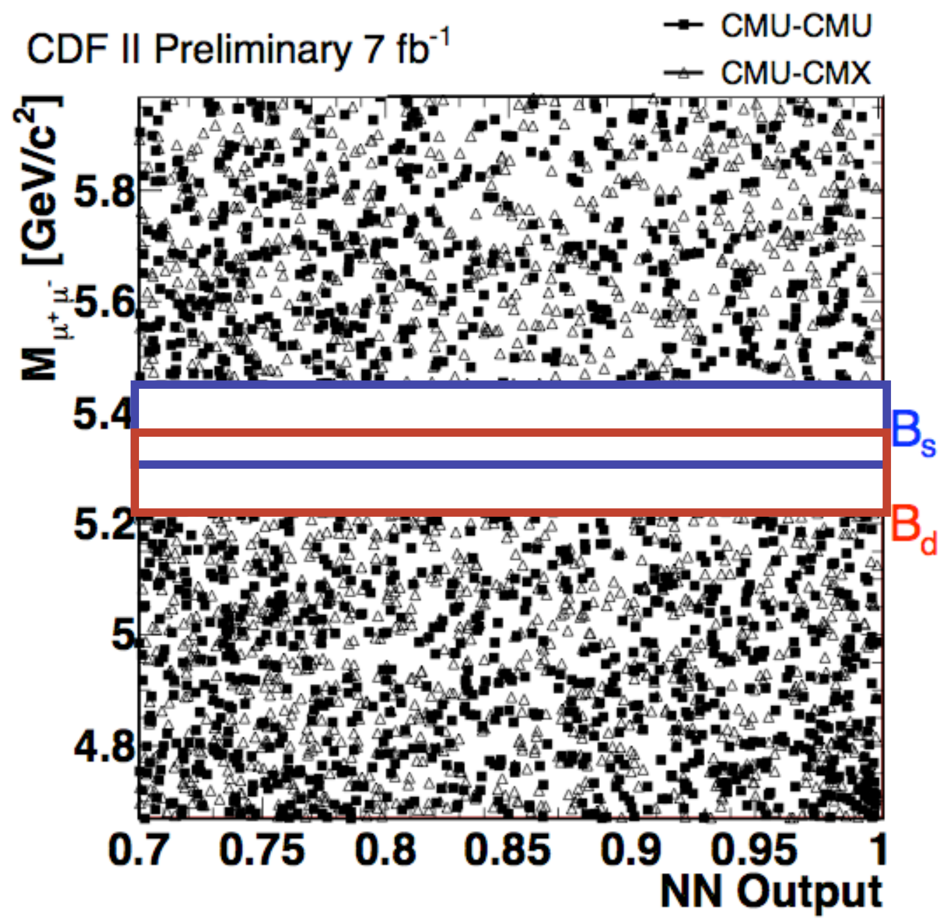
- resulting configuration:
  - 8 bins of NN output between 0.7 and 1.0
  - each NN bin divided into 5 mass bins
  - separately for CC and CF

Highest sensitivity in
3 highest NN bins
0.97 < NN output < 0.987
0.987< NN output < 0.995
0.995< NN output <1

Short recap: Neural Net selection

We are using a 2-dimensional selection:

- a Neural Net is used to select $B \rightarrow \mu^+\mu^-$-like candidates, independent of mass (8 bins)
- B_s and B^0 mass windows are blinded (5 bins)
- CC and CF mode treated separately



Background estimates

- Sources of background events
- Estimation methods
- Cross checks in control samples

$B_{s(d)} \to \mu^+\mu^-$ backgrounds, overview

1) Combinatoric backgrounds:

- continuum $\mu^+\mu^-$ from Drell-Yan
- double semileptonic bb $\to \mu^+\mu^- X$
- b/c $\to \mu$ + fake μ (K,π)

MC predicts a smooth $M_{\mu\mu}$ distribution

2) Two-body hadronic B decays

- B$\to$ hh where h $\to$ fake μ (K,π)

peaking in signal region

1) Combinatoric background

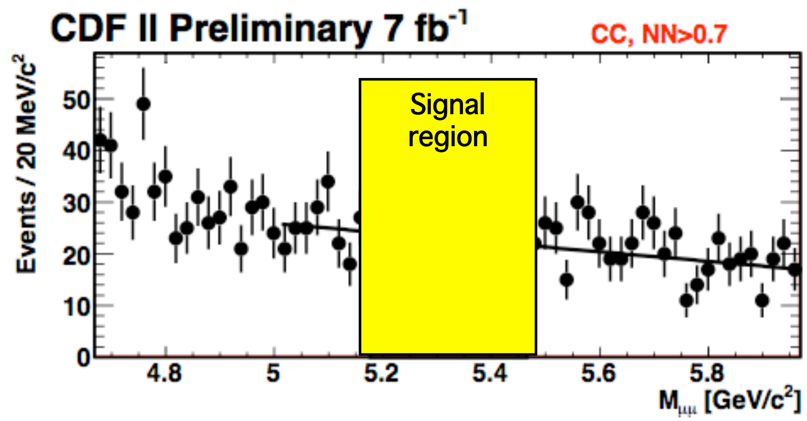
Using our background dominated data sample, fit $M_{\mu\mu}$ to a linear function.

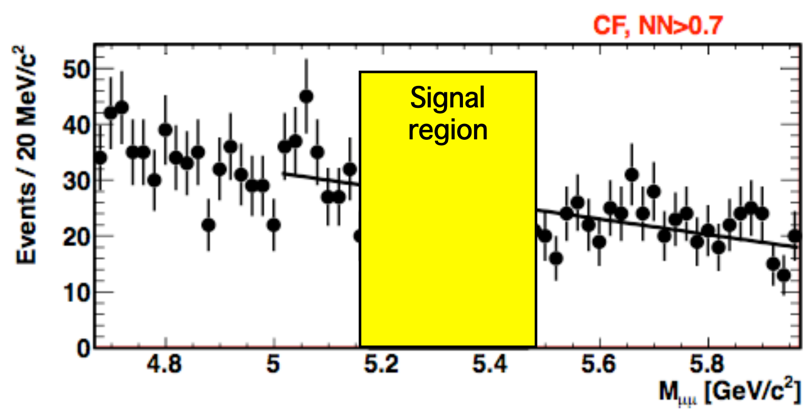
- use distributions of sideband events with NN output >0.7

- only events with $M_{\mu\mu}$>5 GeV used to suppress contributions from $b \rightarrow \mu\mu X$

- slopes then fixed and normalization determined for each NN bin

- systematic uncertainty determined by studying effects of various fit functions and fit ranges
    - between 10-50%





Expected number of combinatoric background events in the signal region

All uncertainties are included

B_s signal window:

NN Bin	CC	CF
$0.700 < NN < 0.970$	129.2±6.5	146.3±7.0
$0.970 < NN < 0.987$	7.9±1.9	11.6±1.8
$0.987 < NN < 0.995$	4.0±1.1	3.3±1.0
$0.995 < NN < 1.000$	0.79±0.52	2.6±1.5

B^0 signal window:

NN Bin	CC	CF
$0.700 < NN < 0.970$	134.0±6.6	153.4±7.3
$0.970 < NN < 0.987$	8.2±2.0	12.1±1.9
$0.987 < NN < 0.995$	4.1±1.2	3.4±1.1
$0.995 < NN < 1.000$	0.8±0.5	2.8±1.6

2) Background from two-body hadronic B decays

Two-body B→ hh decays where h produces a fake muon can contribute to the background

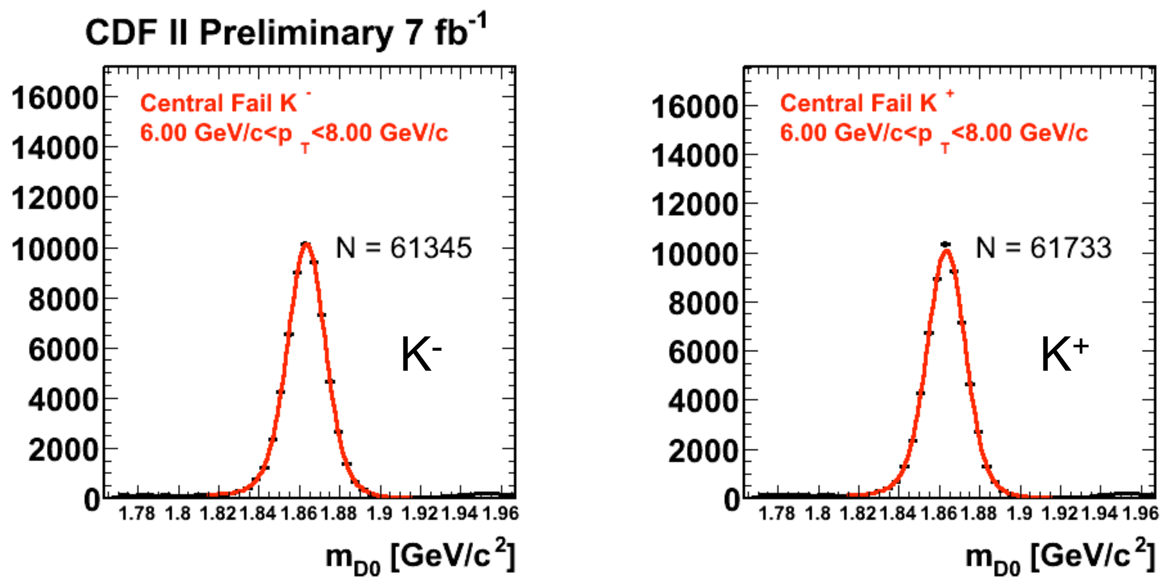
- fake muons dominated by π^+, π^-,K^+, K^-
- fake rates are determined separately using D*-tagged $D \rightarrow K^-\pi^+$ events

Estimate contribution to signal region by:

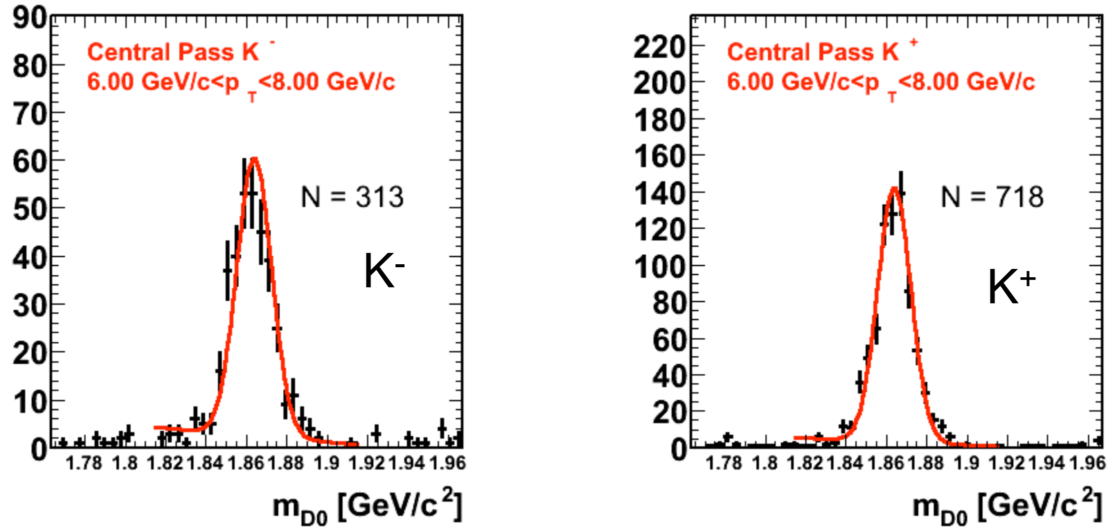
- take acceptance, M_{hh}, $p_T(h)$ from MC samples. Normalizations derived from known branching fractions
- convolute $p_T(h)$ with p_T and luminosity-dependent μ-fake rates. Double fake rate ~0.04%

Fake rates from D*-tagged $D^0 \rightarrow K^-\pi^+$ events

Example of D^0 peaks in one bin of p_T, used to extract a p_T and luminosity-dependent fake rate for K^+ and K^-



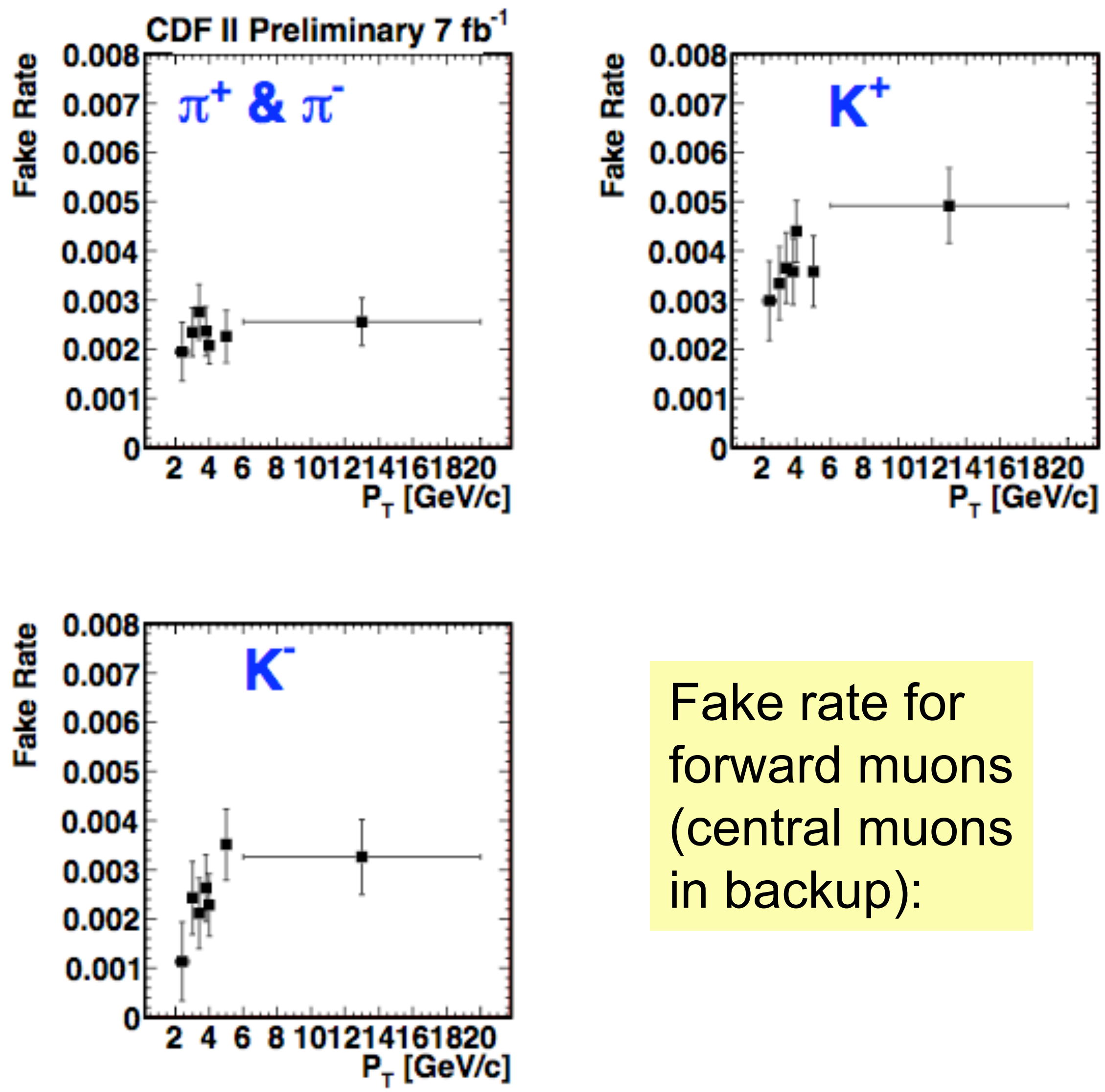
Kaons passing muon selection:



Muon fake rates

- Variations with p_T and luminosity are taken into account

- Total systematic uncertainty (due to both muon legs) dominated by residual run-dependence: ~35%



Fake rate for forward muons (central muons in backup):

Expected number of $B \rightarrow hh$ background events in the signal region

B_s signal window:

NN Bin	CC	CF
$0.700 < NN < 0.970$	0.03±0.01	0.01± <0.01
$0.970 < NN < 0.987$	0.01± <0.01	0.01± <0.01
$0.987 < NN < 0.995$	0.02± <0.01	0.01± <0.01
$0.995 < NN < 1.000$	0.08±0.02	0.03±0.01

← 10x smaller than combinatoric bkg

B_d signal window:

NN Bin	CC	CF
$0.700 < NN < 0.970$	0.31±0.08	0.09±0.02
$0.970 < NN < 0.987$	0.13±0.03	0.05±0.01
$0.987 < NN < 0.995$	0.19±0.05	0.04±0.01
$0.995 < NN < 1.000$	0.72±0.20	0.20±0.05

← Comparable to combinatoric bkg

Cross checks of the total background prediction

Apply background model to statistically independent control samples and compare result with observation. We have investigated 2 groups of samples:

1) Control samples composed mainly of combinatorial backgrounds
   - **OS-**: $\mu^+\mu^-$ events with negative proper decay length
   - **SS+**: loose pre-selection* and same sign muon pairs
   - **SS-**: like SS+ but negative proper decay length

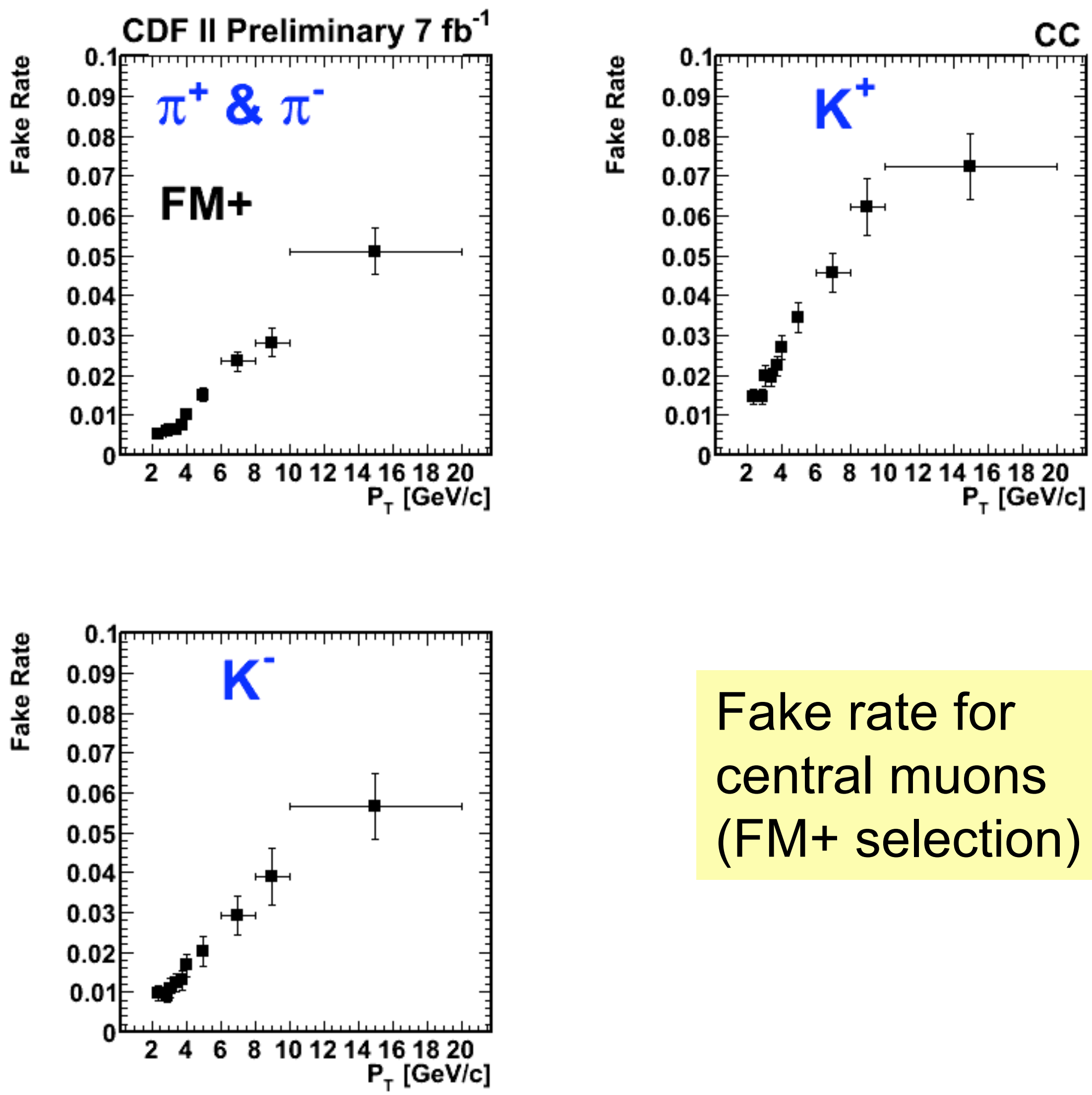
2) Control sample with significant contribution from B->hh background
   - **FM+**: loose pre-selection and at least one muon fails quality requirements

* Loose pre-selection = $p_T(\mu)$>1.5 and $p_T(\mu\mu)$> 4 GeV

Aside: The FM+ control sample

The FM+ control sample has at least one muon which fails our muon quality requirements

➔need a different set of K/π fake rates since the muon ID requirements are different than used in the signal sample. Same method as before is used



Fake rate for central muons (FM+ selection)

Result of background checks in control samples

Control Sample	Prediction	Nobs	Prob(N>=Nobs)
OS-	2140.0 ± 53.9	1999	98%
SS+	19.7 ± 3.4	25	19%
SS-	46.8 ± 5.3	53	25%
FM+	567.8 ± 25.4	593	24%
Sum	2774.3 ± 59.9	2670	91%

Shown are total number of events in all NN bins.

- "Prob(N>=Nobs)" is the Poisson probability for making an observation at least as large given the predicted background

✓ Good agreement across all control samples.

Full table of bkgd checks in control samples

CC only, see backup for CF

- Good agreement in most sensitive NN bins
- ✓ now have sufficient confidence in background estimation

FM+ is rich in B->hh background. Good agreement in highest NN bin shows that we can accurately predict this background

sample	*NN* cut	CC pred	CC obsv	CC prob(%)*
OS-	0.700<NN<0.760	217.4±(12.5)	203	77.7
	0.760<NN<0.850	262.0±(14.1)	213	99.1
	0.850<NN<0.900	117.9±(8.6)	120	44.7
	0.900<NN<0.940	112.1±(8.4)	116	39.4
	0.940<NN<0.970	112.7±(8.4)	108	64.2
	0.970<NN<0.987	80.2±(6.9)	75	68.3
	0.987<NN<0.995	67.6±(6.3)	41	99.8
	0.995<NN<1.000	32.5±(4.2)	35	37.5
SS+	0.700<NN<0.760	3.0±(0.9)	3	55.0
	0.760<NN<0.850	3.3±(1.0)	5	25.4
	0.850<NN<0.900	1.5±(0.7)	2	43.2
	0.900<NN<0.940	0.9±(0.5)	1	56.8
	0.940<NN<0.970	1.2±(0.6)	1	65.9
	0.970<NN<0.987	1.5±(0.7)	2	43.2
	0.987<NN<0.995	0.3±(0.3)	0	74.1
	0.995<NN<1.000	0.3±(0.3)	0	74.1
SS-	0.700<NN<0.760	5.7±(1.3)	8	23.7
	0.760<NN<0.850	8.4±(1.6)	7	69.8
	0.850<NN<0.900	3.3±(1.0)	6	14.3
	0.900<NN<0.940	2.4±(0.8)	4	24.0
	0.940<NN<0.970	2.4±(0.8)	4	24.0
	0.970<NN<0.987	2.1±(0.8)	0	12.2
	0.987<NN<0.995	1.5±(0.7)	0	22.3
	0.995<NN<1.000	0.3±(0.3)	1	30.0
FM+	0.700<NN<0.760	118.3±(8.6)	136	11.1
	0.760<NN<0.850	110.5±(8.3)	121	22.3
	0.850<NN<0.900	52.0±(5.4)	37	96.3
	0.900<NN<0.940	37.3±(4.5)	37	53.0
	0.940<NN<0.970	20.1±(3.3)	20	52.3
	0.970<NN<0.987	8.3±(2.0)	6	77.1
	0.987<NN<0.995	8.7±(2.0)	3	97.5
	0.995<NN<1.000	20.8±(3.5)	24	30.7

*if zero events are observed, "Prob(N>=Nobs)" is the Poisson probability for observing exactly 0

Signal efficiency

- Signal Acceptance
- Dimuon reconstruction efficiency
- Neural Net cut efficiency
- Relative normalization to B^+

Signal efficiency

- Estimate total acceptance times efficiency for $B_s(B^0)\rightarrow \mu^+\mu^-$ decays as $\alpha_{B_s} \cdot \varepsilon_{B_s}^{total} = \alpha_{B_s} \cdot \varepsilon_{B_s}^{trig} \cdot \varepsilon_{B_s}^{reco} \cdot \varepsilon_{B_s}^{NN}$
  - α_{Bs}: geometric and kinematic acceptance of the triggered events, from MC. Trigger performance checked with data
  - ε_{trig}: trigger efficiency for events within the acceptance, from data
  - ε_{reco}: dimuon reconstruction efficiency (incl.baseline cuts) for events that pass the trigger
  - ε^{NN}: efficiency for $B_s(B^0)\rightarrow \mu^+\mu^-$ events to satisfy the NN requirement

Signal efficiency

- Estimate total acceptance times efficiency for $B_s(B^0) \to \mu^+\mu^-$ decays as $\alpha_{B_s} \cdot \varepsilon_{B_s}^{total} = \alpha_{B_s} \cdot \varepsilon_{B_s}^{trig} \cdot \varepsilon_{B_s}^{reco} \cdot \varepsilon_{B_s}^{NN}$
  - α_{Bs}: geometric and kinematic acceptance of the triggered events, from MC. Trigger performance checked with data
  - ε_{trig}: trigger efficiency for events within the acceptance, from data

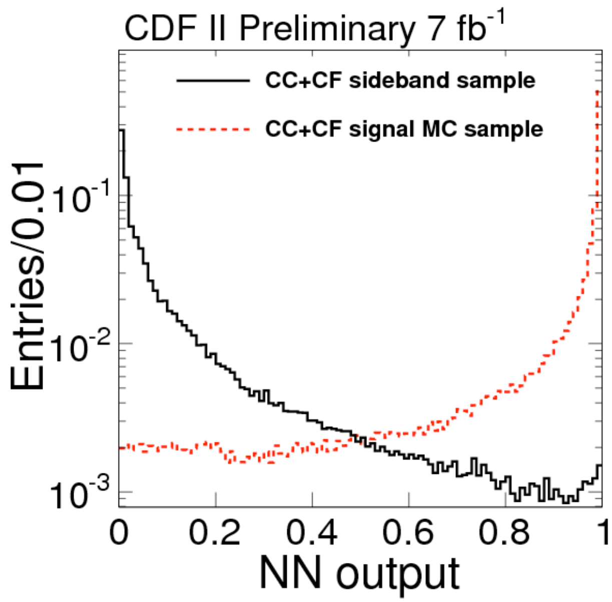
Focus for next few slides

  - ε_{reco}: dimuon reconstruction efficiency (incl.baseline cuts) for events that pass the trigger
  - ε^{NN}: efficiency for $B_s(B^0) \to \mu^+\mu^-$ events to satisfy the NN requirement

Signal efficiency

- Dimuon reconstruction efficiency ε_{reco}
  - ε_{reco} is product of
    - drift chamber track reconstruction efficiency, muon reconstruction efficiency, and vertex detector efficiency
      - Measured in the data using $J/\psi \rightarrow \mu\mu$ and D* tagged $D^0 \rightarrow K\pi$ decays
      - Muons identified using muon likelihood and track dE/dx cut

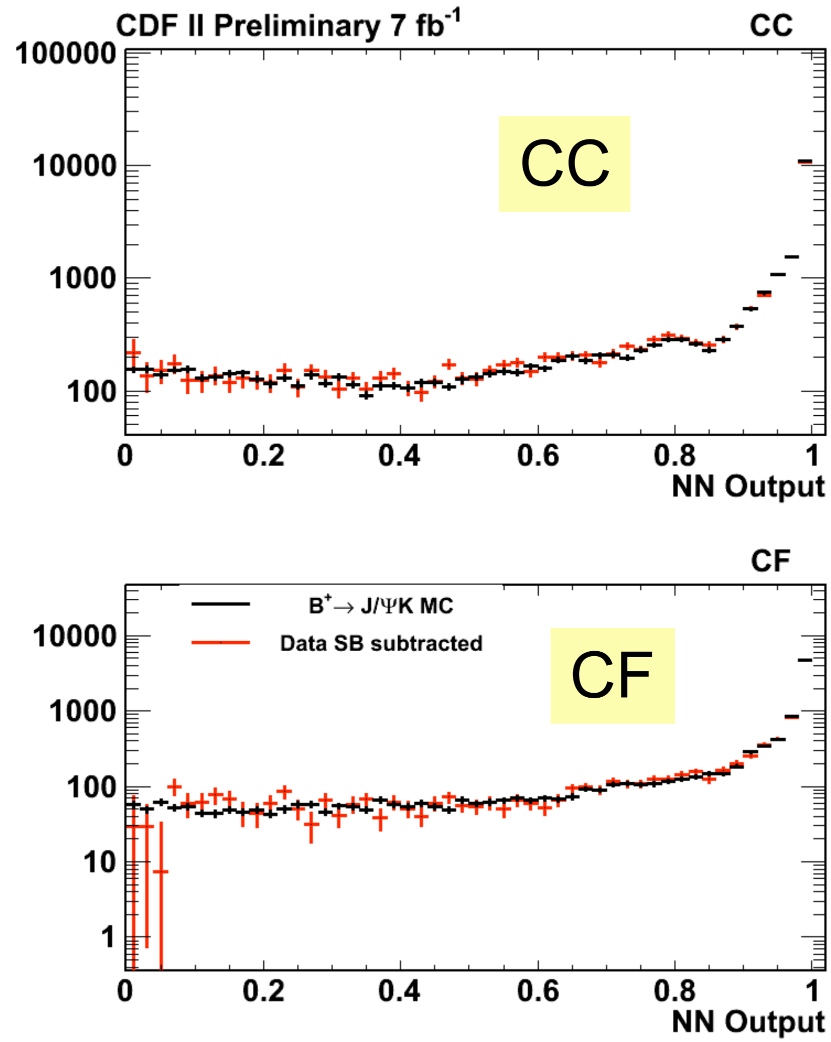
- Neural Network Cut Efficiency ε^{NN}
  - ε^{NN} extracted using signal MC
    - e.g. $\varepsilon_{NN}(NN>0.7)=0.95$
  - Systematic uncertainty based on difference between NN efficiency in B^+ MC and data
    - total: ~6% uncertainty (see next slide for detail)



Neural Network Cut Efficiencies- some more detail

- Compare the NN output efficiency between
  - B^+ MC and
  - sideband subtracted B^+ data
- average 4.6% difference is assigned as a systematic uncertainty

✓ This shows that we can accurately model our NN output efficiency



Relative normalization to $B^+ \to J/\psi K^+$

- We use $B^+ \to J/\psi K^+$ decays as a normalization mode
  - decays very similar, many systematic uncertainties cancel
- Expression for B^+ signal efficiency same as for B_s, except
  - ε_{reco} includes additional K reconstruction efficiency, excludes ε^{NN}

$$BR(B^0_{s(d)} \to \mu^+\mu^-) = \frac{N_{B_{s(d)}}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B_{s(d)}}} \cdot \frac{\varepsilon^{total}_{B^+}}{\varepsilon^{total}_{B_{s(d)}}} \cdot \frac{1}{\varepsilon^{NN}_{B_{s(d)}}} \cdot \frac{f_u}{f_s} \cdot BR(B^+ \to J/\Psi K^+ \to \mu^+\mu^- K^+)$$

Relative uncertainties (stat.+syst.) in parenthesis

	CC		CF	
$(\alpha_{B+}/\alpha_{B_s})$	0.307 ± 0.018	$(\pm 6\%)$	0.197 ± 0.014	$(\pm 7\%)$
$(\epsilon^{trig}_{B+}/\epsilon^{trig}_{B_s})$	0.99935 ± 0.00012	$(<1\%)$	0.97974 ± 0.00016	$(<1\%)$
$(\epsilon^{reco}_{B+}/\epsilon^{reco}_{B_s})$	0.85 ± 0.06	$(\pm 8\%)$	0.84 ± 0.06	$(\pm 9\%)$
$\epsilon^{NN}_{B_s}(NN > 0.70)$	0.915 ± 0.042	$(\pm 4\%)$	0.864 ± 0.040	$(\pm 4\%)$
$\epsilon^{NN}_{B_s}(NN > 0.995)$	0.461 ± 0.021	$(\pm 5\%)$	0.468 ± 0.022	$(\pm 5\%)$
N_{B+}	22388 ± 196	$(\pm 1\%)$	9943 ± 138	$(\pm 1\%)$
f_u/f_s	3.59 ± 0.37	$(\pm 13\%)$	3.59 ± 0.37	$(\pm 13\%)$
$BR(B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+)$	$(6.01 \pm 0.21) \times 10^{-5}$	$(\pm 4\%)$	$(6.01 \pm 0.21) \times 10^{-5}$	$(\pm 4\%)$
SES (All bins)	$(2.9 \pm 0.5) \times 10^{-9}$	$(\pm 18\%)$	$(4.0 \pm 0.7) \times 10^{-9}$	$(\pm 18\%)$

Single Event Sensitivity, for the sum of all NN bins, CC+CF, corresponds to an expected number of SM $B_s \to \mu^+\mu^-$ events of $N(B_s \to \mu^+\mu^-) = 1.9$ in $7 pb^{-1}$

Expected numbers of background and "SM signal" events (B_s)

CC only

NN Bin	ϵ_{NN}	B→hh Bkg	Total Bkg	Exp SM Signal
$0.700 < NN < 0.970$	20%	0.03	129.24±6.50	0.26±0.05
$0.970 < NN < 0.987$	8%	< 0.01	7.91±1.27	0.11±0.02
$0.987 < NN < 0.995$	12%	0.02	3.95±0.89	0.16±0.03
$0.995 < NN < 1.000$	46%	0.08	0.79±0.40	0.59±0.11

CF only

NN Bin	ϵ_{NN}	B→hh Bkg	Total Bkg	Exp SM Signal
$0.700 < NN < 0.970$	21%	0.01	146.29±7.00	0.19±0.04
$0.970 < NN < 0.987$	10%	0.01	11.57±1.57	0.09±0.02
$0.987 < NN < 0.995$	8%	0.01	3.25±0.82	0.08±0.01
$0.995 < NN < 1.000$	46%	0.03	2.64±0.74	0.43±0.08

NN signal efficiency

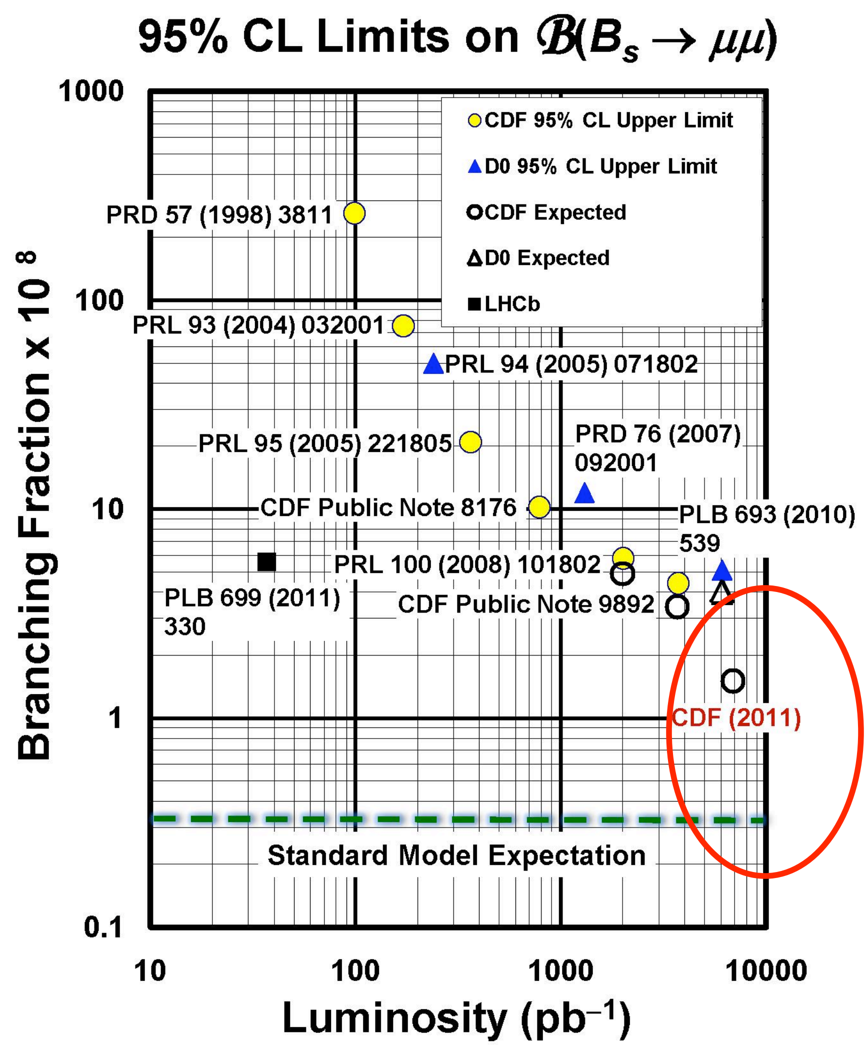
Expected limits

$BR(B_s \to \mu^+\mu^-) < 1.5\times10^{-8}$ @ 95%CL
$BR(B^0 \to \mu^+\mu^-) < 4.6\times10^{-9}$ @ 95%CL

Significant improvement in sensitivity over all previous analyses

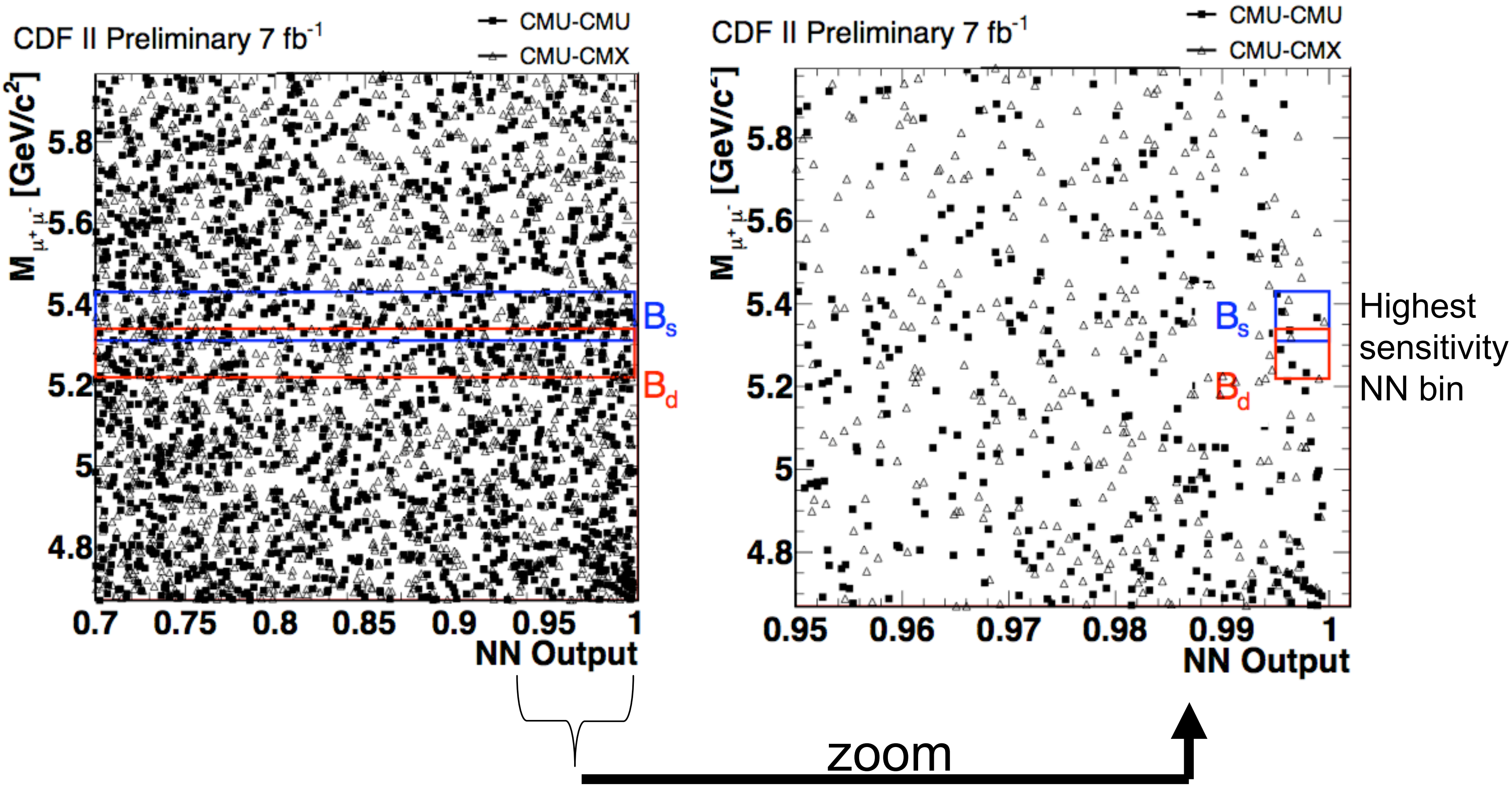
For $BR(B_s \to \mu^+\mu^-)$:

	Expected	Observed
2.0 fb^{-1}	: 4.9×10^{-8}	5.8×10^{-8}
3.7 fb^{-1}	: 3.4×10^{-8}	4.4×10^{-8}
7 fb^{-1}	: 1.5×10^{-8}	

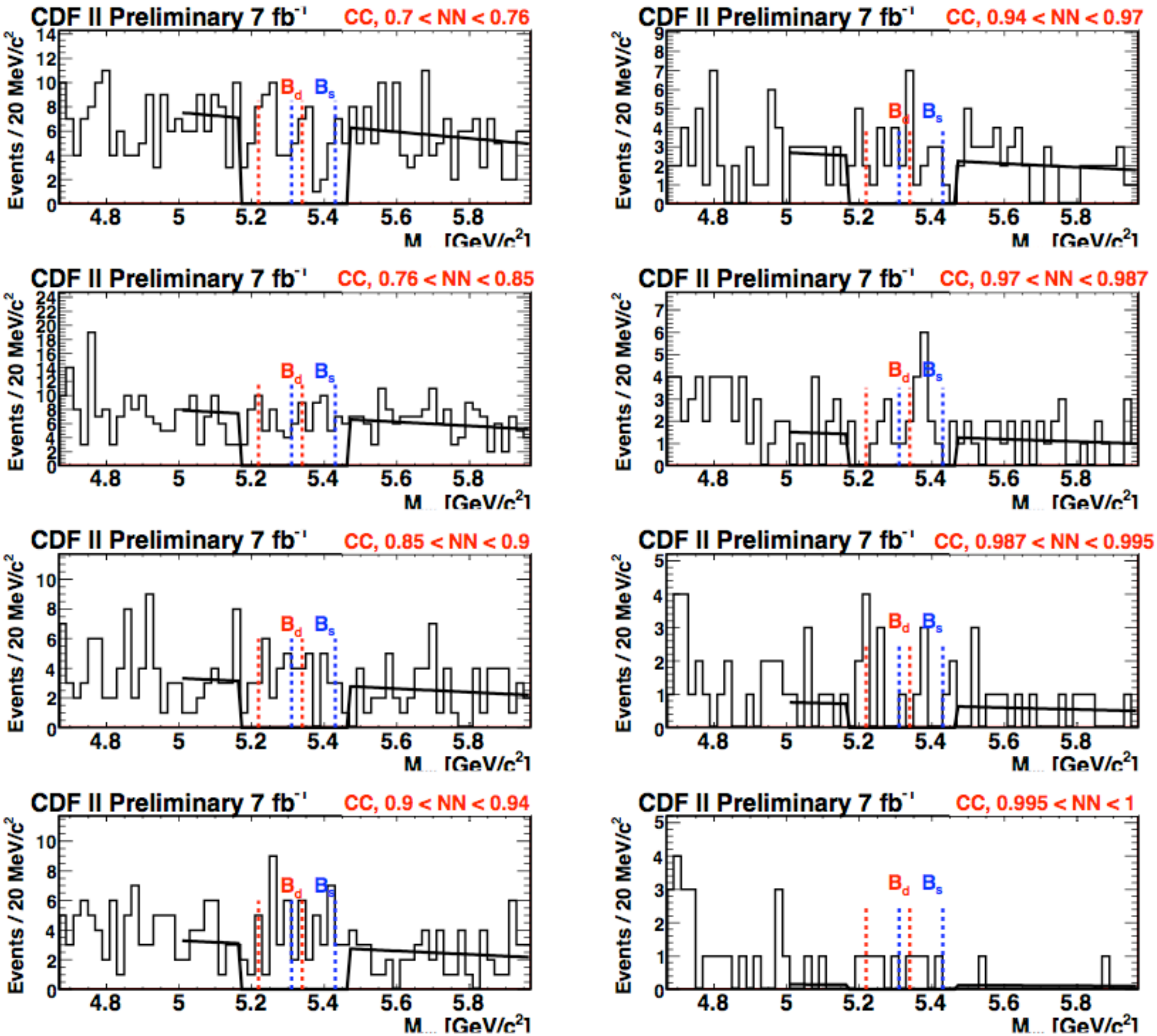


Opening “the box”

$B_s \to \mu^+\mu^-$ search: opening the box

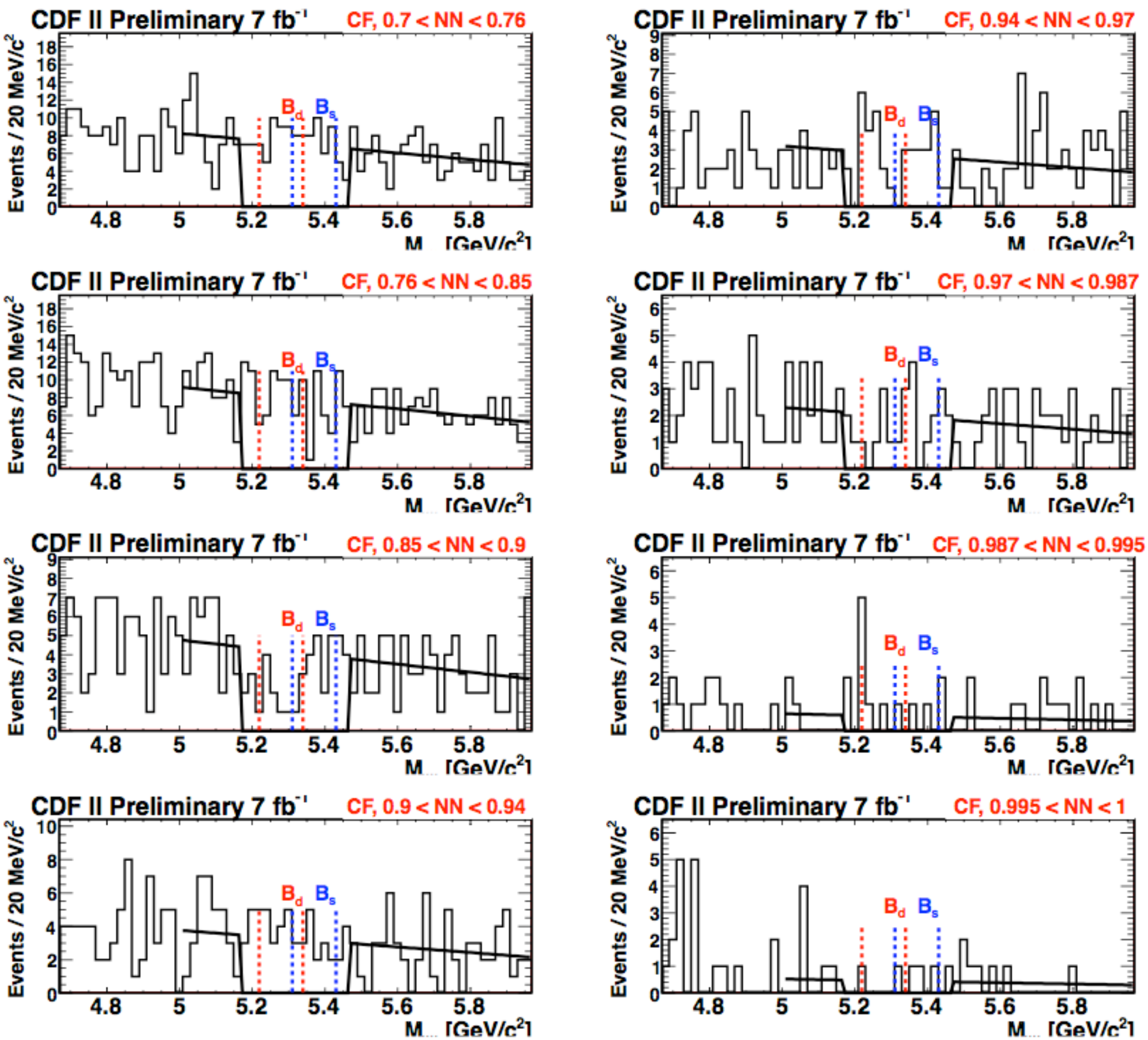


CC only



CDF II Preliminary 7 fb⁻¹
CC, 0.7 < NN < 0.76
CC, 0.94 < NN < 0.97
CC, 0.76 < NN < 0.85
CC, 0.97 < NN < 0.987
CC, 0.85 < NN < 0.9
CC, 0.987 < NN < 0.995
CC, 0.9 < NN < 0.94
CC, 0.995 < NN < 1
B_d
B_s
Events / 20 MeV/c²
M [GeV/c²]

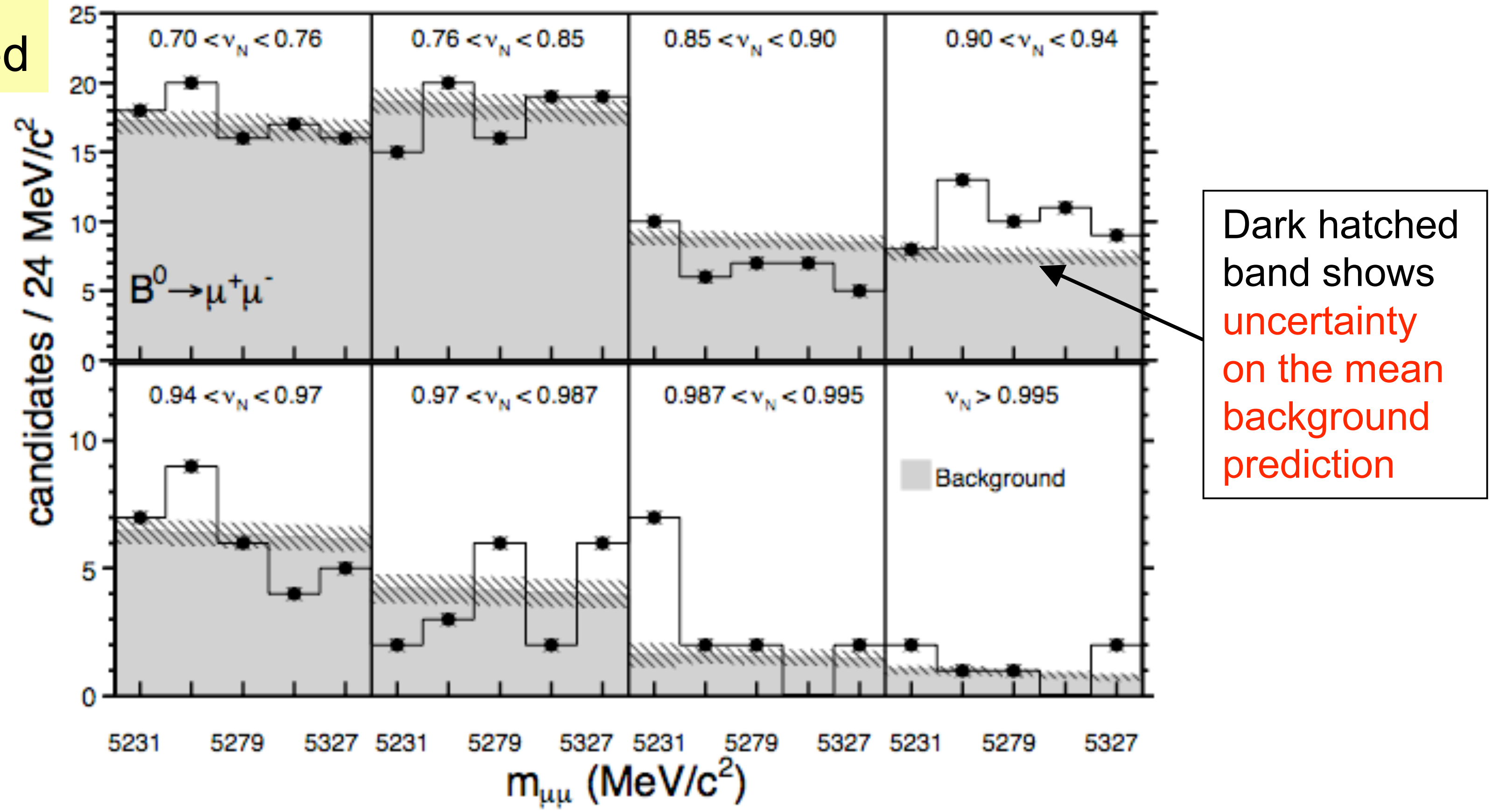
CF only


CDF II Preliminary 7 fb^{-1}
CF, 0.7 < NN < 0.76
CDF II Preliminary 7 fb^{-1}
CF, 0.94 < NN < 0.97
CDF II Preliminary 7 fb^{-1}
CF, 0.76 < NN < 0.85
CDF II Preliminary 7 fb^{-1}
CF, 0.97 < NN < 0.987
CDF II Preliminary 7 fb^{-1}
CF, 0.85 < NN < 0.9
CDF II Preliminary 7 fb^{-1}
CF, 0.987 < NN < 0.995
CDF II Preliminary 7 fb^{-1}
CF, 0.9 < NN < 0.94
CDF II Preliminary 7 fb^{-1}
CF, 0.995 < NN < 1
B_d
B_s
Events / 20 MeV/c^2
M [GeV/c^2]

Focus on B^0 signal window first

B^0 signal window, comparison of observation and background prediction

CC, CF combined

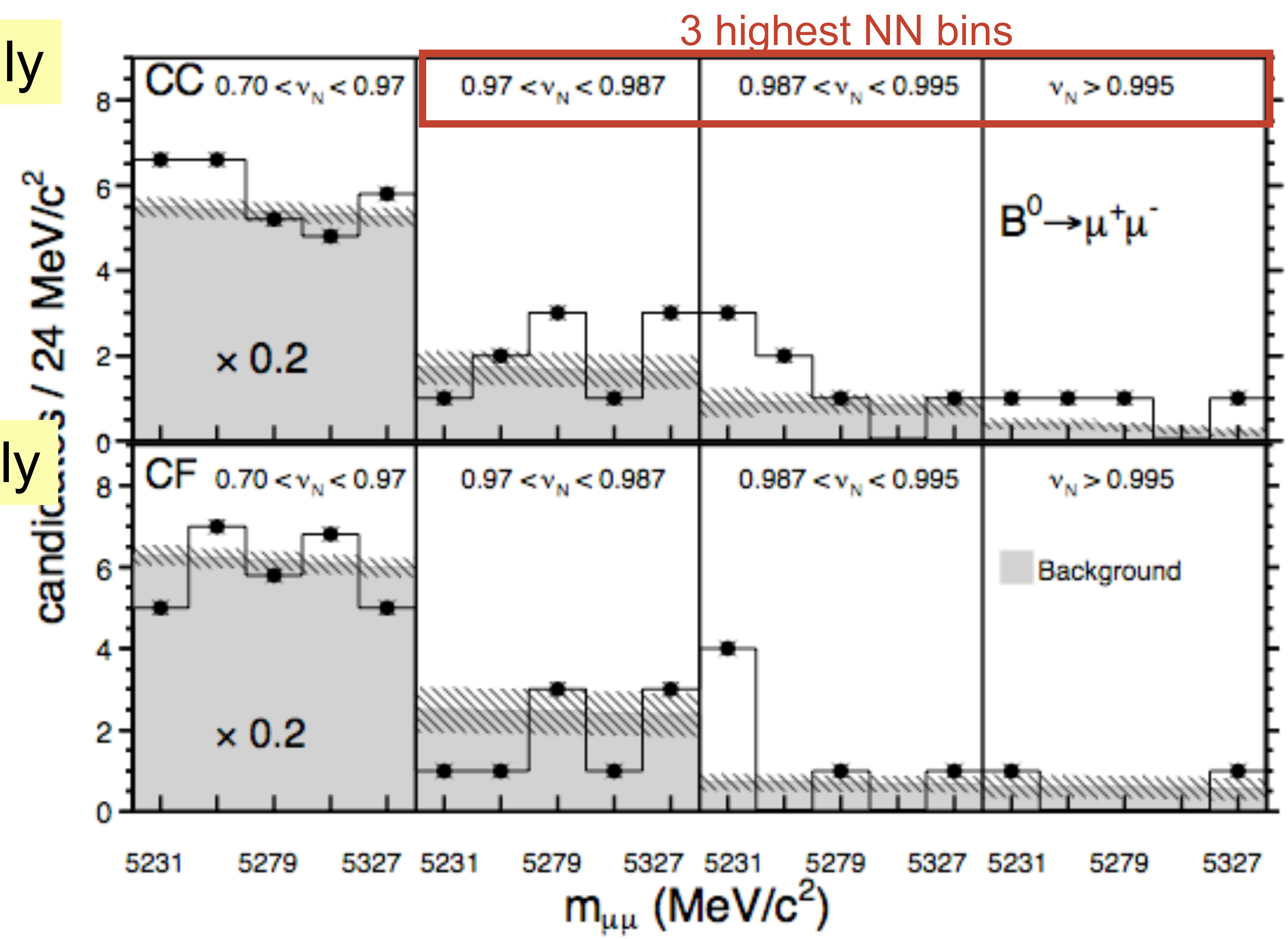


Data and background expectation are in good agreement

B^0 signal window, comparison of observation and background prediction

CC only

CF only



B^0 signal window, comparison of observation and background prediction

3 most sensitive NN bins only

CC only

CC		Mass bins [GeV/c^2]				
NN Bins		5.219-5.243	5.243-5.267	5.267-5.291	5.291-5.315	5.315-5.339
$0.970<NN<0.987$	Exp	3.00 ± 0.65	2.97 ± 0.64	2.93 ± 0.64	2.90 ± 0.63	2.86 ± 0.62
	Obs	2	3	4	3	4
$0.987<NN<0.995$	Exp	0.90 ± 0.28	0.89 ± 0.28	0.86 ± 0.27	0.84 ± 0.27	0.81 ± 0.27
	Obs	3	2	1	0	1
$0.995<NN<1.000$	Exp	0.40 ± 0.21	0.38 ± 0.20	0.32 ± 0.17	0.25 ± 0.15	0.20 ± 0.14
	Obs	1	1	1	0	1

CF only

CF						
$0.970<NN<0.987$	Exp	2.50 ± 0.59	2.47 ± 0.58	2.44 ± 0.58	2.40 ± 0.57	2.37 ± 0.56
	Obs	1	4	3	1	2
$0.987<NN<0.995$	Exp	0.71 ± 0.25	0.70 ± 0.25	0.69 ± 0.25	0.68 ± 0.24	0.67 ± 0.24
	Obs	4	0	1	0	1
$0.995<NN<1.000$	Exp	0.62 ± 0.42	0.62 ± 0.42	0.60 ± 0.41	0.57 ± 0.40	0.55 ± 0.39
	Obs	1	0	0	0	1

Data and background expectation are in good agreement

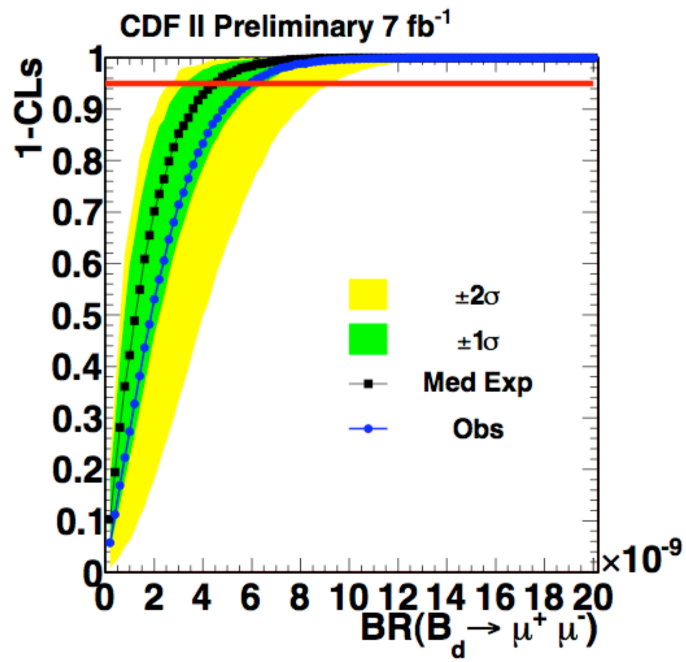
$B^0 \rightarrow \mu^+\mu^-$ search, observed limit

We set a limit (using CLs method) of

$$BR(B^0 \rightarrow \mu^+\mu^-) < 6.0 \times 10^{-9}$$

at 95% C.L.

- world's best limit
- consistent with the expected limit $BR(B^0 \rightarrow \mu^+\mu^-) < 4.6 \times 10^{-9}$



Compare to the SM BR calculation of

$$BR(B^0 \rightarrow \mu^+\mu^-) = (1.0 \pm 0.1) \times 10^{-10}$$

Determination of the p-value

Ensemble of background-only pseudo-experiments is used to determine a p-value for a given hypothesis

- for each pseudo-experiment, we do two fits and form the log-likelihood ratio

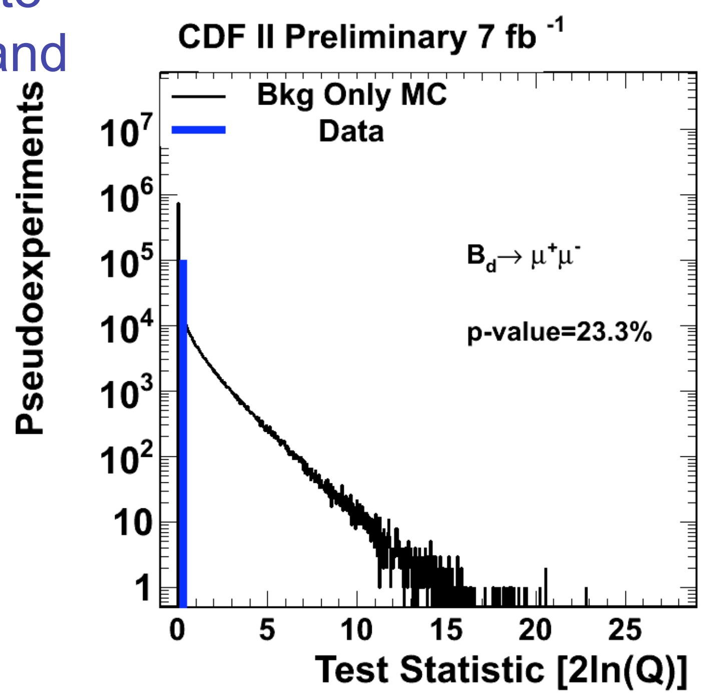
$$2\ln(Q) \quad \text{with} \quad Q = \frac{L(s+b \mid data)}{L(b \mid data)}$$

- in the denominator, the “signal” is fixed to zero (I.e. we assume background only), and in the numerator s floats

- *L(h|x)* is the product of Poisson probabilities over all NN and mass bins

- systematic uncertainties included as nuisance parameters, modeled as Gaussian.

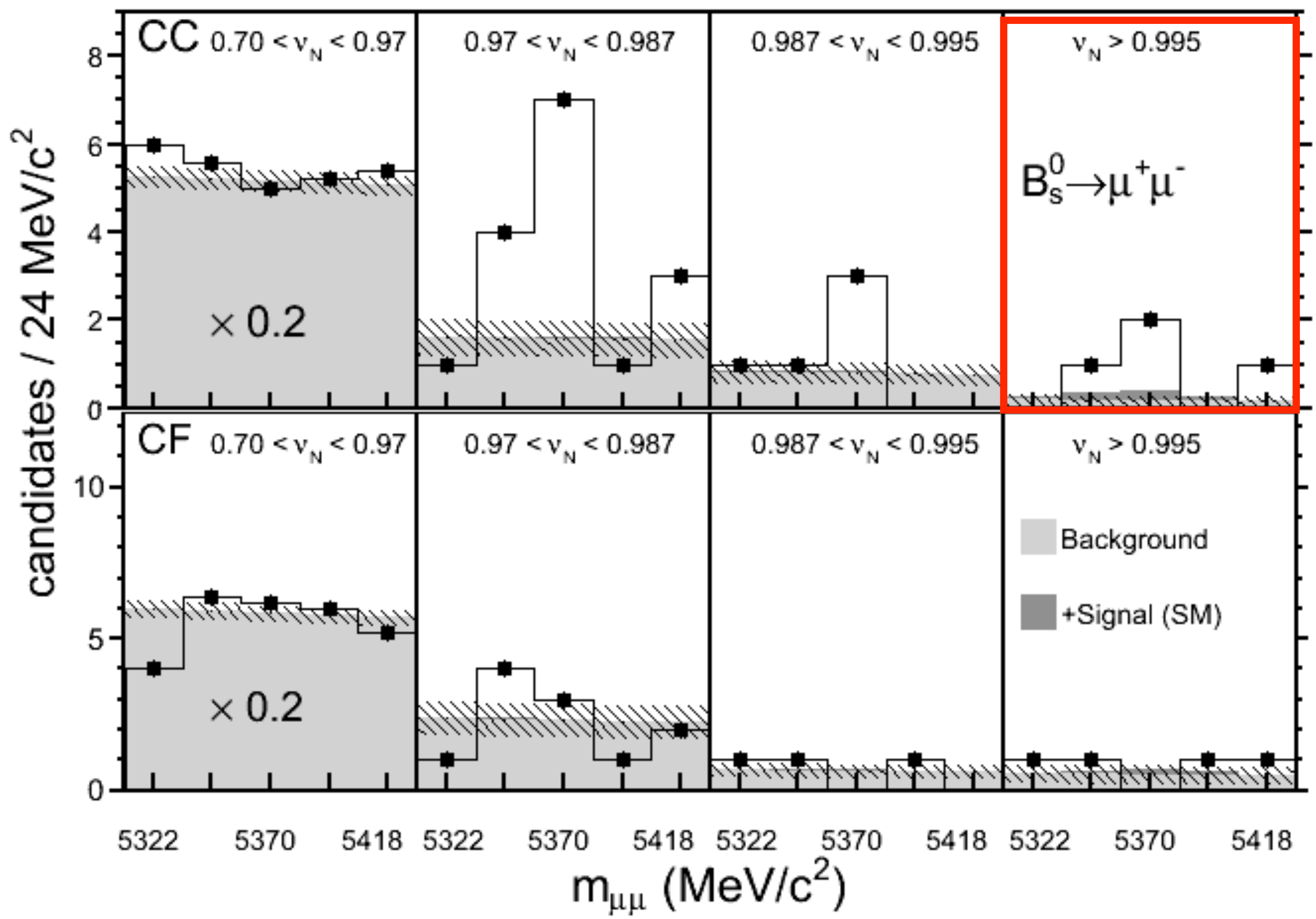
Log Likelihood Distribution of pseudo-experiments for background-only hypothesis for $B^0 \rightarrow \mu^+\mu^-$ signal window



Result: the p-value for the background-only hypothesis is 23.3%

B_s signal window

Data in B_s signal window



B_s signal window, comparison of observation and background prediction

Shown is the total expected background and total uncertainty, as well as number of observed events

	CC	Mass bins [GeV/c^2]				
NN Bins		5.310-5.334	5.334-5.358	5.358-5.382	5.382-5.406	5.406-5.430
0.970<NN<0.987	Exp	1.62±0.49	1.6±0.48	1.58±0.47	1.57±0.47	1.55±0.46
	Obs	1	4	7	1	3
0.987<NN<0.995	Exp	0.82±0.27	0.8±0.27	0.79±0.26	0.78±0.26	0.78±0.26
	Obs	1	1	3	0	0
0.995<NN<1.000	Exp	0.21±0.14	0.18±0.13	0.16±0.12	0.16±0.12	0.16±0.12
	Obs	0	1	2	0	1
	CF					
0.970<NN<0.987	Exp	2.38±0.56	2.34±0.55	2.31±0.54	2.28±0.54	2.25±0.53
	Obs	1	4	3	1	2
0.987<NN<0.995	Exp	0.67±0.24	0.66±0.24	0.65±0.24	0.64±0.23	0.63±0.22
	Obs	1	1	0	1	0
0.995<NN<1.000	Exp	0.56±0.39	0.54±0.38	0.53±0.38	0.52±0.37	0.51±0.36
	Obs	1	1	0	1	1

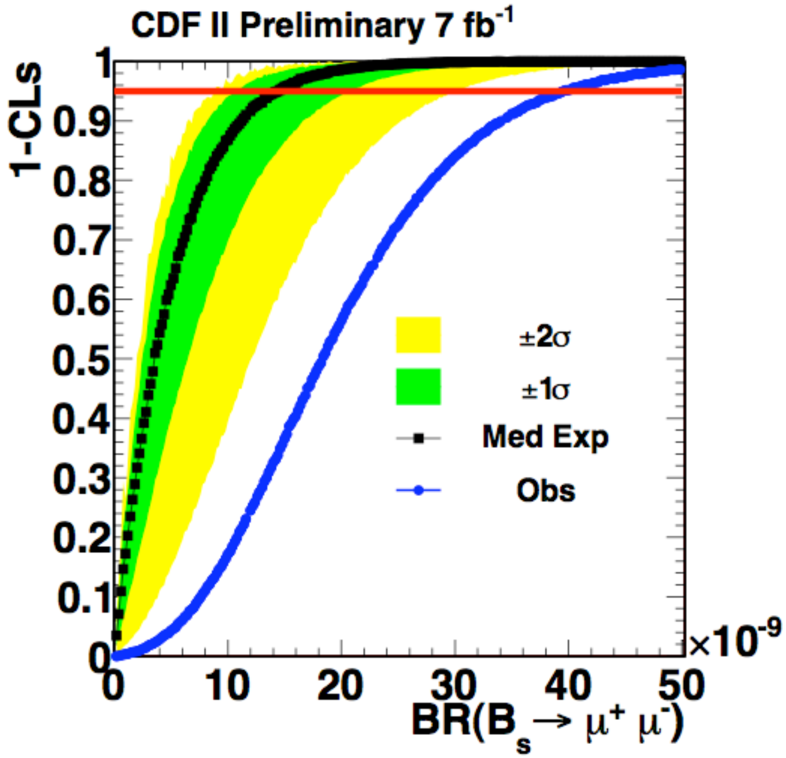
Observe an excess, concentrated in the 3 highest NN bins of the CC sample, over background expectation

$B_s \to \mu^+\mu^-$ search, observed limit

Using the CLs method, we observe

$BR(B_s \to \mu^+\mu^-) < 4.0\times10^{-8}$

at 95% C.L.



- Compare to the expected limit $BR(B^0 \to \mu^+\mu^-) < 1.5\times10^{-8}$

- outside the 2σ consistency band

Need statistical interpretation of the observed excess:

- what is the level of inconsistency with the background?
- what does a fit to the data in the B_s search window yield?

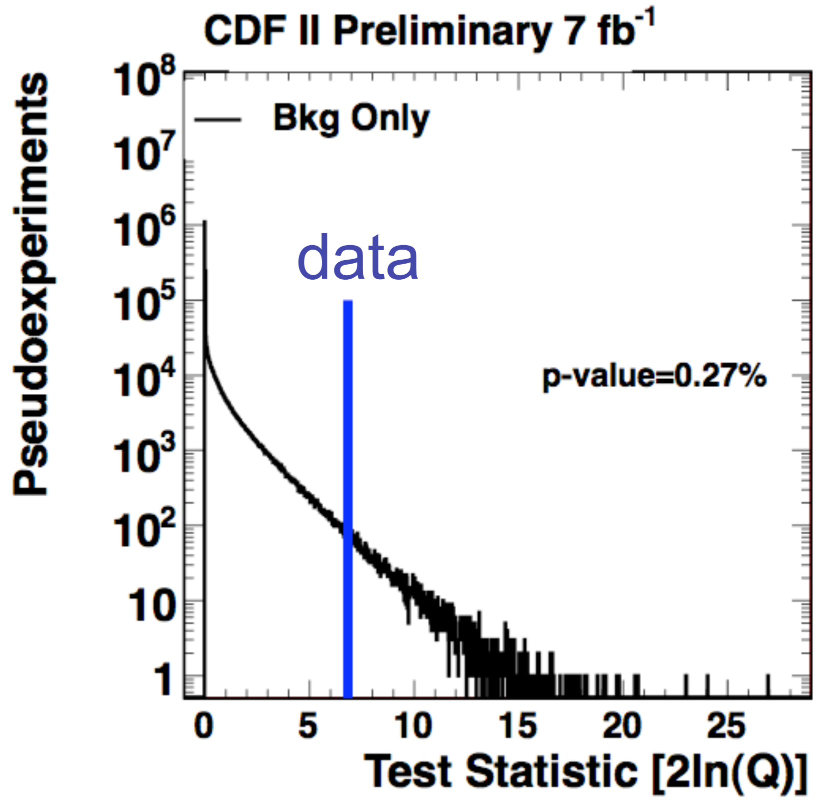
Statistical Interpretation

P value for background-only hypothesis

Log Likelihood Distribution
of pseudo-experiments
for background hypothesis

Observed p-value: 0.27%.

This corresponds to a 2.8σ discrepancy with a background-only null hypothesis (one-sided gaussian)



Fit to the data in the B_s search window

Using the log-likelihood fit described before, we set the first two-sided limit of $B_s \to \mu^+\mu^-$ decay

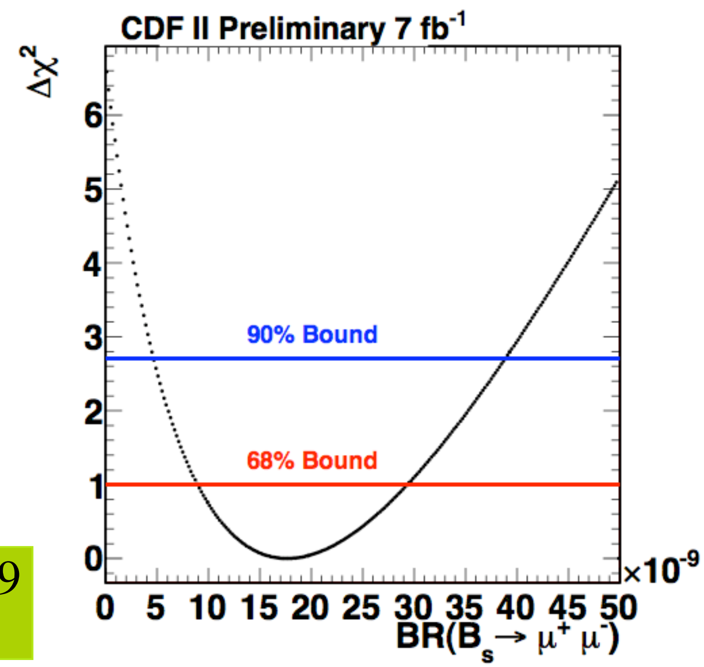
$$4.6 \times 10^{-9} < BR(B_s \to \mu^+\mu^-) < 3.9 \times 10^{-8}$$ @90% C.L.

Our central value is

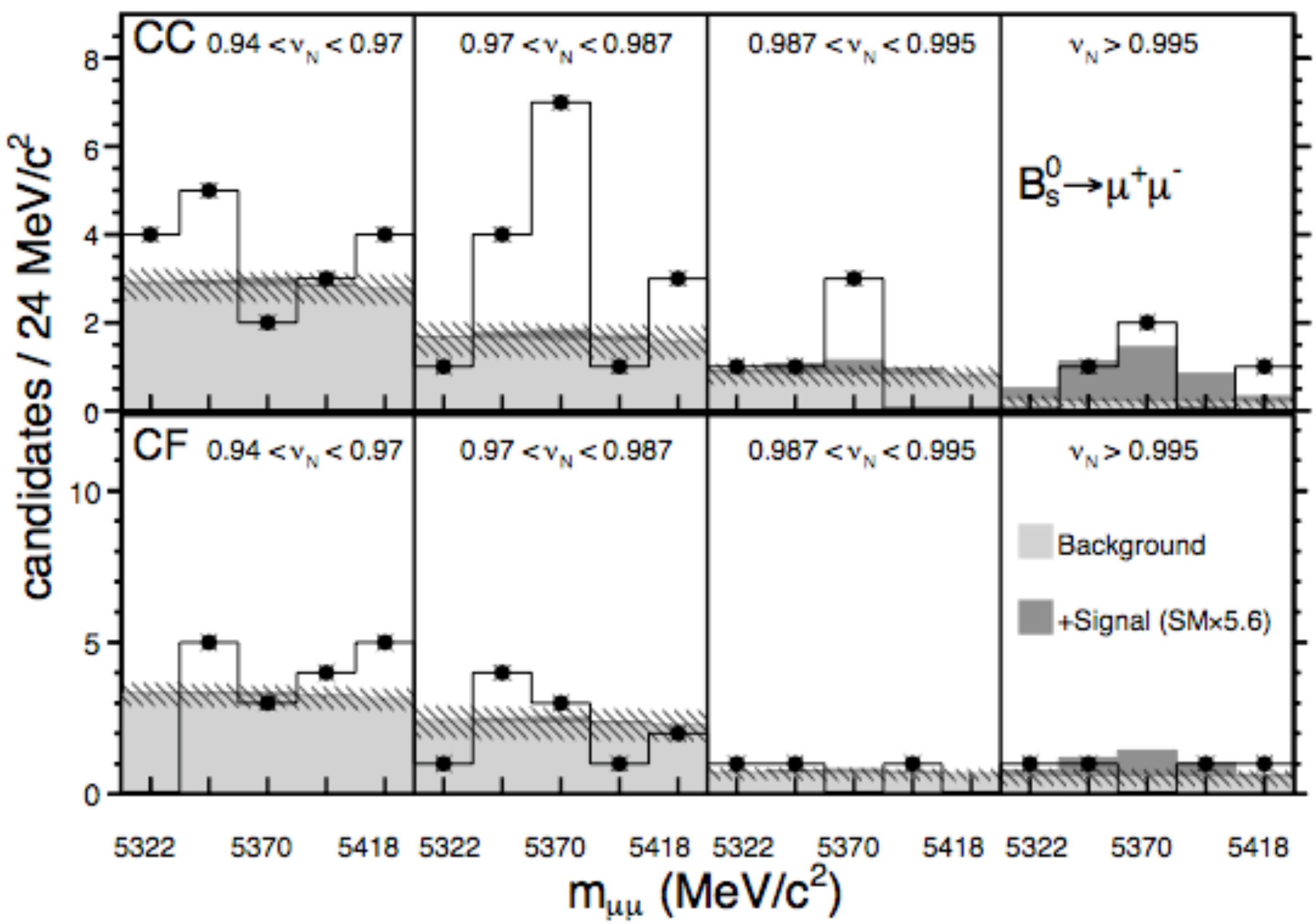
$$BR(B_s \to \mu^+\mu^-) = 1.8^{+1.1}_{-0.9} \times 10^{-8}$$

Compare to SM calculation of

$$BR(B_s \to \mu^+\mu^-) = (3.2 \pm 0.2) \times 10^{-9}$$



Data in B_s signal window

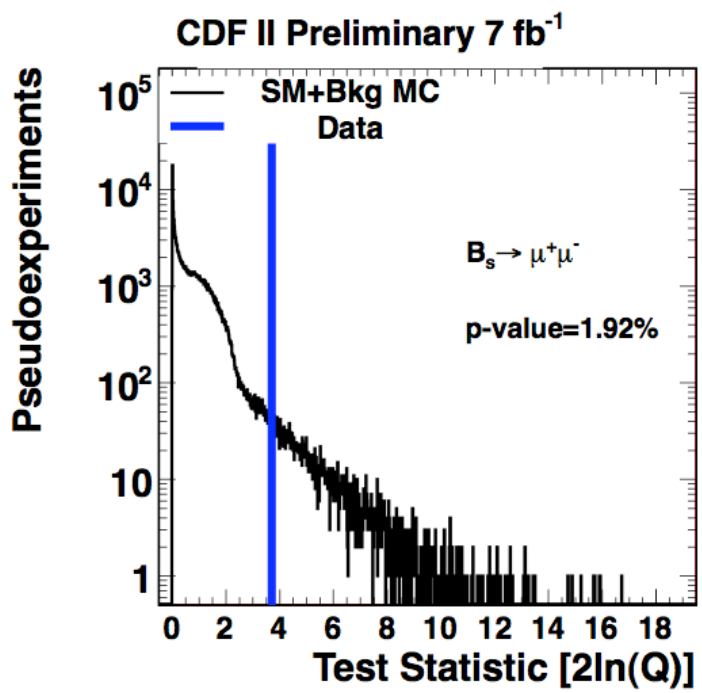


Consistency with the SM prediction of $B_s \to \mu^+\mu^-$ decays

reminder: SM prediction: BR($B_s \to \mu^+\mu^-$)=$(3.2\pm0.2)\times10^{-9}$
A. J. Buras et al., JHEP 1010:009,2010

If we include the SM BR($B_s \to \mu^+\mu^-$) in the background hypothesis, we observe a p-value of 1.9%

"Background" hypothesis now includes the SM expectation of BR($B_s \to \mu\mu$)



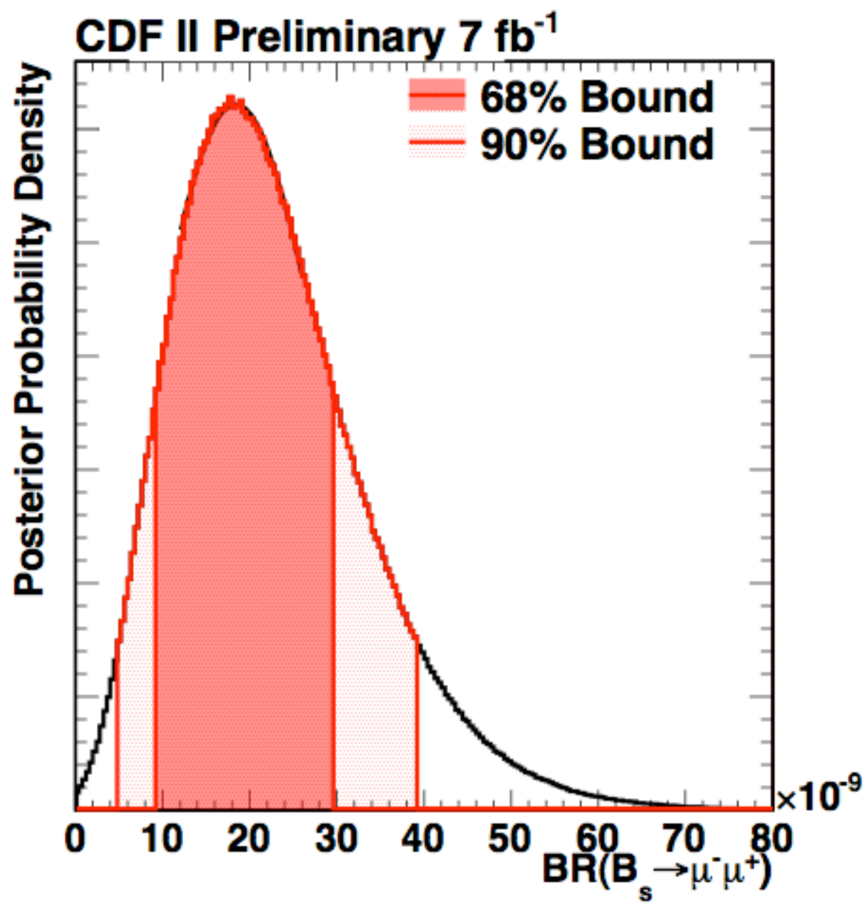
taking into account the small theoretical uncertainty on the SM prediction by assuming +1σ: p-value: 2.1%

Cross Checks

Fit to the data: cross checks

Use Bayesian binned likelihood technique

- assumes a flat prior for BR>0
- integrates over all sources of systematic uncertainty assuming gaussian priors
- best fit value taken at maximum, uncertainty taken as shortest interval containing 68% of the integral.



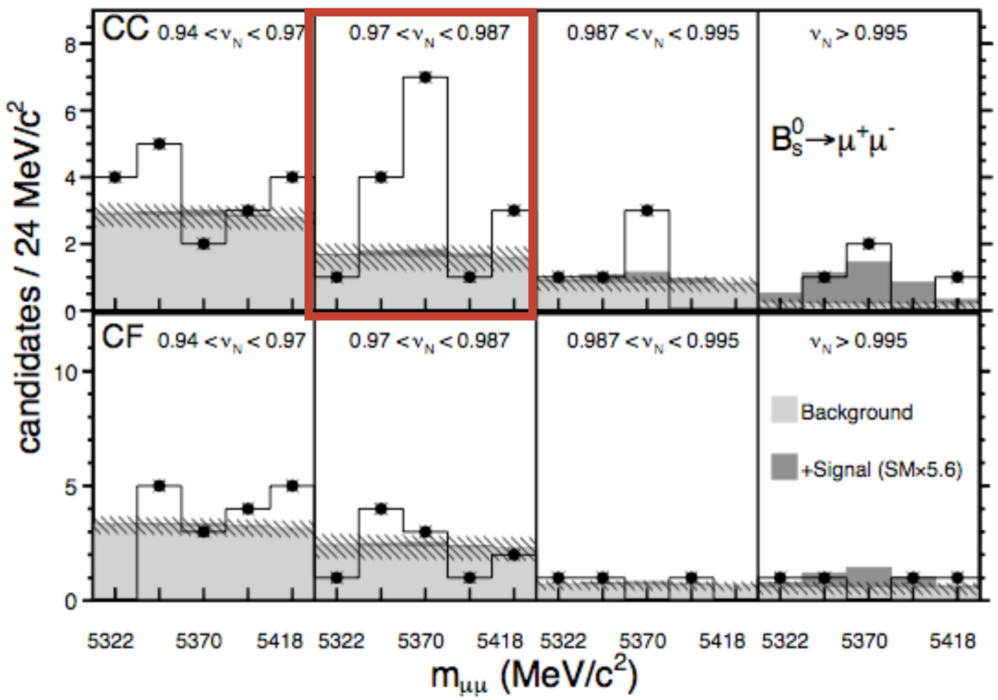
Best fit to the data yields almost identical results as before

$$BR(B_s \to \mu^+\mu^-) = 1.8^{+1.1}_{-0.9} \times 10^{-8}$$

A closer look at the data

- excess observed in CC muons
- in most sensitive NN bin: data looks signal-like
- see a fluctuation in 0.97<NN<0.987- little signal sensitivity in this bin.

B_s signal window, CC and CF separate
Showing only the most sensitive 4 highest NN bins



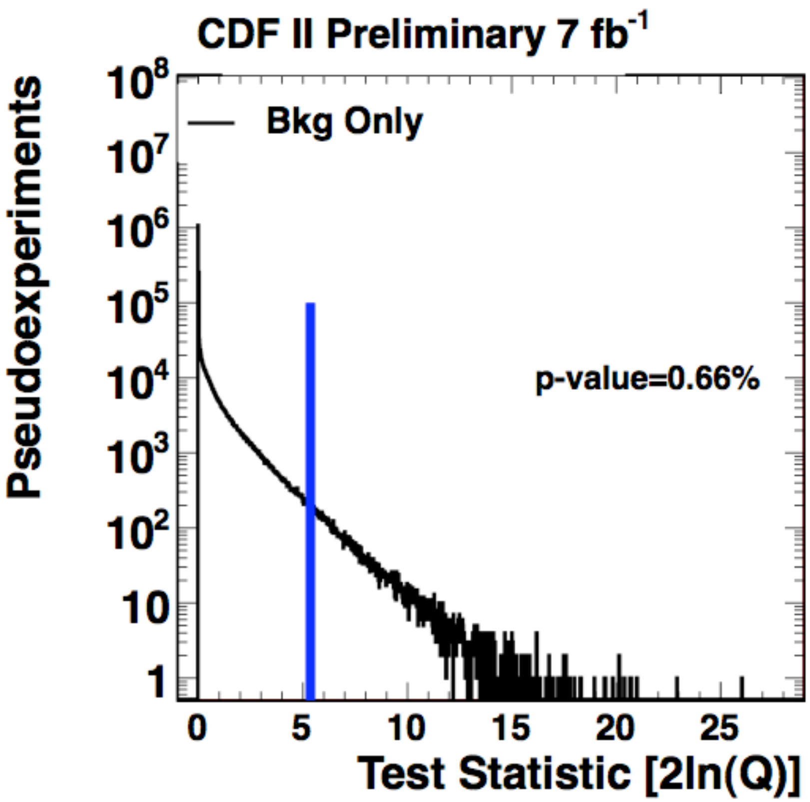
Does the fluctuation in this bin drive the result?
Check how the answer changes if we only look at the two highest NN bins..

Fit to the data, only considering the 2 highest NN bins

- Background-only hypothesis:
  Observed p-value: 0.66%
  (compare to 0.27%)

- Background + SM hypothesis:
  Observed p-value: 4.1%
  (compare to 1.9%)

- Conclusion: “fluctuation” in the lower sensitivity bin adds to the observed discrepancy, but is not the driving contribution.



Residual B → hh background

The number of residual B → hh events are very small. E.g. for the highest NN bins:

	CC	CF
B_s signal window	0.08±0.2	0.03±0.01
B^0 signal window	0.72±0.2	0.2±0.05

Factor 10 higher contribution in B^0 signal window because B → hh peaks closer to the B^0 mass

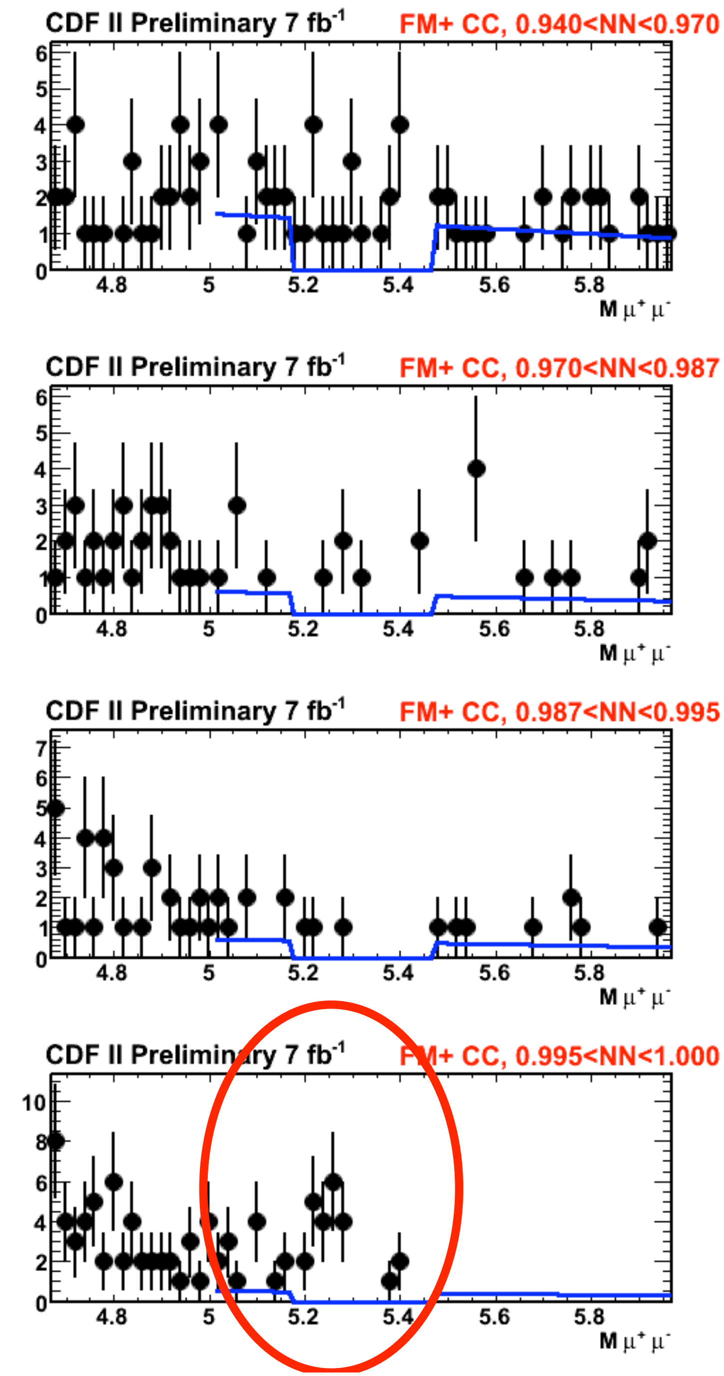
- and we see no excess over the prediction in the B^0 signal window

We carefully checked our predictions in a control region enhanced in B → hh decays (FM+ sample, at least one "muon" has to fail our muon selection)

	Predicted total events	observed	Prob.(%)
$0.700<NN<0.760$	$118.3\pm(8.6)$	136	11.1
$0.760<NN<0.850$	$110.5\pm(8.3)$	121	22.3
$0.850<NN<0.900$	$52.0\pm(5.4)$	37	96.3
$0.900<NN<0.940$	$37.3\pm(4.5)$	37	53.0
$0.940<NN<0.970$	$20.1\pm(3.3)$	20	52.3
$0.970<NN<0.987$	$8.3\pm(2.0)$	6	77.1
$0.987<NN<0.995$	$8.7\pm(2.0)$	3	97.5
$0.995<NN<1.000$	$20.8\pm(3.5)$	24	30.7

Observation in FM+ sample, highest NN bins

In our highest NN bin we clearly select B → hh and can predict it accurately with our background estimate method.



Summary- Cross checks

We have performed cross checks (some shown in the backup slides) to confirm that

✓The results are stable w.r.t. variations in error shape assumptions

- have compared poisson to gaussian statistics for shapes of systematic uncertainties

✓The results are independent of the statistical treatment

- we get the same answers using Bayesian and Likelihood fit

✓The results are not driven by a fluctuation that is observed in the 3rd highest NN bin

- somewhat smaller significance when the 3rd highest NN bin is excluded

✓The excess is not from $B \rightarrow hh$

- 0.08 residual events, carefully checked modeling

Conclusions

We see an excess over the background-only expectation in the B_s signal region and have set the first two-sided bounds on $BR(B_s \rightarrow \mu^+\mu^-)$

$$4.6 \times 10^{-9} < BR(B_s \rightarrow \mu^+\mu^-) < 3.9 \times 10^{-8}$$ at 90% C.L.

A fit to the data, including all uncertainties, yields

$$BR(B_s \rightarrow \mu^+\mu^-) = 1.8^{+1.1}_{-0.9} \times 10^{-8}$$

Data in the B^0 search window are consistent with background expectation, and the world's best limit is extracted:.

$$BR(B^0 \rightarrow \mu^+\mu^-) < 6.0(5.0) \times 10^{-9}\, at 95\%(90\%) C.L.$$

Conclusions

• Although of moderate statistical significance, this may be the first sign of a $B_s \to \mu^+\mu^-$ signal

- great interest in this decay because of its excellent sensitivity to NP

• Maybe the first glimpse of exciting times ahead?

Archive: http://arxiv.org/abs/1107.2304, Fermilab-Pub-11-315-E

Public web page:
/cdf/www/physics/new/bottom/110707.blessed-Bsd2mumu

Since we posted our result on Tuesday, we've had a lot of feedback from Theorists (see next slides)

First email we received:

“Yeeeeeeeeeeees! Just as I predicted!”

Interpretation in an mSUGRA model

$m_0/m_{1/2}$ plane in a mSUGRA interpretation with $\tan\beta=50$

B. Dutta, Y.Mimura, Y. Santoso

Green: Region preferred by $B_s \to \mu^+\mu^-$ 90% range

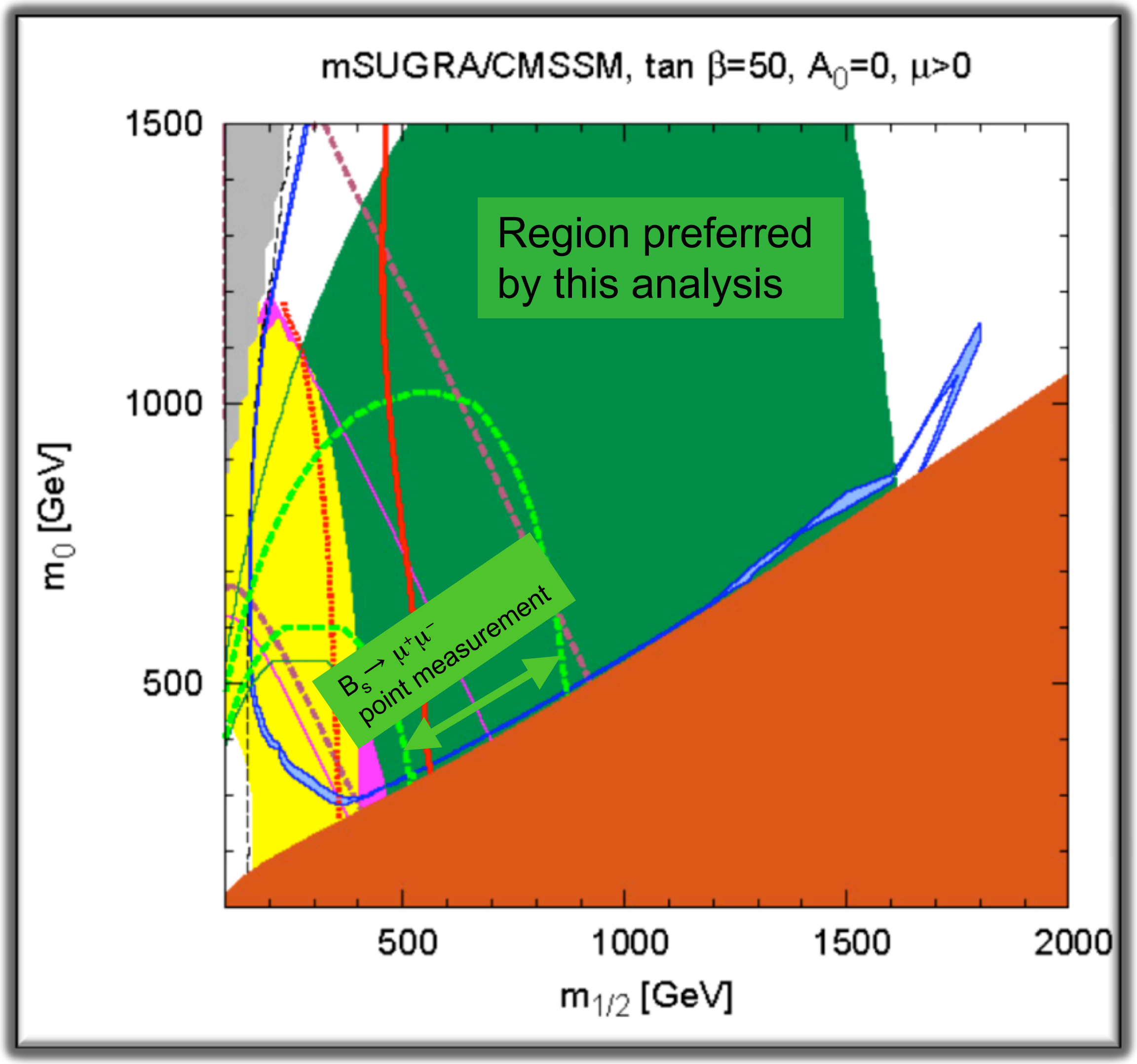
Dashed green: point measurement

Excluded by

a Rare B decay $b \to s\gamma$

b No CDM candidate

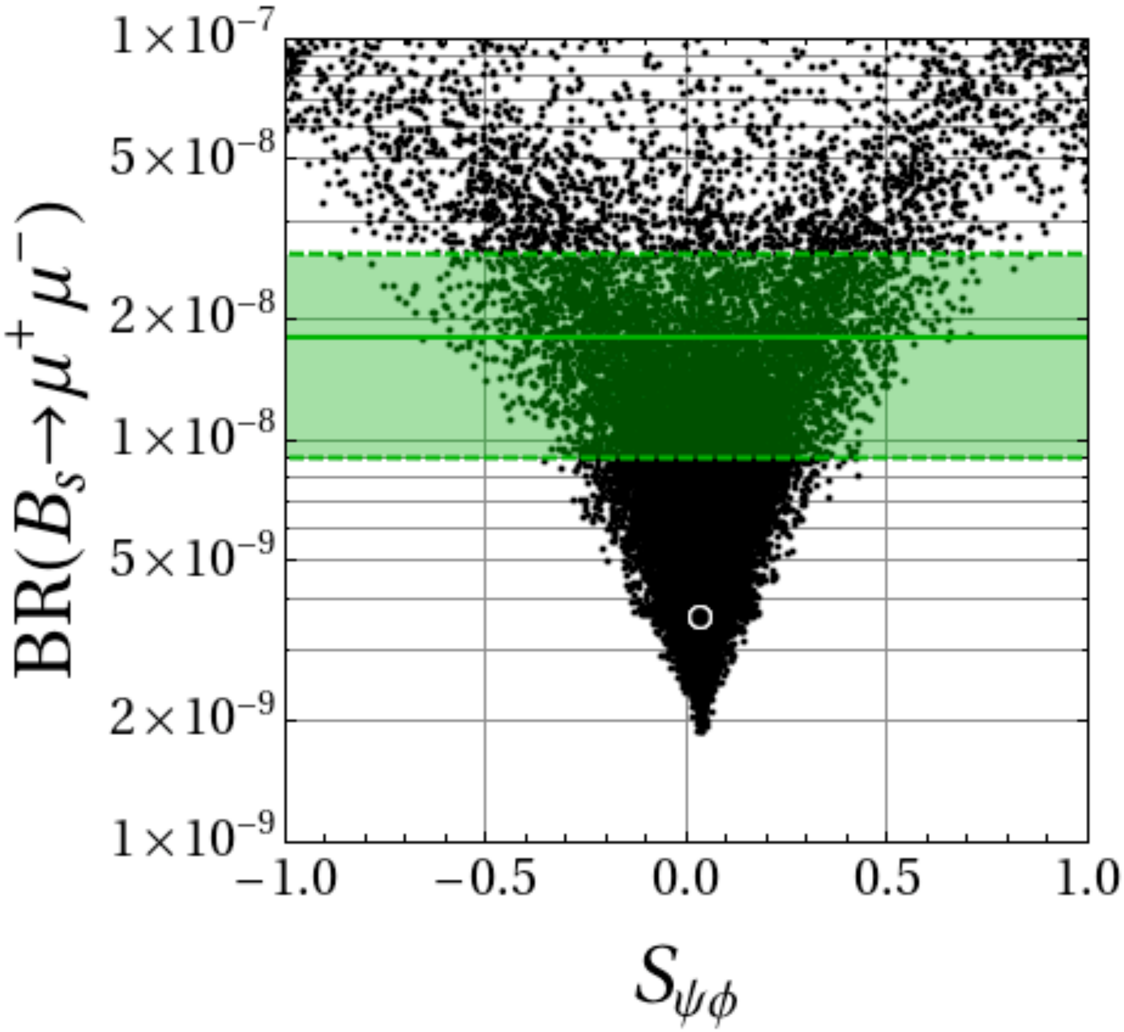
c No EWSB



More Interpretations

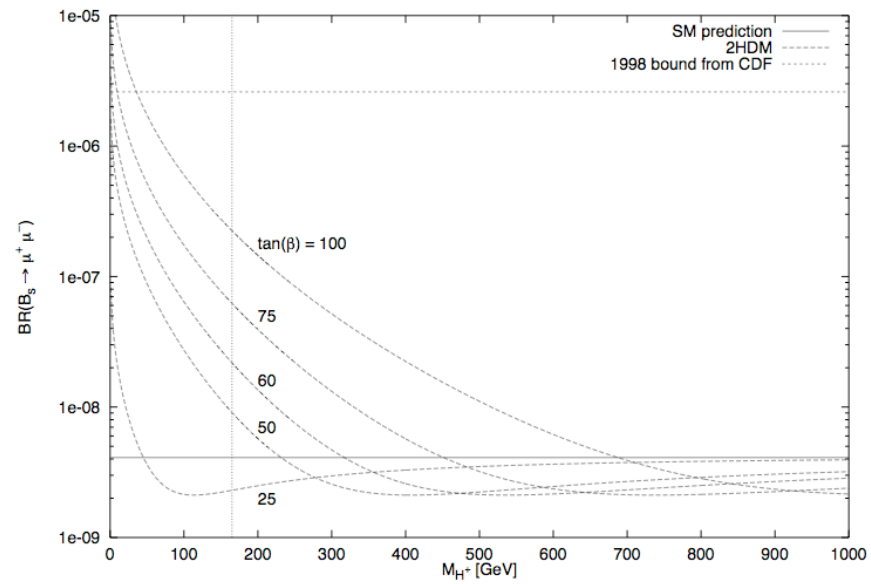
Updated plot from
Altmannshofer, Buras, Gori, Paradisi, Straub, Nucl.Phys.B830:17-94,2010 (arXiv:0909.1333)

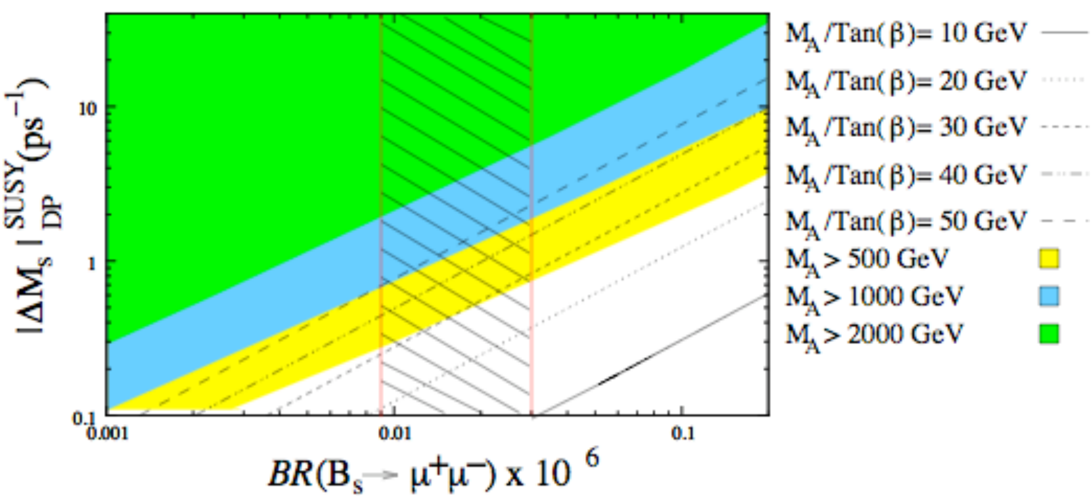
correlation between $BR(B_s \rightarrow \mu^+\mu^-)$ and the CP violating phase in Bs mixing in a SUSY model from *Agashe, Carone Phys.Rev. D68 (2003) 035017 (hep-ph/0304229).*



More Interpretations

U.Nierste, H.Logan: New two-sided limit excludes a significant portion of the allowed parameter space for tanβ and M_{H+} in a two-Higgs Doublet model:





C.Wagner, M. Carena, A.Menon, A. Szynkman and R. Noriega: correlation between $B_s \to \mu^+\mu^-$ and the SUSY contributions to ΔM_s in the MSSM with minimal flavor violation

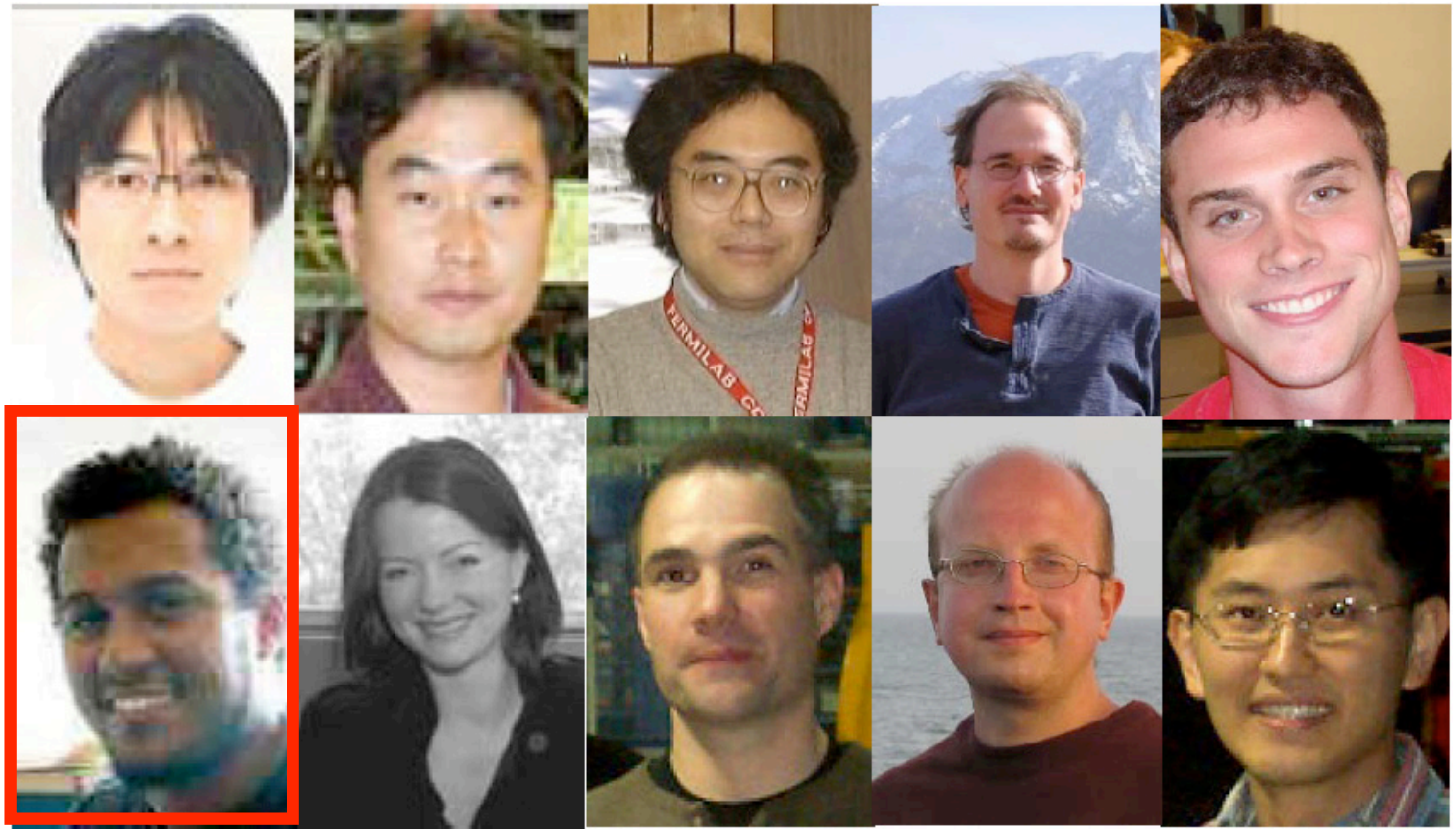
Acknowledgements

Many people have contributed to this result-
the Fermilab and Tevatron staff, many CDF collaborators

and we thank our Theorist colleagues for extremely useful discussion:
A. Buras, U. Nierste, S. Gori, C. Wagner, G. Hou, A. Soni, L. Roskowski, T. Hurth, W. Altmannshofer, C.Davies, and others.

Acknowledgements

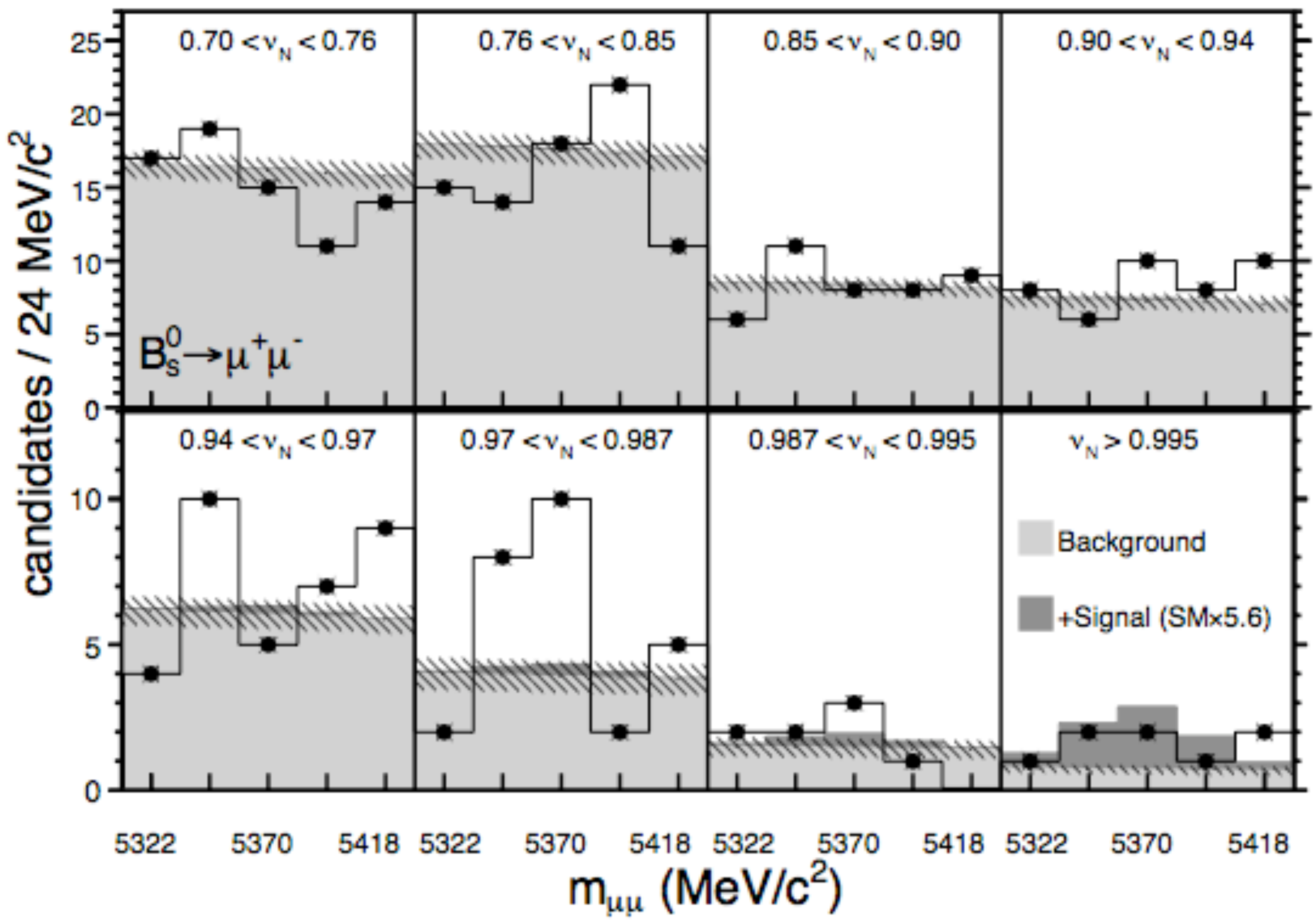
The $B_s \rightarrow \mu^+\mu^-$ group, with special thanks to graduate student Walter Hopkins, Cornell, who carried the lion's share of the work!



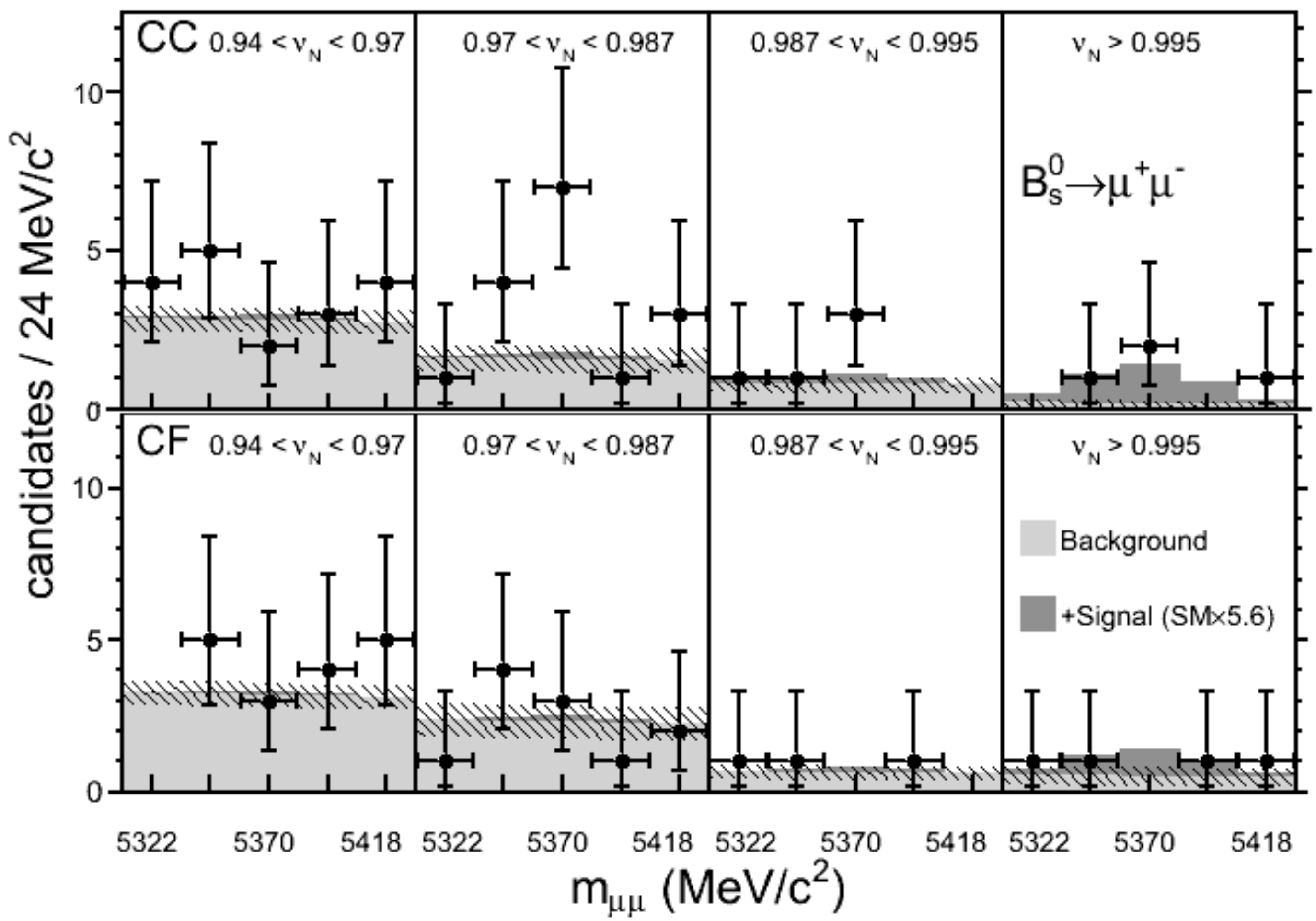
Backup slides

Data in B_s signal window

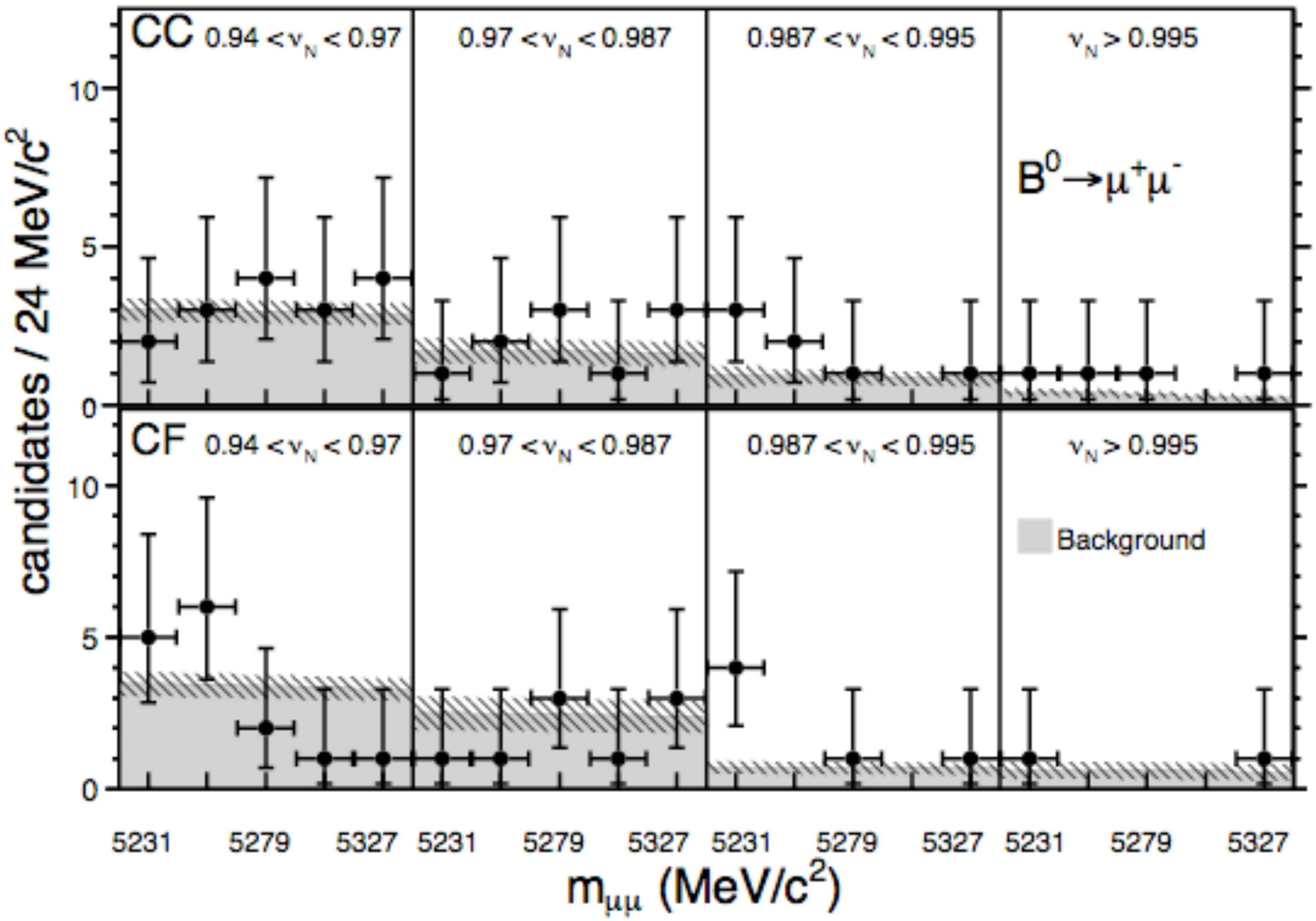
CC and CF combined



Data in Bs signal window



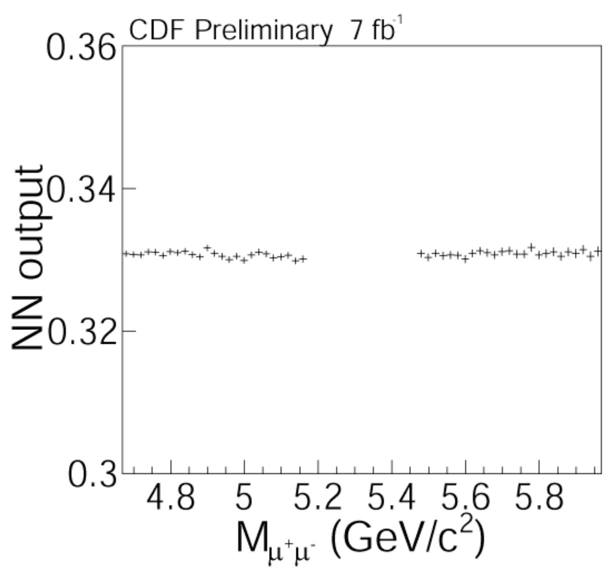
Data in B0 signal window

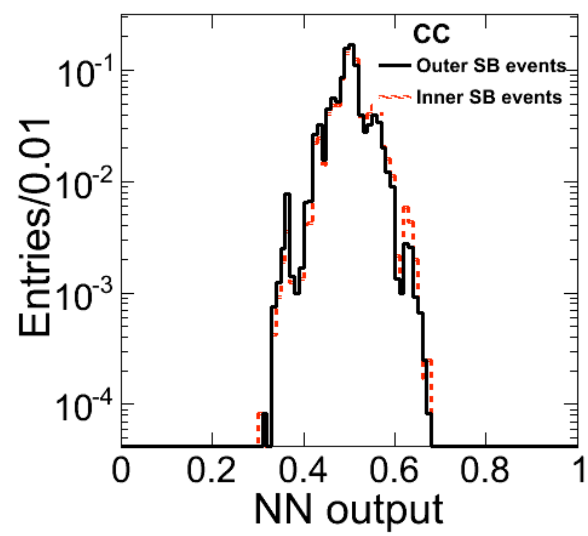


The Neural Net: validation and checks

Test if Neural Net introduces a selection bias

- to check if NN is "overtraining" on features of sideband data we divide the sideband data into training and testing samples
  - ✓variations in relative sample size have no effect
- to check for mass bias we train NN output in bins of dimuon invariant mass
  - ✓observe no mass bias
- train NN on inner and outer sideband events, checking for mass bias
  - ✓observe no difference

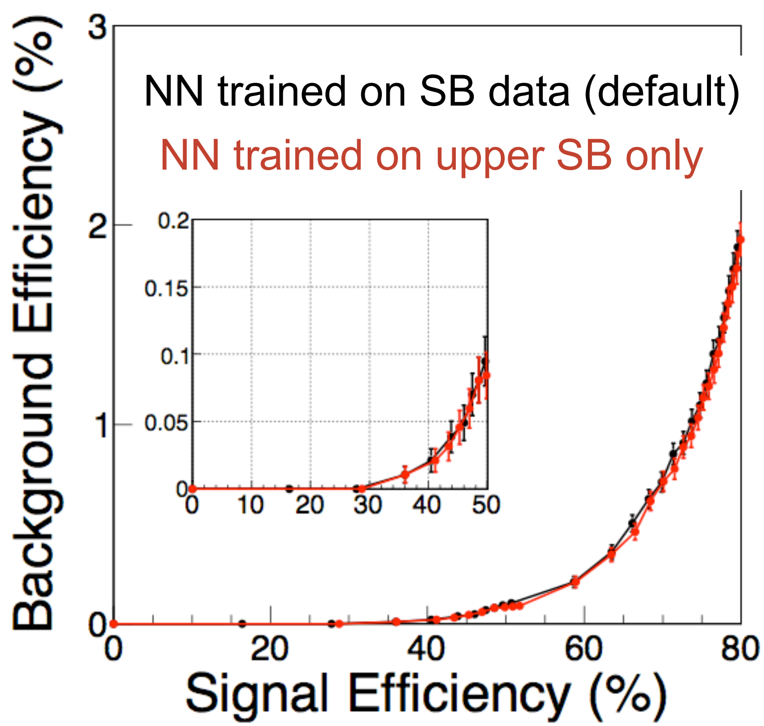


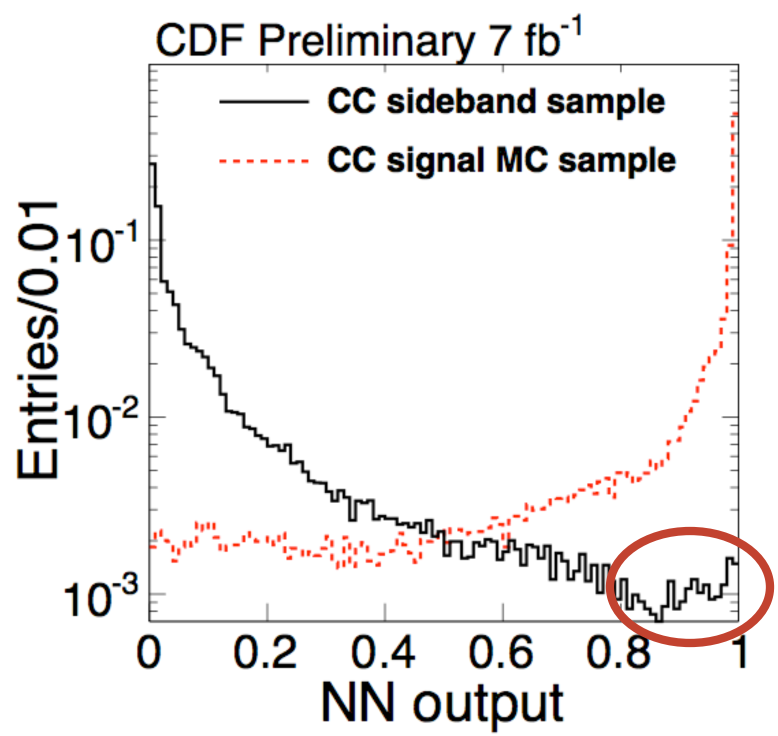


The Neural Net: validation and checks

Where does the increase at large NN output come from?

- caused by low mass events (<5 GeV) from partially reconstructed $b \rightarrow \mu\mu X$ decays.
- to check if the training is affected by these events we repeated it using upper sideband events only
  - ✓No change in NN output efficiencies

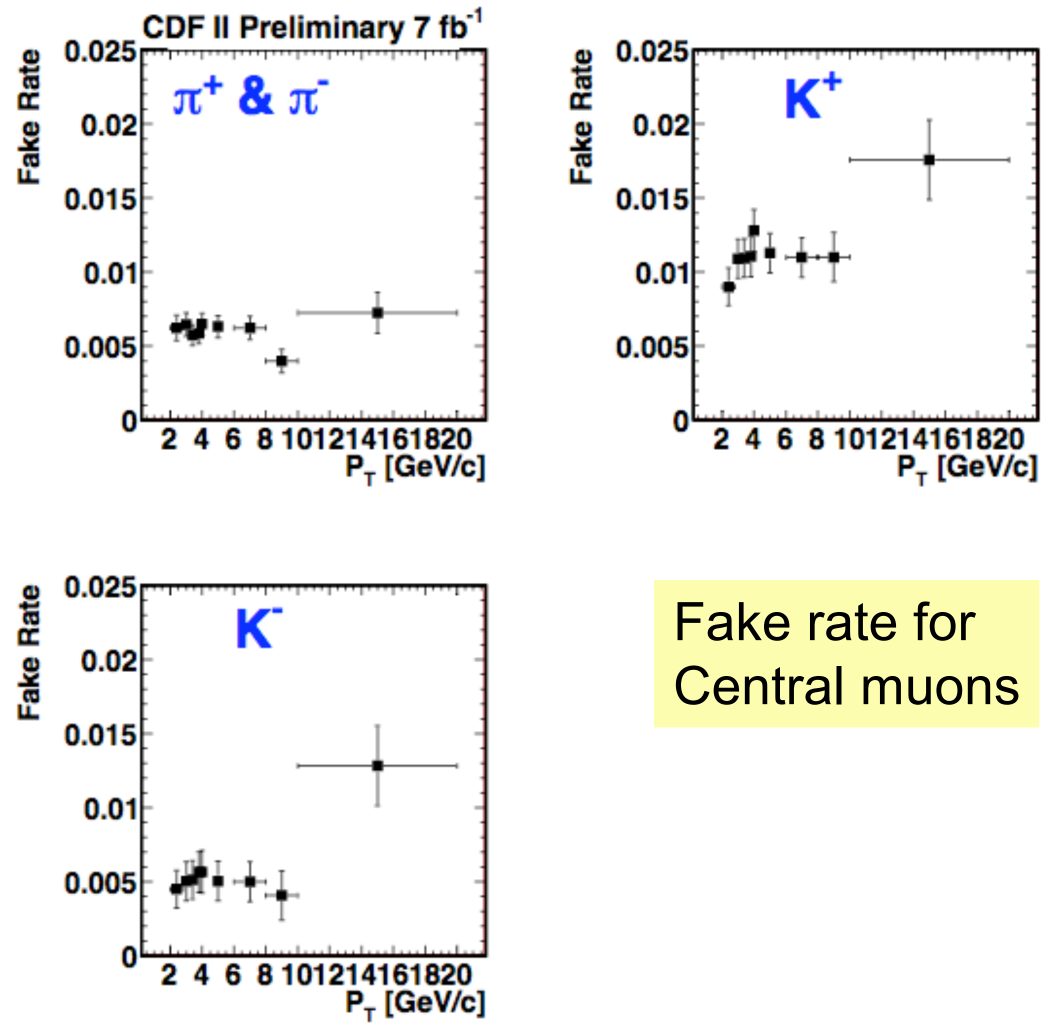




Muon fake rates

- Variations with p_T and luminosity are taken into account

- Total systematic uncertainty (due to both muon legs) dominated by residual run-dependence: ~35%



Fake rate for Central muons

Full table of bkgd checks in control samples

CF only

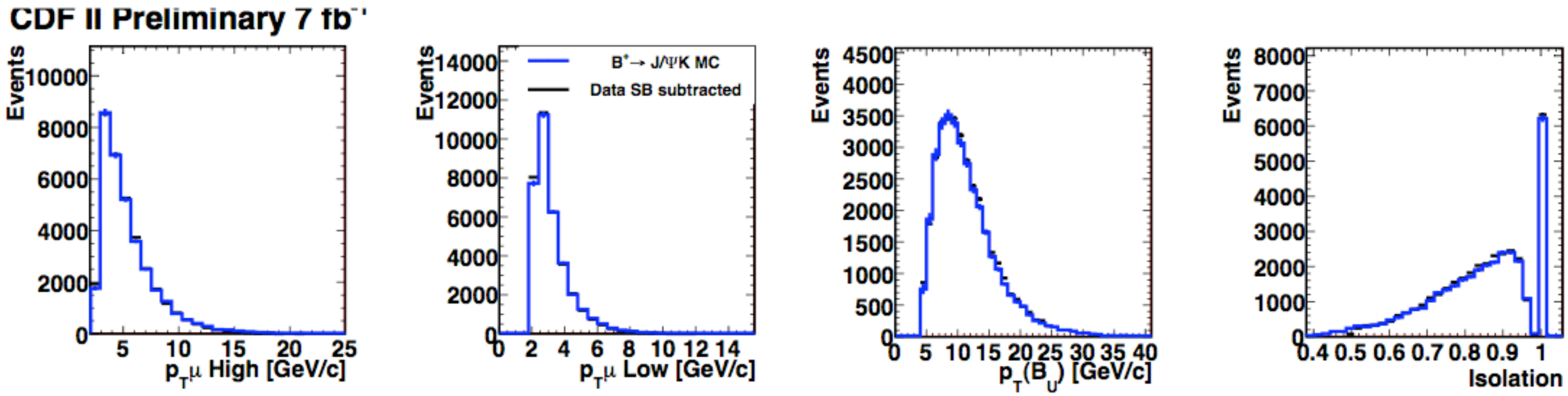
sample	*NN* cut	CF pred	CF obsv	CF prob(%)
OS-	0.700<NN<0.760	209.3±(12.0)	187	88.8
	0.760<NN<0.850	332.3±(16.3)	325	62.0
	0.850<NN<0.900	146.7±(9.7)	144	57.7
	0.900<NN<0.940	144.2±(9.6)	139	63.9
	0.940<NN<0.970	128.6±(8.9)	112	88.4
	0.970<NN<0.987	92.8±(7.4)	89	63.0
	0.987<NN<0.995	45.4±(5.0)	55	14.0
	0.995<NN<1.000	38.3±(4.5)	37	58.2
SS+	0.700<NN<0.760	0.3±(0.3)	1	30.0
	0.760<NN<0.850	4.2±(1.1)	4	57.8
	0.850<NN<0.900	0.3±(0.3)	3	1.3
	0.900<NN<0.940	0.6±(0.4)	1	45.4
	0.940<NN<0.970	0.9±(0.5)	1	56.8
	0.970<NN<0.987	0.6±(0.4)	0	54.9
	0.987<NN<0.995	0.5±(0.4)	0	60.1
	0.995<NN<1.000	0.3±(0.3)	1	30.0
SS-	0.700<NN<0.760	4.2±(1.1)	4	57.8
	0.760<NN<0.850	5.1±(1.2)	7	27.1
	0.850<NN<0.900	2.7±(0.9)	2	71.0
	0.900<NN<0.940	0.9±(0.5)	4	2.8
	0.940<NN<0.970	3.0±(0.9)	1	92.3
	0.970<NN<0.987	2.4±(0.8)	5	12.2
	0.987<NN<0.995	0.6±(0.4)	0	54.9
	0.995<NN<1.000	1.8±(0.7)	0	16.5
FM+	0.700<NN<0.760	54.8±(5.6)	66	12.7
	0.760<NN<0.850	66.3±(6.2)	57	83.1
	0.850<NN<0.900	33.7±(4.3)	25	90.3
	0.900<NN<0.940	17.4±(3.1)	26	6.6
	0.940<NN<0.970	9.5±(2.2)	15	10.2
	0.970<NN<0.987	5.3±(1.7)	9	13.4
	0.987<NN<0.995	2.7±(1.2)	3	49.3
	0.995<NN<1.000	2.1±(1.0)	8	0.7

*

*if zero events are observed, "Prob(N>=Nobs)" is the Poisson probability for observing exactly 0

Neural Network Cut Efficiencies- some more detail on the systematic uncertainty

- Observed differences between B^+ data and MC simulation, resulting in a 4% systematic uncertainty



- Observed difference in NN cut efficiency between B^+ data and MC simulation: average 4.6% difference

NN cut	CC Data	CC MC	CC Diff	CF Data	CF MC	CF Diff
NN>0.90	0.648±0.003	0.666±0.004	0.022	0.654±0.005	0.667±0.005	0.013
NN>0.95	0.571±0.003	0.588±0.004	0.017	0.574±0.005	0.583±0.005	0.007
NN>0.96	0.544±0.003	0.561±0.004	0.015	0.550±0.005	0.562±0.005	0.012
NN>0.97	0.514±0.003	0.530±0.004	0.016	0.515±0.005	0.530±0.005	0.015
NN>0.98	0.476±0.003	0.489±0.004	0.013	0.469±0.005	0.476±0.005	0.007
NN>0.99	0.392±0.003	0.406±0.004	0.014	0.356±0.005	0.380±0.005	0.024
NN>0.992	0.360±0.003	0.374±0.004	0.014	0.338±0.005	0.362±0.005	0.024
NN>0.995	0.304±0.003	0.312±0.004	0.008	0.299±0.005	0.319±0.005	0.020

Relative normalization to $B^+ \to J/\psi K+$: systematics

$$BR(B^0_{s(d)} \to \mu^+\mu^-) = \frac{N_{B_{s(d)}}}{N_{B^+}} \cdot \frac{\alpha_{B^+}}{\alpha_{B_{s(d)}}} \cdot \frac{\varepsilon^{total}_{B^+}}{\varepsilon^{total}_{B_{s(d)}}} \cdot \frac{1}{\varepsilon^{NN}_{B_{s(d)}}} \cdot \frac{f_u}{f_s} \cdot BR(B^+ \to J/\Psi K^+ \to \mu^+\mu^- K^+)$$

$$\frac{\alpha_{B^+}}{\alpha_{B_{s(d)}}} = 0.307 \pm 0.0018(stat) \pm 0.018(syst)$$

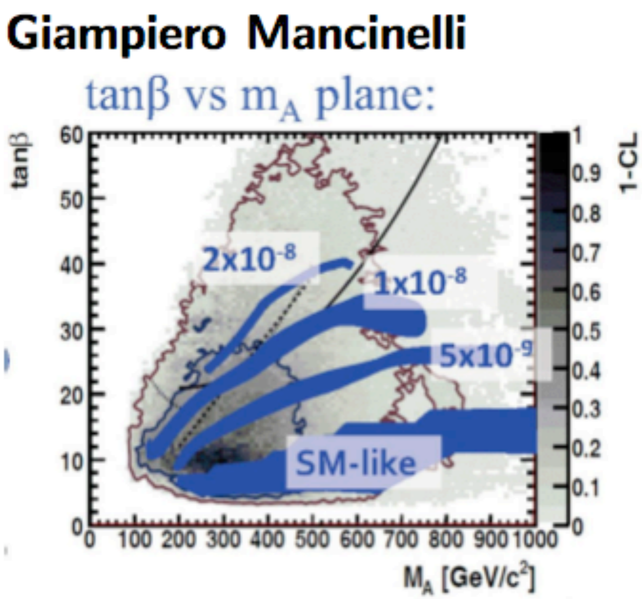
$$\frac{\varepsilon^{total}_{B^+}}{\varepsilon^{total}_{B_{s(d)}}} = 0.849 \pm 0.06(stat) \pm 0.007(syst)$$

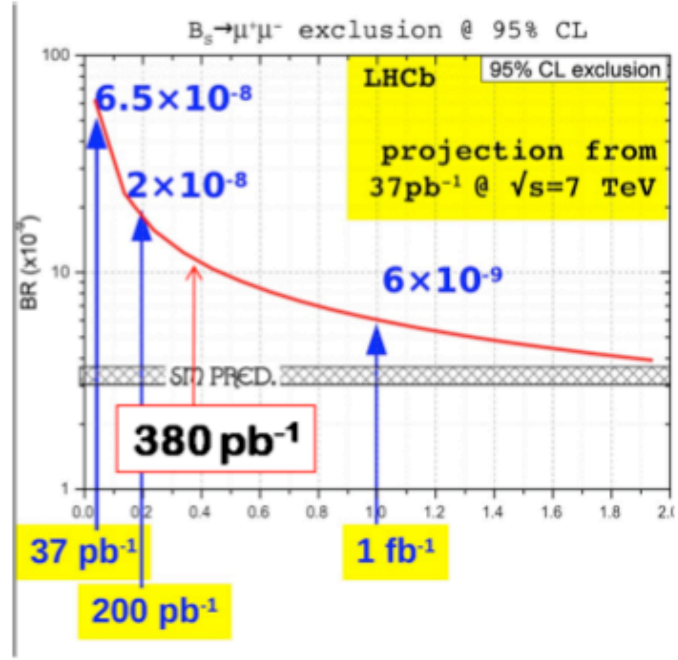
Systematic uncertainties include:

varying fragm. functions, renormalization and factorization scale, and the B-meson masses

kinematic differences between B_s and J/ψ decays, estimated using J/ψ data.
Kaon efficiency, B^+ vertex probability cut, estimated in data.

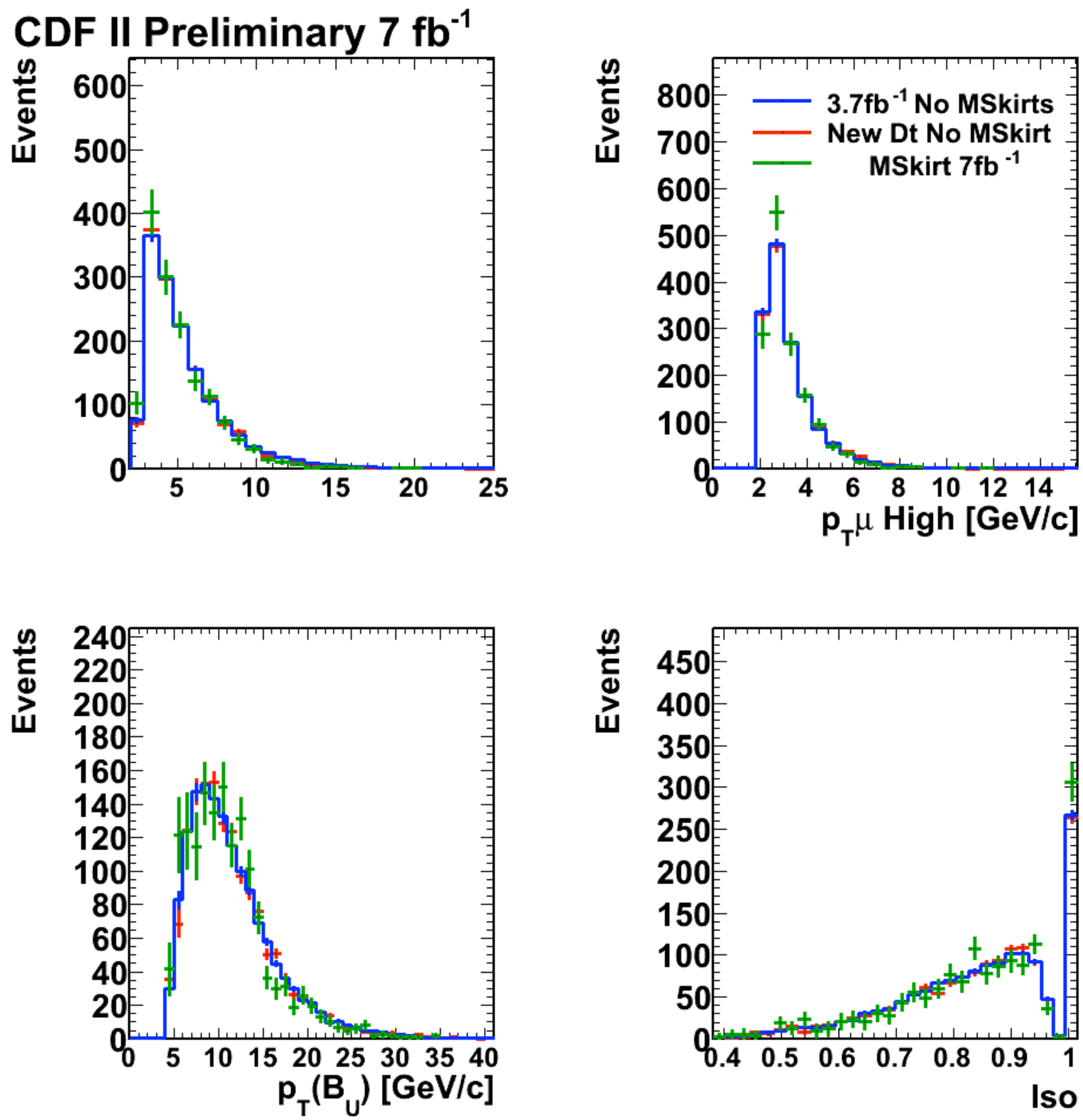
LHCb and CMS/Atlas in 2011- projections



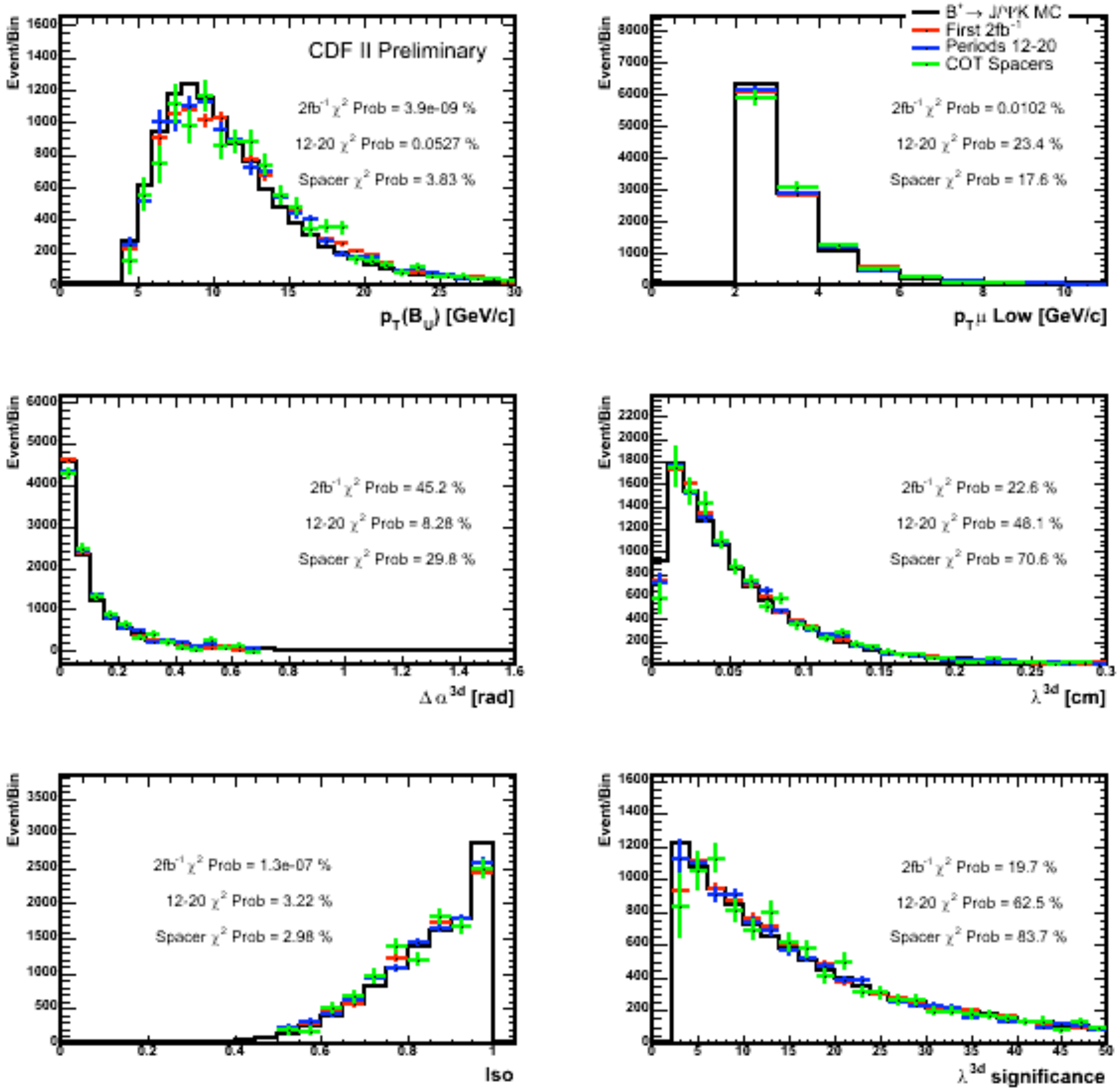


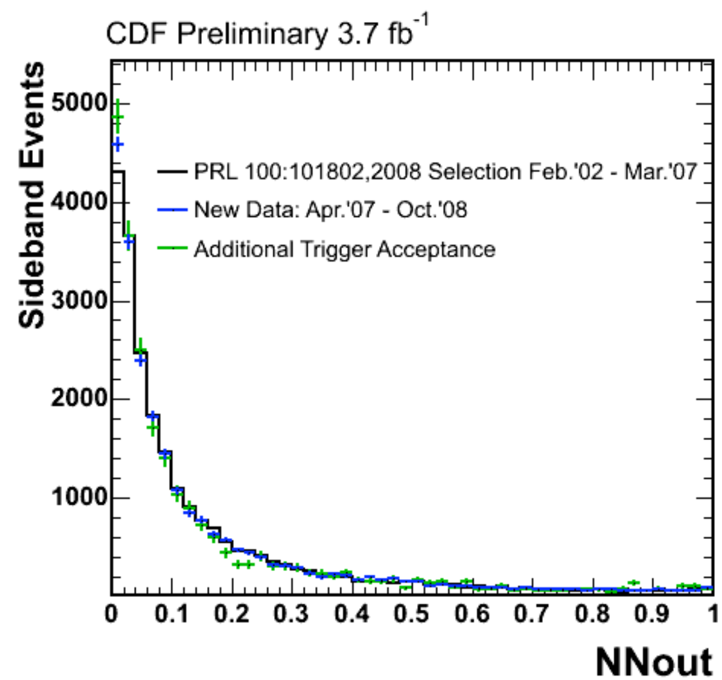
Experiment	N sig	N bkg	90% CL limit
ATLAS (10 fb^{-1}) $\sigma(bb) = 500\mu b$	5.7 evts	14^{+13}_{-10} evts (only bb$\to \mu\mu$)	-
CMS (1 fb^{-1}) $\sigma(bb) = 500\mu b$	2.36 evts	6.53 evts (2.5 bb$\to \mu\mu$)	1.6×10^{-8} (official) $\sim 1.0 \times 10^{-8}$ (LHCb MF est.)

Validation of the “miniskirt” data



Validation of the “COT-spacer” data





B_s signal window, number of expected SM events

	5.310-5.334	5.334-5.358	5.358-5.382	5.382-5.406	5.406-5.430
0.700-0.760	0.002 ± 0.000	0.007 ± 0.001	0.011 ± 0.002	0.006 ± 0.001	0.001 ± 0.000
0.760-0.850	0.004 ± 0.001	0.015 ± 0.003	0.020 ± 0.004	0.011 ± 0.002	0.003 ± 0.001
0.850-0.900	0.004 ± 0.001	0.010 ± 0.002	0.014 ± 0.003	0.008 ± 0.001	0.002 ± 0.000
0.900-0.940	0.005 ± 0.001	0.016 ± 0.003	0.023 ± 0.004	0.012 ± 0.002	0.002 ± 0.000
0.940-0.970	0.008 ± 0.001	0.022 ± 0.004	0.032 ± 0.006	0.016 ± 0.003	0.003 ± 0.001
0.970-0.987	0.010 ± 0.002	0.029 ± 0.005	0.041 ± 0.007	0.022 ± 0.004	0.005 ± 0.001
0.987-0.995	0.013 ± 0.002	0.046 ± 0.008	0.062 ± 0.011	0.031 ± 0.006	0.007 ± 0.001
0.995-1.000	0.052 ± 0.009	0.167 ± 0.030	0.227 ± 0.040	0.119 ± 0.021	0.029 ± 0.005

Table: Expected number SM Signal events in CMU-CMU channel

	5.310-5.334	5.334-5.358	5.358-5.382	5.382-5.406	5.406-5.430
0.700-0.760	0.002 ± 0.000	0.006 ± 0.001	0.007 ± 0.001	0.005 ± 0.001	0.001 ± 0.000
0.760-0.850	0.003 ± 0.001	0.012 ± 0.002	0.015 ± 0.003	0.009 ± 0.002	0.002 ± 0.000
0.850-0.900	0.003 ± 0.001	0.009 ± 0.002	0.012 ± 0.002	0.006 ± 0.001	0.001 ± 0.000
0.900-0.940	0.004 ± 0.001	0.012 ± 0.002	0.017 ± 0.003	0.009 ± 0.002	0.002 ± 0.000
0.940-0.970	0.005 ± 0.001	0.015 ± 0.003	0.021 ± 0.004	0.013 ± 0.002	0.003 ± 0.001
0.970-0.987	0.008 ± 0.002	0.026 ± 0.005	0.036 ± 0.007	0.019 ± 0.003	0.005 ± 0.001
0.987-0.995	0.007 ± 0.001	0.021 ± 0.004	0.029 ± 0.005	0.017 ± 0.003	0.004 ± 0.001
0.995-1.000	0.039 ± 0.007	0.116 ± 0.021	0.159 ± 0.029	0.090 ± 0.016	0.023 ± 0.004

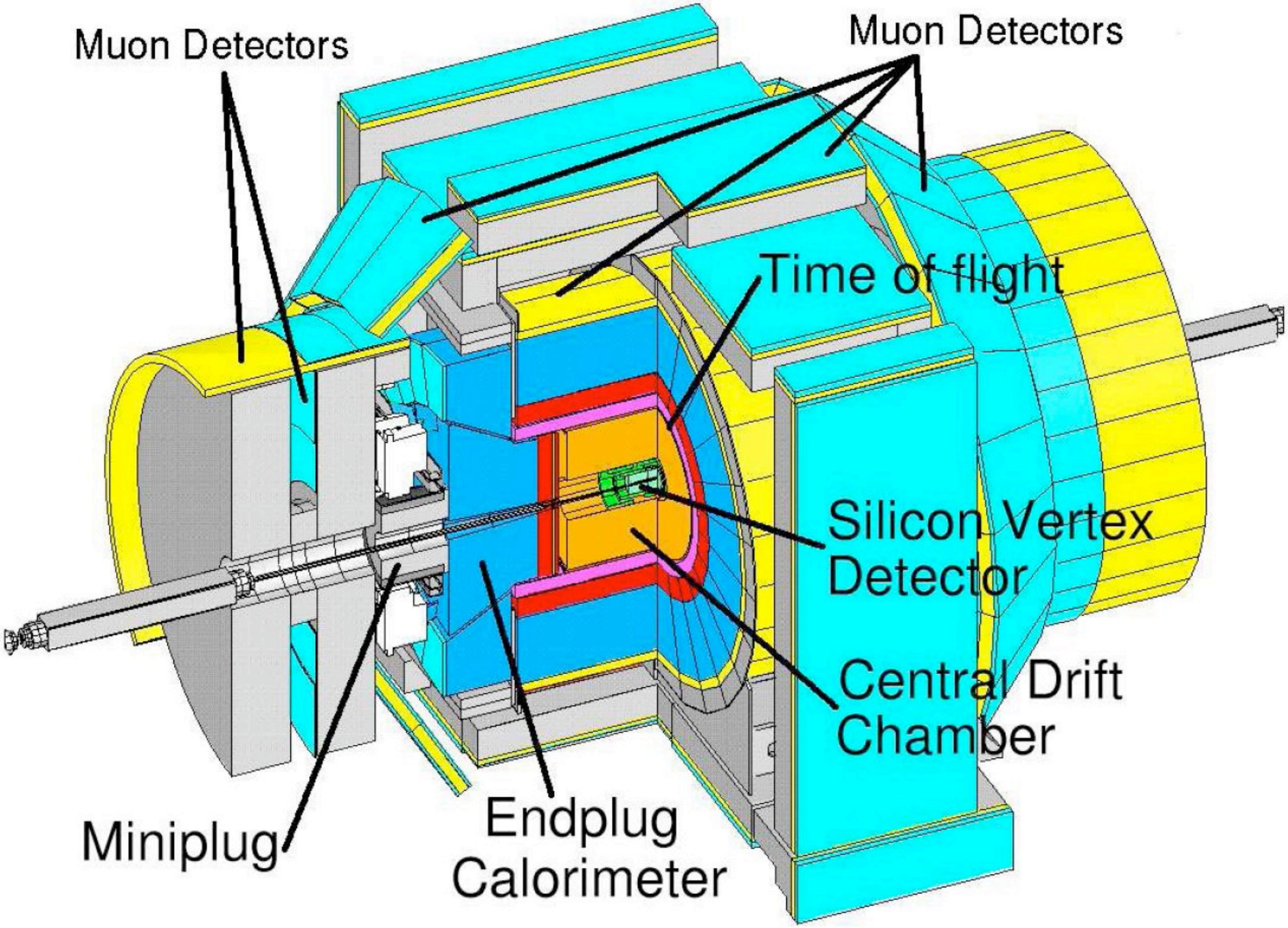
Table: Expected number SM Signal events in CMU-CMX channel

NN: Discriminating variables

Rank	Variable
1	$\Delta\alpha_{3\mathrm{d}}$
2	Isolation
3	Larger $\lvert d_0(\mu)\rvert$
4	$\lvert d_0(B_s^0)\rvert$
5	$L_{2\mathrm{d}}/\sigma_{L_{2\mathrm{d}}}$
6	χ^2_{vtx}
7	$L_{3\mathrm{d}}$
8	Lower $p_T(\mu)$
9	Significance of smaller $\lvert d_0(\mu)\rvert$
10	$\lambda_{3\mathrm{d}}/\sigma_{\lambda 3\mathrm{d}}$
11	$\lambda_{3\mathrm{d}}$
12	Smaller $\lvert d_0(\mu)\rvert$
13	$\Delta\alpha_{2\mathrm{d}}$
14	Significance of larger $\lvert d_0(\mu)\rvert$

Full table of expected and observed data

	Mass Bin (GeV)	5.310-5.334	5.334-5.358	5.358-5.382	5.382-5.406	5.406-5.430	Total
UU NN bin	Exp Bkg	8.38±0.62	8.27±0.61	8.17±0.59	8.07±0.58	7.97±0.56	40.86
0.7-0.76	Obs	9	6	7	2	5	29
UU NN bin	Exp Bkg	7.83±0.62	7.81±0.6	7.79±0.59	7.77±0.57	7.75±0.56	38.96
0.76-0.85	Obs	9	6	11	12	6	44
UU NN bin	Exp Bkg	3.23±0.43	3.22±0.41	3.21±0.4	3.2±0.39	3.19±0.37	16.05
0.85-0.9	Obs	5	6	2	5	4	22
UU NN bin	Exp Bkg	3.6±0.46	3.56±0.44	3.52±0.42	3.48±0.41	3.44±0.39	17.59
0.9-0.94	Obs	4	5	4	5	9	27
UU NN bin	Exp Bkg	3.0±0.4	2.96±0.38	2.91±0.37	2.87±0.36	2.83±0.35	14.58
0.94-0.97	Obs	4	4	2	3	3	16
UU NN bin	Exp Bkg	1.65±0.28	1.63±0.27	1.61±0.26	1.59±0.26	1.57±0.25	8.05
0.97-0.987	Obs	1	5	7	1	3	17
UU NN bin	Exp Bkg	0.96±0.2	0.93±0.19	0.91±0.19	0.89±0.18	0.87±0.18	4.55
0.987-0.995	Obs	1	1	3	0	0	5
UU NN bin	Exp Bkg	0.26±0.08	0.22±0.08	0.2±0.07	0.19±0.07	0.18±0.07	1.03
0.995-1	Obs	0	1	2	0	1	4
UX NN bin	Exp Bkg	8.74±0.63	8.61±0.61	8.48±0.6	8.35±0.58	8.22±0.57	42.39
0.7-0.76	Obs	8	13	9	9	9	48
UX NN bin	Exp Bkg	9.65±0.66	9.52±0.64	9.38±0.62	9.24±0.61	9.1±0.6	46.89
0.76-0.85	Obs	7	8	7	11	4	37
UX NN bin	Exp Bkg	5.07±0.5	4.99±0.49	4.92±0.47	4.84±0.46	4.76±0.44	24.59
0.85-0.9	Obs	1	5	6	3	5	20
UX NN bin	Exp Bkg	3.92±0.47	3.87±0.45	3.82±0.43	3.76±0.42	3.71±0.4	19.08
0.9-0.94	Obs	4	1	6	3	3	17
UX NN bin	Exp Bkg	2.65±0.37	2.67±0.36	2.69±0.35	2.71±0.34	2.74±0.34	13.46
0.94-0.97	Obs	0	5	3	4	5	17
UX NN bin	Exp Bkg	2.4±0.34	2.37±0.33	2.34±0.32	2.3±0.31	2.27±0.3	11.68
0.97-0.987	Obs	1	4	3	1	2	11
UX NN bin	Exp Bkg	0.54±0.16	0.54±0.15	0.55±0.15	0.55±0.15	0.56±0.15	2.74
0.987-0.995	Obs	1	1	0	1	0	3
UX NN bin	Exp Bkg	0.83±0.0	0.78±0.0	0.75±0.0	0.71±0.0	0.68±0.0	3.75
0.995-1	Obs	1	1	0	1	1	4


Muon Detectors
Muon Detectors
Time of flight
Silicon Vertex Detector
Central Drift Chamber
Miniplug
Endplug Calorimeter

Cross checks of B → hh

p_T dependence of fake rate will bias B->hh candidates that survive the muon selection. Will this affect the NN output distribution, e.g. resulting in a bias to high NN output?

CC only (see backup for CF)

NN output distribution of the signal simulation is compared before and after applying the pt dependent fake rate. No difference is observed.