Search for $B_s \rightarrow \mu^+\mu^-$
and $B_d \rightarrow \mu^+\mu^-$ Decays

D.Glenzinski
FNAL / CDF
16-Apr-2004

Outline
- Introduction
- CDF & Tevatron
- Method
- Results
In the Standard Model $B \rightarrow \mu\mu$ is a FCNC decay…
only possible at the loop level

\[
BR(B_s \rightarrow \mu^+\mu^-) = (3.5 \pm 0.9) \times 10^{-9}
\]


Not yet experimentally observed.

\[
BR(B_s \rightarrow \mu^+\mu^-) < 2.0 \times 10^{-6} \text{ @ } 90\% \text{ CL}
\]

(CDF, PRD 57 (1998) 3811R)
Several extensions to SM allow for $\text{BR} \gg \text{BR}(\text{SM})$.
Even modest improvements to limits can give interesting constraints on “relevant” models.
Introduction

B-hadron production cross-sections:
- PEPII : $\sigma(B) \sim 1 \text{ nb}$
- TevII : $\sigma(B) \sim 30000 \text{ nb}$

After trigger and reconstruction:
- $1 \text{ fb}^{-1}(B\text{-factory}) \sim 1 \text{ pb}^{-1}(\text{Tevatron})$

Center of mass energies at B-factories below $B_s$ threshold, but at Tevatron:
#$B^+ : B_d : B_s : \Lambda_b \sim 4 : 4 : 1 : 1$

This decay mode offers the Tevatron experiments a unique opportunity.
- World’s highest energy pp collider
  \[ E_{\text{cm}} = 2 \text{ TeV} \]
- CDF and D0 significantly upgraded
- New data taking since Mar-2001
- Significant accelerator upgrades ongoing
- Have >300 pb-1 on tape
- this analysis based on 171 pb-1
- Tevatron doing well
- expect another >=200 pb-1 FY04
CDF:

Features:

• large radius tracking wire chamber (COT)

• 1.4 T solenoid

• precision silicon vertexing (SVX)

• muon chambers (CMU & CMP, $|\eta| < 0.6$)

$\eta = -\ln(\tan(\theta / 2))$
CDF

- physics data since Feb-2002
- data-taking efficiency >85%
- performing well
Method

\[ BR(B_s \rightarrow \mu^+\mu^-) = \frac{(N_{\text{candidates}} - N_{bg})}{\alpha \cdot \epsilon_{\text{total}} \cdot \sigma_{B_s} \cdot \int Ldt} \]

This measurement requires that we:

- demonstrate understanding of background, \( N_{bg} \)
- accurately estimate \( \alpha \varepsilon \)
- intelligently optimize cuts

Since SM predicts 0 events, this is really a “search”

- more rigorous about testing \( N_{bg} \) estimate
- emphasis on performing an unbiased optimization
**Method**

Collect sample using Di-Muon Triggers

76k events

Make reconstruction requirements & Constrain to a common 3D vertex

2981 events

Apply cuts to discriminate Signal from Background

? events

Estimate BR using:

\[
BR(B_s \rightarrow \mu^+\mu^-) = \frac{(N_{\text{candidates}} - N_{bg})}{\alpha \cdot \epsilon_{\text{total}} \cdot \sigma_{Bs} \cdot \int Ldt}
\]

**Strategy:**

- “blind” ourselves to data in signal region
- use sideband data to understand background
- employ *a priori* optimiztn
- don’t “open box” until expected sensitivity warrants (< 0.5 RunI)

➤ I’ll talk about each piece in turn
Method: Unbiased Optimization

When optimizing the selection criteria, we “blinded” ourselves to the data in an extended search region.

Search Region:
- $5.169 < M_{\mu\mu} < 5.469$ GeV
- corresponds to $\pm 4\sigma(M_{\mu\mu})$
- width included in optimiztn

Sideband Regions:
- additional 0.5 GeV on either side of search region
- used to understand Bkgd
Method: Triggers

Collect sample using Di-Muon Triggers

- **“CMU-CMU”**
  - both muons in CMU
  - $P_T(\mu) > 1.5$ GeV
  - $2.7 < M_{\mu\mu} < 6.0$ GeV
  - $\Delta \phi(\mu\mu) < 2.25$ rad
  - $P_T(\mu^+) + P_T(\mu^-) > 5$ GeV

- **“CMUP-CMU”**
  - 1 muon in CMP, 1 in CMU
  - $P_T(\text{CMP}\mu) > 1.5$ GeV
  - $P_T(\text{CMU}\mu) > 3.0$ GeV
  - $2.7 < M_{\mu\mu} < 6.0$ GeV
  - $\Delta \phi(\mu\mu) < 2.25$ rad

⇒ 76k events satisfy trigger
**Method: Reconstruction Requirements**

Collect sample using Di-Muon Triggers

76k events

Make reconstruction requirements & Constrain to a common 3D vertex

**We Require:**

- “good” COT tracks and CMU/P track-stubs
- $\geq 4$ SVX $r-\phi$ hits
- $4.669 < M_{\mu\mu} < 5.969$ GeV
- “good” vertex
  - $\sigma(L_T) < 150$ $\mu$m
  - $\chi^2 < 15$
  - $L_T < 1$ cm
- $P_T(\mu\mu) > 6$ GeV

$\rightarrow$ 2984 events survive

(expect $< 30 \ B_s \rightarrow \mu+\mu-$… this is bkgd dominated)
Method: Reconstruction Requirements

These requirements:
- keep 92% of Bs→μ+μ−
- reject 50% of the background.

Bs→μ+μ− MC

μ+μ− Data

(4.669 < Mμμ < 5.969 GeV)
Method: Discriminate Signal from Background

At this stage:
- sample is background dominated
- need to find variables that reduce background by a factor of >1000
- ... and keep as much signal as possible

→ let’s think about signal & background characteristics…
Method: Discriminate Signal from Background

Signal Characteristics

- final state is fully reconstructed
- $B_s$ has long lifetime
  ($c\tau = 483 \, \mu m$)
- $B$ fragmentation is hard

For real $B_s \rightarrow \mu^+\mu^-$ expect:
- $M_{\mu\mu} = M(B_s)$
- $\lambda = cL_T \frac{M_{\mu\mu}}{P_T(\mu\mu)}$ to be large
- $L_T$ and $P_T(\mu\mu)$ to be co-linear
- few additional tracks
Method: Discriminate Signal from Background

In general:
- $M_{\mu\mu} \neq M(B_s)$
- $\lambda = cL_T M_{\mu\mu}/P_T(\mu\mu)$ will be smaller

Contributing Backgrounds
- sequential semi-leptonic decay, $b \rightarrow \mu- cX \rightarrow \mu+ \mu- X$
- double semi-leptonic decay, $g \rightarrow b\bar{b} \rightarrow \mu+ \mu- X$
- continuum $\mu+ \mu-$, $\mu$ + fake fake+fake

$\mu+$
$\mu-$
$P_T(\mu\mu)$
$L_T$
primary vertex
di-muon vertex

$L_T$ and $P_T(\mu\mu)$ will not be co-linear
more additional tracks
Method: Discriminating Variables

Discriminating Variables
- Invariant mass, $M_{\mu\mu}$
- $\lambda = cL \times M_{\mu\mu}/P_T(\mu\mu)$
- $\Delta \Phi : \phi(\vec{P}_T(\mu\mu)) - \phi(\vec{L}_T)$
- Isolation
  $= P_T(\mu\mu)/(\Sigma_{\text{trk}} + P_T(\mu\mu))$

$\rightarrow$ need to determine optimal requirements
**Method: Unbiased Optimization**

We used the set of requirements which yielded the minimum *a priori* expected BR Limit:

\[
\langle BR (B_s \rightarrow \mu^+ \mu^-) \rangle = \frac{\langle N^{90\% \text{CL}}_{\text{signal}} \rangle}{\alpha \cdot \mathcal{E}_{\text{total}} \cdot \sigma_{B_s} \int L dt}
\]

where we’ve summed over all possible \( n_{\text{obs}} \):

\[
\langle N^{90\% \text{CL}}_{\text{signal}} \rangle = \sum_{n_{\text{obs}}=0}^{\infty} P(n_{\text{obs}} | n_{\text{bg}}) \cdot N^{90\% \text{CL}}_{\text{signal}} (n_{\text{bg}}, \Delta_{\text{bg}}, \Delta_{\alpha \cdot \epsilon})
\]

- Poisson prob of observing \( n_{\text{obs}} \) when expecting \( n_{\text{bg}} \)
- 90% CL UL on \( N_{\text{signal}} \) when expecting \( n_{\text{bg}} \) bkgd evts using Bayesian Method and including uncertainties
Method: Unbiased Optimization

The a priori expected BR limit is given by:

$$\left\langle BR \left( B_s \rightarrow \mu^+ \mu^- \right) \right\rangle = \frac{\left\langle N_{signal}^{90\% \, CL} \right\rangle}{\alpha \cdot \varepsilon_{total} \cdot \sigma_{B_s} \int Ldt}$$

where:

$$\left\langle N_{signal}^{90\% \, CL} \right\rangle = \sum_{n_{obs}=0}^{\infty} P(n_{obs} \mid n_{bg}) \cdot N_{signal}^{90\% \, CL} (n_{bg}, \Delta_{bg}, \Delta_{\alpha \varepsilon})$$

To perform the optimization we needed:

- background estimate, $n_{bg} +/- \Delta_{bg}$
- total acceptance estimate, $\alpha \varepsilon_{total} +/- \Delta_{\alpha \varepsilon}$

for each set of $(M_{\mu\mu}, \lambda, \Delta \Phi, \text{Isolation})$ requirements.
**Method: Background Estimate**

We estimate the background in the signal region using:

\[ n_{bg} = n_{sb}(\lambda, \Delta \Phi) \cdot f_{Isol} \cdot f_M \]

- \#sideband events surviving (\(\lambda, \Delta \Phi\)) requirements
- fraction of background events expected to survive Isolation req’rmt
- ratio of \#events in signal region, given \#evts in sidebands

- Isol and \(M_{\mu\mu}\) need to be uncorrelated w/ other vars
- background \(M_{\mu\mu}\) needs to be linear
- can determine \(f_{Isol}\) and \(f_M\) on samples w/ loose (no) \(\lambda, \Delta \Phi\) requirements… \(\Delta_{bg}\) reduced
**Method: Background Estimate**

- CDF II 171 pb⁻¹

**ρ (mass - λ) = -0.03**

**ρ (mass - ΔΦ) = 0.05**

**ρ (mass - Iso) = 0.03**

**ρ (Iso - λ) = -0.14**

**ρ (Iso - ΔΦ) = 0.02**

**ρ (ΔΦ - λ) = -0.30**

Using our background dominated data sample...

Estimate linear correlation coefficient for each combination of variables:

\[
\rho_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \hat{x})(y_i - \hat{y}) / \sigma_x \sigma_y
\]

✓ Isol and M_{μμ} are uncorrelated with other variables

\((Δρ(\text{stat}) = +/-0.03 \text{ each})\)
**Method:**

\[ B_s \rightarrow \mu^+ \mu^- \text{ MC} \]

\[ \rho (\text{mass-} \lambda) = 0.03 \]

\[ \rho (\text{mass-} \Delta \Phi) = 0.07 \]

\[ \rho (\text{mass-} \text{Iso}) = 0.06 \]

\[ \rho (\text{Iso-} \lambda) = 0.01 \]

\[ \rho (\text{Iso-} \Delta \Phi) = 0.01 \]

\[ \rho (\Delta \Phi- \lambda) = -0.37 \]

**NOTE:** \( M_{\mu \mu} \) and Isolation are generally uncorrelated with other variables... even for signal.
Method: Background Estimate

Using our background dominated data sample, fit $M_{\mu\mu}$

CDFII 171 pb$^{-1}$

Background dominated $\mu^+\mu^-$

\[ |\eta_\mu| < 0.6, \quad P_T^{\mu\mu} > 6 \text{ GeV/c} \]

\[ \chi^2 / \text{ndf} = 11.41 / 10 \]

Intercept $781.8 \pm 66.45$

Slope $-104 \pm 12.38$

✓ Background $M_{\mu\mu}$ is linear
Method: Background Estimate

Since assumptions satisfied, we can determine $f_{iso}$ and $f_M$ using background dominated sample:

$$f_{iso} = \frac{\# evts(Iso > \text{threshold})}{\# evts}$$

<table>
<thead>
<tr>
<th>threshold</th>
<th>$f_{iso}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iso&gt;0.60</td>
<td>0.535 +/- 0.009</td>
</tr>
<tr>
<td>Iso&gt;0.65</td>
<td>0.450 +/- 0.009</td>
</tr>
<tr>
<td>Iso&gt;0.70</td>
<td>0.362 +/- 0.009</td>
</tr>
<tr>
<td>Iso&gt;0.75</td>
<td>0.283 +/- 0.008</td>
</tr>
<tr>
<td>Iso&gt;0.80</td>
<td>0.214 +/- 0.008</td>
</tr>
<tr>
<td>Iso&gt;0.85</td>
<td>0.160 +/- 0.007</td>
</tr>
</tbody>
</table>

Since background $M_{\mu\mu}$ is linear,

$$f_M = \frac{\# evts(\text{signal})}{\# evts(\text{sideband})}$$

$$f_M = \frac{\Delta M(\text{signal})}{\Delta M(\text{sideband})}$$

(variation in bins of $\lambda$ and $M_{\mu\mu}$ yield a systematic uncertainty of +/- 5%)
Let’s pause here to consider some specific background sources:

1. Two-body B-decays
   - $B \rightarrow h^+h^-$ ($h = \pi$ or K)
   - $M_{\mu\mu}$ not linear

2. Generic $b\bar{b}$ events
   - $M_{\mu\mu}$ linear?
   - Surprises?
Aside: Specific Background Sources

For two-body B-decays, $B \rightarrow h^+ h^-$ ($h = \pi$ or $K$)...

Estimate contribution to signal region by:

1. **Take acceptance, $M_{hh}$ (assuming $\mu$ mass), $P_T(h)$ from MC samples**

2. **Convolute $P_T(h)$ with $\mu$-fake rates derived from $D^*$ tagged $K$, $\pi$ tracks**
   - fake rates binned in $P_T$ and charge
   - separately determined for $\pi$ and $K$
   - yields double fake rates of $2-6 \times 10^{-4}$

$$B_d \rightarrow h^+ h^- : \alpha \cdot \varepsilon_{total} \cdot BR < 4 \times 10^{-11}$$

$$B_s \rightarrow h^+ h^- : \alpha \cdot \varepsilon_{total} \cdot BR < 1 \times 10^{-9}$$

...expected sensitivity in $10^{-7}$ range, can safely ignore these backgrounds
Aside: Specific Background Sources

For generic $b\bar{b}$ events...

Used $b\bar{b}$ MC sample to learn:

1. Mass and Isolation correlations small
2. $M_{\mu\mu}$ is linear

⇒ Will be accurately accounted for

As a check, can use method described above to predict how many $b\bar{b}$ MC events will fall into signal region for a loose set of cuts:

<table>
<thead>
<tr>
<th>$\eta_{bg}$</th>
<th>$\lambda&gt;50\mu m$</th>
<th>$\lambda&lt;50\mu m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted</td>
<td>3.1+/-0.7</td>
<td>8.8+/-1.3</td>
</tr>
<tr>
<td>Observed</td>
<td>2</td>
<td>6</td>
</tr>
</tbody>
</table>
Aside: Specific Background Sources

For generic $b\bar{b}$ events...

As a further check, can use a set of requirements that are near optimal (ie. tight) and look at (N-1) distributions:

<table>
<thead>
<tr>
<th>Cut omitted</th>
<th>#survive</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolation</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1</td>
<td>$\lambda = 6 , \mu$m</td>
</tr>
<tr>
<td>$\Delta \Phi$</td>
<td>1</td>
<td>$\Delta \Phi = 0.91 , \text{rad}$</td>
</tr>
<tr>
<td>$M_{\mu\mu}$</td>
<td>1</td>
<td>$M_{\mu\mu} = 5.559 , \text{GeV}$</td>
</tr>
</tbody>
</table>

- only 3 events (of $1.2 \times 10^9$) fail a single cut and these are far from the cut thresholds

...no special treatment required.
Aside: Specific Background Sources

We paused to consider some specific background sources:

1. Two-body B-decays
   • negligible

2. Generic $b\bar{b}$ events
   • no special treatment required

Let’s compare $n_{bg}$ predictions to observations in control samples.
Method: Background Cross-checks

• Data Samples (statistically independent)
  – OS+ : opposite-sign muon pairs, $\lambda > 0$;
    the signal sample – not used for xchecks
  – OS- : opposite-sign muon pairs, $\lambda < 0$
  – SS+ : same-sign muon pairs, $\lambda > 0$
  – SS- : same-sign muon pairs, $\lambda < 0$

• OS samples pass the default reco+vertex cuts
• SS samples pass looser reco cuts
  – looser == remove trigger matching, and
    $P_T(\mu) > 1.5$ and $P_T(\mu\mu) > 4.0$ GeV
**Method: Background Cross-checks**

- OS- sample “ideal” control sample.
- SS sample not ideal but useful since some B-backgrounds there.
Method: Background Cross-checks

• Compare #predicted vs #observed for three sets of cuts
  – A : \((λ, ΔΦ, Iso) = (>100μm, <0.20 \text{ rad}, >0.60)\)
  – B : \((λ, ΔΦ, Iso) = (>150μm, <0.20 \text{ rad}, >0.70)\)
  – C : \((λ, ΔΦ, Iso) = (>200μm, <0.10 \text{ rad}, >0.80)\)

• “B” corresponds to near optimal cuts, while A (C) correspond to looser (tighter) sets of cuts

• Note: C < B < A (ie. correlated for same sample), but OS-, SS+ and SS- stat. independent
**Method: Background Cross-checks**

<table>
<thead>
<tr>
<th></th>
<th>OS+</th>
<th>OS-</th>
<th>SS+</th>
<th>SS-</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho(\text{Iso-}\lambda)$</td>
<td>-0.14</td>
<td>-0.05</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td>$\rho(\text{Iso-}\Delta\Phi)$</td>
<td>0.02</td>
<td>-0.08</td>
<td>-0.02</td>
<td>-0.02</td>
</tr>
<tr>
<td>$\rho(\text{Iso-M})$</td>
<td>0.03</td>
<td>0.03</td>
<td>-0.02</td>
<td>-0.01</td>
</tr>
<tr>
<td>$\rho(\lambda-M)$</td>
<td>-0.03</td>
<td>-0.05</td>
<td>-0.02</td>
<td>-0.00</td>
</tr>
<tr>
<td>$\rho(\Delta\Phi-M)$</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>$\rho(\Delta\Phi-\lambda)$</td>
<td>-0.30</td>
<td>-0.21</td>
<td>-0.20</td>
<td>-0.20</td>
</tr>
</tbody>
</table>

(uncertainty is +/-0.03 and +/-0.02, per element, for OS and SS samples, respectively)

$\rho(\text{mass-x}), \rho(\text{Iso-x})$ small for all samples
### Method: Background Cross-checks

| Sample | #predicted     | #obsrvd | $P(>=\text{o}|\text{pred})$ |
|--------|----------------|---------|-----------------------------|
| OS-    | 10.43 +/- 1.89 | 16      | 4%                          |
| SS+    | 5.80 +/- 0.98  | 4       | 83%                         |
| SS-    | 6.72 +/- 1.10  | 7       | 51%                         |
| Sum    | 22.94 +/- 3.14 | 27      |                             |

<table>
<thead>
<tr>
<th>A</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OS-</td>
<td>3.69 +/- 0.80</td>
<td>6</td>
<td>17%</td>
</tr>
<tr>
<td>SS+</td>
<td>1.83 +/- 0.35</td>
<td>1</td>
<td>84%</td>
</tr>
<tr>
<td>SS-</td>
<td>2.32 +/- 0.42</td>
<td>4</td>
<td>20%</td>
</tr>
<tr>
<td>Sum</td>
<td>7.84 +/- 1.19</td>
<td>11</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>OS-</td>
<td>0.64 +/- 0.22</td>
<td>1</td>
<td>47%</td>
</tr>
<tr>
<td>SS+</td>
<td>0.29 +/- 0.08</td>
<td>0</td>
<td>75%</td>
</tr>
<tr>
<td>SS-</td>
<td>0.27 +/- 0.08</td>
<td>1</td>
<td>24%</td>
</tr>
<tr>
<td>Sum</td>
<td>1.21 +/- 0.27</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

where $P(>=\text{o}|\text{p})$ is the Poisson prob of observing $>=\text{o}$ when expecting $\text{p}$; when 0 observed give $P(0|\text{p})$. 
**Method: Background Cross-checks**

one last x-check in fake-µ enhanced sample

- require >=1 leg to fail µ quality cuts
- reduces signal efficiency by factor of 50, while increasing background by factor of about 3
- verify $\rho$(mass-x) and $\rho$(Iso-x) are small

<table>
<thead>
<tr>
<th>fake-µ enhanced sample</th>
<th>#predicted</th>
<th>#obsvd</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda &gt; 0$</td>
<td>20.52 +/- 3.17</td>
<td>17</td>
</tr>
<tr>
<td>$\lambda &lt; 0$</td>
<td>22.33 +/- 3.41</td>
<td>22</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda &gt; 0$</td>
<td>6.52 +/- 1.15</td>
<td>4</td>
</tr>
<tr>
<td>$\lambda &lt; 0$</td>
<td>7.33 +/- 1.25</td>
<td>11</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda &gt; 0$</td>
<td>0.83 +/- 0.23</td>
<td>1</td>
</tr>
<tr>
<td>$\lambda &lt; 0$</td>
<td>0.97 +/- 0.25</td>
<td>1</td>
</tr>
</tbody>
</table>

⇒ OK sufficient confidence in background prediction. Let’s consider efficiencies…
\[ \alpha \cdot \varepsilon_{\text{total}} = \alpha \cdot \varepsilon_{\text{trig}} \cdot \varepsilon_{\text{reco}} \cdot \varepsilon_{\text{final}} \]

where,

\[ \varepsilon_{\text{trig}} = \text{trigger} \]

\[ \varepsilon_{\text{reco}} = \varepsilon_{\text{COT}} \cdot \varepsilon_{\text{muon}} \cdot \varepsilon_{\text{SVX}} \cdot \varepsilon_{\text{vtx}} \]

\[ \varepsilon_{\text{final}} = \varepsilon_{\lambda} \cdot \varepsilon_{\Delta\Phi} \cdot \varepsilon_{\text{Iso}} \cdot \varepsilon_{\text{mass}} \]

I’ll briefly describe each.

For optimization, only \( \varepsilon_{\text{final}} \) varies.
**Method: Efficiency and Acceptance**

Acceptance = fraction of $B_s \rightarrow \mu^+\mu^-$ events that fall within the geometric acceptance of CDF and satisfy the kinematic requirements of the trigger used to collect the dataset.

Use Pythia MC to estimate:

<table>
<thead>
<tr>
<th>Method</th>
<th>Acceptance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$(CMU-CMU)</td>
<td>0.64%</td>
</tr>
<tr>
<td>$\alpha$(CMUP-CMU)</td>
<td>0.02%</td>
</tr>
<tr>
<td>$\alpha$(CMU-CMU &amp;&amp; CMUP-CMU)</td>
<td>5.90%</td>
</tr>
<tr>
<td>$\alpha$(CMU-CMU</td>
<td></td>
</tr>
</tbody>
</table>

Systematics include variations of $P_T(B)$ spectrum, detector material in simulation, and modeling of beam profile and offset.
Method: Efficiency

A quick summary of our efficiency estimates:

- determine trigger and reconstruction efficiencies from data (+/-10% syst associated w/ kinematic differences between data J/$\Psi$ and signal B$_s$)

- use realistic MC to determine efficiency of cuts on discriminating variables

- cross-check MC modeling of above by comparing MC to Data in sample of B+ $\rightarrow$ J/$\Psi$K+ (+/-5% syst)

- total uncertainty +/- 11% dominated by syst

(all uncertainties on this slide are relative uncertainties)
Method: Trigger Efficiency

- use $J/\psi \rightarrow \mu^+\mu^-$ samples
  - use triggers that require only one muon
  - unbiased muon used to parameterize $\varepsilon_{\text{trig}}(P_t, \eta)$

$$\varepsilon_{\text{signal}}^{signal} = \varepsilon_{\text{trig}} \otimes (P_T^{\mu^+}, \eta^{\mu^+}, P_T^{\mu^-}, \eta^{\mu^-})_{B_s \rightarrow \mu^+\mu^-}$$

$$= (85 +/- 3)\%$$

- uncertainty dominated by:
  - syst differences $J/\psi \rightarrow \mu^+\mu^-$ and $B_s \rightarrow \mu^+\mu^-$
- uncertainty also includes:
  - syst variations of parameterization, effects of 2-track correlations and statistics of sample
**Method: Trigger Efficiency**

- **0.00 < |\eta| < 0.10**
  - Plateau (A): 0.9029 ± 0.00058
  - Slope (B): 1.154 ± 0.029
  - Intercept (C): -0.06346 ± 0.3019

- **0.10 < |\eta| < 0.20**
  - Plateau (A): 0.9205 ± 0.005818
  - Slope (B): -0.9736 ± 0.1662
  - Intercept (C): 0.2378 ± 0.3359

- **0.20 < |\eta| < 0.30**
  - Plateau (A): 0.9235 ± 0.004446
  - Slope (B): -1.442 ± 0.2625
  - Intercept (C): -0.2942 ± 0.2678

- **0.30 < |\eta| < 0.40**
  - Plateau (A): 0.9414 ± 0.0004217
  - Slope (B): -2.01 ± 0.4274
  - Intercept (C): -0.5771 ± 0.2452

- **0.40 < |\eta| < 0.50**
  - Plateau (A): 0.9453 ± 0.003632
  - Slope (B): -2.01 ± 0.4274
  - Intercept (C): -0.5771 ± 0.2452

**\(\varepsilon_{\text{trig}}(P_T, \eta)\)**
**Method: Trigger Efficiency**

\[ \varepsilon_{\text{trig}}(P_t, \eta) \times \]

- \( B_s \rightarrow \mu^+\mu^- \text{ MC} \)
- \( P_T^B > 6 \text{ GeV/c}, |\eta^B| < 1.0 \)
- \( P_T^{\mu} > 6 \text{ GeV/c} \)
- \( |y^B| < 1.0 \)

\[ = (85 \pm 3)\% \]
**Method: COT Efficiency**

COT Efficiency is estimated by embedding COT hits from MC muons into real data

- occupancy effects correctly accounted for
- need to tune COT hit simulation

\[ \epsilon_{COT} = 99.22 \pm 0.01^{+0.68}_{-1.80} \%

(note: this is a double-leg efficiency)
Method: COT Efficiency

Need to tune COT hit simulation...

- reasonably well tuned at hit level
- have tunes which bracket data (for syst)
**Method:** COT Efficiency  
*(this page is all single-leg efficiency)*

**Systematics**
- Isolation dependence: $+0.14\% -0.86\%$
- Residual run/Pt depend: $\pm 0.29\%$
- 2 track correlations: $-0.27\%$
- Vary simulation tuning: $\pm 0.08\%$

**late 2003:**
$\varepsilon=99.63\pm 0.02\text{(stat)}\%$

**early 2002:**
$\varepsilon=99.61\pm 0.02\text{(stat)}\%$

**dominant systematic**
$\Delta=1.0\%$

**avg efficiency**
Method: Muon Efficiency

Muon Efficiency is estimated using $J/\psi \rightarrow \mu^+\mu^-$ and $Z \rightarrow \mu^+\mu^-$ data events collected with triggers that only require 1 muon

- unbiased muon used to estimate $\mu$ reco efficiency
- can compare $J/\psi$ and $Z$ events

$$\Delta r \Phi (\text{track-\mu stub}) / \sigma$$

$$\varepsilon_{\text{muon}} = 95.9 \pm 1.3 (\text{stat}) \pm 0.6 (\text{syst})\%$$

(note: this is a double-leg efficiency)
Method: SVX Efficiency

SVX Efficiency is estimated using $J/\psi \rightarrow \mu^+\mu^-$ data events
- no SVX requirements in our trigger path
- completely data determined

$\varepsilon_{SVX} = 74.5 \pm 0.3 \text{(stat)} \pm 2.2 \text{(syst)}\%$

(note: this is a double-leg efficiency)
Method: SVX Efficiency

Systematics

- $P_T$ dependence: $\pm 1.0\%$
- 2 track correlations: $\pm 0.7\%$
- Run dependence: $\pm 0.4\%$

(single-leg uncertainties)

variation used to assign a systematic single-leg efficiency

$\Delta \phi(\mu^+\mu^-) [\text{rad}]$

cross-check double-leg efficiency

(measured)

(estimated)

(spring/summer 2002)
Method: SVX Efficiency

Double-leg SVX efficiency sounds low:

\[ \varepsilon_{SVX} = 74.5 \pm 0.3(\text{stat}) \pm 2.2(\text{syst})\% \]

This corresponds to a single-leg efficiency of 86%.

The efficiency approximately breaks-down like this:

**Single-leg SVX efficiency using 2003 data**

- COT track traverses
  - \( \geq 3 \) active SVX layers: 97%
  - has \( \geq 3 \) SVX \( \phi \) hits associated: 91%
  - our more stringent requirements: 88%

average over full dataset: 86%
Have since improved pattern recognition so that:

- traverse $\geq 3$ active layers (unchanged)
- associate $\geq 3$ $r\phi$ hits (+3% absolute)
- our more stringent requirements (+2% absolute and flat)

old avg (88%) w/ same data

• next generation of this analysis will take advantage of these improvements
Method: Vertex Efficiency

Vertex Efficiency is estimated using $J/\psi \rightarrow \mu^+\mu^-$ data events collected with the same triggers as used for search.

- Data determined $J/\psi \rightarrow \mu^+\mu^-$ efficiency agrees with 2%.
- MC determined $B_s \rightarrow \mu^+\mu^-$ efficiency.

$\varepsilon_{vtx} = 94.7 \pm 0.2 (stat) \pm 1.9 (syst) \%$
Method: Efficiency of Final Selection Criteria

Determine efficiency of final selection criteria \((M, \lambda, \Delta \Phi, \text{Isol})\) using realistic MC simulation

- simulation tuned to detector (COT, SVX, etc.) hit level
- check modeling by comparing \(B^+ \rightarrow J/\psi K^+\) in data/MC

\[ \begin{align*}
|\eta| < 0.6 \\
\text{Pt} > 6 \text{ GeV/c} \\
\lambda > 100 \mu m
\end{align*} \]

\(\text{CDF} \)

- compare 2-track \((\mu^+ \mu^-)\) and 3-track distributions
- momentum scale and invariant mass resolution well modeled in MC (ie. \(\varepsilon(M) \text{ OK}\))
**Method:** Efficiency of Final Selection Criteria

Compare relative efficiencies of Iso and $\Delta \Phi$ cuts:

<table>
<thead>
<tr>
<th>Method</th>
<th>Data</th>
<th>MC</th>
<th>(Data/MC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iso $&gt; 0.6$</td>
<td>(95 +/- 2)%</td>
<td>(97 +/- 1)%</td>
<td>0.98 +/- 0.02</td>
</tr>
<tr>
<td>Iso $&gt; 0.7$</td>
<td>(88 +/- 2)%</td>
<td>(92 +/- 1)%</td>
<td>0.96 +/- 0.03</td>
</tr>
<tr>
<td>Iso $&gt; 0.8$</td>
<td>(68 +/- 2)%</td>
<td>(79 +/- 2)%</td>
<td>0.87 +/- 0.04</td>
</tr>
<tr>
<td>$\Delta \Phi &lt; 0.2$</td>
<td>(98 +/- 2)%</td>
<td>(97 +/- 1)%</td>
<td>1.00 +/- 0.02</td>
</tr>
<tr>
<td>$\Delta \Phi &lt; 0.1$</td>
<td>(89 +/- 3)%</td>
<td>(89 +/- 1)%</td>
<td>0.99 +/- 0.03</td>
</tr>
<tr>
<td>$\Delta \Phi &lt; 0.2$</td>
<td>(99 +/- 1)%</td>
<td>(99 +/- 1)%</td>
<td>1.00 +/- 0.01</td>
</tr>
<tr>
<td>$\Delta \Phi &lt; 0.1$</td>
<td>(92 +/- 2)%</td>
<td>(93 +/- 1)%</td>
<td>0.99 +/- 0.02</td>
</tr>
</tbody>
</table>

For search, we make cut at Iso $> 0.65$ $\Delta \Phi < 0.1$:

- ✔ efficiencies agree well
- ✔ $\varepsilon(\text{Iso})$ and $\varepsilon(\Delta \Phi)$ OK
Method: Efficiency of Final Selection Criteria

Monte Carlo slightly more isolated than Data:

- loose Isolation requirements OK, tighter requirements incur larger systematic
**Method: Efficiency of Final Selection Criteria**

Compare relative efficiencies of $\lambda$ cuts:

<table>
<thead>
<tr>
<th>$\lambda$ Cut</th>
<th>Data(obsvd)</th>
<th>MC(pred)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda &gt; 100 \mu m$</td>
<td>473 +/- 15</td>
<td>451 +/- 3</td>
</tr>
<tr>
<td>$\lambda &gt; 150 \mu m$</td>
<td>415 +/- 13</td>
<td>408 +/- 4</td>
</tr>
<tr>
<td>$\lambda &gt; 200 \mu m$</td>
<td>378 +/- 12</td>
<td>369 +/- 4</td>
</tr>
</tbody>
</table>

Prediction normalized to $\lambda > 50 \mu m$ requirement

For search, we make cut at $\lambda > 150$-200 $\mu m$:
- Relative efficiency agrees well
- $\varepsilon(\lambda)$ OK

In general, MC tracks data efficiencies to better than 5%.
Use MC determined $\varepsilon_{\text{final}}$ with +/-5% (relative) systematic.
**Method: Optimization**

We now have in hand:

1. Background estimate
2. Estimate of total acceptance*efficiency
3. Their associated uncertainties

Let’s Optimize!

Considered >100 different sets of \((M_{\mu \mu}, \lambda, \Delta \Phi, \text{Iso})\) requirements with \(\varepsilon_{\text{final}} = 28\% - 78\%:\)

- use set which minimized *a priori* expected limit, \(<\text{BR}>\)
- minima shallow, \(<\text{BR}>\) varying by <5\% over wide range
- same results for integrated luminosities up to 400 pb-1
- same optimal selection criteria for \(B_d \rightarrow \mu^+ \mu^-\) search
Method: Efficiencies and Uncertainties

Using the optimal selection criteria...

**Efficiencies**

Acceptance : 6.6%
- $\varepsilon_{\text{trig}}$ : 85%
- $\varepsilon_{\text{reco}}$ : 71%
- $\varepsilon_{\text{vtx}}$ : 95%
- $\varepsilon_{\text{final}}$ : 54%

$\alpha \times \varepsilon_{\text{total}}$ : 2.0%

(\(\alpha\) is determined for $P_T(B) > 6$ GeV && $|y| < 1$)

**Uncertainties**

Background stat : 27%
syst : 5%

**Total** : 30%

Acceptance : 7%
- $\varepsilon_{\text{trig}}$ : 4%
- $\varepsilon_{\text{reco}} \times \varepsilon_{\text{vtx}}$ : 4%
- $\varepsilon_{\text{final}}$ : 5%

$\alpha \times \varepsilon_{\text{total}}$ : 10%

Luminosity : 6%

Normalization : 17%

$\alpha \times \varepsilon_{\text{total}} \times L \times \sigma_{BS}$ : 21%

(These are all relative uncertainties)
The optimal set of final selection criteria is:

\[ \Delta M_{\mu\mu} = +/- 80 \text{ MeV around } M(B_s) = 5.369 \text{ GeV} \]
\[ \lambda > 200 \mu\text{m} \]
\[ \Delta \Phi < 0.10 \text{ rad} \]
\[ \text{Isolation} > 0.65 \]

which corresponds to:

\[ \alpha \epsilon_{\text{total}} = (2.0 +/- 0.2)\% \]

single event sensitivity = 1.6 x 10^{-7}

\[ <\text{Bgd}> \text{ in } 171 \text{ pb}^{-1} = 1.1 +/- 0.3 \text{ events} \]

(\(\alpha \epsilon \text{ & Bgd are unchanged for mass window centered on 5.279 GeV for the Bd} \rightarrow \mu+\mu- \text{ search})
Results

\[ \text{BR}(B^0_{s(d)} \rightarrow \mu^+ \mu^-) < 1.5 \times 10^{-7} \]

\[ \text{BR}(B_s \rightarrow \mu^+ \mu^-) < 5.8 \times 10^{-7} \]

\[ \text{BR}(B_d \rightarrow \mu^+ \mu^-) < 1.5 \times 10^{-7} \]

These are both the best limits in the world for these decays.
Results

This new limit...

Eliminates this entire plane (raising $M_A$ to $\sim 400$ GeV opens it back up).

Just begins to eat into allowed parameter space in this plane.
Conclusions

• Have searched for $B_s \rightarrow \mu^+\mu^-$ and $B_d \rightarrow \mu^+\mu^-$ decays using 171 pb-1 of CDFII data.

• Observed 1 and expected 1.1 +/- 0.3 background events.

• Established these world best limits at 90 (95)% CL:
  \[
  \text{BR}(B_s \rightarrow \mu^+\mu^-) < 5.8 (7.5) \times 10^{-7} \\
  \text{BR}(B_d \rightarrow \mu^+\mu^-) < 1.5 (1.9) \times 10^{-7}
  \]
  (submitted to PRL)

• Yields significant reduction in allowed parameter space of some models.

• Expect significant improvements with:
  – more data, more acceptance, more Bgd rejection
Backup: COT “Aging”

- recall that COT geometry consists of 8 “Super Layers” (SL)
  - 12 sense wires in each SL
  - SL 1, 3, 5, 6 are stereo
  - SL 2, 4, 6, 8 are axial
  - axial SL used in L1 trigger

- unexpected reduction in gain on inner 4 SL

- outer 4 SL not significantly affected

how does this affect tracking?
use RAREB triggered evts
• remove (XFT-trgd) muons
• look at tracks $P_T > 1$ GeV
• divided into 7 run ranges of approximately 25/pb each
• 3 (early, middle, late) shown
• averaged over phi:
  ✓ $<N_{ax}>$ drops by 0.8 cnt
  ✓ $<N_{st}>$ drops by 1.9 cnts
  ✓ Smaller than drops induced when varying simulation tune parameters to get syst for COT efficiency (< 0.1%)
  ✓ concentrated on inner 3 layers
• same thing, binned in $\phi$, all run ranges shown
• effect dominated by region around $\phi=4$ rad
• multi-track (geometric) correlations important
Backup: COT “Aging”

- generate relative efficiency in bins of $\phi$ using jet triggered events; also bin in run ranges, each corresponding to $\sim$25pb-1
- fold $B_s \rightarrow \mu\mu (\phi,\phi)$ spectrum with each of these curves
- lumi-weighted result: 0.9946
- difference w/ 1. assigned as additional systematic to the double-leg efficiency
Recently swapped some wire planes out of chamber and had them analyzed. Visibly very different than an unused wire. Potential wire from same plane looked like new.
Further analysis reveals that there’s \( \sim 300 \) nm of hydro-carbons on the affected sense wires. No evidence of silicas (ie. gas system is clean).
Backup: COT “Aging”

What are we doing?

• voltage reduced on inner 4 SL while we investigate problem and possible solutions
• have modified trigger to cope with the reduced gains on the inner axial SLs (more on next slide)
• have assembled an international committee of experts
• plan to increase gas flow by x4 soon (hope to return to nominal voltages after that), and by another x5 during summer shutdown (should mitigate effect)
• are investigating the possibility of using a different gas to further mitigate, and perhaps recover, the effect (using a pulled wire plane in test chamber to study)
Is this data useful?

- track trigger modified to accommodate this (10-20% reduction in yields, depending on N_{track})
- COT track efficiency for leptons from W and Z unaffected and >99% (determined using missing energy triggers)
- track efficiency for pions and muons reduced 5-10% in P_T range 1-10 GeV (recall, we started with 99.6%)
- efficiency for adding SVX hits unchanged (thanks to ISL)
- with SVX hits attached, resolutions nearly unchanged
- 50 pb-1 of data like this so far
- this data usable for physics… will require dedicated simulation effort
Backup: Momentum Scale

Absolute scale set by pinning M(J/ψ) to PDG.

Cross-check using other resonances:

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Measured (stat. only)</th>
<th>PDG (stat+sys)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_s \rightarrow \pi^+\pi^-$</td>
<td>497.36 ± 0.04</td>
<td>497.67 ± 0.03</td>
</tr>
<tr>
<td>$\Upsilon \rightarrow \mu^+\mu^-$</td>
<td>9461 ± 5</td>
<td>9460.30 ± 0.26</td>
</tr>
<tr>
<td>$D^0 \rightarrow K^-\pi^+$</td>
<td>1864.15 ± 0.10</td>
<td>1864.5 ± 0.5</td>
</tr>
<tr>
<td>$D^+ \rightarrow K\pi\pi$</td>
<td>1868.65 ± 0.07</td>
<td>1869.3 ± 0.5</td>
</tr>
<tr>
<td>$D^+ \rightarrow \phi\pi^+$</td>
<td>1868.95 ± 0.37</td>
<td>1869.3 ± 0.5</td>
</tr>
<tr>
<td>$D^+_s \rightarrow \phi\pi^+$</td>
<td>1968.20 ± 0.26</td>
<td>1968.6 ± 0.6</td>
</tr>
<tr>
<td>$B^+ \rightarrow J/\psi K^+$</td>
<td>5278.2 ± 2.2</td>
<td>5279.0 ± 0.5</td>
</tr>
</tbody>
</table>

(all values in MeV/c²; charge conjugation implied)
We chose to use an absolute normalization.

\[ \sigma_{B_s} = \frac{f_s}{f_u} \cdot \sigma_{B_d} \]

- \( \sigma_{B_d} \) measured from CDF PRD 65 (2002) 052005.
- straight forward (and same as RunI)
- with present statistics, contributes to total uncertainty at same level as relative normalization to \( B^+ \rightarrow J/\psi K^+ \).