First observation of $B^0_s \to \mu^+\mu^-$ with CMS

Urs Langenegger

(PSI)

EPFL

2014/02/17

- Introduction
  - theory
  - other experiments
- Analysis
  - and results
- Outlook

PRL, 111, 101804
$B^0_s \rightarrow \mu^+ \mu^-$: A (very) rare $B$ decay

- Decays highly suppressed in Standard Model
  - effective FCNC: $Z$ penguin and box graphs
  - helicity suppressed
  - SM expectation:
    \[
    \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.2 \pm 0.2) \times 10^{-9}
    \]
    \[
    \mathcal{B}(B^0 \rightarrow \mu^+ \mu^-) = (1.0 \pm 0.1) \times 10^{-10}
    \]
  - Cabibbo-enhancement ($|V_{ts}| > |V_{td}|$)

- 'Golden' channels
  - clean experimental signature:
    find two muons
    reconstruct their invariant mass
  - small theoretical uncertainties
    . . . or so they tell us
  - high sensitivity to new physics
    orthogonal to
    $b \rightarrow s \gamma$, $\mu \rightarrow e \gamma$, . . .
    direct searches

Urs Langenegger
First Observation of $B_s^0 \rightarrow \mu^+ \mu^-$ with CMS (2014/02/17)
Search for new physics

- Indirect sensitivity to new physics
  - no helicity suppression
  - scalar couplings
  - pseudo-scalar couplings
  - in particular ‘extended’ Higgs boson sectors
    2HDM: $\mathcal{B} \propto (\tan \beta)^4, m_{H^+}$
    MSSM: $\mathcal{B} \propto (\tan \beta)^6$

- Two approaches
  - ‘top-down’: specific model implications
  - model-independent: effective field theory

⇒ Correlations
  - $\mathcal{B}(B_s^0 \to \mu^+\mu^-)$
  - $(g-2)_\mu$
  - $\mathcal{B}(\mu \to e\gamma)$
    historic plots, for illustration only
First Observation of $B_s^0 \rightarrow \mu^+ \mu^-$ with CMS (2014/02/17)
The decay $B^0_s \rightarrow \mu^+ \mu^-$ in the SM

- Branching fraction in the SM
  Box diagram is suppressed by $m_W^2/m_t^2 \approx 1/4$ with respect to Penguin

  \[
  B(B^0_s \rightarrow \mu^+ \mu^-) = \frac{G_F^2 \alpha^2 m_{B^0_s} \tau_{B^0_s} V_{tb} V_{ts}^*}{16\pi^3} |V_{tb} V_{ts}^*|^2 \sqrt{1 - \frac{4m_\mu^2}{m_{B^0_s}^2}} 2m_\mu \left( -i \frac{f_{B^0_s}}{2} C_{10} \right)^2
  \]

  (equivalent expression for $B^0 \rightarrow \mu^+ \mu^-$)

- Two approaches for numerical calculation
  1) decay constant and CKM factors
  2) bag parameter and $\Delta m_s$

- Corrected a factor 2 in the previous literature
  and missed another (small) term in NLO calculation (cf. Misiak and Urban)

- Formula ignores
  - finite width difference in $B^0_s$ system $\rightarrow 10\%$ effect!
  - soft photon radiation $\rightarrow$ all experiments use Photos

NP, B400, 225 (1993)
PL, B451, 161 (1999)
Uncertainties of SM expectation

- Parametric/experimental uncertainties
  - lifetime/mass of $B_s^0$ mesons
  - $|V_{ts}|$, $|V_{tb}|$, and $|V_{cb}|$
  - $m_t$

- Non-perturbative input from lattice QCD calculations
  - $f_{B_s^0} = 227.7 \pm 4.5$ MeV (FLAG)

- NNLO QCD calculation
  - completed recently (Misiak et al)

- NLO EW corrections
  - completed recently (Bobeth et al)

\[
\mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = 3.65 \times 10^{-9} \\
\times \left( \frac{m_t}{173.1 \text{ GeV}} \right)^{3.06} \times \left( \frac{\alpha_s(m_Z)}{0.1184} \right)^{-0.18} \\
\times \left( \frac{f_{B_s^0}}{227.7 \text{ MeV}} \right)^2 \times \left( \frac{\tau_{B_s^0}}{1.615 \text{ ps}} \right) \\
\times \left( \frac{|V_{tb}V_{ts}/V_{cb}|}{0.980} \right)^2 \times \left( \frac{|V_{cb}|}{0.0424} \right)^2 \\
= (3.65 \pm 0.23) \times 10^{-9}
\]

$\Rightarrow$ CKM uncertainties now largest contribution to systematic error!
On theory predictions

• Quotation from A. Buras in 1309.7791:

While experimentalists from CMS and LHCb should be congratulated on the measurements on such low branching ratios, their result for \( B_{s}^{0} \to \mu^{+} \mu^{-} \) has been predicted by theorists more than a decade ago.

In contrast to what is stated usually in the literature, the most important result of this paper was not the reduction of the scale uncertainty due to the choice of the scale in \( m_{t} \) but the inclusion of a factor of two in the branching ratios for \( B_{s(d)}^{0} \to \mu^{+} \mu^{-} \) which was missed in the previous literature. (they also missed a factor)

• parametric changes obvious e.g. lifetime, \( f_{B} \), \( m_{t}(m_{t}) \)

• not complete error! perturbative contribution

Fig. 6. The same as in fig. 5 for \( B(B_{s} \to \mu \bar{\mu}) \) with \( \tau(B_{s}) = 1.28 \text{ ps}, F_{B} = 200 \text{ MeV} \) and \( |V_{ts}| = 0.041 \).

In analogy to (4.5) and (4.6) we have, without and including the QCD correction

\[
1.44 \times 10^{-9} \leq B(B \to \mu \bar{\mu}) \leq 1.91 \times 10^{-9}, \quad (4.13)
\]

\[
1.76 \times 10^{-9} \leq B(B \to \mu \bar{\mu}) \leq 1.79 \times 10^{-9}, \quad (4.14)
\]

respectively.
Theory predictions over the years

(NB: Top quark discovery in 1995)
A highly competitive field

- \( B^0_s \rightarrow \mu^+\mu^- \) is very much sought after
  - Tevatron
    - CDF best sensitivity
    - DØ best result
  - LHC
  - Accelerated activity
    - factor > 100 in the last decade
    - factor 10 in the past year
    - CDF: ‘indication’ in June 2011
    - LHCb: ‘evidence’ in November 2012

⇒ Announcements in July 2013:
  - CMS at 4.3\( \sigma \)
  - LHCb at 4.0\( \sigma \)
CDF’s result

In summary, we have performed a search for $B^0 \to \mu^+\mu^-$ and $B_s^0 \to \mu^+\mu^-$ decays using 7 fb$^{-1}$ of integrated luminosity collected by the CDF II detector at the Fermilab Tevatron. The data in the $B^0$ search region are consistent with background expectations and the world’s most stringent upper limit on $B(B^0 \to \mu^+\mu^-)$ is established. The data in the $B_s^0$ search region are in excess of the background predictions. A fit to the data determines $B(B_s^0 \to \mu^+\mu^-) = (1.8^{+1.1}_{-0.9}) \times 10^{-8}$ including all uncertainties. Although of moderate statistical significance, this is the first indication of a $B_s^0 \to \mu^+\mu^-$ signal.

- Released before EPS-11
  - CMS and LHCb with upper limits
  - combination ‘excluded’ CDF

- Excess diminished
  - more data
  - conclusions toned down

⇒ Statistical fluctuation
  → background can peak at $B_s^0$ mass

\[\text{PRL, 107, 191801} \quad 1107.2304v1(!)\]
LHCb’s results in 2012

- ‘Strong constraint’ (at 95\% C.L.)
  - 1.0 \text{ fb}^{-1} at 7 \text{ TeV}
  - background model (expo)
    - combinatorial bg
      \[ b \rightarrow c\mu\nu \rightarrow \mu\mu X \]
    - \( \mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 4.5 \times 10^{-9} \) (measured)
    - \( \mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) < 7.2 \times 10^{-9} \) (expected)
  - > 1\( \sigma \) downward fluctuation (or systematic issue)

- ‘First evidence’
  - 1.0 \text{ fb}^{-1} at 7 \text{ TeV} + 1.1 \text{ fb}^{-1} at 8 \text{ TeV}
  - changed background model
    - combinatorial bg (expo)
    - \( B \rightarrow h\mu\nu \) and \( B \rightarrow h\mu\mu \)
  - \( \mathcal{B}(B_s^0 \rightarrow \mu^+\mu^-) = (3.2^{+1.5}_{-1.2}) \times 10^{-9} \)
  - 3.5\( \sigma \) evidence
    - no exp. sensitivity given (‘rumours’: \( \approx 3.0\sigma \))
  - 1.5\( \sigma \) (in)consistency between ’11 and ’12
Experimental aspects
Analysis overview

- **Signal** $B_s^0 \rightarrow \mu^+\mu^-$
  - two muons from one decay vertex
  - well reconstructed secondary vertex
  - momentum aligned with flight direction
  - isolated, mass around $m_{B_s^0}$

- **Background**
  - combinatorial (from sidebands)
    - two semileptonic ($B$) decays (gluon splitting)
    - one semileptonic ($B$) decay and one misidentified hadron
  - rare single $B$ decays (from MC simulation)
    - non-peaking, e.g. $B_s^0 \rightarrow K^-\mu^+\nu$, $\Lambda_b \rightarrow p\mu^+\nu$
    - peaking, e.g. $B_s^0 \rightarrow K^+K^-$

- **Blind analysis**
  - Critical issues
    - optimized selection
    - muon misidentification probability
    - pileup (isolation)
Methodology

- **Measurement of** $B^0_s \rightarrow \mu^+ \mu^-$ **relative to normalization channel:**
  - $B^\pm \rightarrow \Upsilon J/\psi K^\pm$, with well-known branching fraction
  - (nearly) identical selection to reduce systematic uncertainties

\[ B(B^0_s \rightarrow \mu^+ \mu^-) = \frac{n_{B^0_s}^{\text{obs}}}{\varepsilon_{B^0_s} N_{B^0_s}} = \frac{n_{B^0_s}^{\text{obs}}}{\varepsilon_{B^0_s} \mathcal{L} \sigma(pp \rightarrow B^0_s)} \]

\[ = \frac{n_{B^0_s}^{\text{obs}}}{N(B^\pm \rightarrow J/\psi K^\pm)} \frac{A_{B^+}^{\text{ana}}}{A_{B^0_s}^{\text{ana}}} \frac{\varepsilon_{B^+}^{\mu \text{ana}}}{\varepsilon_{B^0_s}^{\mu \text{ana}}} \frac{\varepsilon_{B^+}^{\text{trig}}}{\varepsilon_{B^0_s}^{\text{trig}}} f_{B^+} f_s B(B^+ \rightarrow J/\psi [\mu^+ \mu^-]K) \]

- **Reconstructed exclusive decays**
  - $B^\pm \rightarrow J/\psi K^\pm$: normalization and studies
  - $B^0_s \rightarrow J/\psi \phi$: $B^0_s$ signal MC validation
  - $J/\psi, \Upsilon(1S) \rightarrow \mu^+ \mu^-$: mass calibration

- **Two ‘channels’** per dataset
  - **barrel:** both muons with $|\eta| < 1.4$
    - very good mass resolution, low background
  - **endcap:** 1-2 muon(s) with $|\eta| > 1.4$
    - add more statistics

---

Urs Langenegger  
*First Observation of $B^0_s \rightarrow \mu^+ \mu^-$ with CMS (2014/02/17)*
The Large Hadron Collider (LHC)
The CMS detector

- Design prioritization
  - lepton ID
    - $\rightarrow$ muons
  - $b/\tau$ tagging
    - $\rightarrow$ tracking/vertexing
  - jets and $E_T$

<table>
<thead>
<tr>
<th>Component</th>
<th>Characteristics</th>
<th>Resolutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel</td>
<td>3/2 Si layers</td>
<td>$\delta_z \approx 20,\mu$m, $\delta_\phi \approx 10,\mu$m</td>
</tr>
<tr>
<td>Tracker</td>
<td>10/12 Si strips</td>
<td>$\delta(p_\perp)/p_\perp \approx 1%$</td>
</tr>
<tr>
<td>ECAL</td>
<td>PbWO$_4$</td>
<td>$\delta E/E \approx 3% / \sqrt{E} \oplus 0.5%$</td>
</tr>
<tr>
<td>HCAL (B)</td>
<td>Brass/Sc, $&gt; 7.2\lambda$</td>
<td>$\delta E/E \approx 100\sqrt{E%}$</td>
</tr>
<tr>
<td>HCAL (F)</td>
<td>Fe/Quartz</td>
<td>$\delta(E_T) \approx 0.98\sqrt{\sum E_T}$</td>
</tr>
<tr>
<td>Magnet</td>
<td>3.8 T solenoid</td>
<td>$\delta(p_\perp)/p_\perp \approx 10%$ (STA)</td>
</tr>
<tr>
<td>Muons</td>
<td>DT/CSC + RPC</td>
<td></td>
</tr>
</tbody>
</table>

Weight: 12'500 t
Length: 21.6 m
Diameter: 15 m
Magnetic field: 3.8 T

Tracking resolution:
impact parameter $\approx 15\,\mu$m

Primary vertex resolution:
$\Delta z \approx 20 - 80\,\mu$m
Muon reconstruction and identification

- Large muon acceptance $|\eta| < 2.4$
  - drift tubes
  - cathode strip chambers
  - resistive plate chambers
- Muon reconstruction/identification
  - global muon: outside-in reconstruction
  - tight muon: quality criteria against ‘fakes’
  - BDT: reduce ‘fakes’ by another 50%
    - track ‘kinks’
    - inner-outer matching
    - muon detector information

Muon misidentification

$\epsilon(\mu|\pi) \leq 0.15\%$
$\epsilon(\mu|K) \leq 0.20\%$
$\epsilon(\mu|p) \leq 0.10\%$

measured in data and MC

$D^{*+} \rightarrow D^0\pi^+ \rightarrow K^-\pi^+\pi^+_S,$
$K^0_S \rightarrow \pi^+\pi^-, \Lambda \rightarrow p\pi^-$

(50% syst. uncertainty)
3D vertexing

- All silicon tracker
  - high granularity, low occupancy
  - very well described by MC simulation
- Pixel detector
  - $100 \times 150 \mu m^2$ pixel size
  - substantial charge sharing (low $V_{bias}$)
  - excellent resolution in $r\phi$ and $z$

⇒ Essential for vertexing in high-pileup environment!
Trigger: $B_{s}^{0} \rightarrow \mu^{+}\mu^{-}$ and $B^{\pm} \rightarrow J/\psi K^{\pm}$

- **Dimuon trigger**
  - L1 (hardware) trigger
    - a few kHz at peak luminosity
  - High-level trigger
    - full tracking and vertexing

- **HLT $B_{s}^{0} \rightarrow \mu^{+}\mu^{-}$**
  - two muons with opposite charge
  - $p_{\perp} > 4.0(3.5)$ GeV, $\mathcal{P}(\chi^{2}/dof) > 0.5\%$
  - inv. mass $4.8 < m_{\mu\mu} < 6.0$ GeV

- **HLT $B^{\pm} \rightarrow J/\psi K^{\pm}$ and $B_{s}^{0} \rightarrow J/\psi \phi$**
  - two muons with opposite charge
    - $2.9 < m_{\mu\mu} < 3.3$ GeV
  - $\cos \alpha > 0.9$, $\mathcal{P}(\chi^{2}/dof) > 15\%$
  - ‘displaced’ $J/\psi$

**Trigger efficiency 40 − 80%**
- after analysis selection
- time-dependence in MC

**Determination**
- MC simulation
- data
  - systematics from difference
Candidate selection
Discriminating variables

- **Vertexing**
  - primary vertex w/o the two muons
  - secondary vertex of the two muons

\[
\alpha_{3D} \quad p_{\mu\mu} \quad \mu \quad \delta_{3D} \quad l_{3D}
\]

![Graphs showing distributions of vertexing variables](image)

**Urs Langenegger**

*First Observation of $B_S^0 \rightarrow \mu^+\mu^-$ with CMS (2014/02/17)*
Isolation Variables

- **Relative isolation of dimuon**
  \[ I = \frac{p_\perp(\mu^+\mu^-)}{p_\perp(\mu^+\mu^-) + \sum_{\Delta R < 0.7} p_\perp} \]
  - in cone around dimuon momentum
  - for tracks in cone with $\Delta R < 0.7$
    - with $p_\perp > 0.9$ GeV
    - either associated to same PV as candidate
    - or with $d_{ca} < 500$ $\mu$m
      ($d_{ca} =$ distance of closest approach)

- **Muon isolation**
  - $\Delta R < 0.5$, $p_\perp > 0.5$ GeV and $d_{ca} < 1$ mm

- **Number of tracks close to SV**
  - $p_\perp > 0.5$ GeV and $d_{ca} < 300$ $\mu$m

- **Closest track to SV**
  - $d_{ca}^0$
Normalization and control samples

- **Selection of** $B^{\pm} \to J/\psi K^{\pm}$ and $B_{s}^{0} \to J/\psi \phi$
  - $J/\psi$ plus 1-2 tracks assumed to be kaons
  - **identical** variables, except for isolation: ignore additional tracks
  - vertex $\chi^{2}$: use dimuon vertex (nominal vtx w/ all tracks)
  - $\phi$ selection (for $B_{s}^{0}$)

- **Mass parametrization**
  - **signal**
    - barrel: double Gaussian
    - endcap: single Gaussian
  - **background**
    - partial reco: error function
    - CKM-suppressed: Landau function
    - combinatorial: exponential

⇒ **Sideband-subtracted plots!**
  - **essential for MC validation!**
Comparison of sideband-subtracted distributions

in general good agreement

$B^0_s \rightarrow J/\psi \phi$

$B^\pm \rightarrow J/\psi K^\pm$

Urs Langenegger

First Observation of $B^0_s \rightarrow \mu^+ \mu^-$ with CMS (2014/02/17)
MVA selection

- **BDT training**
  - TMVA framework
  - signal: $\psi^0 \rightarrow \mu^+ \mu^-$ MC simulation
  - background: data dimuon sidebands
  - avoid selection bias
    - split data randomly into three subsets (0,1,2)
    - train on 0, test on 1, apply on 2. etc.
  - in each channel, have 3 BDTs

- **Studies**
  - selection efficiency in high and low mass sidebands
  - signal MC with shifted mass
  - pileup for dimuons (and normalization/control samples)

- **Selection of normalization and control samples with identical BDTs**
  - slightly modified variables (e.g. dimuon vertex fit quality)
  - isolation variables: ignore hadronic particles from $B$ decay

4 channels
- 2011 (5 fb$^{-1}$)
  - barrel
  - endcap
- 2012 (20 fb$^{-1}$)
  - barrel
  - endcap

→ 12 BDTs
output per channel combined
**MC simulation vs. data (II)**

- Differences between data and MC used as systematic uncertainties
  - $B^\pm \rightarrow J/\psi K^\pm$: <3.0%
  - $B^0_s \rightarrow J/\psi \phi$: <9.5% (2011)
    - <3.5% (2012)
      - used for signal $\varepsilon$ uncertainty

- Mass scale uncertainty
  - $\psi$ and $\Upsilon$ to dimuon decays
    - barrel: <6 MeV
    - endcap: <7 MeV
  - correction applied
Pileup dependence?
**Pileup independence**

- **Average number of interactions per bunch crossing**
  - 2011: \( \approx 9 \), 2012: \( \approx 21 \)

- **Pileup independence checked**
  - Signal MC event samples with pileup
    - every single variable used in BDT is shown to be pileup independent
  - Data studies with BDT output distribution vs. \( N_{PV} \)
    - mean and RMS, efficiency of BDT requirement

⇒ **No significant pileup dependence observed**

\[ \chi^2/\text{dof} = 7.7/15 \text{ (pol0)} \]
\[ p_0 = 0.0728 \pm 0.0005 \]
\[ p_1 = 0.0002 \pm 0.0001 \]
Result determination

- Two approaches for determination of result
  - 1D-BDT
    - (optimized) cut on BDT output
      
      |   | barrel | endcap |
      |---|-------|-------|
      | 2011 | 0.29  | 0.29  |
      | 2012 | 0.38  | 0.39  |

  independent data set used

  → 4 mass distributions

  - categorized-BDT
    - per channel 2-4 categories (BDT bins)
      
      |   |   |   |   |
      |---|---|---|---|
      | 2011 barrel | 0.10 | 0.31 | -  |
      | 2011 endcap  | 0.10 | 0.29 | -  |
      | 2012 barrel  | 0.10 | 0.23 | 0.33| 0.44|
      | 2012 endcap  | 0.10 | 0.22 | 0.29| 0.45|

  equalized expected signal yield

  → 12 mass distributions

- Strategy (decided before unblinding)
  - $B^0 \rightarrow \mu^+\mu^-$ 1D-BDT: UL with CL$_S$
  - $B^0_s \rightarrow \mu^+\mu^-$ categorized-BDT: UML fit
    (best expected sensitivity with categorized-BDT)
Unbinned maximum likelihood fit

- Probability distribution functions
  - peaking components: Crystal Ball (w/ and w/o Gaussian)
  - combinatorial background: polynomial of first degree
  - \( b \rightarrow u \mu \bar{\nu} \) background: Gaussian kernels for MC-predicted mixture
  - per-event mass resolution included (excellent data/MC agreement)

- Fit for \( B_s^0 \) and \( B^0 \) simultaneously
  - peaking background
    - constrained to expectation
    - normalized to measured \( B^+ \) yield
    - yield cross checked on independent data set
  - semileptonic background
    - fixed shape
    - floating normalization within uncertainties (dominated by unknown \( \Lambda_b \rightarrow p \mu \nu \))
  - combinatorial background
    - no constraint on slope
    - validated with independent data set
    - varied functional form

---

Urs Langenegger First Observation of \( B_s^0 \rightarrow \mu^+ \mu^- \) with CMS (2014/02/17)
Expectations and observation (1D-BDT)

Summary for 1D-BDT approach in $B^0$ and $B^0_S$ ‘signal’ regions

- ‘signal’ regions

| $B^0$ | $5.20 < m < 5.30$ GeV |
| $B^0_S$ | $5.30 < m < 5.45$ GeV |

<table>
<thead>
<tr>
<th></th>
<th>2011 barrel</th>
<th>2012 barrel</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^0 \rightarrow \mu^+\mu^-$</td>
<td>$B^0_S \rightarrow \mu^+\mu^-$</td>
<td>$B^0 \rightarrow \mu^+\mu^-$</td>
</tr>
<tr>
<td>$\varepsilon_{\text{tot}}[%]$</td>
<td>$0.33 \pm 0.03$</td>
<td>$0.30 \pm 0.04$</td>
</tr>
<tr>
<td>$N_{\text{signal}}^{\exp}$</td>
<td>$0.27 \pm 0.03$</td>
<td>$2.97 \pm 0.44$</td>
</tr>
<tr>
<td>$N_{\text{total}}^{\exp}$</td>
<td>$1.3 \pm 0.8$</td>
<td>$3.6 \pm 0.6$</td>
</tr>
<tr>
<td>$N_{\text{obs}}$</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>2011 endcap</th>
<th>2012 endcap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^0 \rightarrow \mu^+\mu^-$</td>
<td>$B^0_S \rightarrow \mu^+\mu^-$</td>
<td>$B^0 \rightarrow \mu^+\mu^-$</td>
</tr>
<tr>
<td>$\varepsilon_{\text{tot}}[%]$</td>
<td>$0.20 \pm 0.02$</td>
<td>$0.20 \pm 0.02$</td>
</tr>
<tr>
<td>$N_{\text{signal}}^{\exp}$</td>
<td>$0.11 \pm 0.01$</td>
<td>$1.28 \pm 0.19$</td>
</tr>
<tr>
<td>$N_{\text{total}}^{\exp}$</td>
<td>$1.5 \pm 0.6$</td>
<td>$2.6 \pm 0.5$</td>
</tr>
<tr>
<td>$N_{\text{obs}}$</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Unbinned maximum likelihood fit (II)

- Illustration of the UML fits
  - highest (2\textsuperscript{nd} highest) S/B categories for (barrel, endcap) × (2011, 2012)

- Signal candidate distributions
  - consistent with (signal!) expectations
    - kinematic variables
    - vertexing variables
    - number of $pp$ interactions

Urs Langenegger

First Observation of $B^0_S \rightarrow \mu^+ \mu^-$ with CMS (2014/02/17)
All BDT bins

- CMS - $L = 5 \text{ fb}^{-1}$ $\sqrt{s} = 7 \text{ TeV}$ - Barrel
  - $0.1 < \text{BDT} < 0.31$
- CMS - $L = 5 \text{ fb}^{-1}$ $\sqrt{s} = 7 \text{ TeV}$ - Barrel
  - $0.31 < \text{BDT} < 1.00$
- CMS - $L = 20 \text{ fb}^{-1}$ $\sqrt{s} = 8 \text{ TeV}$ - Barrel
  - $0.1 < \text{BDT} < 0.23$
- CMS - $L = 20 \text{ fb}^{-1}$ $\sqrt{s} = 8 \text{ TeV}$ - Barrel
  - $0.23 < \text{BDT} < 0.33$
- CMS - $L = 20 \text{ fb}^{-1}$ $\sqrt{s} = 8 \text{ TeV}$ - Endcap
  - $0.1 < \text{BDT} < 0.22$
- CMS - $L = 20 \text{ fb}^{-1}$ $\sqrt{s} = 8 \text{ TeV}$ - Endcap
  - $0.22 < \text{BDT} < 0.33$
Results

- Results of the UML fit in the categorized-BDT approach

\[ \mathcal{B}(B^+_s \rightarrow \mu^+\mu^-) = (3.0^{+1.0}_{-0.9}) \times 10^{-9} \]
\[ \mathcal{B}(B^0 \rightarrow \mu^+\mu^-) = (3.5^{+2.1}_{-1.8}) \times 10^{-10} \]

\[ \mathcal{B}(B^0_s \rightarrow \mu^+\mu^-) = [3.0^{+0.9}_{-0.8} \text{(stat)} + 0.6 \text{(syst)}] \times 10^{-9} \]

- Significances
  - \( B^0_s \rightarrow \mu^+\mu^- : 4.3\sigma \) [expected 4.8\( \sigma \) (median)]
  - \( B^0 \rightarrow \mu^+\mu^- : 2.0\sigma \)

- UML fit in the 1D-BDT approach
  - \( B^0_s \rightarrow \mu^+\mu^- : 4.8\sigma \) [expected 4.7\( \sigma \) (median)]

- External input
  - \( \mathcal{B}(B^+) = (6.0 \pm 0.2) \times 10^{-5} \) (PDG)
  - \( f_s/f_u = 0.256 \pm 0.020 \pm 0.013_{\text{extrapol. (LHCb)}} \)

- Upper limit on \( B^0 \rightarrow \mu^+\mu^- \)

\[ \mathcal{B}(B^0 \rightarrow \mu^+\mu^-) < 1.1 \times 10^{-9} \text{ 95\%CL} \]
Systematics

- Hadronization probability ratio $f_s/f_u$ from LHCb [JHEP 04, 001 (2013)]
  - additional 5% systematics for possible $p_\perp$ or $\eta$ dependence
  - in-situ studies show no $p_\perp$ dependence
  - ratio of $B^\pm \to J/\psi K^\pm$ vs $B_s^0 \to J/\psi \phi$

- Rare decays
  - hadron to muon misidentification probability
    $K_S^0 \to \pi^+ \pi^-$, $\Lambda \to p\pi$, and $D^{*+} \to D^0 (K^- \pi^+) \pi^+$
    50% uncertainty, pions/kaons/protons uncorrelated
  - branching fraction uncertainties
  - $\Lambda_b \to p\mu\bar{\nu}$:
    large range of predictions in literature (w/o JHEP, 1109, 106)
    take average $(6.5 \times 10^{-4})$ and assign 100% uncertainty
    (note that invariant mass covers $B_s^0$ signal region, using EvtGen ‘phase space’ model for decay)

- Normalization
  - 5%, varied functional forms and mass-constrained fits

---

Urs Langenegger

First Observation of $B_s^0 \to \mu^+ \mu^-$ with CMS (2014/02/17)
Outlook and Conclusions
CMS in Run 2

- ‘Simple’ scaling of current analysis
  - officially; reality probably better
  - 2015-2017: ca 100 fb$^{-1}$ at $\approx 13$ TeV
    i.e. $\approx 6 - 8$ times more $B^0_s$ mesons
- $B^0_s \rightarrow \mu^+\mu^-$
  - effective lifetime
- $B^0 \rightarrow \mu^+\mu^-$
  - excess in CMS (and LHCb)?
  - requires extended/improved analysis
    - fake rate
    - peaking background
    - rare sl background

<table>
<thead>
<tr>
<th>L (fb$^{-1}$)</th>
<th>No. of $B^0_s$</th>
<th>No. of $B^0$</th>
<th>$\delta B/B(B^0_s \rightarrow \mu^+\mu^-)$</th>
<th>$\delta B/B(B^0 \rightarrow \mu^+\mu^-)$</th>
<th>$B^0$ sign.</th>
<th>$\frac{\delta B(B^0 \rightarrow \mu^+\mu^-)}{B(B^0 \rightarrow \mu^+\mu^-)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>16.5</td>
<td>2.0</td>
<td>35%</td>
<td>$&gt;100%$</td>
<td>0.0–1.5 $\sigma$</td>
<td>$&gt;100%$</td>
</tr>
<tr>
<td>100</td>
<td>144</td>
<td>18</td>
<td>15%</td>
<td>66%</td>
<td>0.5–2.4 $\sigma$</td>
<td>71%</td>
</tr>
<tr>
<td>300</td>
<td>433</td>
<td>54</td>
<td>12%</td>
<td>45%</td>
<td>1.3–3.3 $\sigma$</td>
<td>47%</td>
</tr>
<tr>
<td>3000</td>
<td>2096</td>
<td>256</td>
<td>12%</td>
<td>18%</td>
<td>5.4–7.6 $\sigma$</td>
<td>21%</td>
</tr>
</tbody>
</table>

CMS Simulation - Scaled to L = 300 fb$^{-1}$

- data
- full PDF
- $B_{s} \rightarrow \mu^+\mu^-$
- $B_{s} \rightarrow \mu^+\mu^-$
- combinatorial bkg
- semileptonic bkg
- peaking bkg

Urs Langenegger

First Observation of $B^0_s \rightarrow \mu^+\mu^-$ with CMS (2014/02/17)
Conclusions

- First observation of $\mathcal{B}(B^0_s \rightarrow \mu^+\mu^-)$ at $\geq 4$ standard deviations
  - substantial improvements to previous analysis
    - muon identification with BDT
    - analysis selection with BDT
    - UML fit to mass distributions
  - consistent with SM
  - and upper limit on $\mathcal{B}(B^0 \rightarrow \mu^+\mu^-)$

- Exciting future
  - Run 2
    - up to 100 fb$^{-1}$
      - large increase of $B$ sample!
      - no trigger problems
    - $B^0_s \rightarrow \mu^+\mu^-$
      - differential measurements
    - $B^0 \rightarrow \mu^+\mu^-$
      - increase sensitivity to reach evidence for SM signal?
Backup
LHCb’s results (II)

- $\mathcal{B}(B^0_s \rightarrow \mu^+ \mu^-) = (2.9^{+1.1}_{-1.0}) \times 10^{-9}$
  ▶ significance of $4.0\sigma$ (rounded up)

- Upper limit on $\mathcal{B}(B^0_s \rightarrow \mu^+ \mu^-)$
  (1- and 2-sided) 95%CL $[10^{-9}]$

<table>
<thead>
<tr>
<th>$\mathcal{B}$</th>
<th>Ref.</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4.5$</td>
<td>PRL, 108, 231801</td>
<td>$1 \text{ fb}^{-1}$, $&gt;1\sigma$ low</td>
</tr>
<tr>
<td>$6.4$</td>
<td>PRL, 110, 021801</td>
<td>$2 \text{ fb}^{-1}$, excess in 2012</td>
</tr>
<tr>
<td>$5.1$</td>
<td>PRL, 111, 101805</td>
<td>$3 \text{ fb}^{-1}$, $4.0(5.0)\sigma$ obs (exp)</td>
</tr>
</tbody>
</table>

PRL, 111, 101805
LHCb-PAPER-2013-046 suppl. material

Urs Langenegger
First Observation of $B^0_s \rightarrow \mu^+ \mu^-$ with CMS (2014/02/17)
$y_s$ correction to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$

- **LHCb measured significant $B_s^0$ decay width difference**
  \[ y_s = (\Gamma_L^s - \Gamma_H^s)/(\Gamma_L^s + \Gamma_H^s) = 0.088 \pm 0.014 \]

- **Untagged $B_s^0$ decay distribution is sum of two contributions**
  \[ \langle \Gamma(B_s^0(t) \rightarrow f) \rangle = \Gamma(B_s^0(t) \rightarrow f) + \Gamma(\bar{B}_s^0(t) \rightarrow f) \]
  \[ = R_H^f e^{-\Gamma_H^s t} + R_L^f e^{-\Gamma_L^s t} \]

  experiments measure (in untagged events)

  \[ \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} = \frac{1}{2} \int_0^\infty \langle \Gamma(B_s^0(t) \rightarrow f) \rangle dt \]
  \[ = \frac{1}{2} \left[ \frac{R_H^f}{\Gamma_H^s} - \frac{R_L^f}{\Gamma_L^s} \right] \]

  (in SM: $A_{\Delta \Gamma}^{\mu^+\mu^-} = 1$)

  \[ \rightarrow 10\% \text{ correction(!)} \]

  (with ‘effective’ lifetime $\tau_f$)

  \[ = \left[ 2 - (1 - y_s^2) \frac{\tau_f}{\tau_{B_s^0}} \right]^{-1} \times \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{theo}} \]
$B^0 \rightarrow K^{*0} \mu^+ \mu^-$ angular analysis

- Recently published
  - 5 fb$^{-1}$ at $\sqrt{s} = 7$ TeV
  - good statistics at high-$Q^2$
  - 23 - 106 signal events/$Q^2$-bin

- Selection w/o PID
  - C-n-C optimized for $S/\sqrt{S+B}$
  - best $B$-vertex fit $\chi^2$
  - $B^0$ if $K^+\pi^-$ mass closest to $K^{*0}$ mass
  - reject event if both $K^+\pi^-$ and $K^-\pi^+$ within 50 MeV of $K^{*0}$ (8% wrong choice)