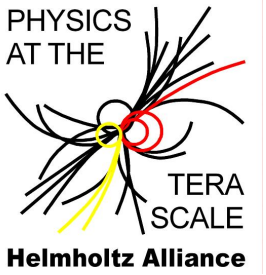
Rare Decays of $B^0_{(s)}$ Mesons to Muon Pairs with the ATLAS Detector (Run 1)

Wolfgang Walkowiak
University of Siegen

on behalf of the ATLAS collaboration



Rencontres de Blois 2016

May 29th - June 3rd, 2016

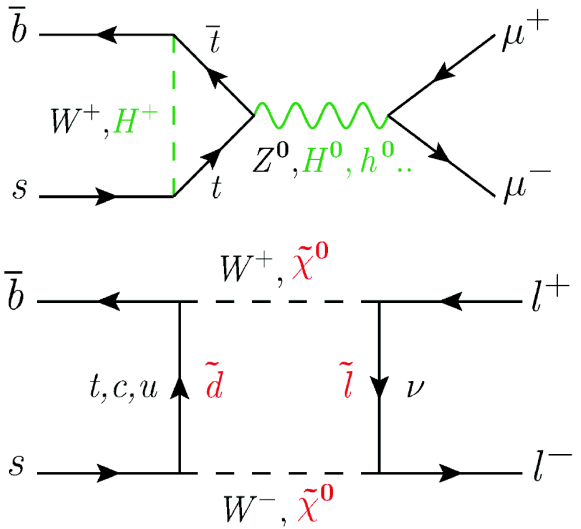
New Physics in Rare B Decays

Study Flavor Changing Neutral Currents (FCNC)

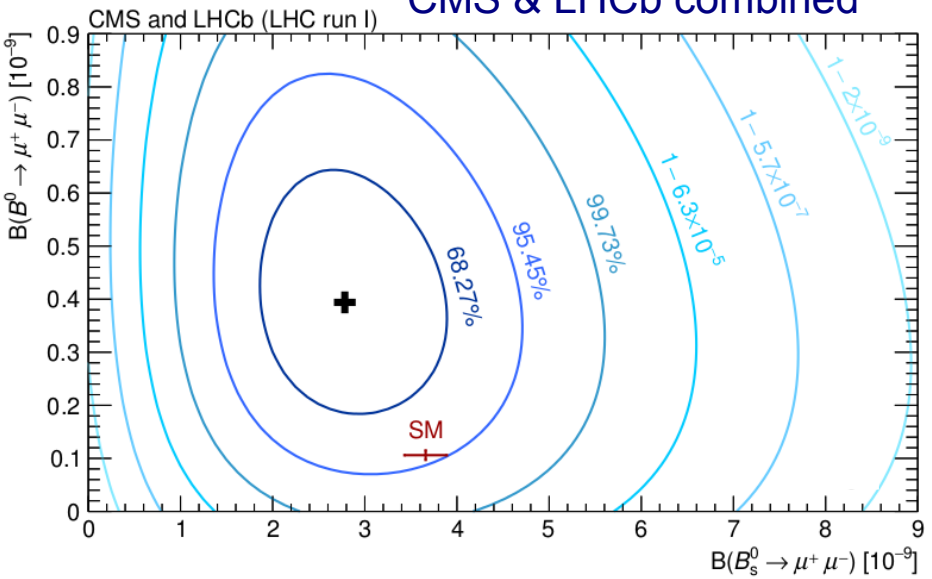
- Rare decays $B^0_{(s)} \rightarrow \mu^+\mu^-$
  - Highly suppressed in SM
  - Non-SM particles → modify BR
  - Powerful indirect search for NP
- BR expectations (SM)
  - $BR(B^0_s \rightarrow \mu^+\mu^-) = (3.65 \pm 0.23)\cdot 10^{-9}$
  - $BR(B^0 \rightarrow \mu^+\mu^-) = (1.06 \pm 0.09)\cdot 10^{-10}$

    [C. Bobeth et al., PRL 112 (2014) 101801]
- CMS & LHCb combined (Run 1):
  - $BR(B^0_s \rightarrow \mu^+\mu^-) = (2.8\ ^{+0.7}_{-0.6})\cdot 10^{-9}$
  - $BR(B^0 \rightarrow \mu^+\mu^-) = (3.9\ ^{+1.6}_{-1.4})\cdot 10^{-10}$

    [Nature 522 (2015) 68-72]



CMS & LHCb combined



$B^0_{(s)} \to \mu^+ \mu^-$ Analysis Strategy: Overview

- ATLAS data at E_{CMS} = 7 and 8 TeV (4.9 + 20.0 fb^{-1})
- Di-μ triggers:
  - 2011: Two μ with $p_T > 4$ GeV
  - 2012: Three trigger categories
    (μ with $p_T > 4$ and 6 GeV, one barrel μ in $|\eta| < 1.05$)
- μ tracks reconstructed in inner detector and muon spectrometer
- For signal $m_{\mu\mu}$: [5166 MeV, 5526 MeV] blinded
- $p_T(B) > 8$ GeV, $|\eta(B)| < 2.5$
- Background reduction with MVA classifiers:
  - Continuum-BDT against combinatorial background
  - Fake-BDT against hadrons mis-id'd as μ's
- Signal extraction: ML fit to $m_{\mu\mu}$ in three intervals of continuum-BDT
- Normalization with $B^\pm \to J/\psi\, K^\pm$
  - Yield, fragmentation and efficiency ratios

Analysis Strategy: Normalization Channel (1)

Normalization with $B^{\pm} \rightarrow J/\psi\, K^{\pm}$:

$$BR\left(B^0_{(s)} \rightarrow \mu^+\mu^-\right) = N_{d(s)} \cdot \frac{1}{N_{J/\psi K^+}} \cdot \frac{\epsilon_{J/\psi K^+}}{\epsilon_{\mu\mu}} \cdot \frac{f_u}{f_{d(s)}}$$

$$\cdot BR\left(B^+ \rightarrow J/\psi K^+\right) \cdot BR\left(J/\psi \rightarrow \mu^+\mu^-\right)$$

[PDG 2014]

- $B^0_s \rightarrow \mu^+\mu^-$ signal $N_{d(s)}$
- Efficiency corrections $\epsilon_{J/\psi K^+}/\epsilon_{\mu\mu}$
- Hadronization probabilities $B^{\pm}$ vs. $B^0_{(s)}$: $f_u/f_{d(s)}$

- Three trigger categories (2012 data) + 2011 data
  → Four trigger/data categories

Analysis Strategy: Normalization Channel (2)

Modified for trigger/data categories (k):

$$BR\left(B^0_{(s)} \to \mu^+\mu^-\right) = N_{d(s)} \cdot \frac{1}{D_{norm}} \cdot \frac{f_u}{f_{d(s)}}$$

$$\cdot BR\left(B^+ \to J/\psi K^+\right) \cdot BR\left(J/\psi \to \mu^+\mu^-\right)$$

[PDG 2014]

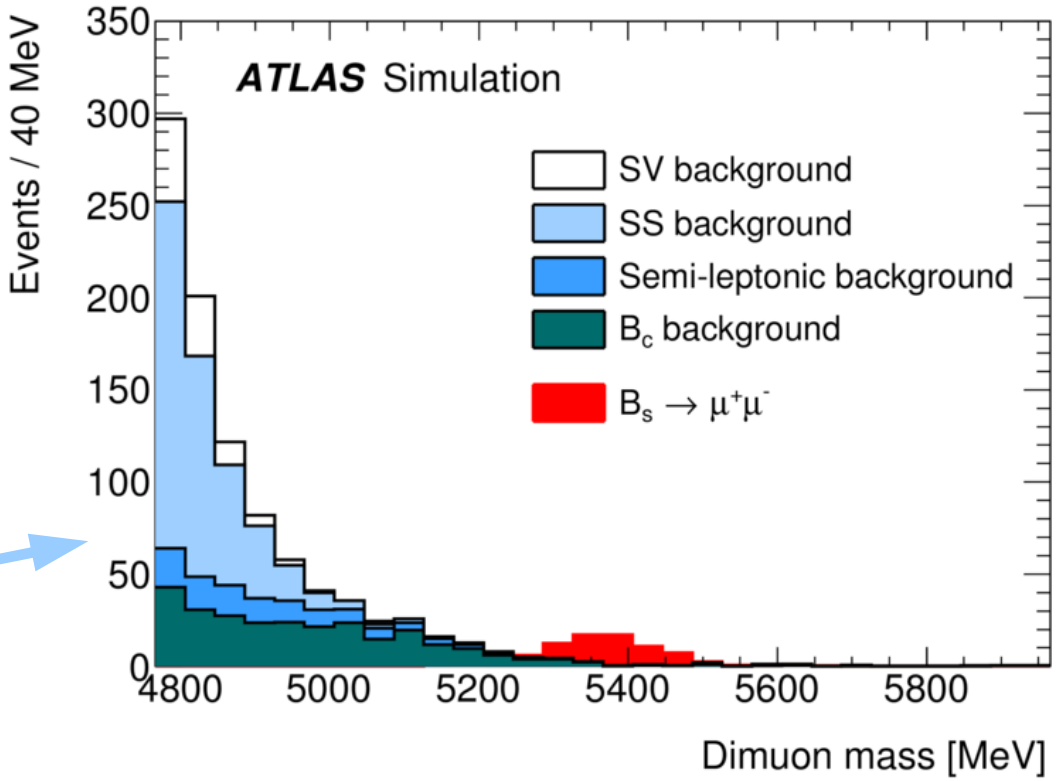
with

$$D_{norm} = \sum_k N^k_{J/\psi K^+} \alpha_k \left\{ \frac{\epsilon_{\mu\mu}}{\epsilon_{J/\psi K^+}} \right\}_k$$

- Normalization $N^k_{J/\psi K^+}$
- Efficiency corrections $\{\epsilon_{J/\psi K^+}/\epsilon_{\mu\mu}\}_k$
- Prescaling factors α_k

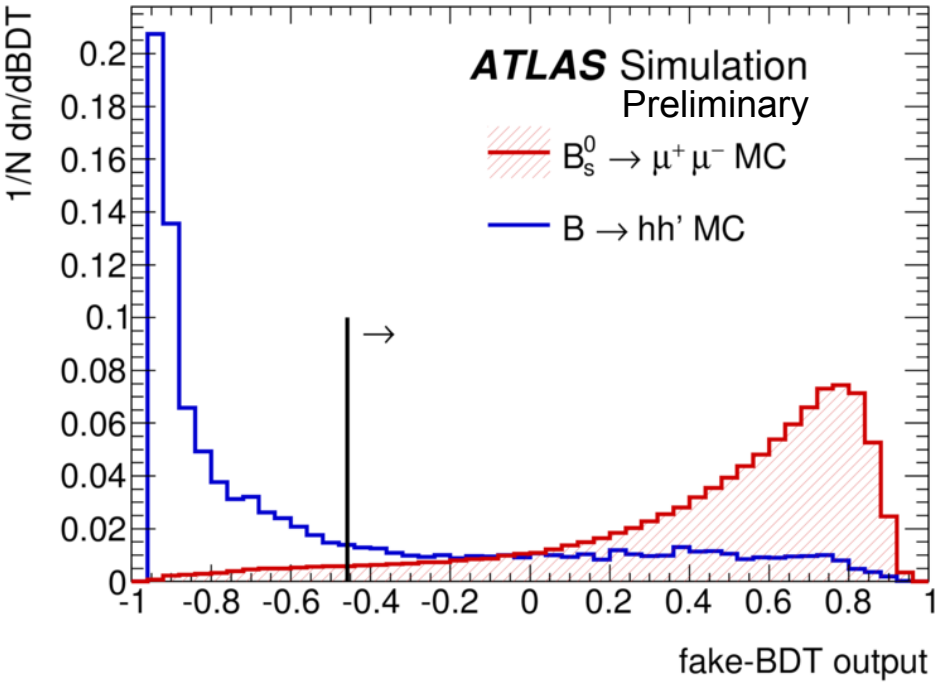
Background Contributions

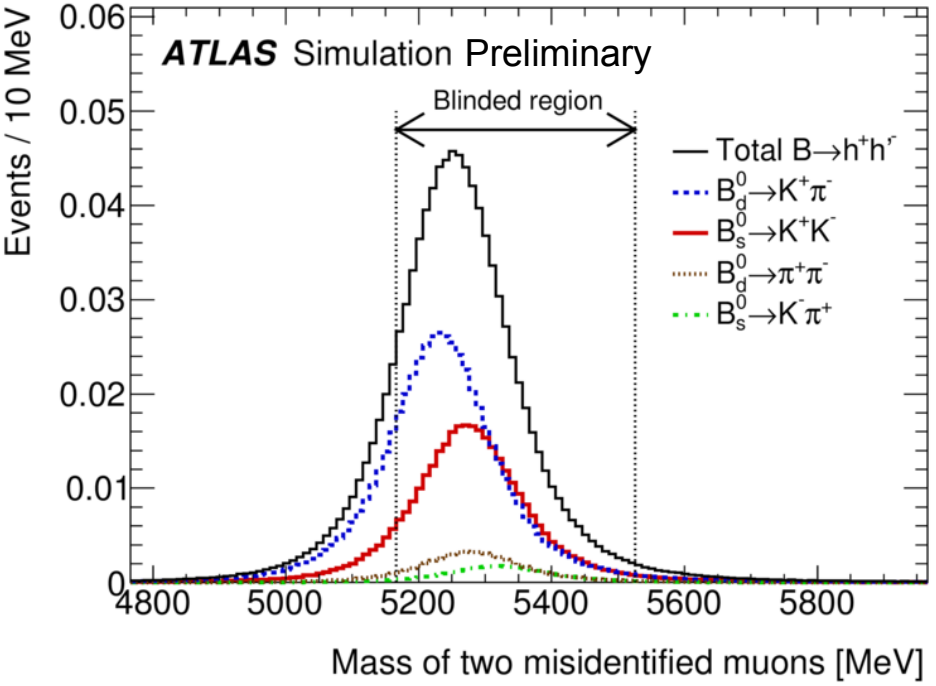
- Combinatorial background (opposite-side μ's):
  - Dominant
  - Smoothly distributed in $m_{\mu\mu}$
- Partially reco'd B decays:
  - Same Vertex (SV):
    $B \to \mu\mu X$ (e.g. $B \to K^* \mu\mu$)
  - Same Side (SS)
    semileptonic decay cascades
    ($b \to c\mu\nu \to s(d)\mu\mu\nu\nu$)
  - B_c decays (e.g. $B_c \to J/\psi\, \mu\nu$)
  - → all real μ's at low $m_{\mu\mu}$
- Semileptonic $B_{(s)}$ decays:
  - Real μ + charged hadron
- Peaking background of charmless hadronic $B^0_{(s)}$ decays:
  - $B^0_{(s)} \to hh'$ ($h = K^{\pm}$ or $\pi^{\pm}$)
  - → Small, but in $m_{\mu\mu}$-signal region!



Hadron Misidentification: Fake-BDT

- Studied using simulated $B \to hh'$, $\Lambda_b \to ph$
- Validated on data ($\phi \to K^+ K^-$, $B^\pm \to J/\psi\, K^\pm$)
- Low $P^{mis}(\mu^\pm)$ = 0.4% ($K^\pm$), 0.2% ($\pi^\pm$), < 0.01% (protons)
- Reduced by x 0.4 by fake-BDT with $\varepsilon(\mu^\pm)$ = 95%.





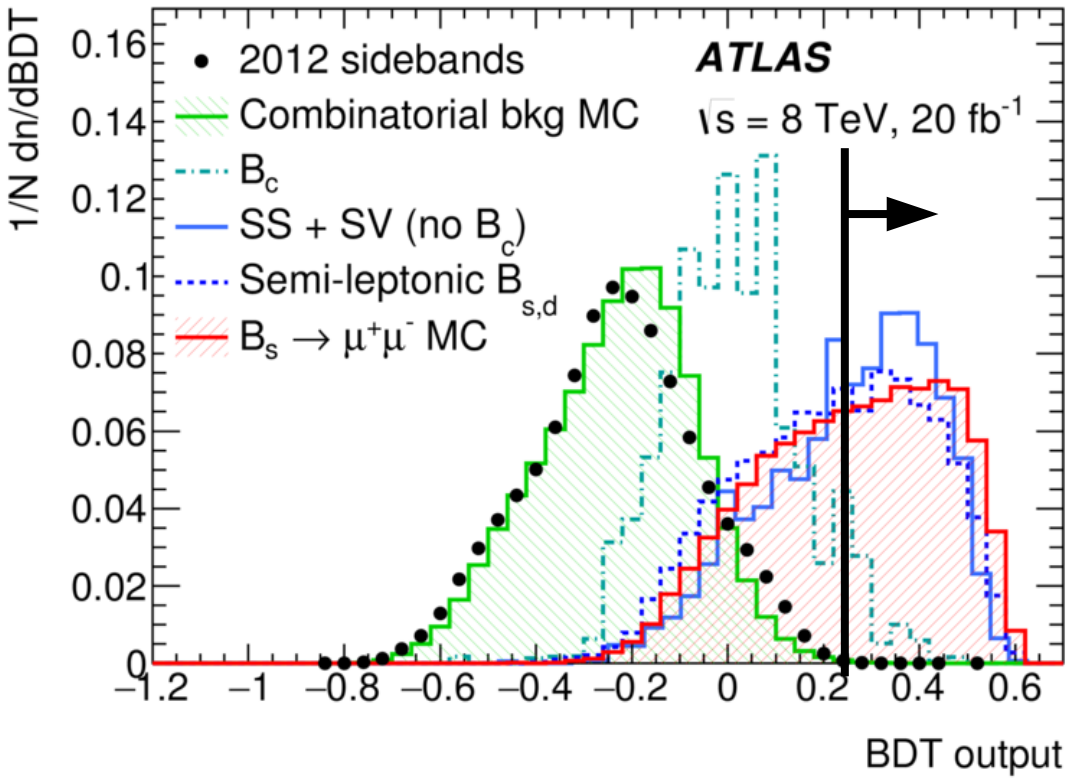
Normalized using $B^\pm \to J/\psi\, K^\pm$ yield and efficiency ratio:

- $N_{peak\text{-}bkg}$ = 1.0 ± 0.4 events

Combinatorial Background: Continuum-BDT

Pairs of $\mu^{\pm}$ from uncorrelated hadron decays (from b and c & c.c.):

- Separated by BDT classifier (15 variables) on properties of
  - B candidate
  - μ's from B candidate
  - other tracks in event
  - pile-up vertices



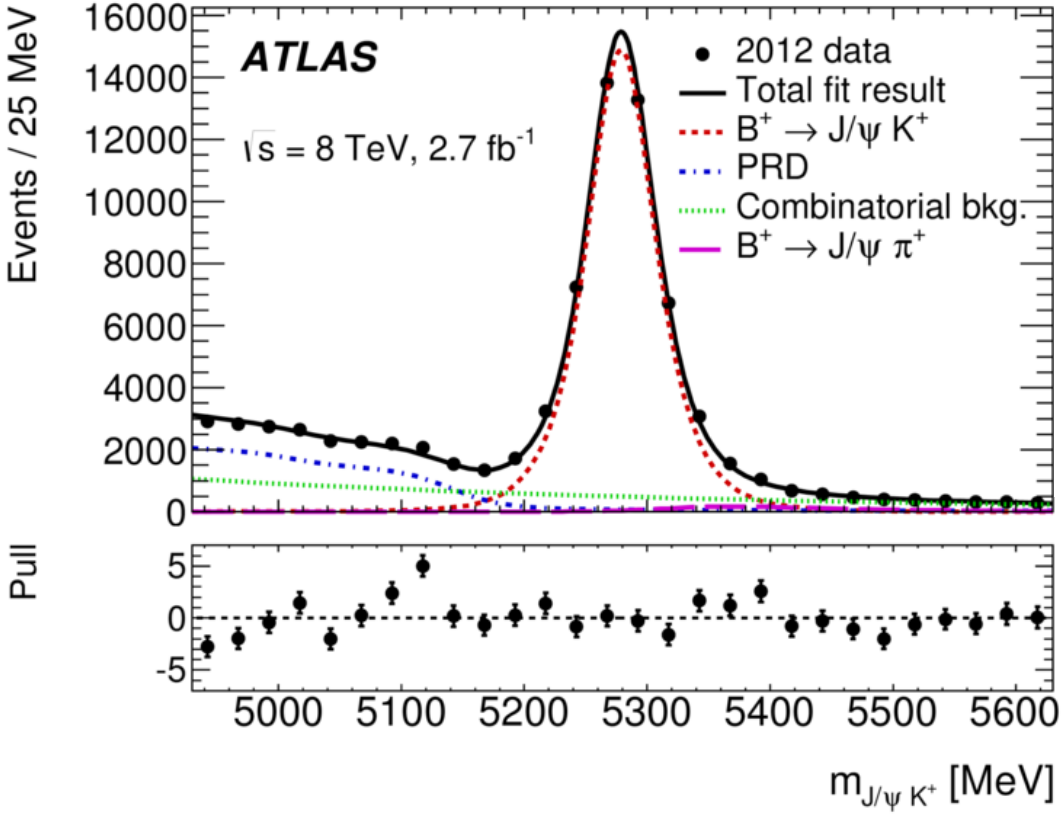
- Trained continuum-BDT on large MC sample (1.4 G) of uncorrelated b-(c-) and $\overline{b}$-($\overline{c}$-) hadrons, decays forced into μ final states
  - Validated on sideband data
- Bkg suppression ~ 1000
- $\varepsilon(B_s^0 \to \mu^+\mu^-) = 54\%$

- B-related backgrounds (factor ~ 2000 less)
  - SS+SV
  - Semileptonic decays (~ 30% of SV)
  - peaking background

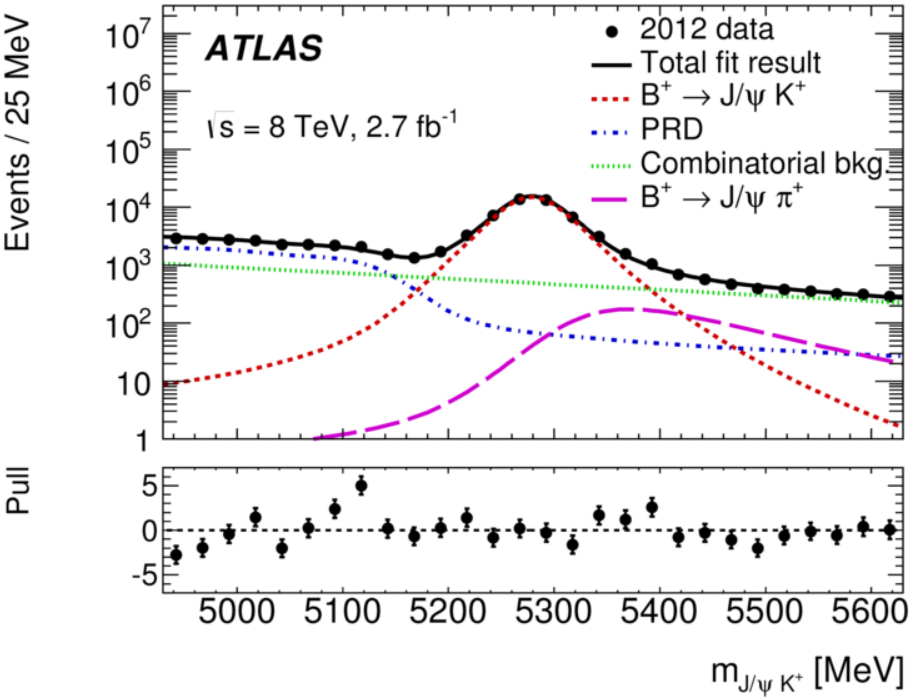
$B^{\pm}$ Yield Extraction

- Applied fake-BDT and continuum-BDT selections
- Separately in 4 data/trigger categories
- Unbinned ML fit of $m_{\mu\mu K}$ distribution

Main 2012 trigger category



Logarithmic scale



Efficiency Ratio $\varepsilon_{\mu\mu}/\varepsilon_{J/\psi K}$

$$D_{norm} = \sum_k N^k_{J/\psi K^+} \alpha_k \left\{ \frac{\epsilon_{\mu\mu}}{\epsilon_{J/\psi K^+}} \right\}_k$$

Per category (k): $\{\varepsilon_{J/\psi K^+}/\varepsilon_{\mu\mu}\}_k$ from MC:

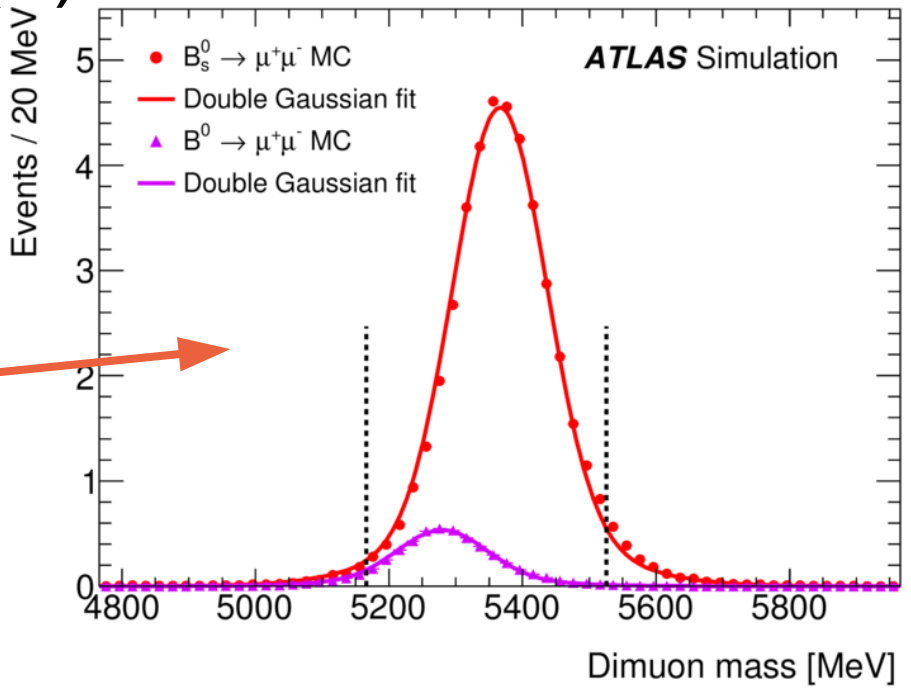
- p_T and η spectra tuned on reference channel data
- Systematic uncertainties:
  - Finite MC statistics for p_T- & η-reweighting
  - Data-MC discrepancies (±4.2%, dominant) in 15 variables of continuum-BDT
  - Isolation ($I_{0.7}$) tuned for $B^{\pm}$ and $B^0_{(s)}$ modes (2012 only)
    → corrections
- For B^0_s:
  - Lifetime difference $\Delta\tau$ between B^0_s mass eigenstates
    → correction (+4%) from MC (τ from SM prediction)
- Total corrections to $\varepsilon_{J/\psi K^+}/\varepsilon_{\mu\mu}$: +3% for B^0 and -1% for B^0_s
- Total ΔD_{norm}^{sys} = ±5.9%

Signal Yield Extraction (1)

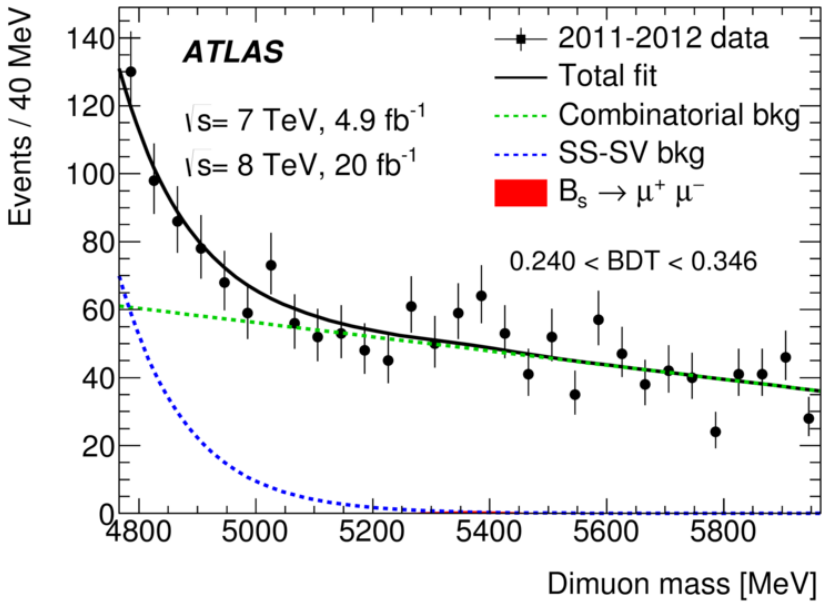
N_d and N_s extracted by ML fit in $m_{\mu\mu}$

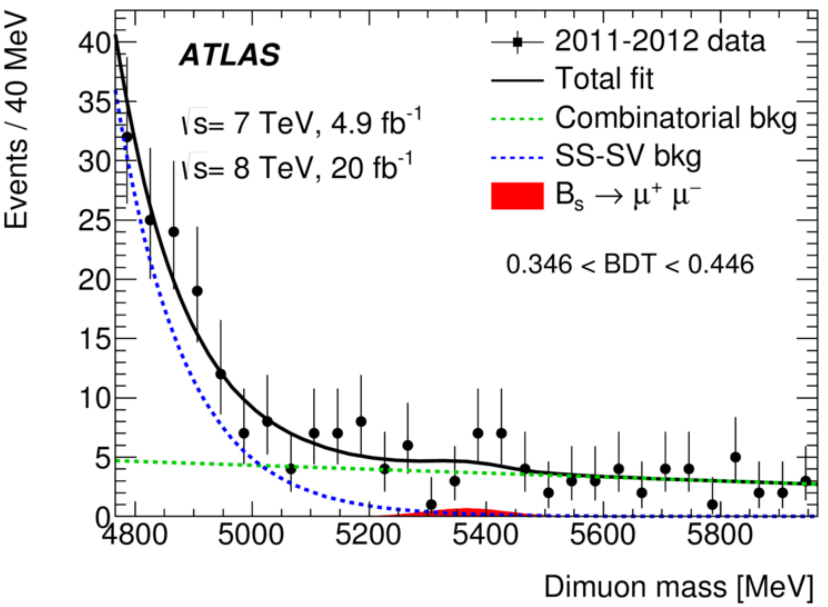
- In three intervals in continuum-BDT output (each ε_{sig} = 18%)
- Signal and background model:
  - Signal: two Gaussians, avg. width 80 MeV, independent of continuum-BDT
  - Low mass (SV+SS) background: exponential in $m_{\mu\mu}$, determined from sideband data, independent of continuum-BDT
  - Peaking background: two Gaussians, 1.0 ± 0.4 events total, equal amplitude in continuum-BDT bins
  - Continuum background: linear in $m_{\mu\mu}$, minimal correlation between $m_{\mu\mu}$ and continuum-BDT, sideband data consistent with MC
  - Systematics of fit:

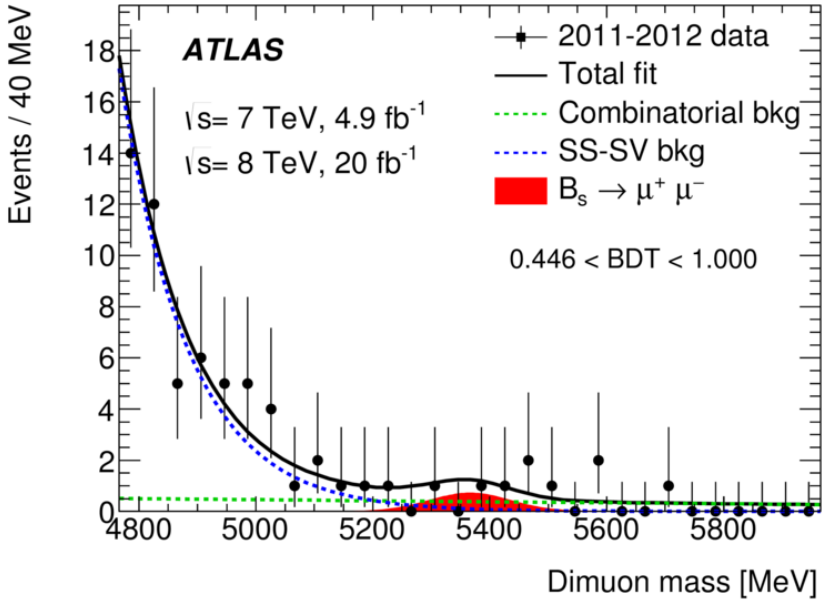
$$\sigma_{syst}(N_s)=\sqrt{2^2+(0.06\times N_s)^2} \quad \text{and} \quad \sigma_{syst}(N_d)=3 \text{ events}$$



Signal Yield Extraction (2)







- Yields constrained > 0:
  - N_s = 11 and N_d = 0
- Unconstrained yields:
  - N_s = 16 ± 12 and N_d = -11 ± 9
- Expected from SM predictions:
  - N_s = 41 and N_d = 5
- Fewer B^0_s, no B^0 events!

Branching Fraction Extraction

$$BR\left(B^0_{(s)} \to \mu^+\mu^-\right) = N_{d(s)} \cdot \frac{1}{D_{norm}} \cdot \frac{f_u}{f_{d(s)}}$$
$$\cdot BR\left(B^+ \to J/\psi K^+\right) \cdot BR\left(J/\psi \to \mu^+\mu^- K^+\right)$$

$$D_{norm} = \sum_k N^k_{J/\psi K^+} \alpha_k \left\{ \frac{\epsilon_{\mu\mu}}{\epsilon_{J/\psi K^+}} \right\}_k$$

Ingredients to normalization:

- $B^{\pm}$ and J/ψ branching fractions [PDG 2014]
- f_u/f_s from ATLAS $f_s/f_d = 0.240 \pm 0.020$ [PRL 115 (2015) 262001]
  assuming $f_u/f_d = 1$ (same p_T & η range!)
- $\{\varepsilon_{J/\psi K^+}/\varepsilon_{\mu\mu}\}_k$ and $N^k_{J/\psi K^+}$ in D_{norm}

- Total normalization uncertainty:
  - ± 11% for BR($B^0_s \to \mu^+\mu^-$)
  - ± 7% for BR($B^0 \to \mu^+\mu^-$)

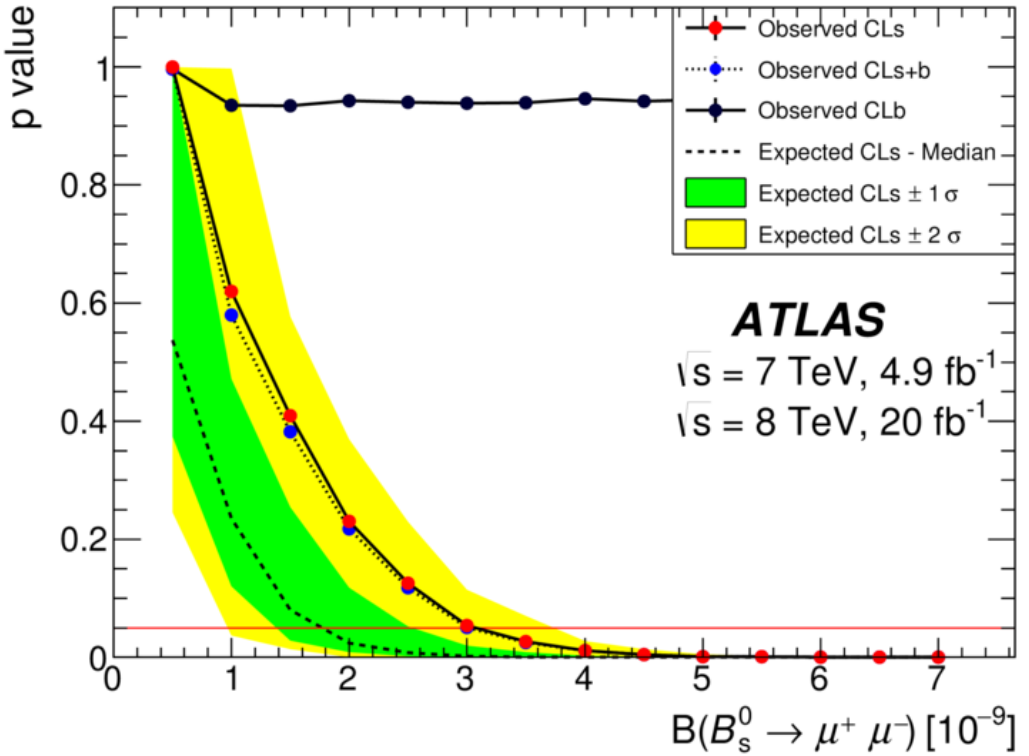
$B^0_s \to \mu^+\mu^-$ Branching Fraction Result

$BR(B^0_s \to \mu^+\mu^-) = 0.9\ ^{+1.1}_{-0.8} \times 10^{-9}$

- Obtained within boundary of non-negative branching fractions

- Upper limit (CLs method):
  $BR(B^0_s \to \mu^+\mu^-) < 3.0 \times 10^{-9}$
  at 95% CL
- Expected upper limit:
  $BR(B^0_s \to \mu^+\mu^-) < 1.8\ ^{+0.7}_{-0.4} \times 10^{-9}$

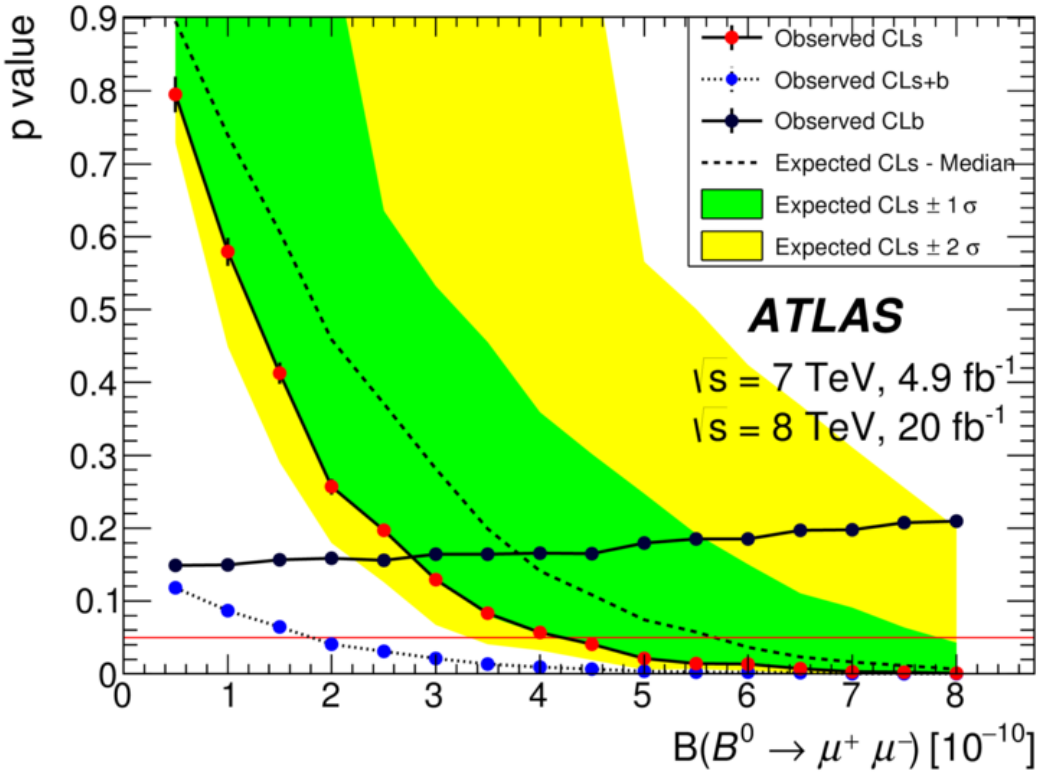
- Compatibility with null hypothesis:
  $p = 0.08\ (1.4\ \sigma)$



- Expected significance for SM $BR(B^0_s \to \mu^+\mu^-)$: 3.1 σ

$B^0 \rightarrow \mu^+\mu^-$ Branching Fraction Result

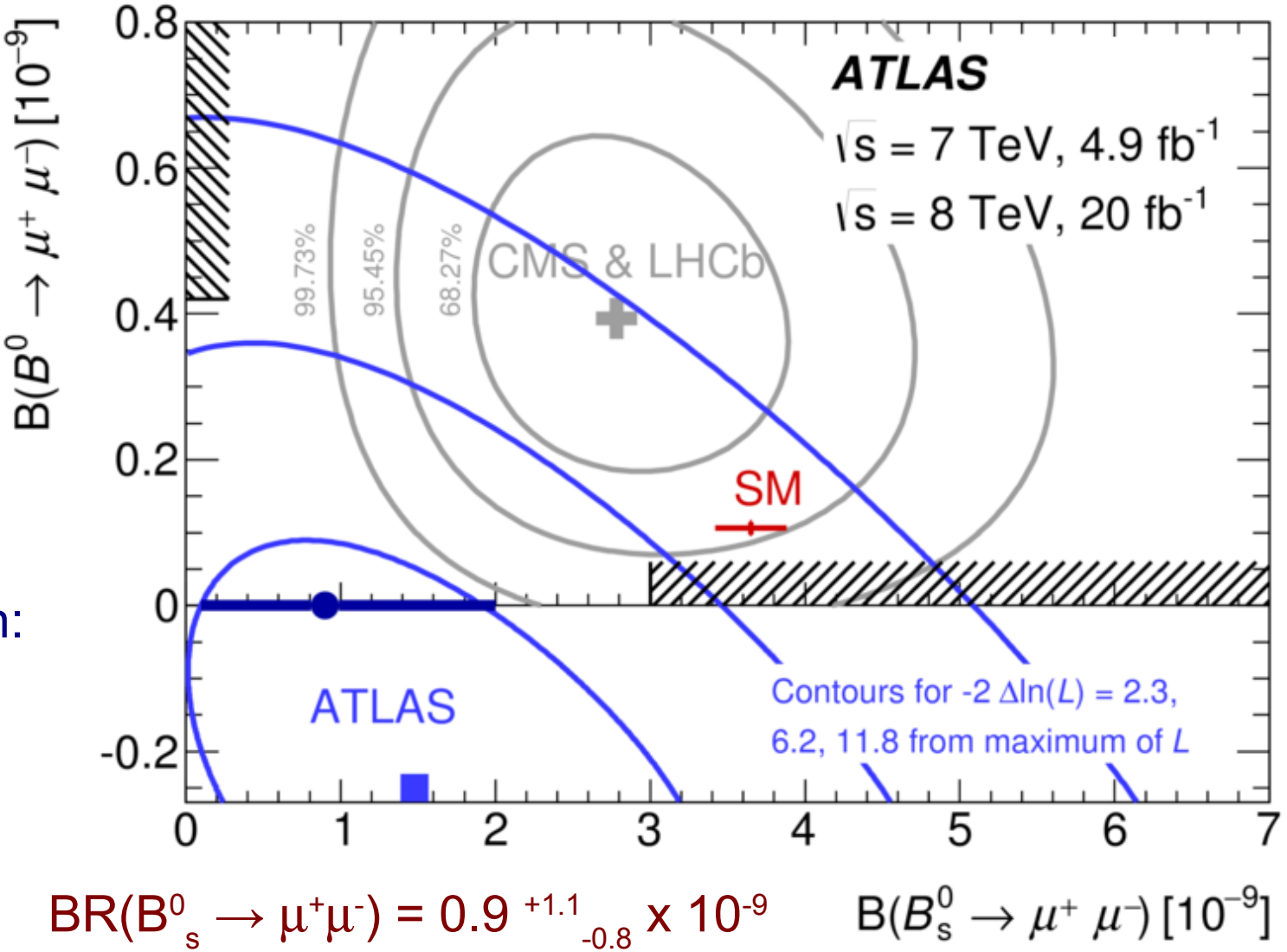
- Upper limit (CLs method):
  $BR(B^0 \rightarrow \mu^+\mu^-) < 4.2 \times 10^{-10}$ at 95% CL
- Expected upper limit:
  $BR(B^0 \rightarrow \mu^+\mu^-) < 5.7\ ^{+2.1}_{-1.2} \times 10^{-10}$
- CLb ≈ 0.15 for $BR(B^0 \rightarrow \mu^+\mu^-)$ near 0:
  → -1 σ fluctuation of background
- Expected significance for SM:
  0.2 σ



Summary of Results

$BR(B^0 \rightarrow \mu^+\mu^-) < 4.2 \times 10^{-10}$ at 95% CL

CMS & LHCb contours from [Nature 522 (2015) 68-72]



Compatibility of simultaneous fit with SM prediction:

P = 0.048 (2.0 σ)

$BR(B^0_s \rightarrow \mu^+\mu^-) = 0.9\ ^{+1.1}_{-0.8} \times 10^{-9}$

[ATLAS: arXiv:1604.04263 [hep-ex]]

Conclusions

Submitted to Eur. Phys. J. C
arXiv:1604.04263 [hep-ex]

Results for $B^0_{(s)} \to \mu^+\mu^-$ with ATLAS Run-1 data:

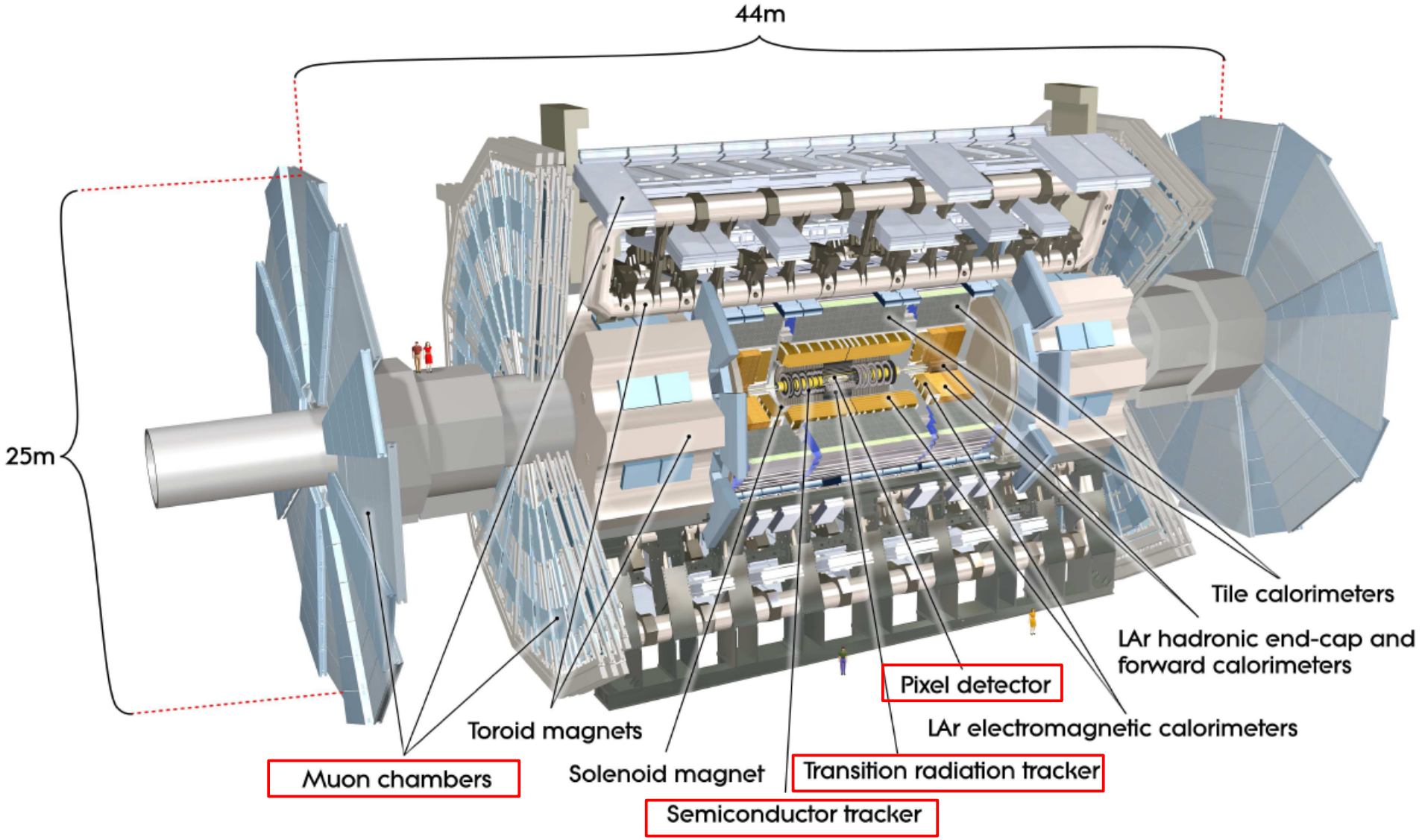
- $BR(B^0_s \to \mu^+\mu^-) = 0.9^{+1.1}_{-0.8} \times 10^{-9}$
  $< 3.0 \times 10^{-9}$ at 95% CL
  - Lower than SM prediction
  - Lower than central value of CMS & LHCb combination
    $[BR(B^0_s \to \mu^+\mu^-)_{CMS\&LHCb} = 2.8^{+0.7}_{-0.6} \times 10^{-9}]$
  - Closer to CMS & LHCb than to SM prediction

- $BR(B^0 \to \mu^+\mu^-) < 4.2 \times 10^{-10}$ at 95% CL
  - Includes SM prediction
  - Reaches central value of CMS & LHCb combination
    $[BR(B^0 \to \mu^+\mu^-)_{CMS\&LHCb} = 3.9^{+1.6}_{-1.4} \times 10^{-10}]$
  - Closer to SM prediction than to CMS & LHCb

- Compatibility for the simultaneous fit with SM is 2.0 σ.
  $\to$ Looking forward to more data from LHC run 2!

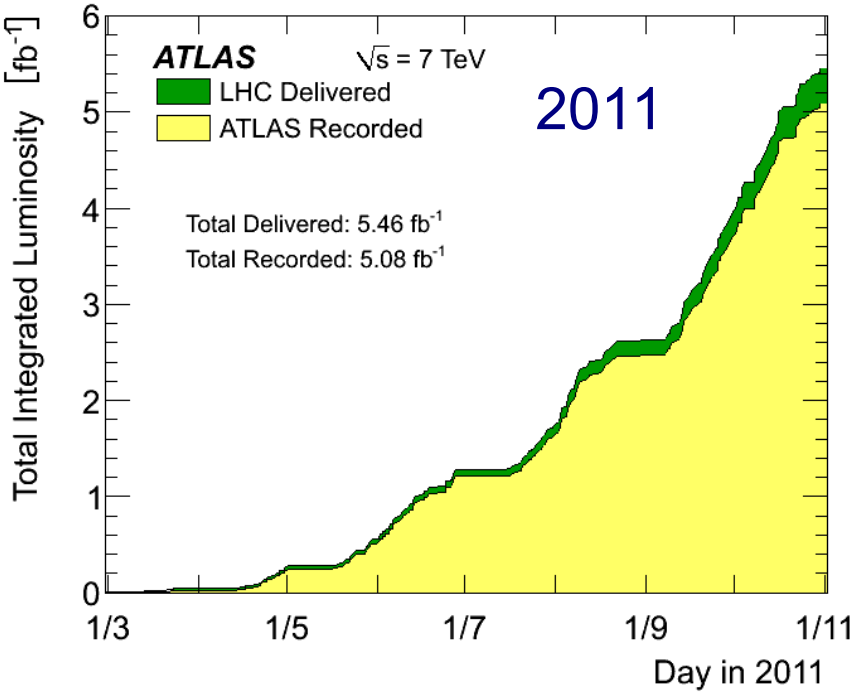
Supporting Material

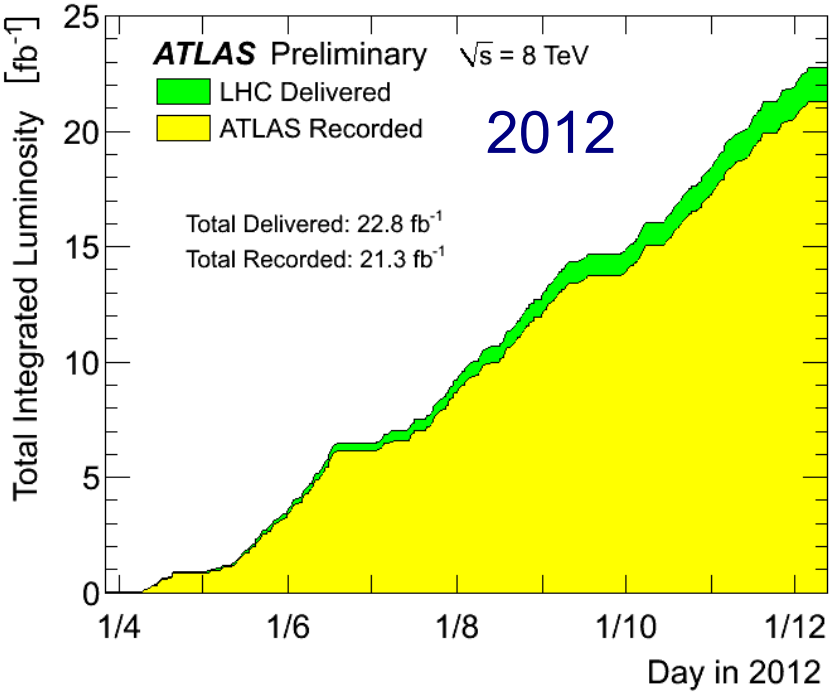
A Toroidal LHC ApparatuS (ATLAS)



Inner detector and muon chambers most important for analysis presented

ATLAS Run-1 Data (2011/2012)





Data taken in 2011:

- E_{CM} = 7 TeV
- $\int L\,dt$ = 5.08 fb^{-1} recorded
- $\langle\mu\rangle$ = 9.1

$B^0_{(s)} \rightarrow \mu^+\mu^-$ analysis:

- $\int L\,dt$ = 4.9 fb^{-1} used

Data taken in 2012:

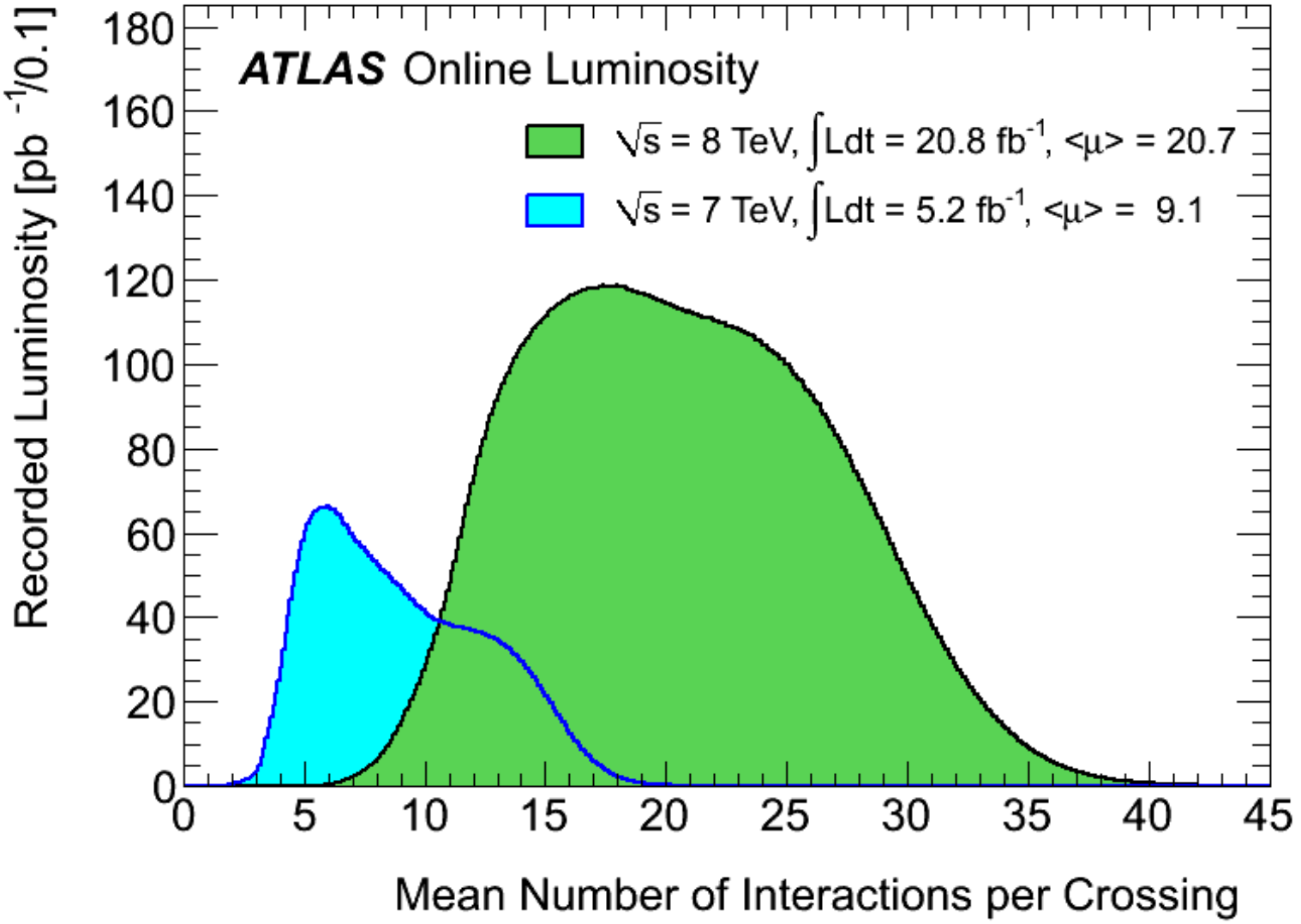
- E_{CM} = 8 TeV
- $\int L\,dt$ = 21.3 fb^{-1} recorded
- $\langle\mu\rangle$ = 20.7

$B^0_{(s)} \rightarrow \mu^+\mu^-$ analysis:

- $\int L\,dt$ = 20 fb^{-1} used

ATLAS Run-1 – Mean Number Of Interactions

- 2011 data: <μ> = 9.1, 2012 data: <μ> = 20.7



Di-μ Triggers for Low p_T Di-μ Events

Full $m_{\mu\mu}$ range:

- Dimu 1.5 – 14.0 GeV

$J/\psi \rightarrow \mu^+\mu^-$:

- Jpsimumu 2.5 – 4.3 GeV

Intermediate $m_{\mu\mu}$ range:

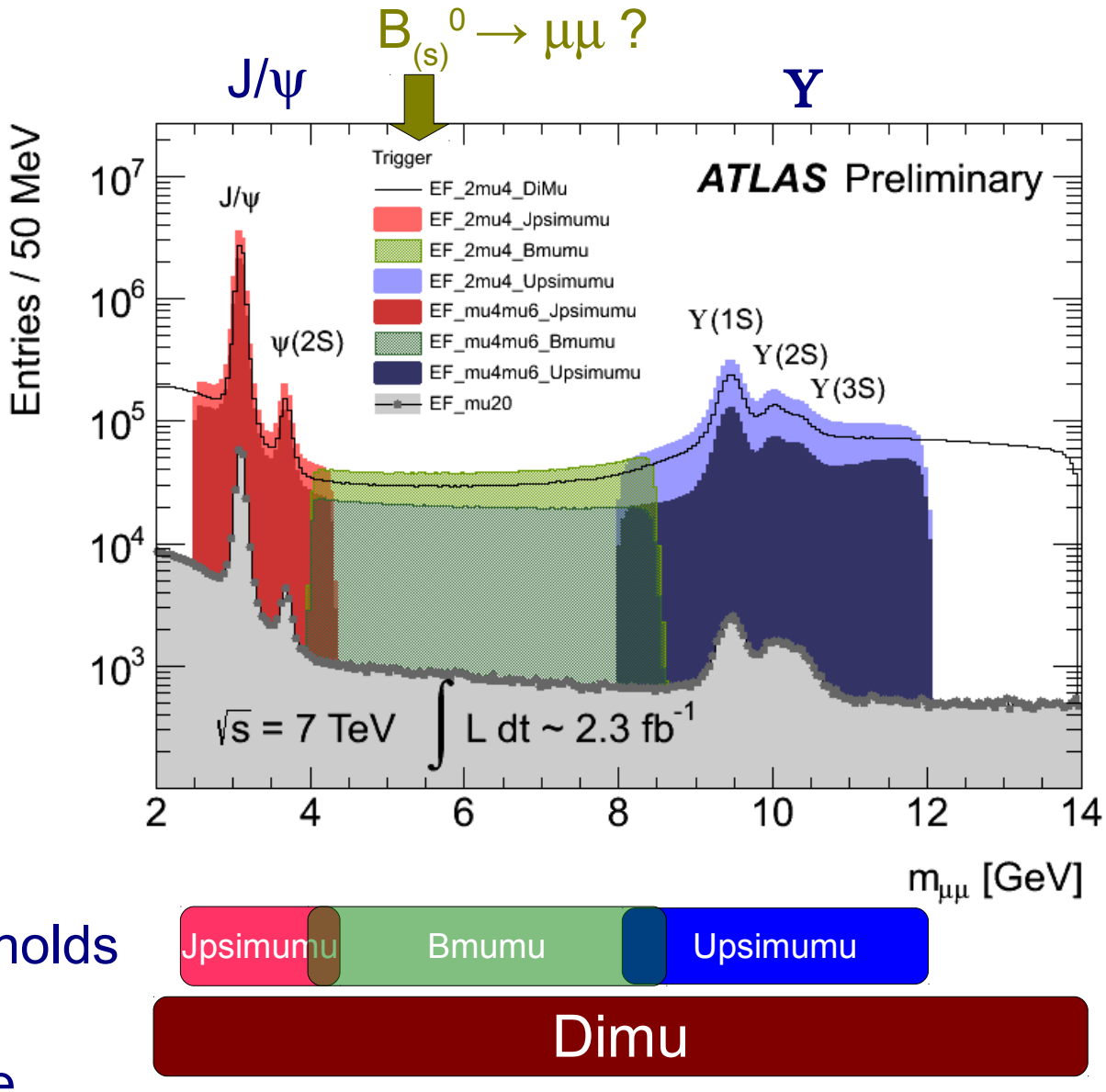
- Bmumu 4.0 – 8.5 GeV

$\Upsilon \rightarrow \mu^+\mu^-$:

- Upsimumu 8.0 – 12.0 GeV

Adjust trigger rates by

- Increasing μ p_T trigger thresholds
- Adding prescale factors
- Lifetime cuts at HLT possible



Di-μ Triggers for 2011 and 2012 Data

Exclusive trigger/data categories:

- 2011: di-μ triggers with p_T > 4 GeV (22%)

- 2012: three mutually exclusive categories: (78%)
  - T1: 1 μ p_T > 6 GeV & 1 μ p_T > 4 GeV (68%)
  - T2: both μ p_T > 4 GeV, at least 1 μ $|\eta|$ < 1.05 & not in T1 (6%)
  - T3: both μ p_T > 4 GeV & not in T1 or T2 (4%)

Preselection Cuts & Trigger Selection

General
- $p_{T,\mu} > 4$ GeV, $|\eta_\mu| < 2.5$ (both μ "combined")
- $p_{T,K} > 1.0$ GeV, $|\eta_K| < 2.5$ ($B^\pm$)
- Tracks: # pixel hits >0, # SCT hits > 4, #TRT hits + #TRT outliers > 5 and #TRT outliers < 0.9 (#TRT hits + #TRT outliers) for $0.1 < |\eta| < 1.9$
  - μ tracks: good muon track quality requirements

J/ψ specific ($B^\pm$):
- $2.915 < m_{J/\psi} < 3.275$ GeV
- J/ψ vertex $\chi^2/\text{ndf} < 10$

B_s ($B^\pm$) specific:
- B_s and $B^\pm$ vertex $\chi^2/\text{ndf} < 6$
- $p_T > 8$ GeV, $|\eta| < 2.5$
- B_s: $4.766 < m_{\mu\mu} < 5.966$ GeV
  $B^\pm$: $4.930 < m_{\mu\mu K} < 5.630$ GeV
  B_s: $5.050 < m_{\mu\mu KK} < 5.650$ GeV
- PV closest in z to point of closest approach to beam axis of backward extrapolated B momentum

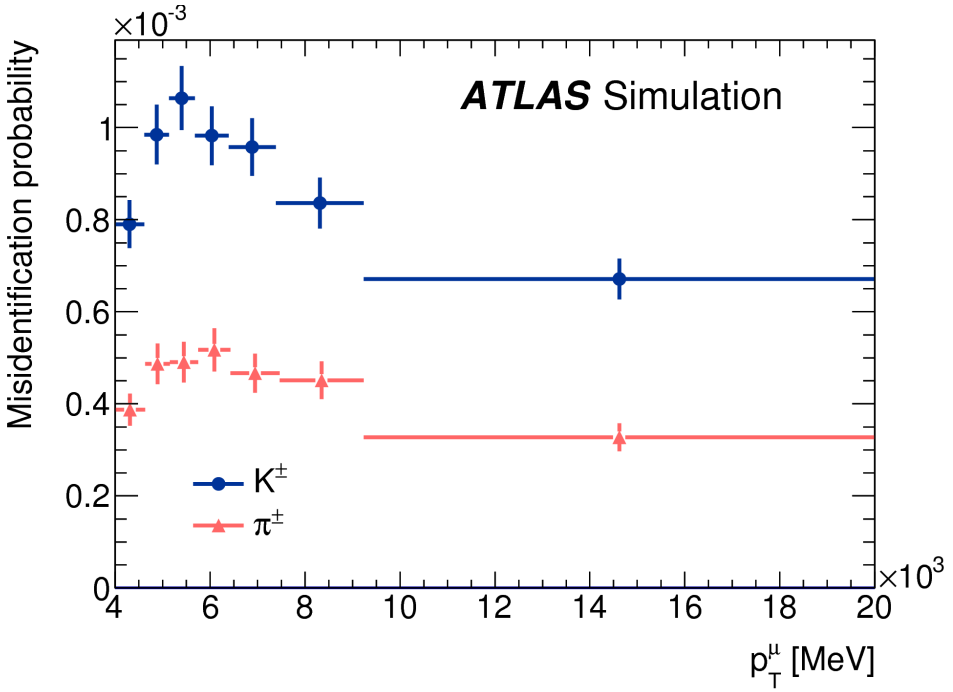
Trigger selection (2012 data):
- T1: $p_T > 6$ GeV and > 4 GeV for 2 μ "higher threshold"
- T2: $p_T > 4$ GeV for both μ and one μ with $|\eta| < 1.05$ and not in T1; "barrel"
- T3: $p_T > 4$ GeV for 2 μ and not in T1, T2

2011 data:
- $p_T > 4$ GeV for 2 μ

Fake-BDT variables

1. Absolute value of the track rapidity measured in the ID.
2. Ratio q/p (charge over momentum) measured in the MS.
3. Scattering curvature significance: maximum variation of the track curvature between adjacent layers of the ID.
4. χ^2 of the track reconstruction in the MS.
5. Number of hits used to reconstruct the track in the MS.
6. Ratio of the values of q/p measured in the ID and in the MS, corrected for the average energy loss in the calorimeter.
7. χ^2 of the match between the tracks reconstructed in the ID and MS.
8. Energy deposited in the calorimeters along the muon trajectory obtained by combining ID and MS tracks.

Misidentification probability for $K^{\pm}$ and $\pi^{\pm}$



BDT Input Variables

Variable	Description
p_{T}^{B}	Magnitude of the B candidate transverse momentum $\overrightarrow{p_{\mathrm{T}}}^{B}$.
$\chi^2_{\mathrm{PV,DV}\ xy}$	Significance of the separation $\overrightarrow{\Delta x}$ between production (*i.e.* associated PV) and decay (DV) vertices in the transverse projection: $\overrightarrow{\Delta x}_{\mathrm{T}} \cdot \Sigma^{-1}_{\overrightarrow{\Delta x}_{\mathrm{T}}} \cdot \overrightarrow{\Delta x}_{\mathrm{T}}$, where $\Sigma_{\overrightarrow{\Delta x}_{\mathrm{T}}}$ is the covariance matrix.
ΔR	three-dimensional opening between $\overrightarrow{p}^{B}$ and $\overrightarrow{\Delta x}$: $\sqrt{{\alpha_{\mathrm{2D}}}^2 + \Delta\eta^2}$
$\lvert\alpha_{\mathrm{2D}}\rvert$	Absolute value of the angle between $\overrightarrow{p_{\mathrm{T}}}^{B}$ and $\overrightarrow{\Delta x}_{\mathrm{T}}$ (transverse projection).
L_{xy}	Projection of $\overrightarrow{\Delta x}_{\mathrm{T}}$ along the direction of $\overrightarrow{p}_{\mathrm{T}}^{B}$: $(\overrightarrow{\Delta x}_{\mathrm{T}} \cdot \overrightarrow{p_{\mathrm{T}}}^{B})/\lvert\overrightarrow{p_{\mathrm{T}}}^{B}\rvert$.
$\mathrm{IP}_{B}^{\mathrm{3D}}$	three-dimensional impact parameter of the B candidate to the associated PV.
$\mathrm{DOCA}_{\mu\mu}$	Distance of closest approach (DOCA) of the two tracks forming the B candidate (three-dimensional).
$\Delta\phi_{\mu\mu}$	Difference in azimuthal angle between the momenta of the two tracks forming the B candidate.
$\lvert d_0\rvert^{\mathrm{max}}$-sig.	Significance of the larger absolute value of the impact parameters to the PV of the tracks forming the B candidate, in the transverse plane.
$\lvert d_0\rvert^{\mathrm{min}}$-sig.	Significance of the smaller absolute value of the impact parameters to the PV of the tracks forming the B candidate, in the transverse plane.
$P_{\mathrm{L}}^{\mathrm{min}}$	Value of the smaller projection of the momenta of the muon candidates along $\overrightarrow{p_{\mathrm{T}}}^{B}$.
$I_{0.7}$	Isolation variable defined as ratio of $\lvert\overrightarrow{p_{\mathrm{T}}}^{B}\rvert$ to the sum of $\lvert\overrightarrow{p_{\mathrm{T}}}^{B}\rvert$ and of the transverse momenta of all additional tracks contained within a cone of size $\Delta R < 0.7$ around the B direction. Only tracks with $p_{\mathrm{T}} > 0.5$ GeV and matched to the same PV as the B candidate are included in the sum.
$\mathrm{DOCA}_{\mathrm{xtrk}}$	DOCA of the closest additional track to the decay vertex of the B candidate. Tracks matched to a PV different from the B candidate are excluded.
$N_{\mathrm{xtrk}}^{\mathrm{close}}$	Number of additional tracks compatible with the decay vertex (DV) of the B candidate with $\ln(\chi^2_{\mathrm{xtrk,DV}}) < 1$. The tracks matched to a PV different from the B candidate are excluded.
$\chi^2_{\mu,\mathrm{xPV}}$	Minimum χ^2 for the compatibility of a muon in the B candidate with a PV different from the one associated with the B candidate.

Isolation Variable

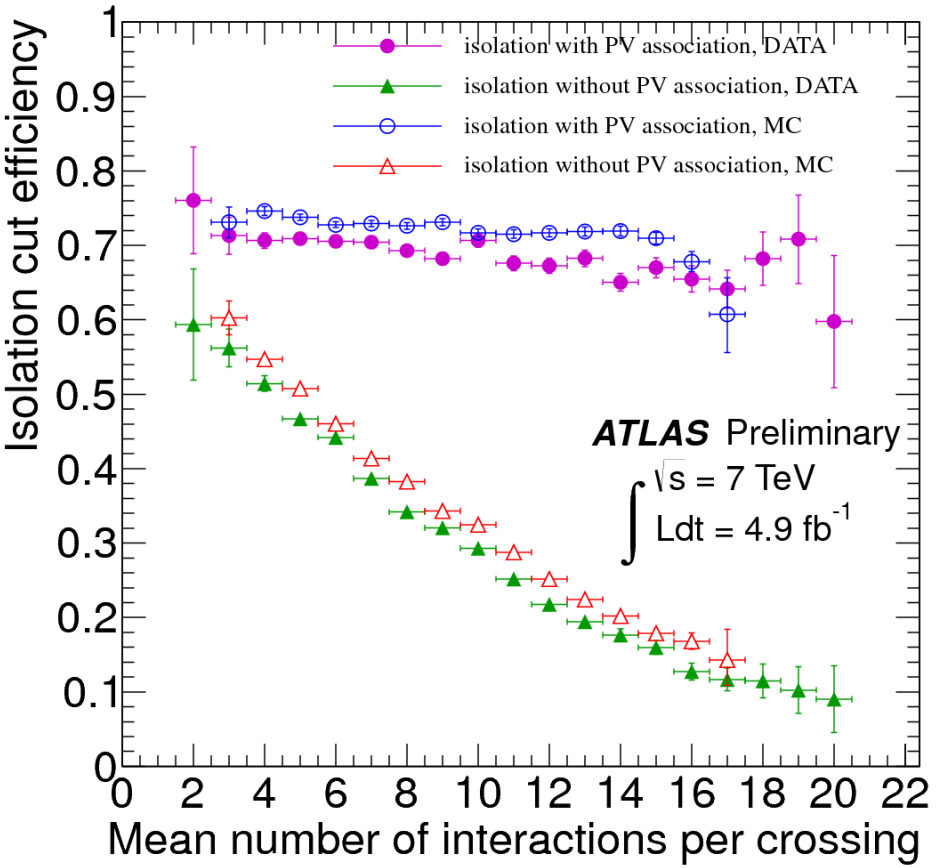
- Isolation variable:

$$I_{\Delta R} = \frac{p_{\mathrm{T}}^{\mathrm{B}}}{p_{\mathrm{T}}^{\mathrm{B}} + \sum_{i_{\mathrm{track}} \in \mathrm{cone}(\Delta R)} p_{\mathrm{T}}^{i_{\mathrm{track}}}}$$

  - Tracks with p_T > 0.5 GeV excluding B daughters in $\Delta R < 0.7$ with

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$$

- PV association of tracks:
  - Removes interference from other interactions
  - Isolation cut efficiency is independent of pile-up



[ATLAS-CONF-2013-076]

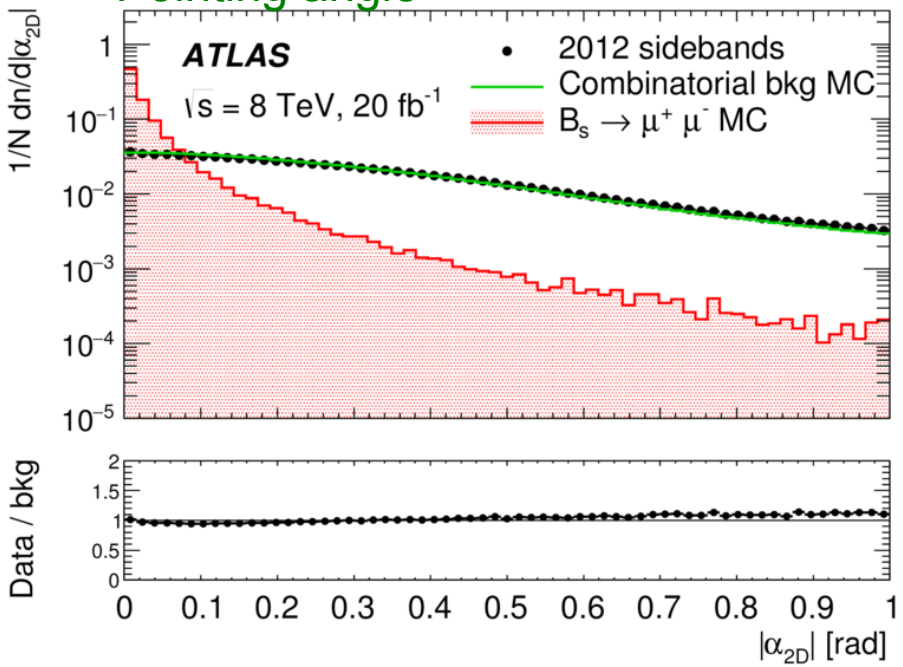
Background Discrimination (1)

Continuum background:

- Dominated by $b\bar{b} \rightarrow \mu\mu X$
- Continuum-BDT:
  - 15 discriminating variables
  - Trained on MC

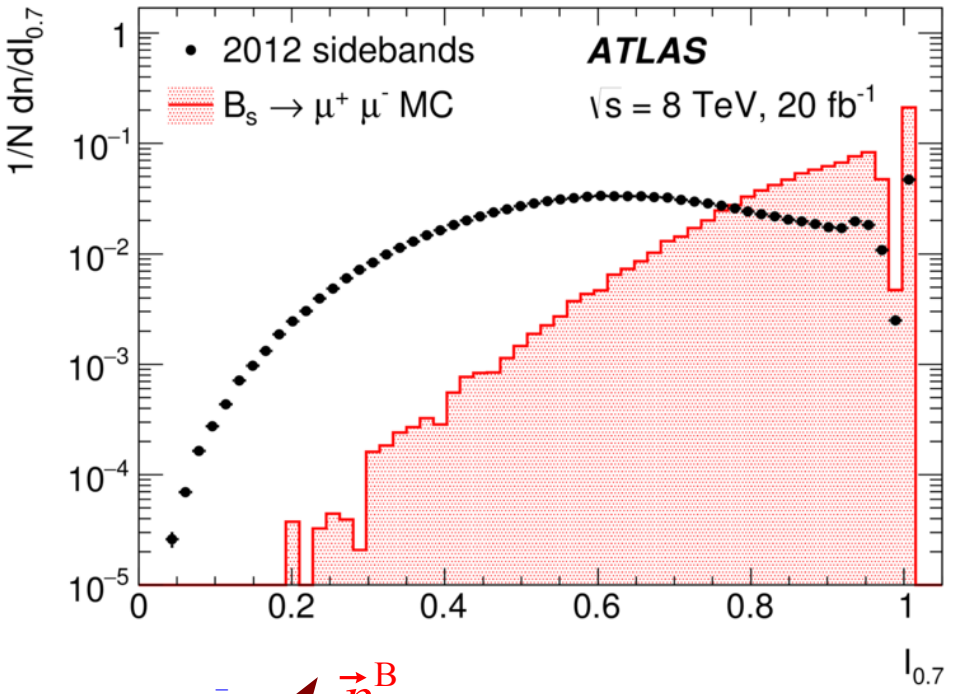
Example of two variables

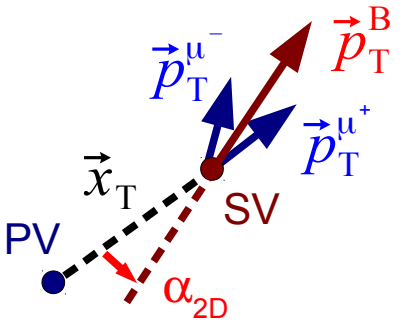
Pointing angle



$$I_{\Delta R} = \frac{p_{\mathrm{T}}^{\mathrm{B}}}{p_{\mathrm{T}}^{\mathrm{B}} + \sum_{i_{\mathrm{track}} \in \mathrm{cone}(\Delta R)} p_{\mathrm{T}}^{i_{\mathrm{track}}}}$$

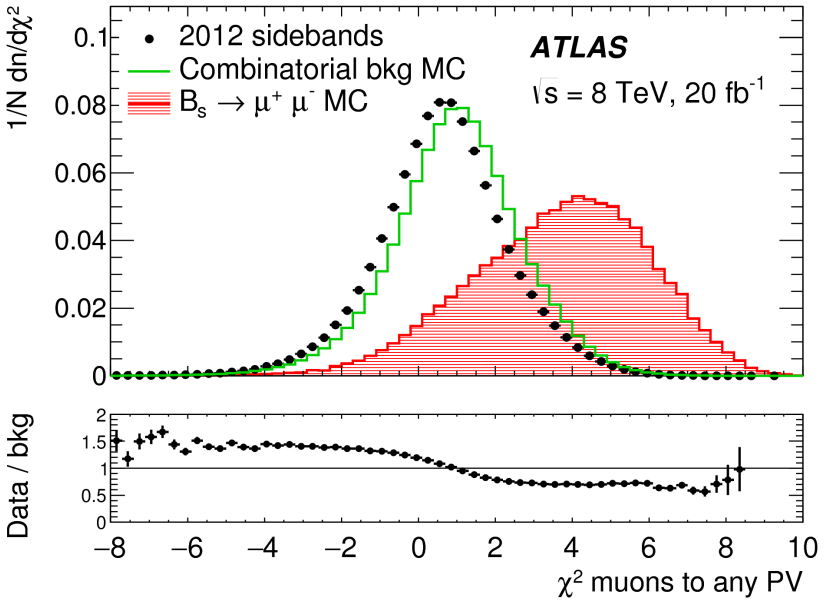
Isolation



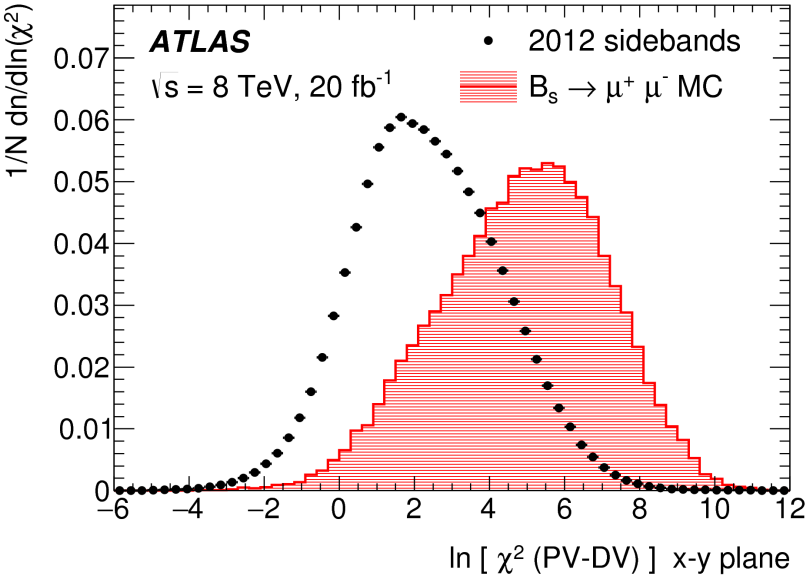


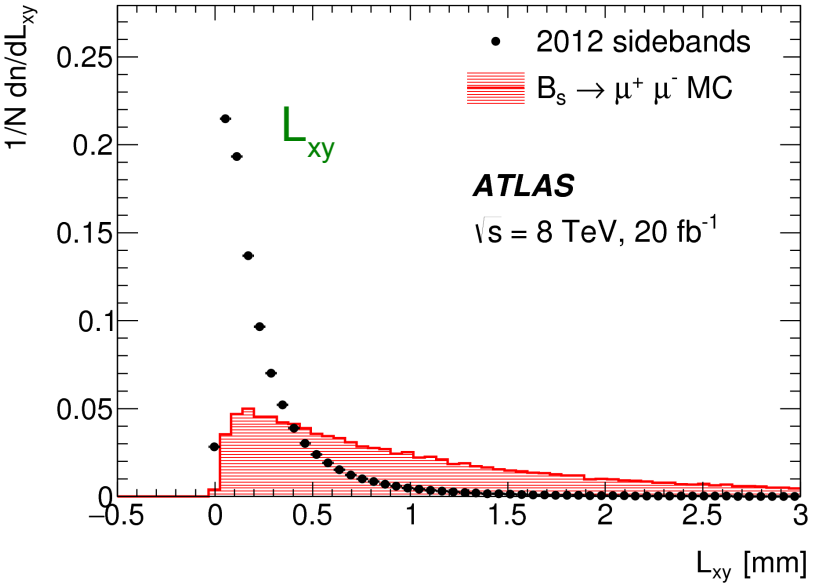
Background Discrimination (2)

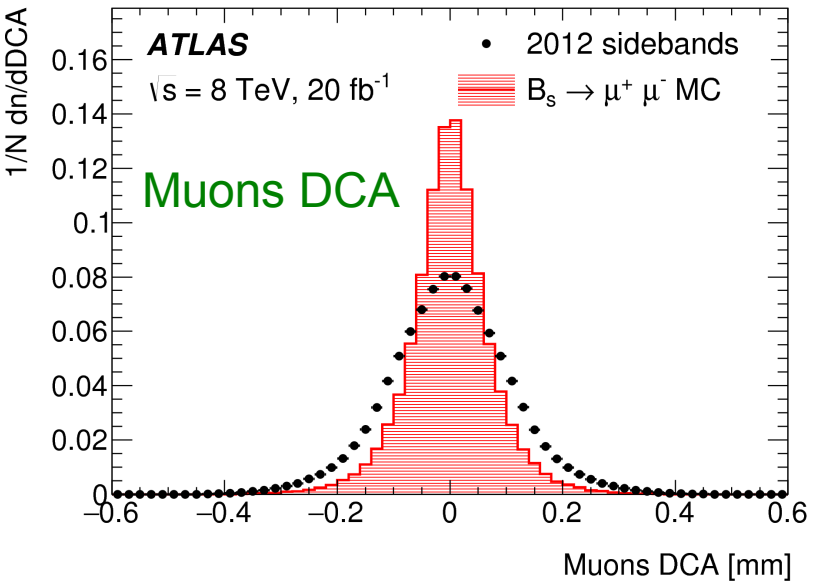
χ^2 muons to any PV



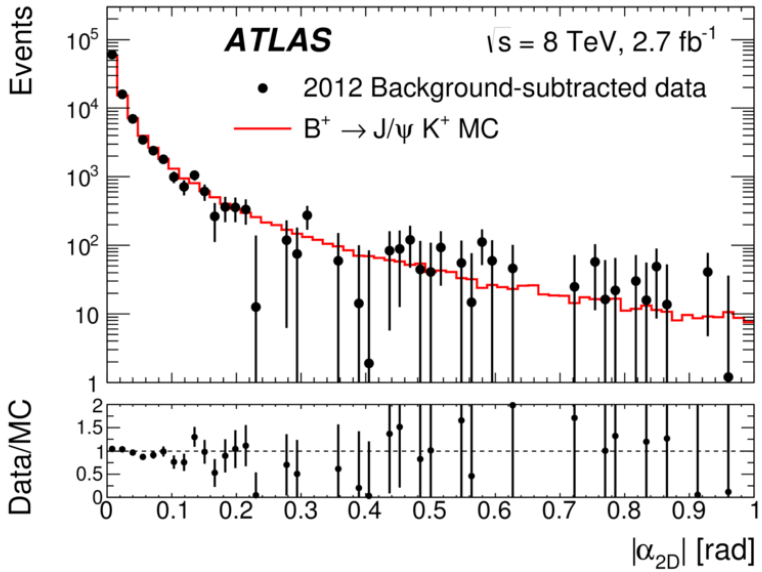
Ln [χ^2 (PV-DV)] x-y plane

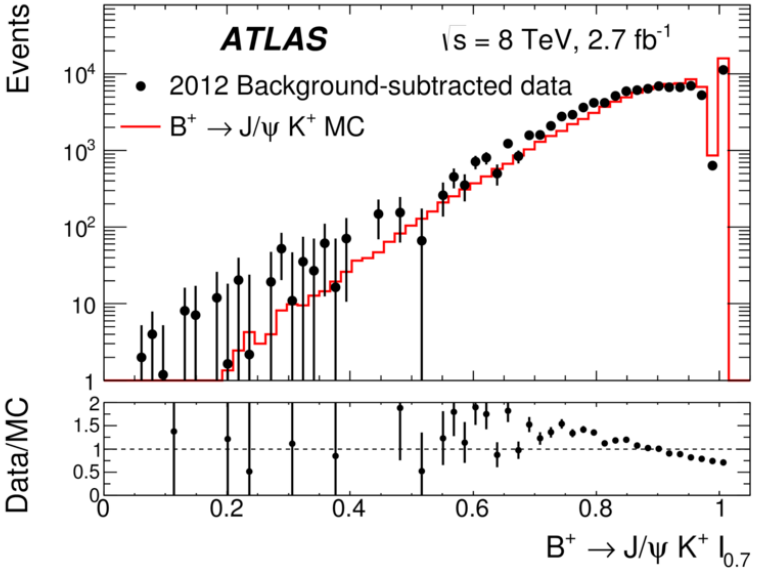


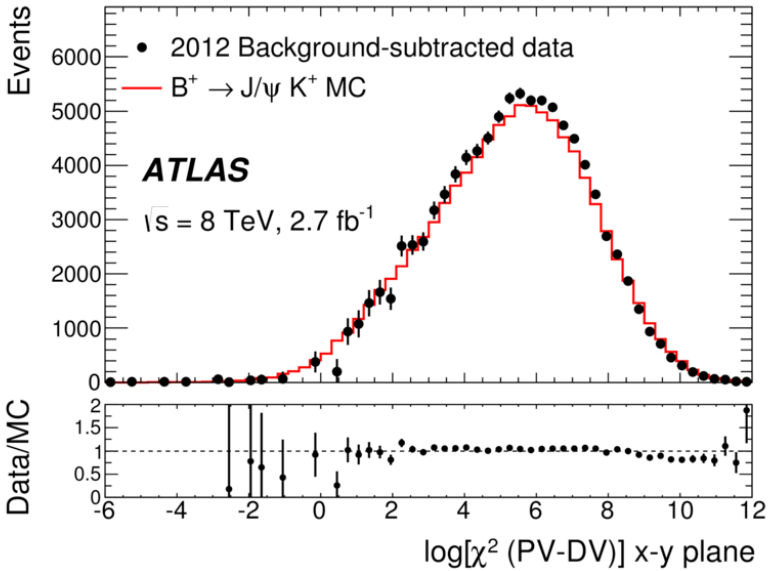


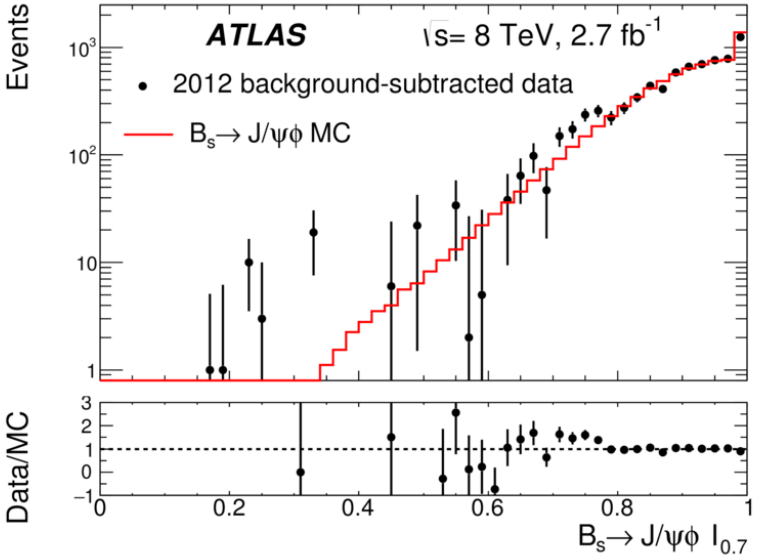


Data – MC Comparisons





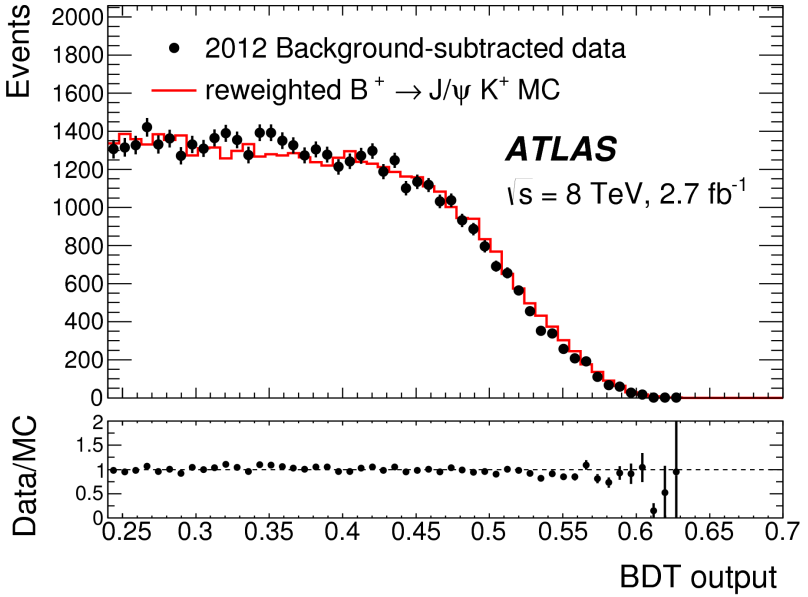




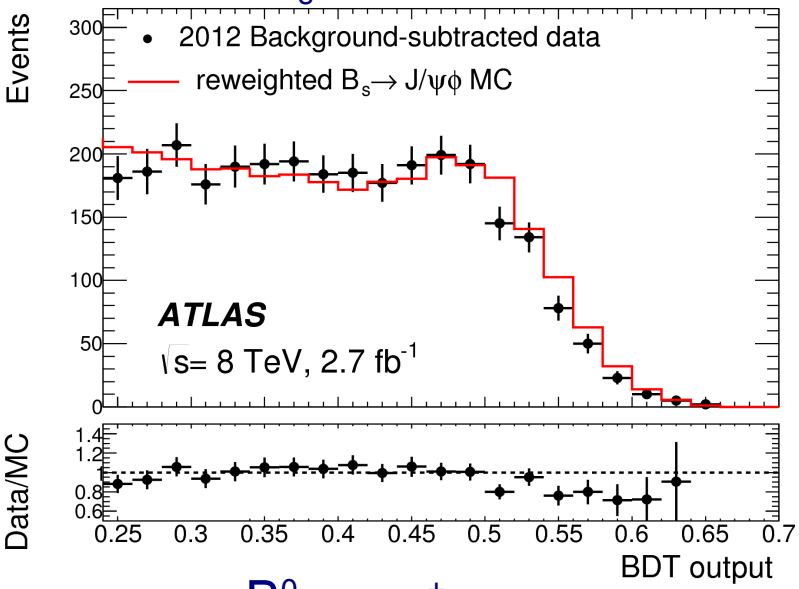
- Good data-MC agreement (except for isolation)
  - Reweight MC samples (B^0_s/B^0 and $B^{\pm}$)

Data – MC Comparisons for c-BDT Output

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

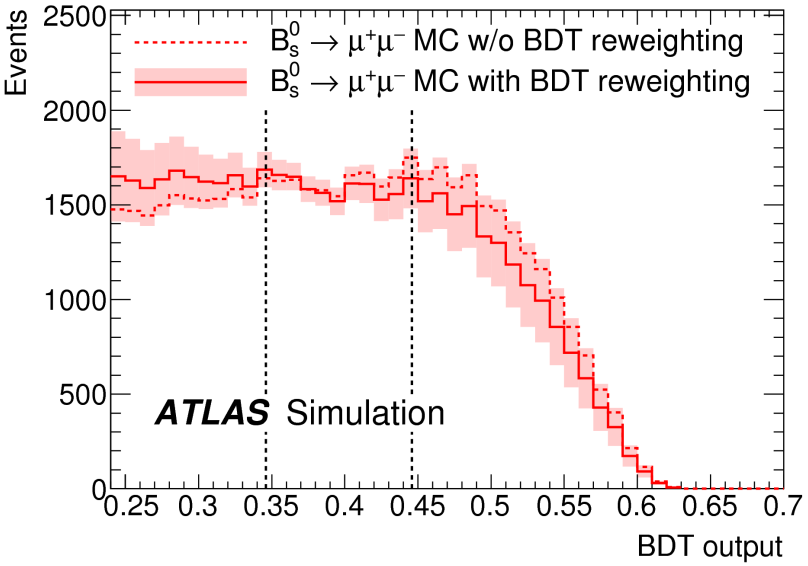


$B^0_s \rightarrow J/\psi\, \Phi$



- MC samples normalized to data
- Linear correction to MC distributions (equal for all channels)
- Systematic uncertainty indicated for $B^0_s \rightarrow \mu^+\mu^-$ MC
- Vertical lines: 3 c-BDT intervals

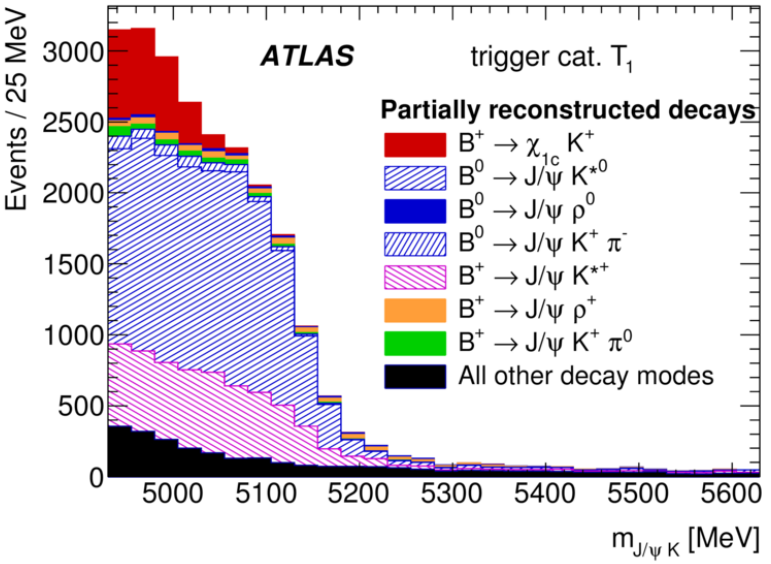
$B^0_s \rightarrow \mu^+\, \mu^-$



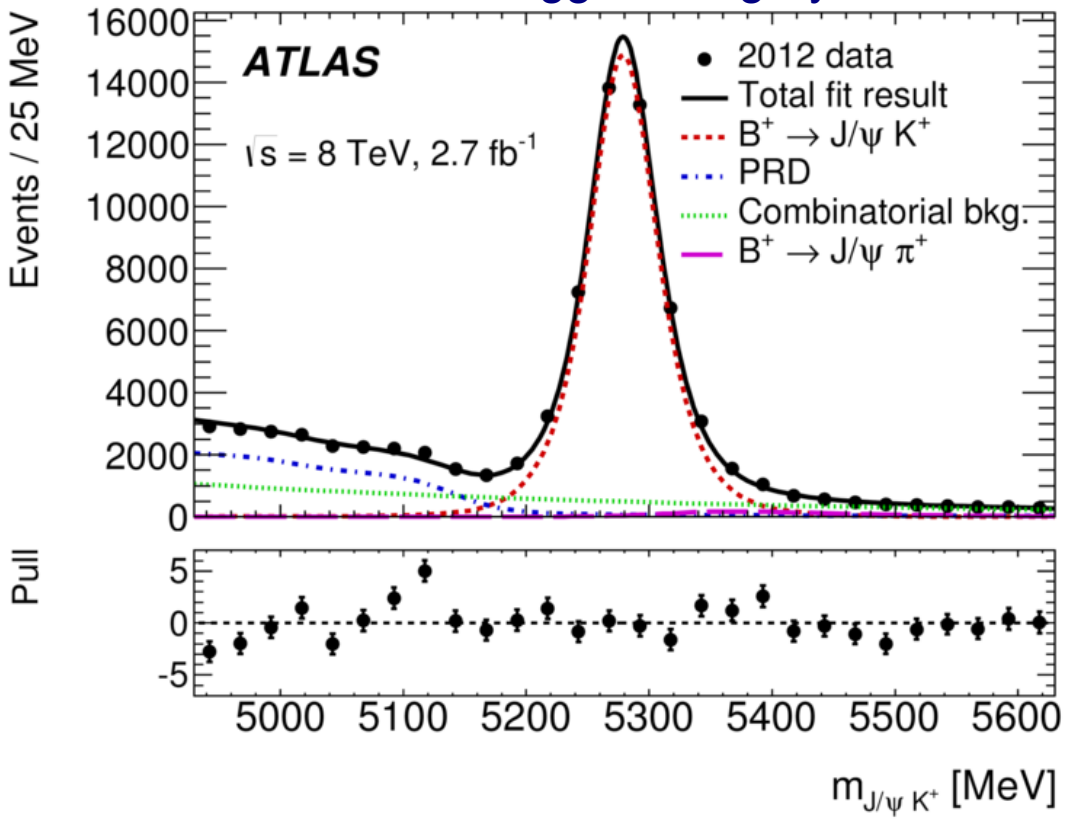
$B^{\pm}$ Yield Extraction

- Applied fake-BDT and continuum-BDT selections
- Separately in 4 data/trigger categories
- Unbinned ML fit of $m_{\mu\mu K}$ distribution

Partially reco'd B decays



Main 2012 trigger category



- Fit results:

Category	$N_{J/\psi K^+}$	$N_{J/\psi \pi^+}$
T_1	$46\,860 \pm 290 \pm 280$	$1\,420 \pm 230 \pm 440$
T_2	$5\,200 \pm 84 \pm 100$	$180 \pm 51 \pm 89$
T_3	$2\,512 \pm 91 \pm 42$	$85 \pm 77 \pm 30$
2011	$95\,900 \pm 420 \pm 1\,100$	$3\,000 \pm 340 \pm 1\,140$

- $B(B^{\pm} \to J/\psi \pi^{\pm})/B(B^{\pm} \to J/\psi K^{\pm}) = 0.035 \pm 0.003 \pm 0.012$

Efficiency Ratio

- Separately computed efficiencies for B^0_s/B^0 and $B^\pm$
- Simulated MC samples used:
  - $p_T(B)$ and $\eta(B)$ spectra corrected using $B^\pm$ data
  - $\varepsilon_{Trigger}$ corrected using J/ψ and Υ data
  - $B^0_s \to \mu^+\mu^-$ lifetime and B isolation corrections
- Fiducial volume: $p_T(B) > 8$ GeV, $|\eta(B)| < 2.5$
- Split into
  - acceptance (cuts on $\mu^\pm$ and $K^\pm$ kinematics)
  - ε terms (identification, trigger, reconstruction and selection cuts)
- Systematic uncertainties (on D_{norm}):

Statistical uncertainty in simulation	0.5%
p_T, η reweighting and trigger efficiency	1.3%
Data to MC discrepancy in discriminating variables	4.2%
K^+ and B^+ reconstruction	3.6%
Residual trigger efficiency systematic uncertainty	1.5%
B^+ yield	0.8%
Total uncertainty	5.9%

Systematic Uncertainties for $B(B^0_{(s)} \to \mu^+\mu^-)$

	$\mathcal{B}(B^0_s \to \mu^+\mu^-)$	$\mathcal{B}(B^0 \to \mu^+\mu^-)$
Scale uncertainties		
$\mathcal{B}(B^+ \to J/\psi K^+) \times \mathcal{B}(J/\psi \to \mu\mu)$ branching fractions	3.1%	3.1%
$B^0_{(s)}/B^+$ production ratio	8.3%	0
B^+ yield and $B^0_{(s)}/B^+$ efficiency ratio	5.9%	5.9%
Relative efficiency of continuum-BDT intervals	9%	9%
Signal and background model	6%	0
Total scale uncertainty	16%	11%
Offset uncertainties		
Signal and background model	0.2×10^{-9}	0.7×10^{-10}

- Completely correlated uncertainties between both channels:
  - Branching fractions
  - $B^{\pm}$ yield
  - Efficiency ratio
  - Efficiency of c-BDT intervals
- All uncertainties included in profile-likelihood fit by Gaussian terms

Interpolated Background in Cont.-BDT Bins

- Number of background events interpolated in the blinded region per continuum-BDT output bin

Bin	Continuum BDT range		# background events
1	0.240	0.346	509 ± 28
2	0.346	0.446	32 ± 6
3	0.446	1.0	5 ± 2

- Expected signal from SM prediction:
  N_s = 41 events
  N_d = 5 events

B^0_s/B^0 Signal Fit Model

N_s and N_d by unbinned extended simultaneous maximum likelihood fit in $m_{\mu\mu}$:

- In three intervals in continuum-BDT output (each ε_{sig} = 18%):
  [0.242-0.351], [0.351-0.454], [0.455, ∞]
- Extract N_s and N_d

- Signal model:
  - B^0_s signal: two Gaussians,
    (same mean, avg. width 80 MeV, independent of continuum-BDT)
  - Shape parameters from MC
  - B^0 signal → similar shape from MC ($BF(B^0_s)/BF(B_0)$ = 0.113 in SM)
  - Mass shape and relative fraction fixed in 3 c-BDT intervals

B^0_s/B^0 Background Fit Models

N_s and N_d by unbinned extended simultaneous ML fit in $m_{\mu\mu}$:

- Combinatorial $\mu\mu$ background:
  - 1rst order Chebychev polynomial
  - $m_{\mu\mu}$ slope and relative fractions different in c-BDT intervals
- SS-SV background
  - Tails reach into signal region
  - Exponential function
  - Fixed $m_{\mu\mu}$ shape in 3 c-BDT intervals
  - Fractions in 3 c-BDT intervals fitted
- Semi-leptonic background
  - Mainly $B^0 \to \pi\mu\nu$, $B^0_s \to K\mu\nu$, $\Lambda_b \to p\mu\nu$
  - Contribute at low $m_{\mu\mu}$, like tail of a Gaussian
  - No separate PDF, only for studies of systematics
- "Peaking" background
  - Signal-like PDF, normalization fixed to 1.0 ± 0.4 events
  - Fixed mass shape and relative fractions in 3 c-BDT intervals
- Components normalized independently in each c-BDT interval

Branching Fraction Extraction

$$BR\left(B^0_{(s)} \to \mu^+\mu^-\right) = N_{d(s)} \cdot \frac{1}{D_{norm}} \cdot \frac{f_u}{f_{d(s)}}$$
$$\cdot BR\left(B^+ \to J/\psi K^+\right) \cdot BR\left(J/\psi \to \mu^+\mu^- K^+\right)$$

$$D_{norm} = \sum_k N^k_{J/\psi K^+} \alpha_k \left\{\frac{\epsilon_{\mu\mu}}{\epsilon_{J/\psi K^+}}\right\}_k$$

Ingredients to normalization:

- $B^{\pm}$ and J/ψ branching fractions [PDG 2014]
- f_u/f_s from ATLAS $f_s/f_d = 0.240 \pm 0.020$ [PRL 115 (2015) 262001]
  assuming $f_u/f_d = 1$ (same p_T & η range!)
- $\{\epsilon_{J/\psi K^+}/\epsilon_{\mu\mu}\}_k$ and $N^k_{J/\psi K^+}$ in D_{norm}

D_{norm} = (2.88 ± 0.17) x 10^{-6} (B^0_s)
= (2.77 ± 0.16) x10^{-6} (B^0)

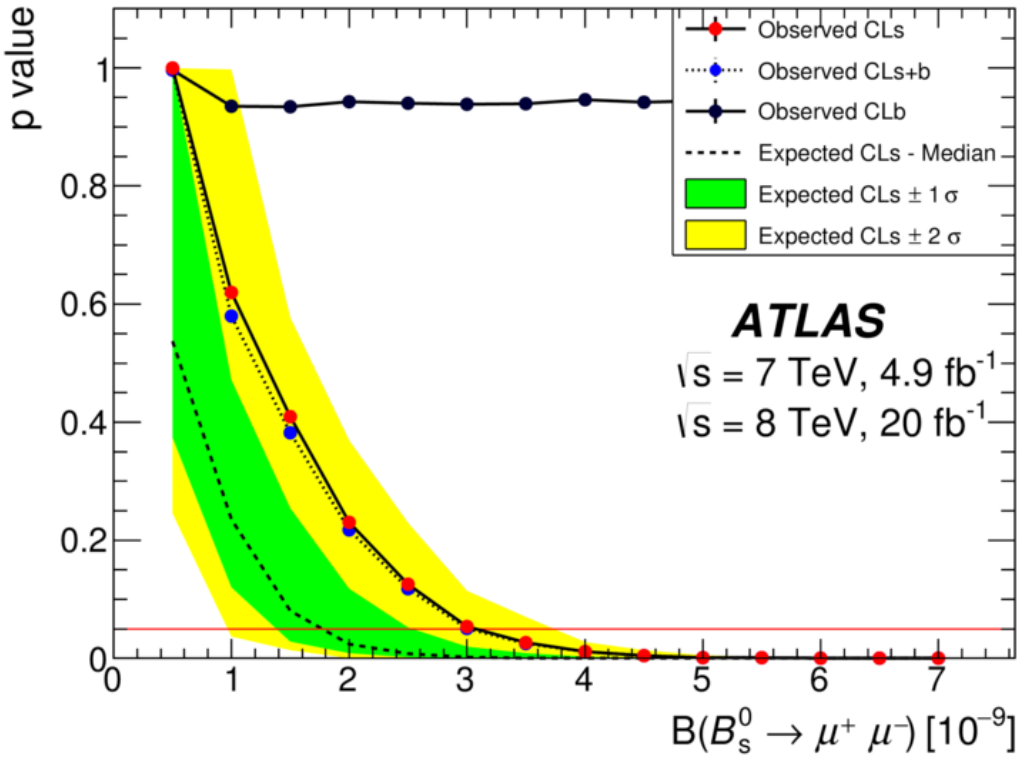
- Total normalization uncertainty:
  - ± 11% for BR($B^0_s \to \mu^+\mu^-$)
  - ± 7% for BR($B^0 \to \mu^+\mu^-$)

SES = (8.9 ± 1.0) x 10^{-11} (B^0_s)
= (2.21 ± 0.15) x10^{-11} (B^0)

$B^0_s \to \mu^+\mu^-$ Branching Fraction Result

$BR(B^0_s \to \mu^+\mu^-) = 0.9\ ^{+1.1}_{-0.8} \times 10^{-9}$

- Obtained within boundary of non-negative branching fractions
- Errors by frequentist belt using pseudo-MC experiments
  - $\sigma_{syst} = \pm 0.3 \times 10^{-9}$ (included above)
- Upper limit (CLs method):
  $BR(B^0_s \to \mu^+\mu^-) < 3.0 \times 10^{-9}$
  at 95% CL
  Expected upper limit
  (no signal, $BR(B^0 \to \mu^+\mu^-)$ free):
  $BR(B^0_s \to \mu^+\mu^-) < 1.8\ ^{+0.7}_{-0.4} \times 10^{-9}$
- Compatibility with null hypothesis
  (no signal, $BR(B^0 \to \mu^+\mu^-)$ free):
  $p = 0.08\ (1.4\ \sigma)$
  - Expected significance assuming SM $BR(B^0_s \to \mu^+\mu^-)$: 3.1 σ



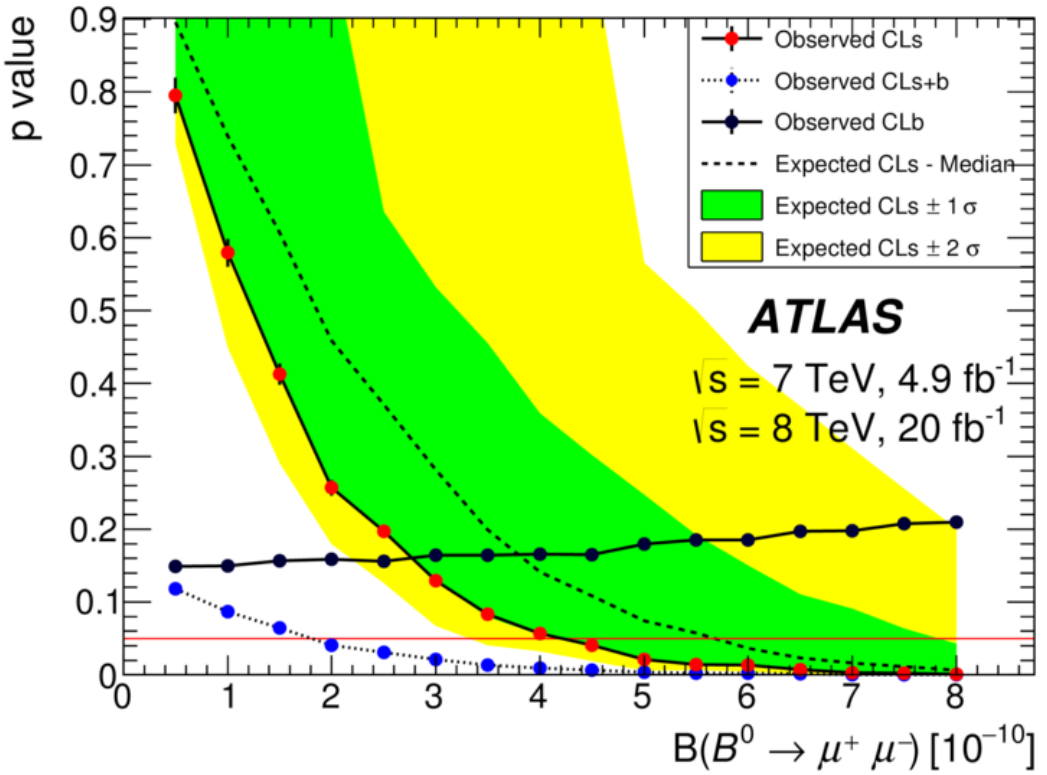
$B^0 \to \mu^+\mu^-$ Branching Fraction Result

- Upper limit (CLs method, no signal, BR($B^0_s \to \mu^+\mu^-$) free):
  BR($B^0 \to \mu^+\mu^-$) < 4.2 x 10^{-10} at 95% CL

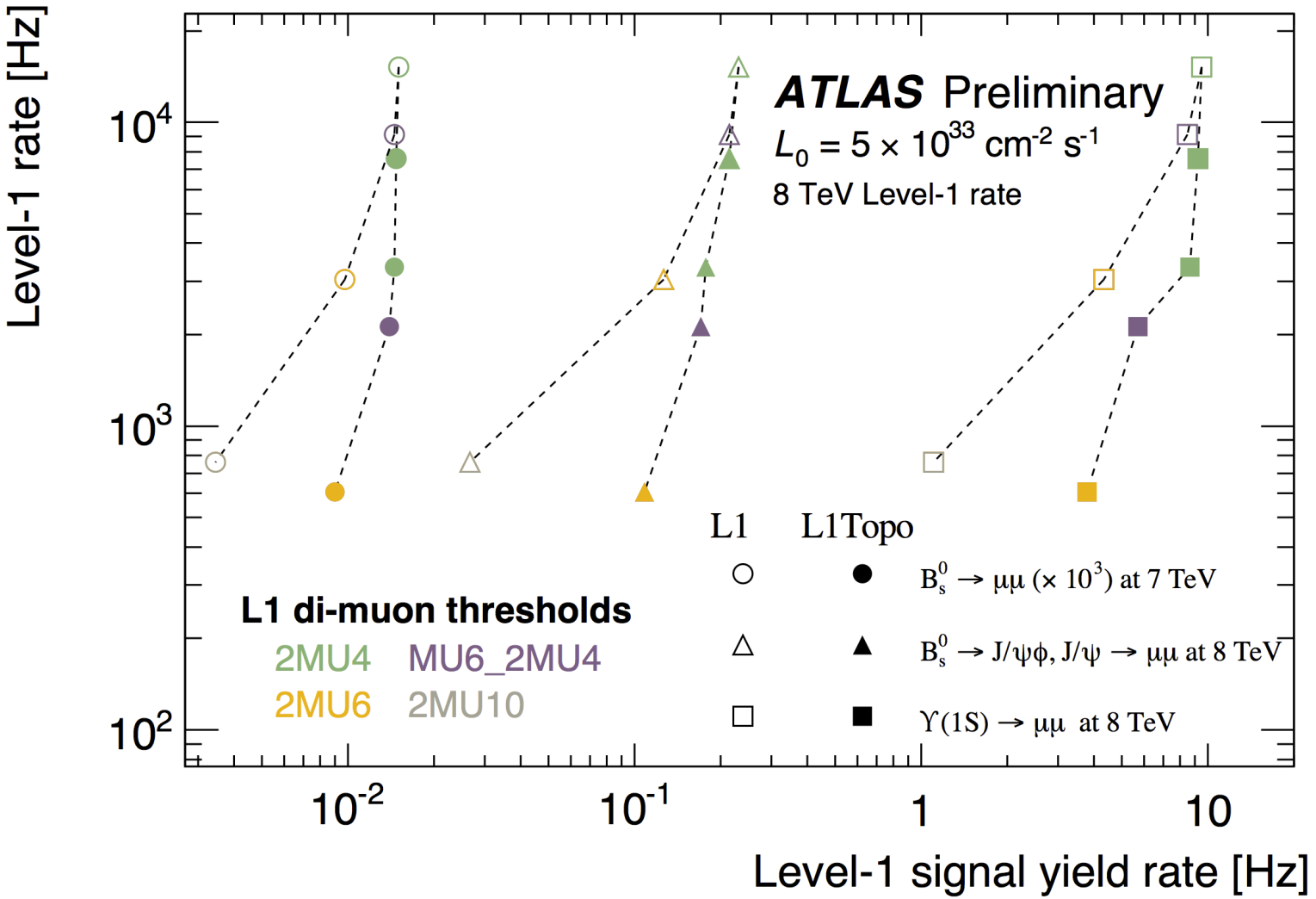
- CLb ≈ 0.15 for BR($B^0 \to \mu^+\mu^-$) near 0:
  → -1 σ fluctuation of background

- Expected upper limit:
  BR($B^0 \to \mu^+\mu^-$) < $5.7\ ^{+2.1}_{-1.2}$ x 10^{-10}

- Limit larger than SM prediction:
  BR($B^0 \to \mu^+\mu^-$)$_{SM}$ =
  (1.06 ± 0.09) x 10^{-10}
- Expected significance assuming SM: 0.2 σ



Topological L1 Trigger



Menu optimized for all B physics channels

- under commissioning for 2016 data