BPH-13-007

CMS+LHCb $B_{s,d} \to \mu\mu$ Combination Approval

CMS Contact Persons/Editors: K.-F. Chen, J. Bendavid, J. Butler

September 3, 2014

INTRODUCTION

- The previous combination PAS (under the same BPH-13-007) was based on
  - A naive combination with toy MC.
  - The f_s/f_u had been factorized out and synchronized with LHCb.
- **This is a more thorough combination down to the fitter level:**
  - Produce a combined UML fitter. The two $B_{s,d}\to\mu\mu$ branching fractions are fitted as common parameters of interests.
  - Synchronize the input variables as well as the correlated systematics.
  - $B_d\to\mu\mu$ limit is evaluated with RooFit/RooStats code adapted from Higgs Combine (CMS) or a custom made FC tool (LHCb).
  - CWR of the paper draft ends on Sep/8.
- Contact persons and editors:
  - CMS: Jack Kai-Feng Chen, Josh Bendavid, Joel Butler
  - LHCb: Marc-Olivier Bettler, Francesco Dettori, Val Gibson

THE APPROACH

- **Based on the RooWorkspace outputs from the two experiments**
  - Fitters have been prepared with common $B_{s,d}\to\mu\mu$ branching fractions, f_s/f_u (5% Gaussian constraint to a common LHCb measurement), and $B^+\to J/\psi K^+$ branching fraction.
  - PDF has been “re-structured” in order to match the requirement of the RooStats-based FC code.
  - Joint CMS and LHCb PDFs with `RooSimultaneous` (CODE A) or `RooNLL+RooMinuit` (CODE B).
- “Harmonisation” between two analyses:
  - Maintaining the same analyses/fitting strategy as in original publications, adopt the same external inputs (physics parameters, branching fractions, models) as much as possible.
  - CMS side: introduce the new **$\Lambda_b\to p\mu\nu$ model**, as well as the **lifetime efficiency correction**.
  - Many thanks to Urs and Luca for providing data and help for this work!

REVIEW MATERIALS

- Link to the paper draft (shared version): https://cds.cern.ch/record/1752194/files/BPH-13-007-paper-v3.pdf
- Link to the AN (shared version + CMS only appendix): http://cms.cern.ch:80/iCMS/jsp/openfile.jsp?tp=draft&files=AN2014_167_v2.pdf
- Link to the twiki: https://twiki.cern.ch/twiki/bin/view/CMS/B2MuMuCombinationReview

Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data

The CMS and LHCb collaborations

To be submitted to Nature

In this presentation:

PAPER

= figures included in paper.

EXTENDED

= figures included in extended data (not printed, online only).

CMS SIDE:
Updates since summer 2013

- Input branching fractions and parameters
- Modeling of $\Lambda_b \rightarrow p\mu\nu$ background
- Time-dependent correction
- Updated fitting results

INPUT PARAMETERS

	Value or branching fraction	Ref.
f_s/f_u	0.256±0.025 (stat+syst)	LHCb
$B^\pm \to J/\psi\, K^\pm$	$(6.02 \pm 0.20) \times 10^{-5}$	[1]
$B_s \to J/\psi\, \phi$	$(3.16 \pm 0.76) \times 10^{-5}$	[1]
$B_s \to K^+K^-$	$(2.54 \pm 0.38) \times 10^{-5}$	[2]
$B_s \to \pi^+K^-$	$(5.00 \pm 1.10) \times 10^{-6}$	[2]
$B_s \to \pi^+\pi^-$	$(7.30 \pm 1.40) \times 10^{-7}$	[2]
$B^0 \to K^+K^-$	$(1.30 \pm 1.00) \times 10^{-7}$	[2]
$B^0 \to K^+\pi^-$	$(1.96 \pm 0.06) \times 10^{-5}$	[2]
$B^0 \to \pi^+\pi^-$	$(5.11 \pm 0.22) \times 10^{-6}$	[2]
$\Lambda_b \to p\pi^-$	$(3.50 \pm 1.00) \times 10^{-6}$	[1]
$\Lambda_b \to pK^-$	$(5.50 \pm 1.40) \times 10^{-6}$	[1]
$B_s \to K^-\mu^+\upsilon$	$(1.42 \pm 0.07) \times 10^{-4}$	[2]
$B^0 \to \pi^-\mu^+\upsilon$	$(1.42 \pm 0.07) \times 10^{-4}$	[2]
$\Lambda_b \to p\mu^-\nu$	$(6.50 \pm 6.50) \times 10^{-4}$	[3]
$B^+ \to \pi^+\mu^+\mu^-$	$(2.30 \pm 0.60) \times 10^{-8}$	[4]
$B^+ \to K^+\mu^+\mu^-$	$(4.80 \pm 0.40) \times 10^{-7}$	[1]
$B^0 \to \pi^0\mu^+\mu^-$	$(1.20 \pm 0.60) \times 10^{-8}$	[1]
$B^0 \to K^0\mu^+\mu^-$	$(3.80 \pm 0.80) \times 10^{-7}$	[1]
$B^0 \to \mu^+\mu^-\gamma$	$(1.34 \pm 0.27) \times 10^{-10}$	[5]
$B_s \to \mu^+\mu^-\gamma$	$(1.89 \pm 0.38) \times 10^{-8}$	[5]

UPDATES

➡ $f_d/f_s = 3.86 \pm 0.22 \otimes 5\%$ Gaussian

➡ $(6.10 \pm 0.19) \times 10^{-5}$ [PDG 2013/2014]

➡ $(1.27 \pm 0.59) \times 10^{-4}$ [PRD 86 (2012) 114025]

➡ $(1.44 \pm 0.05) \times 10^{-4}$ [PDG 2013/2014]

➡ $(4.94 \pm 2.19) \times 10^{-4}$ [JHEP 1109 (2011) 106]

➡ $(3.4 \pm 0.5) \times 10^{-7}$ [PDG 2013/2014]

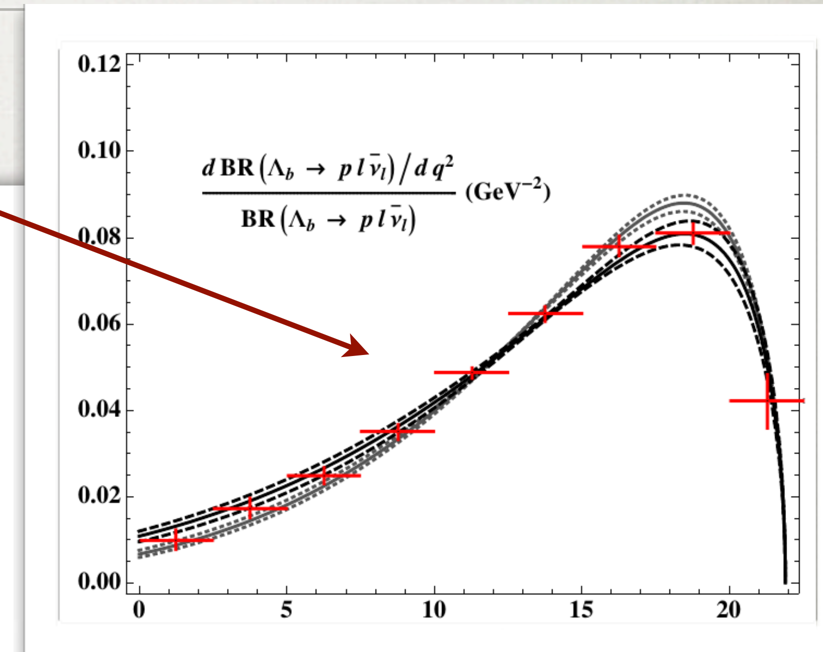
[1] PDG 2012
[2] HFAG
[3] Average of published calculations.
[4] LHCb JHEP 1212, 125 (2012).
[5] D. Melikhov and N. Nikitin, PRD 70, 114028 (2004).

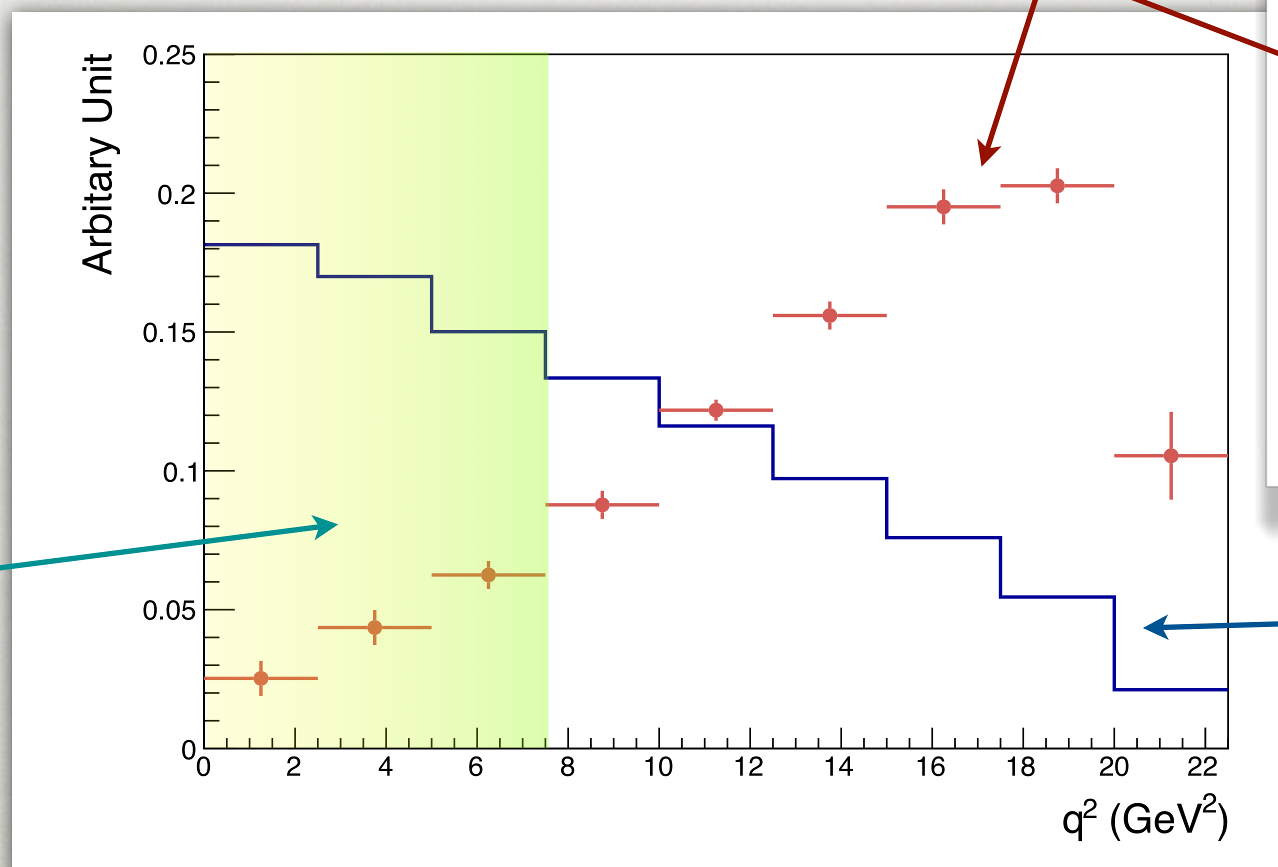
$\Lambda_b \rightarrow p\mu\nu$ MODEL

- This is the dominant semileptonic B decay background in our analysis. Huge uncertainty due to the totally unknown branching fraction.
- The branching fraction and model used in summer paper:
  - BF = 6.5 x 10^{-4}, ±100% uncertainty.
  - Model: phase-space decay from EvtGen.
- Updated BF & model (synchronized with LHCb):
  - Based on A. Khodjamirian et al, JHEP 1109 (2011) 106 + some other patches (lifetime, V_{ub}).
  - BF = (4.94 ± 2.19) x 10^{-4}
  - Model: based on the TH reference next slide (calculated as a function of q^2)

$\Lambda_b \to p\mu\nu$ MODEL

Prediction



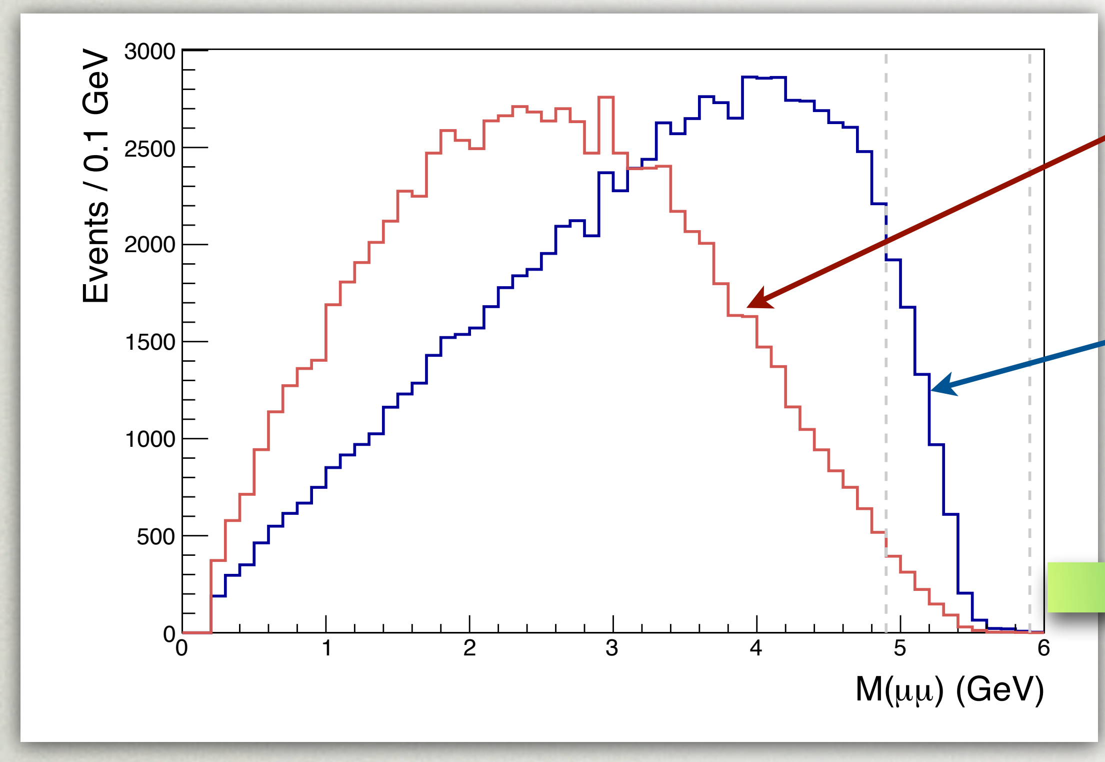


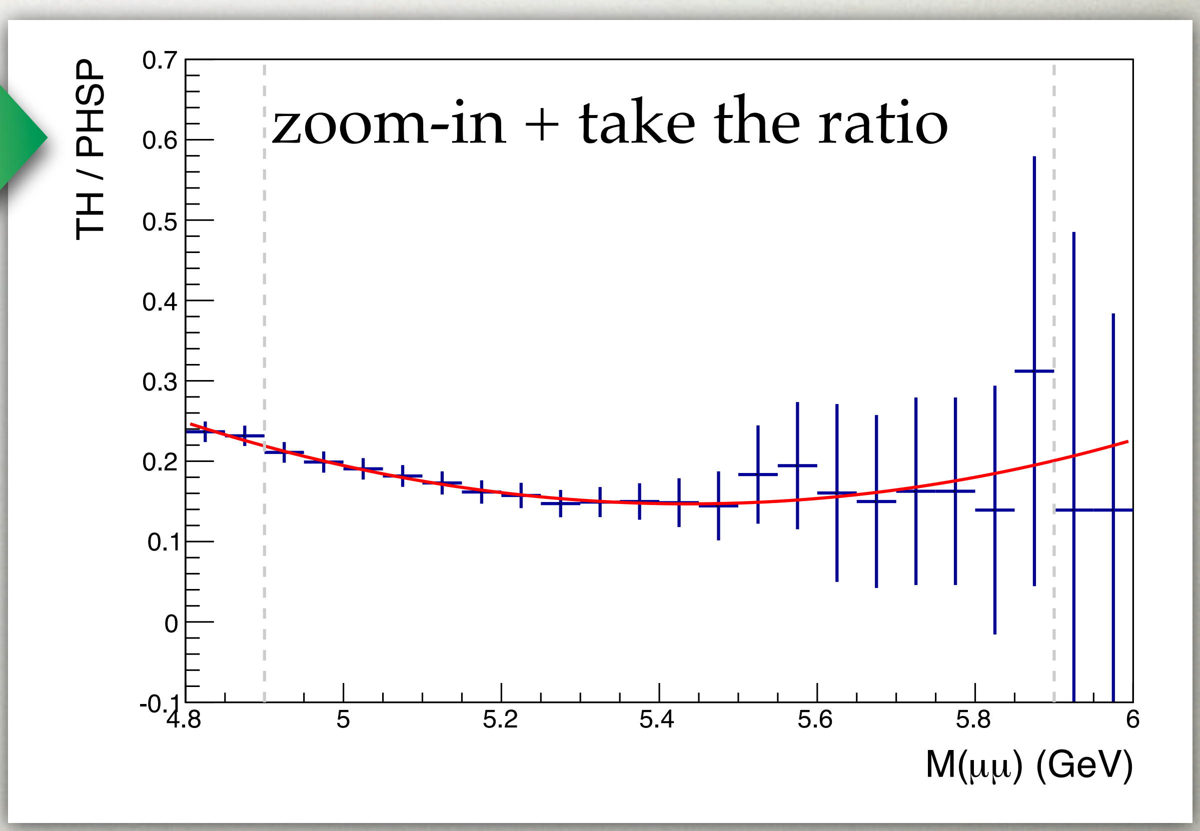
The first 3 bins are relevant to $B \to \mu\mu$ analysis.

EvtGen/PHSP model

- $q^2 = M^2(\mu\nu)$: Lower q^2 = higher $M(p\mu)$ = closer to $B \to \mu\mu$ signal region.
- The quoted branching fractions are not that different (6.5×10^{-4} versus 4.94×10^{-4}), the new q^2 dependent model will reduce the $\Lambda_b \to p\mu\nu$ contribution by a large factor.

$\Lambda_b \to p\mu\nu$ MODEL





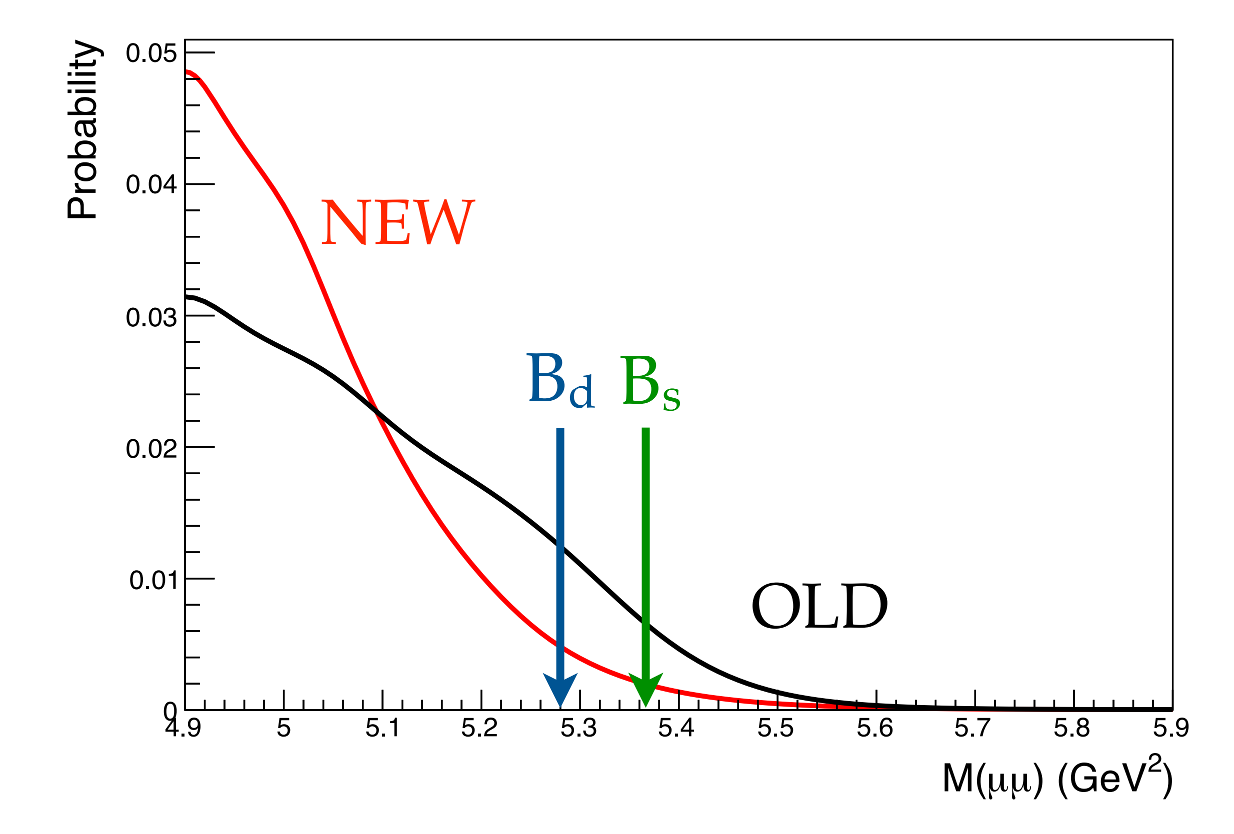
- The average contribution in the signal region should be scaled by ~0.179 (from model) and by 0.76 (from BF), resulting the factor of ~7.5 reduction.
- Inject this correction (the curve) into the PDF construction.

MODIFICATION TO THE SEMILEPTONIC PDF

(as an example)

Yields in 2012/barrel/highest BDT bin
low mass sideband

Channel	Yield (old)	Yield (new)
$B_s \to K\mu\nu$	2.04±1.02	1.82±1.24
$B^0 \to \pi\mu\nu$	3.31±1.66	3.36±1.68
$\Lambda_b \to p\mu\nu$	13.31±14.88	1.81±1.21
$B^+ \to \pi\mu\mu$	1.14±0.30	1.14±0.30
$B^+ \to K\mu\mu$	-	-
$B^0 \to \pi^0\mu\mu$	0.73±0.36	0.73±0.36
$B^0 \to K^0\mu\mu$	-	-
$B^0 \to \mu\mu\gamma$	0.014±0.003	0.014±0.003
$B_s \to \mu\mu\gamma$	0.54±0.11	0.54±0.11
Sum	**21.1±15.0**	**9.4±2.5**



- Resulting the shape and total expected yield change on the semileptonic component.
- A similar update to all of the bins.

LIFETIME CORRECTION

- Objective: the B_s lifetime (and the time-dependent decay rate) is slightly different between the MC we used and the best up-to-date knowledge. This changes the $B_s \to \mu\mu$ efficiencies slightly.

This is what we used in the MC generation, $\tau_{gen} = 1.461$ ps

$$\Gamma(B_s \to \mu\mu) \propto e^{-t/\tau_{\rm gen}}$$

This is the time-dependent untagged decay rate

$$\Gamma(B_s \to \mu\mu) \propto e^{-t/\tau_{B_s}} \left[\cosh\left(\frac{y_s t}{\tau_{B_s}}\right) + A_{\Delta\Gamma} \sinh\left(\frac{y_s t}{\tau_{B_s}}\right)\right]$$

From PDG & HFAG

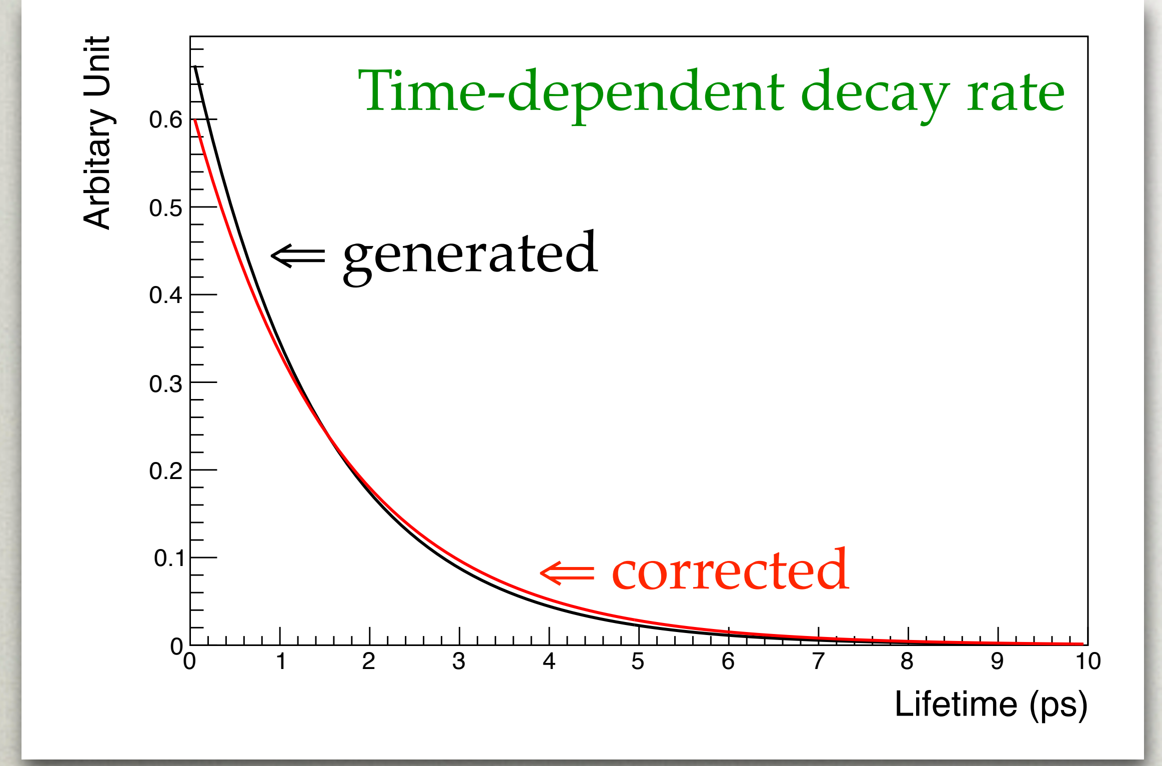
$$\tau_{B_s} = 1.516 \pm 0.011 \text{ ps}$$

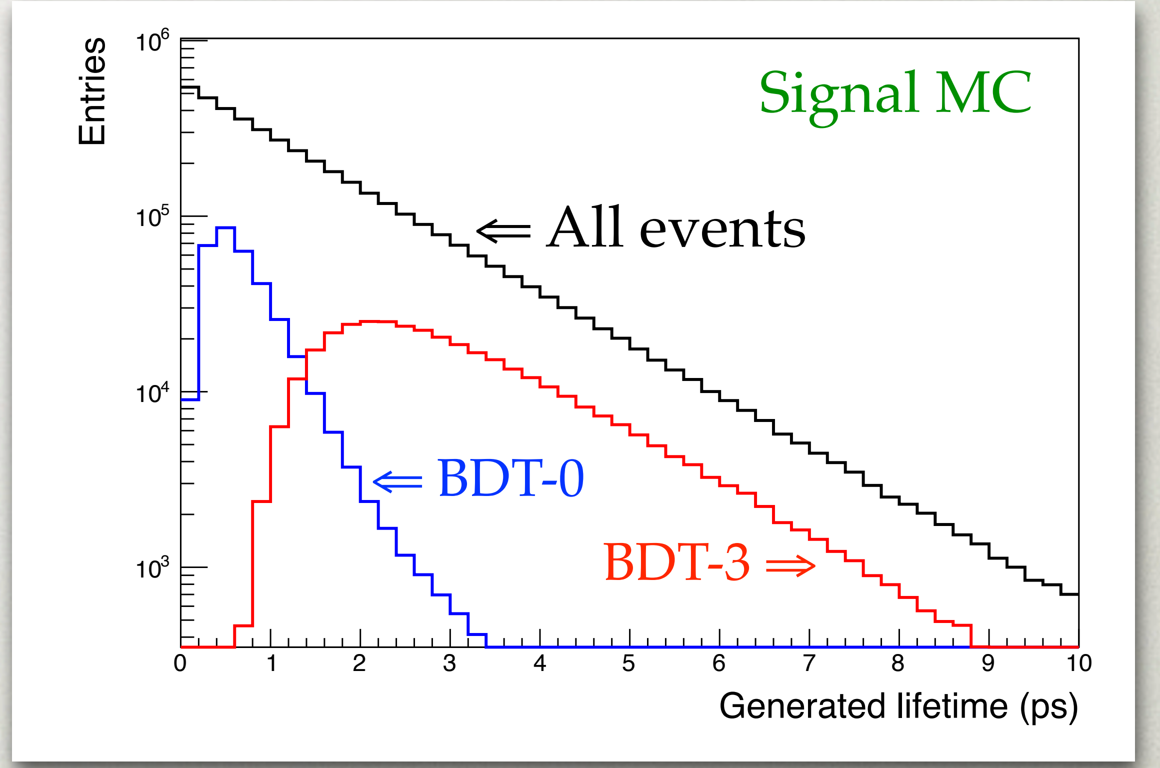
$$y_s = \frac{\Gamma_L - \Gamma_H}{\Gamma_L + \Gamma_H} = 0.0615 \pm 0.0085$$

SM value $A_{\Delta\Gamma} = 1$

LIFETIME CORRECTION

- The approach: re-weight the signal MC events with the time-dependent decay rate function according to the generated lifetime.
- Estimate the correction in each `[2011/2012]-[barrel/endcap]-[BDT]` bin.





It is clear that the lifetime is **correlated** with BDT.

LIFETIME CORRECTION

Efficiency correction in each bin:

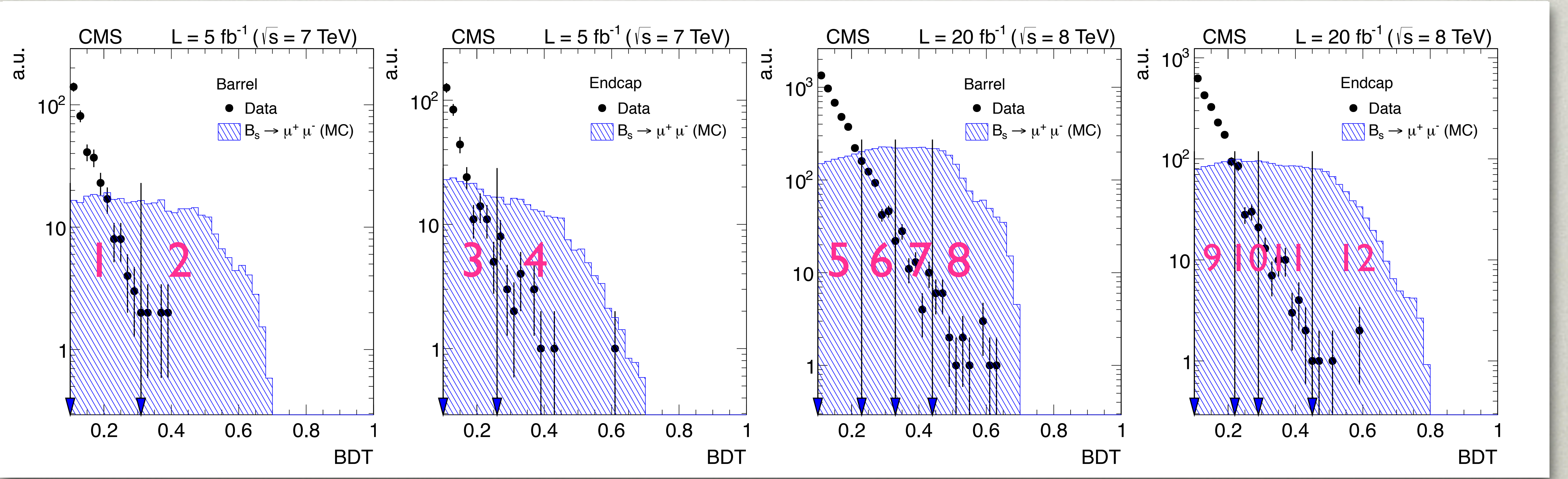
	Correction	$\delta\tau_B$	δy_s
2011/barrel/BDT-0	–0.0250	±0.0020	±0.0025
2011/barrel/BDT-1	+0.0905	±0.0060	±0.0075
2011/endcap/BDT-0	–0.0213	±0.0018	±0.0022
2011/endcap/BDT-1	+0.0883	±0.0058	±0.0073
2012/barrel/BDT-0	–0.0487	±0.0037	±0.0047
2012/barrel/BDT-1	–0.0180	±0.0015	±0.0019
2012/barrel/BDT-2	+0.0366	±0.0024	±0.0030
2012/barrel/BDT-3	+0.1185	±0.0077	±0.0097
2012/endcap/BDT-0	–0.0402	±0.0031	±0.0039
2012/endcap/BDT-1	–0.0067	±0.0007	±0.0008
2012/endcap/BDT-2	+0.0479	±0.0032	±0.0039
2012/endcap/BDT-3	+0.1088	±0.0071	±0.0088

(uncertainties propagated from PDG/HFAG τ_B and y_s)

- Low BDT bin
  ⇒ shorter candidate lifetime
  ⇒ negative correction (~ –3%)
- High BDT bin
  ⇒ longer candidate lifetime
  ⇒ positive correction (~ +10%)
- **The major correction is from the difference of τ_B in the generation (1.461 ps) and PDG (1.516 ps).**
- The uncertainties propagated from τ_B and y_s are fairly small (<1%). Can be neglected safely.

EVENTS CATEGORIZING

- As a reminder: the events are categorized according to the beam energy and detector region, as well as the analysis BDT value.
- This is the same main analysis strategy used for the summer 2013 publication – there are 12 BDT-categories in total.

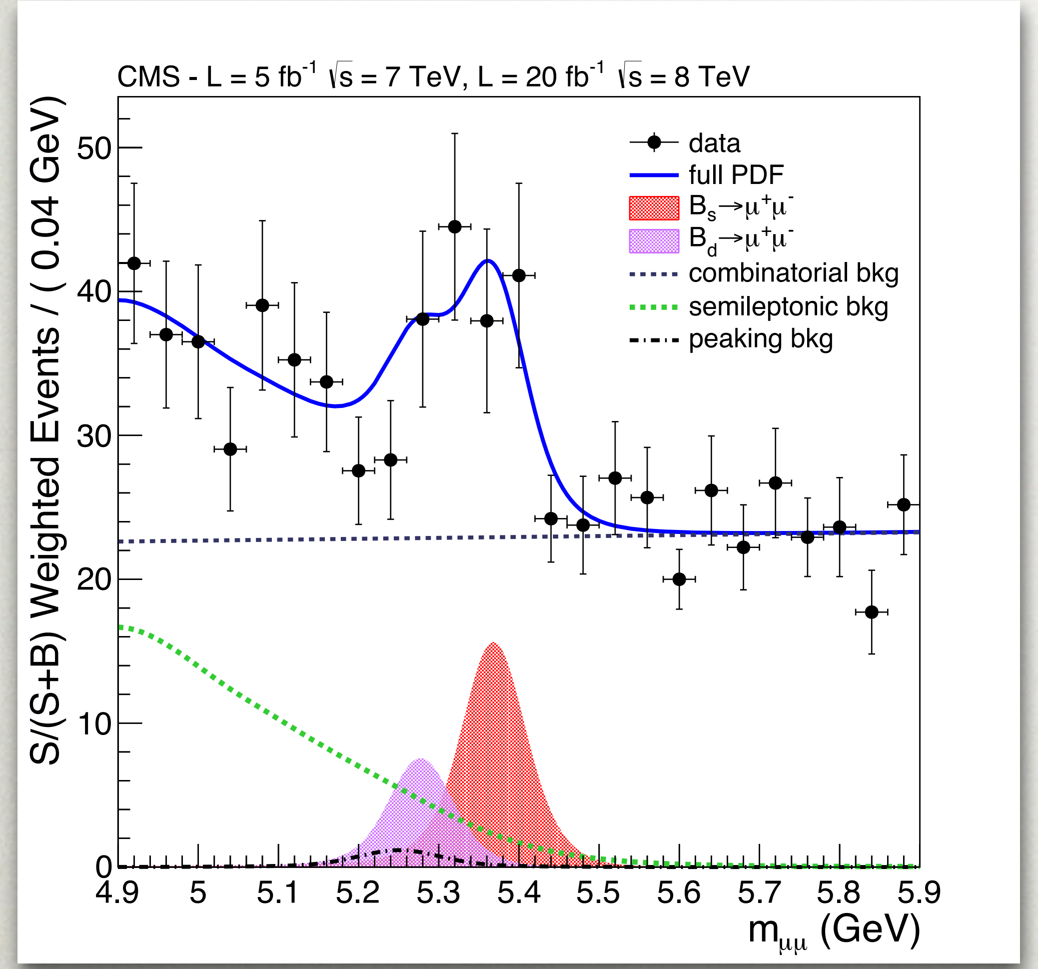


A similar strategy is used by the LHCb analysis with 8 BDT-categories.

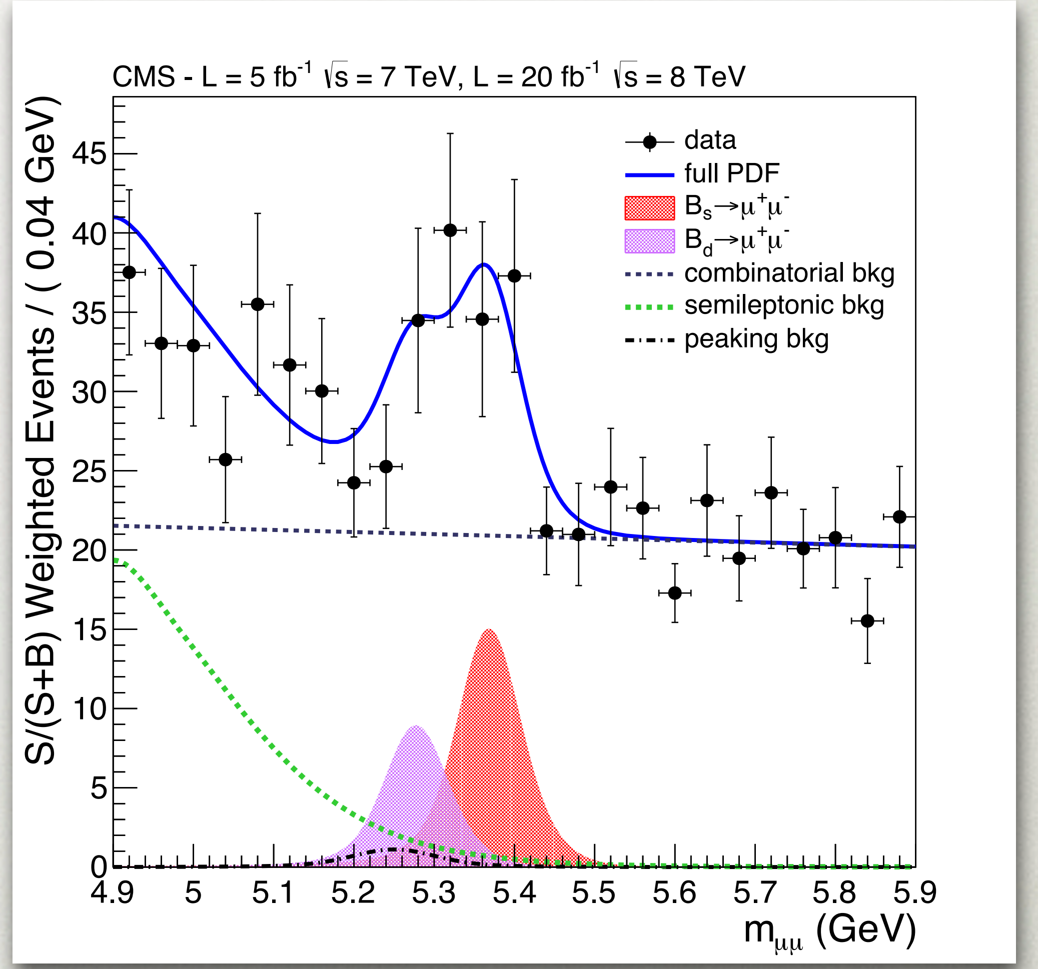
UPDATED FIT RESULTS

- A simultaneous fit to 12 BDT-categories:

ORIGINAL (published)



UPDATED
(Λ_b model+lifetime correction)



$BF(B_s) = (2.99\ +1.04/-0.88) \times 10^{-9}$
$BF(B_d) = (3.48\ +2.13/-1.81) \times 10^{-10}$

$BF(B_s) = (2.80\ +0.95/-0.81) \times 10^{-9}$
$BF(B_d) = (4.36\ +2.23/-1.91) \times 10^{-10}$

SEMILEPTONIC BACKGROUND

Semileptonic background yields in each bin:

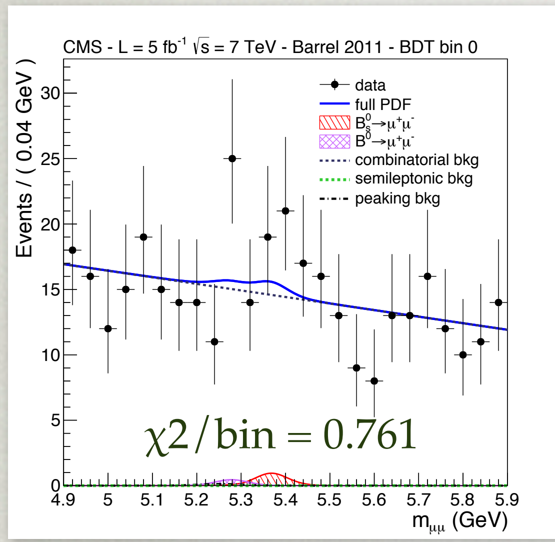
	Expected	Fitted	Expected (new)	Fitted (new)
2011/barrel/BDT-0	43.1 ± 39.5	0.0 ± 7.3	12.5 ± 4.1	10.6 ± 3.9
2011/barrel/BDT-1	27.2 ± 23.0	3.2 ± 2.4	9.4 ± 2.9	5.5 ± 2.2
2011/endcap/BDT-0	18.0 ± 16.0	13.1 ± 12.9	5.6 ± 1.8	5.6 ± 1.8
2011/endcap/BDT-1	11.9 ± 9.9	9.4 ± 4.6	4.3 ± 1.3	4.5 ± 1.2
2012/barrel/BDT-0	87.6 ± 79.4	56.4 ± 46.5	26.1 ± 8.3	26.5 ± 8.0
2012/barrel/BDT-1	71.2 ± 63.2	0.0 ± 8.4	22.2 ± 6.9	16.9 ± 6.1
2012/barrel/BDT-2	55.9 ± 47.6	31.4 ± 13.1	18.9 ± 5.7	20.4 ± 3.9
2012/barrel/BDT-3	29.0 ± 23.4	10.4 ± 5.3	10.8 ± 3.2	9.6 ± 2.5
2012/endcap/BDT-0	21.2 ± 18.6	30.3 ± 15.6	6.8 ± 2.1	7.0 ± 2.1
2012/endcap/BDT-1	23.8 ± 20.3	23.6 ± 11.2	8.0 ± 2.4	8.8 ± 2.2
2012/endcap/BDT-2	15.0 ± 12.2	11.5 ± 7.4	5.5 ± 1.6	6.0 ± 1.4
2012/endcap/BDT-3	8.6 ± 6.6	1.6 ± 2.0	3.4 ± 1.0	2.9 ± 0.9

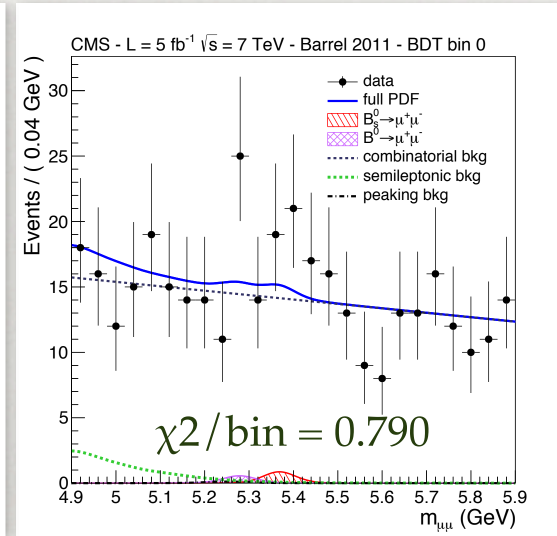
The expected yields are constrained with Gaussians in the fit.

PROJECTIONS 1-ON-1

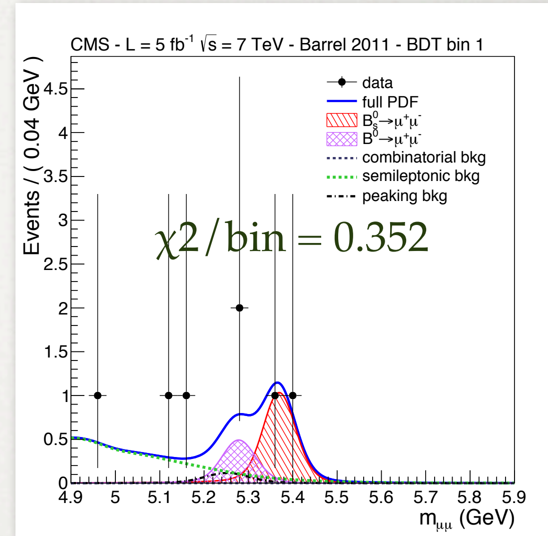
PUB | NEW | PUB | NEW

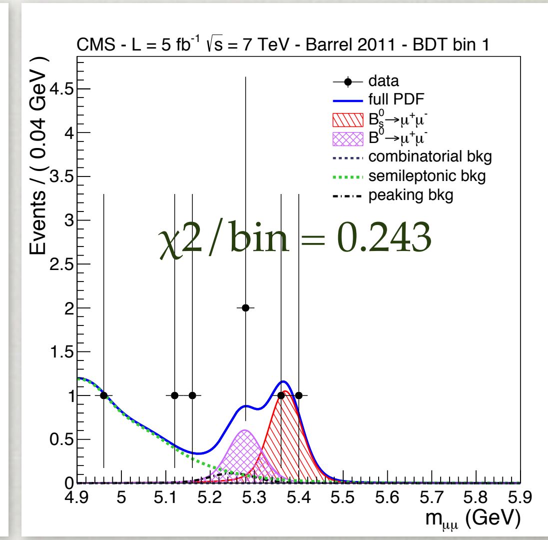
2011/barrel/BDT-0



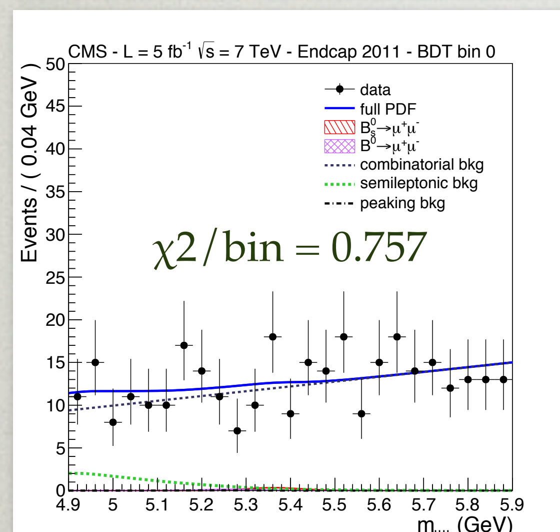


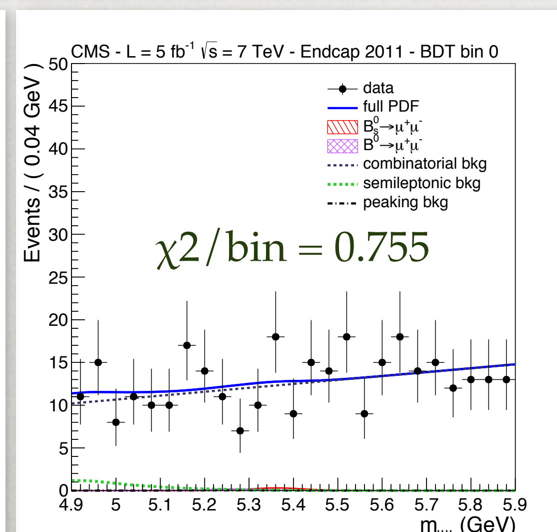
2011/barrel/BDT-1



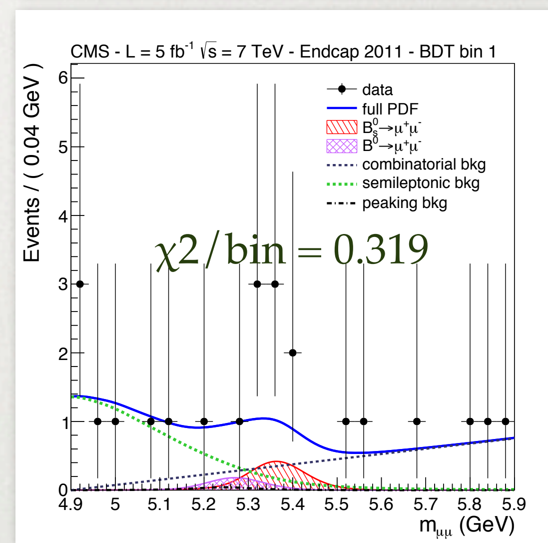


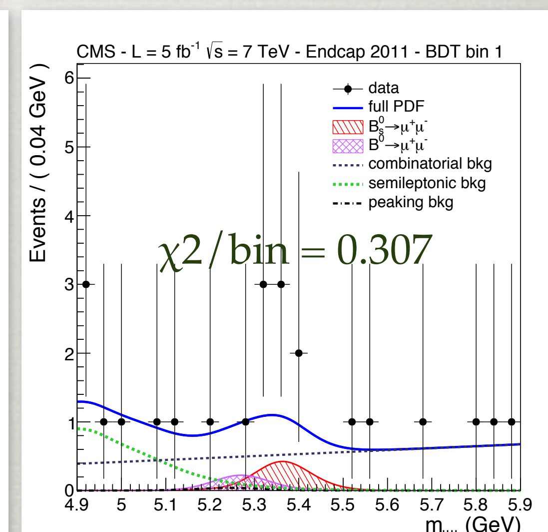
2011/endcap/BDT-0



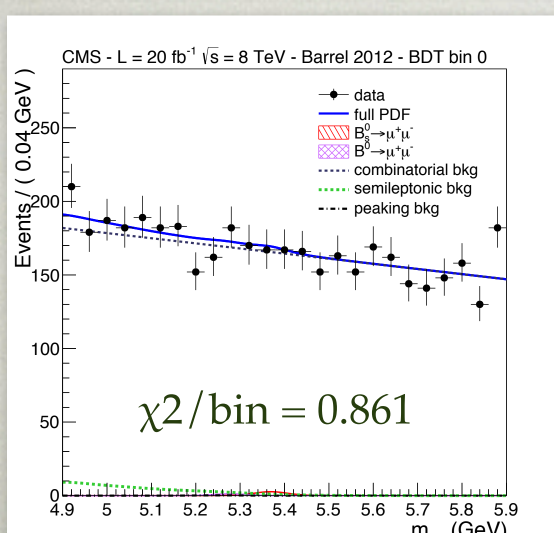


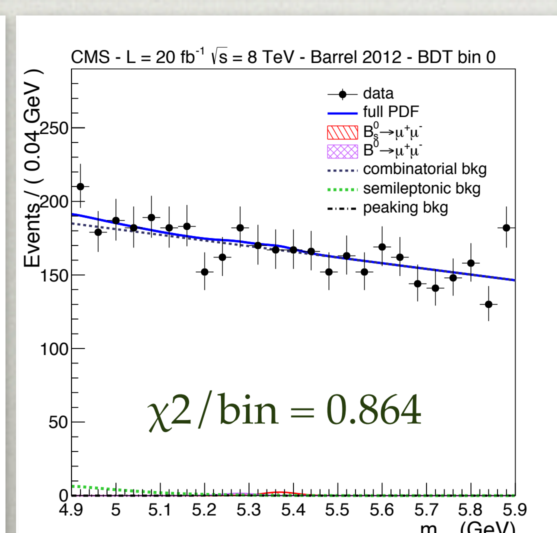
2011/endcap/BDT-1



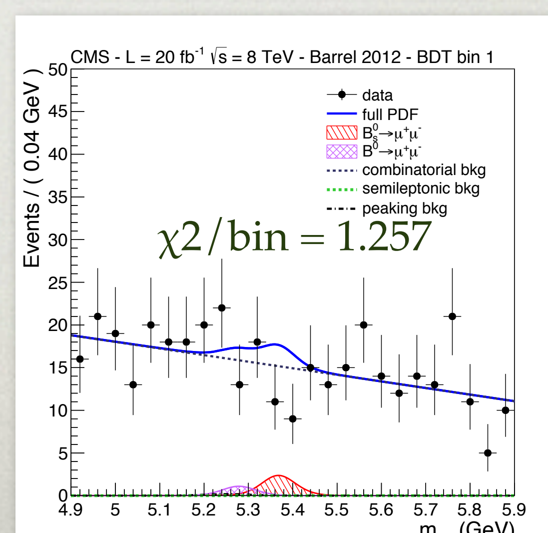


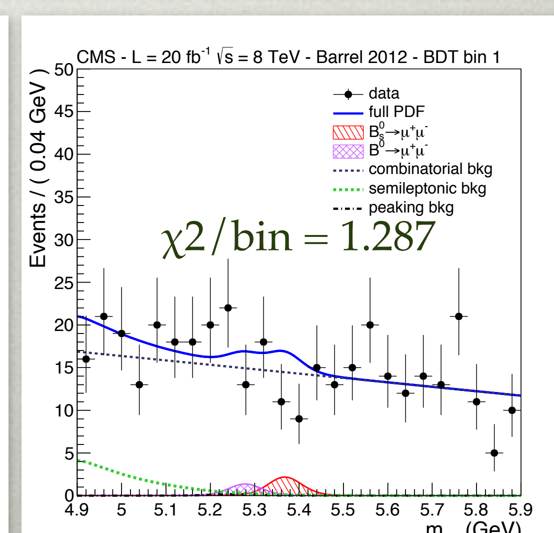
2012/barrel/BDT-0





2012/barrel/BDT-1

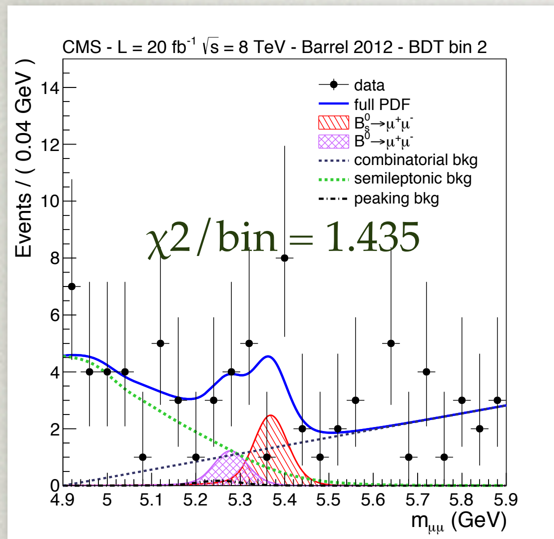


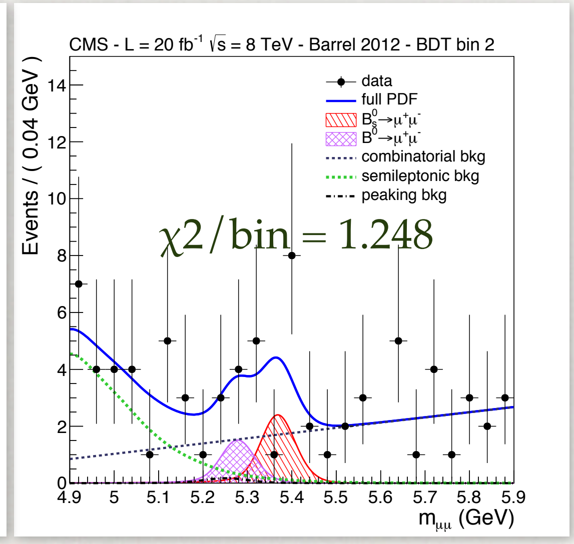


PROJECTIONS 1-ON-1

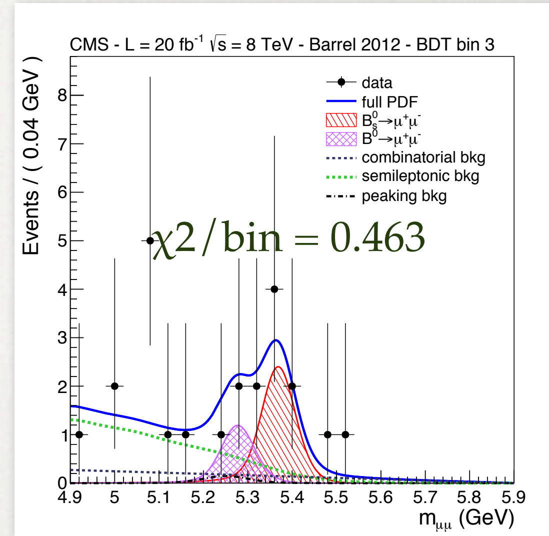
PUB NEW PUB NEW

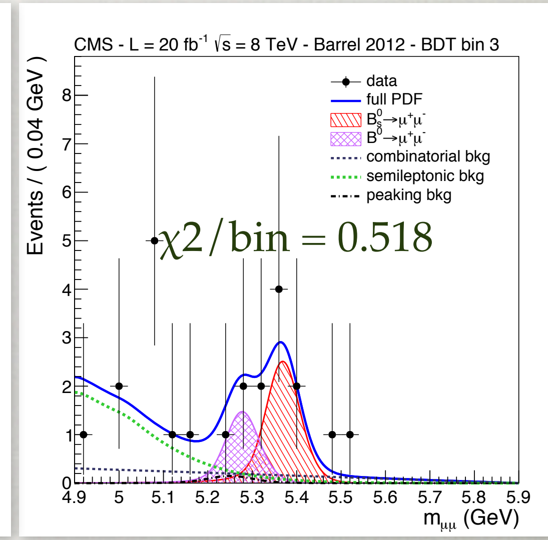
2012/barrel/BDT-2



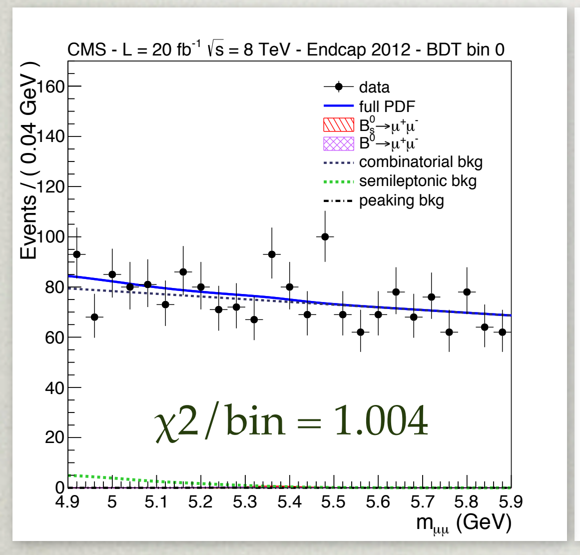


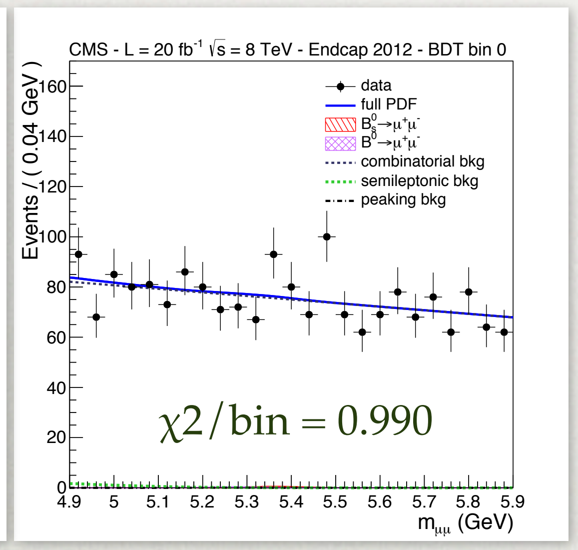
2012/barrel/BDT-3



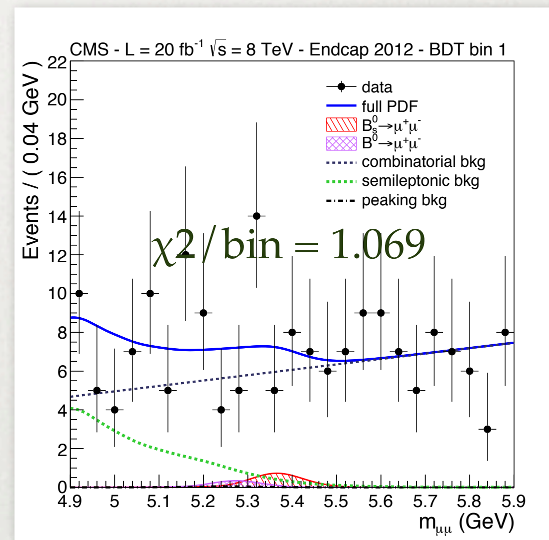


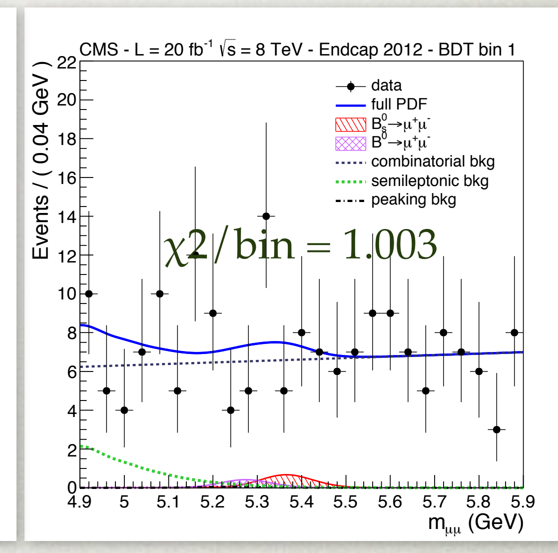
2012/endcap/BDT-0



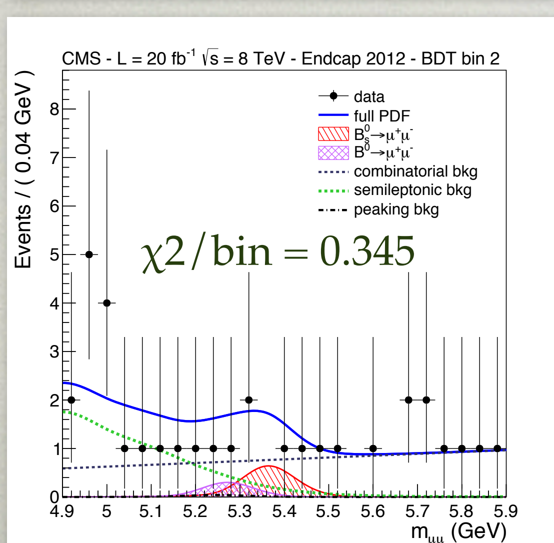


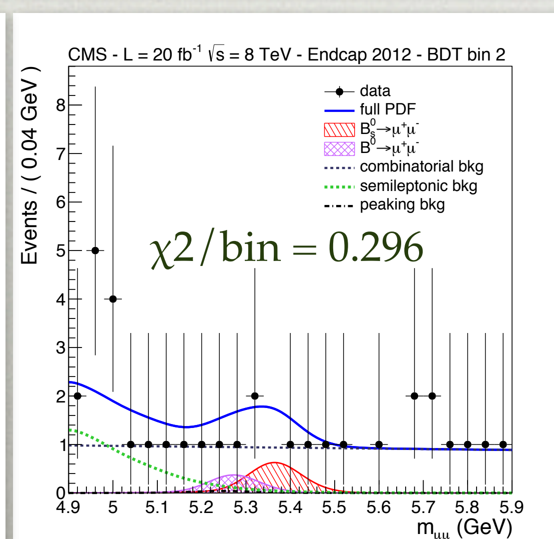
2012/endcap/BDT-1



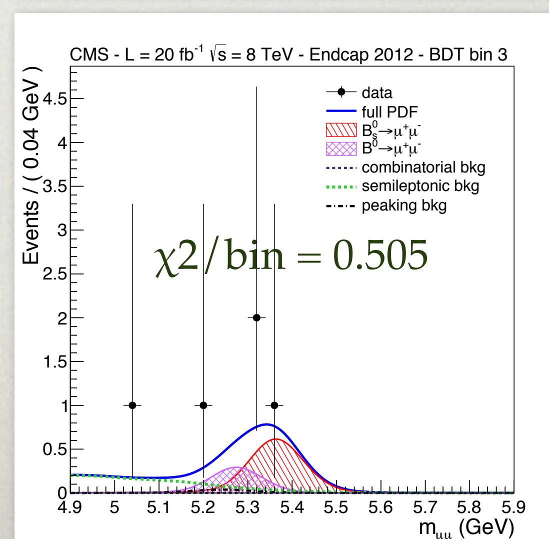


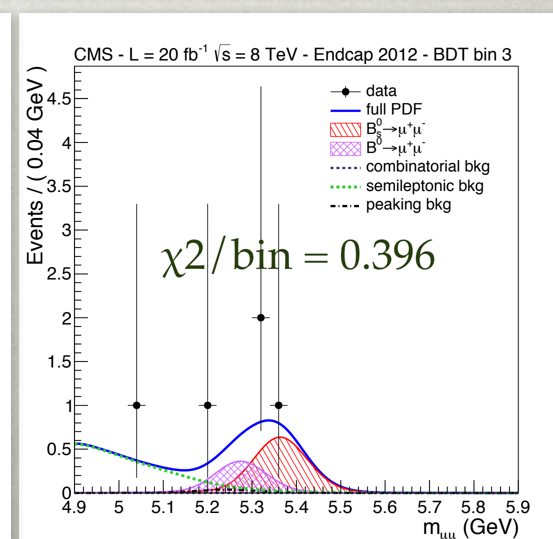
2012/endcap/BDT-2





2012/endcap/BDT-3





In preparation of publication

- Combined fit results
- Likelihood profile scan
- Feldman-Cousins scan
- Combined mass plot

THE COMBINED FITTER

- The setup:
  - Loading the RooWorkspace (containing the data & PDF) from two experiments.
  - Construct **a global likelihood or a global PDF** with common fitting parameters and nuisance parameters if applicable:
    1) B_s and B_d branching fractions
    2) f_s/f_u: CMS value is constrained to LHCb value with a 5% Gaussian to account for p_T-η dependence.
    3) $J/\psi\ K^+$ and $J/\psi \to \mu\mu$ branching fractions
    4) TH $B_{s(d)} \to \mu\mu$ branching fractions, for the ratio fits.
- Run on the common version of gcc 4.8 and root 5.34.10 with 1-core only (in order to avoid the random round-off error in multi-core mode).
- Total number of floated parameters (mostly the nuisance parameters) already reaches 152.

THE BEST FITTED VALUES

CMS	$BF(B_s) = (2.8\ {+0.9}/{-0.8}) \times 10^{-9}$ $BF(B_d) = (4.4\ {+2.2}/{-1.9}) \times 10^{-10}$
LHCb	$BF(B_s) = (2.7\ {+1.1}/{-0.9}) \times 10^{-9}$ $BF(B_d) = (3.3\ {+2.4}/{-2.1}) \times 10^{-10}$

Remark: individual results will not be shown in the paper

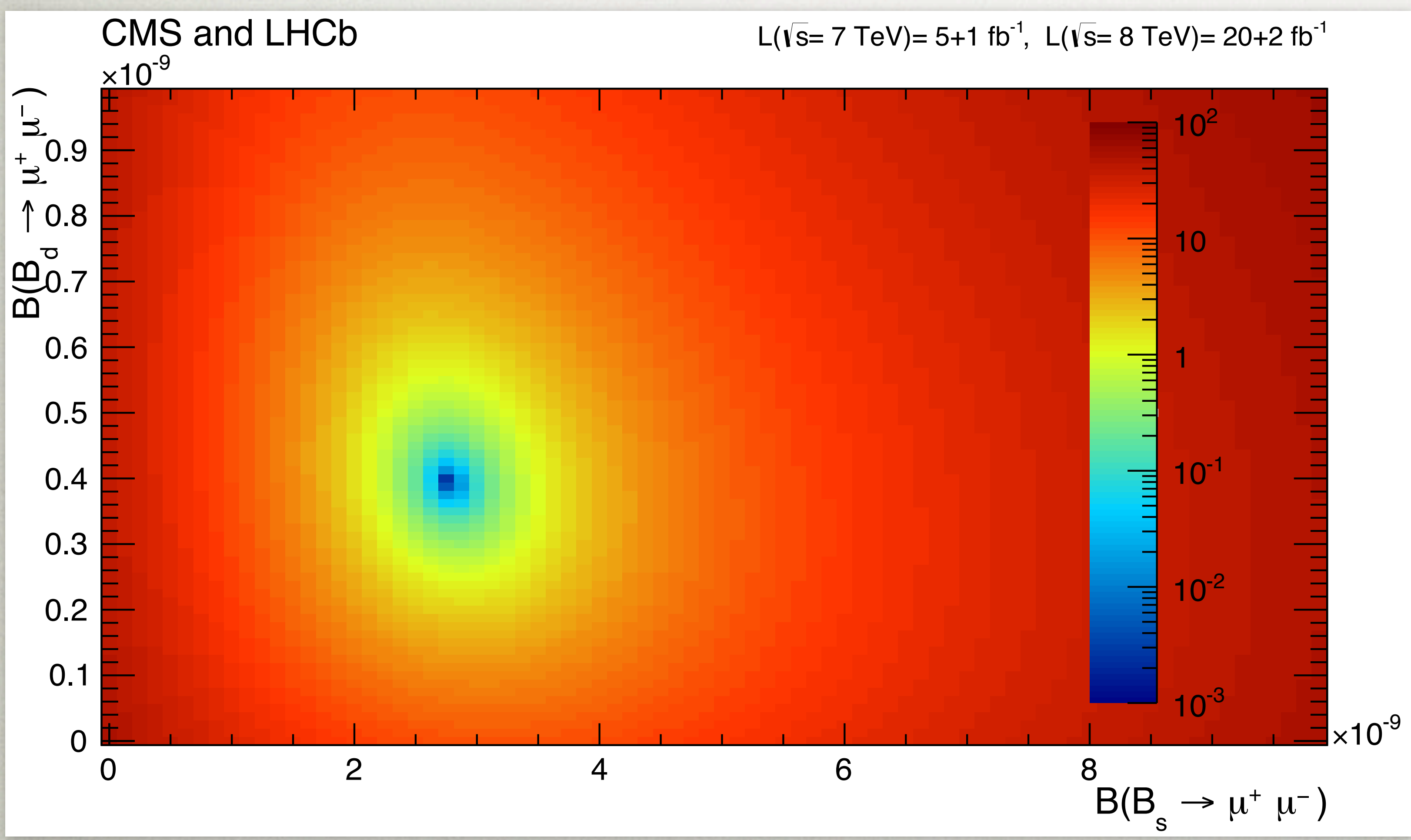
COMBINED	$BF(B_s) = (2.8\ {+0.7}/{-0.6}) \times 10^{-9}$ $BF(B_d) = (3.9\ {+1.6}/{-1.4}) \times 10^{-10}$
RATIO TO SM	$S(B_s) = 0.76\ {+0.20}/{-0.18}$ $S(B_d) = 3.7\ {+1.6}/{-1.4}$
BF RATIO	$R = BF(B_d)/BF(B_s)$ $= 0.14\ {+0.08}/{-0.06}$
TH PRL 112, 101801 (2014)	$BF(B_s) = (3.65 \pm 0.23) \times 10^{-9}$ $BF(B_d) = (1.06 \pm 0.09) \times 10^{-10}$

LIKELIHOOD PROFILE SCAN

- The setup:
  - Based on the combined fitter described above.
  - Simple/classical way – calculate the values of minimized $-2\log(L)$ for each given branching fractions on the grid.
  - Re-run the MIGRAD (+MINOS) commands until the fit converges for each point.
  - Compare the output from CODE A [a RooSimultaneous implementation] and CODE B [a RooNLL+RooMinuit implementation].
- 2D scan over both $B_{s,d}\to\mu\mu$ branching fractions.
- 1D scan over $B_{s(d)}\to\mu\mu$ with $B_{d(s)}\to\mu\mu$ floated.
- Also scanning over the ratios to the SM branching fractions, as well as the ratio of the two branching fractions

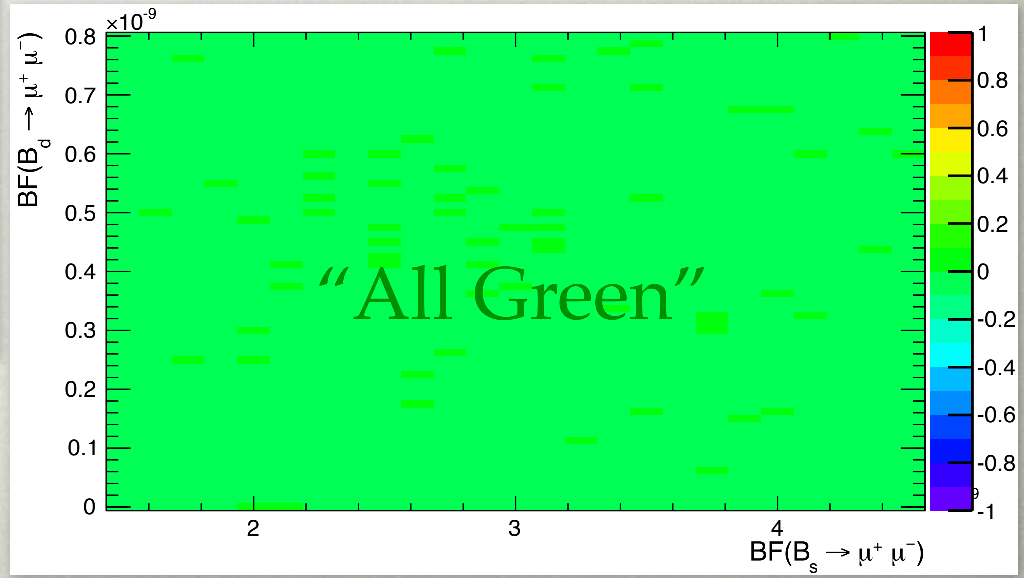
LIKELIHOOD SCANS

- 2D profile likelihood scan for BF(B_d) vs BF(B_s).
- Excellent agreement between the two implementations of fitter.



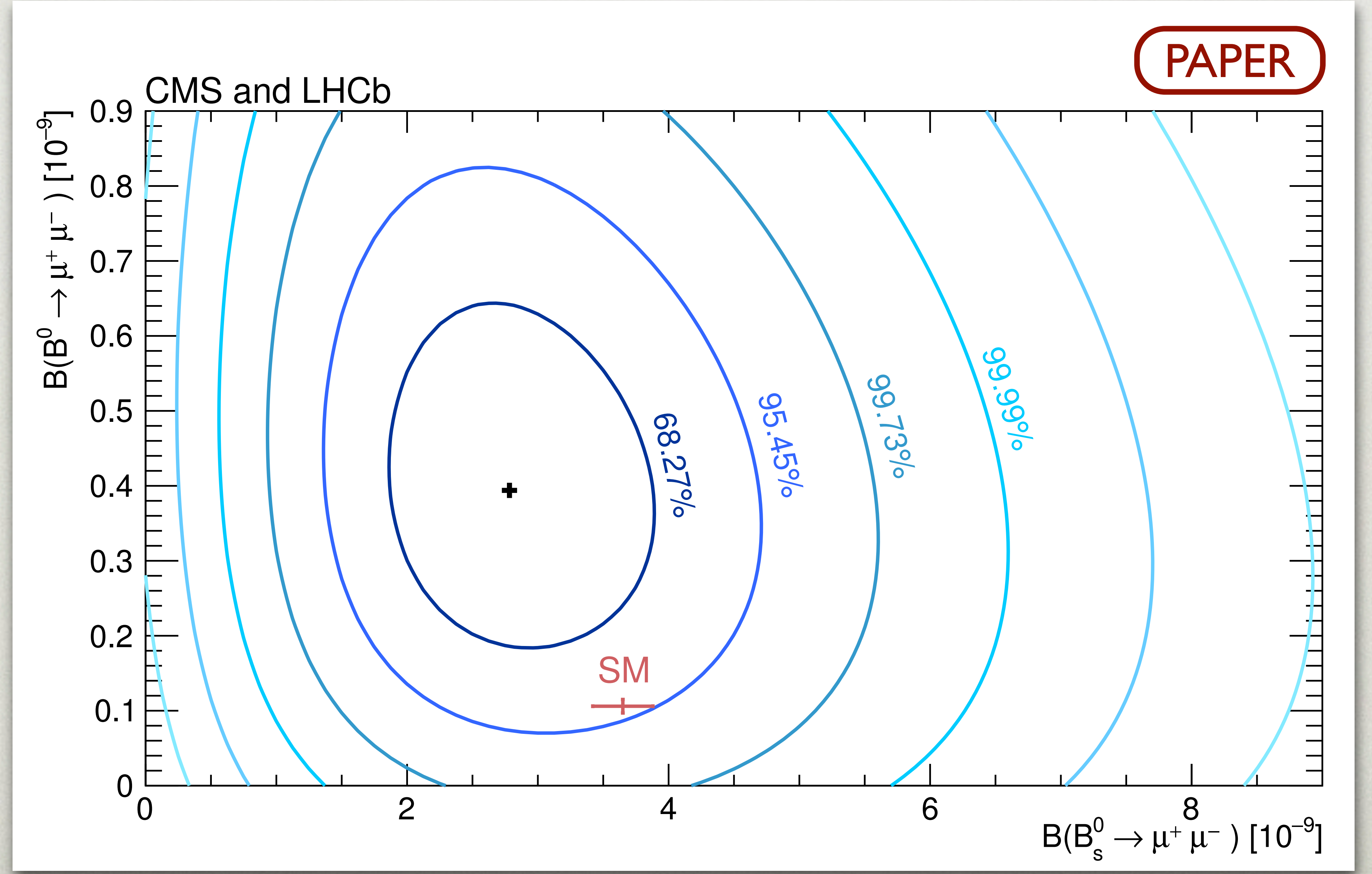
2D SCAN

The absolute difference between the log-likelihood values from two implementations.

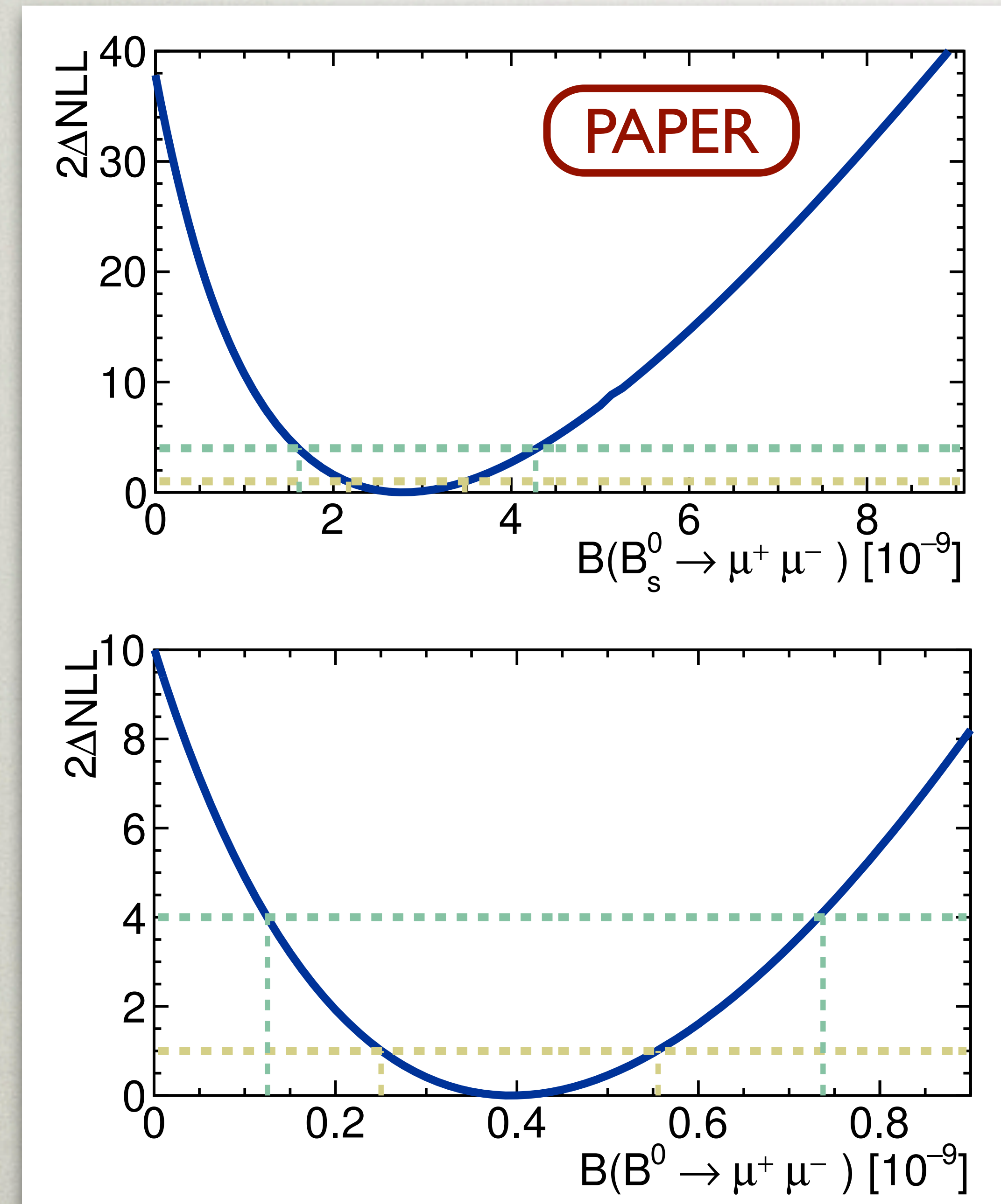


LIKELIHOOD SCANS

- The 2D contour plot for BF(B_d) versus BF(B_s):

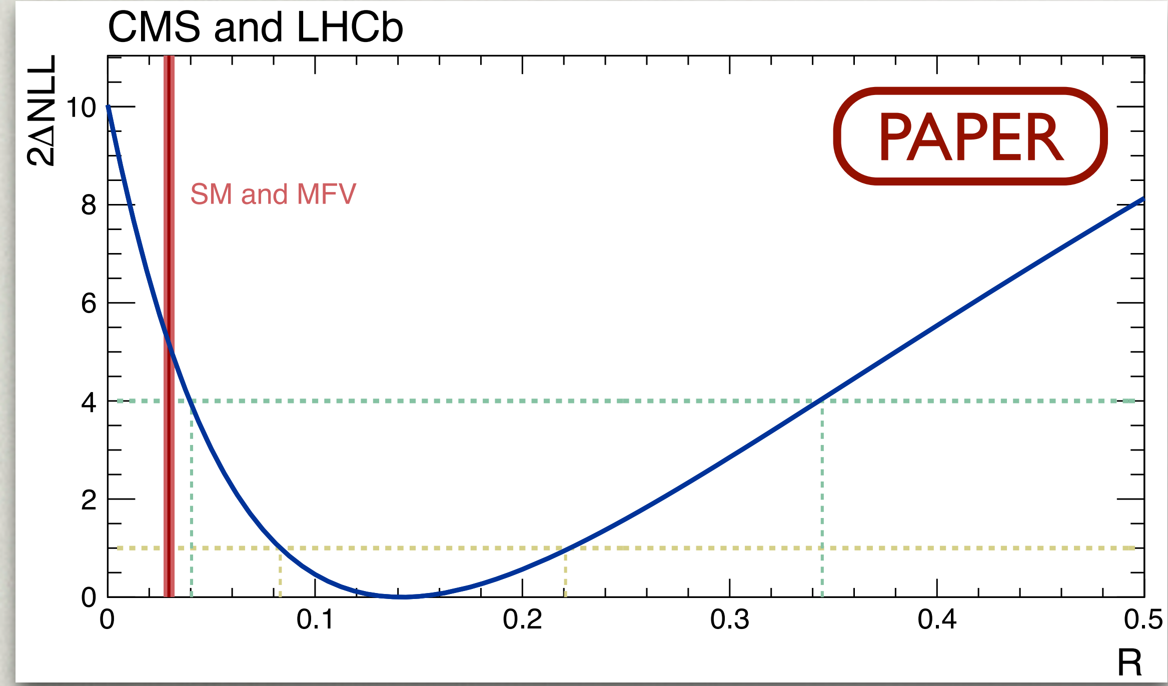


LIKELIHOOD SCANS



- 1D likelihood scan over BF(B_s) is performed with profiling BF(B_d), and vice versa.
- Significance given by likelihood scan:
  - for B_s channel:
    Away from zero: **6.2σ**
    compatibility with SM: **1.2σ**
  - as the search for B_d:
    Away from zero: **3.2σ**
    *(Feldman-Cousins toys: **3.0σ**, see following slides.)*
    compatibility with SM: **2.2σ**

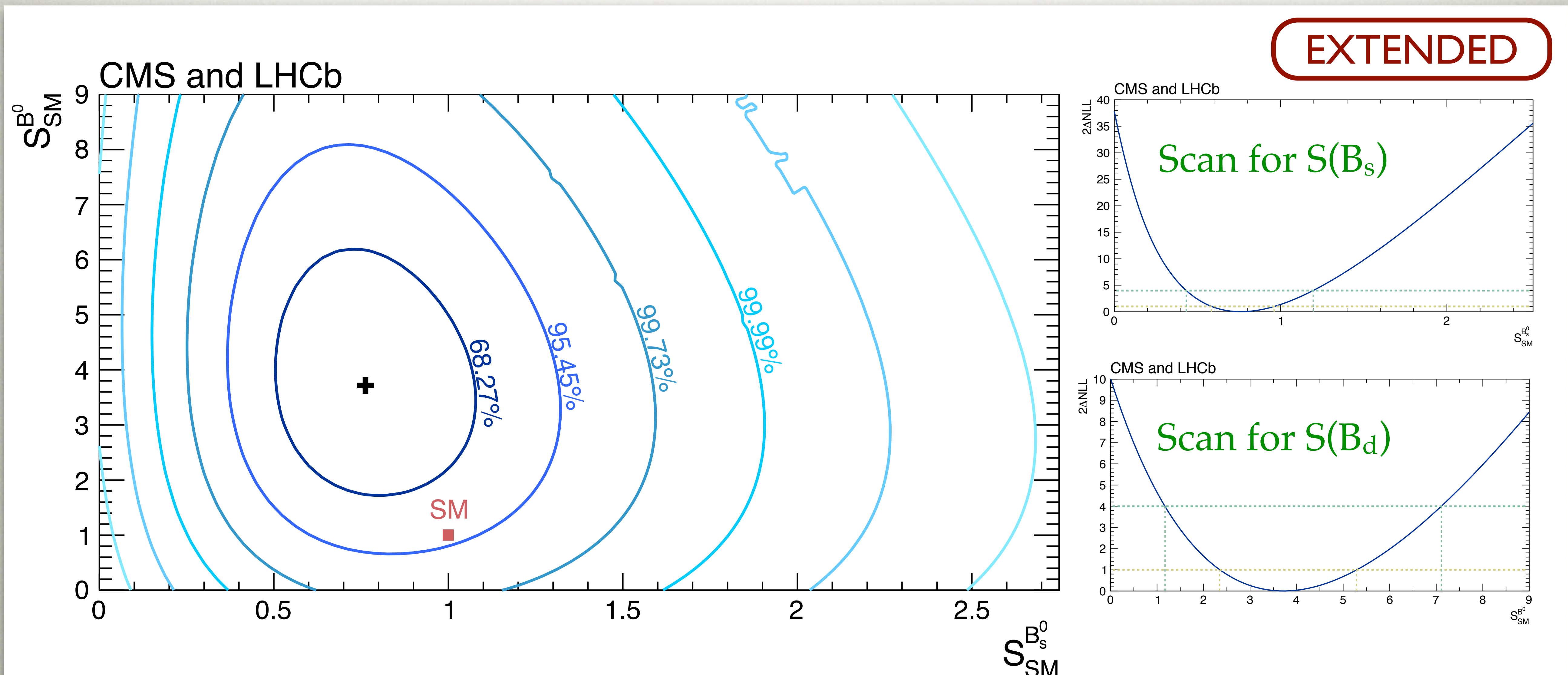
LIKELIHOOD SCANS



- The scan over the ratio of branching fraction $R = BF(B_d) \ / \ BF(B_s)$ with profiling $BF(B_s)$.
- Compatibility with SM: 2.3σ

LIKELIHOOD SCANS

- Scans for **S(B$_s$)** and **S(B$_d$)**, the ratios to the SM branching fractions.
- Note the SM branching fractions are also treated as nuisance parameters.



FELDMAN-COUSINS INTERVAL

- Likelihood scans are only an asymptotic limit, which may not be valid near physical boundaries, e.g. BF ≥ 0.
  - Cross-check confidence intervals for BF(B_d) using the **Feldman-Cousins procedure**.
  - Corresponding toys also provide a estimate of the significance.
- Feldman-Cousins procedure uses the **likelihood ratio** $\mathcal{L}(x|\mu)/\mathcal{L}(x|\hat{\mu})$ as a test statistic to determine the confidence level for a given parameter of interest at fixed value μ, with respect to observed data x and the best fit value for parameter of interest $\hat{\mu}$.
- Key point: for a given sized confidence interval (eg. 68% or 95%), this procedure automatically chooses between one-sided confidence limits and two-sided confidence intervals, with proper frequentist coverage.

"FULLY FREQUENTIST" FELDMAN COUSINS WITH NUISANCES

- In the presence of additional nuisance parameters, the test statistic becomes the **profile likelihood ratio**

$$R = \mathcal{L}(\bar{x}|\mathcal{B}(B_d))/\mathcal{L}(\bar{x}|\hat{\mathcal{B}}(B_d))$$

  where nuisance parameters are profiled in the minimization of the likelihood.

- Since the test statistic distribution is not known a priori, sampling distributions are constructed by throwing **pseudo-data (toys)**.

“FULLY FREQUENTIST” FELDMAN COUSINS WITH NUISANCES

- Toys are generated using **“fully-frequentist” prescription**:

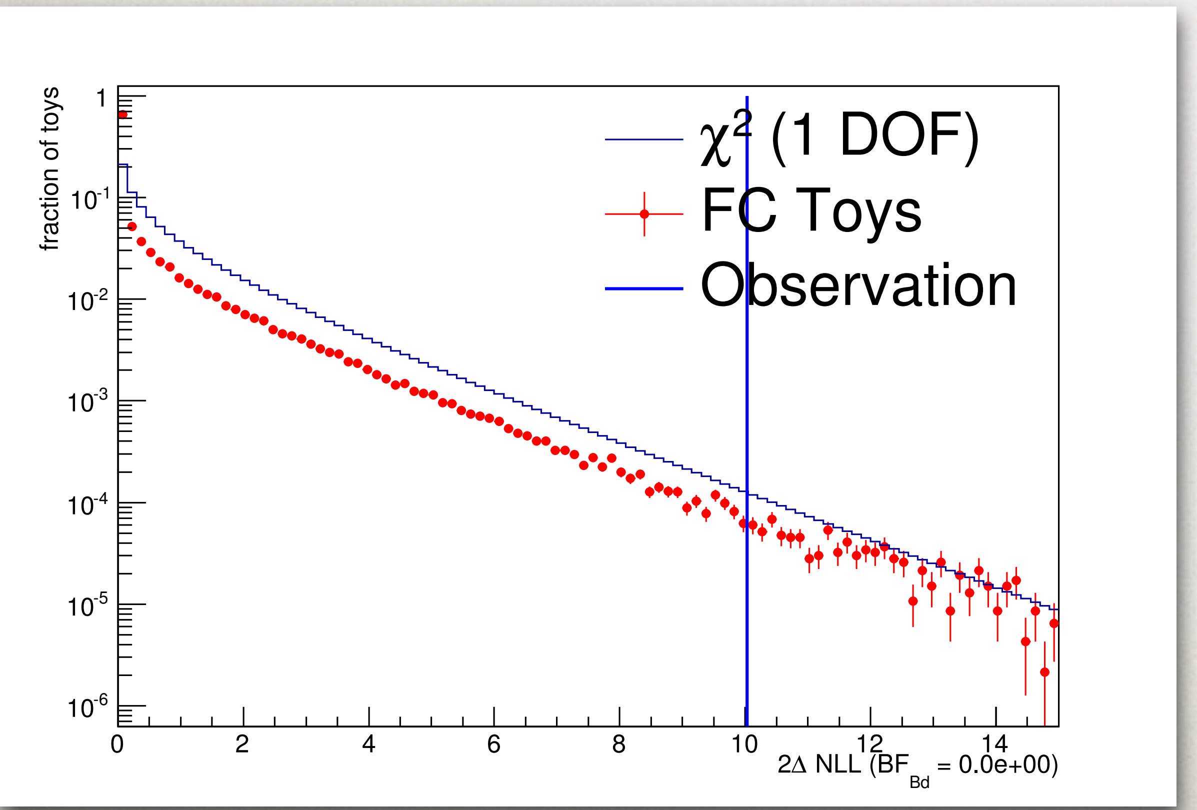
1) For a given value of BF(B_d) being tested, a fit is performed to the data with BF(B_d) fixed and nuisances [including BF(B_s)] profiled.
2) Best-fitted nuisances values are used to generate toy samples. The nuisances values are **not** randomized at the generation; but for the constrained nuisances, minimum of the constrained term is randomized around the best-fitted value.
3) For each toy, two fits are performed, one for BF(B_d) fixed to the test point, and a second with BF(B_d) profiled, the likelihood ratio $-2\Delta L$ used to build the sampling distribution for the test-statistic.

- For each test point of BF(B_d), the observed value of the test statistic in data is used to evaluate the confidence level from the sampling distribution.

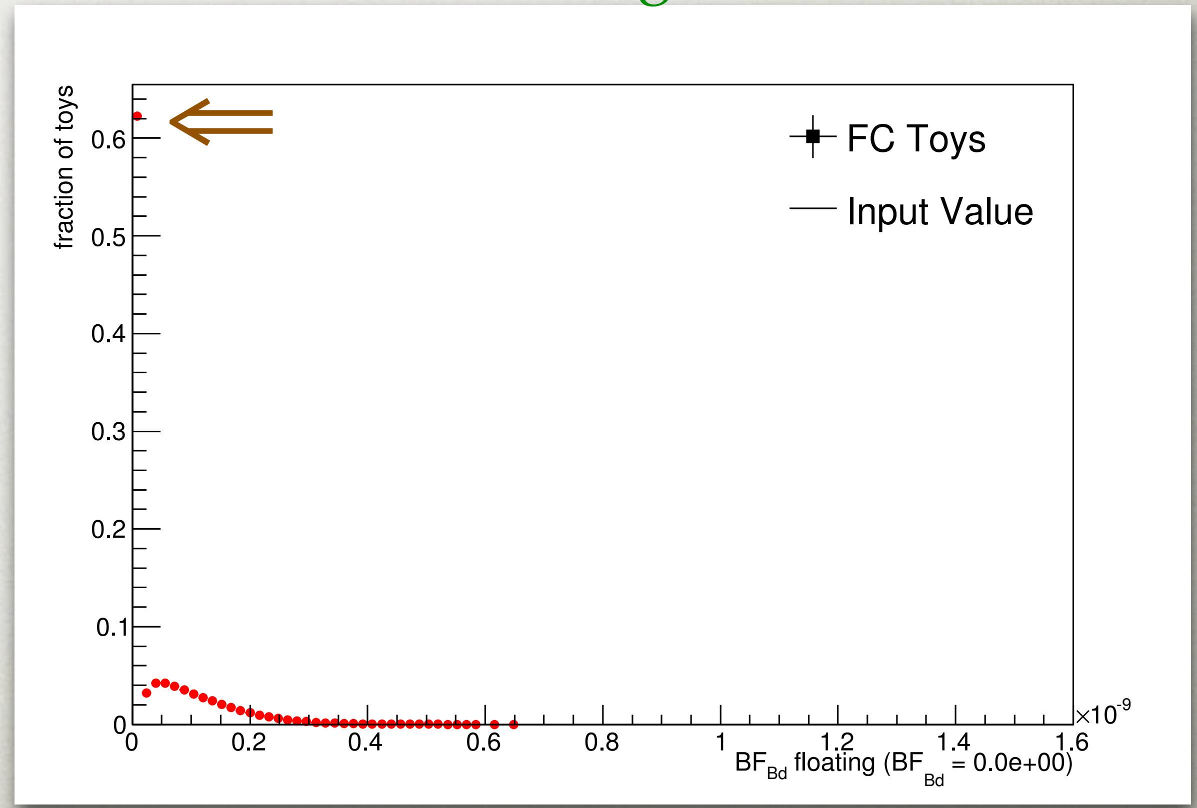
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = 0.0×10^{-10}

- For the scan point at BF(B_d)=0, the accumulation of toys at the boundary strictly produces a spike at 2ΔNLL = 0.

Test Statistics



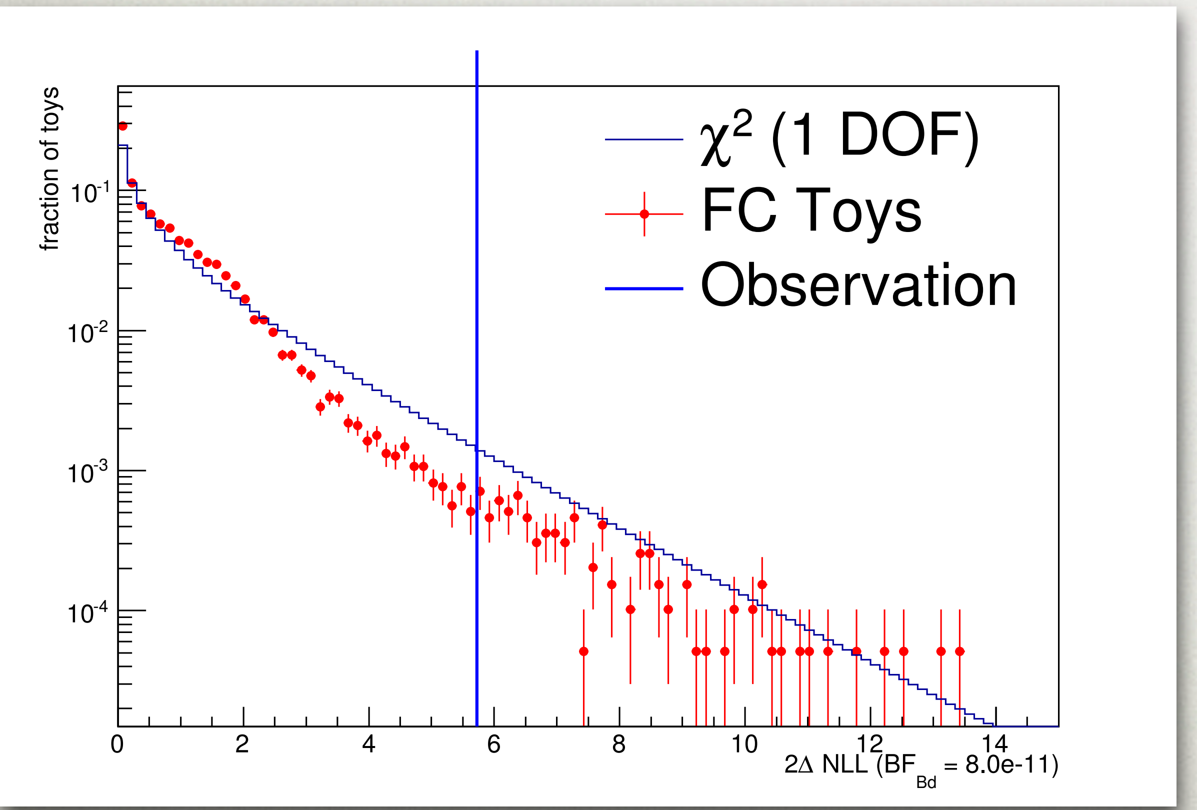
B_d Branching fraction



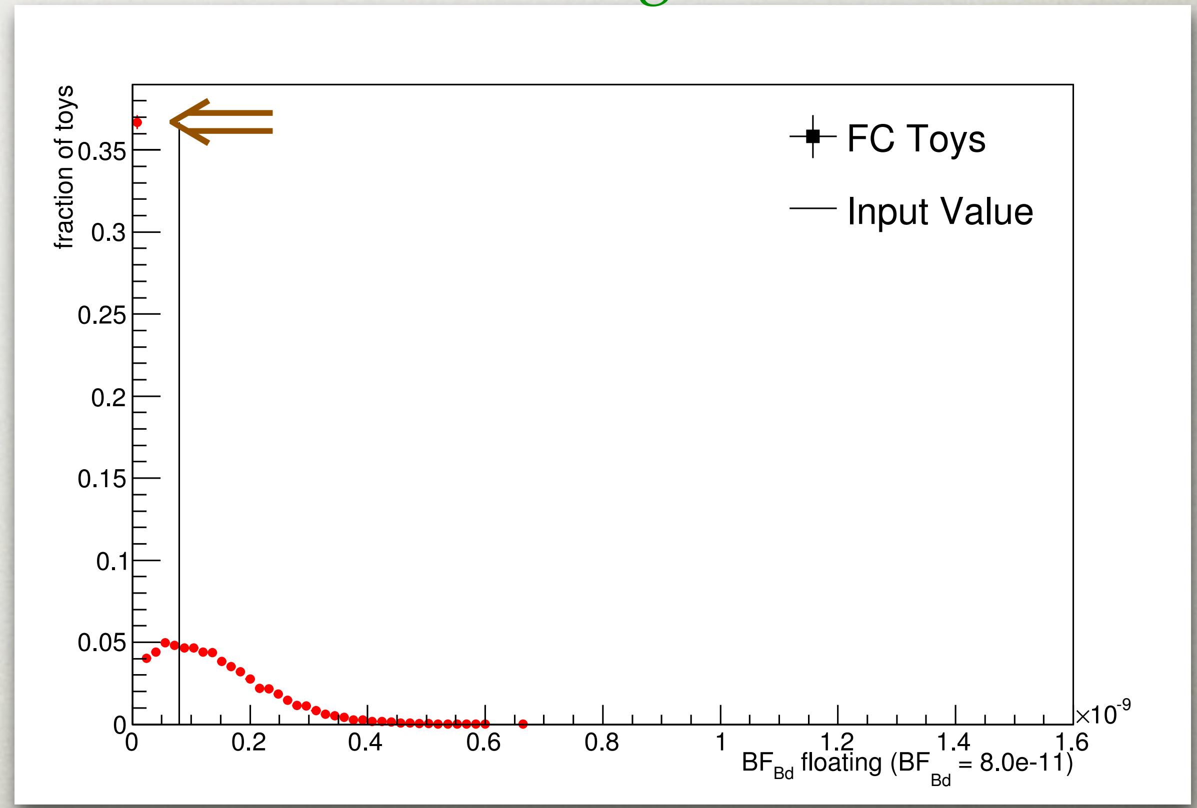
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = 0.8×10^{-10}

- For the scan point at BF(B_d) near 0, the accumulation of toys at the boundary produces a distorted test statistic distribution clustered at lower values with respect to the χ^2 distribution.

Test Statistics



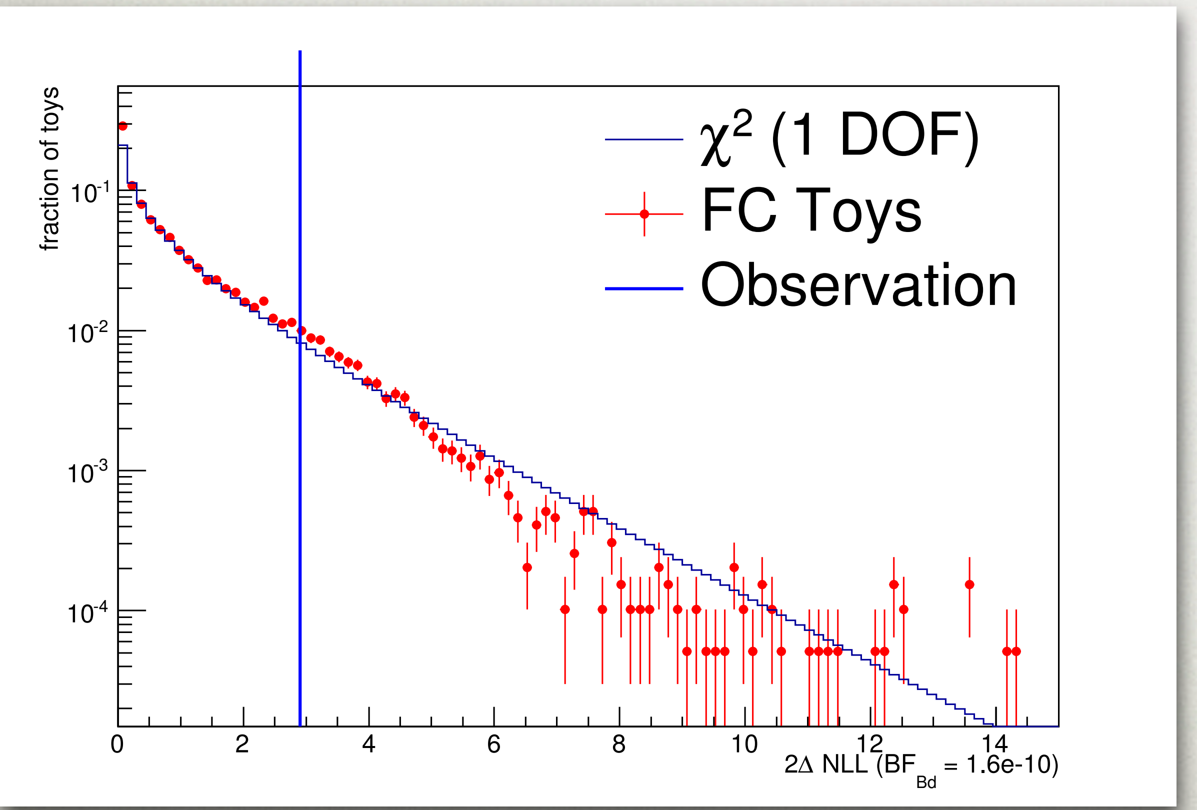
B_d Branching fraction



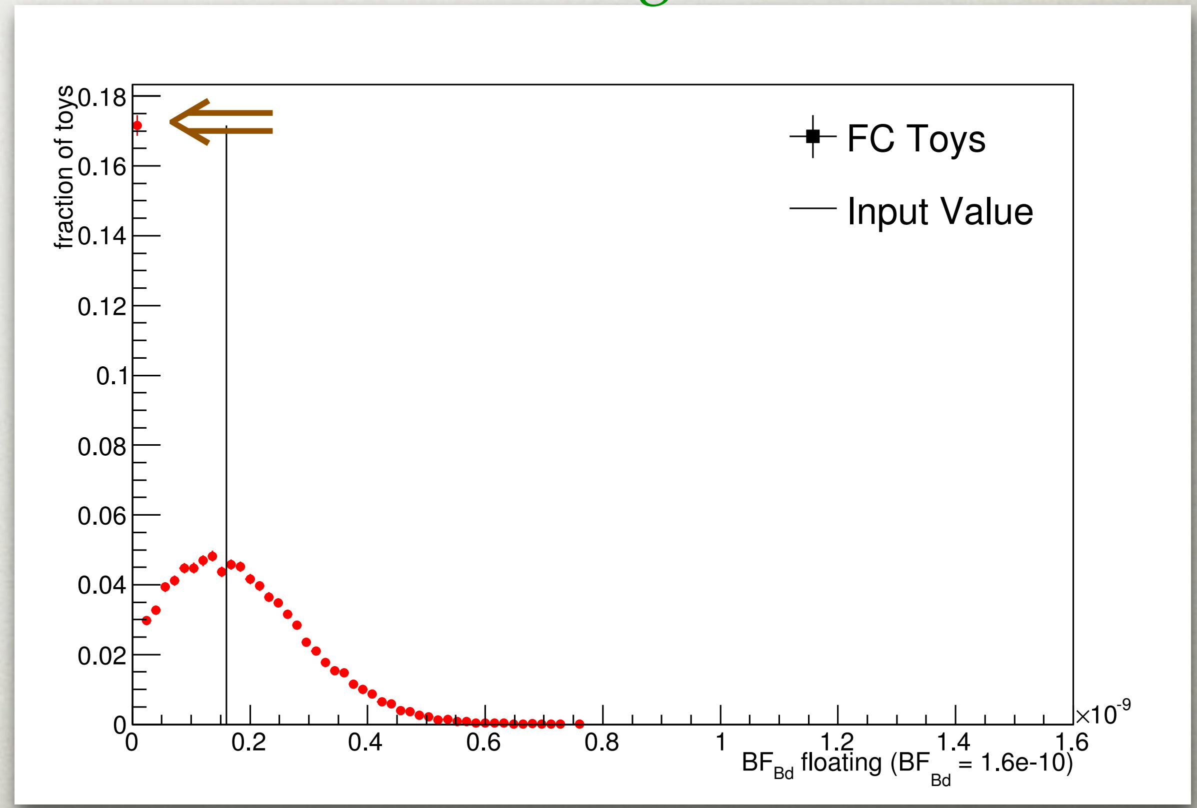
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = 1.6×10^{-10}

- For the scan points with BF(B_d) farther from 0, the remaining accumulation of toys at the boundary is far enough away from the input value that the test stat distribution is no longer much distorted with respect to the χ^2 distribution.

Test Statistics



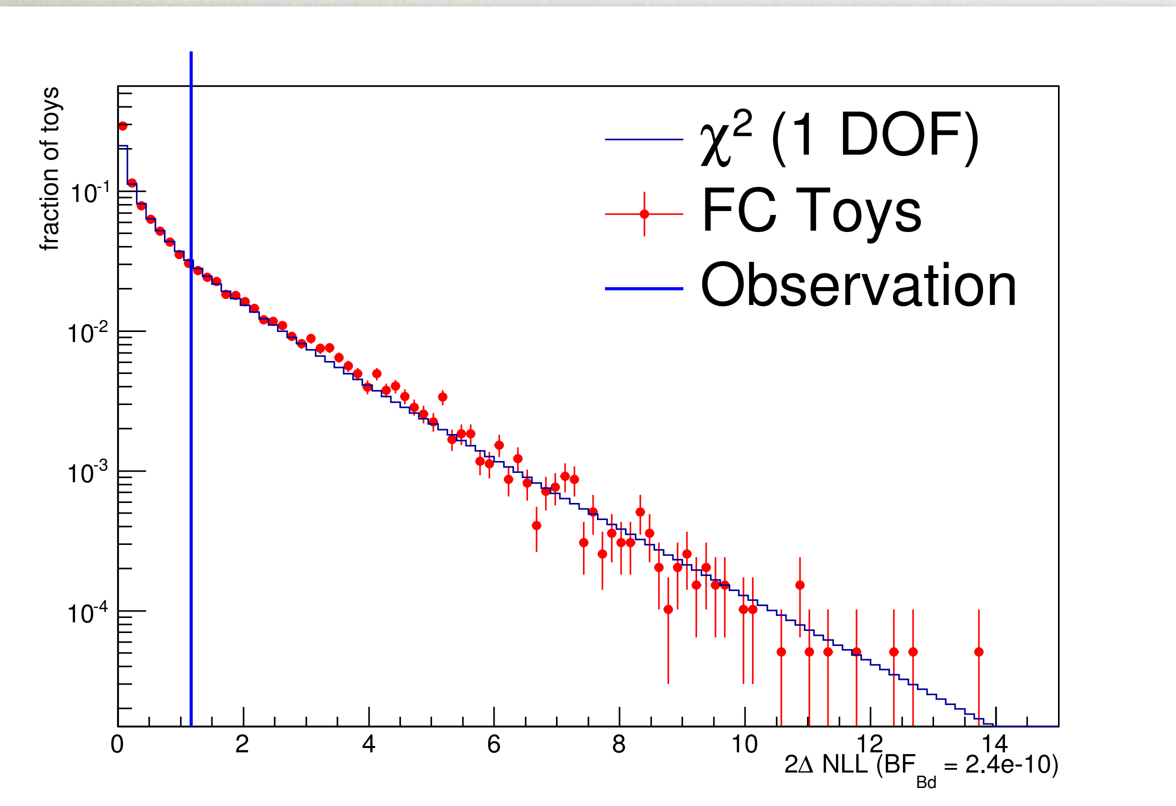
B_d Branching fraction



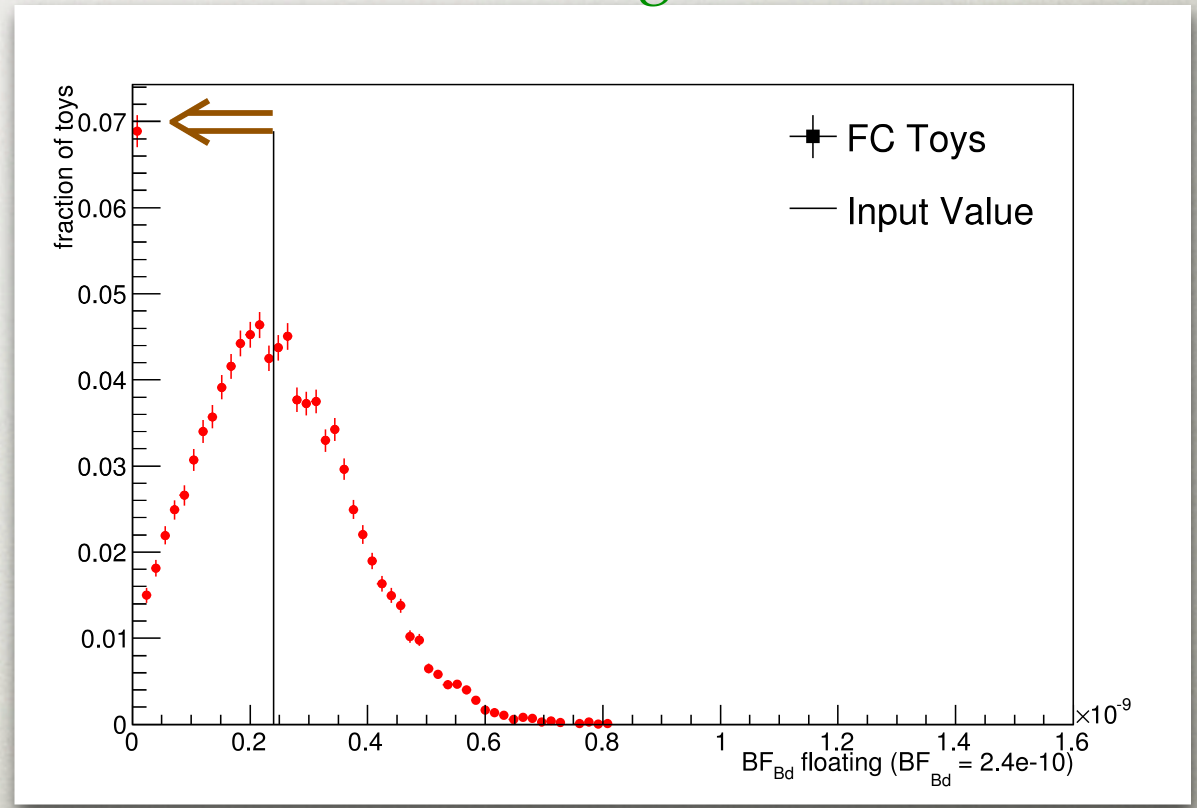
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = 2.4×10^{-10}

- For the scan points with BF(B_d) sufficiently far from 0, the fraction of toys accumulating at the boundary becomes smaller and the distribution follows closely the χ^2 distribution (ie. consistent with likelihood scan/Wilk's theorem).

Test Statistics



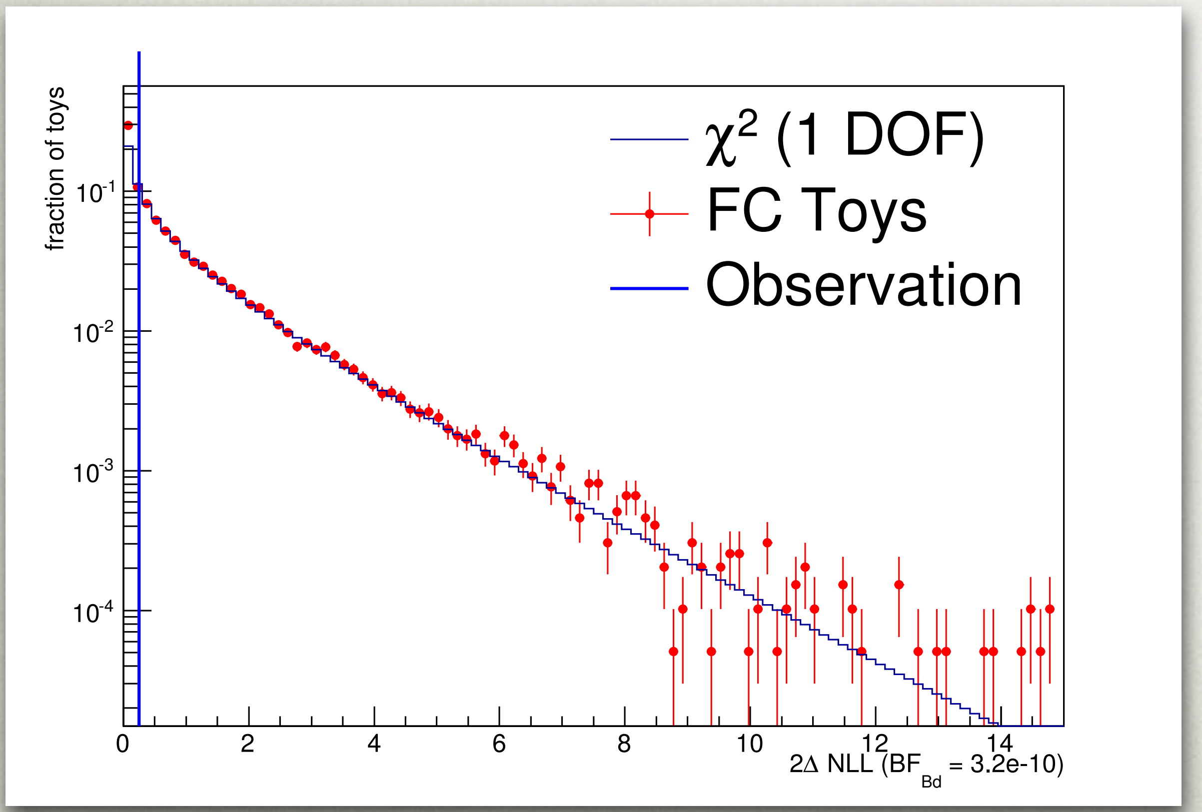
B_d Branching fraction



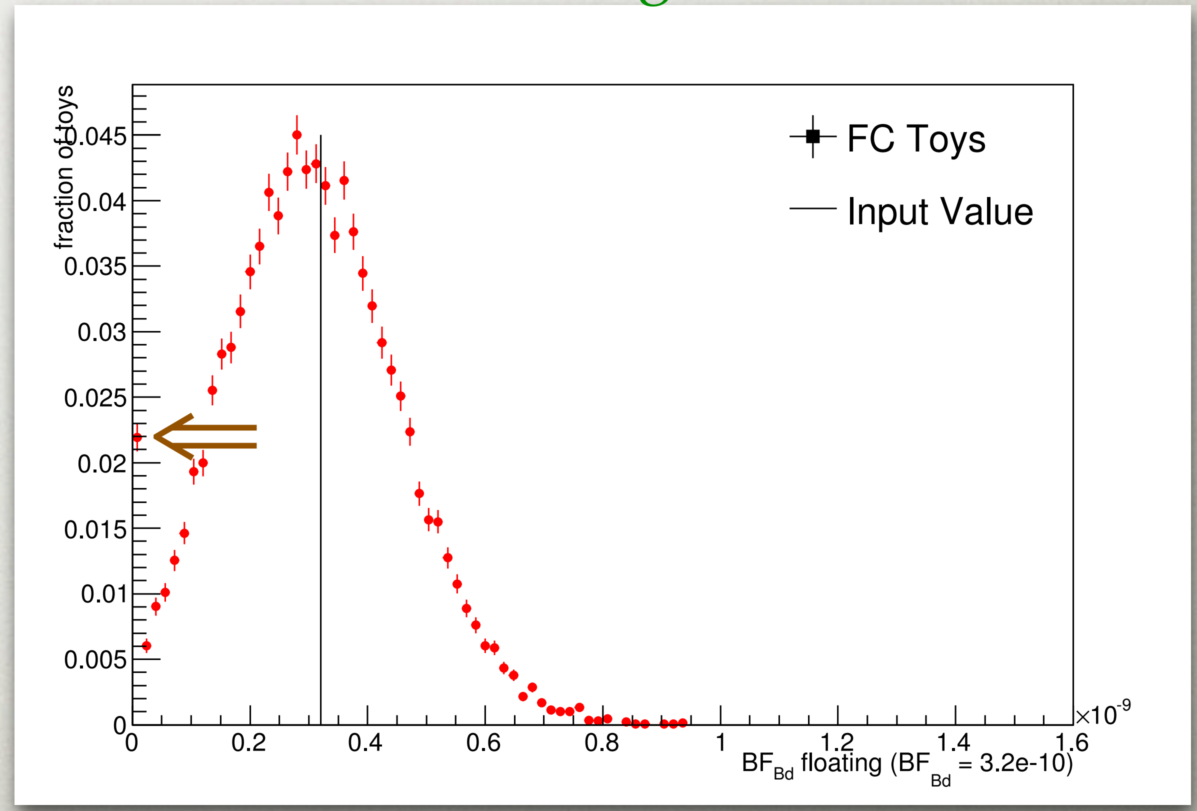
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = $3.2{\times}10^{-10}$

- For the scan points with BF(B_d) sufficiently far from 0, the fraction of toys accumulating at the boundary becomes smaller and the distribution follows closely the χ^2 distribution (ie. consistent with likelihood scan/Wilk's theorem).

Test Statistics



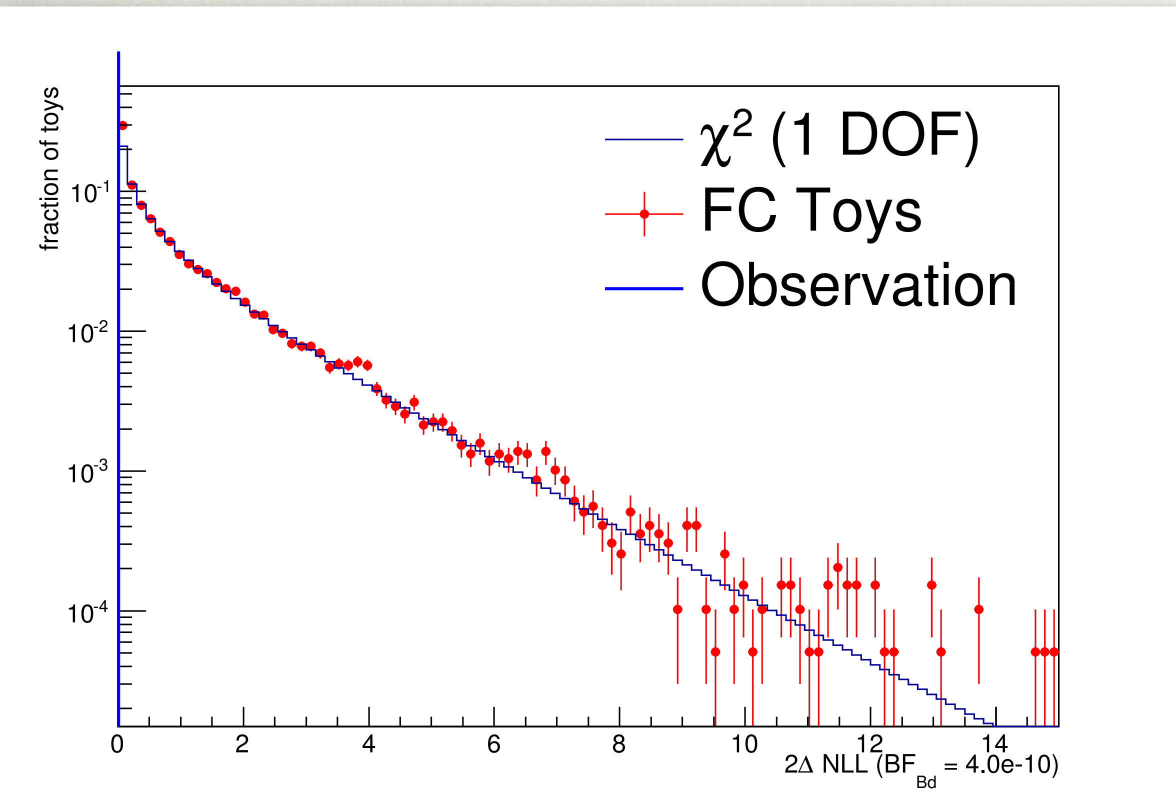
B_d Branching fraction



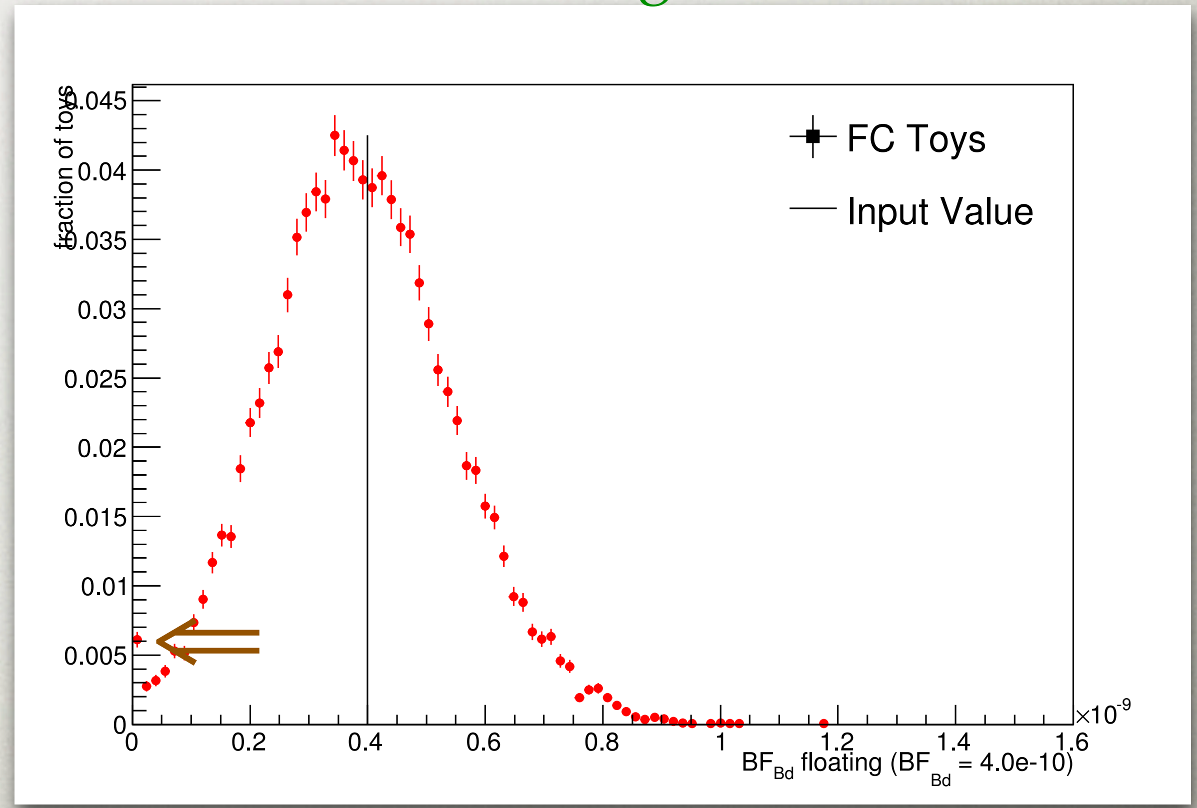
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = 4.0×10^{-10}

- For the scan points with BF(B_d) sufficiently far from 0, the fraction of toys accumulating at the boundary becomes smaller and the distribution follows closely the χ^2 distribution (ie. consistent with likelihood scan/Wilk's theorem).

Test Statistics



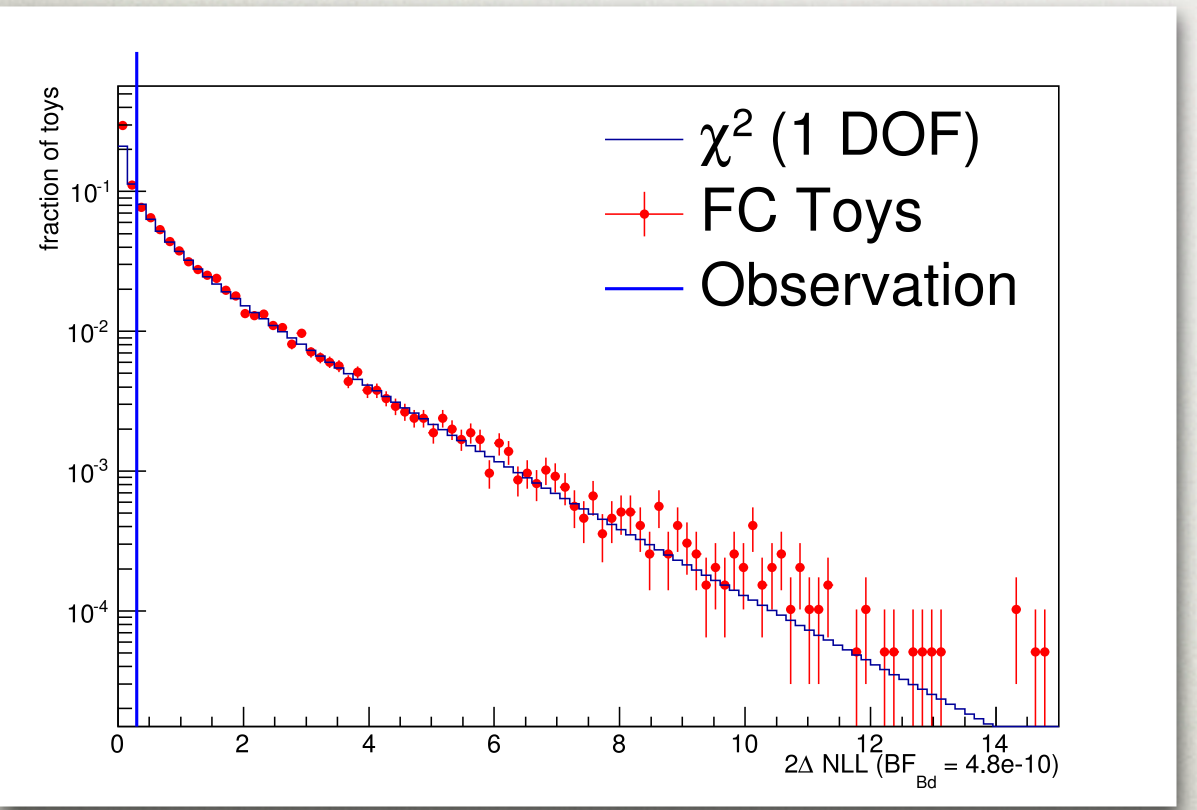
B_d Branching fraction



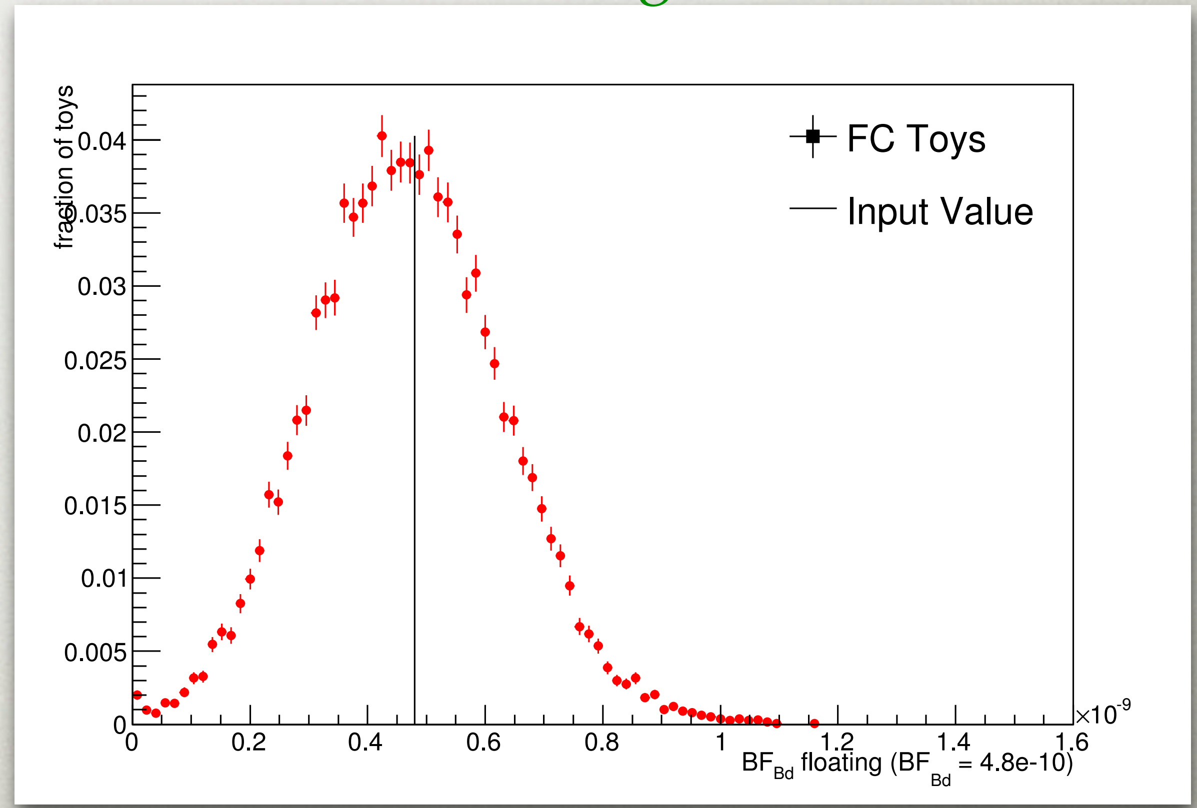
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = 4.8×10^{-10}

- For the scan points with BF(B_d) sufficiently far from 0, the fraction of toys accumulating at the boundary becomes smaller and the distribution follows closely the χ^2 distribution (ie. consistent with likelihood scan/Wilk's theorem).

Test Statistics



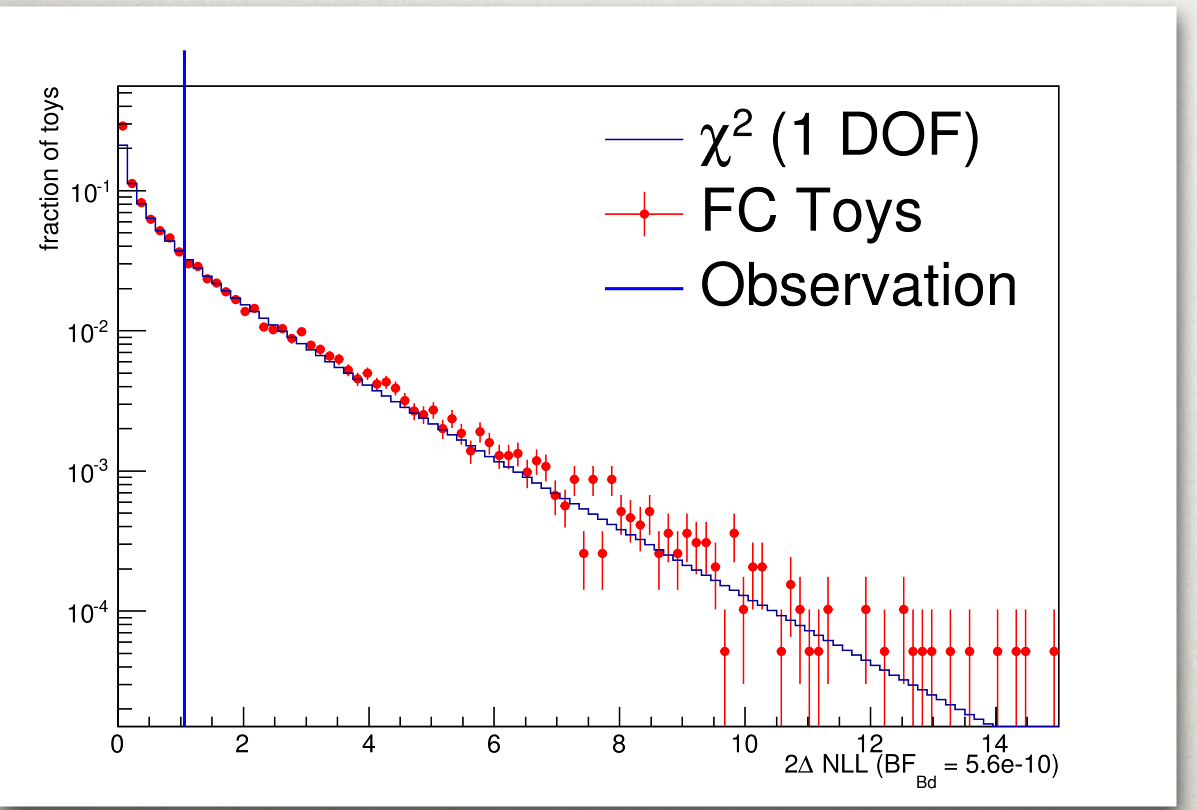
B_d Branching fraction



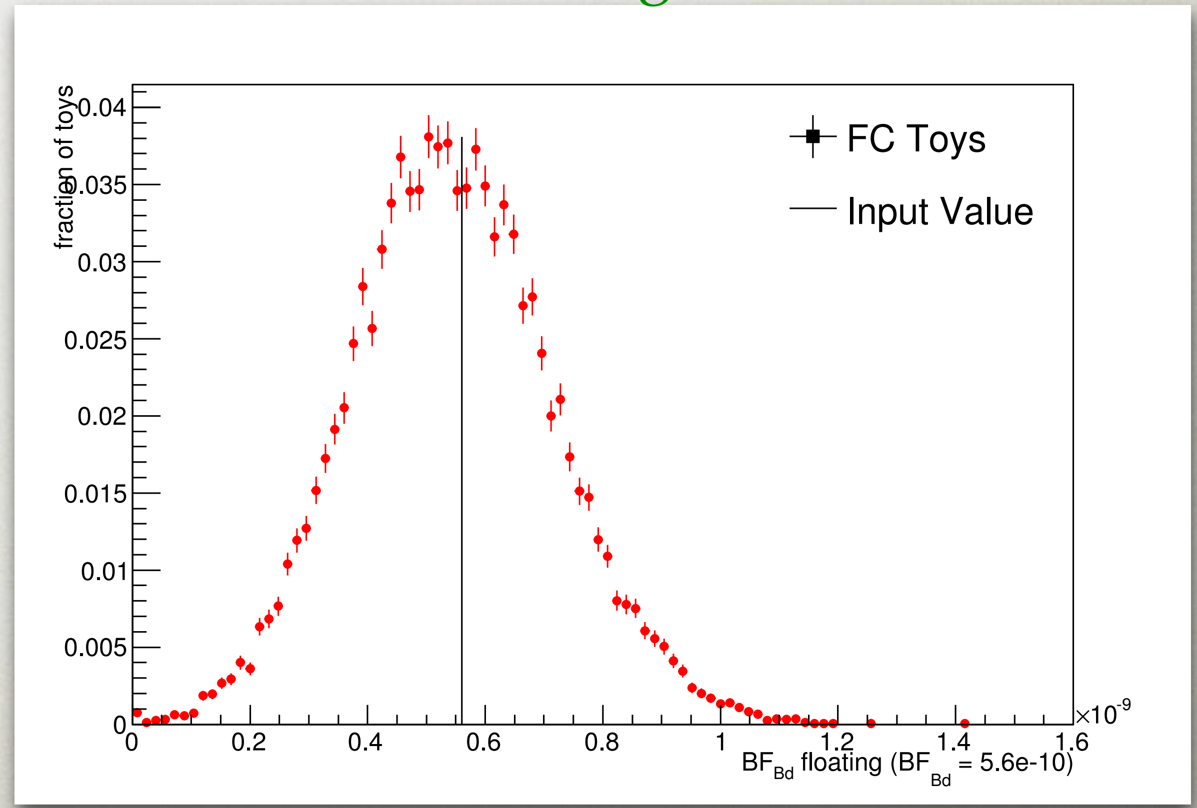
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = 5.6×10^{-10}

- For the scan points with BF(B_d) sufficiently far from 0, the fraction of toys accumulating at the boundary becomes smaller and the distribution follows closely the χ^2 distribution (ie. consistent with likelihood scan/Wilk's theorem).

Test Statistics



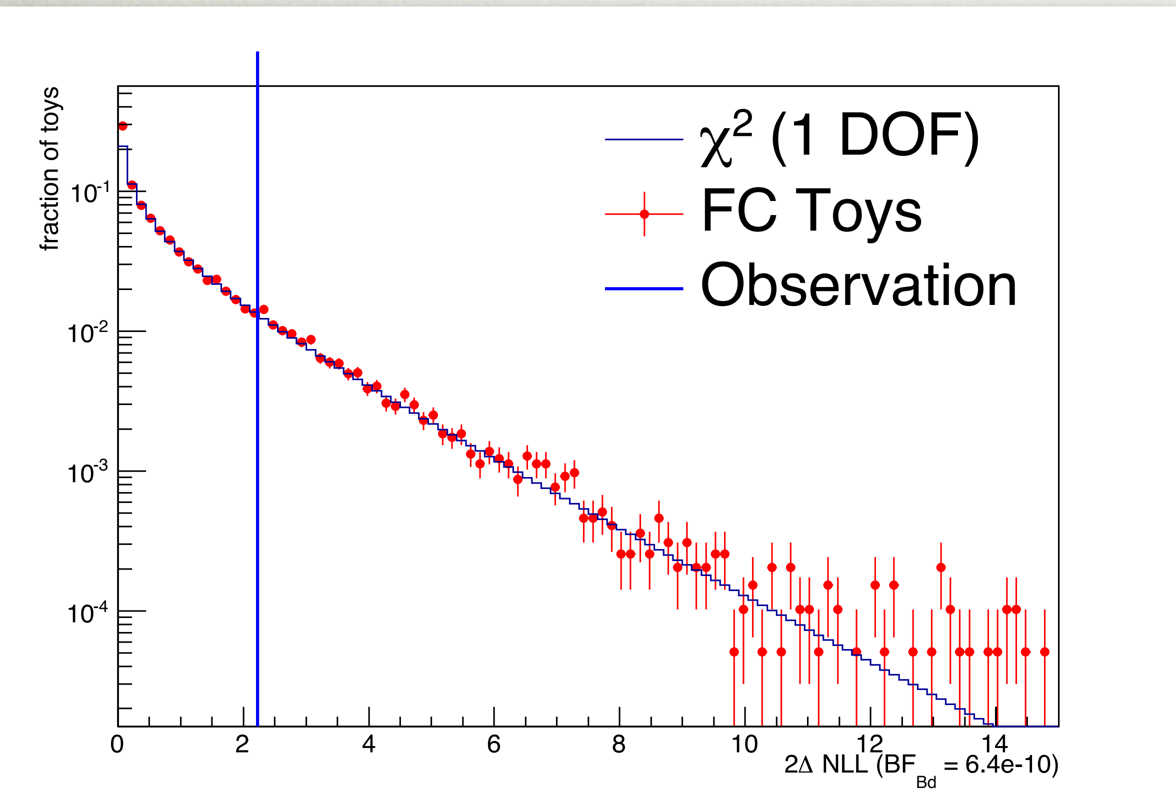
B_d Branching fraction



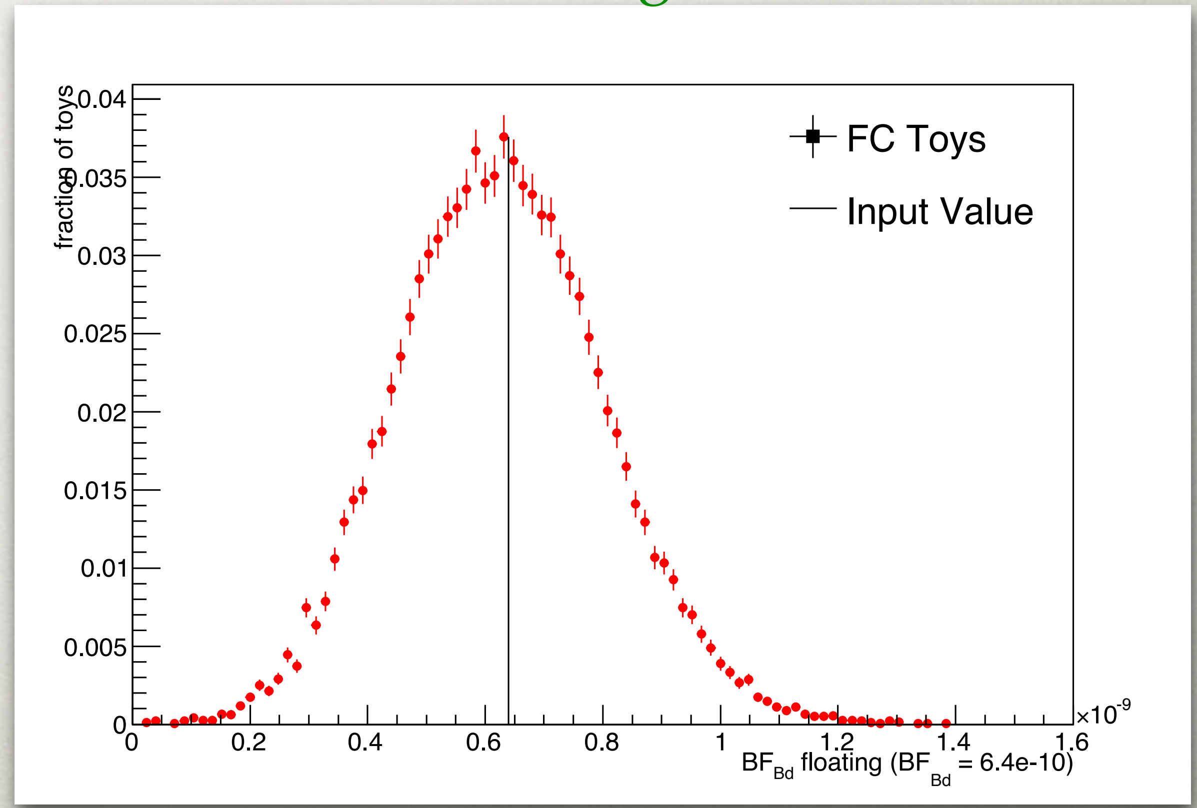
TEST STATISTIC & BF(B_d) DISTRIBUTIONS: BF(B_d) = 6.4×10^{-10}

- For the scan points with BF(B_d) sufficiently far from 0, the fraction of toys accumulating at the boundary becomes smaller and the distribution follows closely the χ^2 distribution (ie. consistent with likelihood scan/Wilk's theorem).

Test Statistics

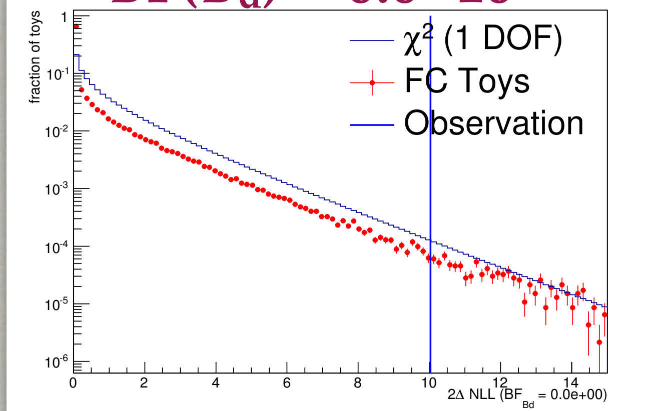


B_d Branching fraction

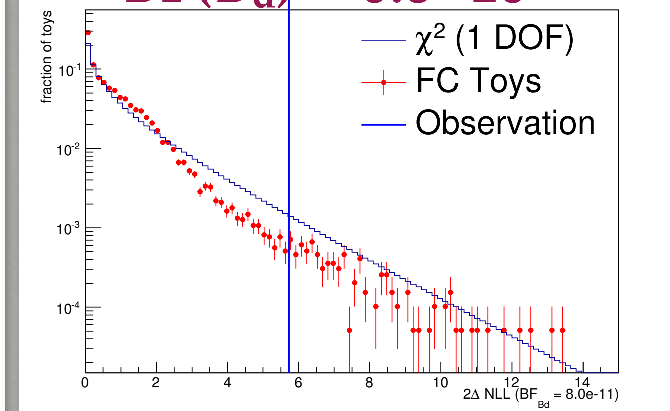


STEP-BY-STEP
TEST STATISTIC DISTRIBUTIONS

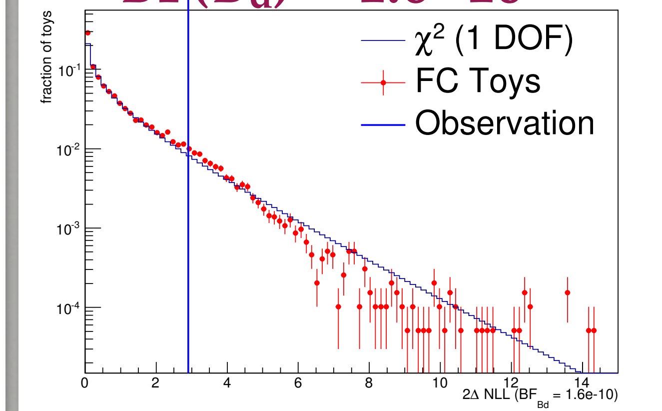
$BF(B_d) = 0.0\times10^{-10}$



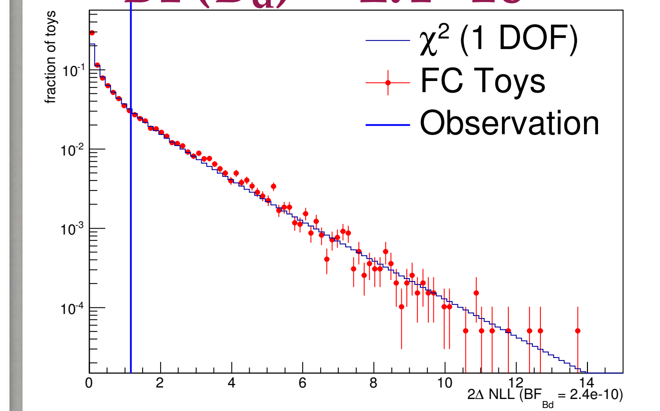
$BF(B_d) = 0.8\times10^{-10}$



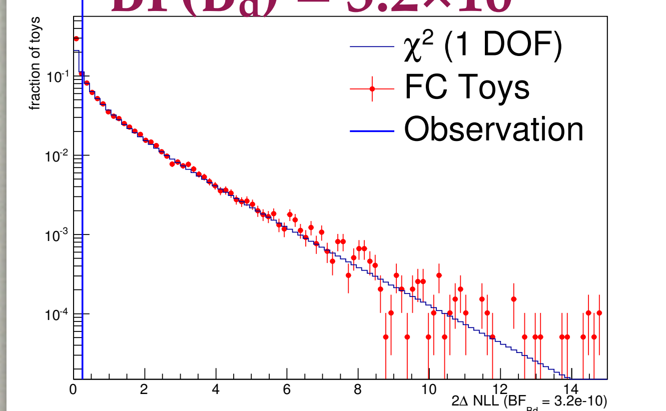
$BF(B_d) = 1.6\times10^{-10}$



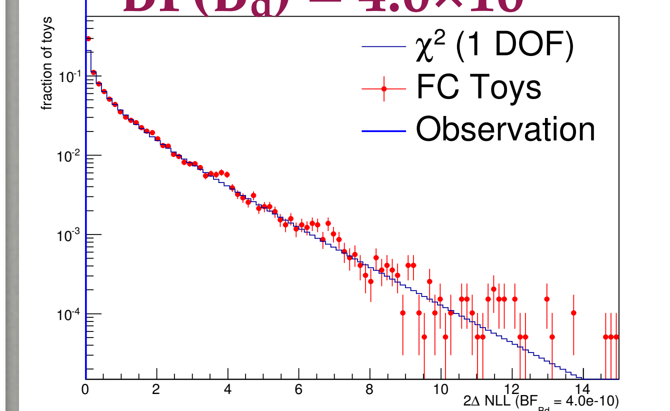
$BF(B_d) = 2.4\times10^{-10}$



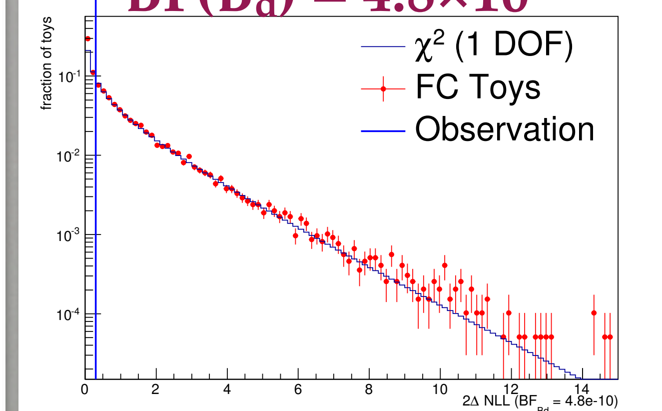
$BF(B_d) = 3.2\times10^{-10}$



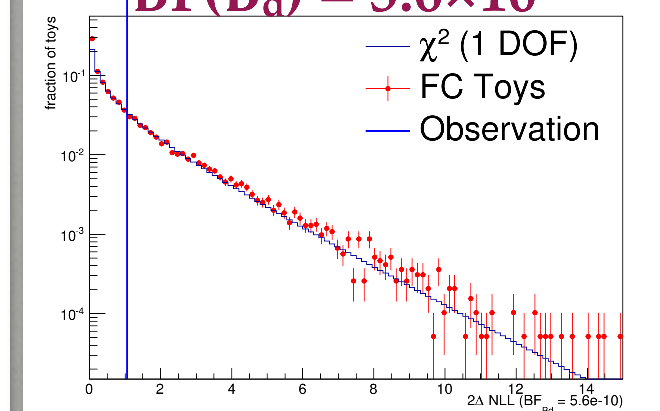
$BF(B_d) = 4.0\times10^{-10}$



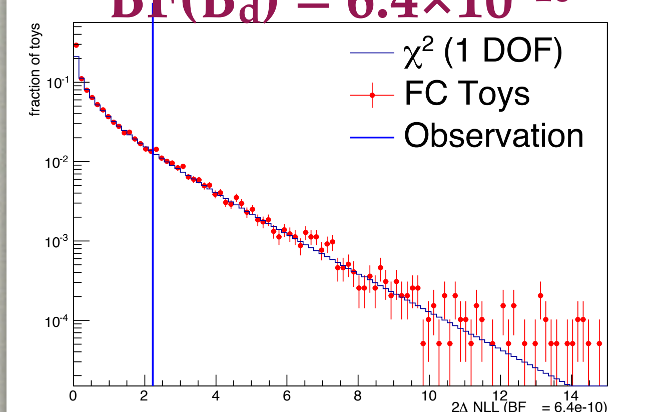
$BF(B_d) = 4.8\times10^{-10}$



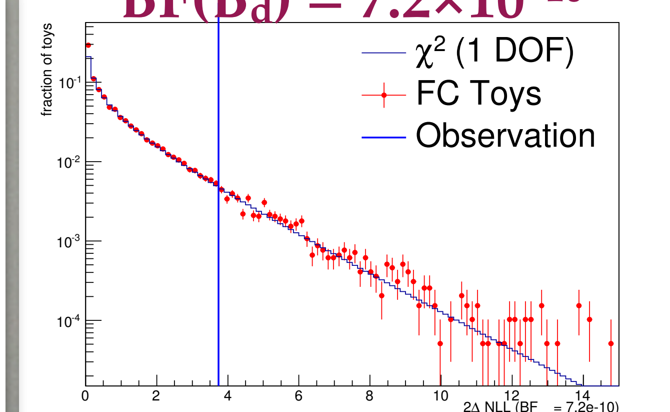
$BF(B_d) = 5.6\times10^{-10}$



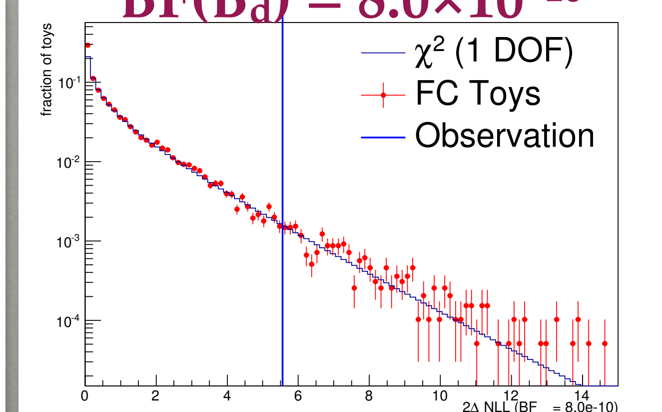
$BF(B_d) = 6.4\times10^{-10}$



$BF(B_d) = 7.2\times10^{-10}$

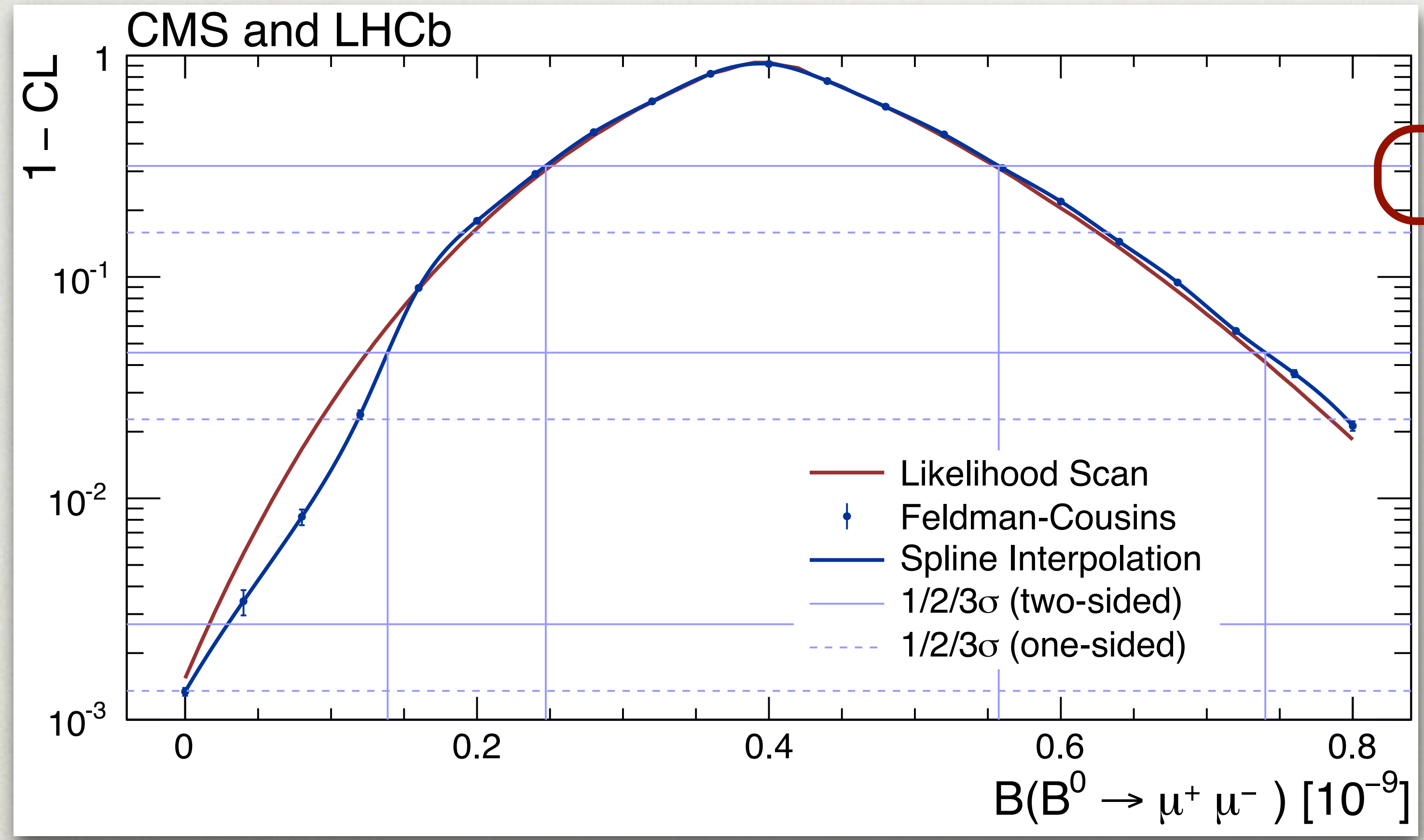


$BF(B_d) = 8.0\times10^{-10}$



There are more scanning steps in between.

FELDMAN-COUSINS RESULTS



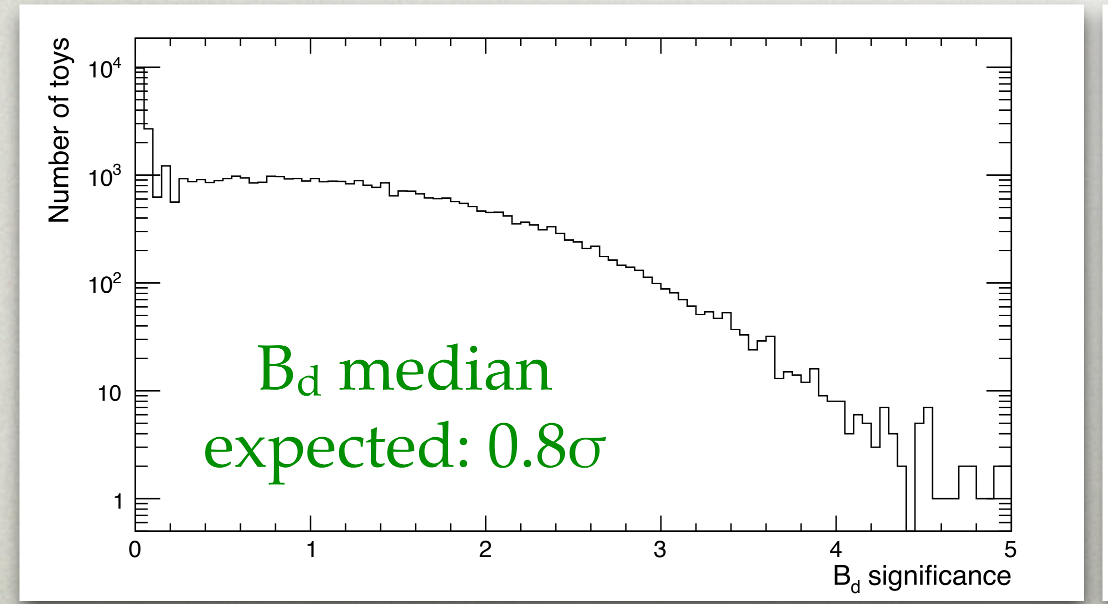
EXTENDED

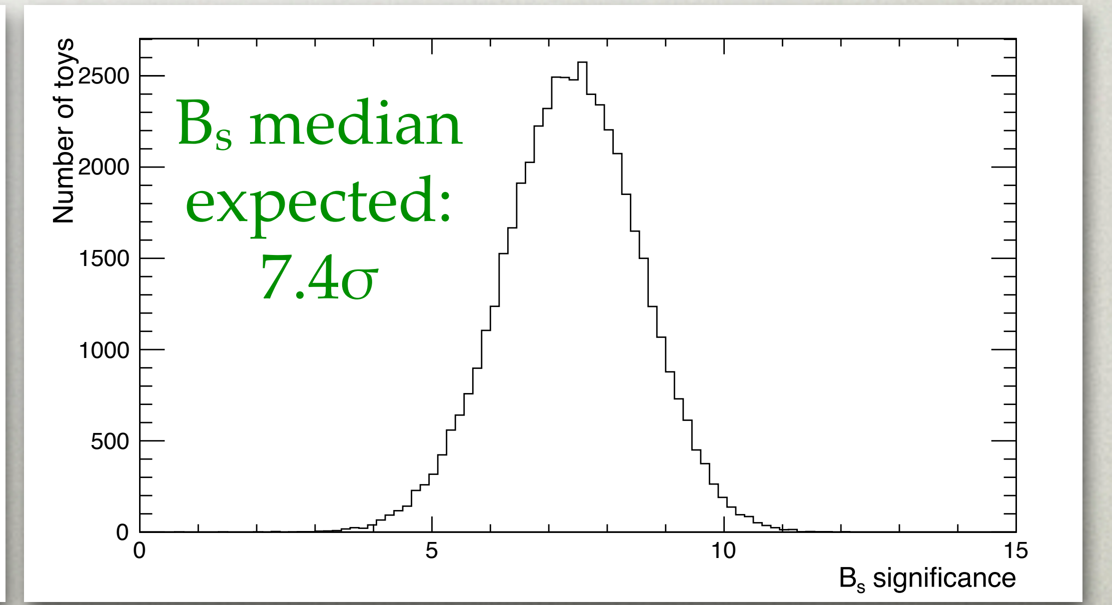
- Excellent agreement with likelihood scan / asymptotic limit, except near physical boundary $BF(B_d) \geq 0$.
- p-value with respect to the background only hypothesis with F-C toys: **$1.34\ {+0.06}/{-0.05} \times 10^{-3}$ [significance: $(3.00\pm0.01)\sigma$]**
- with respect to SM: **2.2σ**

⇒ Consistent with likelihood scan (3.2σ)

EXPECTED SENSITIVITIES

- In order to assess the expected sensitivities for the signals, the F-C tool is also used to simulate with SM signals included:
  - First fit to the data with BF(B_s) & BF(B_d) fixed to SM values.
  - Frequentist toys have been generated using the nuisance parameters obtained from the former fit.
  - Each toy is fitted allowing both branching fractions to be free, and determine the significance for each toy using the asymptotic formula.





COMBINED MASS PLOT

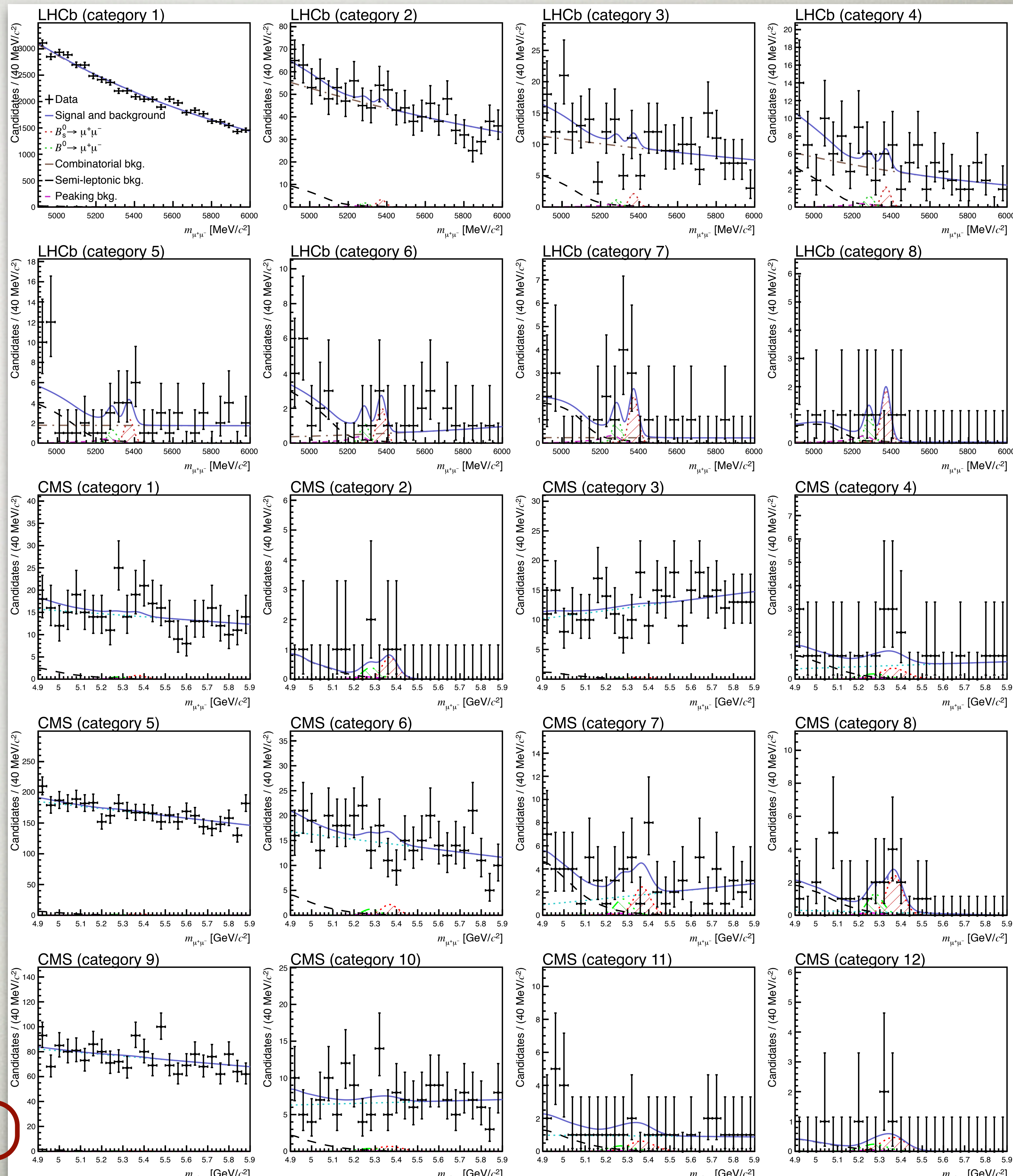
- Seeking for an unified way to present our signal:
  - Unit has been unified to MeV.
  - The range of the mass distribution has been restricted to the common part between the two experiments, ie. [4900, 5900] MeV.
  - Signals are modeled by the Crystal Ball functions accounting for the FSR tail. The ISR contributions have been estimated to be tiny (<2%, details in the backup).
  - Neglect the small difference in terms of peak positions [no real visual effect actually].

	LHCb	CMS	Shift
$M(B_s)$ [MeV]	5371.9	5368.0	–3.9
$M(B_d)$ [MeV]	5284.9	5277.5	–7.4

  - The mass resolutions are also different [LHCb ~ 25 MeV, CMS ~ 32-75 MeV], which has been taken into account in S/(S+B) ranking estimation (see the following slides for details).

INDIVIDUAL CATEGORIES

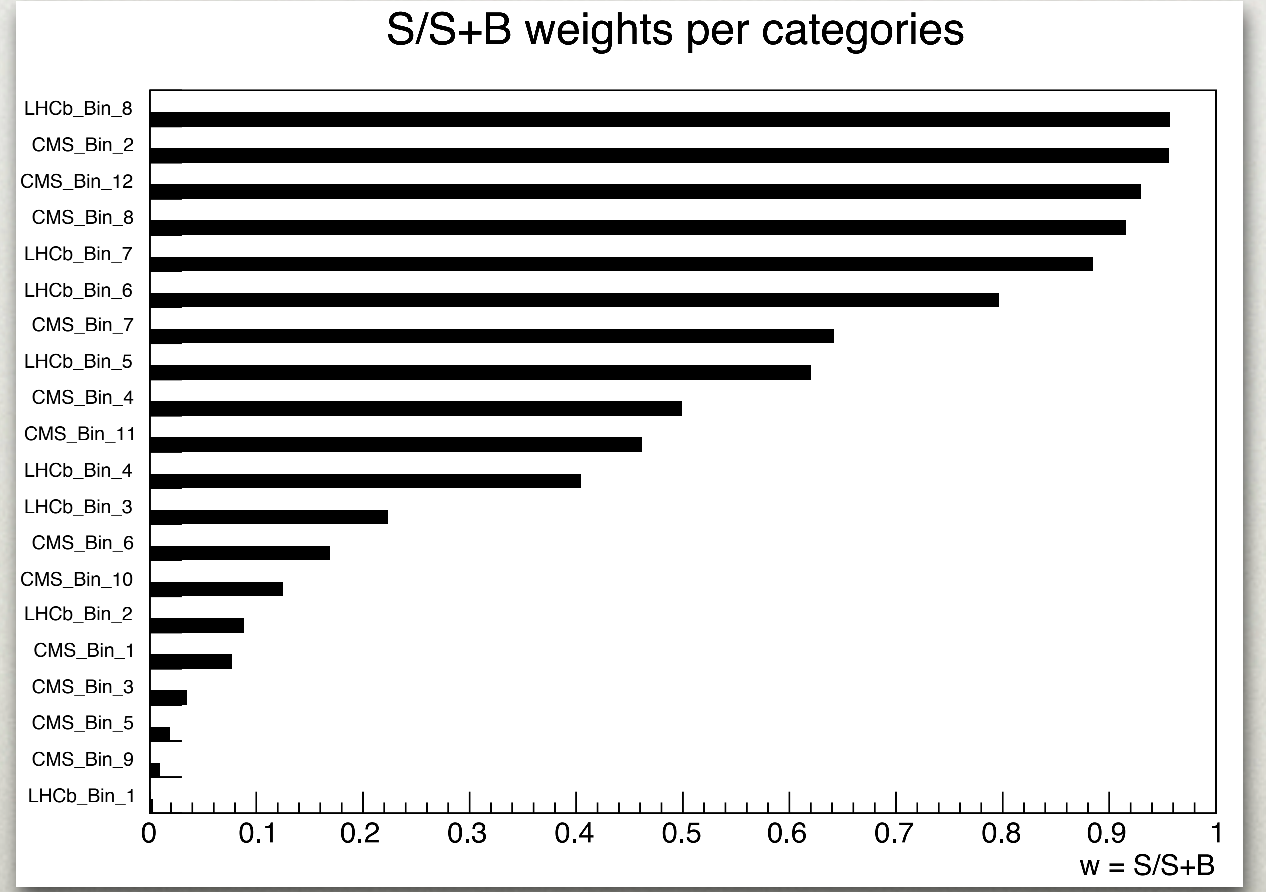
- LHCb 8 categories.
- CMS 12 categories.
- Seeking for a single representative plot.



EXTENDED

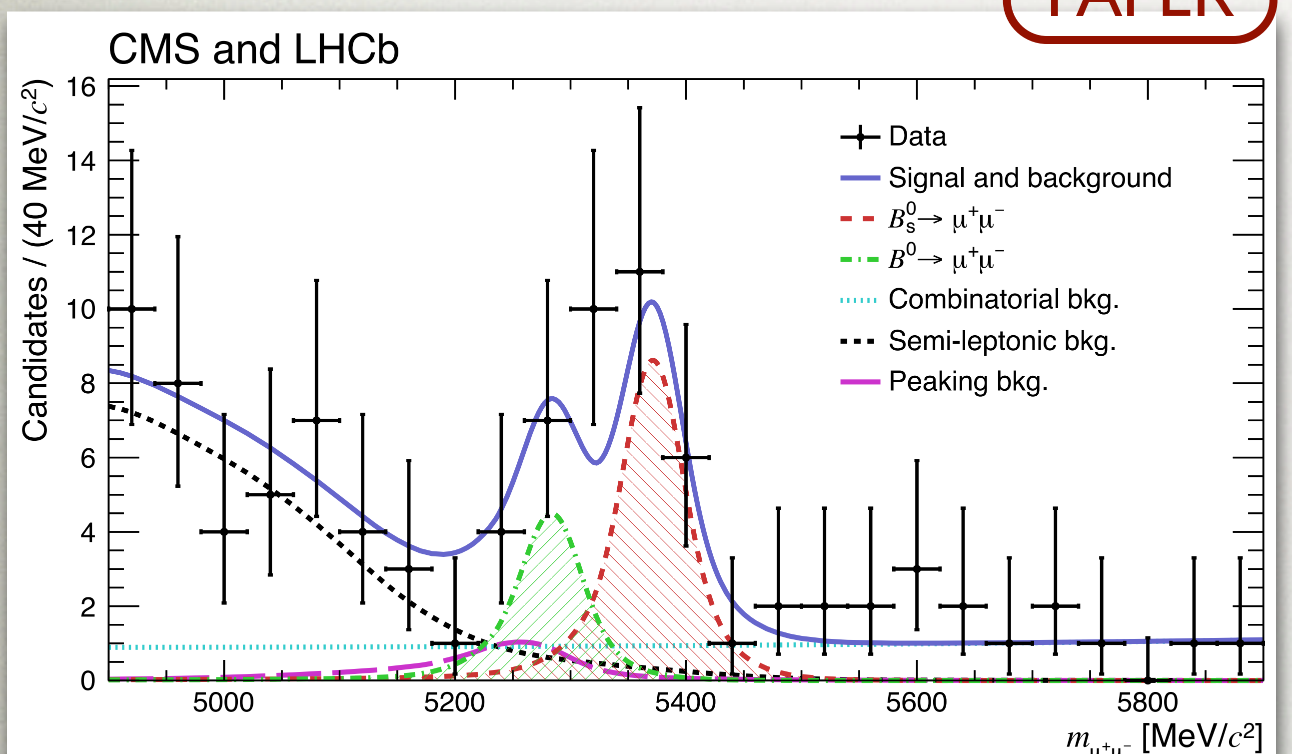
CATEGORY RANKING

- Weights are computed for each category using **S/(S + B)** under the B_s peak (thereby taking into account the varying mass resolution across categories).
- For CMS categories, integrate over per-event mass resolution to compute the weights (as for weighted plot in CMS paper).

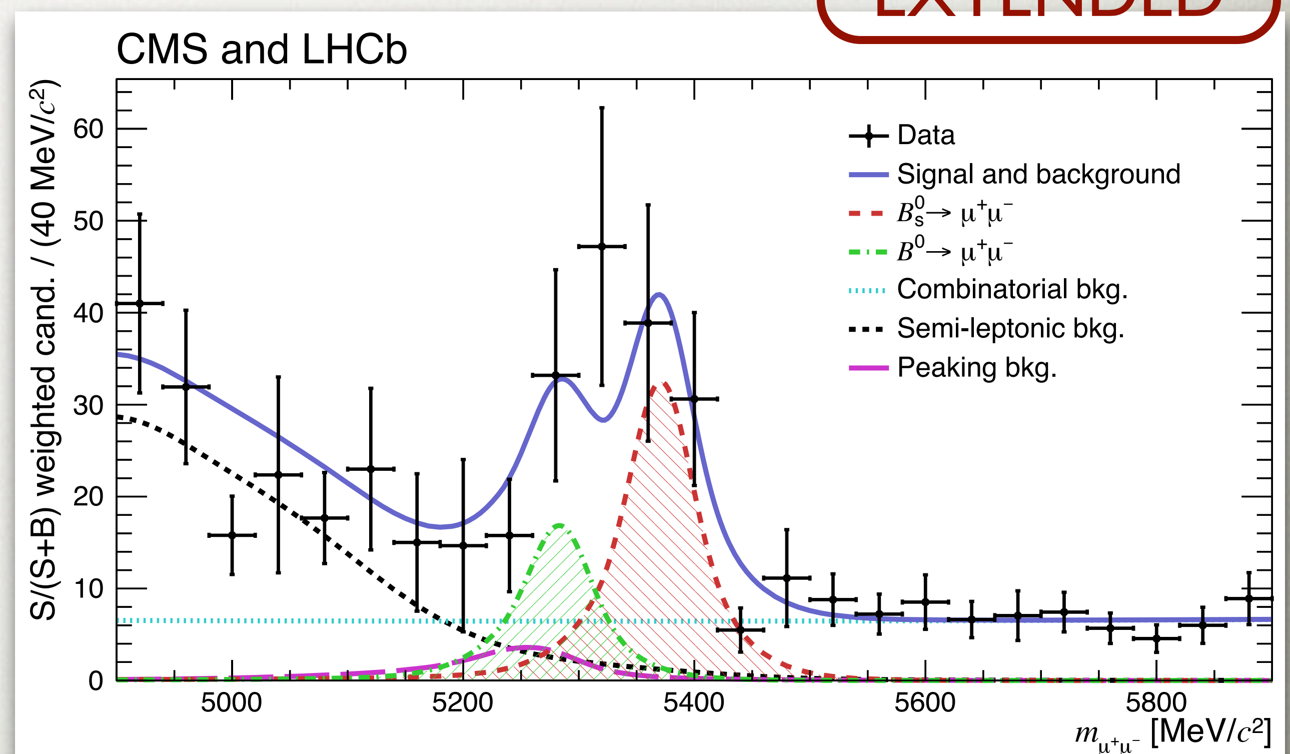


REPRESENTATIVE MASS PLOT

PAPER



EXTENDED

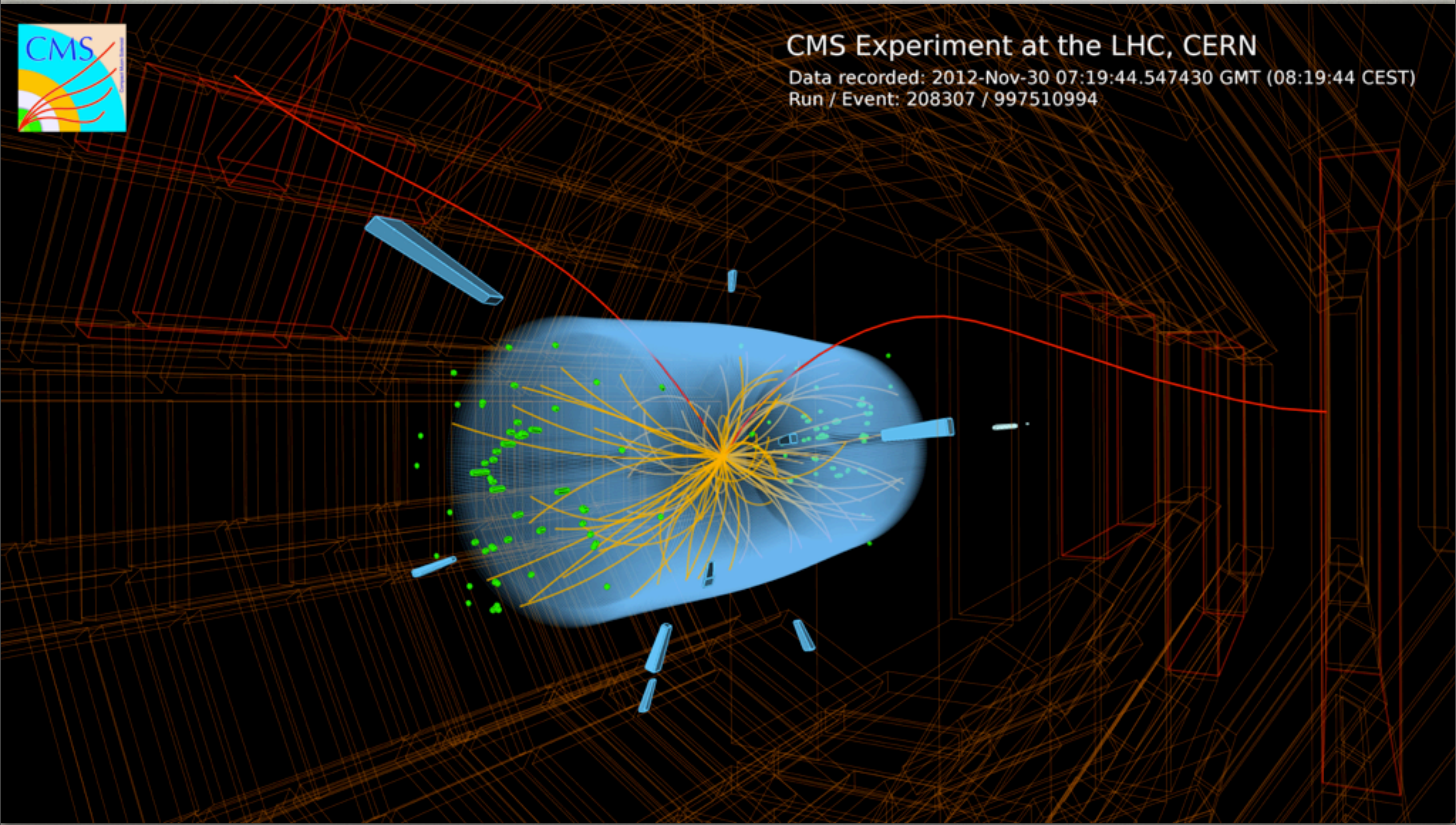


- Stack of Best 6 categories (3 from CMS, 3 from LHCb)
- S/(S+B) weighted plot, with total Bs yield conserved.

(as for the CMS paper)

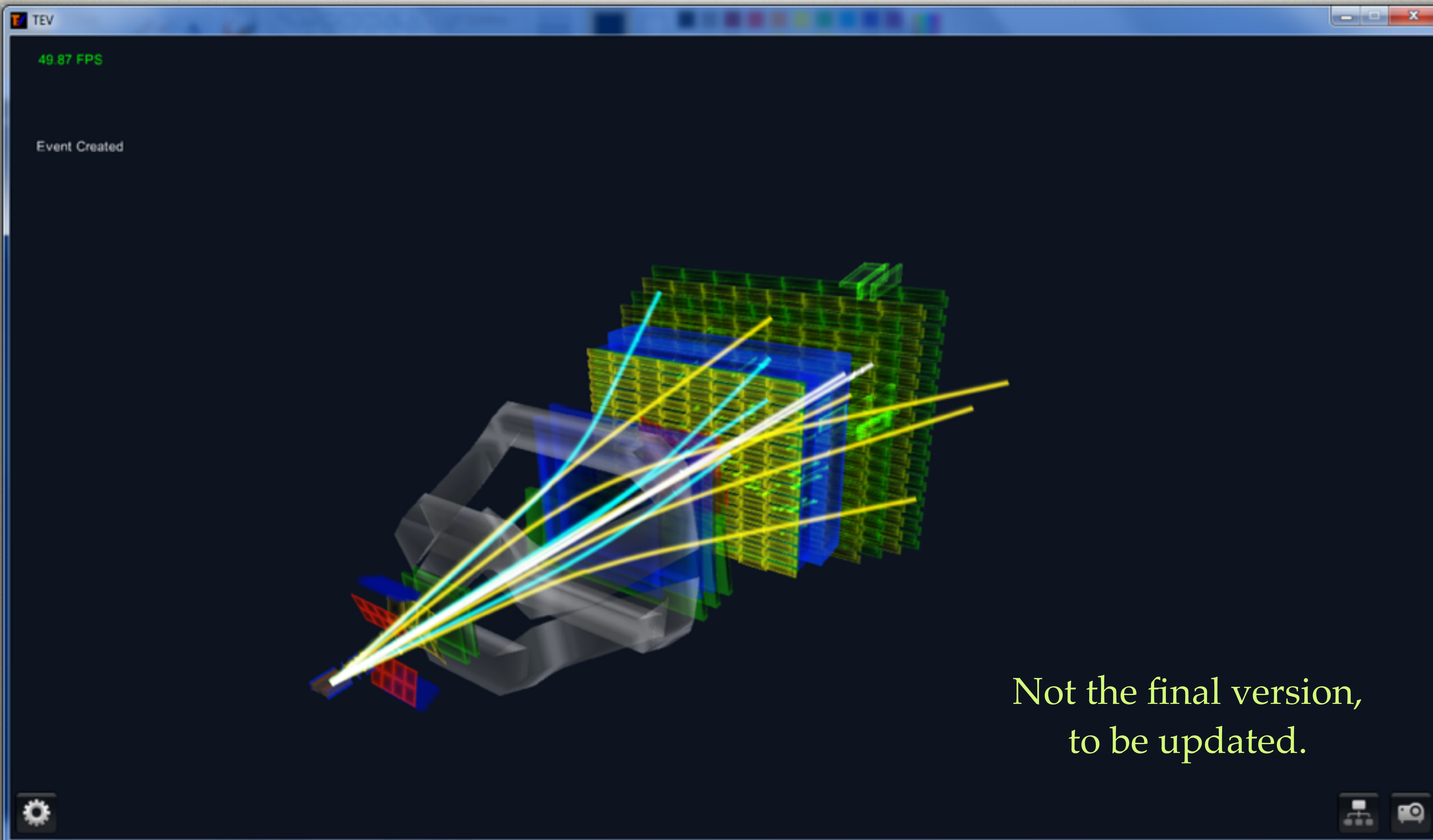
REPRESENTATIVE EVENT DISPLAY

EXTENDED



REPRESENTATIVE EVENT DISPLAY

EXTENDED



SUMMARY

- A full combination for CMS and LHCb analyses for $B_{s,d} \to \mu^+\mu^-$ at likelihood level has been carried out.
- The best fitted branching fractions are

```
BF(Bs) = (2.8 +0.7/−0.6) x 10^−9
BF(Bd) = (3.9 +1.6/−1.4) x 10^−10
```

- The observed significance for $B_s \to \mu^+\mu^-$ is 6.2σ.
- It also produces 3 standard deviation evidence for an excess of events in the search for $B_d \to \mu^+\mu^-$ decays.
- Based on the combined fitter, the 1D/2D likelihood scans, Feldman-Cousins scans, the combined mass plots have been produced.

Backup materials

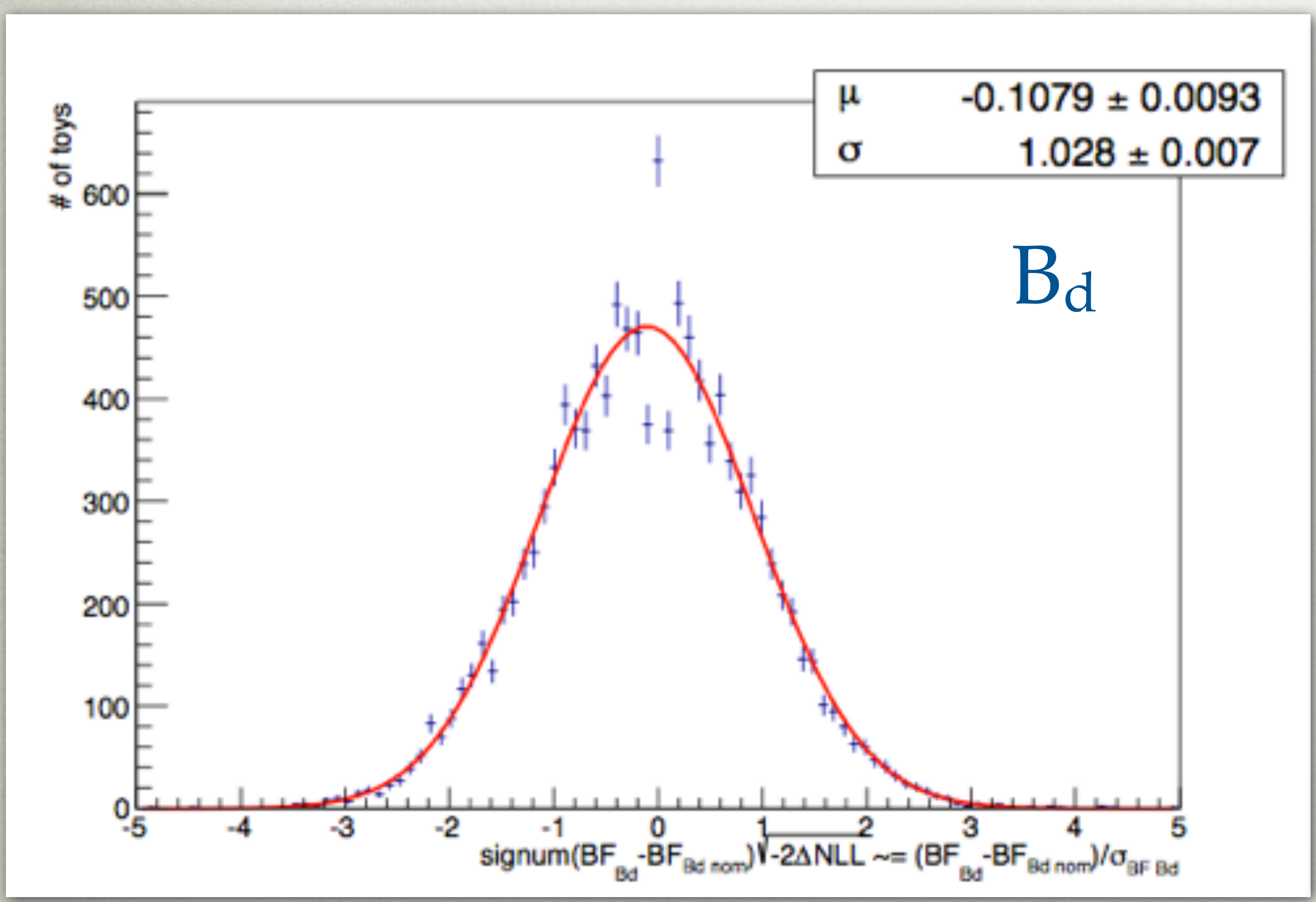
- Split expected / observed significances
- Classical pull distributions
- Cross checks for semileptonic B PDF
- ISR contributions

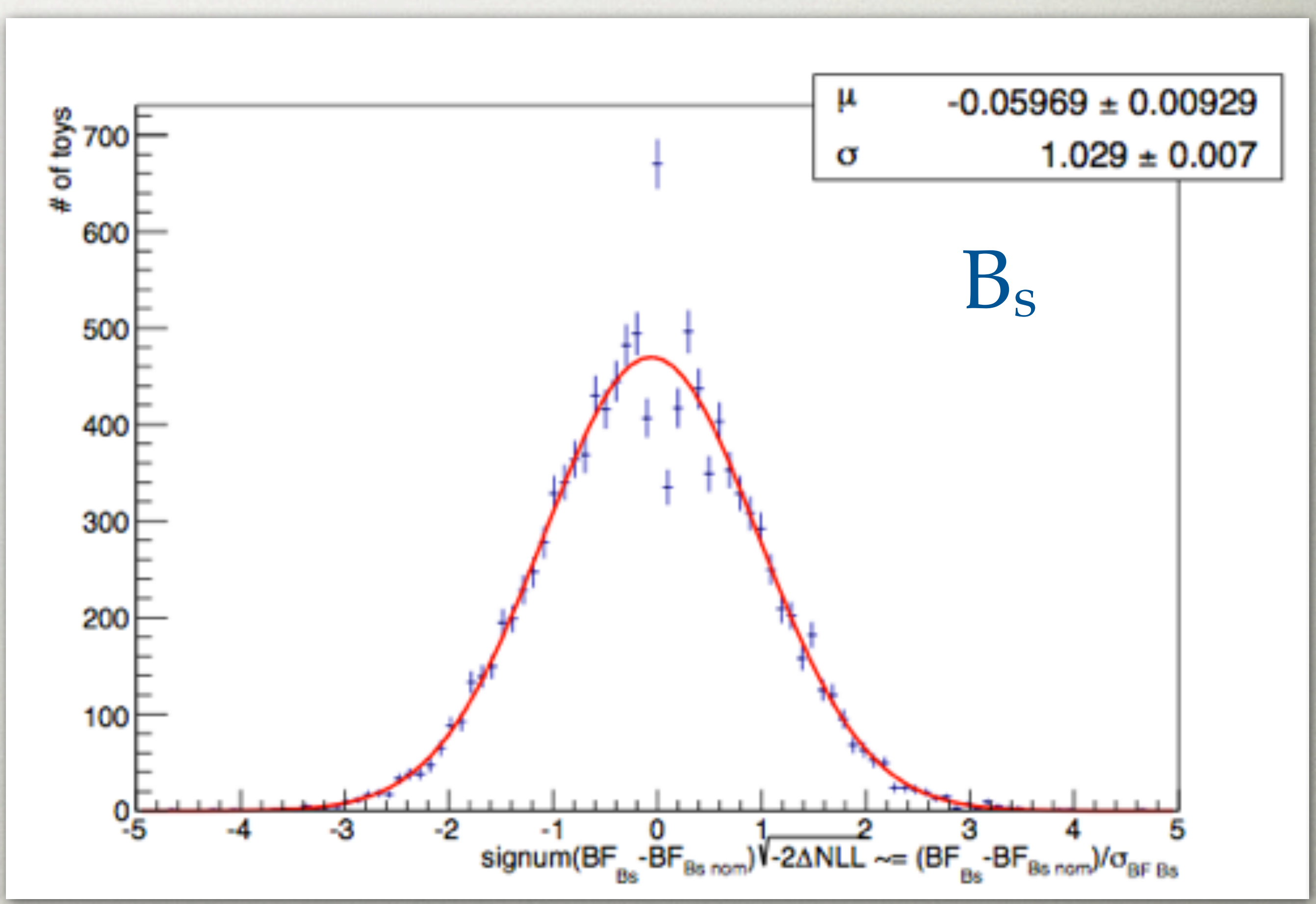
SPLIT EXPECTED/OBSERVED SIGNIFICANCES

- The full set of numbers for expected and observed significance, as computed with Wilk's theorem.
- Observed Significance:
  - CMS-only: B_s: 4.7σ, B_d: 2.6σ
  - LHCB-only: B_s: 3.8σ, B_d: 1.7σ
  - Combined: B_s: 6.2σ, B_d: 3.2σ
- Expected Significance (median expected for SM branching fractions):
  - CMS-only: B_s: 5.3σ, B_d: 0.56σ
  - LHCB-only: B_s: 5.1σ, B_d: 0.53σ
  - Combined: B_s: 7.4σ, B_d: 0.78σ

CLASSICAL PULL DISTRIBUTIONS

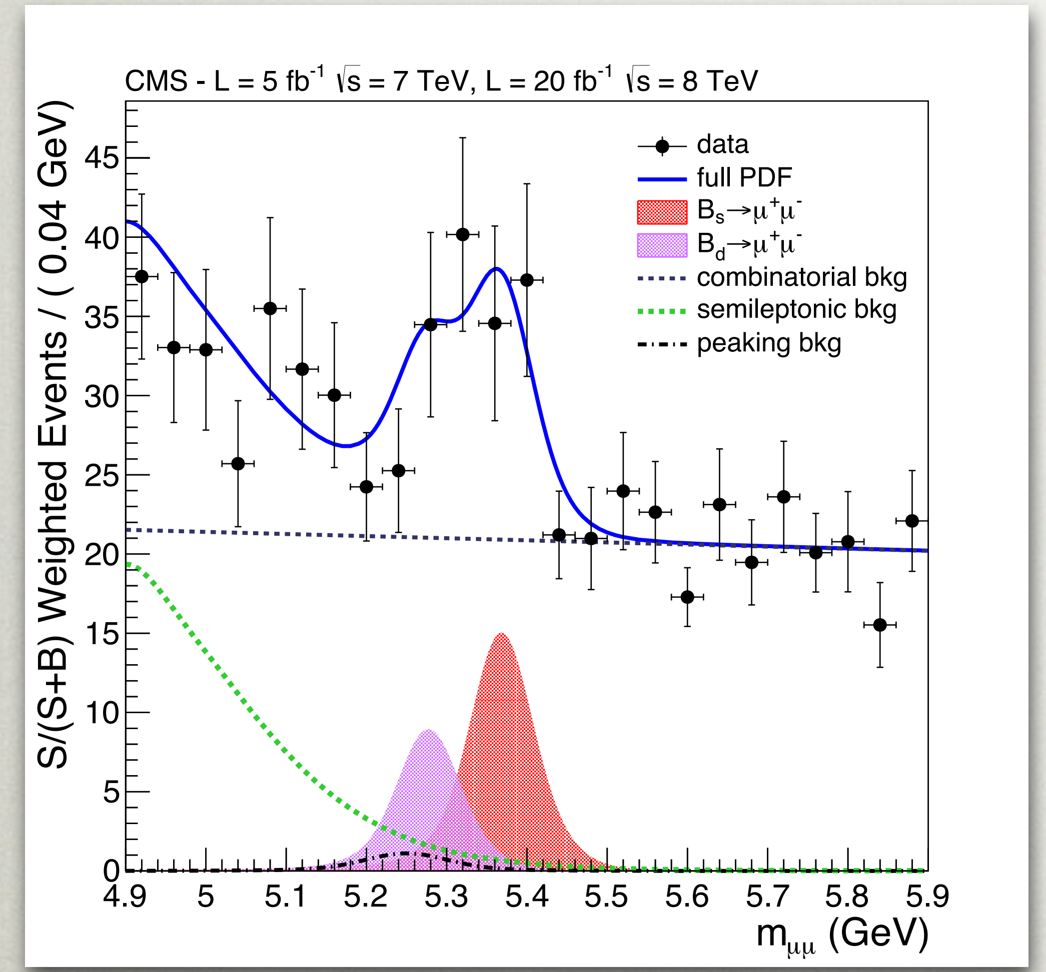
- ARC comment: could we see some pulls (such as Figs 7/8 shown as pulls or pulls comparing the results of both frameworks) to judge whether the fit frameworks have any bias?





CROSS CHECKS FOR SEMILEPTONIC B PDF

Reference fit
(Λ_b model+lifetime corrected)



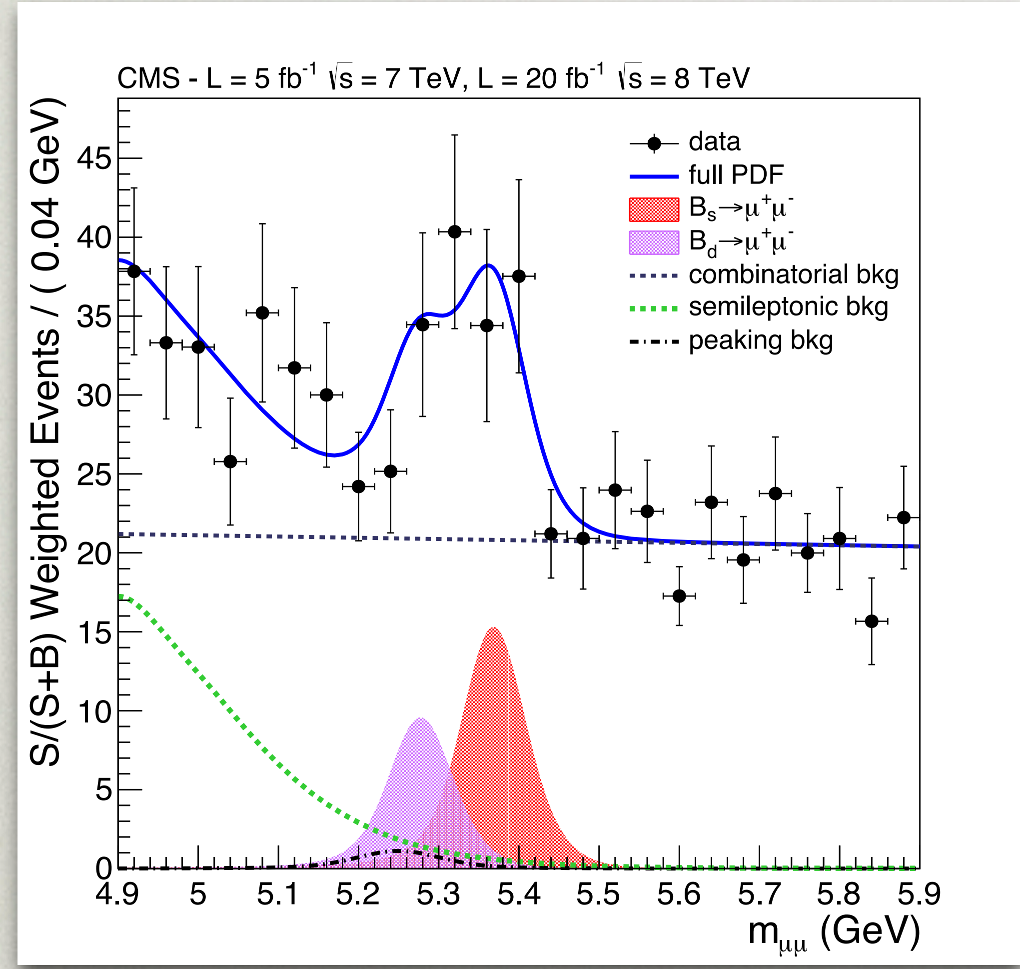
BF(B_s) $= (2.80 +0.97/-0.82) \times 10^{-9}$

BF(B_d) $= (4.31 +2.20/-1.89) \times 10^{-10}$

- Checks done:
  - Enlarge the uncertainty to 100%
  - Enlarge the yield/uncertainty to the old numbers
  - Fit with different PDF (fixed)
  - Fit with PDF morphing

SEMILEPTONIC UNCERTAINTIES & YIELDS

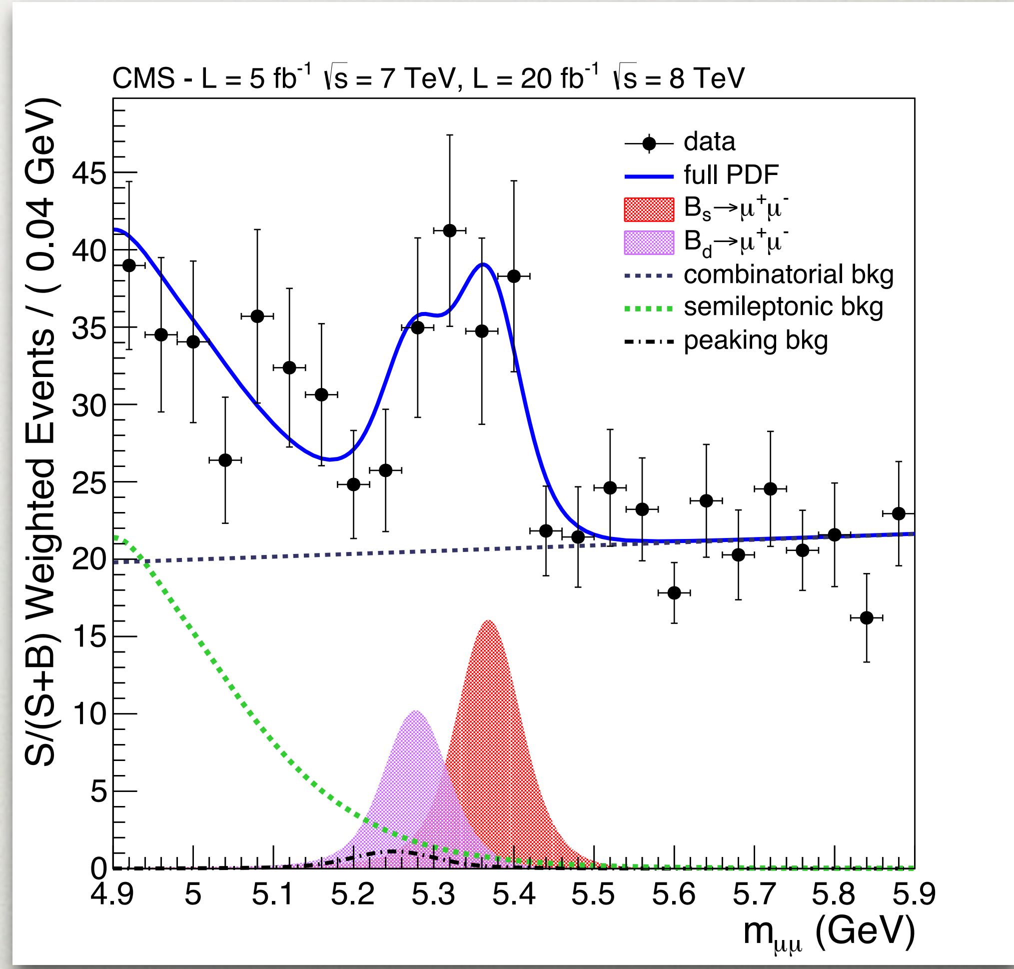
Set semileptonic uncertainties to 100%



$BF(B_s) = (2.85 +0.99/-0.84) \times 10^{-9}$

$BF(B_d) = (4.62 +2.24/-1.99) \times 10^{-10}$

Use old (large) yields & uncertainties



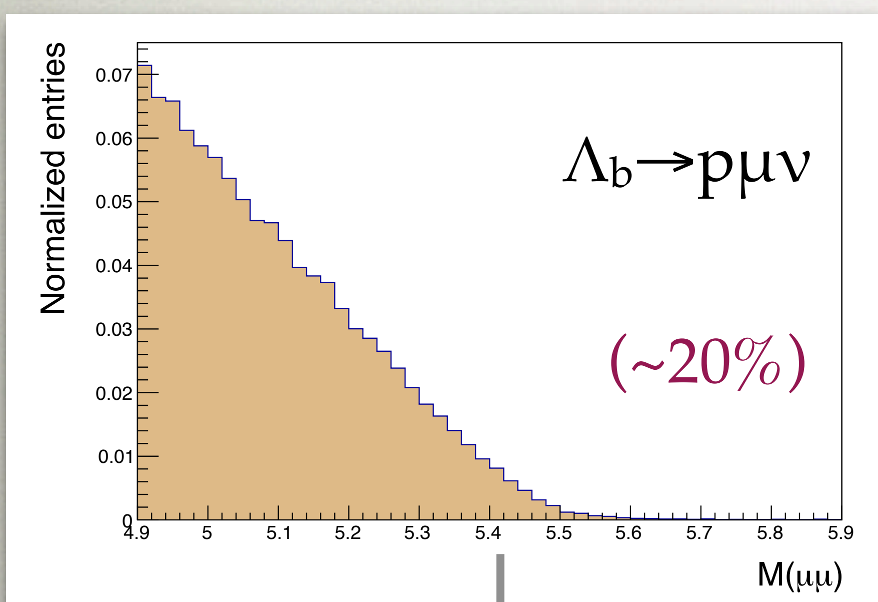
$BF(B_s) = (3.00 +1.00/-0.85) \times 10^{-9}$

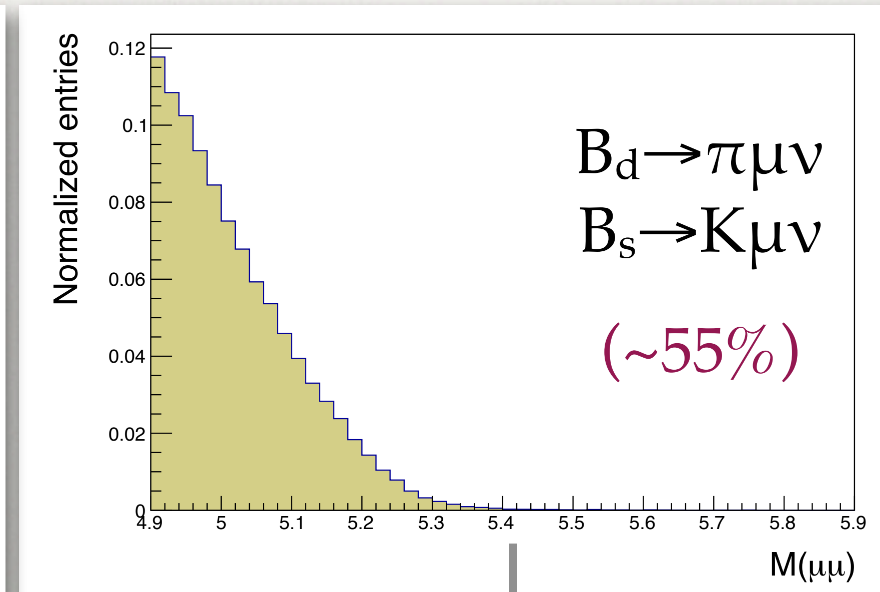
$BF(B_d) = (4.95 +2.24/-1.98) \times 10^{-10}$

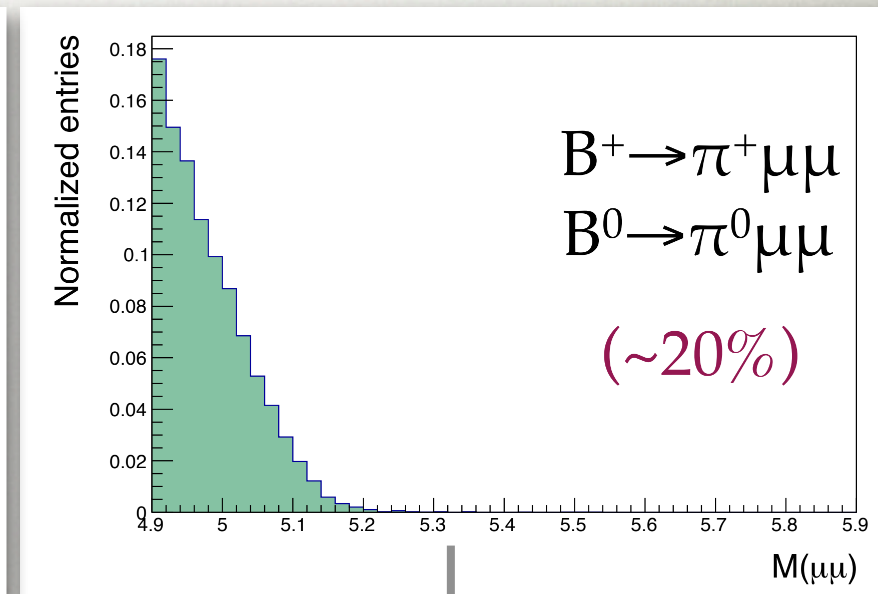
SEMILEPTONIC MODELING

- The current semileptonic PDFs are implemented with 2 RooKeysPdf for mass and mass resolution (it's a 2D function).
- Construction of PDF was carried out with weighted MC events, ie. mix the MC events from each process with proper weights.
- Cannot change the PDF on the fly — have to update the weights of MC events and rebuild the PDF from beginning.
- Checks are carried out in two ways:
  - Fit with alternative PDFs (fixed line shapes, but with different weights between subprocesses).
  - Fit with template morphing.

SEMILEPTONIC COMPONENTS



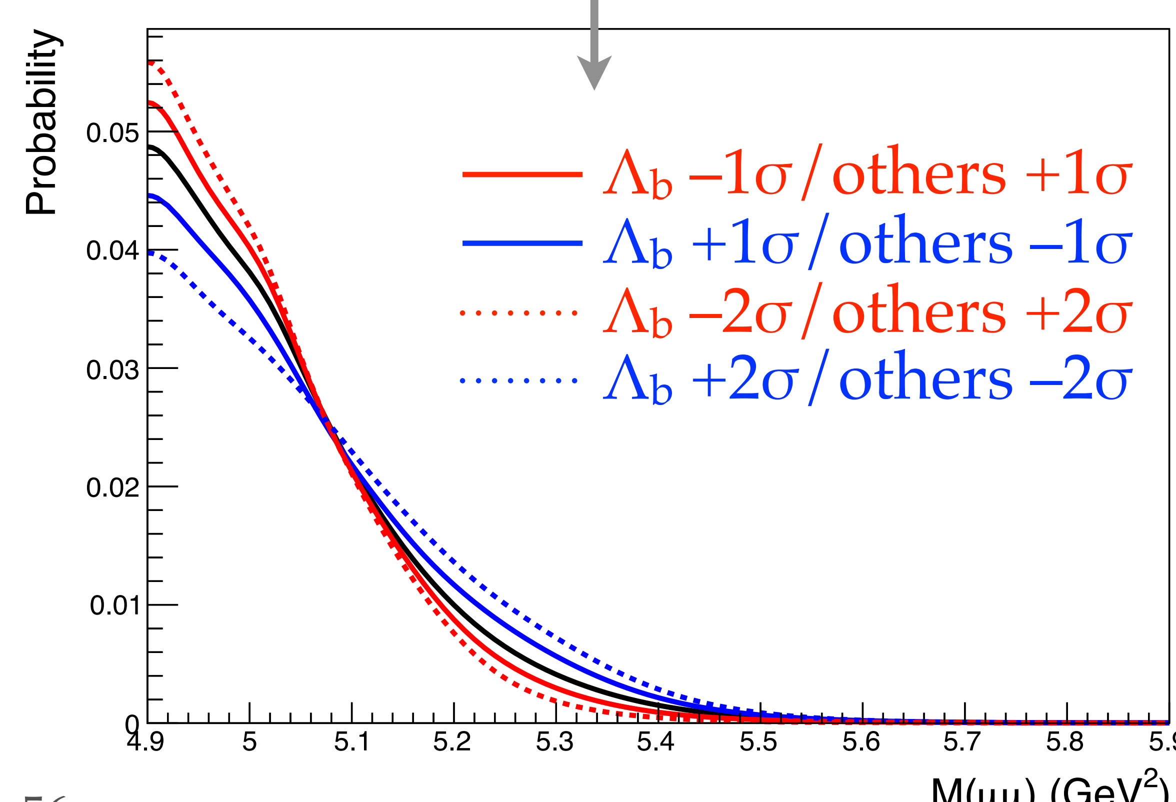




Basically the Λ_b shape has the smallest slope — simply try the extreme cases, ie. anti-correlated Λ_b and other components. [Remark: 1σ ~ O(50%), so maximal variation goes to roughly 2σ...]

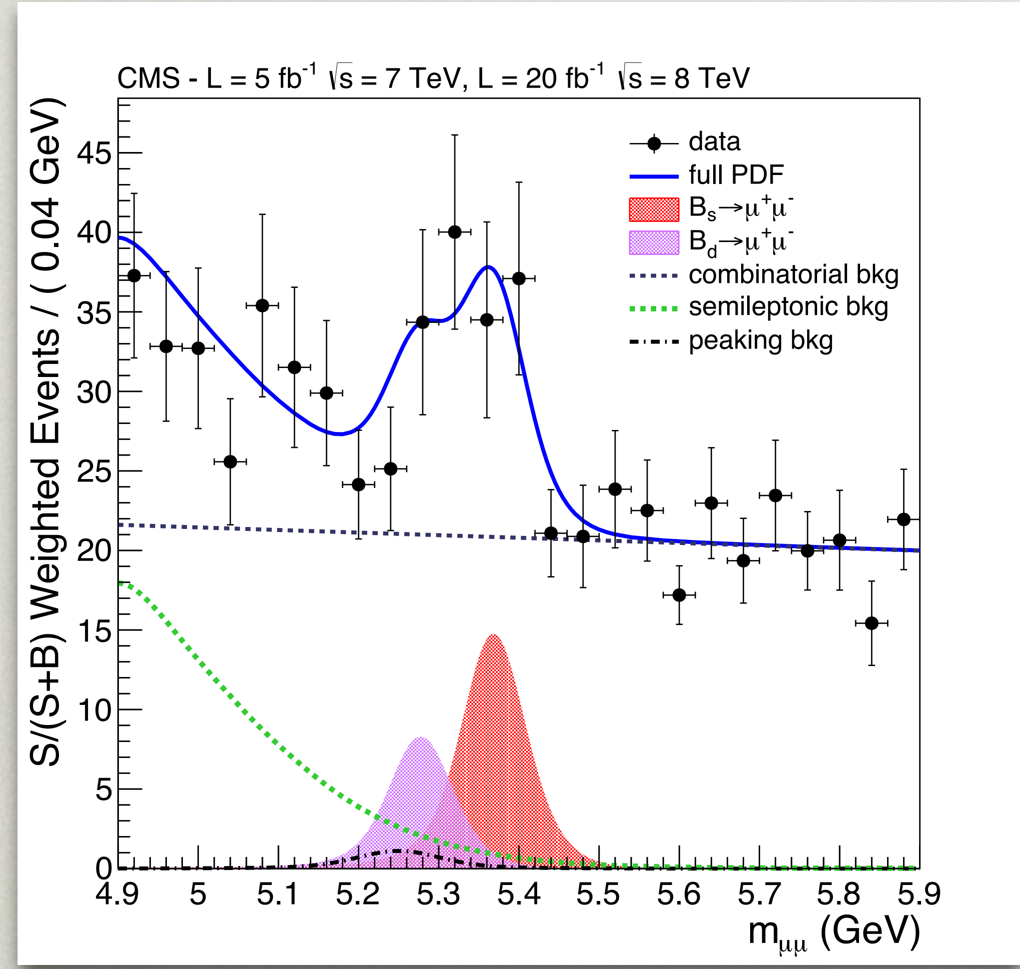
A leading order check only ➡

(since all subprocess can vary independently in principle...)



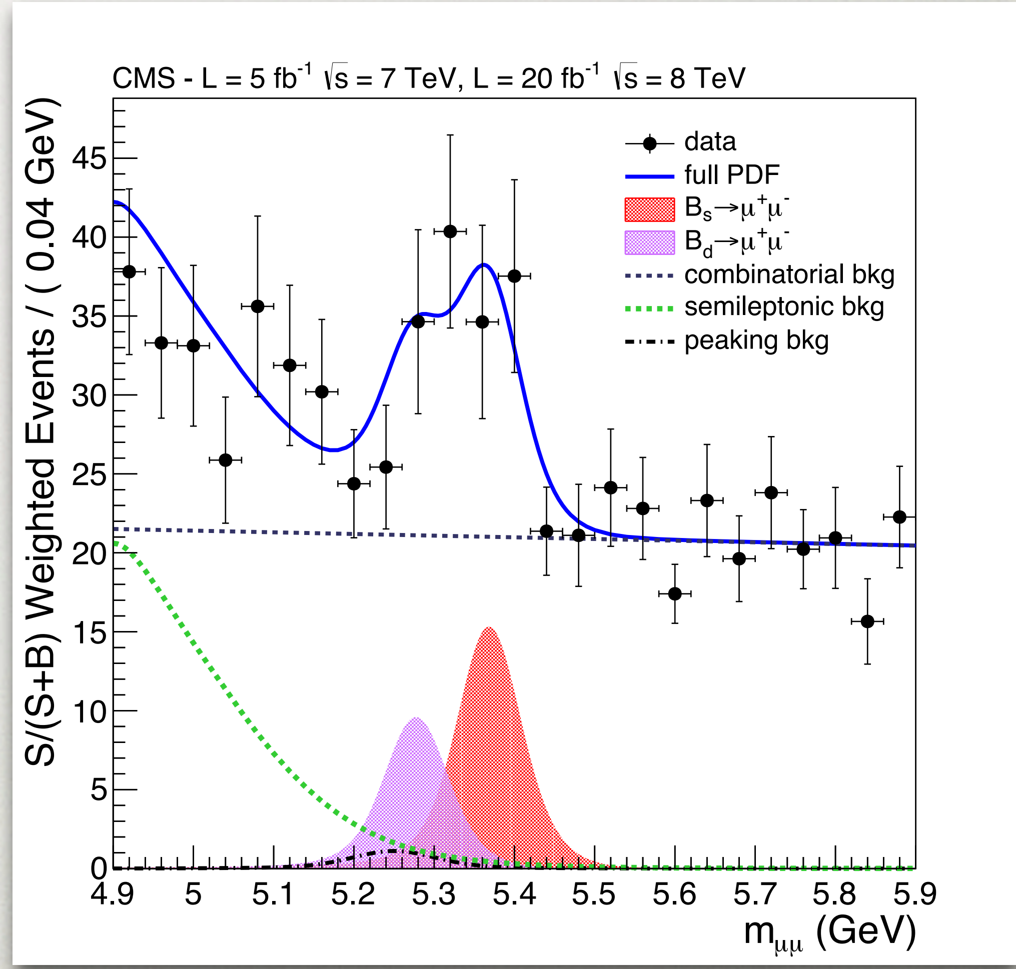
PDF REPLACEMENT

Replacing semileptonic PDF:
Λ_b +1σ / all others −1σ
no changes of yields



$BF(B_s) = (2.74 +0.97/-0.81) \times 10^{-9}$
$BF(B_d) = (3.99 +2.17/-1.86) \times 10^{-10}$

Replacing semileptonic PDF:
Λ_b −1σ / all others +1σ
no changes of yields



$BF(B_s) = (2.85 +0.98/-0.82) \times 10^{-9}$
$BF(B_d) = (4.62 +2.23/-1.93) \times 10^{-10}$

PDF MORPHING

- Adding a new nuisance parameter (β) that can vary in [–2,+2], which maps to the $+2\sigma/-2\sigma$ semileptonic line shapes. Common to all BDT categories.
- Constraining β to a Gaussian of unit width and zero mean (ie. $\beta=\pm1$ are corresponding to $\pm1\sigma$ variations of the semileptonic line shape).
- Implemented with an adhoc fitter with output workspaces only.

REFERENCE

$BF(B_s) = (2.80\ {+0.97}/{-0.82}) \times 10^{-9}$

$BF(B_d) = (4.31\ {+2.20}/{-1.89}) \times 10^{-10}$

$\delta BF(B_d) = +51.0\% / -43.9\%$

+SHAPE SYSTEMATICS

$BF(B_s) = (2.79\ {+0.97}/{-0.82}) \times 10^{-9}$

$BF(B_d) = (4.23\ {+2.22}/{-1.90}) \times 10^{-10}$

$\beta = 0.25\ {+0.91}/{-1.05}$

$\delta BF(B_d) = +52.5\% / -44.9\%$

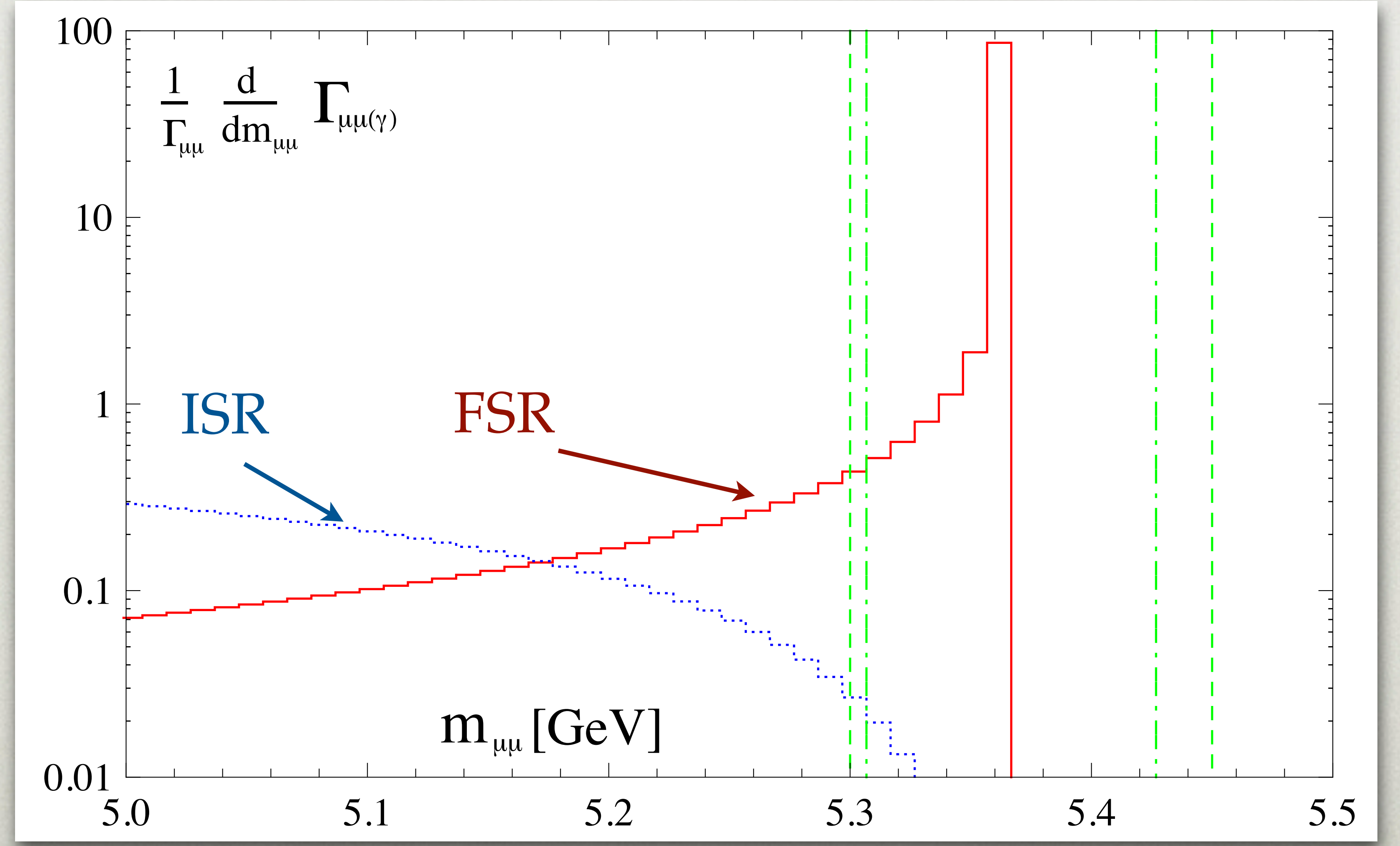
$\Rightarrow$ Best fitted: $+0.25\sigma$ line-shape; resulting variations are minor.

SEMILPETONIC PDF: COMMENTS

- Enlarge the semileptonic yield uncertainties to 100% resulting a variation of ~2% on B_s and ~7% on B_d.
- Fit with alternative PDFs also shows a variation of ~2% of B_s and ~7% on B_d branching fractions.
- Fit with template morphing shows no real difference in the resulting central values. The relative uncertainties on B_d is slightly increased (51% $\Rightarrow$ 52%, roughly equivalent to add ~10% in quadrature).

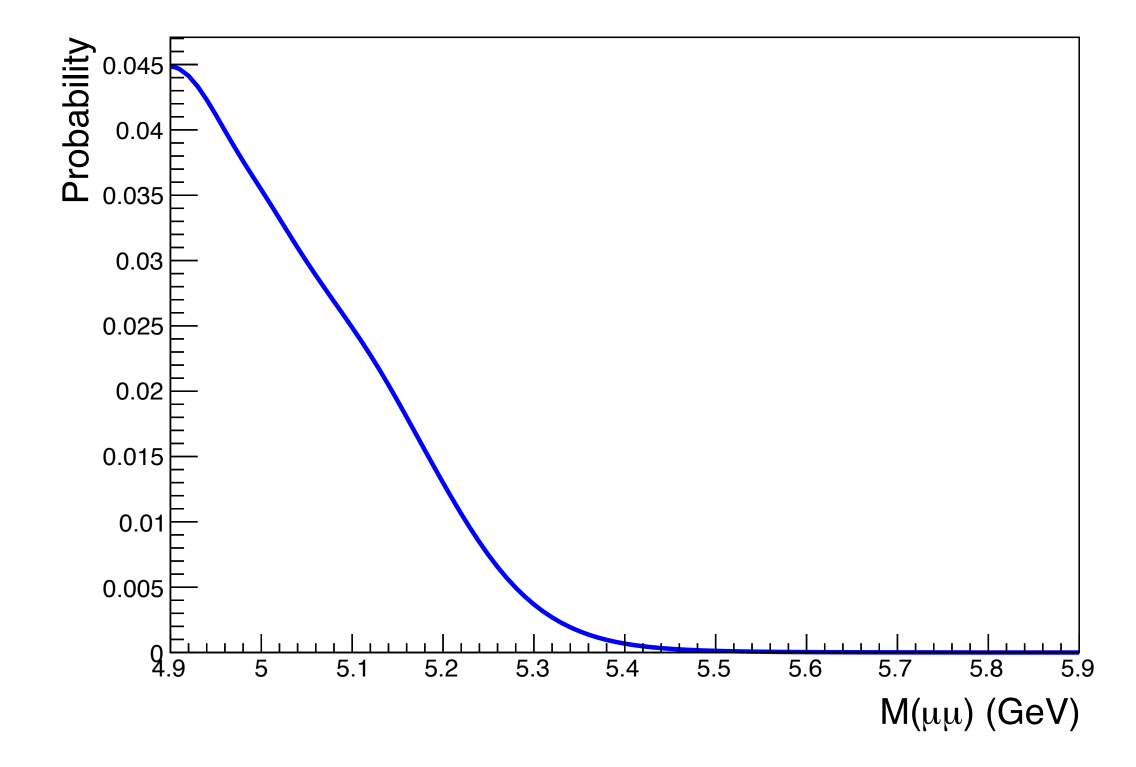
ISR CONTRIBUTIONS

- It has been pointed out that the ISR contribution might be an issue, while the FSR contribution has been included in the PDF modeling already.
- Source: M. Misiak's talk at Beauty 2014: https://indico.cern.ch/event/308116/session/13/contribution/49



ISR MODELING

- Digitize the histogram from the talk, and found that it can be modeled with a linear function: a + b*x with a = 4.758, b = –0.892.
- Integrate over the curve in the fitting region [4.9 GeV, 5.9 GeV] gives a rate of 8.3% (normalized to the total $B_s \to \mu\mu$ decay rate)
- Use the per-event mass resolution from B_s MC to smear the model, and produce the corresponding PDF:



Then include this contribution in the UML fit as a cross check.

TEST WITH UML FIT

REFERENCE	BF(B_s) = (2.79 +0.97/-0.82) x 10^{-9} BF(B_d) = (4.31 +2.20/-1.89) x 10^{-10}
Constraining to (8.3±8.3)% of B_s yield	BF(B_s) = (2.77 +0.97/-0.81) x 10^{-9} BF(B_d) = (4.33 +2.20/-1.90) x 10^{-10}
Doubled contribution: (16.6±16.6)% of Bs	BF(B_s) = (2.74 +0.96/-0.81) x 10^{-9} BF(B_d) = (4.36 +2.20/-1.91) x 10^{-10}
Adding B_d ISR as well (~5.4% of B_d yield)	BF(B_s) = (2.79 +0.97/-0.81) x 10^{-9} ⇒ 1% effect BF(B_d) = (4.27 +2.20/-1.89) x 10^{-10}
Doubled contribution: 16.6% B_s + 10.8% B_d	BF(B_s) = (2.78 +0.98/-0.82) x 10^{-9} ⇒ 2% effect BF(B_d) = (4.23 +2.20/-1.89) x 10^{-10}

The effect should be small enough and can be neglected.