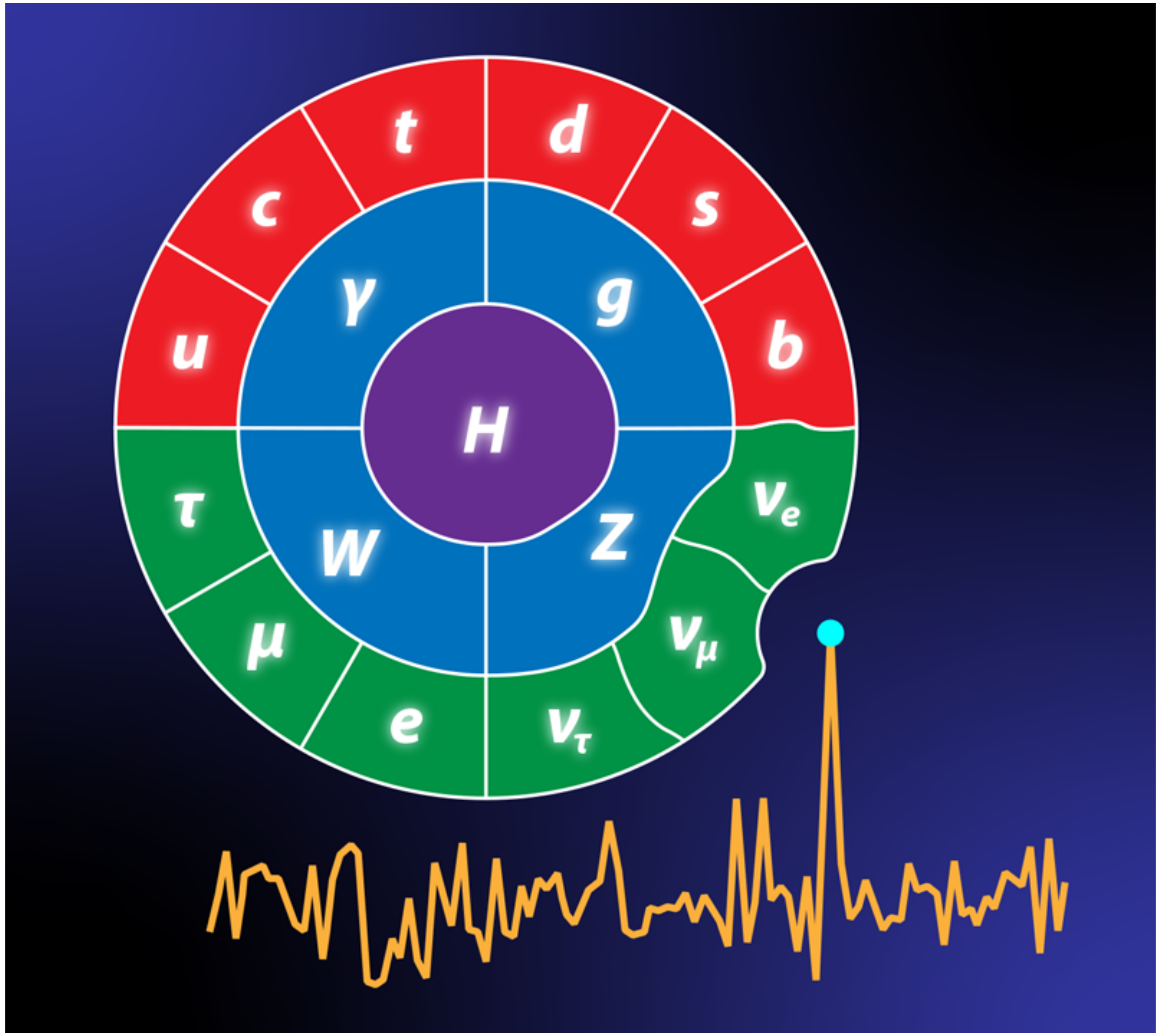
Measurements of $B_{(s)} \to \mu\mu$ decays

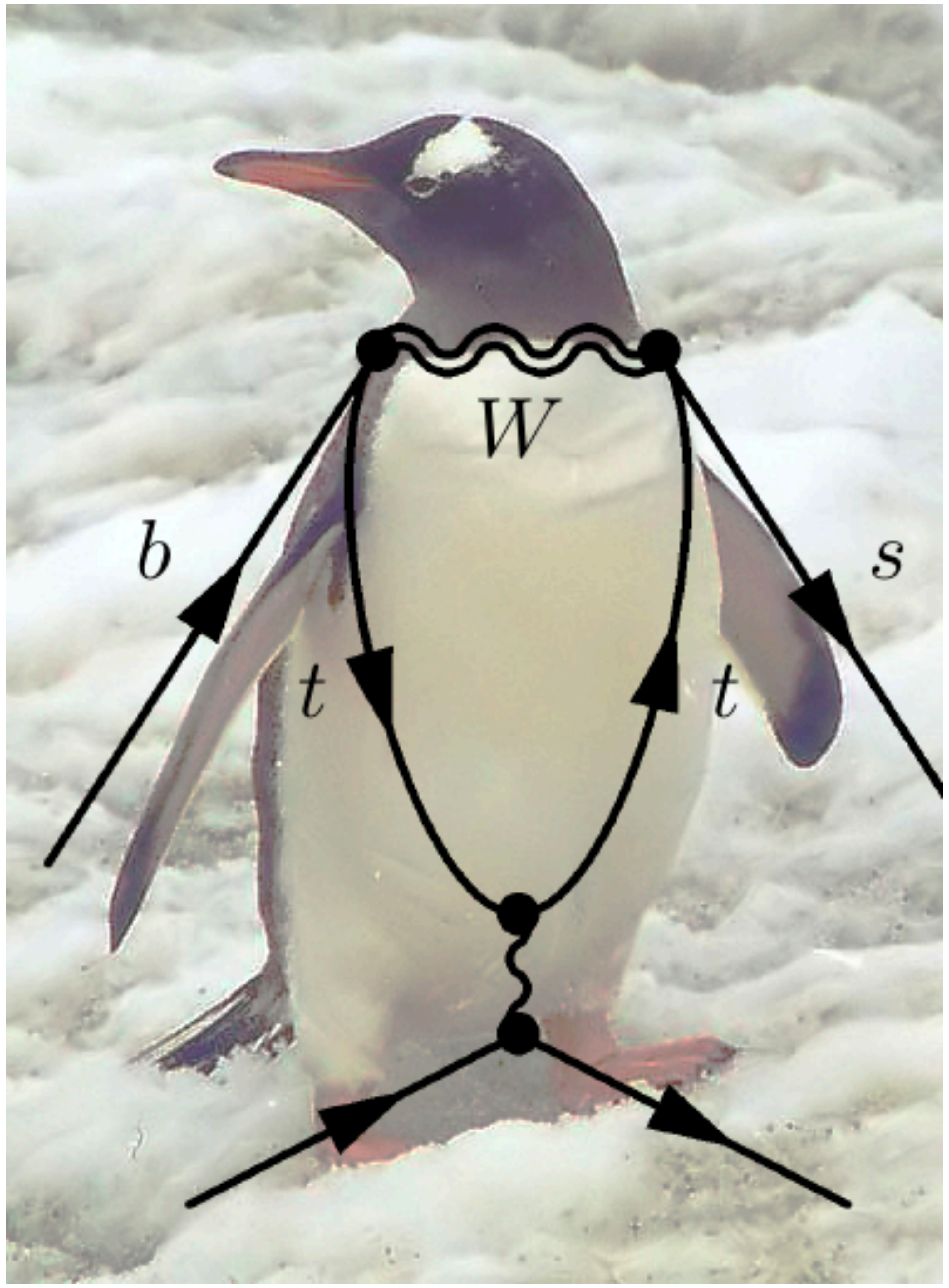
Dmytro Kovalskyi, MIT

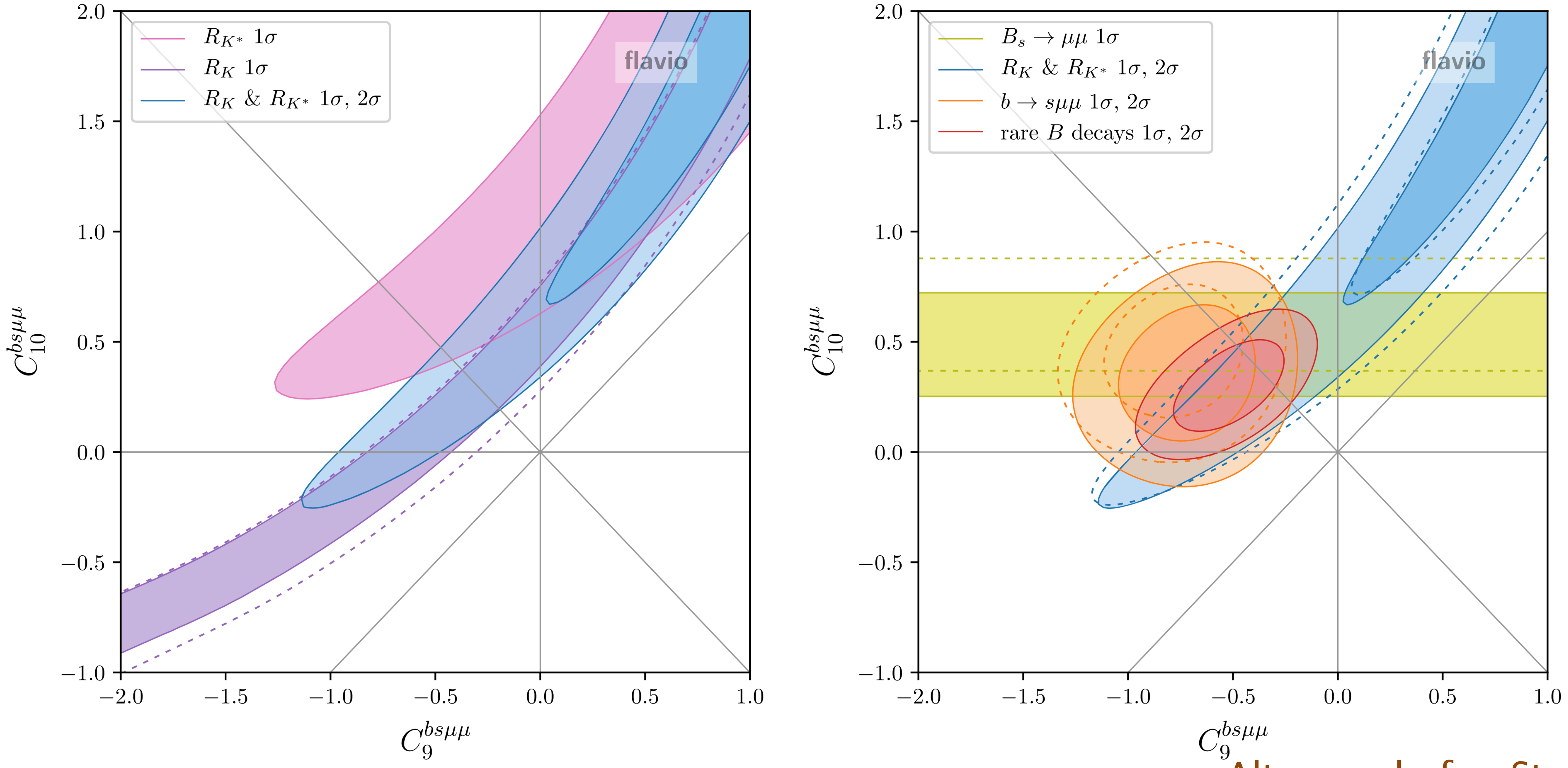
The Era of Anomalies



- Higgs discovery in 2012 made the Standard Model complete
  - Most likely it's not the final model
  - It doesn't describe all effects
- Anomalies may be a hint of something NEW
  - In most cases they are just statistical fluctuations, but not always

Rare B decay Anomalies



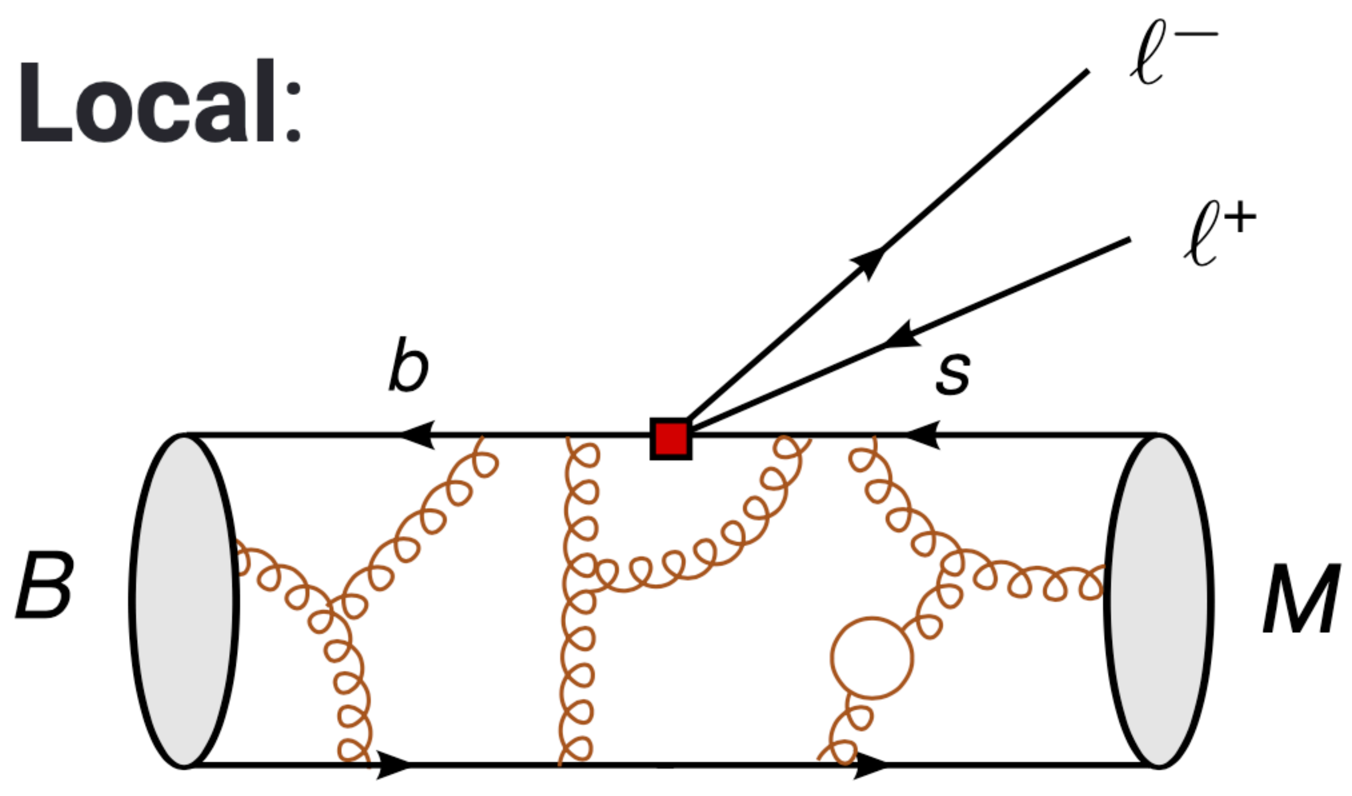


Altmannshofer, Stangl, arXiv:2103.13370

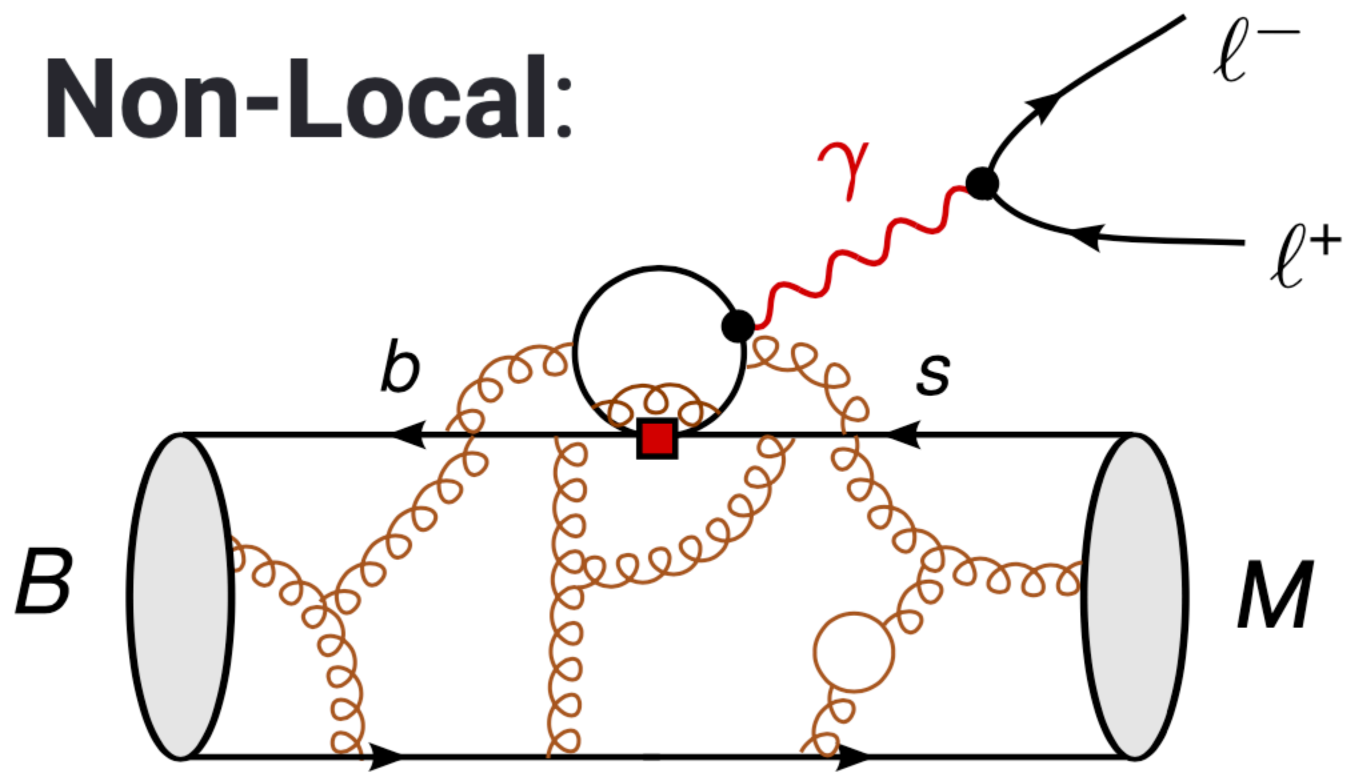
- Multiple discrepancies are observed in rare B decays
  - 3.1σ Lepton Flavour Universality violation in R(K) and R(K*)
  - 2-3σ discrepancies in branching fraction and angular observables
- Global fits provide good description of data in the effective 4-fermion interaction

Interpretation of Data

Local:

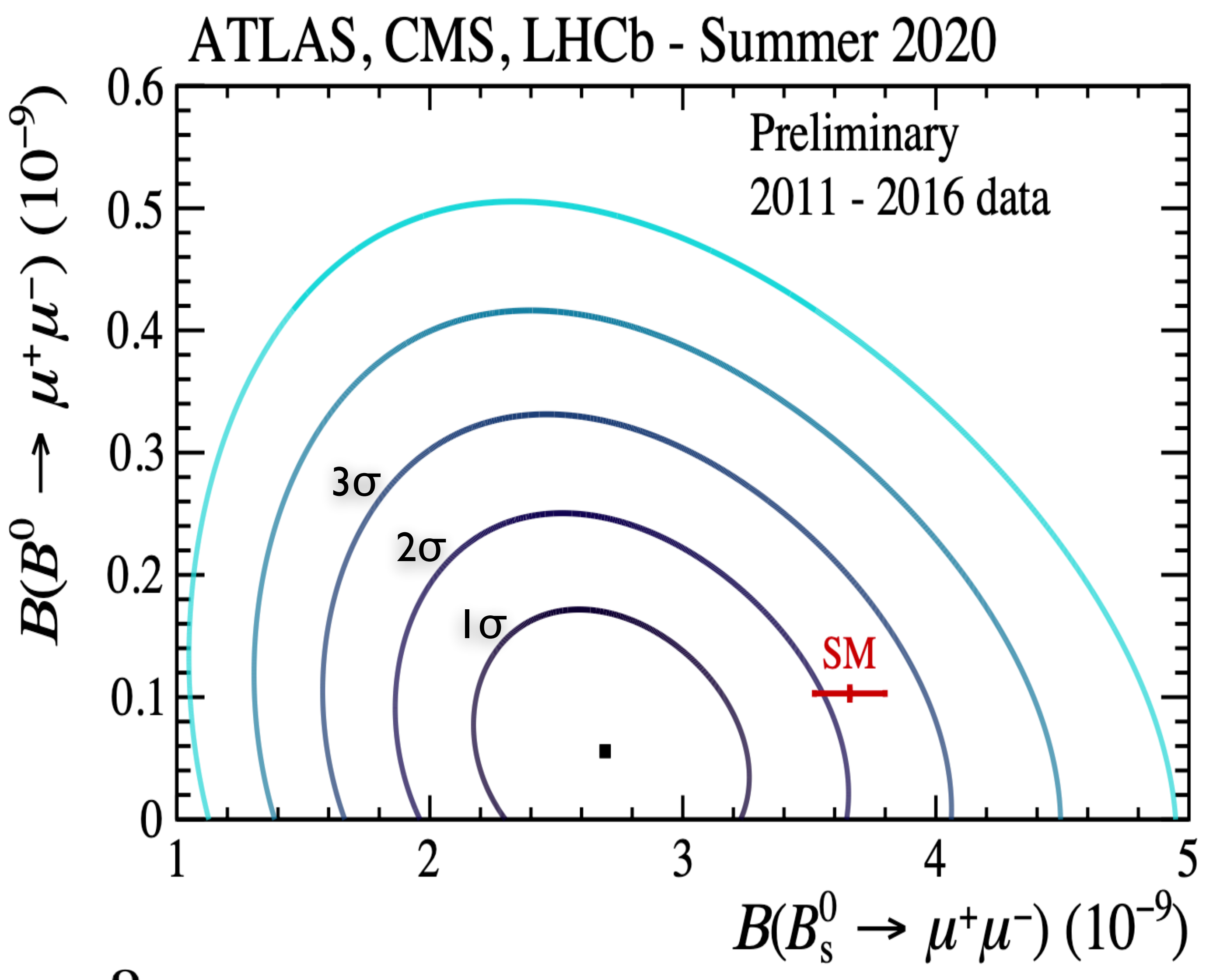


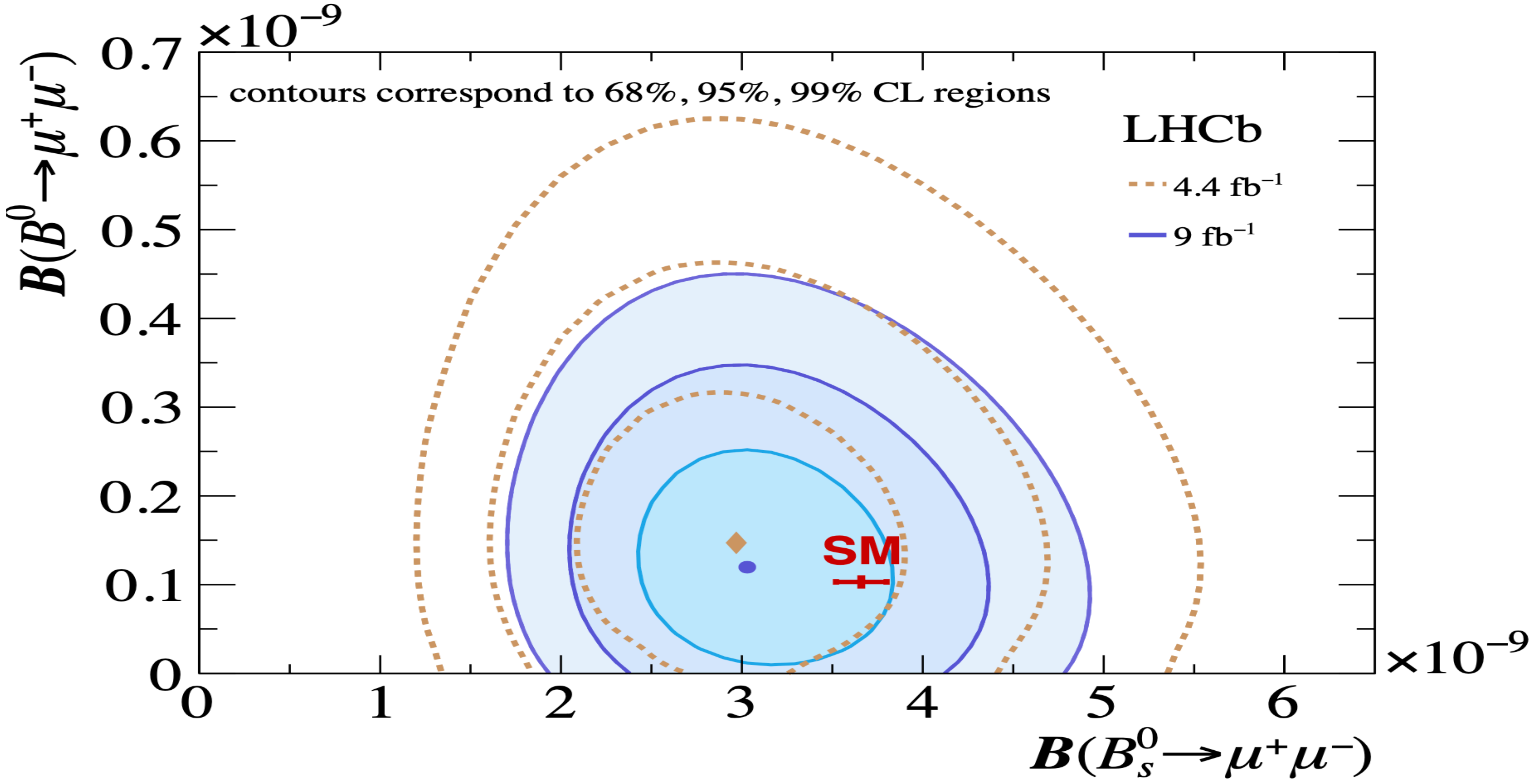
Non-Local:



- Main challenge
  - What we can measure in most cases has non-trivial non-perturbative QCD contributions and hard to interpret
  - Clean theoretical observables in most cases are very hard to measure experimentally
- Theory Overview
  - Local contributions
    - Contributions from new physics at higher energy scale - perturbative
    - Local matrix elements (form factors) - non-perturbative, but well know for large q^2
  - Non-local contributions
    - Enters at low q^2
    - Dominated by $b \to sc\bar{c}$ - can mimic New Physics contributions

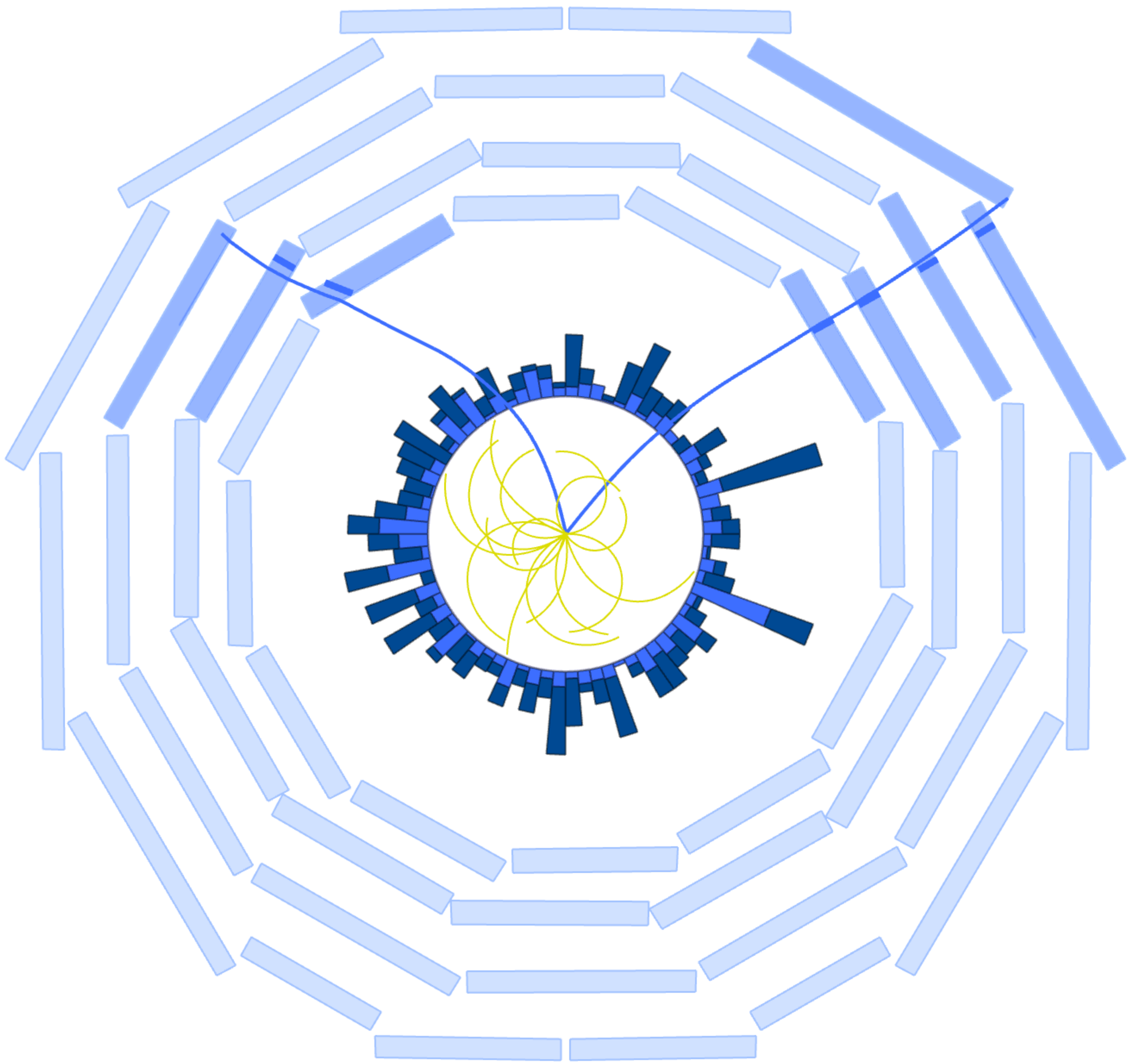
Why $B_{(S)} \to \mu\mu$?





- Unique rare $b \to s\ell\ell$ processes
  - Sensitive to New Physics effects
  - Non perturbative hadronic contributions enter via B(s) decay constant
    - Well known from Lattice QCD
- Theory uncertainty
  - Dominant contribution comes from uncertainty on $|V_{cb}|$
    - Difference between inclusive and exclusive measurements
      - Buras, Venturini, arXiv:2203:11960
- Sensitive to a subset of effective interactions associated with the rate B anomalies
  - Critical for understanding of the anomalies

Event Selection

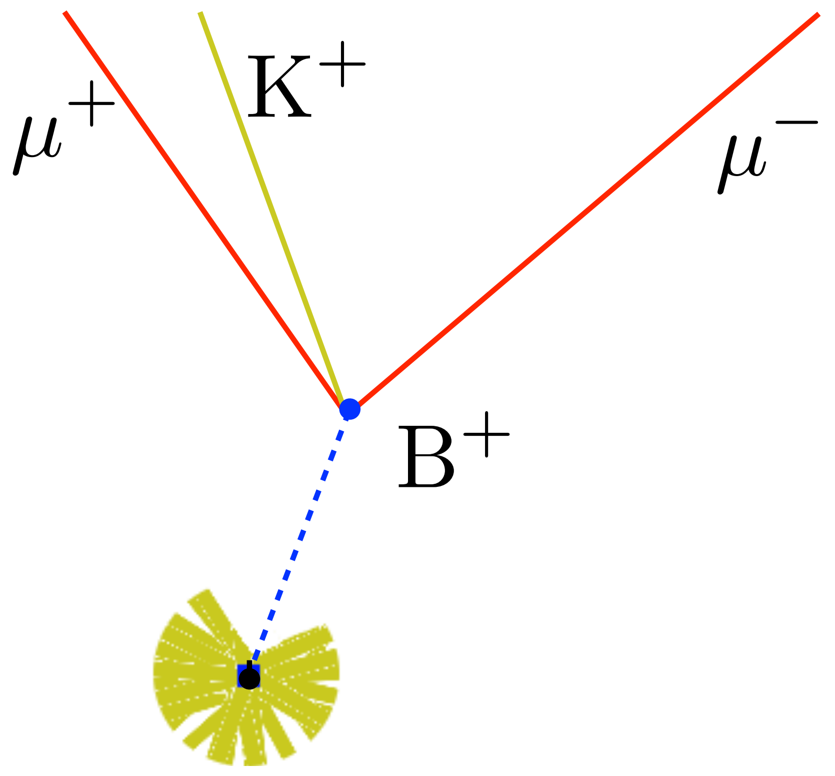


Selection	$B_s^0 \to \mu^+\mu^-$	$B^+ \to J/\psi K^+$	$B_s^0 \to J/\psi\phi$
B candidate mass [GeV]	[4.90,5.90]	[4.90,5.90]	[4.90,5.90]
Blinding window [GeV]	[5.15,5.50]		
$p_{T\mu}$ [GeV]	> 4	> 4	> 4
$\lvert\eta_\mu\rvert$	< 1.4	< 1.4	< 1.4
3D SV displacement significance	> 6	> 4	> 4
$p_{T\mu\mu}$ [GeV]	> 5	> 7	> 7
$\mu\mu$ SV probability	> 0.025	> 0.1	> 0.1
J/ψ candidate mass [GeV]		[2.9,3.3]	[2.9,3.3]
Kaon p_T [GeV]		> 1	> 1
Mass-constrained fit probability		> 0.025	> 0.025
2D $\mu\mu$ pointing angle [rad]		< 0.4	< 0.4
ϕ candidate mass [GeV]			[1.01, 1.03]

- Selection requirements are as loose as possible
  - Provide more data to MultiVariate Analysis (MVA)
  - Limited by trigger requirements
- Normalization channel selection is optimized to match kinematics of signal

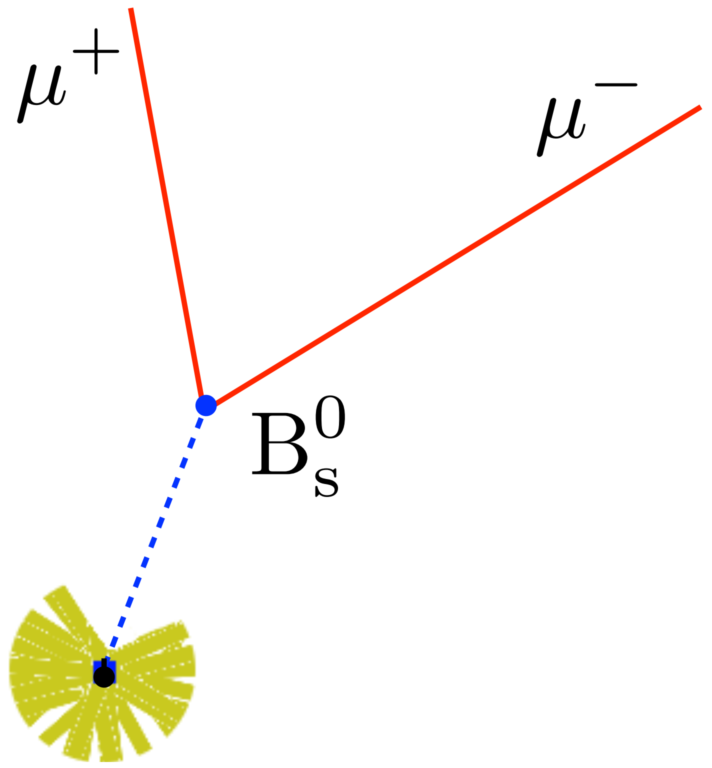
Dominant Backgrounds

3-body and partial decays

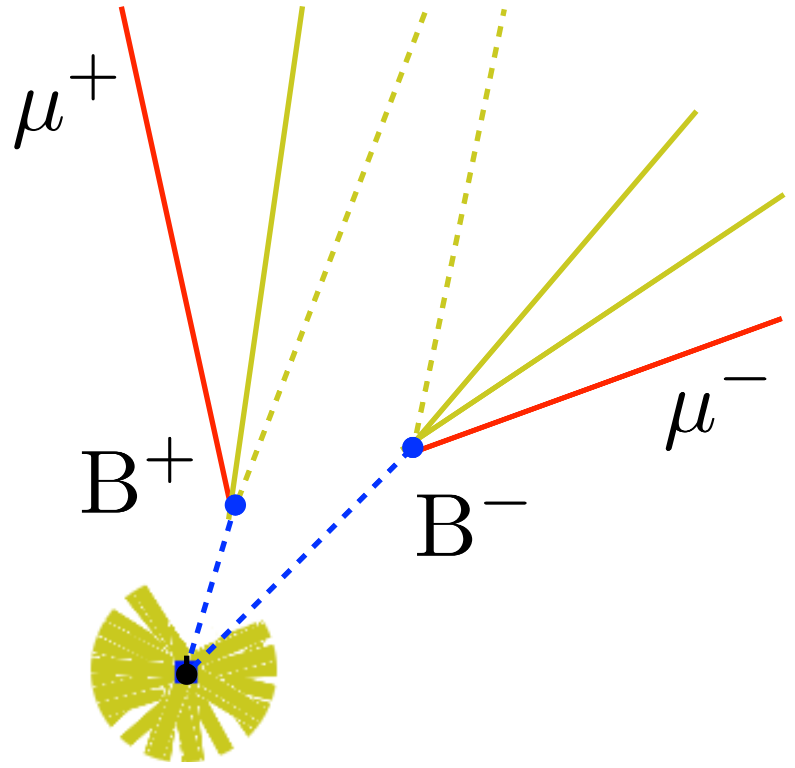


Muons from the same B hadron

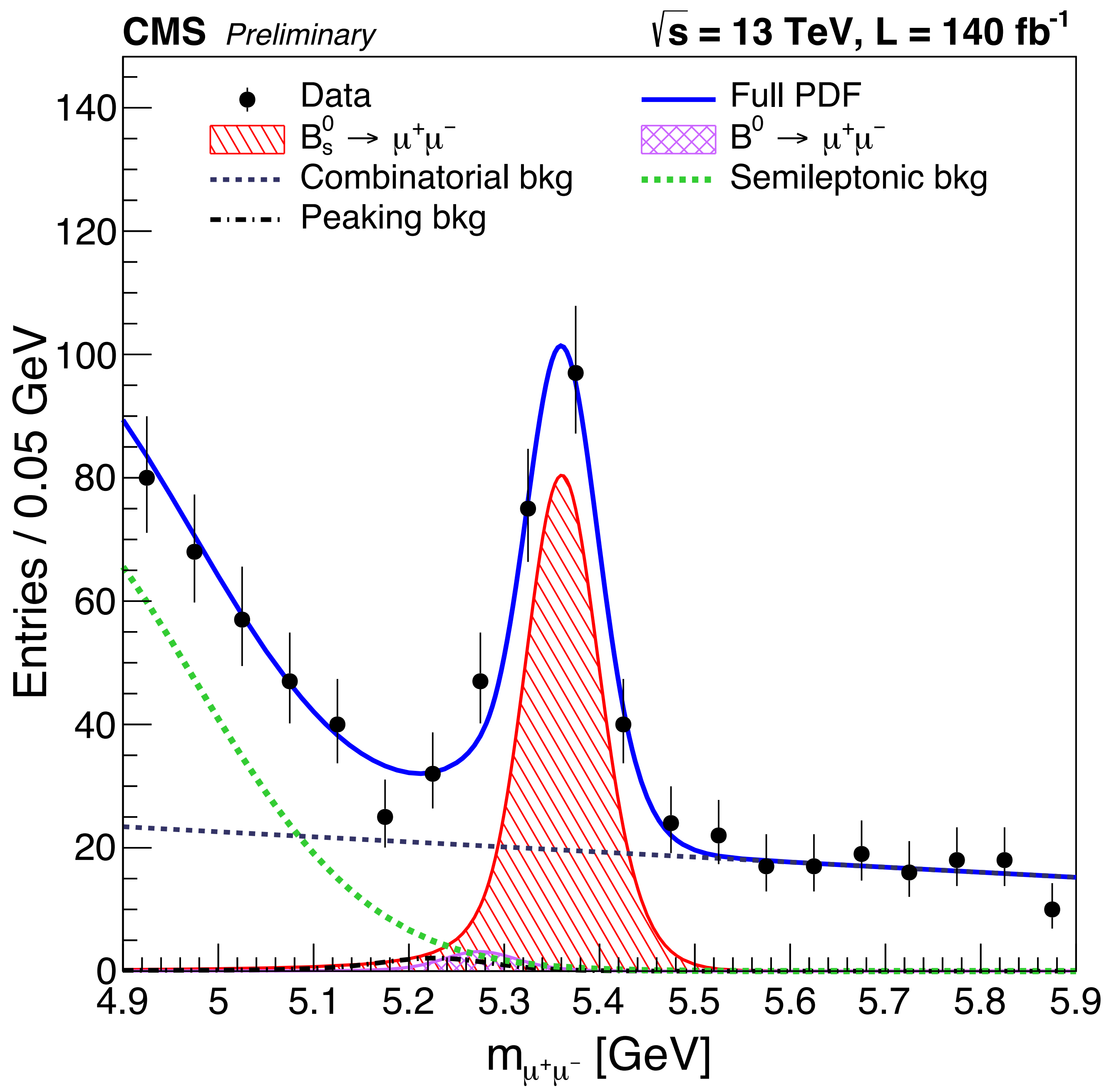
Signal $B_s \to \mu\mu$



Combinatorial Background

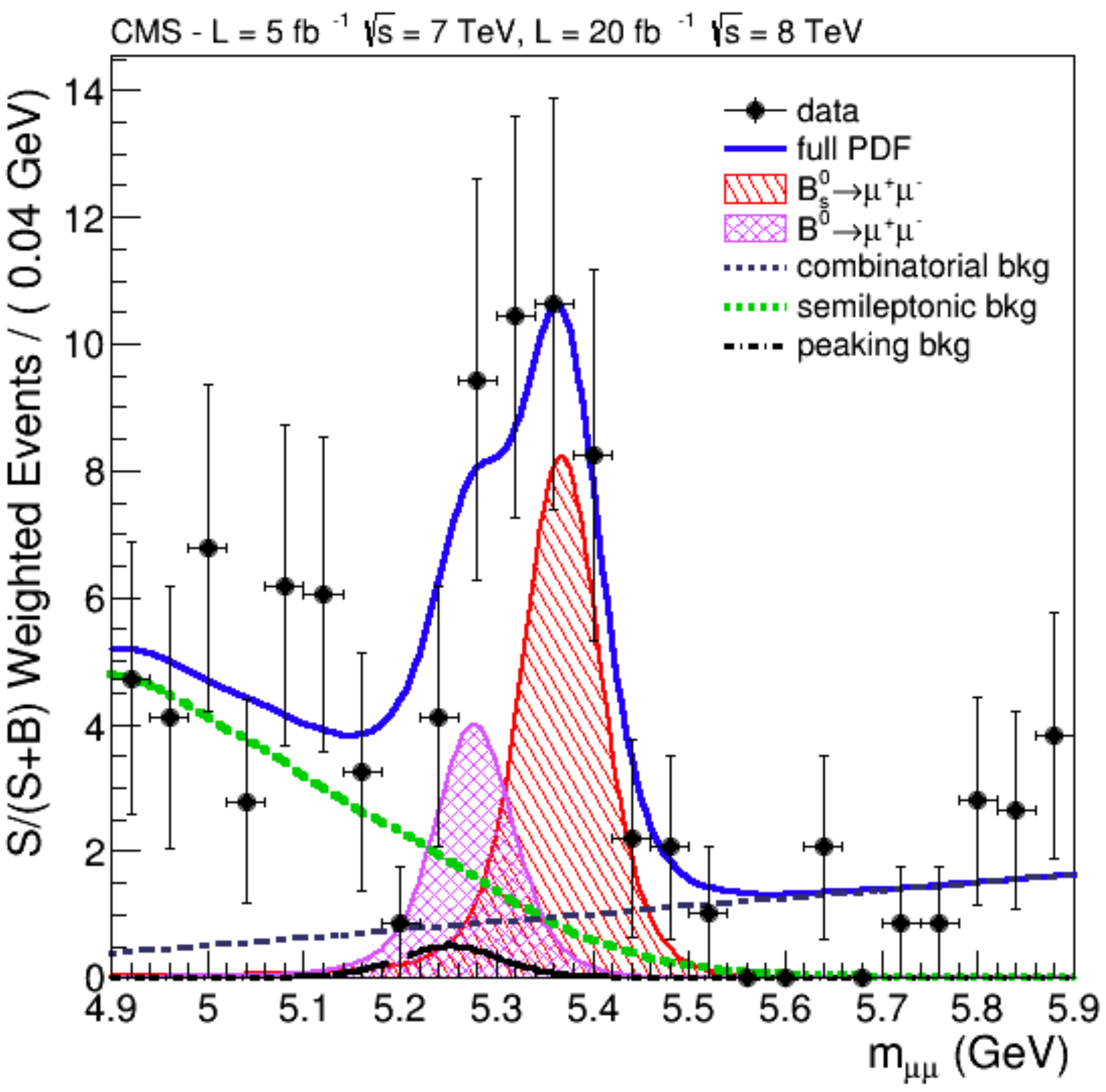


Muons originate from different B hadrons



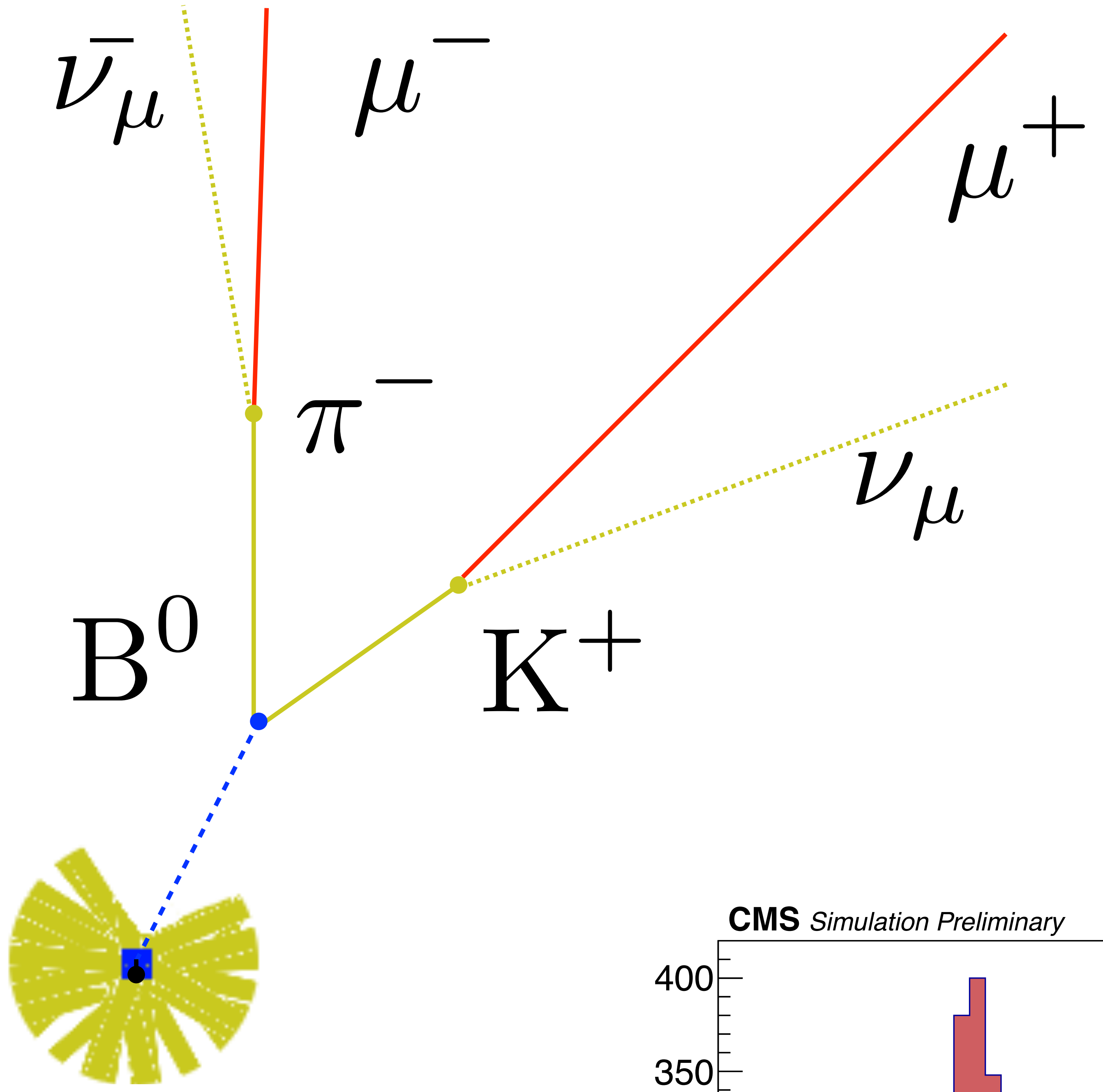
Charmless Two-body Background

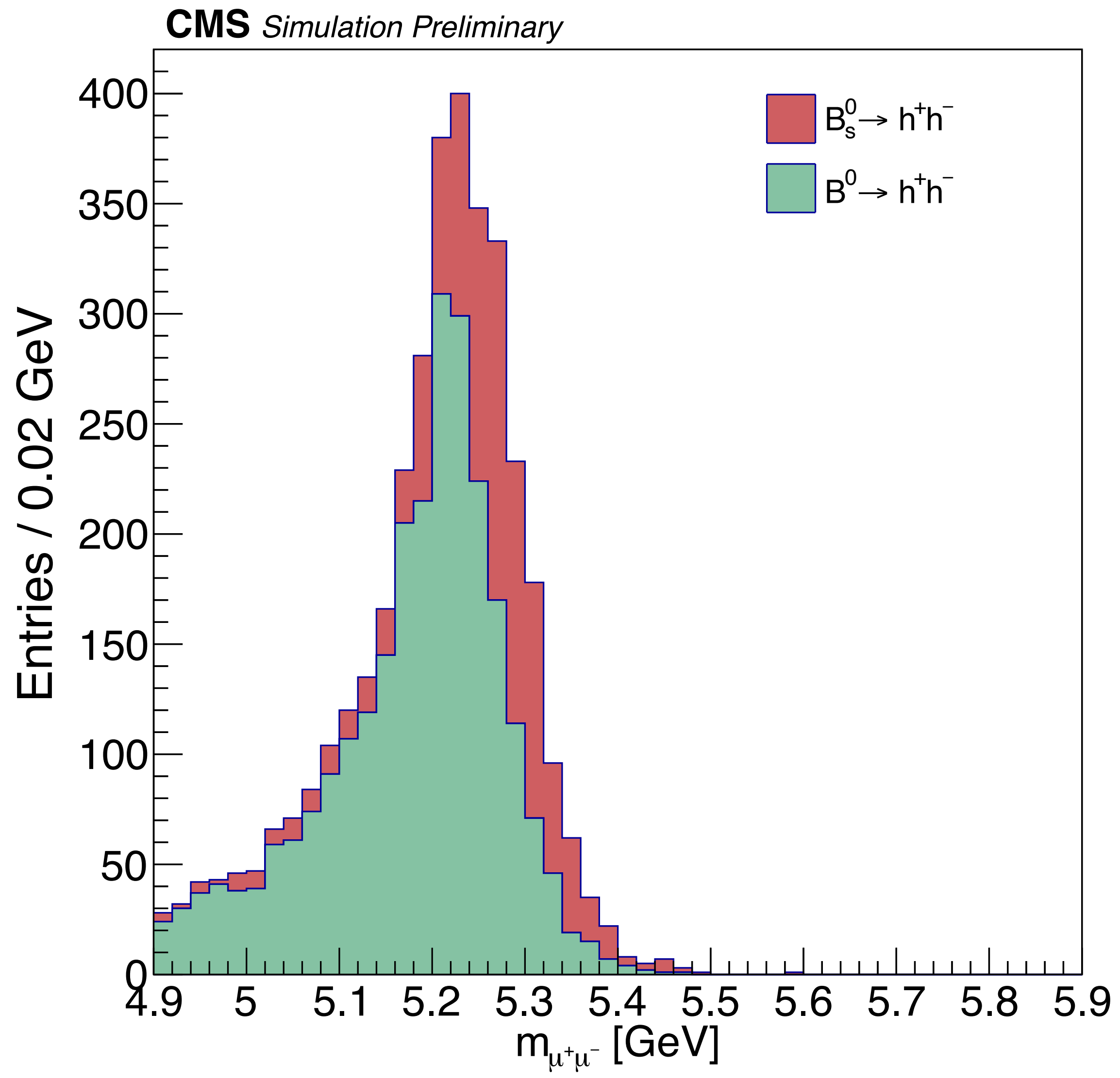
CMS Run1



- Run1 publication with 25/fb of data had a hint of very large $B^0 \to \mu\mu$ signal
  - It was just a 2σ anomaly, which happens quite often
- The result was questioned
  - How do we know it's not just B→hh?
    - Charmless two-body decays with hadrons reconstructed as muons
- Next publication with 61/fb of data used a new MVA based muon id to suppress the background as much as possible
  - Lower signal efficiency

Muon Fakes

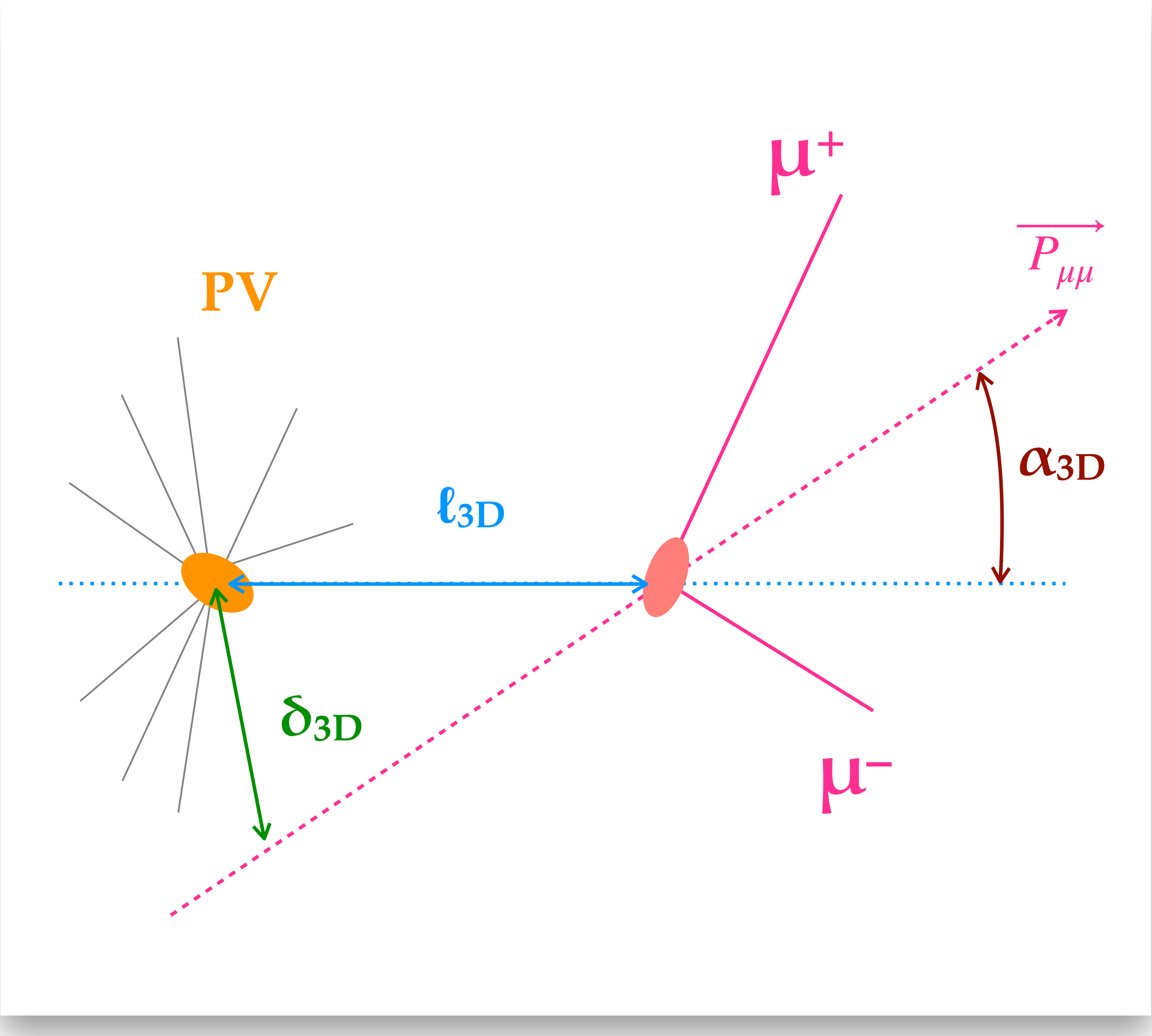




- Double muon fakes from B→hh - non-trivial background
  - Looks like signal
  - Rate is comparable to $B^0 \to \mu\mu$
  - B→Kπ and $B_s \to KK$ are dominant contribution
- Primary sources of fakes
  - Pion and kaon decays in flight to muon and neutrino
  - Other contributions are negligible and easy to reject
- Used MVA based muon identification
  - Detect minor imperfections in the muon candidate trajectory
  - Factor of 2-3 better rejection of fakes than the standard muon selection
    - Kaon decays are easier to reject
- Fake rates are measured in $K_s \to \pi\pi$ and $\phi \to KK$ control samples
  - Simulated reasonably well: ~25% systematic per hadron

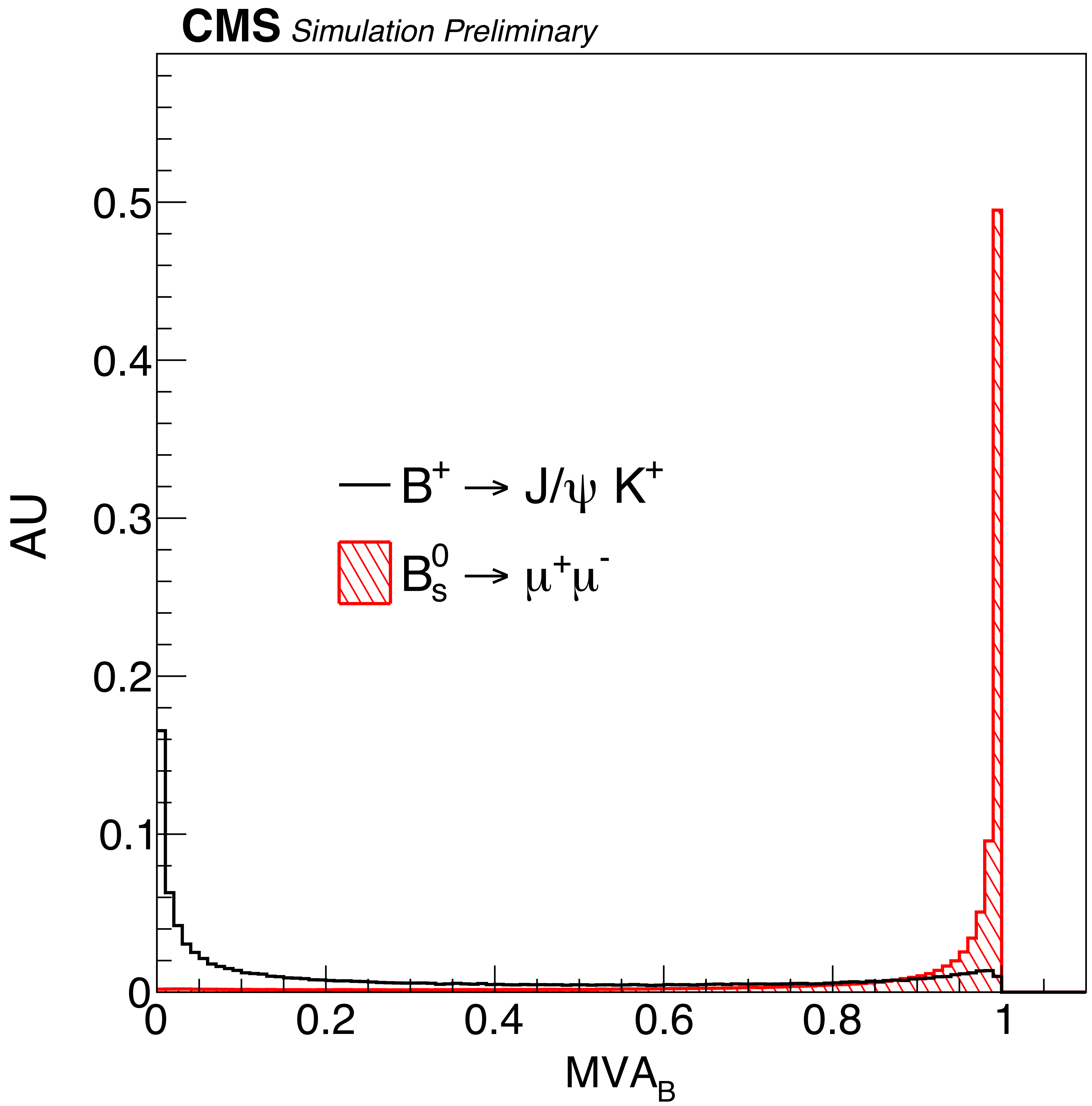
Multivariate Analysis

Multivariate Analysis



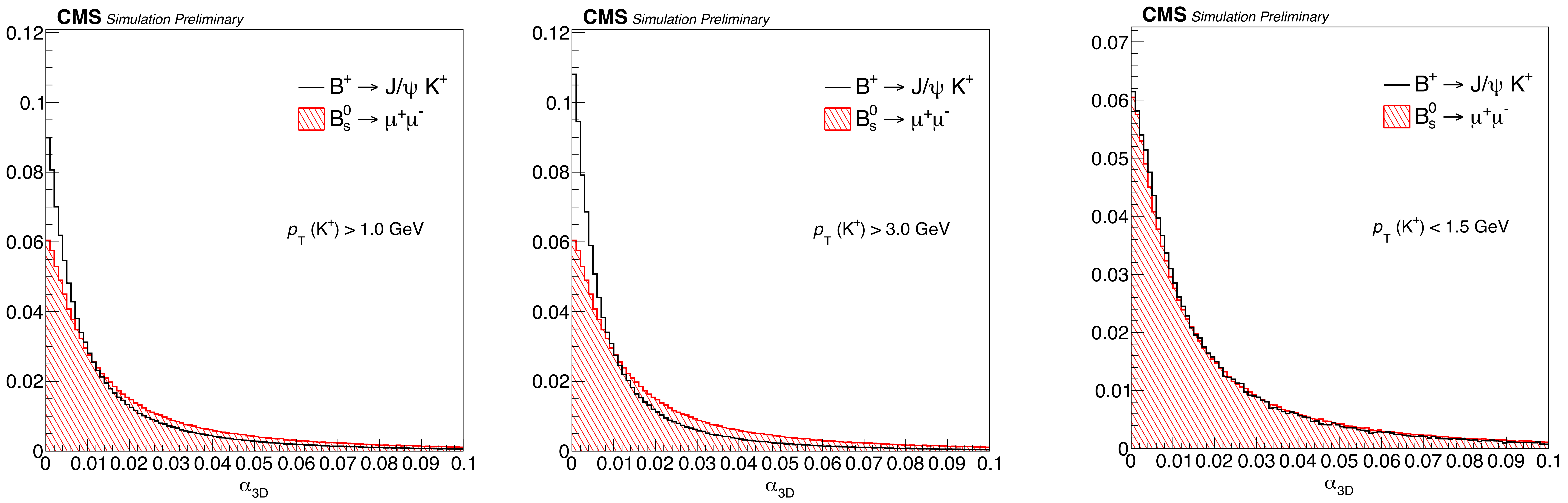
- New multivariate analysis (MVA_B) used to suppress the dominant backgrounds
  - Trained with signal MC and mass sideband data with the XGBoost package (advanced gradient boosting algorithm)
- Most discriminating variables
  - Pointing angles: α_{2D}, α_{3D}
  - Impact parameter and its significance: δ_{3D}, $\delta_{3D}/\sigma(\delta_{3D})$
  - Flight length and its significance: $\ell_{3D}/\sigma(\ell_{3D})$
  - Isolation for B candidate and muons
  - Dimuon vertex quality

MVA in Data



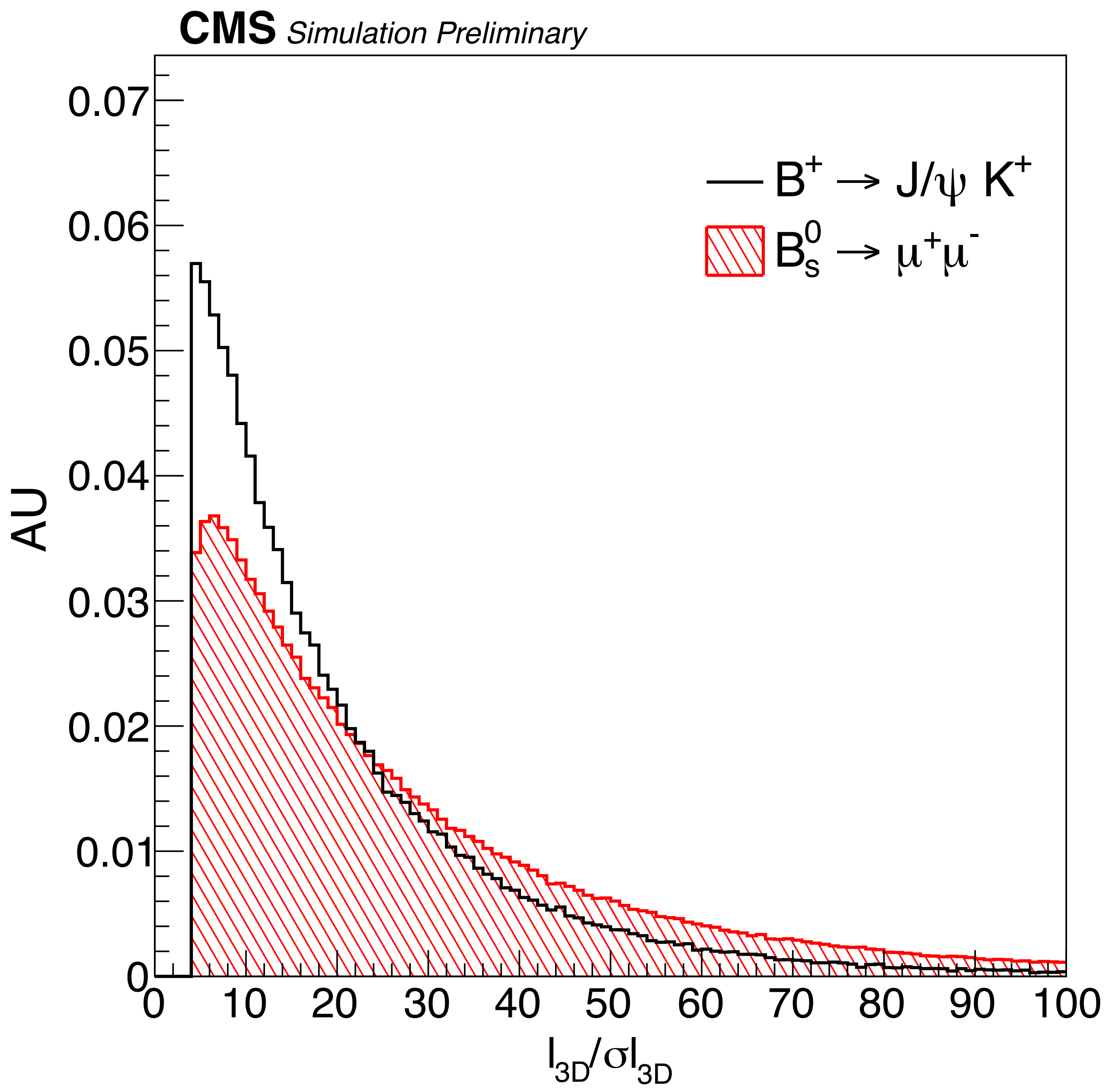
- MVA mismodeling can be a major source of systematics
  - Need a data control sample
  - $B^+ \to J/\psi K^+$ is the best candidate
- MVA is trained to rejected μμK events
  - Extra track, wrong pointing angle etc
- Need to use correct input to get signal-like response
  - μμK: pointing angle, impact parameter
  - μμ: vertex probability, displacement, isolation (ignore kaon)

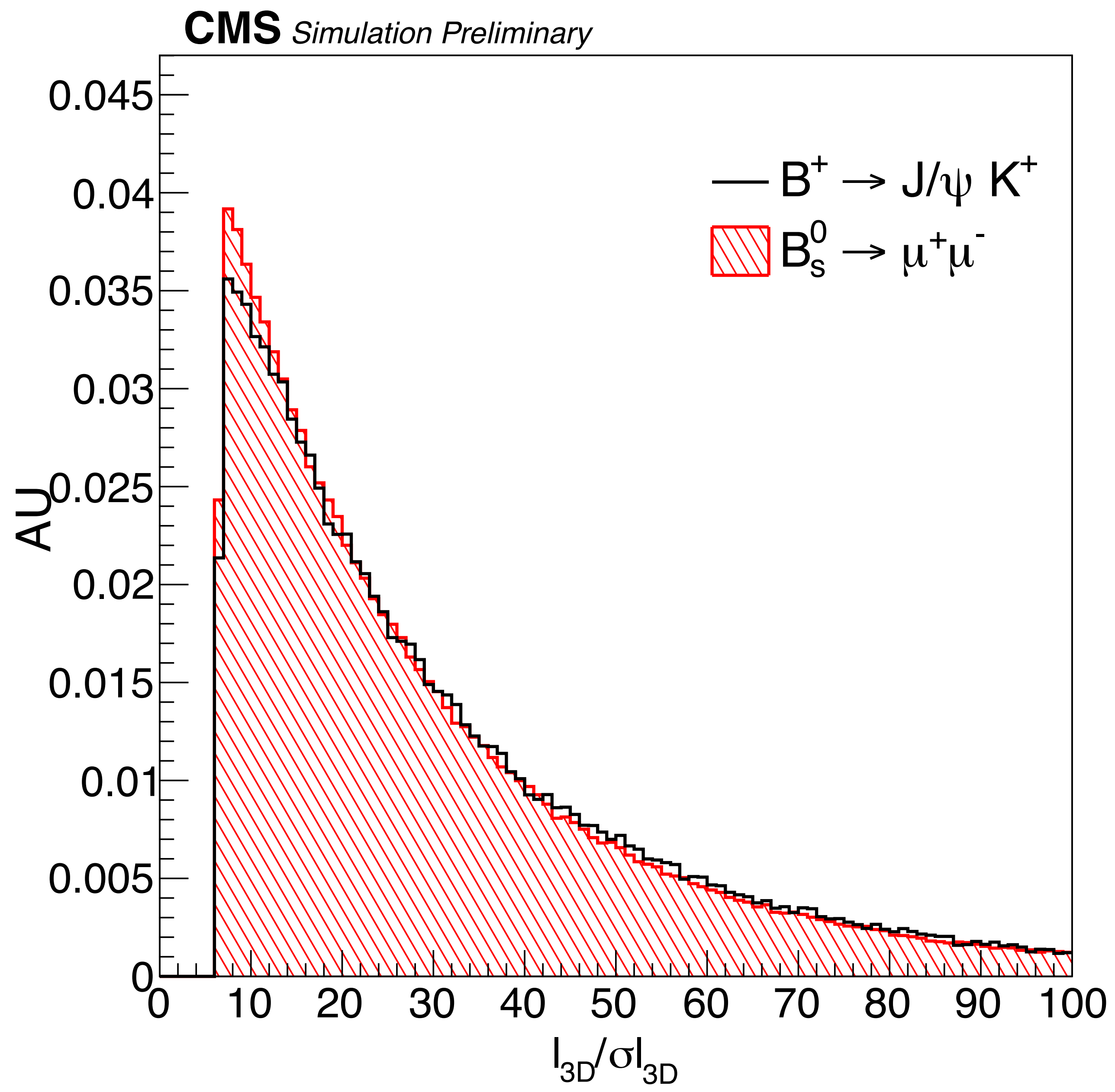
$B^+ \to J/\psi K^+$ Event Selection



- The pointing angle is one of the most powerful discriminators
- The distribution has a non-trivial dependence on the decay kinematics
  - Need to match $B^+ \to J/\psi K^+$ phase space to $B_s \to \mu\mu$
- The agreement between α(μμ) and α(μμK) is getting better when the kaon is soft
  - Leads to larger background, but it's manageable thanks to large number of events
    - Use sPlot technique to extract distributions in Data

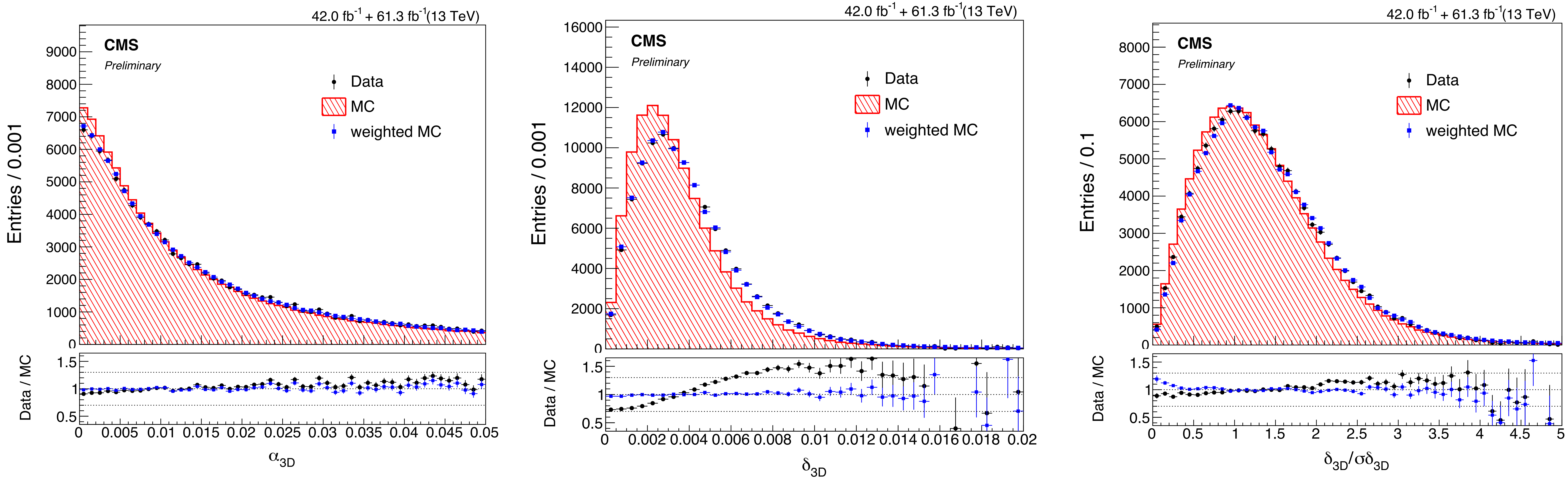
Flight Length Significance





- Soft kaon requirement gives matching distributions for all input variables but the flight length significance
- It correlates with the dimuon mass
  - Smaller mass → smaller opening angle → larger uncertainty on vertex position along the trajectory → smaller significance
- Scaling flight length significance by 1.6 provides a decent matching

XGBoost Reweighting



- With the special $B^+ \to J/\psi K^+$ control sample we can correct for mismodeling in MC simulation
- Trained XGBoost classifier on MC vs Data with sWeights
  - Capture the difference between MC and Data
  - Use $w = \frac{\mathrm{Prob}_{\mathrm{Data}}}{\mathrm{Prob}_{\mathrm{MC}}}$ as a weight to reweight $B_{(s)} \to \mu\mu$ MC samples

Branching Fraction Measurements

Branching Fraction Measurement

$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \mathcal{B}(B^+ \to J/\psi K^+) \times \frac{N_{B_s^0\to\mu^+\mu^-}}{N_{B^+\to J/\psi K^+}} \times \frac{\epsilon_{B^+\to J/\psi K^+}}{\epsilon_{B_s^0\to\mu^+\mu^-}} \times \frac{f_u}{f_s}$$

external B production fraction ratio

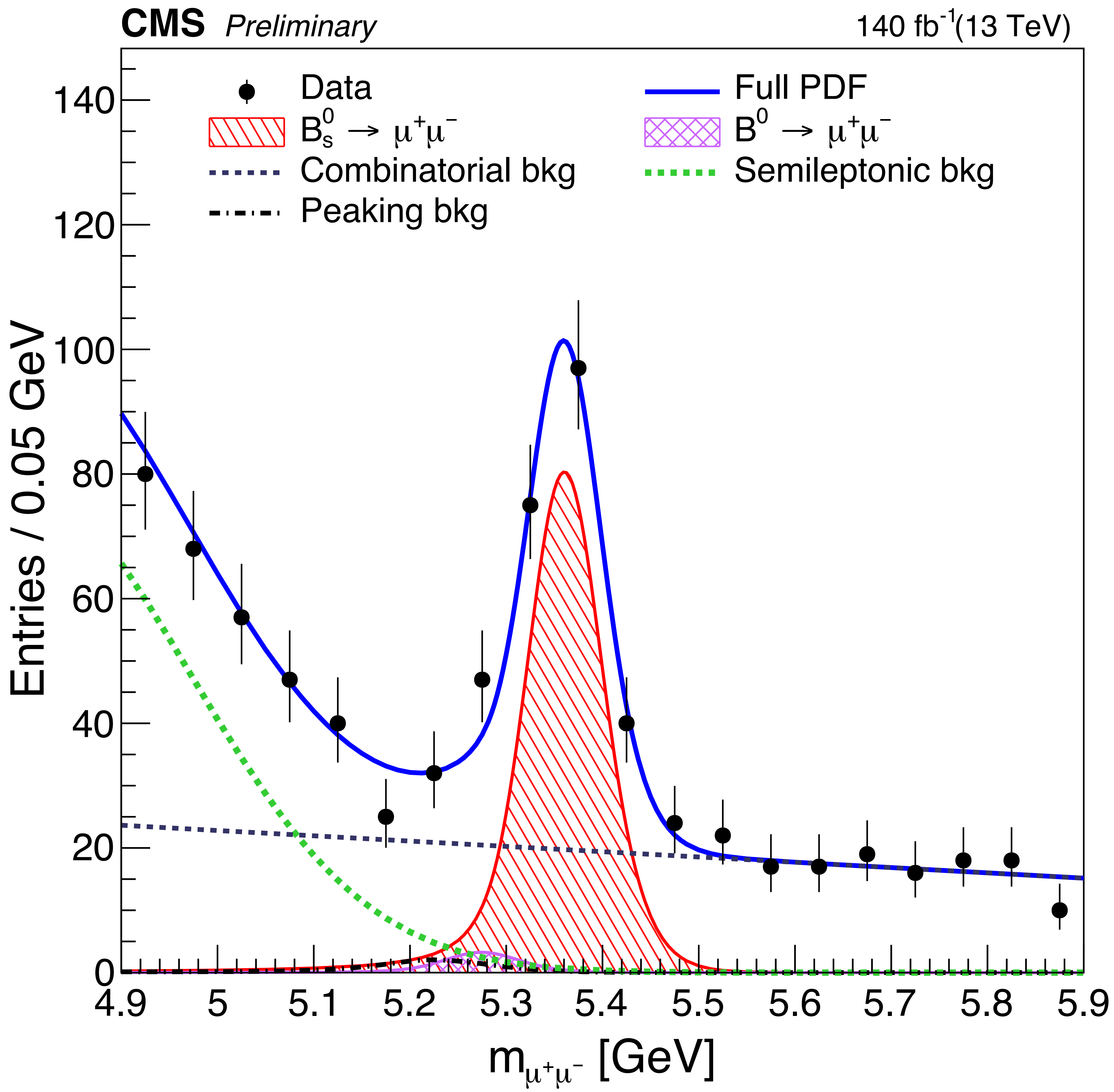
$$or \left\{= \mathcal{B}(B_s^0 \to J/\psi\phi) \times \frac{N_{B_s^0\to\mu^+\mu^-}}{N_{B_s^0\to J/\psi\phi}} \times \frac{\epsilon_{B_s^0\to J/\psi\phi}}{\epsilon_{B_s^0\to\mu^+\mu^-}}\right\}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = \mathcal{B}(B^+ \to J/\psi K^+) \times \frac{N_{B^0\to\mu^+\mu^-}}{N_{B^+\to J/\psi K^+}} \times \frac{\epsilon_{B^+\to J/\psi K^+}}{\epsilon_{B^0\to\mu^+\mu^-}} \times \frac{f_u}{f_d}$$

=1

- Branching Fraction are normalized using $B^+ \to J/\psi K^+$ and $B_s \to J/\psi\phi$ decays
  - Most systematic effects cancel in the ratio
  - $B^+ \to J/\psi K^+$ provides nominal normalization
    - $B_s \to J/\psi\phi$ - alternative normalization with different systematics that can benefit from independent BF measurement of $B_s \to J/\psi\phi$

Unbinned Maximum Likelihood Fit



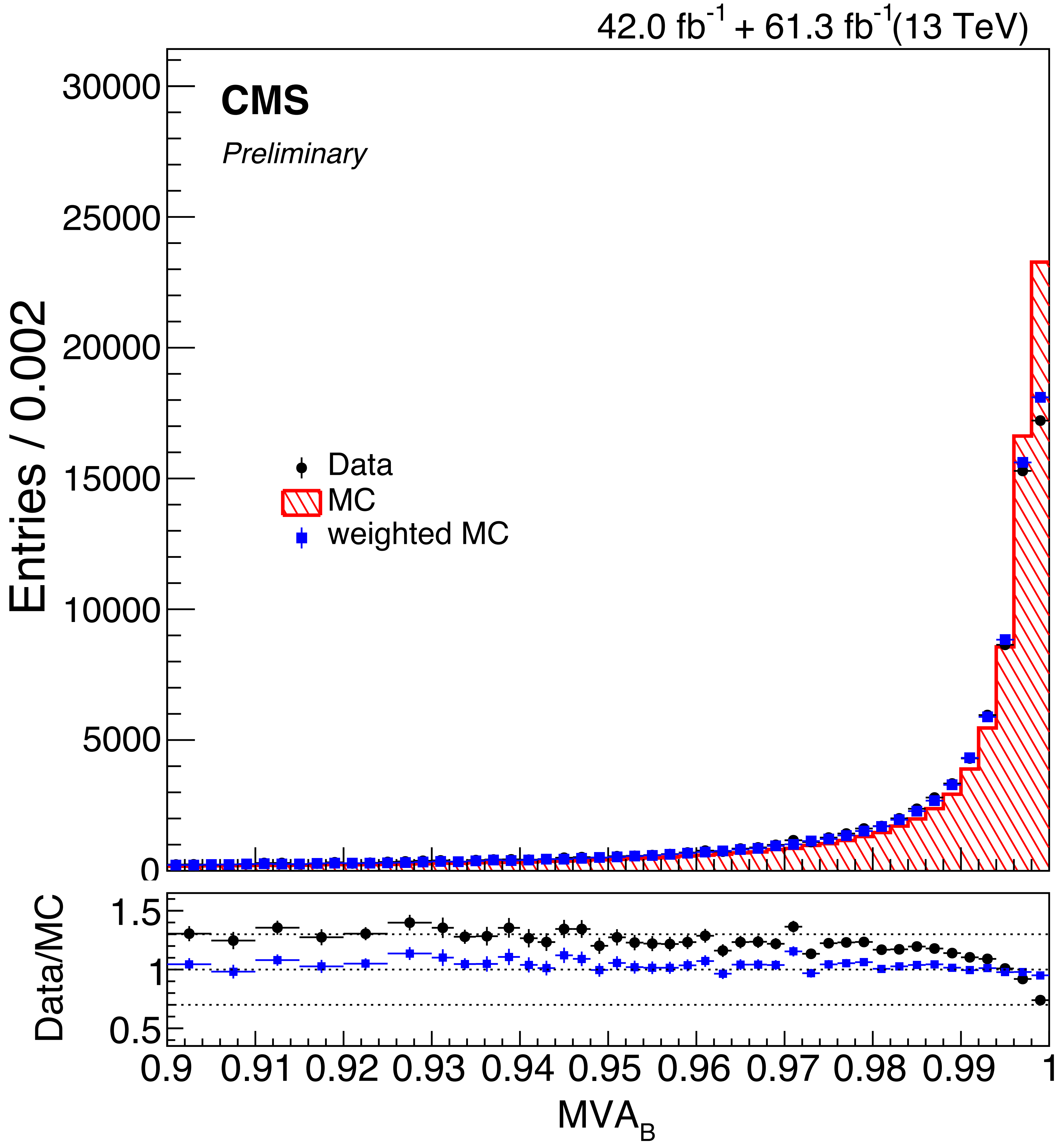
- 2D unbinned ML fit in mass and its uncertainty
- 16 categories
  - rapidity of forward muon: [0,0.7] and [0.7,1.4]
  - data taking period: 2016a, 2016b, 2017, 2018
  - MVA: loose and tight
- Shapes
  - Signal - fixed CrystalBall with mass scale and resolution corrected using J/ψ and ϒ(1S)
  - Combinatorial - 1st order polynomial with free floating parameters
  - Semileptonic - Gaussian with floating parameters

Postfit Event Yields by Category

Data set	Channel	$N(B_s^0)$	$N(B^0)$	$N(\text{comb})$	$N(\text{peak})$	$N(\text{semi})$	N(total)	Data
				$MVA_B > 0.99$				
2016a	0	5.3	0.2	2.8	0.2	6.0	14.5	16
2016a	1	9.4	0.4	16.2	0.4	9.9	36.3	35
2016b	0	6.3	0.3	1.7	0.2	7.9	16.4	12
2016b	1	9.9	0.4	8.6	0.4	13.3	32.6	32
2017	0	23.5	1.0	51.4	0.8	29.6	106.3	114
2017	1	33.9	1.3	89.6	1.4	44.0	170.2	165
2018	0	34.5	1.4	64.8	1.3	38.4	140.4	143
2018	1	50.0	2.0	151.0	2.5	50.9	256.4	252
				$0.99 > MVA_B > 0.9$				
2016a	0	4.8	0.2	118.0	0.2	8.4	131.6	132
2016a	1	8.9	0.4	324.8	0.4	16.5	351.0	352
2016b	0	5.6	0.2	107.6	0.2	10.9	124.5	126
2016b	1	9.2	0.4	257.1	0.4	18.2	285.3	287
2017	0	15.2	0.6	637.7	0.7	26.4	680.6	683
2017	1	21.7	0.9	1430.5	1.1	44.3	1498.5	1498
2018	0	23.3	1.0	936.2	1.2	52.5	1014.2	1017
2018	1	34.2	1.4	2222.5	1.8	79.7	2339.6	2340

Channel 0 and 1 refer to forward muon rapidity ranges of [0.0,0.7] and [0.7,1.4]

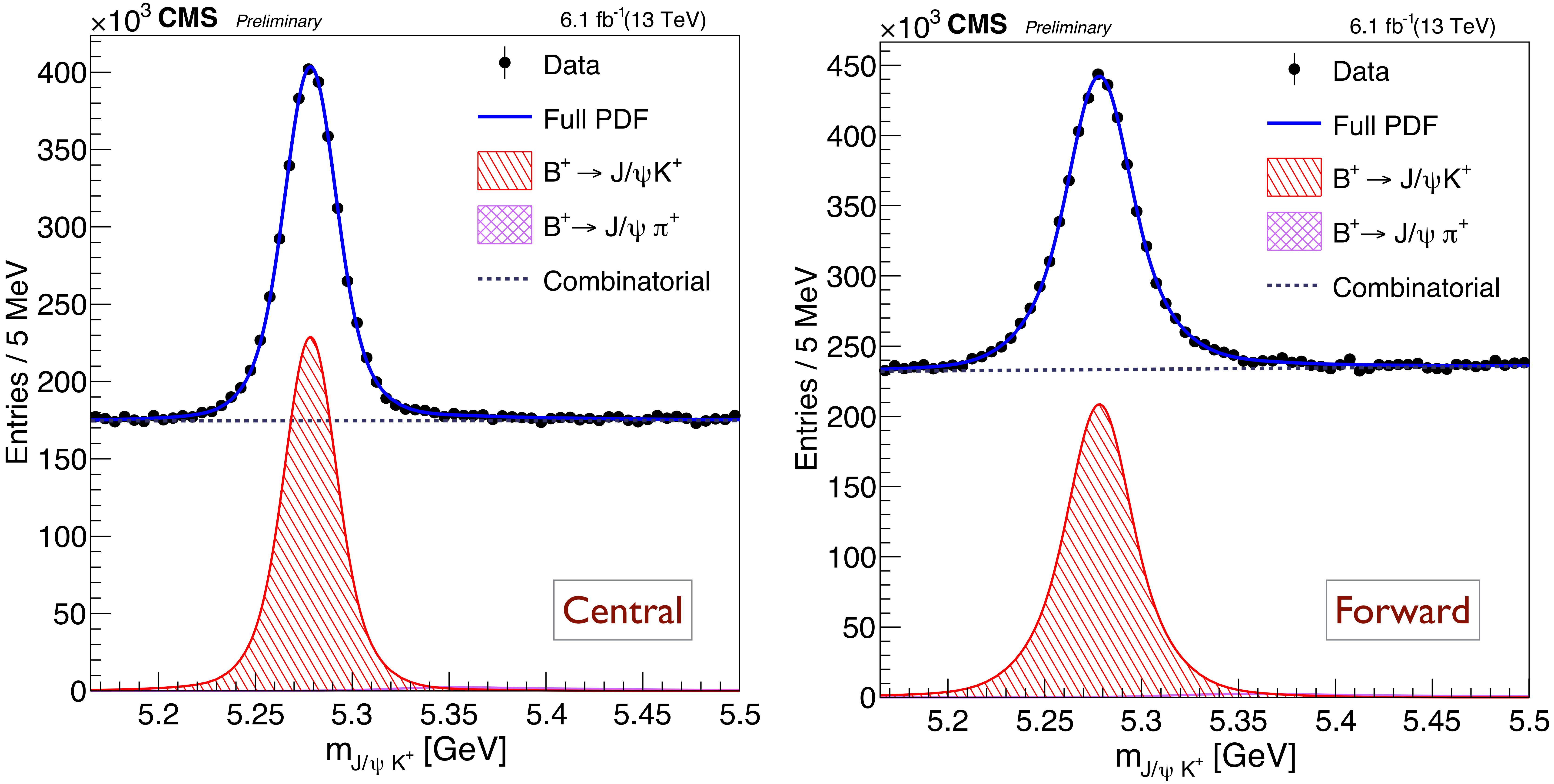
Efficiency Corrections



Method	MVA$_B$>0.9			MVA$_B$>0.99		
	2016	2017	2018	2016	2017	2018
Ratio	1.011 ± 0.013	0.939 ± 0.007	0.903 ± 0.008	1.058 ± 0.019	0.891 ± 0.008	0.885 ± 0.010
XGBoost	0.991 ± 0.008	0.949 ± 0.003	0.917 ± 0.002	1.008 ± 0.011	0.905 ± 0.004	0.908 ± 0.002

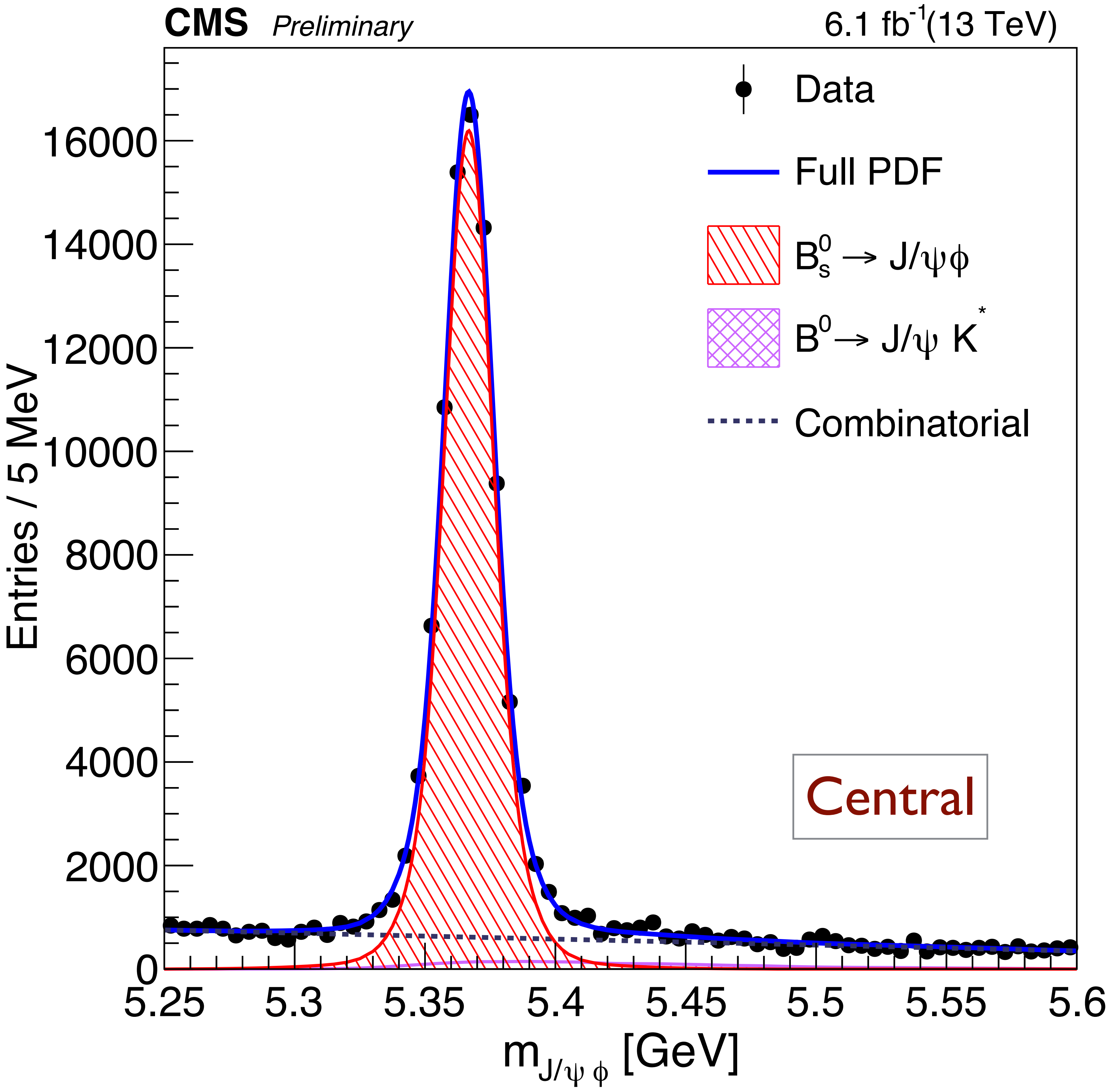
- Use $B^+ \to J/\psi K^+$ events to model $B_{(s)} \to \mu\mu$ decays in Data and MC
  - Additional selection requirements
    - Kaon P_T < 1.5 GeV
  - Rescaled flight length significance
- Data/MC correction factors
  - Ratio method: efficiency ratio between sWeighted data/MC
  - XGBoost: train a XGBoost classifier to reweight MC to match to the data

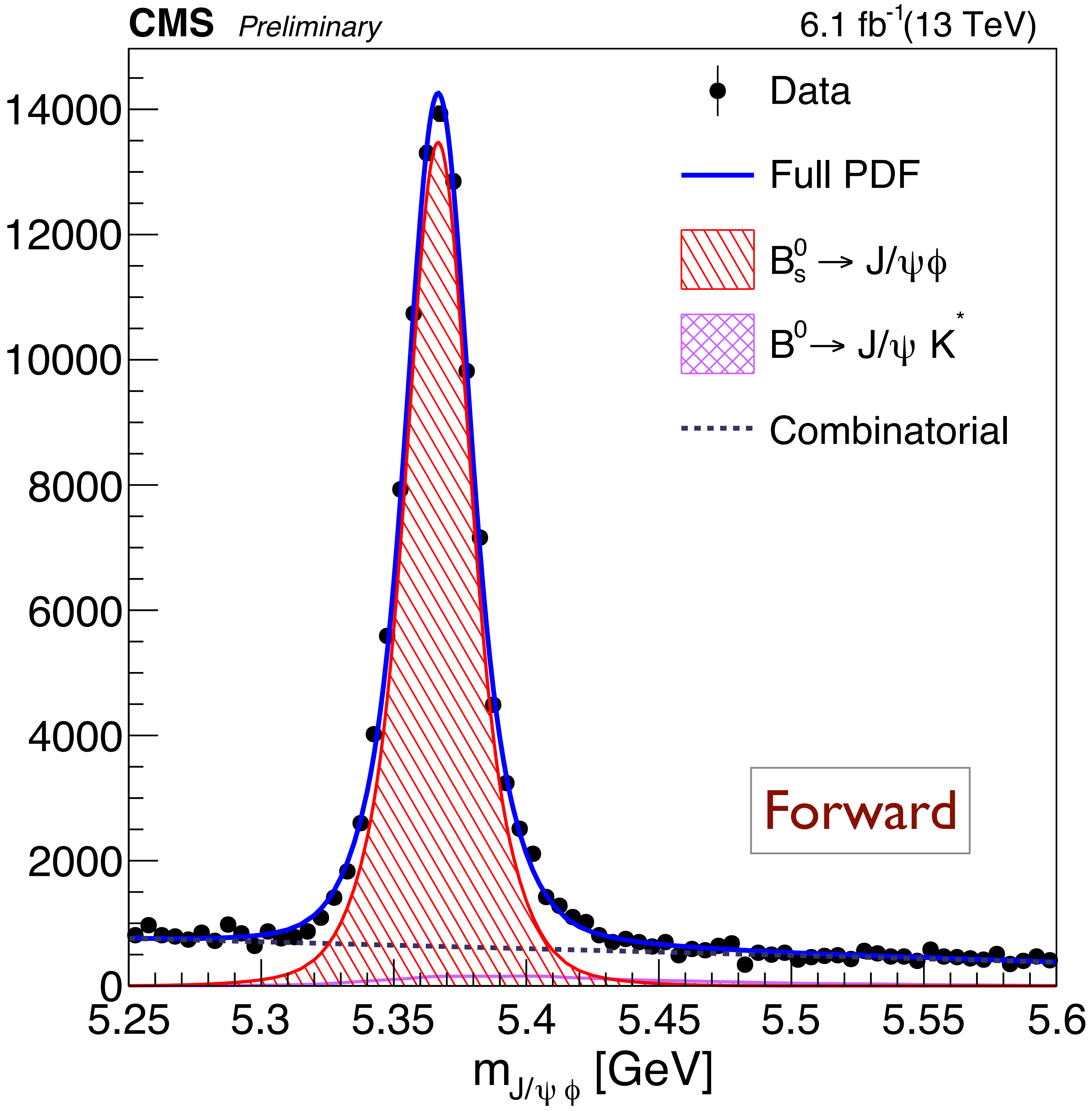
$B^+ \to J/\psi K^+$ Yield Fits



- Two different signal models are used to extract signal normalization in unbinned maximum likelihood fits
  - Nominal model is built using analytical functions
  - Alternative one is using non-parametric signal model convolved with a resolution model
- The difference between the two estimates is taken as systematics (1%)

$B_s \to J/\psi\varphi$ Yield Fits





- $B_s \to J/\psi\varphi$ is used as an alternative normalization for Bs BF
  - At the moment it's primarily a cross-check
- Data is fitted to two different signal models, taking the difference as systematics

Branching Fraction Systematics

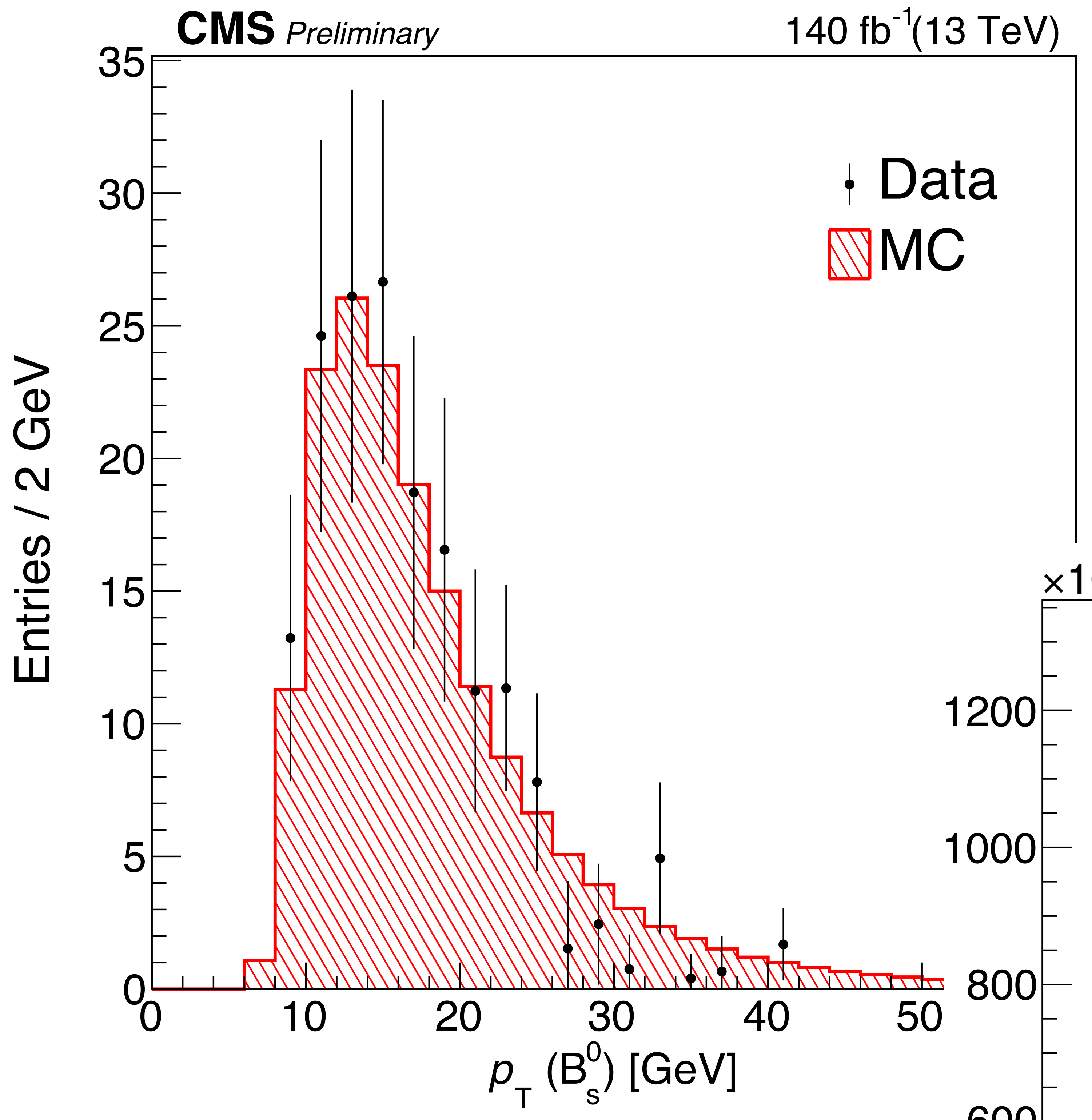
Effect	$BF(B_s \to \mu^+\mu^-)$	$BF(B^0 \to \mu^+\mu^-)$
Trigger efficiency	2–4%	
Pileup	1%	
Vertex quality	1%	
MVA_B correction	2–3%	
Tracking efficiency	2.3%	
$J/\psi K^+$ shape	1%	
Fit bias	2.2%	4.5%
f_s/f_u ratio	3.5%	-

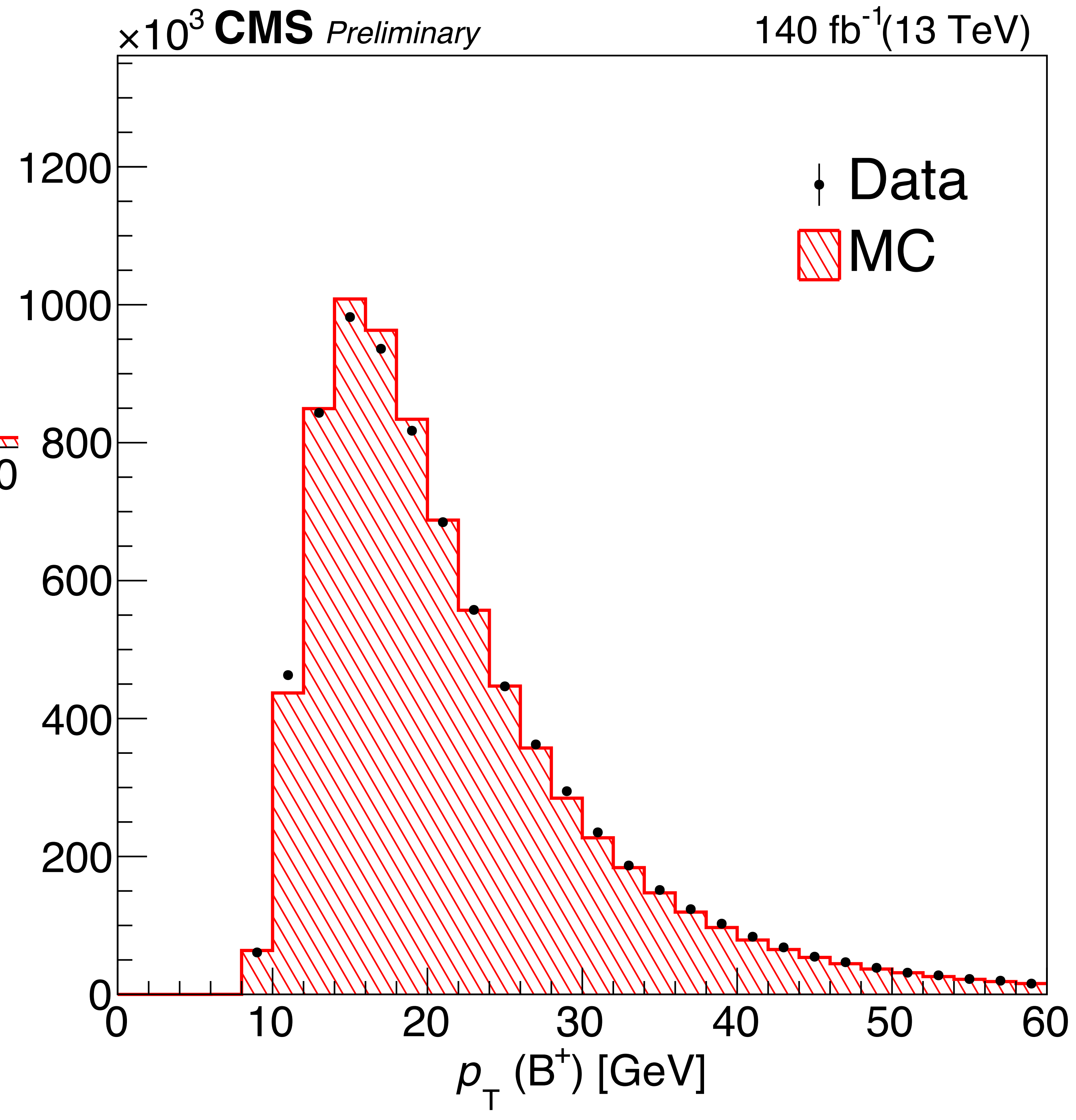
- Signal efficiency is correlated with the B_s lifetime
  - $B_s \to \mu\mu$ branching fractions are measured assuming the SM lifetime value
  - For alternative hypothesis scale BF using the following scale factor

$$\alpha_{BF} = 1.577 - 0.358\,\tau$$

  - τ is in ps
    - Example: $\alpha_{BF}(1.61) = 1.0, \quad \alpha_{BF}(1.43) = 1.065$

f_s/f_u ratio in Branching Fraction fit





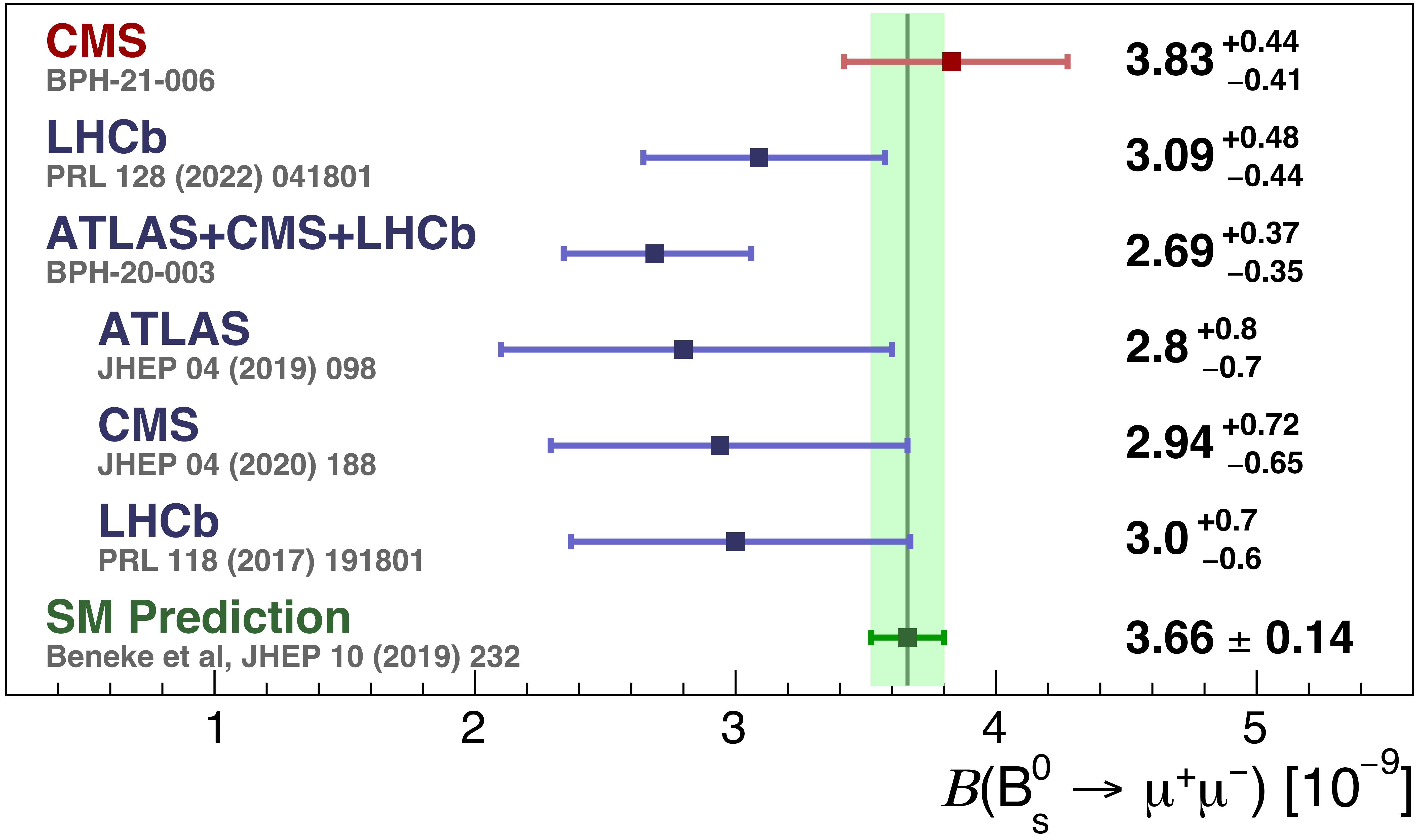
$$\mathcal{B}(B_s^0 \to \mu^+\mu^-) = \mathcal{B}(B^+ \to J/\psi K^+) \frac{N_{B_s^0\to\mu^+\mu^-}}{N_{B^+\to J/\psi K^+}} \frac{\varepsilon_{B^+\to J/\psi K^+}}{\varepsilon_{B_s^0\to\mu^+\mu^-}} \frac{f_u}{f_s},$$

- f_s/f_u = 0.231±0.008
  - Based on P_T-dependent results from LHCb
    - PRD 104 (2021) 032005
    - Integrate with the effective P_T distribution
  - Previous measurement used 0.252±0.032
- Resulting BF can be rescaled:
  - One can use a different f_s/f_u value
  - Treated as an external uncertainty
    - not as a constrained nuisance parameter

$B_s \to \mu\mu$ Branching Fraction Result

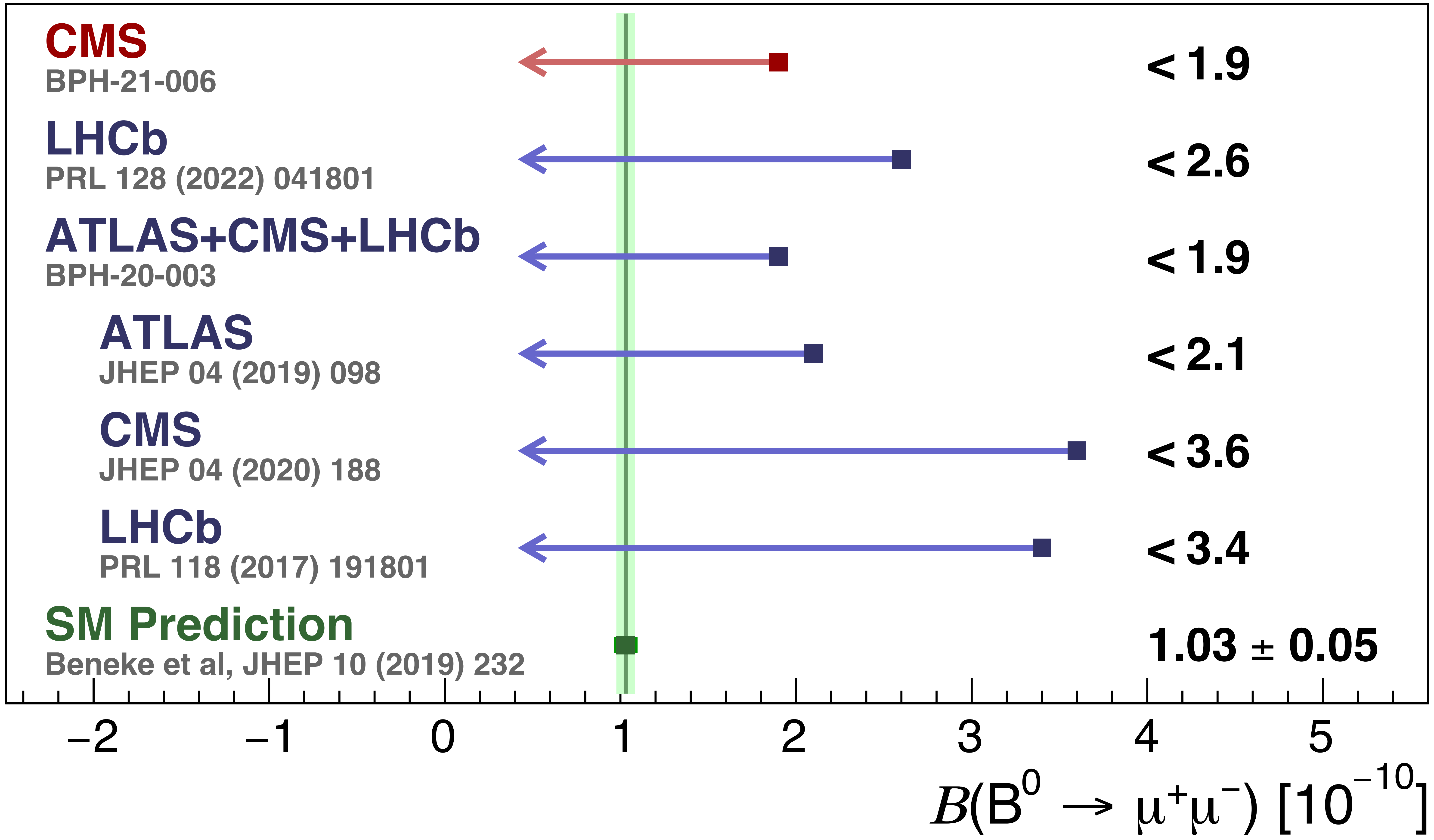
$$\mathcal{B}(B^0_s \to \mu^+\mu^-) = \left[3.83^{+0.38}_{-0.36}\ \text{(stat)}\ ^{+0.19}_{-0.16}\ \text{(syst)}\ ^{+0.14}_{-0.13}\ (f_s/f_u)\right] \times 10^{-9}$$

Alternative using $B_s \to J/\psi\phi$: $\mathcal{B}(B^0_s \to \mu^+\mu^-) = \left[3.95^{+0.39}_{-0.37}\ \text{(stat)}\ ^{+0.27}_{-0.22}\ \text{(syst)}\ ^{+0.21}_{-0.19}\ \text{(BF)}\right] \times 10^{-9}$

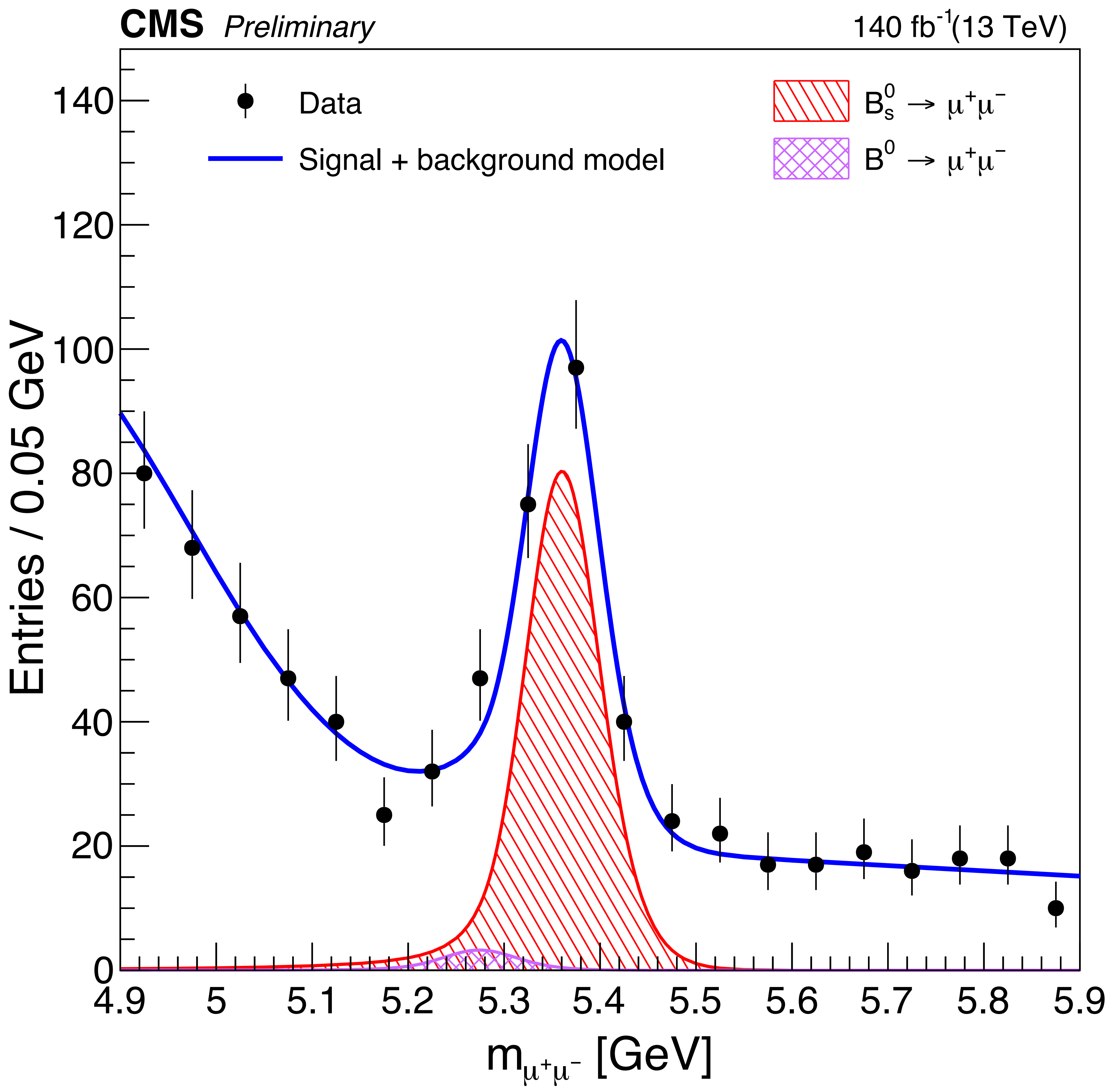


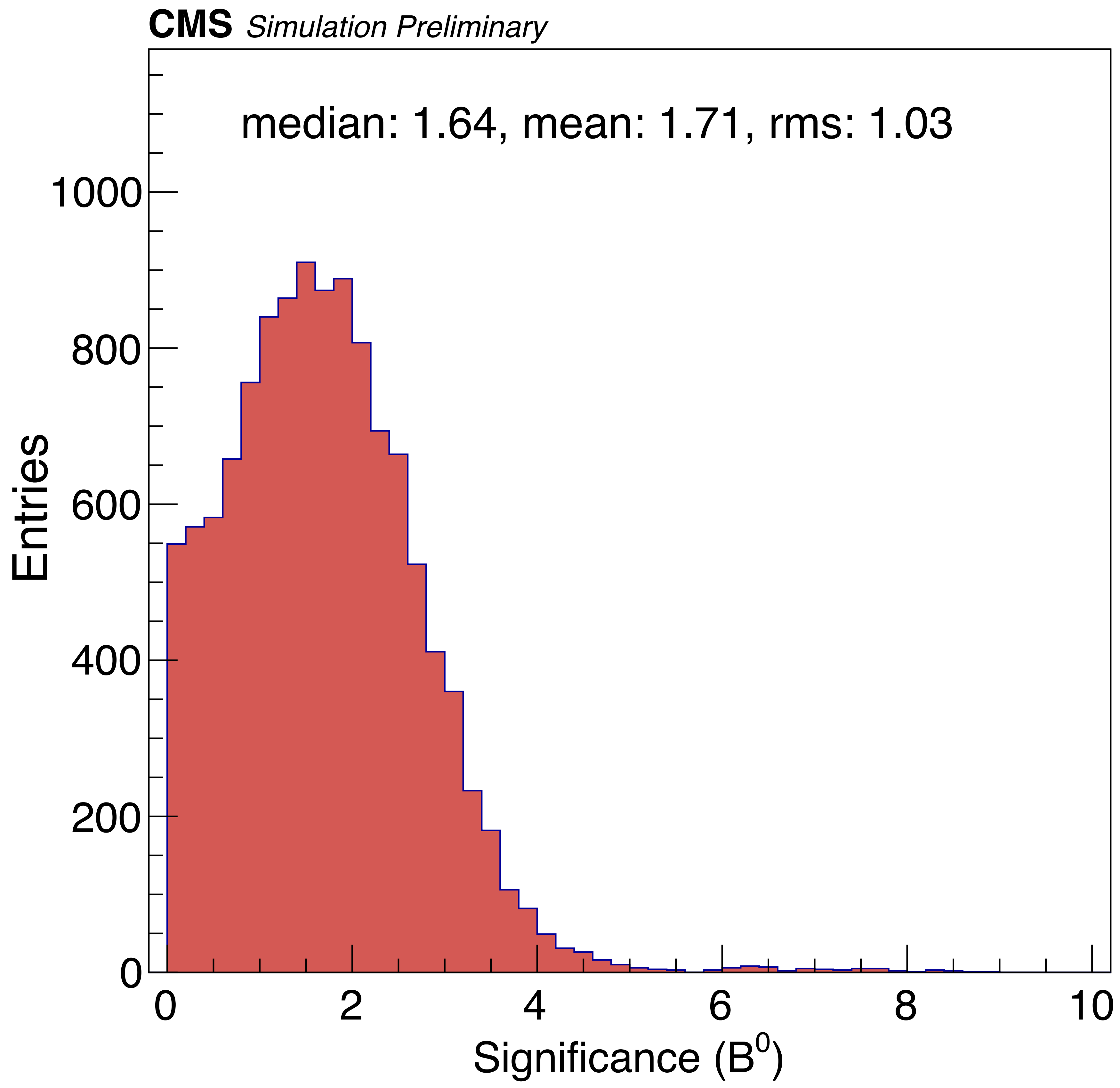
$B^0 \to \mu\mu$ Branching Fraction Result

$$\mathcal{B}(B^0 \to \mu^+\mu^-) = \left[0.37^{+0.75}_{-0.67}\ (\text{stat})\ ^{+0.08}_{-0.09}\ (\text{syst})\right] \times 10^{-10}$$



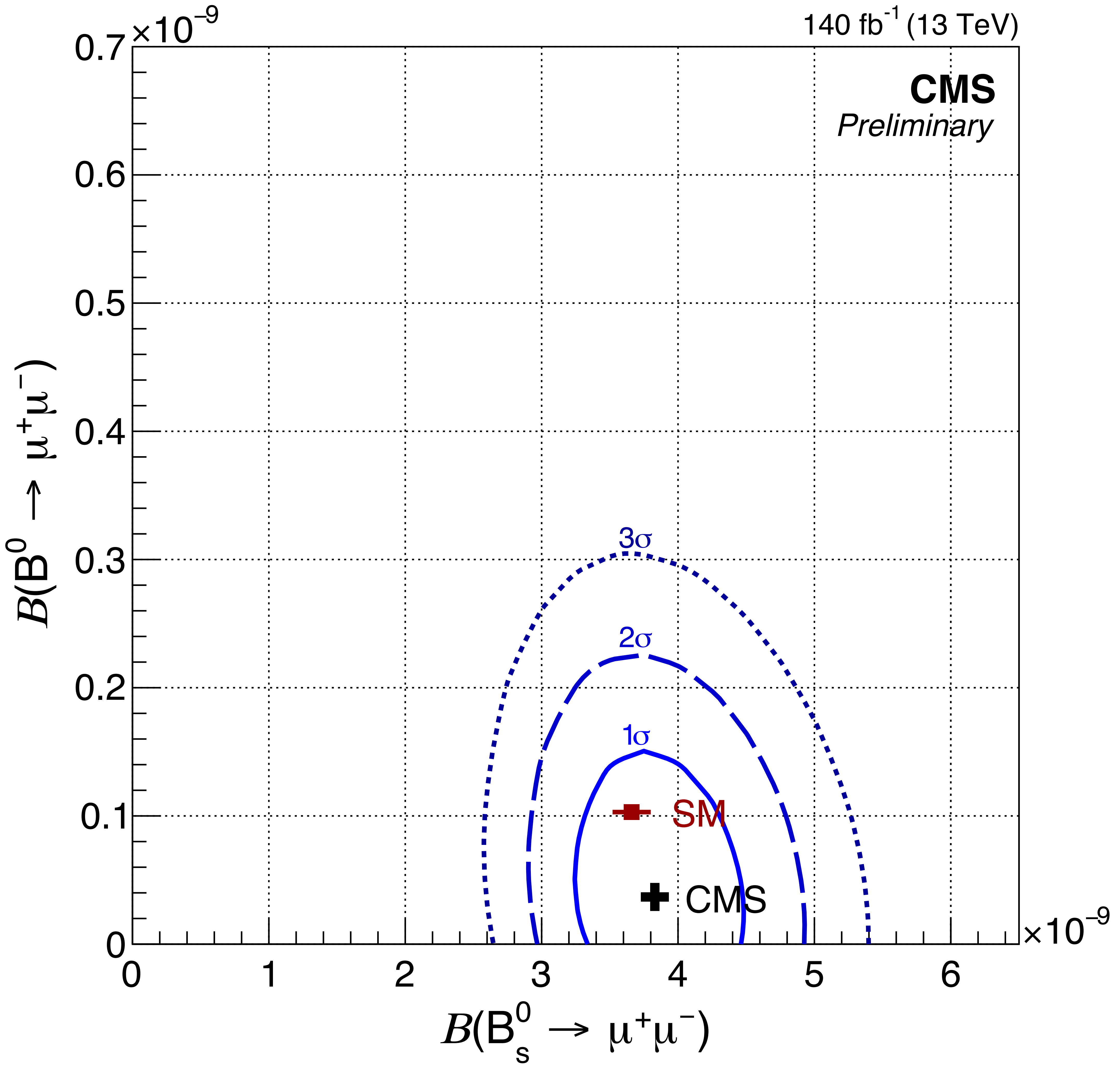
Sensitivity to $B^0 \to \mu\mu$

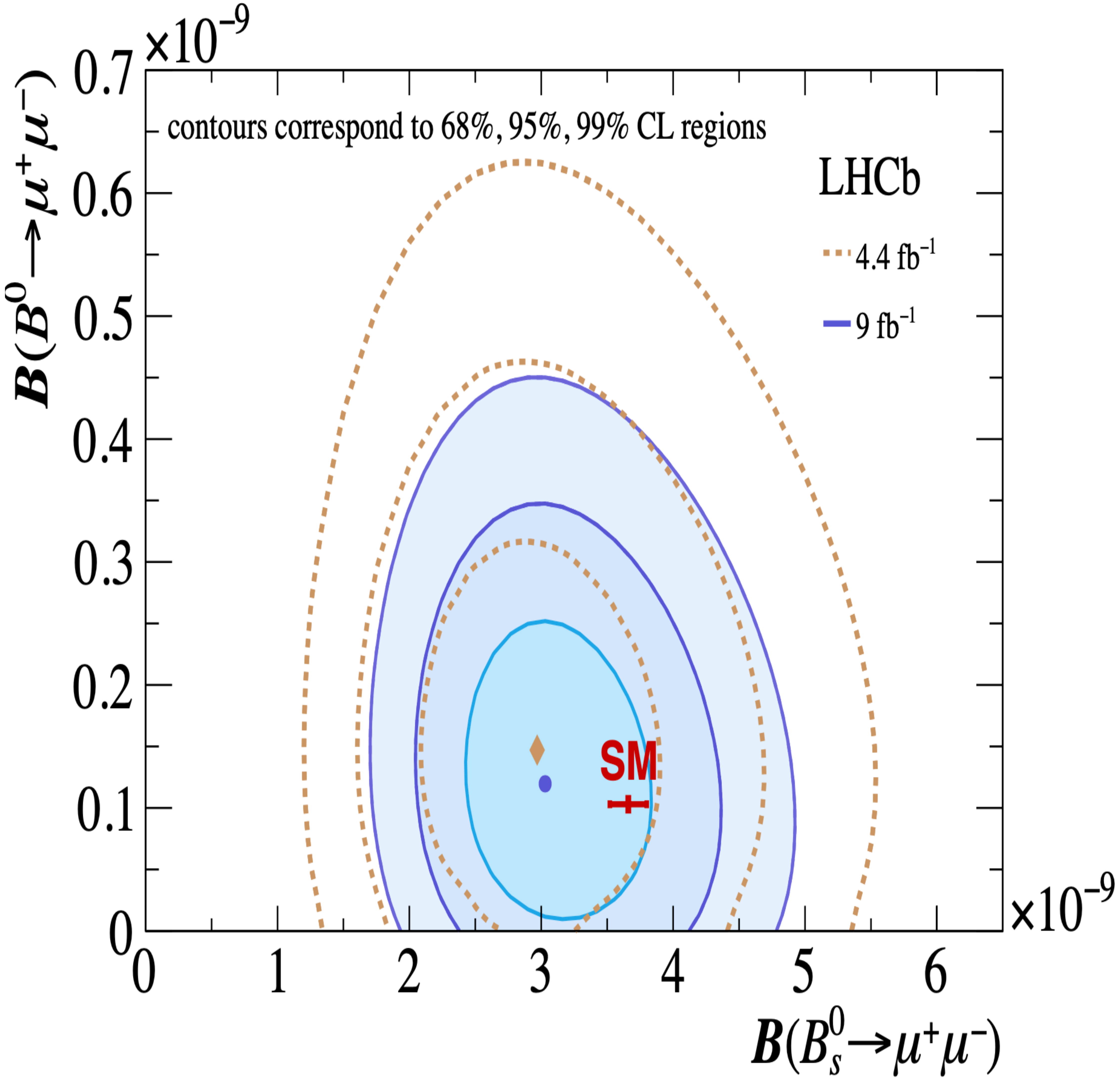




- The main challenge with $B0 \to \mu\mu$ is the combinatorial background
- It will require more data and analysis improvements to reach discovery level

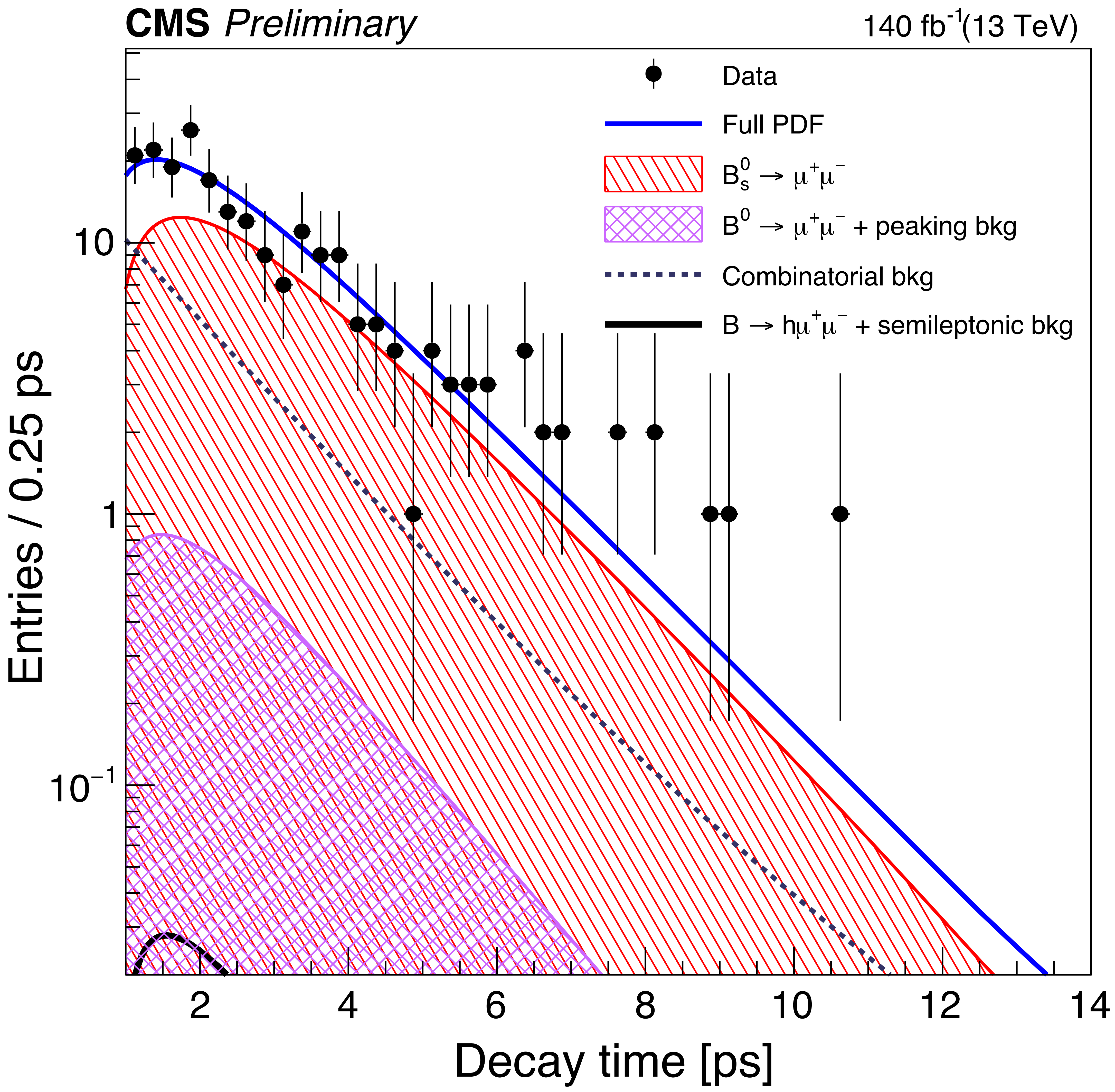
2D Contour Plots





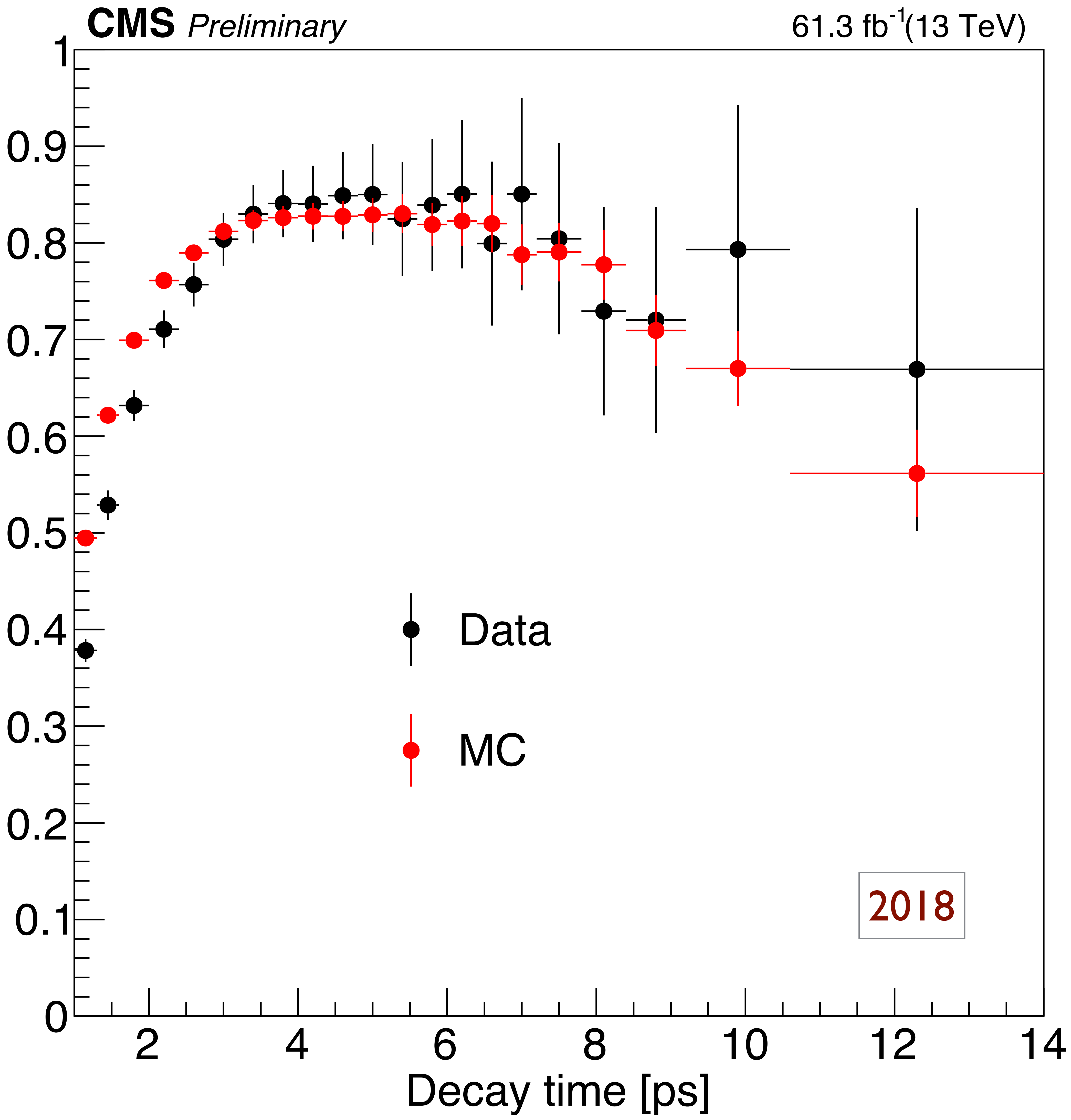
Lifetime Measurement

Effective Lifetime Measurement



- In the absence of CP violation only the heavy Bs state decays into dimuon
  - Different composition of states may be allowed by New Physics.
- Plot shows decay time projection
- Efficiency correction
  - Decay time efficiency derived from MC
  - Corrected by $B^+ \to J/\psi K^+$ data to mitigate the bias from tight MVA_B requirement.
    - The residual bias and the difference between $B_s \to \mu\mu$ and $B^+ \to J/\psi K^+$ are considered as a systematic uncertainty.

Lifetime Systematics

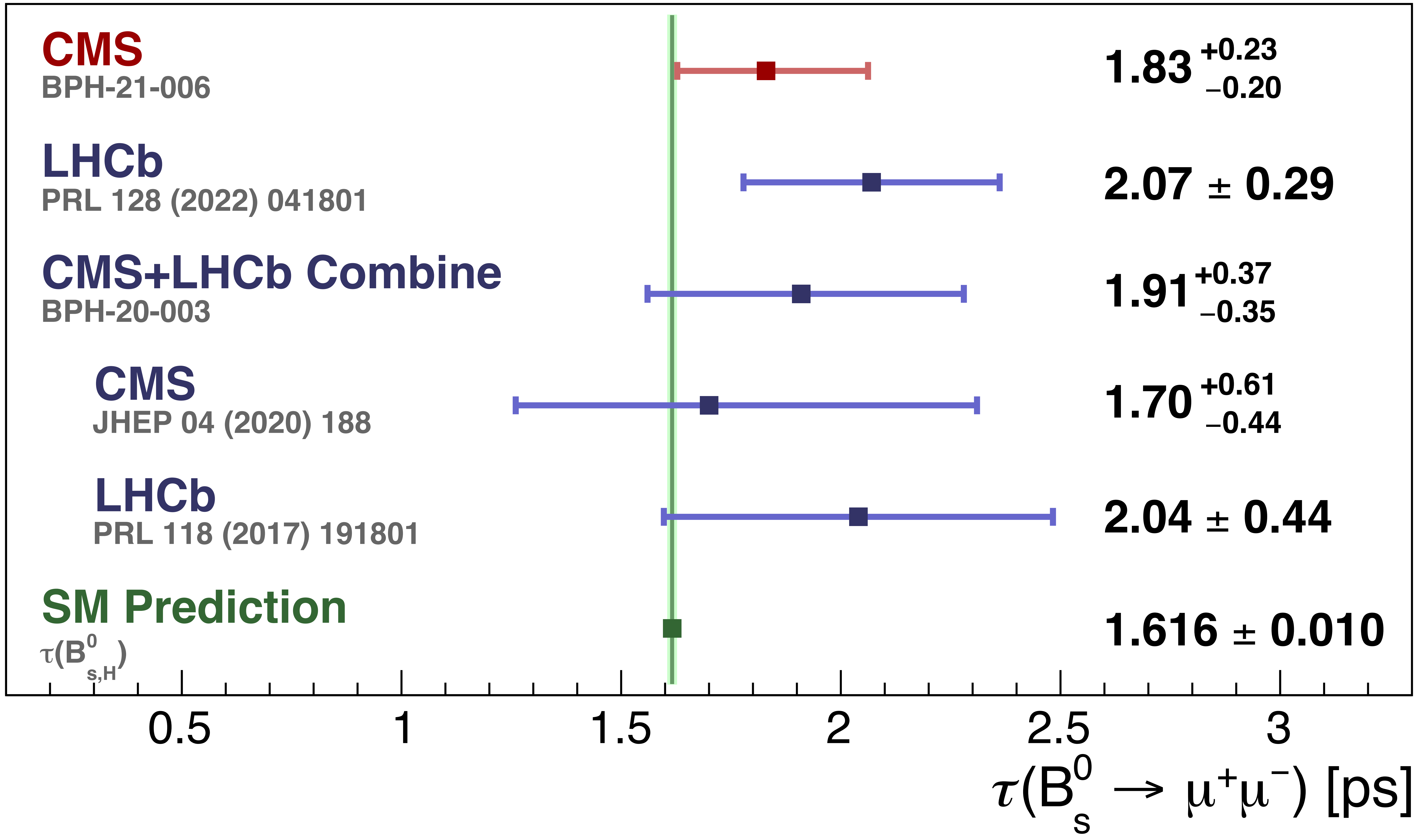


Effect	2016a	2016b	2017	2018
Efficiency modeling	0.01 ps			
Gen lifetime bias	0.01 ps			
Decay time mis-modeling	0.10 ps	0.06 ps	0.02 ps	0.02 ps
Lifetime bias	0.04 ps	0.04 ps	0.05 ps	0.04 ps
Total	**0.11 ps**	**0.07 ps**	**0.05 ps**	**0.04 ps**

- Dominant systematics comes from a strong correlation between MVA_B and decay time, which are hard to model well
  - Corrections can be derived in data and uncertainty is mostly limited by the size of the control sample

Effective Lifetime

$$\tau = 1.83\,^{+0.23}_{-0.20}\,(\text{stat})\,^{+0.04}_{-0.04}\,(\text{syst})\text{ ps}$$



Summary

- Studies of rare $B_{(s)} \to \mu^+\mu^-$ decays provides a unique tool to explore and understand rare B decay anomalies
  - Theoretically clean
  - Sensitive to the same processes
- CMS finalized analysis of 140 fb^{-1} data collected during LHC Run-2
  - All results are consistent with SM predictions
- Relative uncertainty on $BF(B_s \to \mu^+\mu^-)$ has been reduced to 11%
  - The best single measurement to date
- Statistical uncertainties dominate in all measurements
  - Good perspectives for further improvements with Run-3 data
- A conference note and other materials are available at
  - https://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/BPH-21-006/index.html