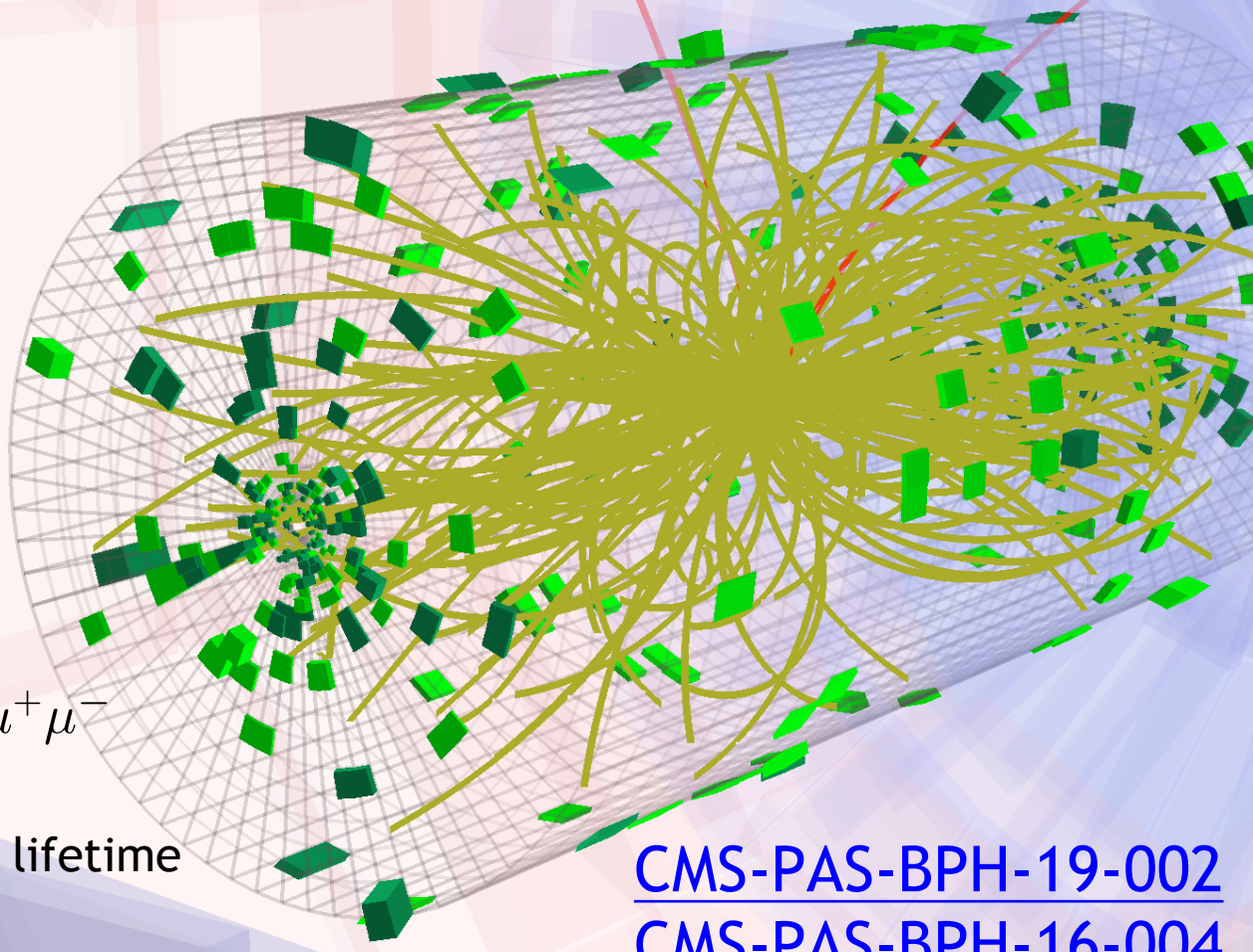
Recent results on $B \to \mu^+\mu^-$ decays and first observation of $\Lambda_b \to J/\psi\phi\Lambda$ with the CMS experiment

Urs Langenegger
(Paul Scherrer Institute)

on behalf of the CMS collaboration

- Introduction
  CMS detector
- $\Lambda_b^0 \to J/\psi\,\phi\Lambda$
  first observation
- $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$
  branching fractions
  $B_s^0 \to \mu^+\mu^-$ effective lifetime

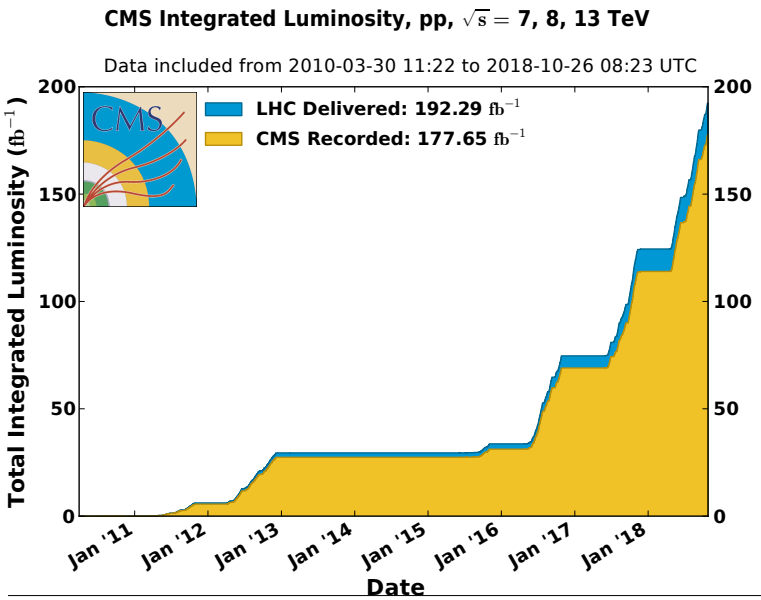


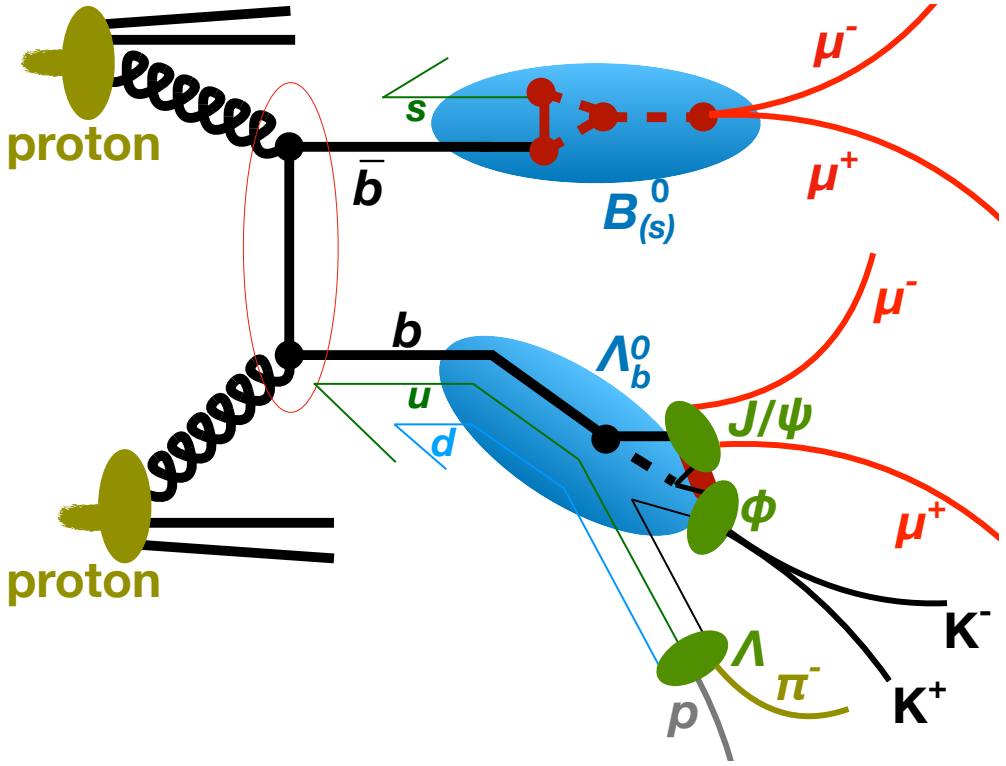
CMS-PAS-BPH-19-002
CMS-PAS-BPH-16-004

LHC seminar, CERN, 2019/09/17

B physics in CMS

- B physics activities in CMS cover wide area of topics
  - production of b hadrons
    - polarization
    - correlations
  - decays of b hadrons
    - lifetime
    - branching fractions
    - 'new physics'
      - heavy intermediate states
      - exotic states





- Datasets covered today
  - 2018 for $\Lambda_b^0 \to J/\psi\,\phi\Lambda$
  - 2011/2012/2016 for $B \to \mu^+\mu^-$
  - $\approx 60\,\text{fb}^{-1}$ in both cases

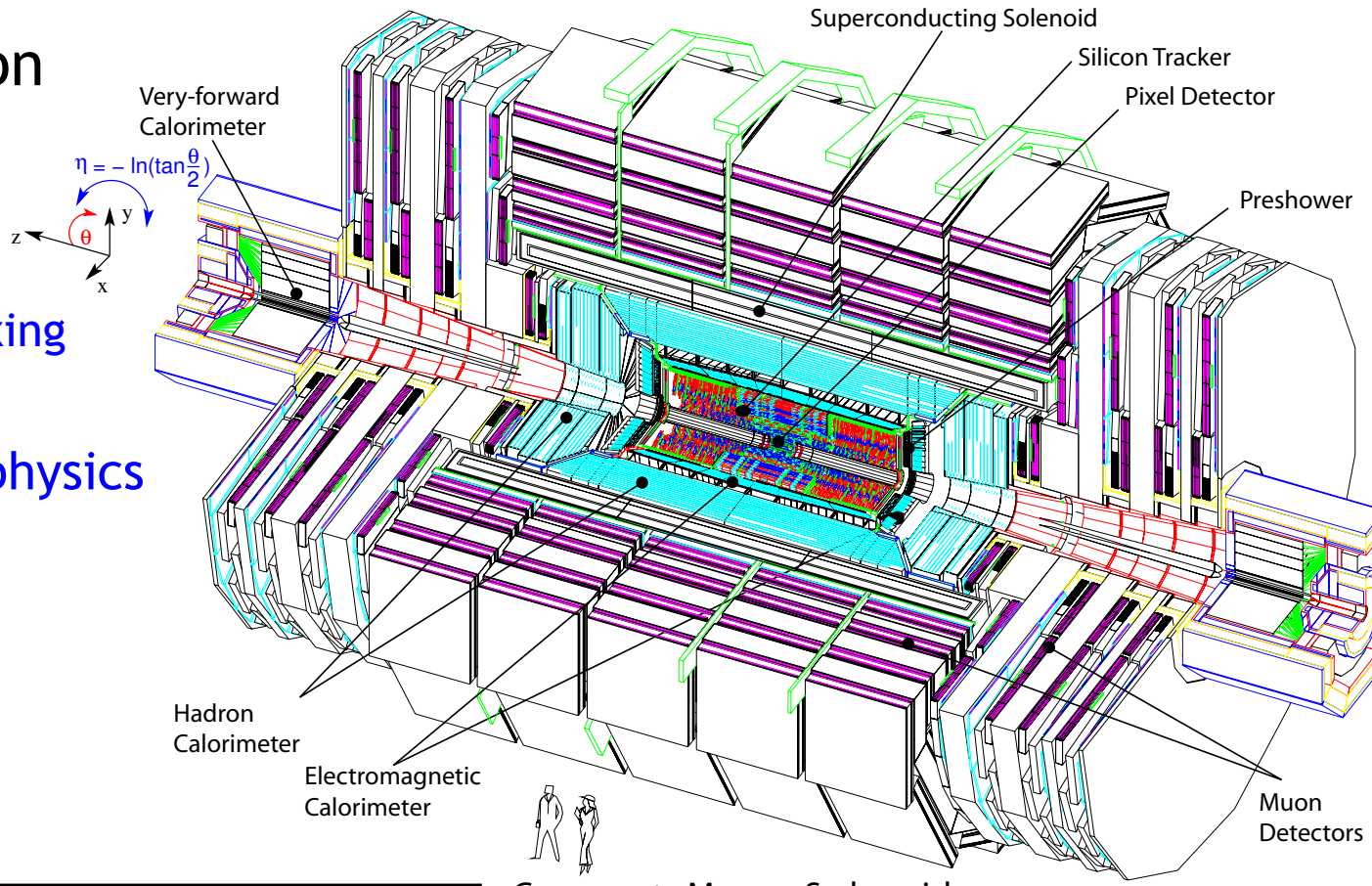
The CMS detector

JINST 3, S08004

The CMS detector

- Design prioritization
  - ▷ lepton ID
    - → muons
  - ▷ b/τ tagging
    - → tracking/vertexing
  - ▷ jets and $\not{E}_T$
  - → well suited for B physics

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



Component	Characteristics	Resolutions
Pixel	3/2 (4/3) Si layers	$\delta_z \approx 20\,\mu$m, $\delta_\phi \approx 10\,\mu$m
Tracker	10/12 Si strips	$\delta(p_\perp)/p_\perp \approx 1\%$
ECAL	$PbWO_4$	$\delta E/E \approx 3\%/\sqrt{E} \oplus 0.5\%$
HCAL (B)	Brass/Sc, $> 7.2\lambda$	$\delta E/E \approx 100\sqrt{E}\%$
HCAL (F)	Fe/Quartz	$\delta(\not{E}_T) \approx 0.98\sqrt{\sum E_T}$
Magnet	3.8 T solenoid	
Muons	DT/CSC + RPC	$\delta(p_\perp)/p_\perp \approx 10\%$ (STA)

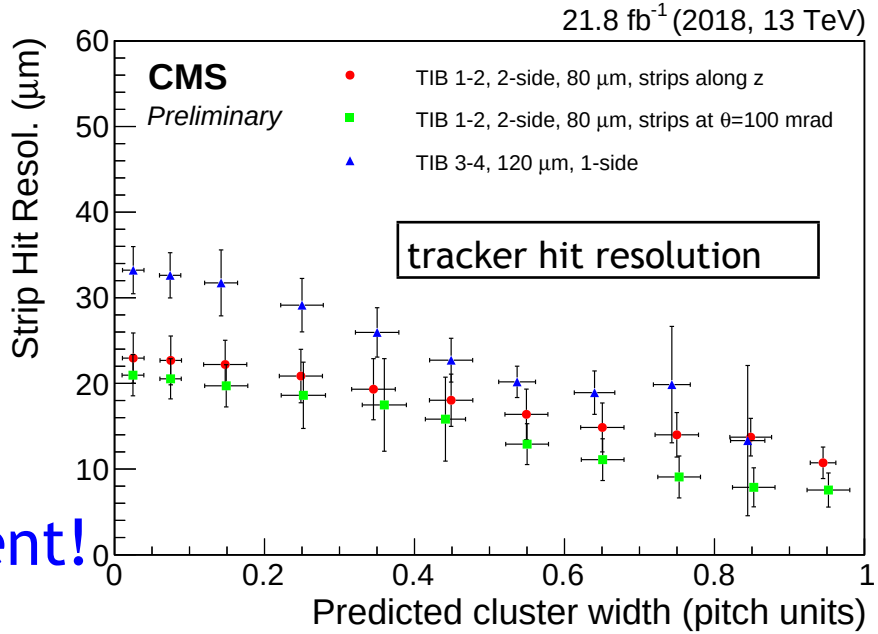
Tracking resolution:
impact parameter $\approx 15\,\mu$m

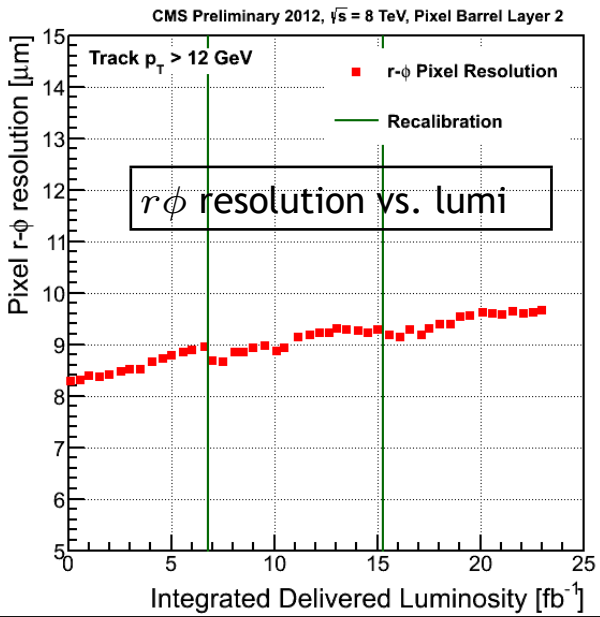
Primary vertex resolution:
$\Delta z \approx 20 - 80\,\mu$m

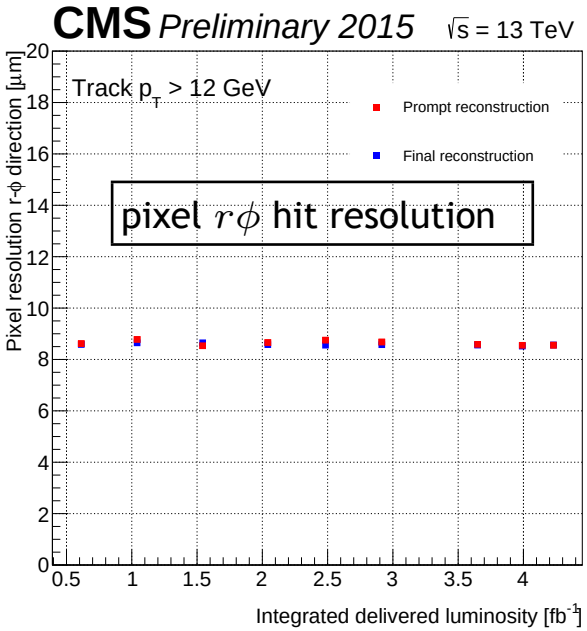
3D tracking and vertexing

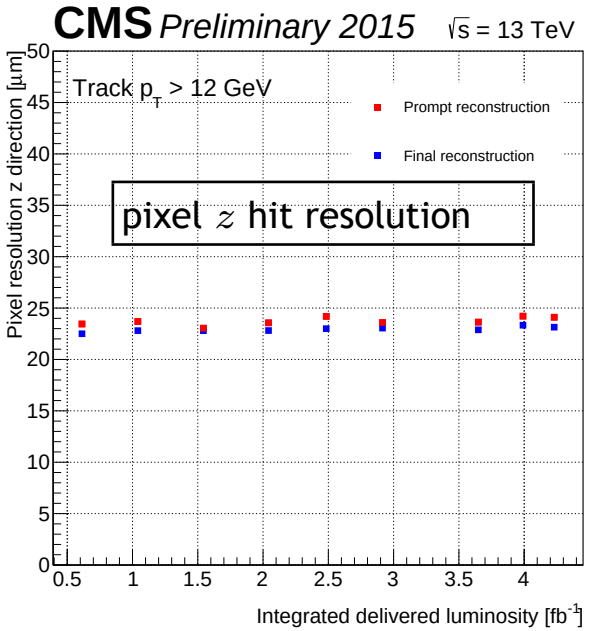
- All silicon tracker
  - ▷ high granularity, low occupancy
- Pixel detector
  - ▷ $100 \times 150\,\mu\text{m}^2$ pixel size
  - ▷ substantial charge sharing (low V_{bias})
  - $\rightarrow$ excellent resolution in $r\phi$ and z

$\Rightarrow$ Essential in high-pileup environment!



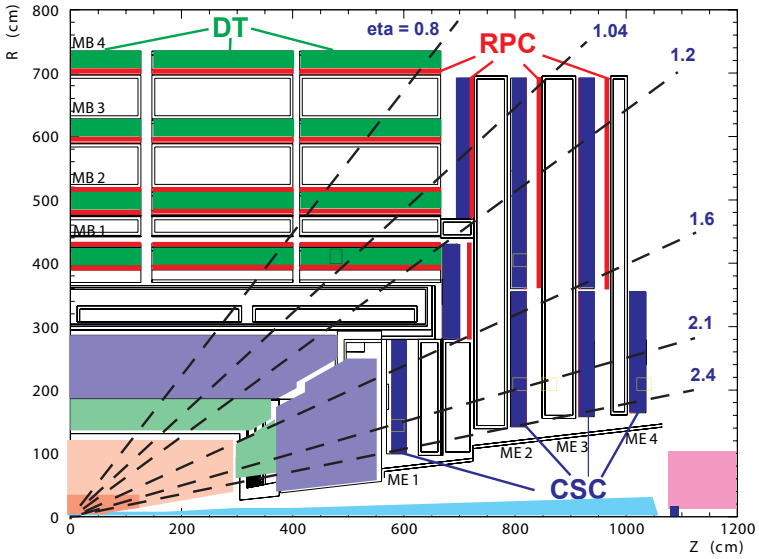






Muon reconstruction

JINST,13,P06015
JINST,7,P10002

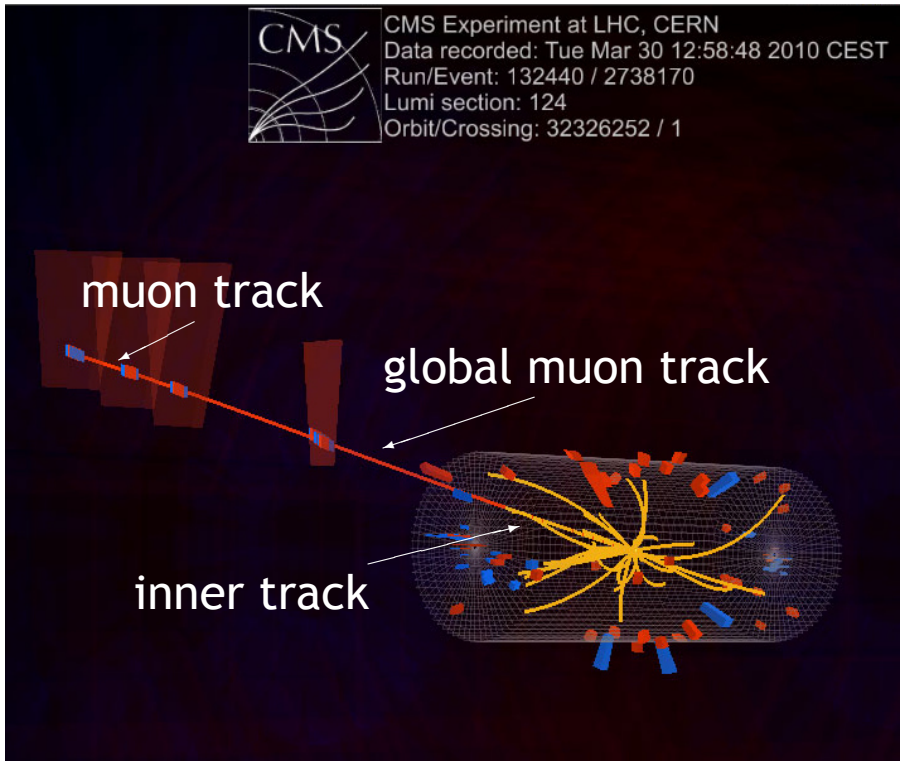


- Large muon acceptance $|\eta| < 2.4$
  - drift tubes
  - cathode strip chambers
  - resistive plate chambers
- Muon reconstruction
  - standalone muon:
    in muon system (trigger ingredient)
  - 'soft': high efficiency
    for J/ψ analyses
  - 'BDT': low misidentification
    for $B \to \mu^+\mu^-$ analyses

Muon misidentification for BDT muons

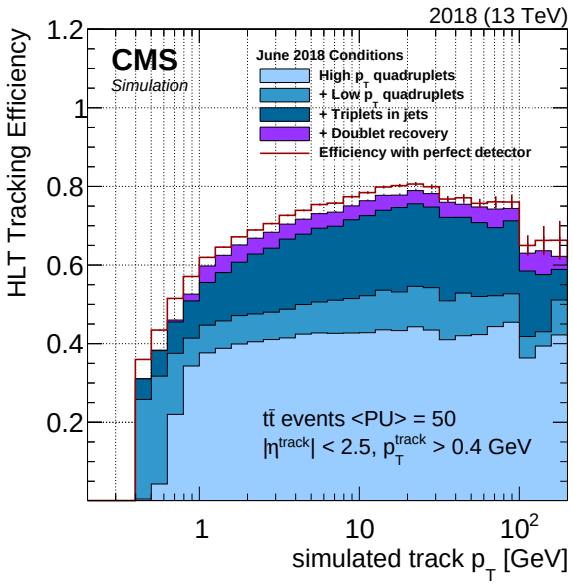
$$\begin{aligned} \varepsilon(\mu|\pi) &\approx 0.06\% \\ \varepsilon(\mu|K) &\approx 0.10\% \\ \varepsilon(\mu|p) &\leq 0.01\% \end{aligned}$$

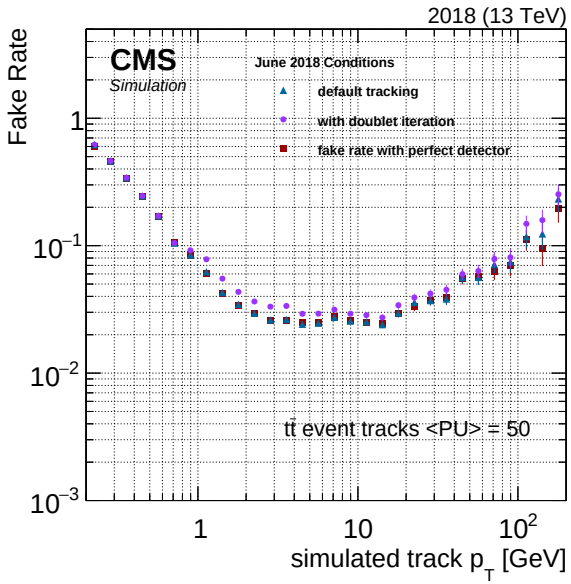
measured/validated in data:
$K^0_S \to \pi^+\pi^-$, $\phi \to K^-K^+$, $\Lambda \to p\pi^-$
$D^{*+} \to D^0\pi^+_s \to K^-\pi^+\pi^+_s$

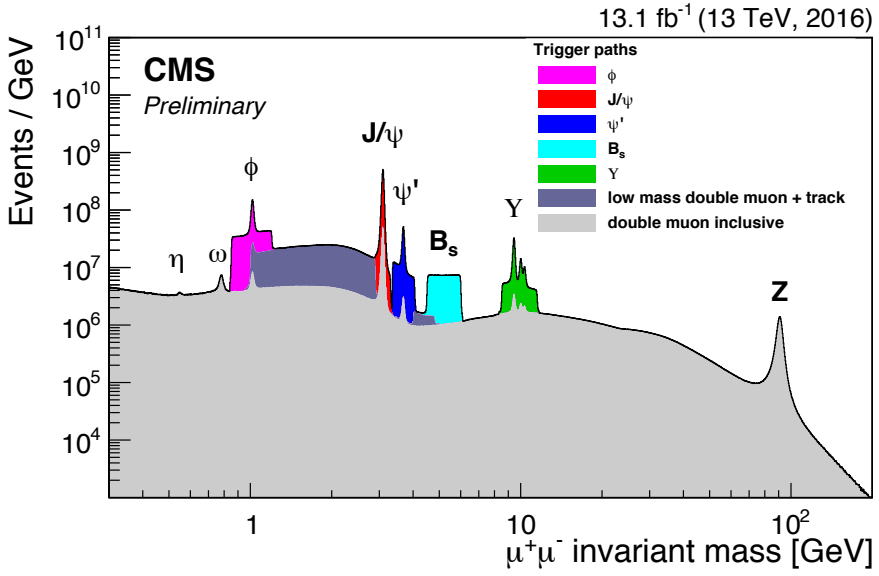


B physics trigger

- B physics mostly triggered with (displaced) dimuon $(+X)$ triggers
  (displaced from pp collisions; B hadrons have a lifetime of about 1.5 ps)
  - ▷ other setups are in progress/under analysis
- L1: hardware trigger based on muons
  - ▷ $< 4\,\mu$s latency
  - ▷ no explicit muon $p_\perp$ threshold (strong B field implies $p_\perp > 3$ GeV in barrel)
- HLT: high-level trigger
  - ▷ full tracking and vertexing
  - ▷ specific trigger paths

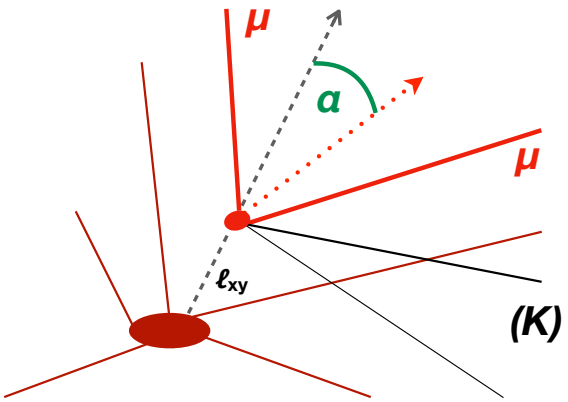


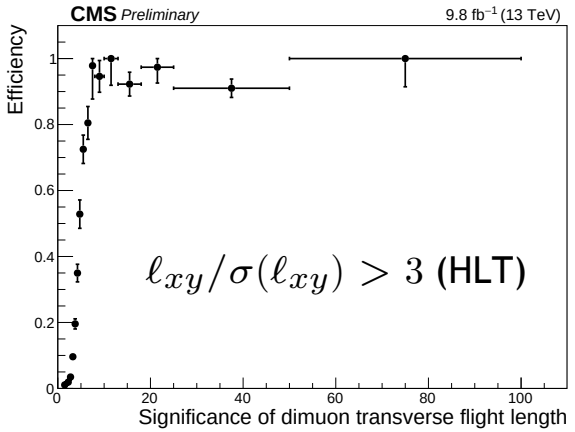




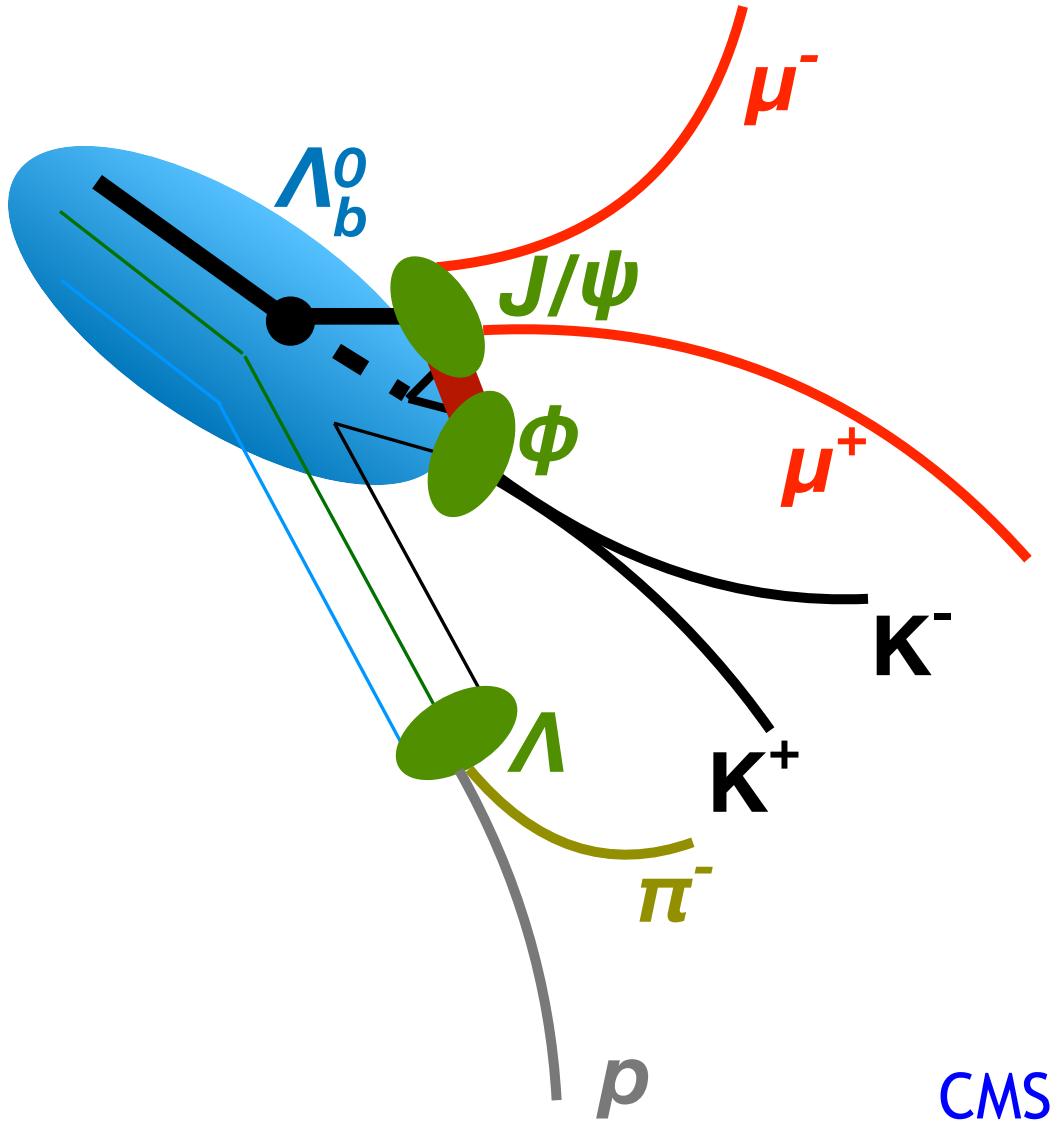
Displaced J/ψ and $B_s^0 \to \mu^+\mu^-$ triggers

- HLT 'displaced' J/ψ
  - two muons with opposite charge
  - $2.9 < m_{\mu\mu} < 3.3$ GeV
  - $\ell_{xy}/\sigma(\ell_{xy}) > 3$
  - $\cos\alpha > 0.9$, $\mathcal{P}(\chi^2/dof) > 10\%$
- HLT 'displaced' J/ψ + track(s)
  - two muons with opposite charge
    $2.9 < m_{\mu\mu} < 3.3$ GeV, $\ell_{xy}/\sigma(\ell_{xy}) > 3$
  - $\cos\alpha > 0.9$, $\mathcal{P}(\chi^2/dof) > 10\%$
  - invariant mass requirements on tracks
    (targeted towards $\phi \to K^+K^-$)
- HLT $B_s^0 \to \mu^+\mu^-$
  - two muons with opposite charge
  - inv. mass $4.8 < m_{\mu\mu} < 6.0$ GeV
  - $p_\perp > 4.0(3.5)$ GeV, $\mathcal{P}(\chi^2/dof) > 0.5\%$
  - no displacement requirement!





First observation of $\Lambda_b^0 \to J/\psi\,\phi\Lambda$

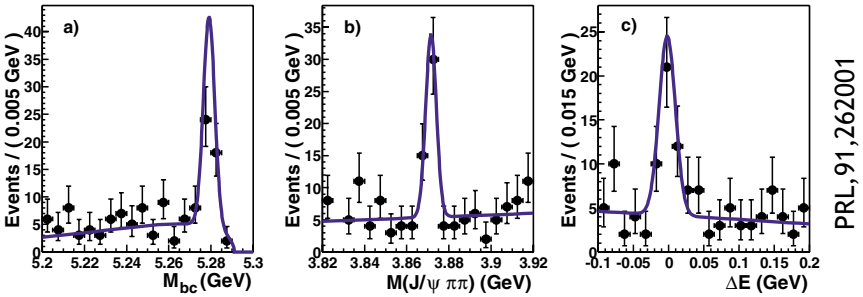


CMS-PAS-BPH-19-002

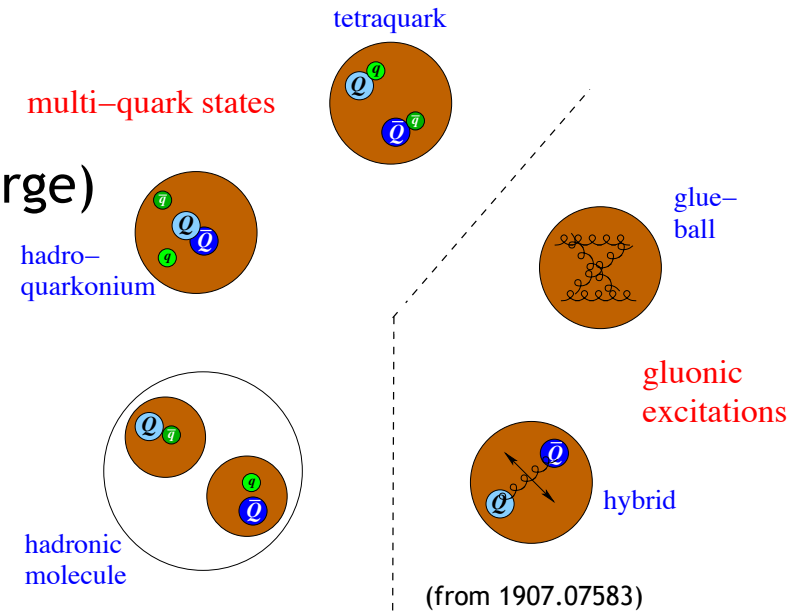
‘New physics’

- Known (confirmed) hadrons until 2003
  - ▷ mesons$(q\,\overline{q})$ and baryons$(q\,q\,q)$
  - ▷ other possibilities not excluded by ‘quark model’: e.g. $\overline{q}\,\overline{q}\,q\,q$, $\overline{qqq}qqq$, . . .
- 2003: Belle discovery
  - ▷ $X(3872)$, now known as $\chi_{c1}(3872)$ among decay products of

    $B^{\pm} \to J/\psi\,\pi^+\pi^-K^{\pm}$ $(J/\psi \to \mu^+\mu^-)$

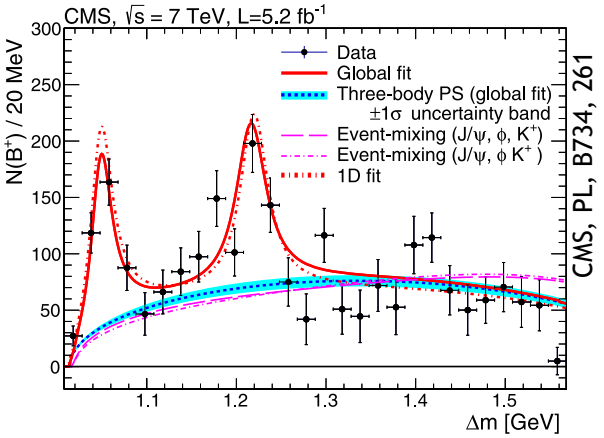


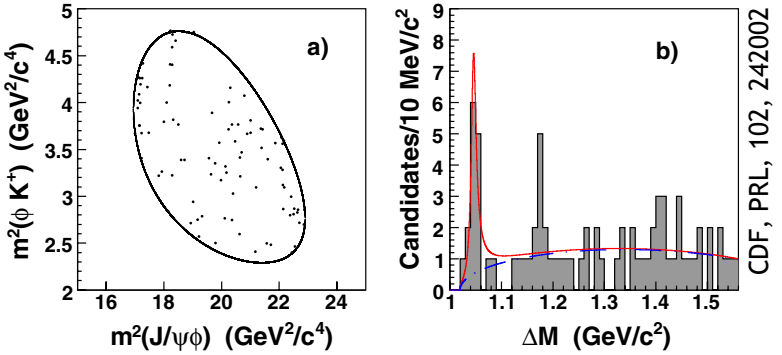
  - → many more neutral and charged(!) states discovered in mean time inconsistent with unassigned onia states (mass, decay characteristics, electric charge)
- Interpretations of new states
  - ▷ tetraquarks and hadroquarkonium
  - ▷ hadronic molecules
  - ▷ hybrids and glueballs
  - ▷ pentaquarks



Exotic states in $J/\psi\,\phi$ mass spectrum

- Rich resonant substructure observed in $B^+ \to J/\psi\,\phi K^+$
  - 2008: CDF with evidence for $X(4140)$, a.k.a. $\chi_{c1}(4140)$
    contributions by Belle, LHCb, D0 . . .
  - 2014: CMS with observation of $X(4140)$
    and evidence for $X(4274)$

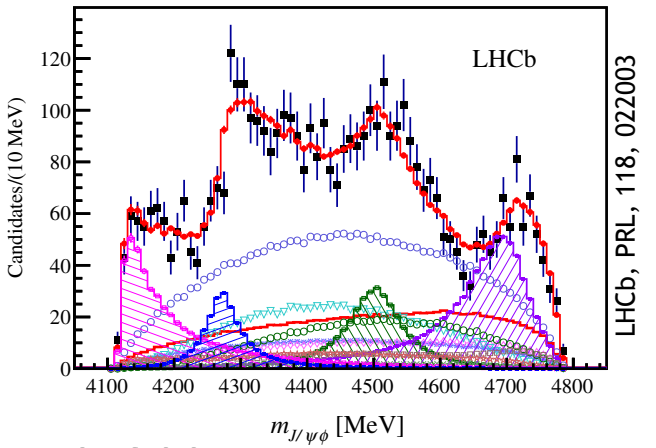




$$\Delta M \equiv m(\mu^+\mu^- K^+ K^-) - m(\mu^+\mu^-)$$

  - 2016: LHCb full amplitude analysis

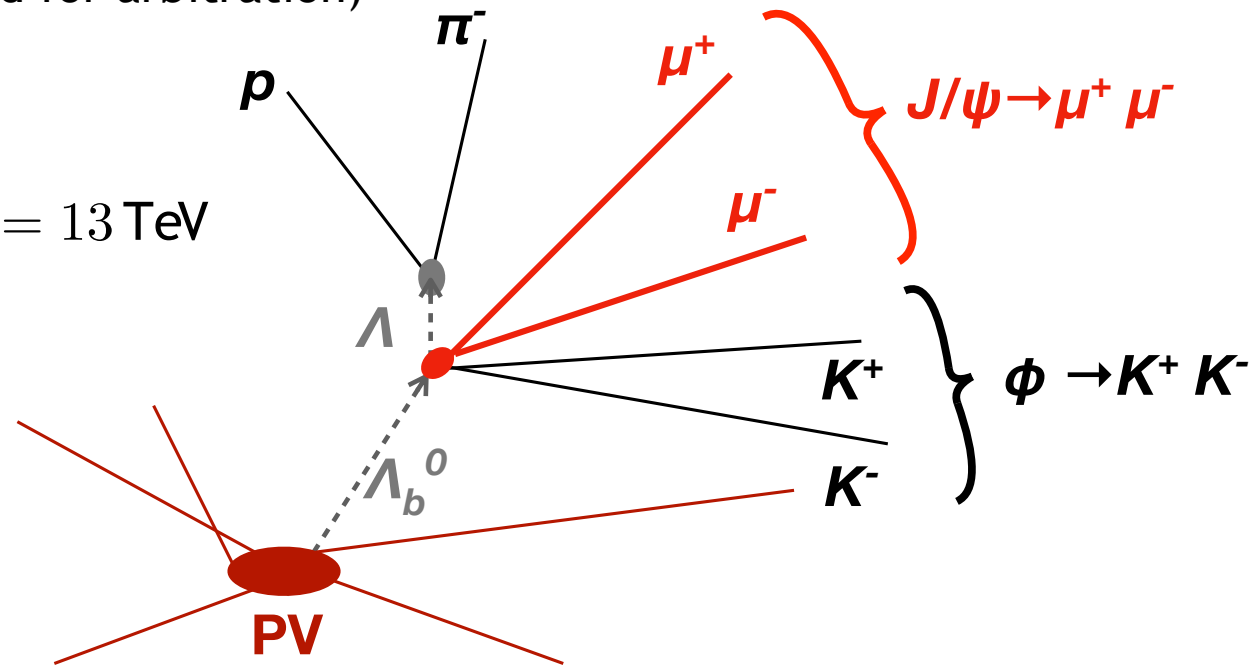
$\to$ $X(4140)$, $X(4274)$, $X(4500)$, and $X(4700)$



- What about baryons?
  - LHCb observed pentaquark candidates (PRL, 115, 072001)
  - $\Lambda_b^0 \to J/\psi\,\phi\Lambda$ is baryonic equivalent $B^+ \to J/\psi\,\phi K^+$

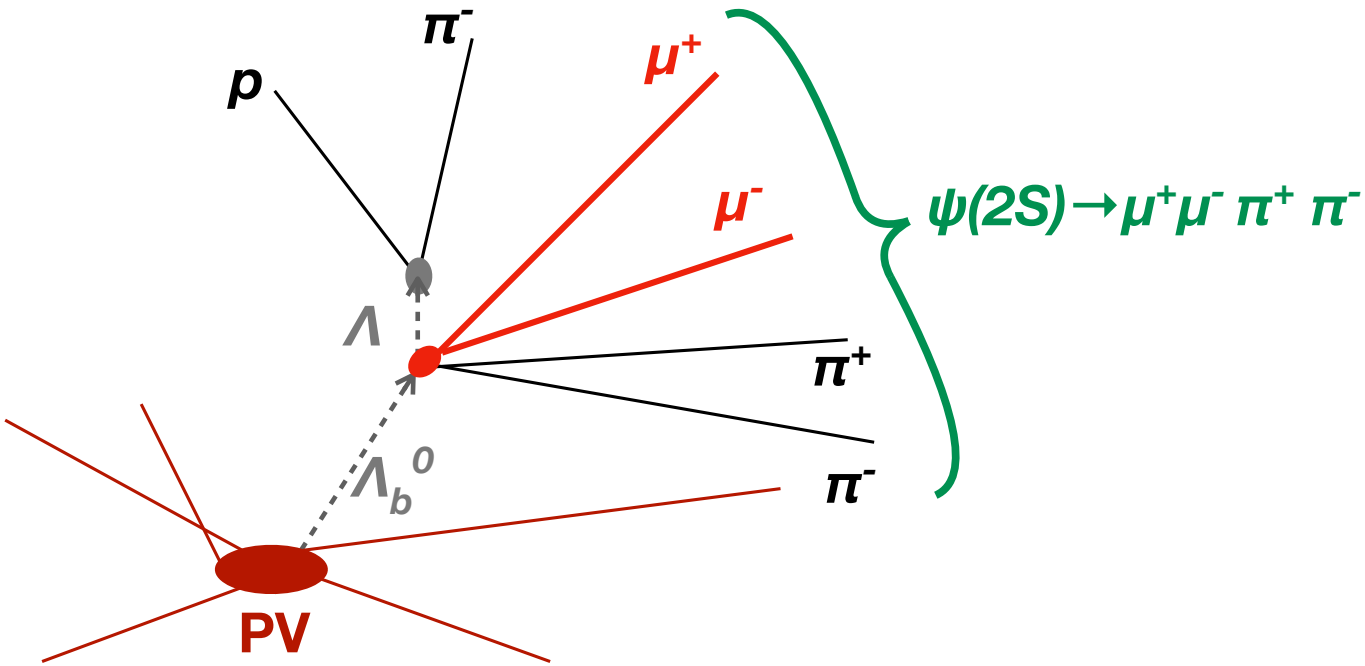
Selection of $\Lambda_b^0 \to J/\psi\,\phi\Lambda$

- Signal $\Lambda_b^0 \to J/\psi\,\phi\Lambda$
  - trigger: displaced $J/\psi\,KK$ with $0.95 < m(KK) < 1.3\,\text{GeV}$
  - J/ψ: good vertex fit and $p_\perp > 6.9\,\text{GeV}$
  - Λ: good vertex fit, $p_\perp > 1\,\text{GeV}$
  - tracks: $p_\perp > 0.8\,\text{GeV}$
  - Λ_b candidate
    displaced vertex
    $\cos(\alpha) > 0.99$ (used for arbitration)
    $p_\perp > 10\,\text{GeV}$
- Dataset
  - pp collisions at $\sqrt{s} = 13\,\text{TeV}$
  - $60\,\text{fb}^{-1}$ from 2018



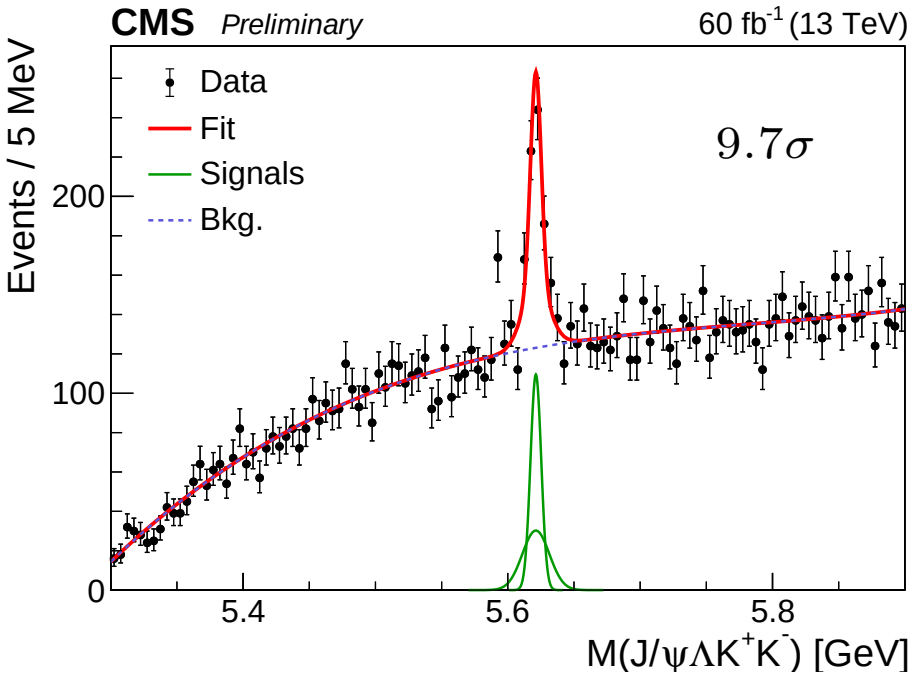
Normalization: $\Lambda_b \to \Lambda\psi(2S), \psi(2S) \to J/\psi\,\pi^+\pi^-$

- For branching fraction use a normalization sample
  - identical final state topology $\to$ reduction of systematic errors
  - trigger and selection: same as for signal
  - $|m(\mu^+\mu^-\pi^+\pi^-) - m(\psi(2S)^{\text{PDG}}| < 15\,\text{MeV}$
    (approximately three effective widths of reconstructed $\psi(2S)$ candidates)
- Validation of $\psi(2S)$ selection
  - no reflections/other background in $m(J/\psi\,\pi^+\pi^-)$



Yield extraction

- Unbinned extended maximum likelihood fit to $m(J/\psi\,\Lambda K^+K^-)$
  - not yet the final signal yield: still includes non-resonant K^+K^-
- Signal: double Gaussian with common mean
  - mean and total normalization floated
  - widths and fraction fixed to MC expectation
- Background: Bernstein polynomial $\mathcal{P}^{(3)}$ of 3rd degree



$$N_{J/\psi\,\Lambda K^+K^-} = 380 \pm 32_{\text{stat}}$$

Systematics

- signal systematics (0.8%)
  triple Gaussian, sum of 2 Crystal Ball fct's
- background systematics (0.6%)
  $(x - x_0)^\alpha \times \mathcal{P}^{(1,2)}$ and $\mathcal{P}^{(2,4)}$
- mass resolution in data vs. MC (6.6%)
  fix fraction only

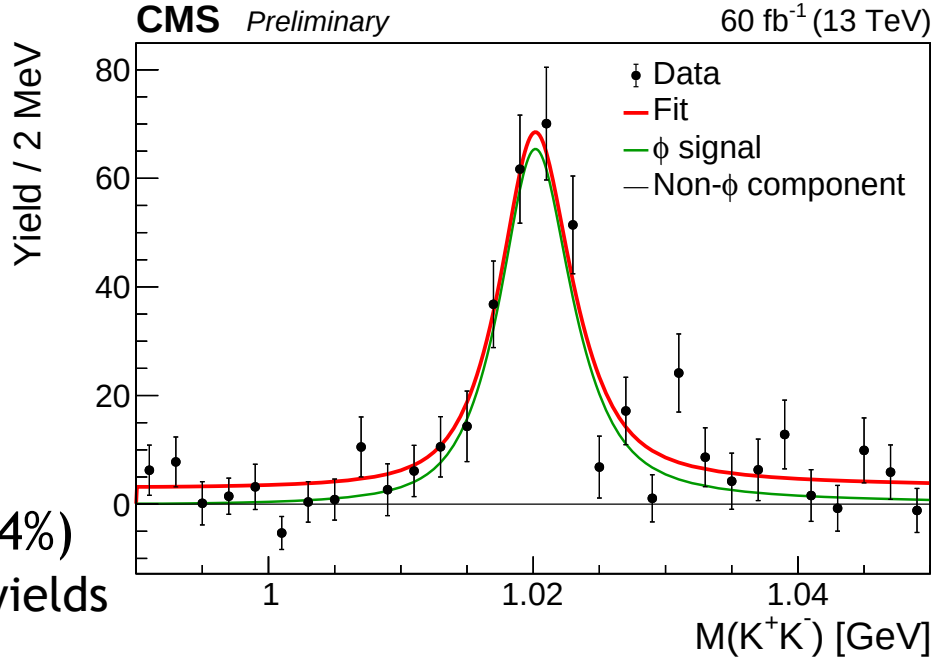
Signal yield extraction

- Determination of resonant $\phi \to K^+K^-$ component
  - $\Lambda_b^0 \to J/\psi\,\phi\Lambda$ background subtraction with $sPlot$
  - unbinned extended maximum likelihood fit to $m(K^+K^-)$ distribution
- Signal: double Gaussian convolved with rel. Breit-Wigner function
  - natural width fixed to PDG
  - detector resolution fixed to MC expectation
- Background: Bernstein polynomial $\mathcal{P}^{(1)}$

$$N_{J/\psi\,\Lambda\phi} = 286 \pm 29_{\text{stat}}$$

Systematics

- signal systematics (0.5%)
  double Crystal Ball $\otimes$ rel. BW fct's
- background systematics (0.8%)
  exponential function
- mass resolution in data vs. MC (1.4%)
  fixed width fractions and Gaussian yields



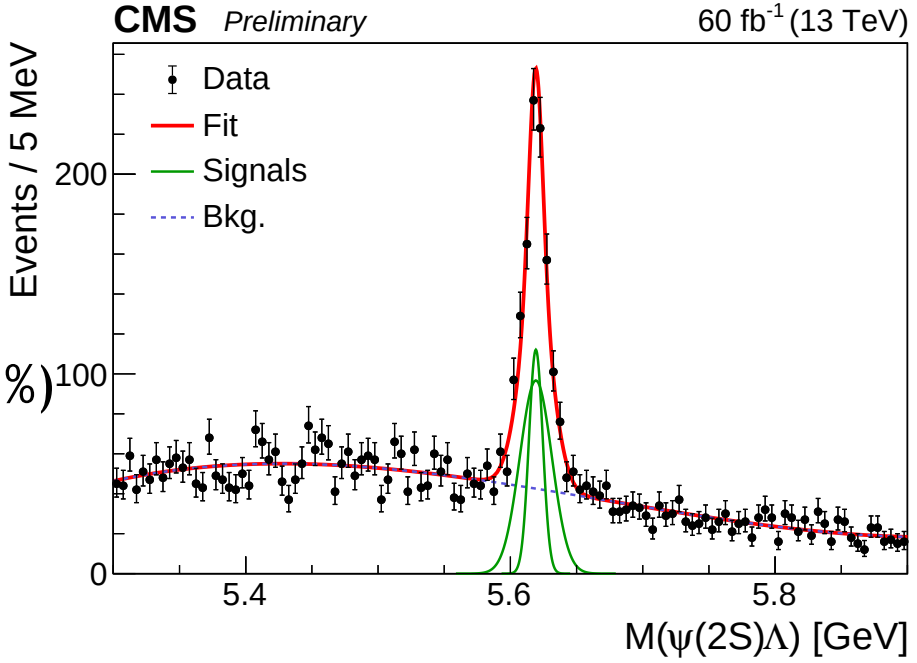
Normalization yield extraction

- Unbinned extended maximum likelihood fit to $m(\Lambda\psi(2S))$
- Signal: double Gaussian with common mean
  - mean and total normalization floated
  - widths and fraction fixed to MC expectation
- Background: Bernstein polynomial $\mathcal{P}^{(3)}$

$$N_{\psi(2S)\Lambda} = 884 \pm 37_{\text{stat}}$$

Systematics

- signal systematics (1.1%)
  sum of 2 Crystal Ball fct's
- background systematics (0.8%)
  $\mathcal{P}^{(2,4)}$
- mass resolution in data vs. MC (3.1%)
  fix fraction only



Results

- Branching fraction relative to normalization

$$\frac{\mathcal{B}(\Lambda_b^0 \to J/\psi\,\phi\Lambda)}{\mathcal{B}(\Lambda_b \to \psi(2S)\Lambda)} = \frac{N(\Lambda_b^0 \to J/\psi\,\phi\Lambda)}{N(\Lambda_b \to \psi(2S)\Lambda)} \frac{\varepsilon(\Lambda_b \to \psi(2S)\Lambda)}{\varepsilon(\Lambda_b^0 \to J/\psi\,\phi\Lambda)} \frac{\mathcal{B}(\psi(2S) \to J/\psi\,\pi^+\pi^-)}{\mathcal{B}(\phi \to K^+K^-)}$$

- Efficiency ratio determined from MC simulation

$$\frac{\varepsilon(\Lambda_b \to \psi(2S)\Lambda)}{\varepsilon(\Lambda_b^0 \to J/\psi\,\phi\Lambda)} = 0.363 \pm 0.011$$

systematics (2.9%)

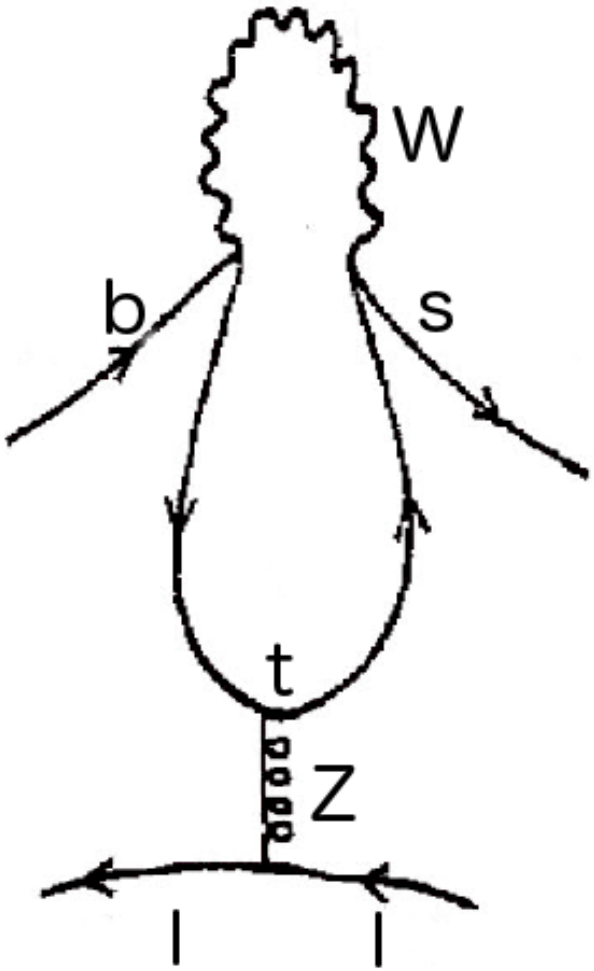
- ▷ dominated by MC sample size

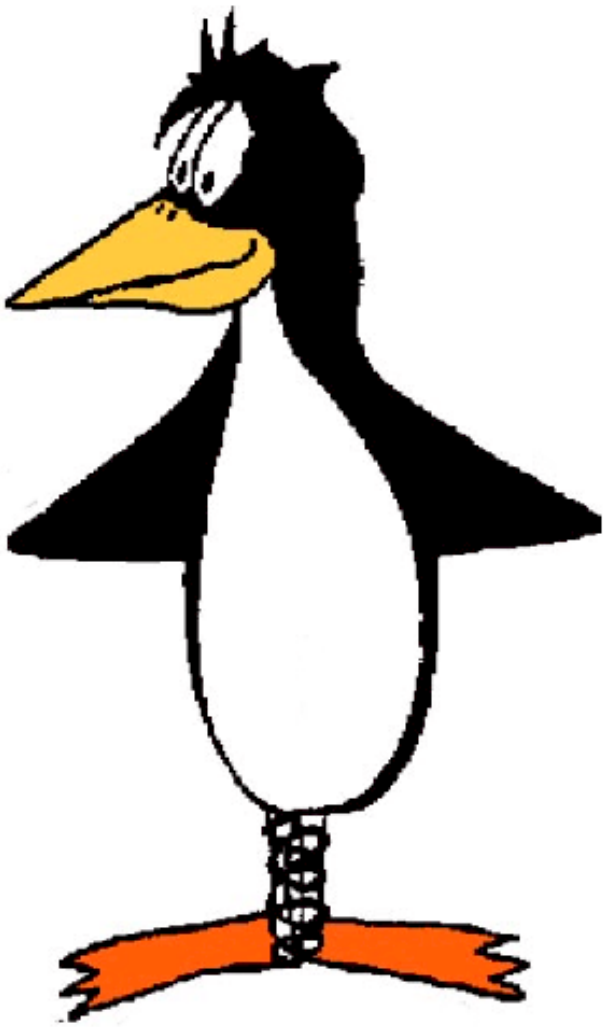
$\Rightarrow$ First observation of $\Lambda_b^0 \to J/\psi\,\phi\Lambda$ with (relative) branching fraction

$$\frac{\mathcal{B}(\Lambda_b^0 \to J/\psi\,\phi\Lambda)}{\mathcal{B}(\Lambda_b \to \psi(2S)\Lambda)} = (8.26 \pm 0.90_{\text{stat}} \pm 0.68_{\text{syst}} \pm 0.11_{\mathcal{B}})\%$$

$\Rightarrow$ Next: study $J/\psi\,\phi$ mass structure with more data from Run 2

$B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$



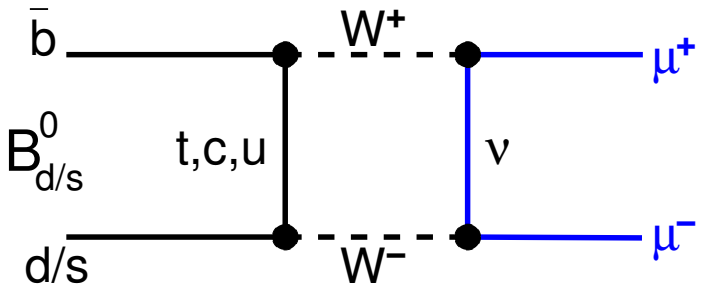


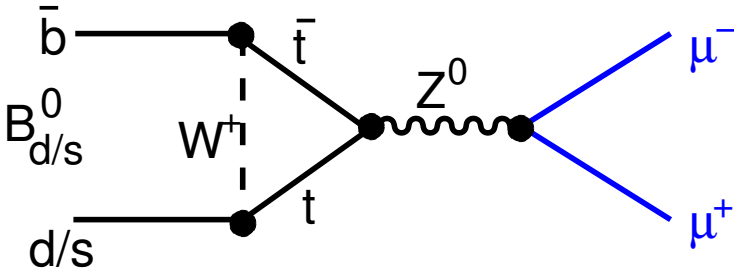
CMS-PAS-BPH-16-004

Leptonic B decays

PRL, 112,101801
PRL, 120,111801
1908.07011

- Leptonic B decays have only leptons in the final state
  - ▷ $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$
  - ▷ many other modes possible (or 'forbidden') as well
  - $\to$ no hadronic complication from final state

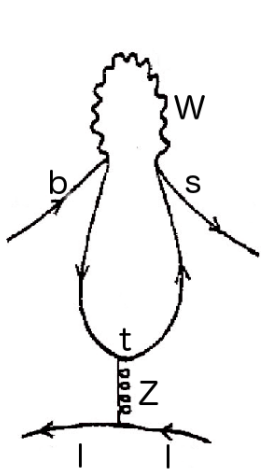


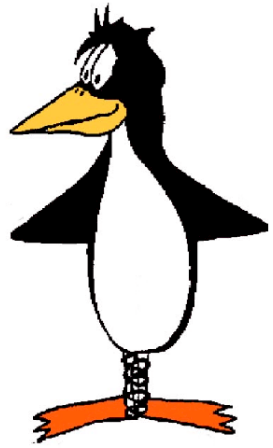


- They are strongly suppressed
  - ▷ SM branching fractions are small

    (ignoring tiny contributions from Higgs boson exchanges)

$$\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)_{\text{SM}} = (3.66 \pm 0.14) \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-)_{\text{SM}} = (1.03 \pm 0.05) \times 10^{-10}$$



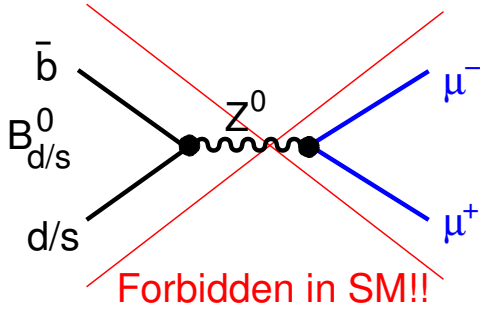


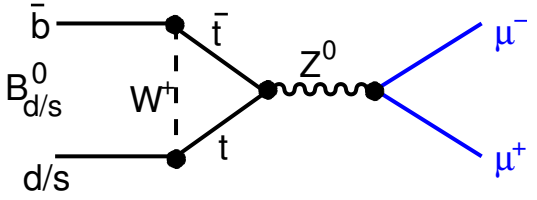
  - ▷ SM expectation: 4-5% theoretical uncertainty!
  - ▷ $\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$: decay time-integrated $\mathcal{B}$

$|B^0_{s\text{H,L}}\rangle = p|B^0_s\rangle \pm q|\bar{B}^0_s\rangle$
with different lifetimes

Why are they suppressed in the SM?

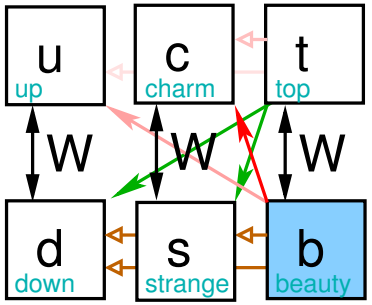
- Effective flavor-changing neutral currents
  - no flavor-changing neutral currents in SM
  - Penguin and box diagrams, but no tree-level process
  - CKM-suppression of $B^0 \to \mu^+\mu^-$ vs. $B_s^0 \to \mu^+\mu^-$: $|V_{td}|^2 < |V_{ts}|^2$
  - Cabibbo-Kobayashi-Maskawa matrix



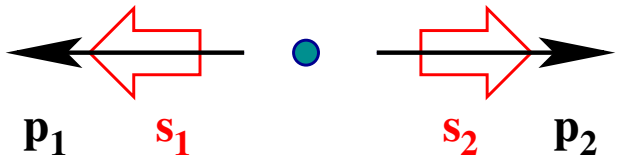


$$\begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{weak}} = \underbrace{\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}}_{V_{\text{CKM}}} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\text{mass}}$$

$$|V_{\text{CKM}}| = \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.224 & 0.974 & 0.042 \\ 0.009 & 0.041 & 0.999 \end{pmatrix}$$

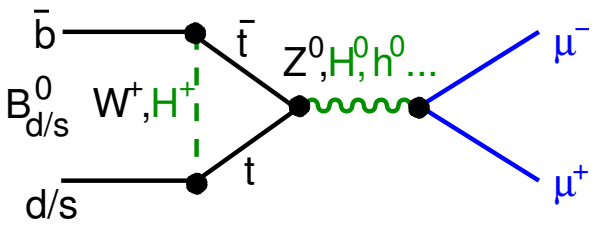


- Helicity suppressed ($V - A$ interaction in SM)
  - B mesons have no spin
  - μ have spin $1/2$
  - weak interaction is left handed

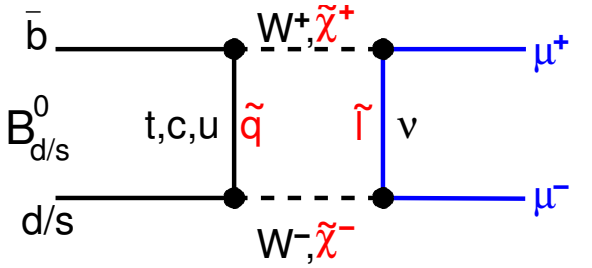


Search for 'BSM' physics

- Already small additional decay width contributions visible
  - because the SM decay width is so small



- Sensitivity to 'beyond-SM' physics
  - no helicity suppression
    - scalar couplings
    - pseudo-scalar couplings
  - 'extended' Higgs boson sectors
  - flavor 'violation': $B_s^0 \to \mu^+\mu^-$ vs $B^0 \to \mu^+\mu^-$



- Two approaches
  - model-independent: effective field theory
    - parametrize new physics with operators and (Wilson) coefficients
    - can include correlations to other processes $R_K^{(*)}, P_5', \ldots$
  - 'top-down': specific model
    - new particles extending the SM world
    - correlations between many processes precisely calculable
    - more specific than above, but very model dependent

$B_s^0 \to \mu^+\mu^-$ effective lifetime

PRL, 109,041801
PRL, 118,191801

- A second independent observable: $B_s^0 \to \mu^+\mu^-$ effective lifetime
  - measure B_s^0 lifetime with $B_s^0 \to \mu^+\mu^-$ decays

$$\tau_{\mu^+\mu^-} \equiv \frac{\int_0^\infty t\,\Gamma(B_s(t) \to \mu^+\mu^-)\,dt}{\int_0^\infty \Gamma(B_s(t) \to \mu^+\mu^-)\,dt} = \frac{\tau_{B_s^0}}{1-y_s^2}\left(\frac{1+2\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} y_s + y_s^2}{1+\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} y_s}\right)$$

  - allows determination of $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$

    B_s^0 mean lifetime $\tau_{B_s^0} = 1.510 \pm 0.005$ ps

    B_s^0 decay width difference $\Delta\Gamma_s$,

    $$\Delta\Gamma_s \equiv \Gamma_{sL} - \Gamma_{sH} = 0.088 \pm 0.006\,\text{ps}^{-1}$$

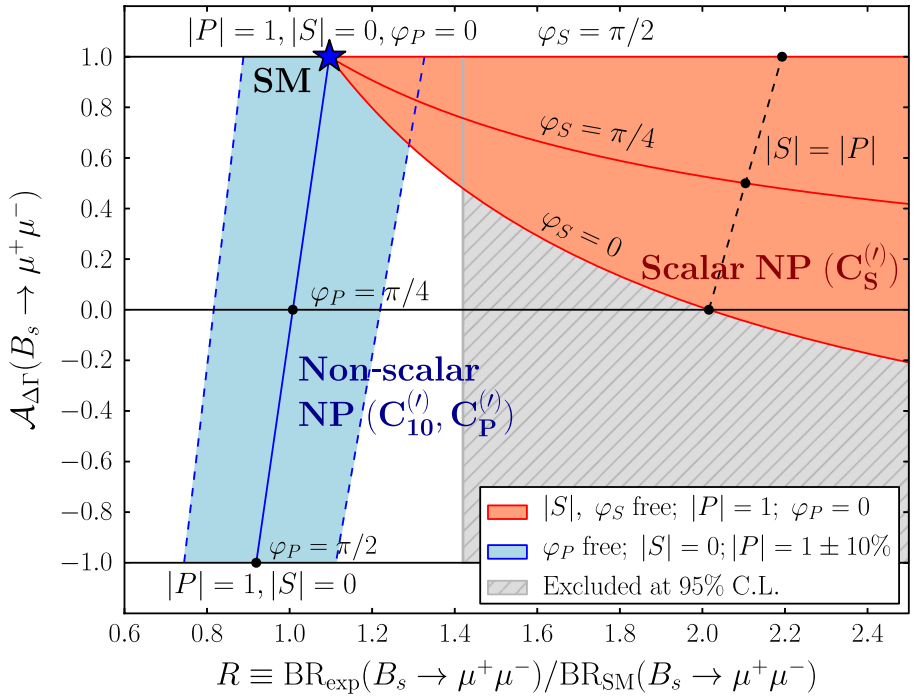
    $$y_s \equiv \tau_{B_s^0}\Delta\Gamma_s/2 = 0.062 \pm 0.006$$

  → scalar vs. non-scalar 'new physics'

  - SM prediction:

    only the $B_{s\,\mathrm{H}}^0$ decays to dimuons:

    $$\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-} = +1 \to \tau_{\mu^+\mu^-}^{\text{SM}} = 1.615\,\text{ps}$$

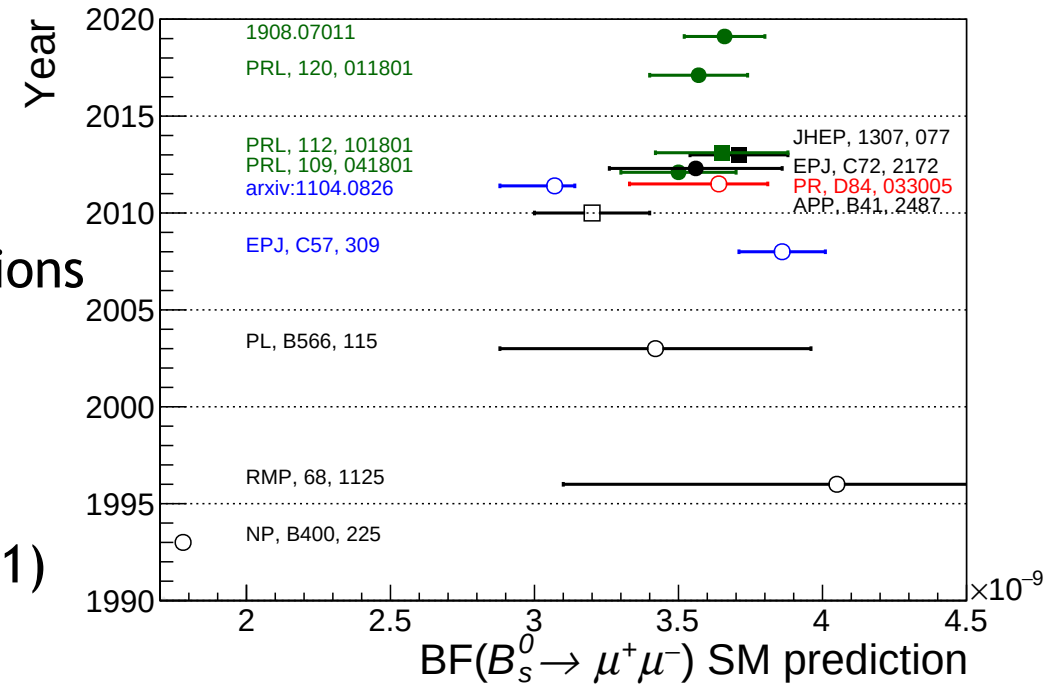


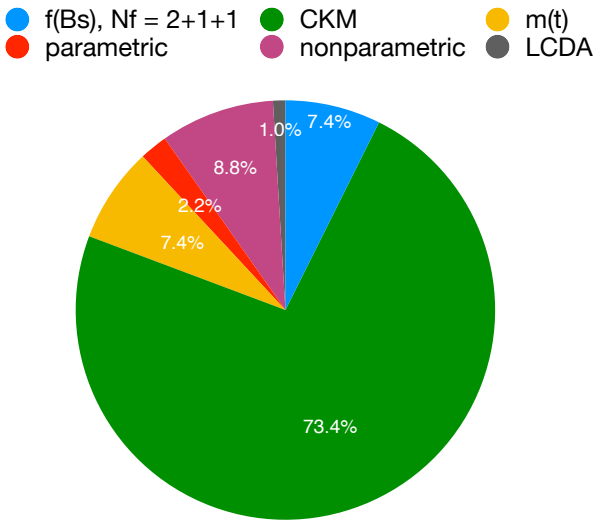
- One measurement to date:
  - LHCb 2017: $\tau_{\mu^+\mu^-} = 2.04 \pm 0.44 \pm 0.05$ ps

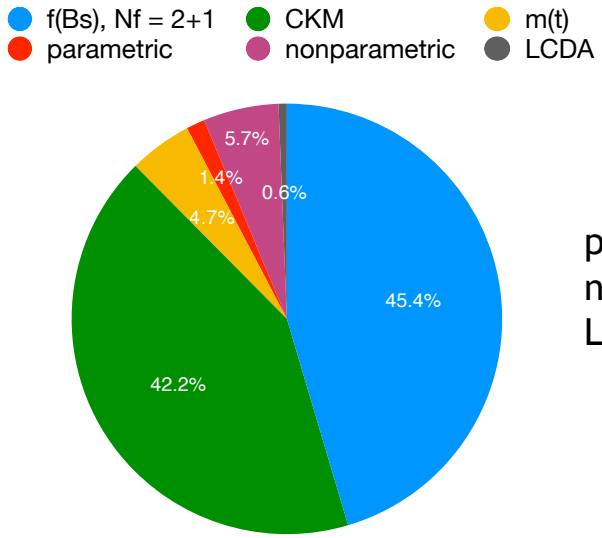
Theoretical context

1908.07011
1902.08191

- SM predictions based on
  - NNLO QCD
  - NLO EW
  - power-enhanced ll QED corrections (larger than anticipated)
  - external input (important!)
- Error budget (relative)
  - latest SM prediction (1908.07011)
  - lattice QCD: $f_{B_s^0}$ at 0.6%! (FNAL/MILC, tbc)



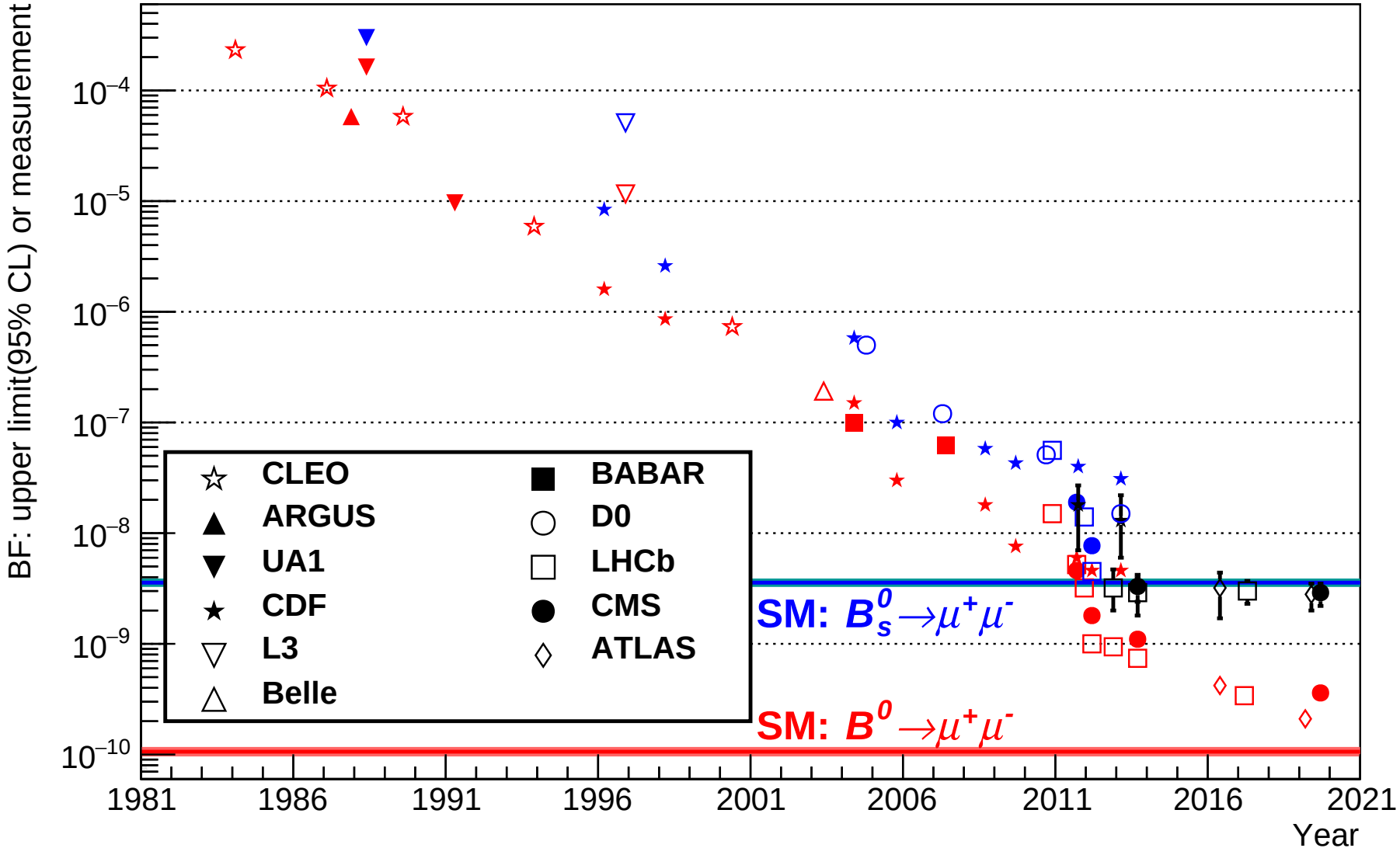




parametric: Γ_s, α_s
nonparametric: μ_W, μ_b, h.o.
LCDA: λ_B, $\sigma_{1,2}$

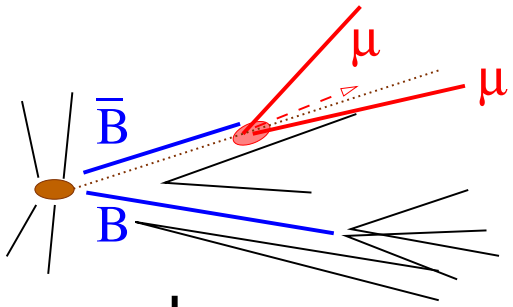
Experimental context

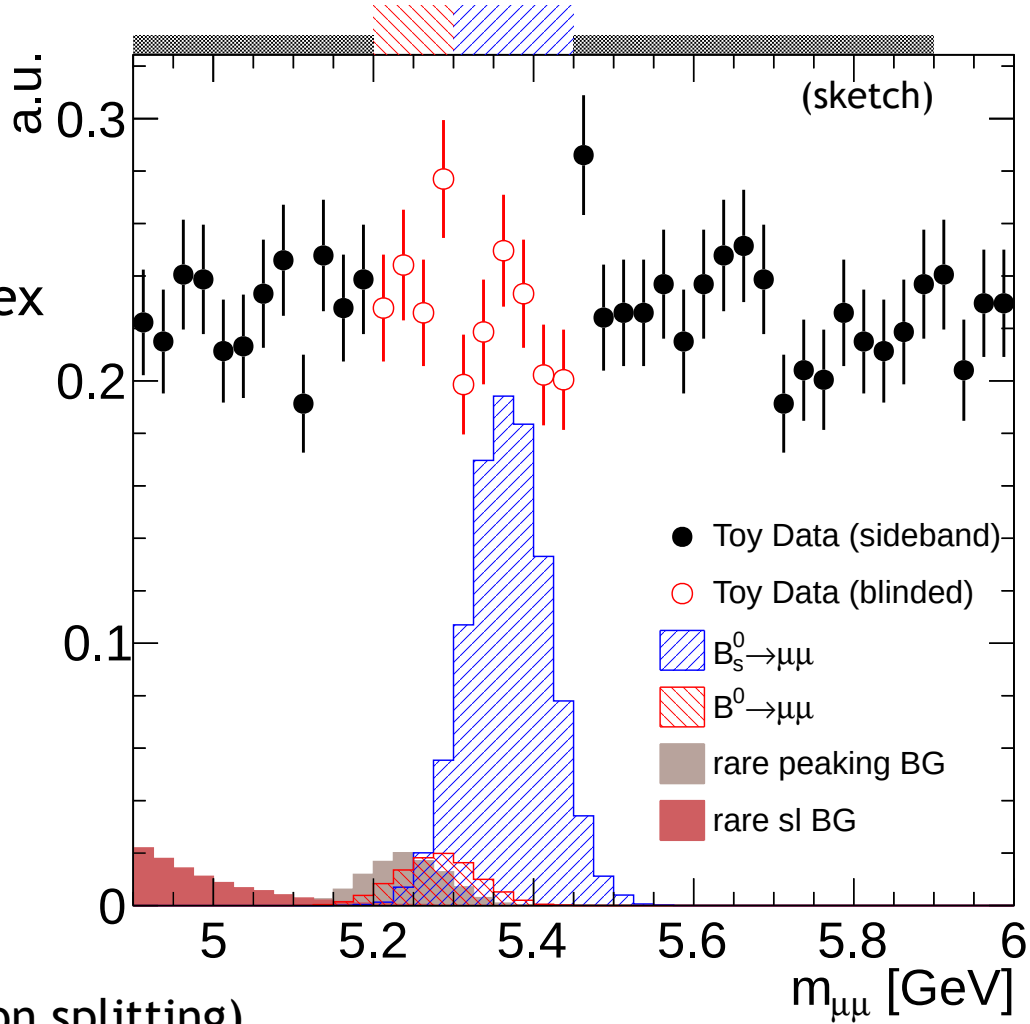
- $\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$ with long history:



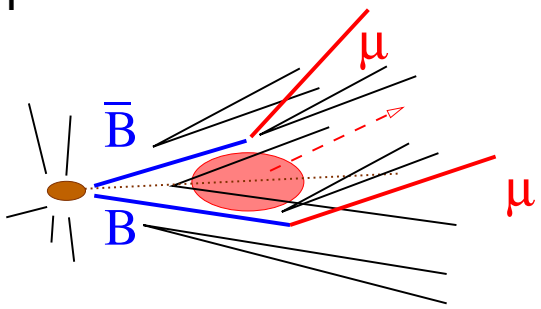
Analysis overview

- Signal $B_s^0 \to \mu^+\mu^-$
  - ▷ two muons from one decay vertex
    - invariant mass
    - secondary vertex, isolated
    - momentum in flight direction





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



- The goal: keep signal, very strongly reduce background!

Methodology

- Measurement of $B_s^0 \to \mu^+\mu^-$ relative to normalization channel:

$$\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-) = \frac{n^{\rm obs}_{B_s^0}}{N(B^+ \to J/\psi\, K^+)} \frac{\varepsilon^{tot}_{B^+}}{\varepsilon^{tot}_{B_s^0}} \frac{f_u}{f_s} \mathcal{B}(B^+ \to J/\psi\,[\mu^+\mu^-]K)$$

  - $B^+ \to J/\psi\, K^+$, $J/\psi \to \mu^+\mu^-$, with $\mathcal{B}(B^+ \to J/\psi\, K^+) = (1.01 \pm 0.03) \times 10^{-3}$

- Reconstructed decays for this result:
  - $B \to \mu^+\mu^-$: 'signal' sample
  - $B^+ \to J/\psi\, K^+$: 'normalization' sample
  - $B_s^0 \to J/\psi\, \phi$: 'control' sample for B_s^0 mesons

- Analysis steps
  - strict muon identification with boosted decision tree
  - tight candidate selection with (another) boosted decision tree
  - unbinned (extended) maximum likelihood fits to selected events
    - branching fractions $\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$
    - effective lifetime $\tau_{\mu^+\mu^-}$

Candidate selection

Dataset

- Combined (re-)analysis of Run 1 and 2016 datasets
  - 2011: 5 fb^{-1} at $\sqrt{s} = 7$ TeV
  - 2012: 20 fb^{-1} at $\sqrt{s} = 8$ TeV
  - 2016: 36 fb^{-1} at $\sqrt{s} = 13$ TeV, subdivided into 2 blocks (2016A and 2016B)

  → 4 running periods
- Trigger in 2016 very similar to 2012
  - except for

    restriction to $|\eta_\mu^{\mathrm{f}}| < 1.4$ instead of $|\eta_\mu^{\mathrm{f}}| < 2.1$ (η_μ^{f}: pseudorapidity of most forward muon)

    efficiency loss at larger decay times

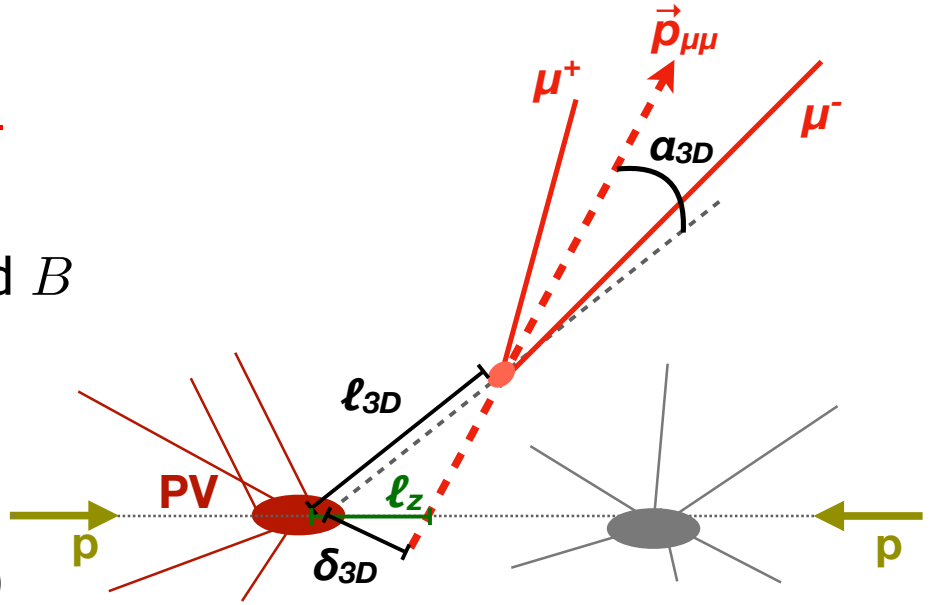
    because of trigger rate limitations in 2016
- Separation of data into 'channels'
  - 'central' and 'forward'

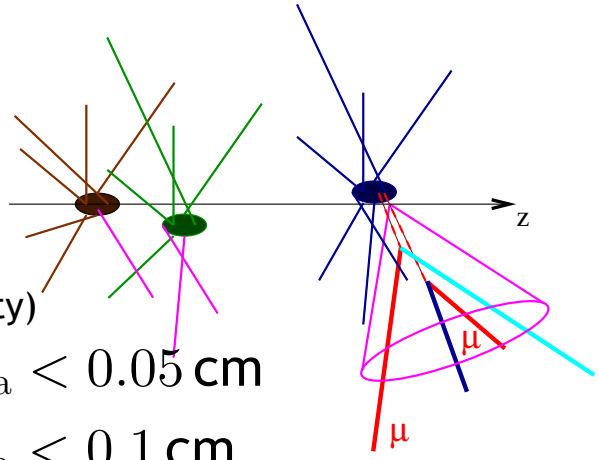
  → NB: 2016 without 'endcaps'
- Blind analysis
  - 'reblinded' Run 1

Year	central	forward
2011	$0 < \lvert\eta_\mu^{\mathrm{f}}\rvert < 1.4$	$1.4 < \lvert\eta_\mu^{\mathrm{f}}\rvert < 2.1$
2012	$0 < \lvert\eta_\mu^{\mathrm{f}}\rvert < 1.4$	$1.4 < \lvert\eta_\mu^{\mathrm{f}}\rvert < 2.1$
2016	$0 < \lvert\eta_\mu^{\mathrm{f}}\rvert < 0.7$	$0.7 < \lvert\eta_\mu^{\mathrm{f}}\rvert < 1.4$

Variables

- B candidate and its tracks
  - kinematics: η and $p_\perp$ of muons and B
- Vertexing
  - primary vertex (pp collision)
    - choice with ℓ_z
    - impact parameter: δ_{3D}, $\delta_{3D}/\sigma(\delta_{3D})$
  - secondary vertex (B-meson decay)
    - pointing angle: α_{3D}
    - flight length: ℓ_{3D}, $\ell_{3D}/\sigma(\ell_{3D})$
    - track $d_{\rm ca}$, vertex fit $\chi^2/{\rm dof}$





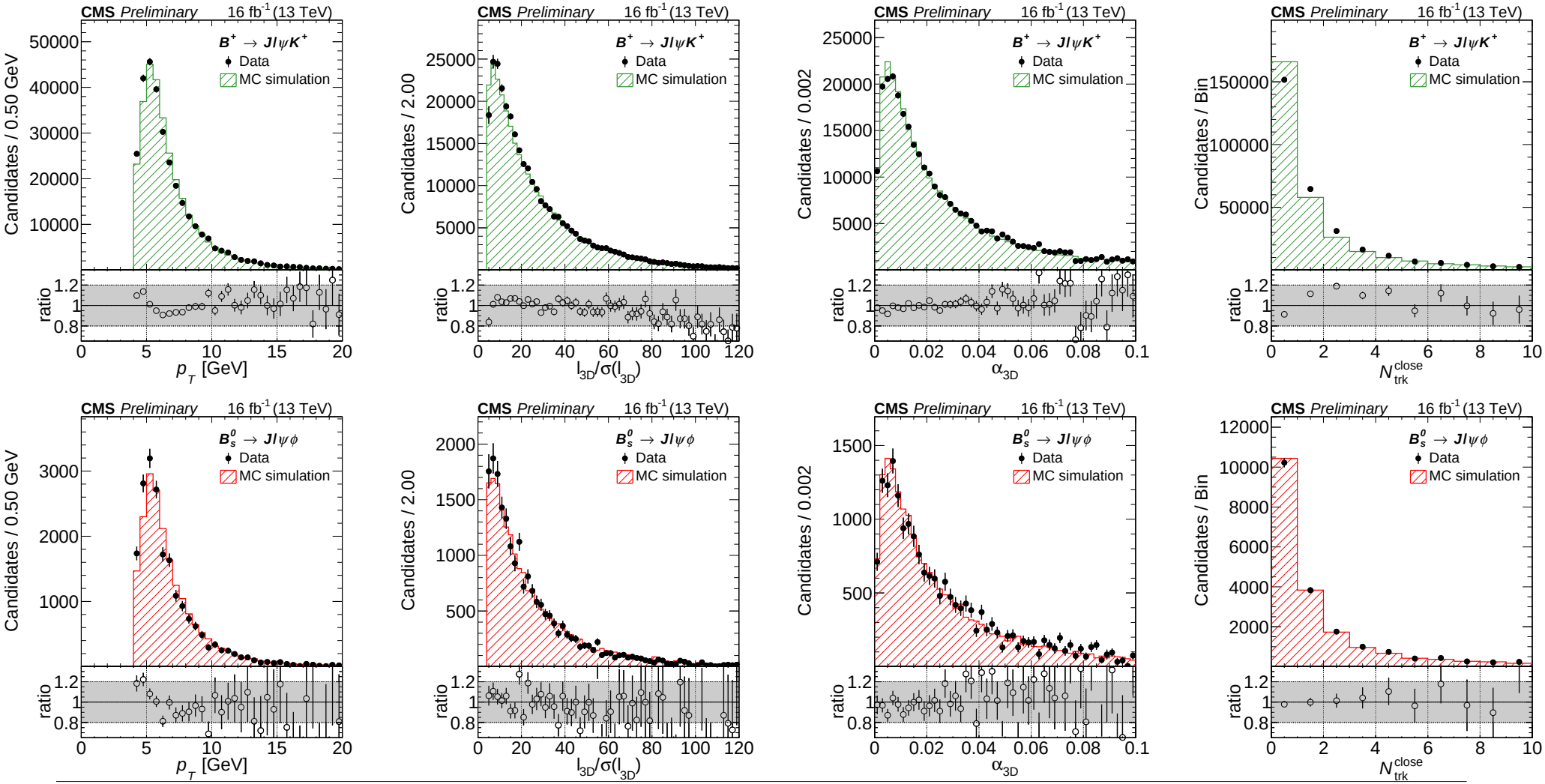
- Isolation (optimized cuts for background rejection and data/MC similarity)
  - $I \equiv p_{\perp B}/(p_{\perp B} + \sum_{\rm trk} p_\perp)$: $p_\perp > 0.9, \Delta R < 0.7, d_{\rm ca} < 0.05\,$cm
  - $I_\mu \equiv p_{\perp \mu}/(p_{\perp \mu} + \sum_{\rm trk} p_\perp)$: $p_\perp > 0.5, \Delta R < 0.5, d_{\rm ca} < 0.1\,$cm
  - $N_{\rm trk}^{\rm close}$: count tracks with $p_\perp > 0.5\,$GeV and $d_{\rm ca} < 0.03\,$cm
  - $d_{\rm ca}^0$: minimum $d_{\rm ca}$ of these tracks to B-SV

    ($\sum_{\rm trk}$ w/ tracks from B-PV or no other PV, but passing $d_{\rm ca}$ requirement)

Simulation vs reality

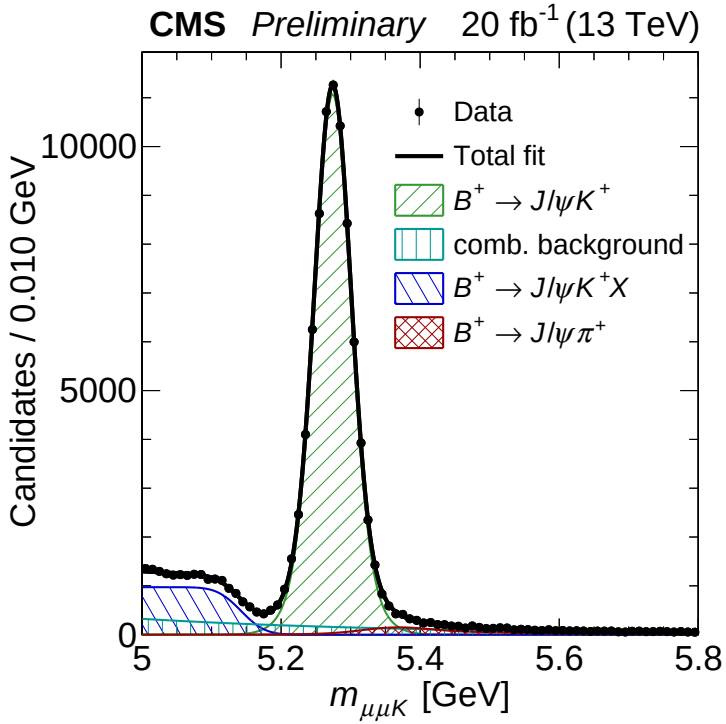
- Comparison of normalization and control samples in data and MC
  - differences accounted for in systematic uncertainty

(three most discriminating variables)

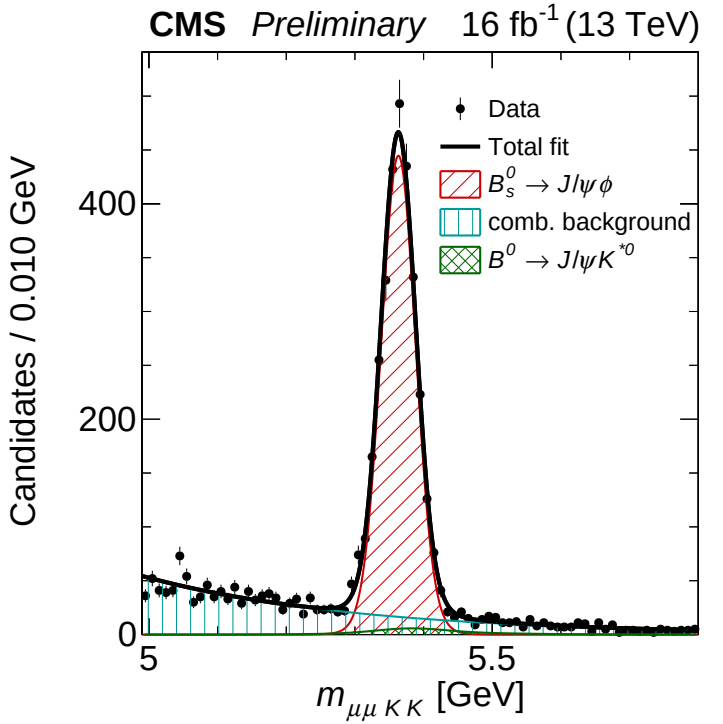


$B \to J/\psi X$ fitting

- Fits to invariant mass
  - signal: double Gaussians with common mean
  - background $B^+ \to J/\psi K^+$:
    - combinatorial: exponential
    - partially reco'ed bg: error function
    - $B^+ \to J/\psi \pi^+$: triple Gaussian, constrained to 4% of signal, shape fixed to MC expectation
  - background $B_s^0 \to J/\psi \phi$:
    - combinatorial: exponential
    - $B^0 \to J/\psi K^{*0}$: triple Gaussian, plus small linear component, shape fixed to MC expectation
- 4% systematic uncertainty on yield
  - difference to fit with J/ψ mass constraint



2016A central B^+

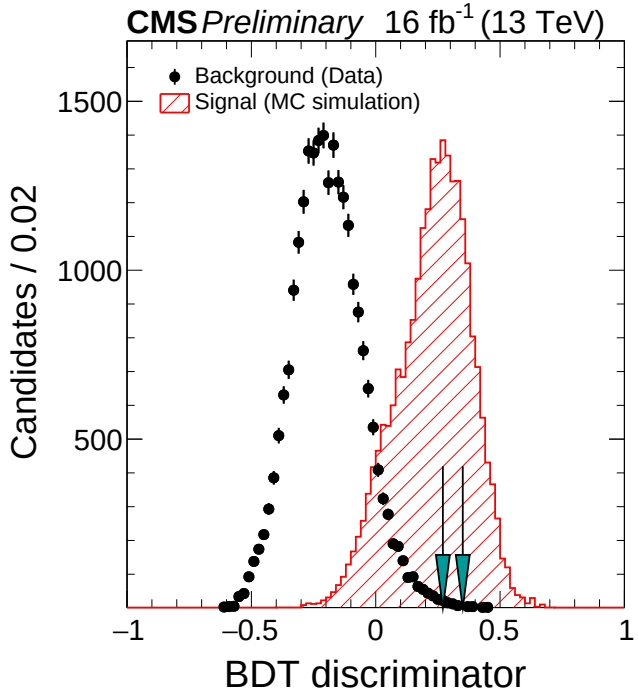


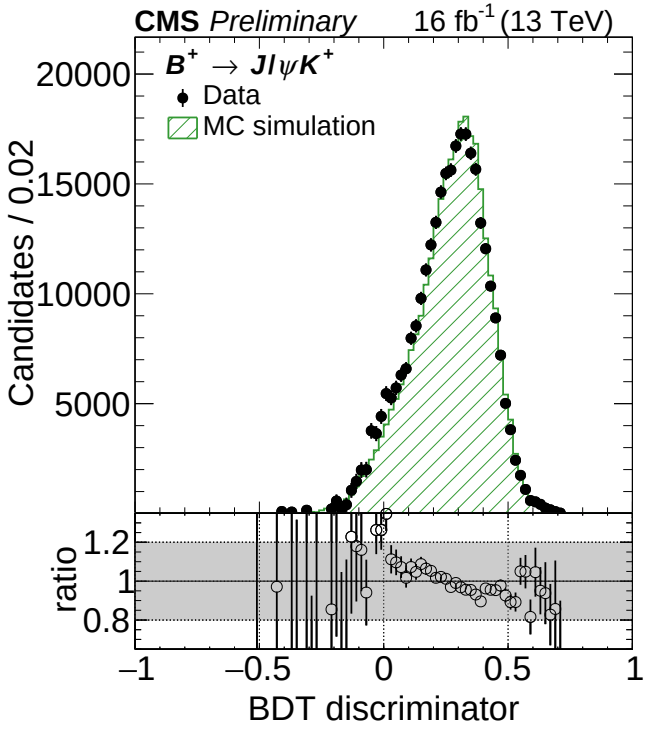
2016B central B_s^0

Multi-variate analysis

- Boosted decision tree
  - ▷ Run 1: BDT unchanged wrt PRL,111,101804
  - ▷ 2016: new BDT trained (same variables)
- BDT training (TMVA)
  - ▷ signal: $B_s^0 \to \mu^+\mu^-$ MC simulation
  - ▷ background: data dimuon sidebands
  - ▷ avoid selection bias
    - split data randomly into three subsets (0,1,2)
    - train on 0, test on 1, apply on 2. etc.
  - $\to$ in each channel, have 3 BDTs
  - ▷ many validation studies
- Systematic uncertainty
  - ▷ double ratio D $\quad D = \dfrac{\left[\frac{\varepsilon(B^+\to J/\psi\, K^+)}{\varepsilon(B_s^0\to J/\psi\,\phi)}\right]_{data}}{\left[\frac{\varepsilon(B^+\to J/\psi\, K^+)}{\varepsilon(B_s^0\to J/\psi\,\phi)}\right]_{MC}}$
  - ▷ 5-10% on efficiency ratio
  - ▷ 0.07 ps on effective lifetime

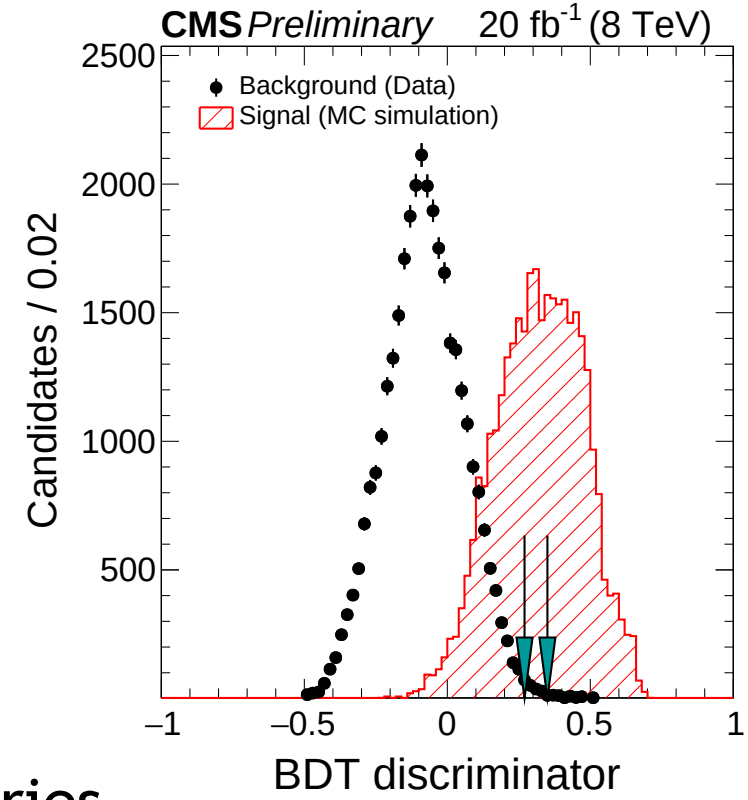
  (largest component in systematic error, but small compared to statistical error)





Final candidate selection

- Optimal BDT discriminator binning
  - ▷ full fit frameworks
  - ▷ repeated for Run 1 data
  - ▷ Asimov data for significance estimation
  - → exclusive ‘categories’



- $\overline{\mathcal{B}}(B_s^0)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$: 14 categories

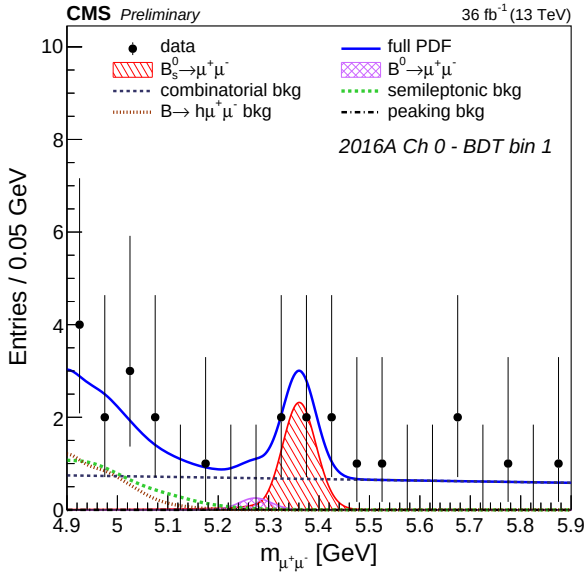
	2011		2012		2016A		2016B	
category	central	forward	central	forward	central	forward	central	forward
low	n/a	n/a	0.27	0.23	0.19	0.19	0.18	0.23
high	0.28	0.21	0.35	0.32	0.30	0.30	0.31	0.38

- effective $B_s^0 \to \mu^+\mu^-$ lifetime: 8 categories

	2011		2012		2016A		2016B	
channel	central	forward	central	forward	central	forward	central	forward
BDT	0.22	0.19	0.32	0.32	0.22	0.30	0.22	0.29

Branching fractions $\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$

Fit model



- 3D Fit for $\bar{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$

$$P(m_{\mu\mu};\sigma(m_{\mu\mu})) \times P(\sigma(m_{\mu\mu})/m_{\mu\mu}) \times P(\mathcal{C})$$

  - dimuon mass $m_{\mu\mu}$
  - per-event dimuon mass resolution $\sigma(m_{\mu\mu})$
  - $\mathcal{C}$: binary distribution for dimuon bending configuration (against possible bias)
    $\mathcal{C}(\pm 1)$: bending towards (away from) each other

- Components of model

Component	Mass	Width	Mass resolution
Signal	CB	cond., $\sigma_{\text{CB}} = \kappa \times \sigma(m_{\mu\mu})$	KEYS
Background hh	CB+G	cond.	KEYS
Background $h\mu\mu$, $h\mu\nu$	KEYS	n/a	KEYS
Combinatorial background	Bernstein pol1	n/a	KEYS (sideband)

(CB: crystal-ball, G: Gaussian)

  - 2 parameters of interest: $\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$ and $\mathcal{B}(B^0 \to \mu^+\mu^-)$
  - constraints on nuisance parameters
    - gaussian: f_s/f_u, $B^+ \to J/\psi\, K^+$, efficiency ratios
    - lognormal: rare background yields

- 1-3% systematic error from unknown $B_s^0 \to \mu^+\mu^-$ eff. lifetime

Rare background yields

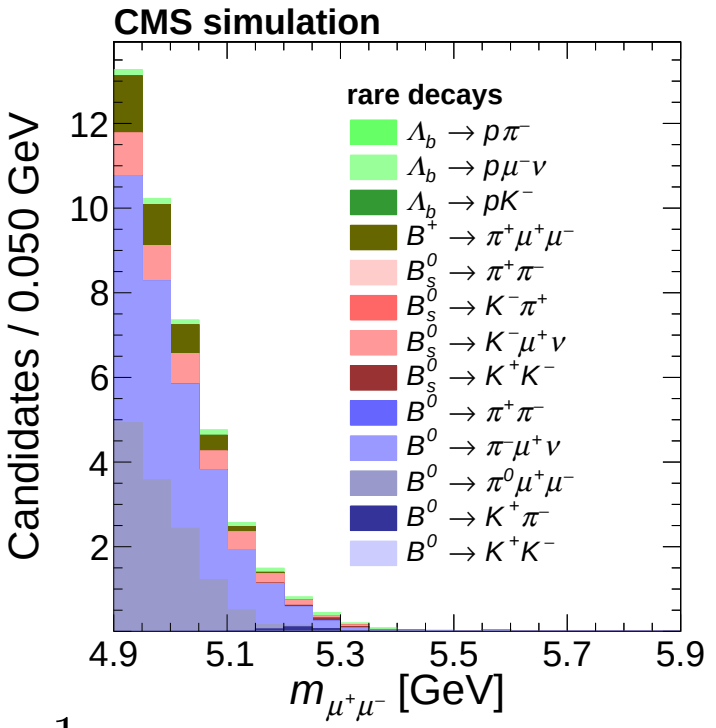
- Rare background yield expectations
  - ▷ known branching fractions
  - ▷ absolute yield from normalization sample:

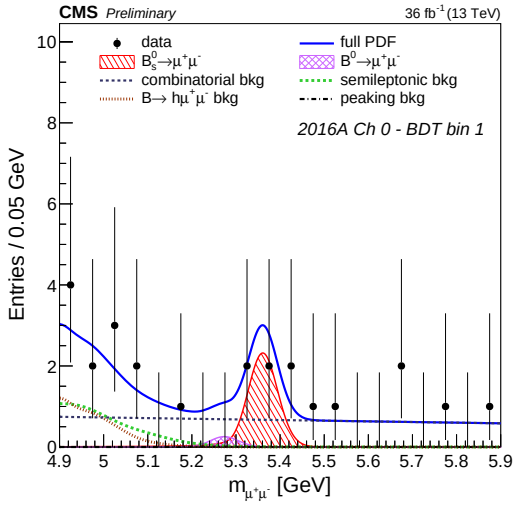
$$n^{\exp}_{B_x\to hh} = \frac{\varepsilon^{tot}_{B}}{\varepsilon^{tot}_{B^+}} \frac{f_u}{f_x} \frac{\mathcal{B}(B_x\to hh)}{\mathcal{B}(B^+\to J/\psi\,[\mu^+\mu^-]K)} \times N(B^+\to J/\psi\,[\mu^+\mu^-]K)$$

with

$$\varepsilon^{B\to hh}_{tot} = w_+(p_\perp,\eta)\times w_-(p_\perp,\eta)\times \varepsilon^{(BDT)}_{ana}\times A\times \frac{1}{2}\varepsilon^{signal}_{trig}$$

  - ▷ rare hadronic decays: complete set
  - ▷ rare sl decays: incomplete set/low statistics
    - → scale factor in low sideband
  - ▷ extensive validation with inverted muon ID selection
  - ▷ new muon ID: peaking background is very small





Results I

CMS-PAS-BPH-16-004

- Branching fractions from 3D UML fit:

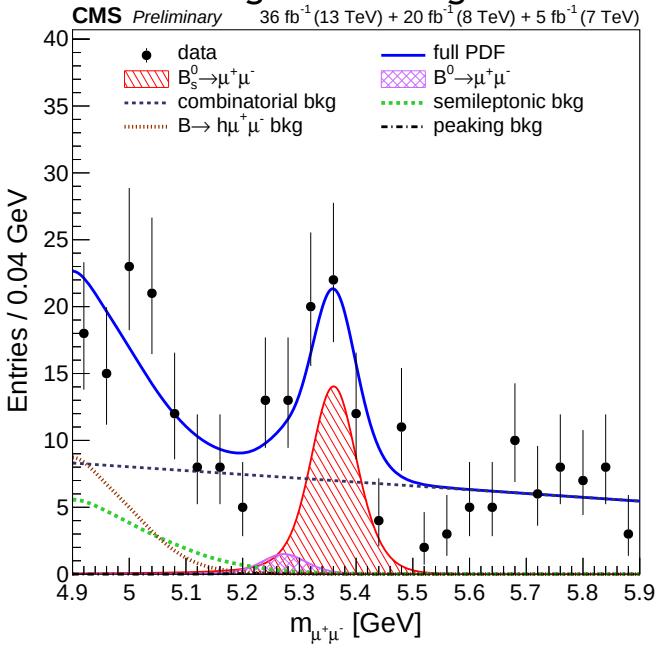
$$\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-) = [2.9^{+0.7}_{-0.6}(\text{exp}) \pm 0.2(f_s/f_u)] \times 10^{-9}$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) = (0.8^{+1.4}_{-1.3}) \times 10^{-10}$$

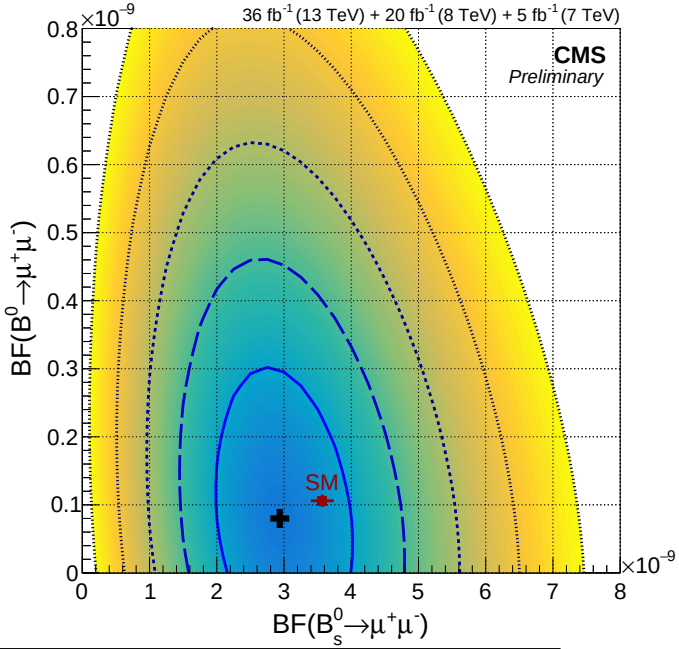
  - Observed (expected) significance:
    $B_s^0 \to \mu^+\mu^-$: $5.6(6.5)\,\sigma$
    $B^0 \to \mu^+\mu^-$: $0.6(0.8)\,\sigma$
  - average $p_\perp(B_s^0 \to \mu^+\mu^-)$: 17.2 GeV
    for future rescaling (f_s/f_u)
  - correlation coefficient: -0.181

$\Rightarrow$ Consistent with SM

- Result on Run-1 subset
  - $\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-) = (2.3^{+1.0}_{-0.8}) \times 10^{-9}$
  - obs (exp) significance: $3.3(4.5)\sigma$
  - consistent with Nature,522,68 (restricted to CMS subset)

Combined mass projection for high-BDT categories



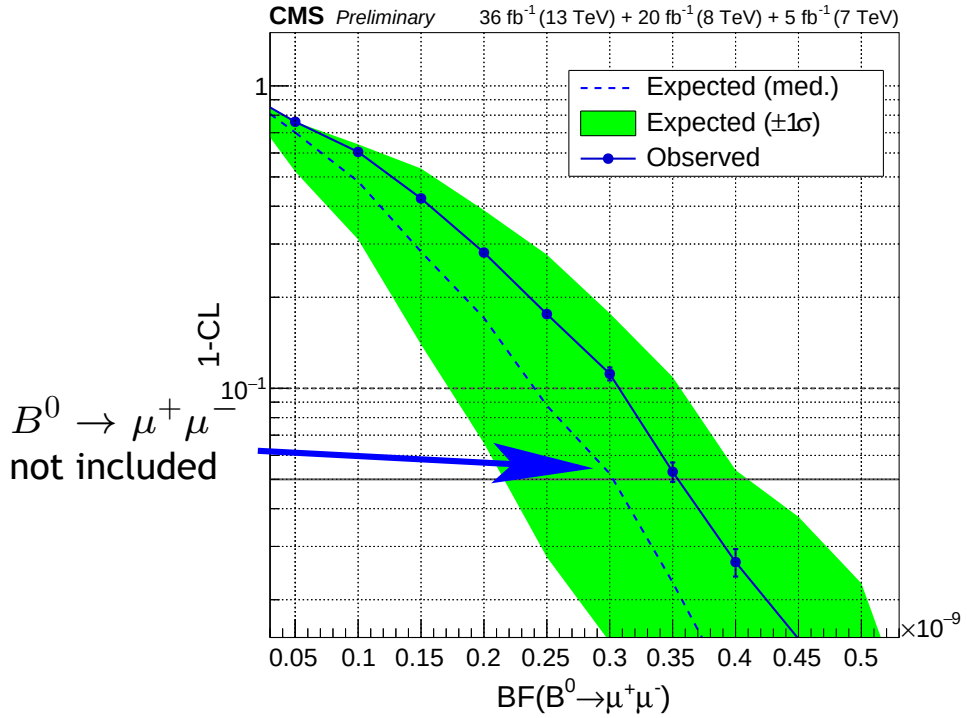


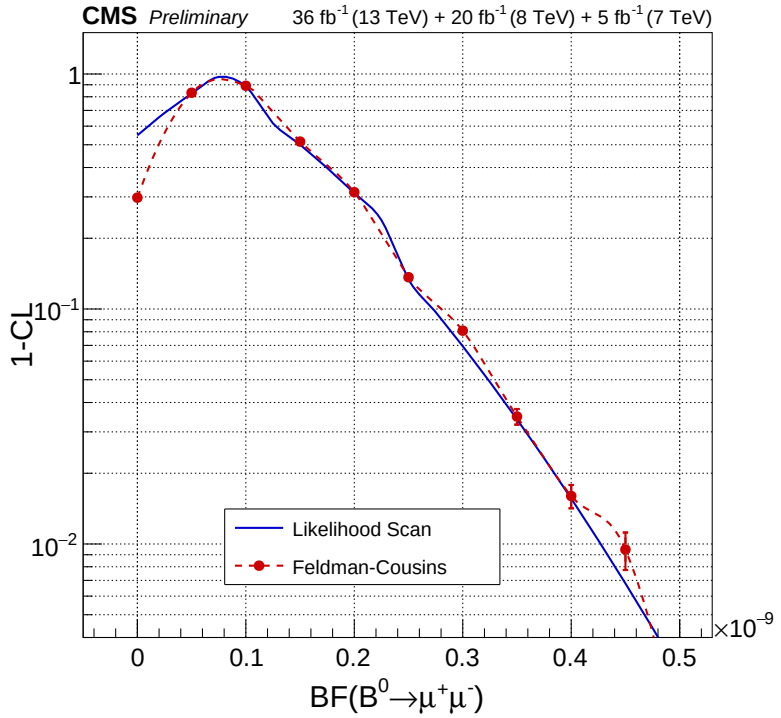
Results II

- $\mathcal{B}(B^0 \to \mu^+\mu^-)$ upper limit
  - insensitive to systematic errors (at given precision)

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.6 \times 10^{-10} \quad (95\%\,\text{CL}), \quad (\text{expected: } 3.0 \times 10^{-10})$$
$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.1 \times 10^{-10} \quad (90\%\,\text{CL}), \quad (\text{expected: } 2.4 \times 10^{-10})$$

NB: this is the main result for $\mathcal{B}(B^0 \to \mu^+\mu^-)$





Summary of systematic errors

- Uncertainties dominated by small signal sample size
  - ▷ relative errors for $\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$, absolute for $\tau_{\mu^+\mu^-}$

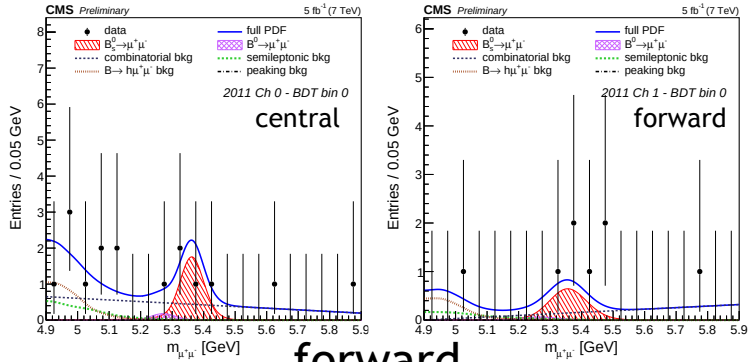
Source	$\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$ [%]	$\tau_{\mu^+\mu^-}$ [ps]	
		2D UML	*sPlot*
Kaon tracking	2.3 – 4	–	–
Normalization yield	4	–	–
Background yields	1	0.03	(*)
Production process	3	–	–
Muon identification	3	–	–
Trigger	3	–	–
Efficiency (data/MC simulation)	5 – 10	–	(*)
Efficiency (functional form)	–	0.01	0.04
Efficiency lifetime dependence	1 – 3	(*)	(*)
Era dependence	5 – 6	0.07	0.07
BDT discriminator threshold	–	0.02	0.02
Silicon tracker alignment	–	0.02	–
Finite size of MC sample	–	0.03	–
Fit bias	–	–	0.09
$\mathcal{C}$-correction	–	0.01	0.01
Total systematic uncertainty	$\binom{+0.3}{-0.2} \times 10^{-9}$	0.09	0.12
Total uncertainty	$\binom{+0.7}{-0.6} \times 10^{-9}$	$^{+0.61}_{-0.44}$	$^{+0.52}_{-0.33}$

(*) included in other item

  - ▷ successful cross check of σ_{syst} with measurement of $\mathcal{B}(B_s^0 \to J/\psi\,\phi)$

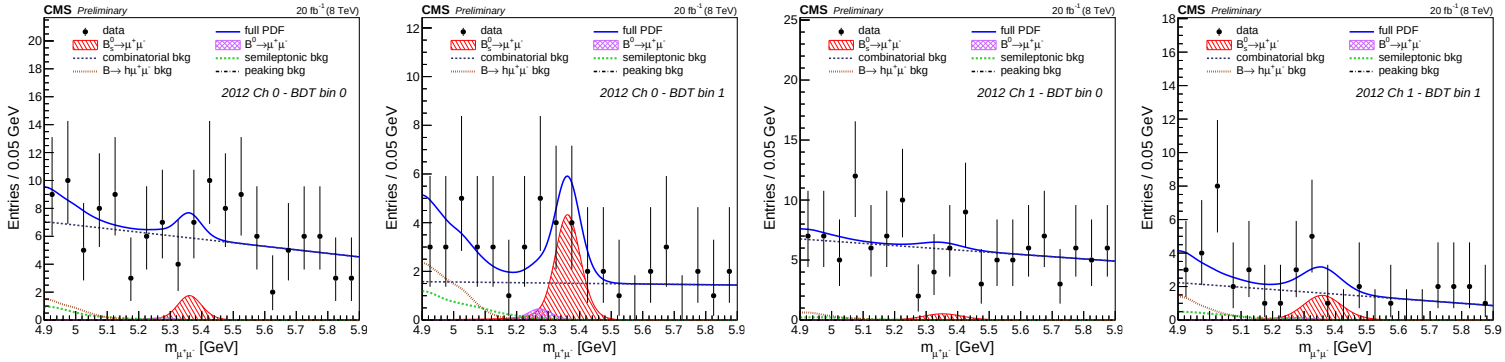
Fit details (plots)

2011

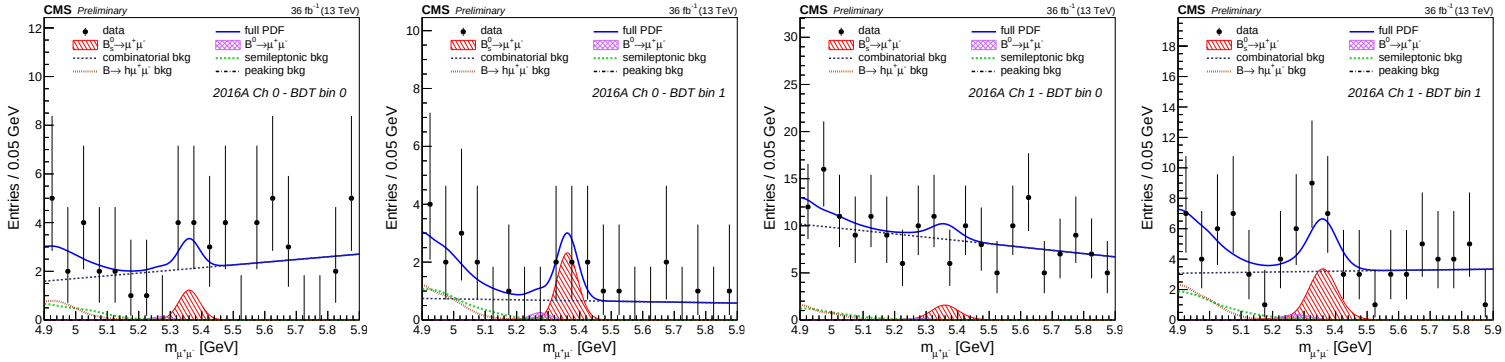


central forward

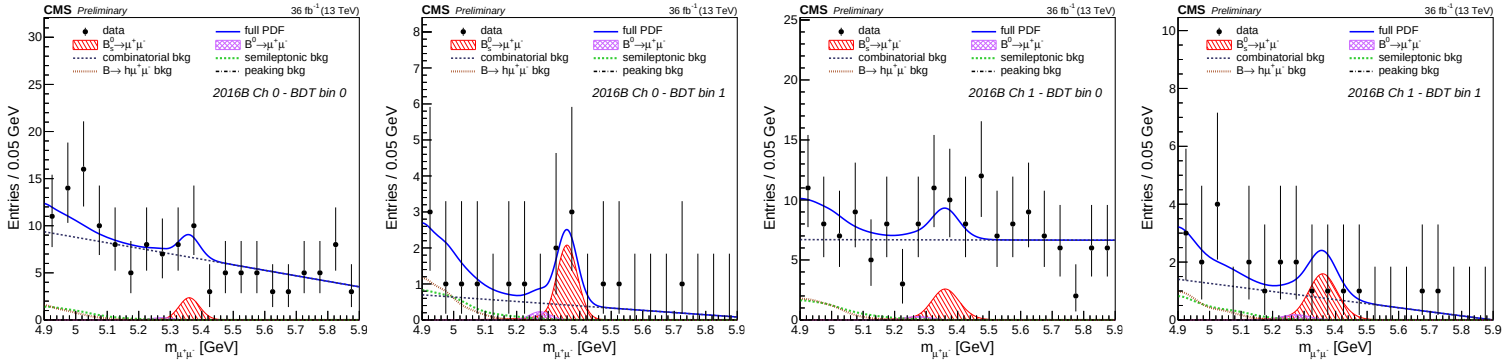
2012



2016A



2016B



Fit details (numbers)

- Obs signal yield $60.8^{+14.5}_{-13.3}$ with $\langle p_\perp \rangle = 17.2\,\text{GeV}$
  - peaking background is $\approx 5 - 10\%$ of $B^0 \to \mu^+\mu^-$ yield
  - uncertainties include statistical and systematic errors
  - signal yields (and errors) determined from $\mathcal{B}$ (and include normalization errors)

Category	$N(B_s^0)$	$N(B^0)$	N_{comb}	$N_{\text{obs}}^{B^+}/100$	$\langle p_\perp(B_s^0)\rangle$ [GeV]	$\varepsilon_{\text{tot}}/\varepsilon_{\text{tot}}^{B^+}$
2011/central/high	$3.6^{+0.9}_{-0.8}$	$0.4^{+0.7}_{-0.6}$	8.4 ± 3.8	750 ± 30	16.4	3.9 ± 0.5
2011/forward/high	$2.0^{+0.5}_{-0.4}$	$0.2^{+0.4}_{-0.3}$	3.2 ± 2.2	220 ± 12	14.9	7.5 ± 0.8
2012/central/low	$3.7^{+0.9}_{-0.8}$	$0.4^{+0.6}_{-0.6}$	115.8 ± 11.3	790 ± 32	16.1	3.8 ± 0.5
2012/central/high	$9.3^{+2.3}_{-2.1}$	$1.0^{+1.7}_{-1.6}$	30.2 ± 7.3	2360 ± 95	17.3	3.2 ± 0.4
2012/forward/low	$1.7^{+0.4}_{-0.4}$	$0.2^{+0.3}_{-0.3}$	116.7 ± 11.0	190 ± 9	14.3	7.3 ± 1.0
2012/forward/high	$4.7^{+1.2}_{-1.1}$	$0.5^{+0.9}_{-0.8}$	31.0 ± 6.5	660 ± 27	15.5	5.9 ± 0.8
2016BF/central/low	$2.2^{+0.5}_{-0.5}$	$0.2^{+0.4}_{-0.4}$	43.0 ± 7.1	580 ± 23	17.5	3.1 ± 0.4
2016BF/central/high	$4.0^{+1.0}_{-0.9}$	$0.4^{+0.8}_{-0.7}$	13.3 ± 4.7	1290 ± 57	19.3	2.5 ± 0.3
2016BF/forward/low	$3.7^{+0.9}_{-0.8}$	$0.4^{+0.7}_{-0.7}$	168.8 ± 13.5	780 ± 31	15.8	3.9 ± 0.5
2016BF/forward/high	$8.1^{+2.0}_{-1.8}$	$0.8^{+1.5}_{-1.4}$	64.2 ± 9.7	1920 ± 78	17.5	3.4 ± 0.4
2016GH/central/low	$4.1^{+1.0}_{-0.9}$	$0.4^{+0.8}_{-0.7}$	128.8 ± 12.0	1020 ± 44	17.2	3.3 ± 0.4
2016GH/central/high	$3.6^{+0.9}_{-0.8}$	$0.4^{+0.7}_{-0.6}$	7.8 ± 3.6	1320 ± 54	20.8	2.2 ± 0.2
2016GH/forward/low	$6.1^{+1.5}_{-1.4}$	$0.6^{+1.1}_{-1.0}$	133.4 ± 12.5	1260 ± 51	16.2	3.9 ± 0.4
2016GH/forward/high	$3.9^{+1.0}_{-0.9}$	$0.4^{+0.8}_{-0.7}$	14.1 ± 4.6	1180 ± 49	19.5	2.7 ± 0.3

A note on f_s/f_u

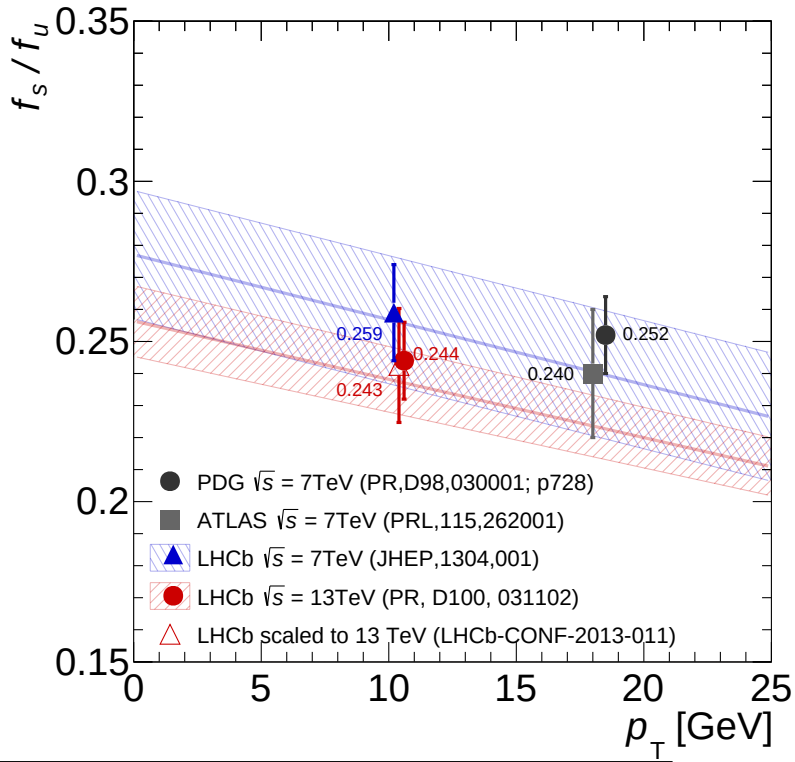
PRL,115,262001
PR,D100,031102

- f_s/f_u is external input for $\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$
  - ▷ experimental situation not entirely clear
    - LHCb sees $p_\perp$-dependence (PR,D100,031102)
    - ATLAS does not see $p_\perp$-dependence (PRL,115,262001)
    - CMS does not see $p_\perp$-dependence (internal study with control sample)
  - ▷ fragmentation fraction x_B not measured
    (x_B: fraction of b momentum $\to$ B)

- Ad hoc error added
  - ▷ PDG $f_s/f_u = 0.252 \pm 0.012$, based on
    $\sqrt{s} = 7\,\text{TeV}$ results of LHCb/ATLAS
  - ▷ additional ad-hoc error
    difference between PR,D100,031102 and PDG
    $p_\perp$ dependence from PR,D100,031102

$$f_s/f_u \quad = \quad 0.252 \pm 0.012(\text{exp.}) \pm 0.015(\text{CMS})$$

$\Rightarrow$ Our result can be rescaled
  - ▷ $\sqrt{s}$ and $p_\perp$ of signal candidates provided
    for each category/channel/running period



Effective lifetime $\tau_{\mu^+\mu^-}$

two methods:
(1) 2D UML fit
(2) sPlot ML fit

Decision on primary method based on expected median performance
(decided before unblinding)

Lifetime fitting

- Determination of proper decay time $t = m\,\ell_{3D}/p$ in 3D space
- 2D unbinned extended maximum likelihood fit to
  - B mass and t decay time in the range $1 < t < 11$ ps
    (σ_t as conditional parameter, complete propagation of uncertainties)
  - Efficiency correction (mostly HLT)
  - model components

mass	shape	source	fit params
Signal	CB	MC	fixed
BG $h\mu\nu$, $h\mu\mu$	G	w8-MC	fixed
BG hh, $B^0 \to \mu^+\mu^-$	CB+G	w8-MC	fixed
Combinatorial BG	Bernstein pol1	sideband	floating
decay time	**shape**	**source**	**fit params**
Signal	expo$\otimes$res$^{(*)}$	MC	floating
BG $h\mu\nu$, $h\mu\mu$	expo$\otimes$res	w8-MC	fixed
BG hh, $B^0 \to \mu^+\mu^-$	expo$\otimes$res	w8-MC	fixed
Combinatorial BG	expo$\otimes$res	sideband	floating

$(*)$ 'res' includes resolution and efficiency (no efficiency correction for the combinatorial background)

- $sPlot$ lifetime fit
  - $sPlot$ weights from $\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-)$ model
  - binned maximum likelihood fit with resolution and efficiency modeling
  - custom algorithm for correct (asymmetric) uncertainties

$\Rightarrow$ Consistent results between the two setups

Results

- Primary result from 2D UML fit:

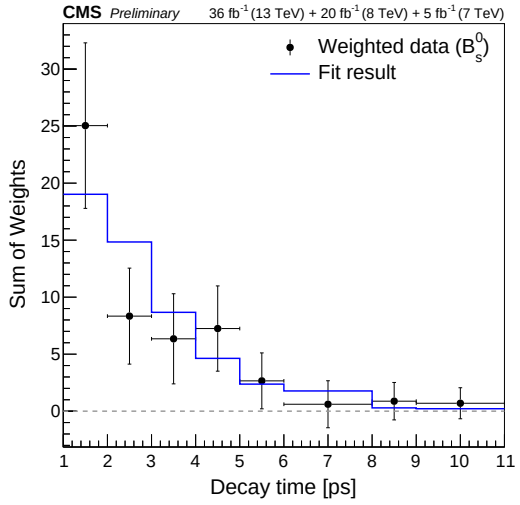
$$\tau(B_s^0 \to \mu^+\mu^-) = 1.70^{+0.61}_{-0.44}\,\text{ps}$$

  - systematic error small: 0.09 ps
  - expected error: $(^{+0.39}_{-0.30})$ ps

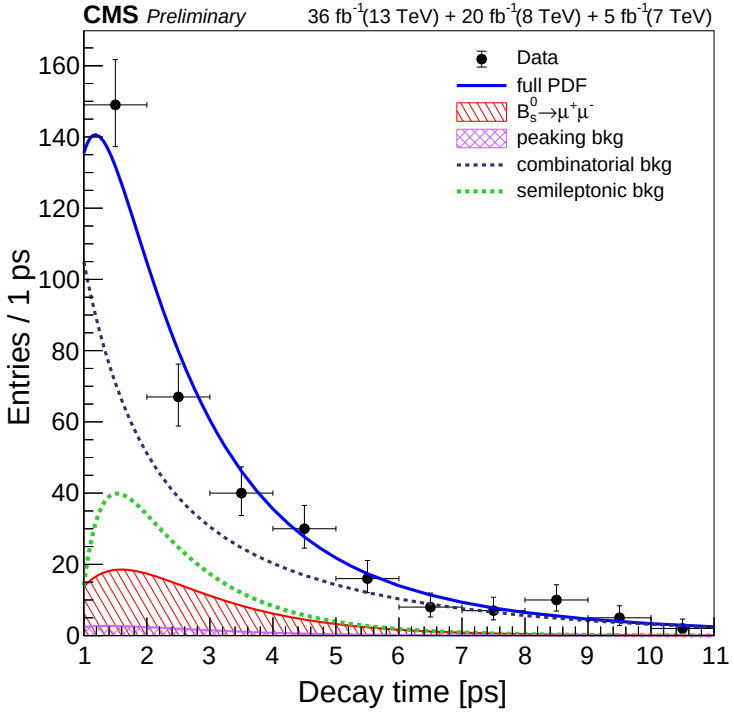
$\Rightarrow$ Consistent with SM

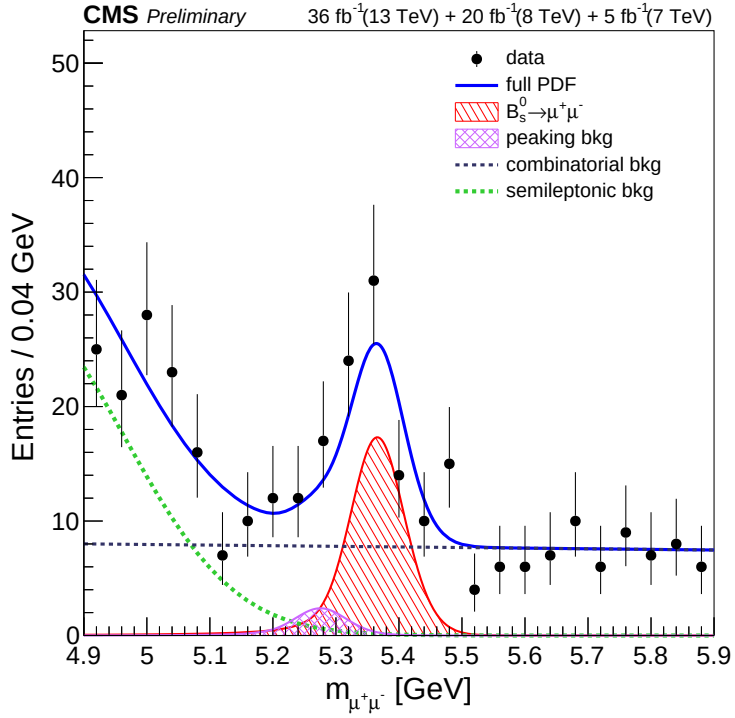
- Result from $sPlot$ method:

$$\tau(B_s^0 \to \mu^+\mu^-) = 1.55^{+0.52}_{-0.33}\,\text{ps}$$



expected error: $(^{+0.49}_{-0.31})$ ps

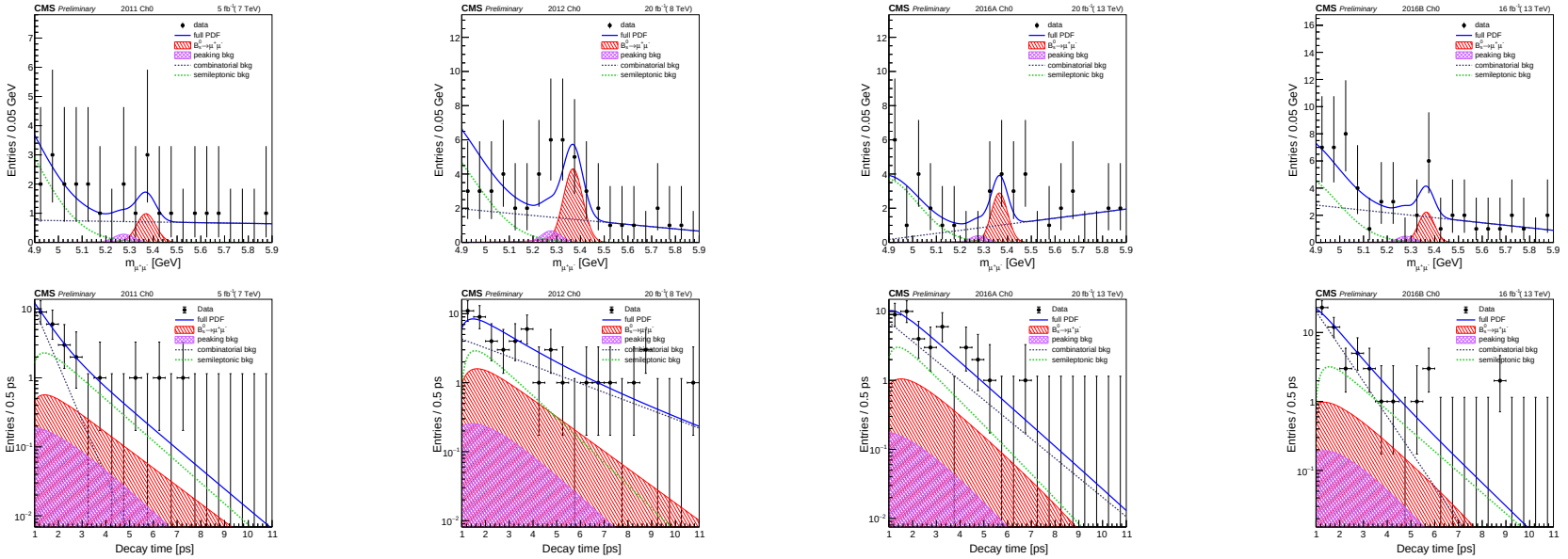




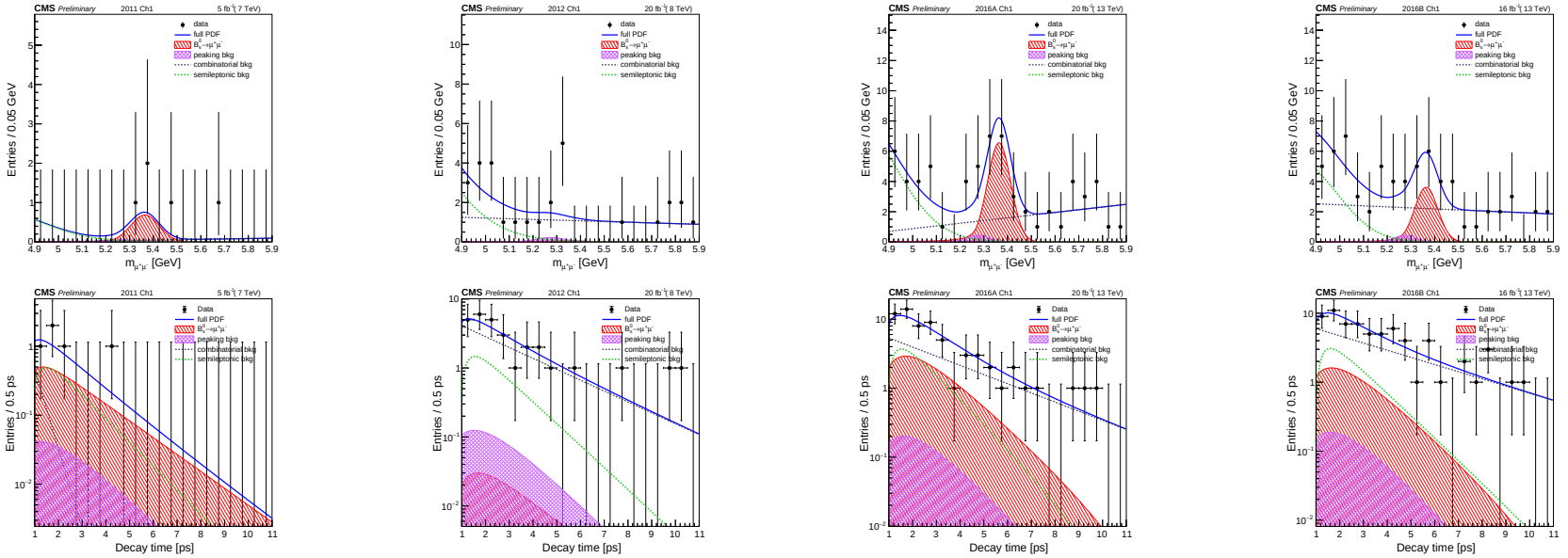
In all channels/eras

2011 2012 2016A 2016B

central



forward



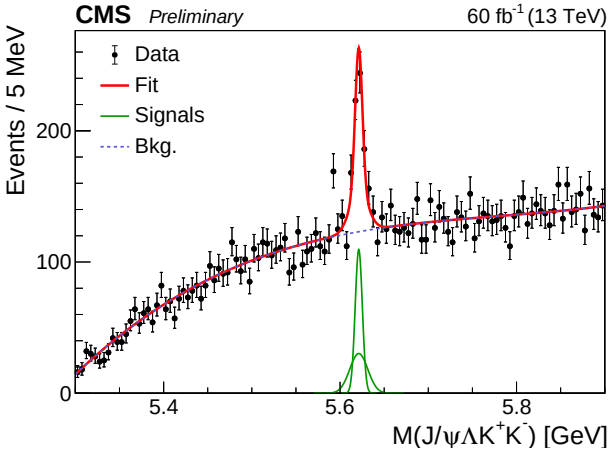
Conclusions

CMS-PAS-BPH-19-002
CMS-PAS-BPH-16-004

- First observation of $\Lambda_b^0 \to J/\psi\,\phi\Lambda$

$$\frac{\mathcal{B}(\Lambda_b^0 \to J/\psi\,\phi\Lambda)}{\mathcal{B}(\Lambda_b \to \psi(2S)\Lambda)} = (8.26 \pm 0.90_{\text{stat}} \pm 0.69_{\text{syst}+\mathcal{B}})\%$$

  - Study substructure



- $B_s^0 \to \mu^+\mu^-$ and $B^0 \to \mu^+\mu^-$ decays with Run 1 and 2016 data
  - update of branching fraction measurements

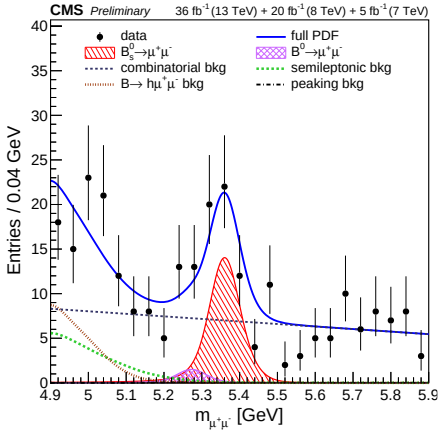
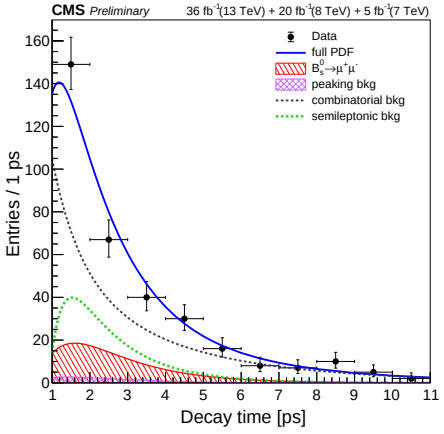
$$\overline{\mathcal{B}}(B_s^0 \to \mu^+\mu^-) = [2.9^{+0.7}_{-0.6}(\text{exp}) \pm 0.2(f_s/f_u)] \times 10^{-9}$$

$$\mathcal{B}(B^0 \to \mu^+\mu^-) < 3.6 \times 10^{-10} \quad (95\%\text{CL})$$

($B_s^0 \to \mu^+\mu^-$ significance: $5.6\,\sigma$ obs, $6.5\,\sigma$ exp, these results supersede PRL,111,101804)

  - first $\tau_{\mu^+\mu^-}$ measurement of CMS

$$\tau_{\mu^+\mu^-} = 1.70^{+0.61}_{-0.44}\ \text{ps}$$

  - all results consistent with SM

Complete Run-2 analyses underway