Measurement of the CP-violating phase ϕ_s in the $B_s^0 \rightarrow J/\psi\, \phi$ channel at 13 TeV by CMS

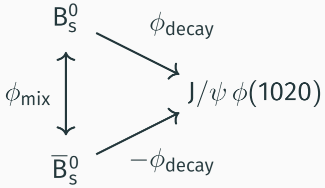
Alberto Bragagnolo[a], on behalf of the CMS Collaboration

LHCP 2020 – 29/05/2020

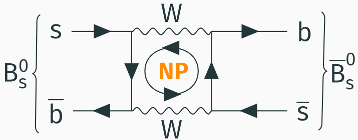
[a]University & INFN Padova (IT)

Introduction

- ϕ_s is a CPV phase arising from the **interference** between B^0_s decays proceeding directly and through B^0_s-$\overline{B}^0_s$ mixing to a CP final state
- **SM prediction**: $\phi_s \simeq -2\beta_s = -36.96^{+0.84}_{-0.72}$ mrad [CKMfitter]
- New Physics can **change** the value of ϕ_s up to $\sim 10\%$ via new particles contributing to the B^0_s-$\overline{B}^0_s$ mixing [JHEP04(2010)031]
- **$B^0_s \to J/\psi\,\phi$ is the golden channel to measure ϕ_s**
  - No direct CPV
  - Only one CPV phase
  - Easy to reconstruct with high S/B
- **Several other interesting observables** measurable with the same analysis: Γ_s, $\Delta\Gamma_s$, $|\lambda|$, Δm^2_s
  - $\lambda = \frac{q}{p}\frac{\overline{A}_{f.s.}}{A_{f.s.}}$, $|B_{L,H}\rangle = p|B^0_s\rangle \pm q|\overline{B}^0_s\rangle$



$$\boxed{\phi_s = \phi_{mix} - 2\phi_{decay}}$$

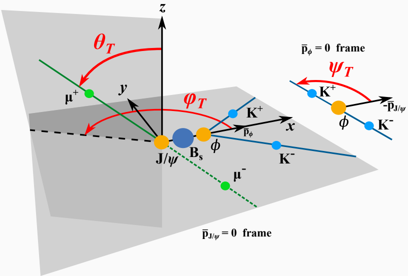


Measurement ingredients

$$\boxed{a_{CP}(t) \propto \eta_{\mu\mu KK} \sin(\phi_s) \sin(\Delta m_s t)}$$

$$\boxed{\text{sensitivity} = f\left(\sqrt{\frac{P_{tag} S}{2}}\sqrt{\frac{S}{S+B}} \cdot e^{-\frac{\sigma_t{}^2 \Delta m_s^2}{2}}\right)}$$

1. **Angular analysis** to separate the different CP eigenstate of the final state
   - ψ_T: helicity angle of K^+ in the ϕ rest frame
   - θ_T: polar angle of μ^+ in the J/ψ rest frame
   - ϕ_T: azimuthal angle of μ^+ in the J/ψ rest frame
2. **Excellent time resolution** to see the fast B_s^0-$\overline{B}_s^0$ oscillation
3. **Highly efficient flavour tagging** to infer the initial B_s^0 flavour
4. **As much statistics as possible** (with good S/N)

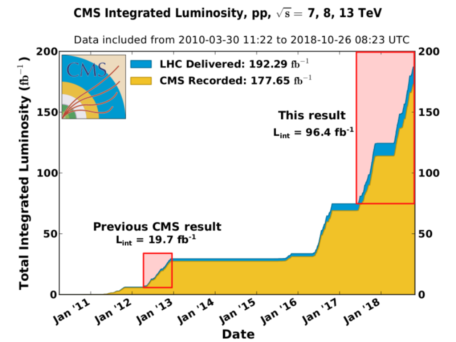


Candidate selection

Trigger strategy

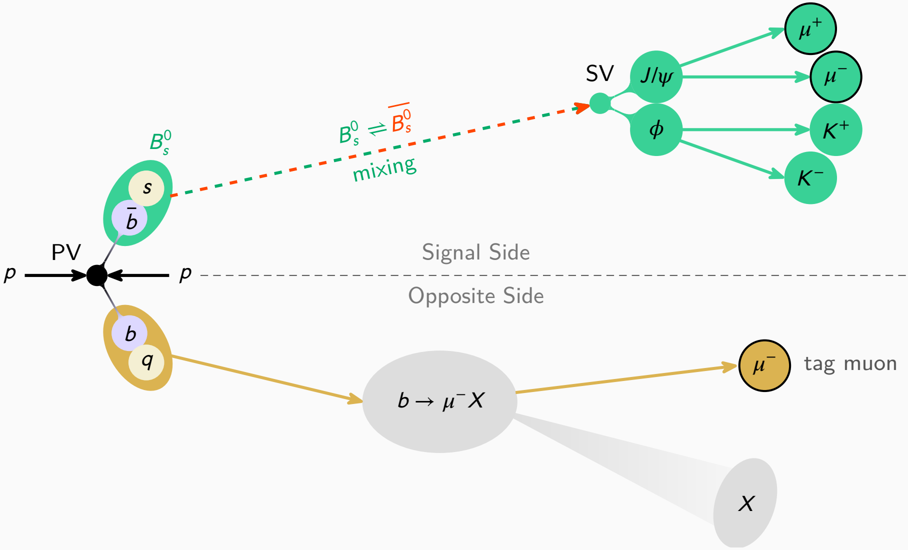
Trigger: $J/\psi \to \mu^+\mu^-$ candidate **plus an additional muon**

- The additional muon is used to **tag** the flavour of the B_s^0, via $b \to \mu^- X$ decays of the other b
- However, the requirement for a third muon **lowers** the rate of selected events
- **Not to apply a displacement cut** on the $J/\psi \to \mu^+\mu^-$ at HLT level



This trigger improves the tagging efficiency at the cost of the reduced number of signal events

Schematic representation of an event



Offline selection

Offline selection

$p_T(\mu)$	≥ 3.5 GeV
$\lvert\eta(\mu)\rvert$	≤ 2.4
$p_T(K)$	≥ 1.2 GeV
$\lvert\eta(K)\rvert$	≤ 2.5
$\left\lvert m(\mu^+\mu^-) - m^{PDG}_{J/\psi}\right\rvert$	< 150 MeV
$\left\lvert m(K^+K^-) - m^{PDG}_{\phi(1020)}\right\rvert$	< 10 MeV
$p_T(B^0_s)$	≥ 11 GeV
$\mathbf{ct(B^0_s)}$	$\mathbf{\geq 70\ \mu m}$
$B^0_s \to J/\psi\,\phi$ Vtx prob	$\geq 0.1\%$
$m(\mu^+\mu^- K^+K^-)$	$[5.24, 5.49]$ GeV

- Vertex fit performed with J/ψ mass constraint
- Cuts to enhance purity S/(S+B)

- **$\mathcal{L}_{int} = 96.4\ fb^{-1}$ collected in 2017 and 2018**
- **Number of signal B^0_s = 48 500**
  - Number of candidates in Run-1: **49 200**

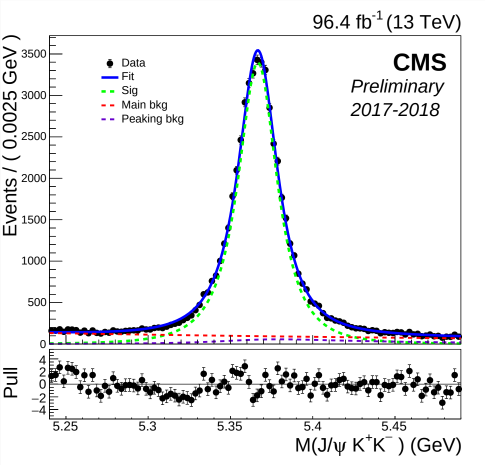
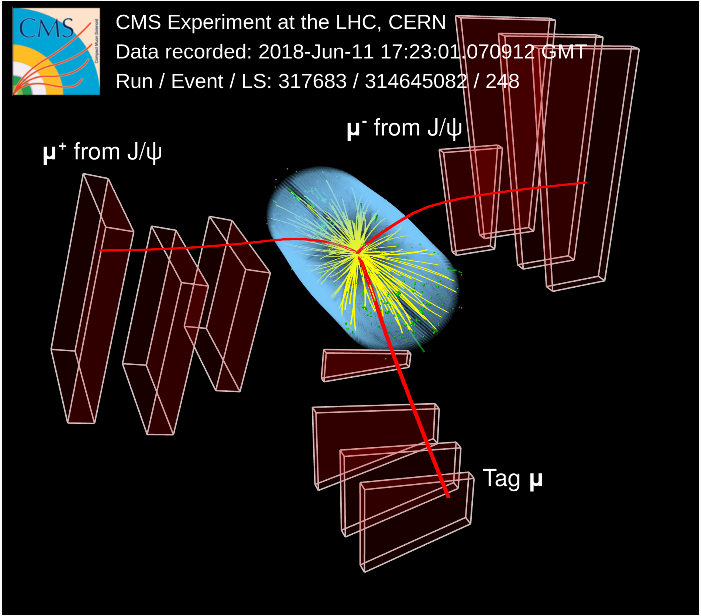


Figure: fit to data 1D projection

Example of a candidate event



Efficiencies

Proper decay length efficiency

- **The efficiency in selecting and reconstructing a B_s^0 decay** **depends on** **of the decay length**
- **To proper fit the decay rate model we need a parametrization of the decay length efficiency**
- Efficiency is evaluated with simulated samples, **separately** for 2017 and 2018, and fitted in the ct range **0.007–0.5 cm**

$$\boxed{\epsilon(ct) = e^{-a \cdot ct} \cdot \mathrm{Chebychev4}(ct)}$$

- The procedure is **validated** by fitting the $B^{\pm}$ lifetime in the $B^{\pm} \to J/\psi\, K^{\pm}$ **control channel**, in eight different data taking periods, each roughly equivalent in statistics to the B_s^0 sample

Data set	$c\tau_{B+}$ [μm]	Pull w.r.t PDG [s.d.]
2018A	489.3 ± 2.0	-0.4
2018B	495.7 ± 2.7	$+1.5$
2018C	489.2 ± 1.4	-1.4
2018D	493.2 ± 1.3	$+1.2$
2018	492.78 ± 0.97	$+1.1$
2017A	493.8 ± 2.4	$+1.0$
2017B	494.8 ± 3.5	$+1.0$
2017C	494.7 ± 2.3	$+1.4$
2017D	489.5 ± 1.7	-0.8
2017	492.9 ± 1.1	$+0.5$

Angular efficiency

- **Detector acceptance and event selection lead to non uniform angular efficiency**
- 3D angular efficiency is evaluated in bins of $\cos\theta_T$, $\cos\psi_T$ and ϕ_T, separately for 2017 and 2018, using simulated samples
  - **Binning**: 70 bins for $\cos\theta_T$ and $\cos\psi_T$, and 30 for ϕ_T
- The efficiency function is parameterized with spherical harmonics and Legendre polynomials up to order six

Flavour tagging

Tagging overview

- **Tagger:** opposite-side (OS) muon
  - Tagging feature: muon charge
  - The muon is selected already at trigger level $\rightarrow$ **very high efficiency**
- **Optimized** in $B_s^0 \rightarrow J/\psi\,\phi$ simulated events and **calibrated** in data using $B^{\pm} \rightarrow J/\psi\,K^{\pm}$ self-tagging decays
- The **figure of merit** is the tagging power $P_{tag} = \epsilon_{tag}\,\mathcal{D}^2_{tag} = \epsilon_{tag}\,(1 - 2\,\omega_{tag})^2$
  - $\epsilon_{tag} = N_{tag}/N_{tot}$, **tagging efficiency** $(N_{tag} = N_{corr.tag} + N_{mistag})$
  - $\omega_{tag} = N_{mistag}/N_{tag}$, **mistag fraction**
- **Mistag probability is evaluated on per-event basis** with a dedicated Deep Neural Network

OS-muon selection

Reconstruction	Global muon[a]
p_T	≥ 2.0 GeV
$\|\eta\|$	≤ 2.4
IP_z w.r.t. PV	≤ 1.0 cm
$\Delta R_{\eta,\phi}$ wrt B_s^0	≥ 0.4
DNN vs fakes from hadrons	Loose WP[b]

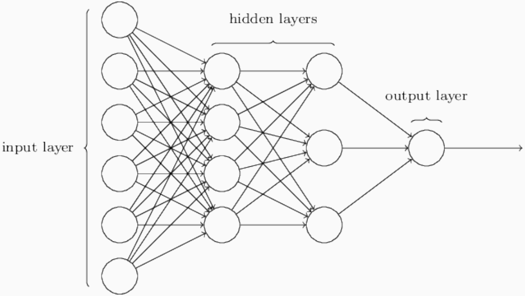
[a] Global muon = reconstructed with information from both tracker and muon system

[b] $\epsilon(\text{muons}) = 98\%,\ \epsilon(\text{hadrons}) = 33\%$ evaluated on Global muon candidates

- Reconstructed b meson tracks **excluded**
- **Dedicated discriminator for soft muons**, trained with muons from simulated samples
  - Signal: genuine muon from b hadron
  - Background: fake muons (mostly $K^{\pm},\ \pi^{\pm}$)
- The muon selection is overall **loose** for maximum efficiency
- **Performance using the muon charge as tagging feature** (without per-event mistag)
  - $\epsilon_{tag} \sim 50\%$
  - $\omega_{tag} \sim 30\%$
  - $P_{tag} \sim 7\%$

Per-event mistag probability

- **Per-event mistag probability enhances the total tagging performance**
- A fully connected **Deep Neural Network** is used to distinguish mistagged events and evaluate per-event mistag probability **at the same time**
    - **Input features:** muon variables (p_T, d_{xy}, $\sigma_{d_{xy}}$, ΔR, ...) and "cone" variables (Iso_{μ}, Q_{cone}, $p_{T,rel}$, energy ratio, ...)
- The DNN is constructed in such a way that the **output score f_{dnn} is equal to the probability of tagging the event correctly**



$$\boxed{f_{dnn} = 1 - \omega_{evt}}$$

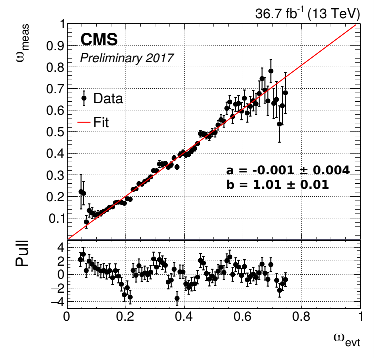
Figure: schematic representation of a fully connected DNN

Per-event mistag calibration

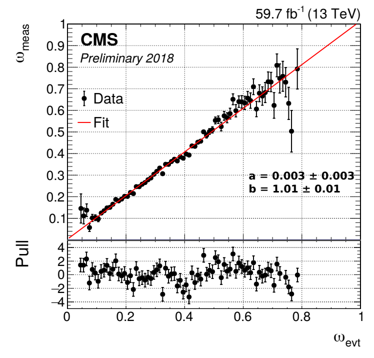
- ω_{evt} **is calibrated in data with self-tagging** $B^{\pm} \to J/\psi\, K^{\pm}$ **decays with a linear function**

$$w_{\text{fit}} = a + b \cdot w_{\text{evt}}$$

2017 tagging calibration



2018 tagging calibration



- **Excellent agreement between prediction and measurement**
- **DNN and calibration are stable $\to$ very small systematic uncertainties**

Tagging performance

- **Tagging performances evaluated in $B^{\pm} \to J/\psi\, K^{\pm}$ data**

Data set	ϵ_{tag}	ω_{tag}	P_{tag}
2017	$(45.7 \pm 0.1)\,\%$	$(27.1 \pm 0.1)\,\%$	$(\mathbf{9.6} \pm 0.1)\,\%$
2018	$(50.9 \pm 0.1)\,\%$	$(27.3 \pm 0.1)\,\%$	$(\mathbf{10.5} \pm 0.1)\,\%$
Run-1	$(8.31 \pm 0.03)\,\%$	$(30.2 \pm 0.3)\,\%$	$(1.31 \pm 0.03)\,\%$

- **High efficiency** due to the additional muon required at trigger level
- **Low dilution** thanks to the DNN based per-event mistag probability
- Final performance, normalized by the event rate, $\sim$ **50% higher** w.r.t. Run-1

Maximum likelihood fit and results

Fit model

$$\boxed{P = N_{sgn} P_{sgn} + N_{bkg} P_{bkg} + N_{peak} P_{peak}}$$

$$\boxed{P_{sgn} = \epsilon(ct)\,\epsilon(\Theta)\,[f(\Theta, ct, \alpha) \otimes G(ct, \sigma_{ct})]\, P_{sgn}(m_{B_s^0})\, P_{sgn}(\sigma_{ct})\, P_{sgn}(\xi)}$$

- $\epsilon(ct)\,\epsilon(\Theta)$: efficiency functions
- $f(\Theta, ct, \alpha)$: differential decay rate PDF
- $G(ct, \sigma_{ct})$: Gaussian resolution function
- $P(m_{B_s^0})$: mass PDFs
- $P(\sigma_{ct})$: decay length uncertainty PDFs
- $P(\xi)$: tag distribution

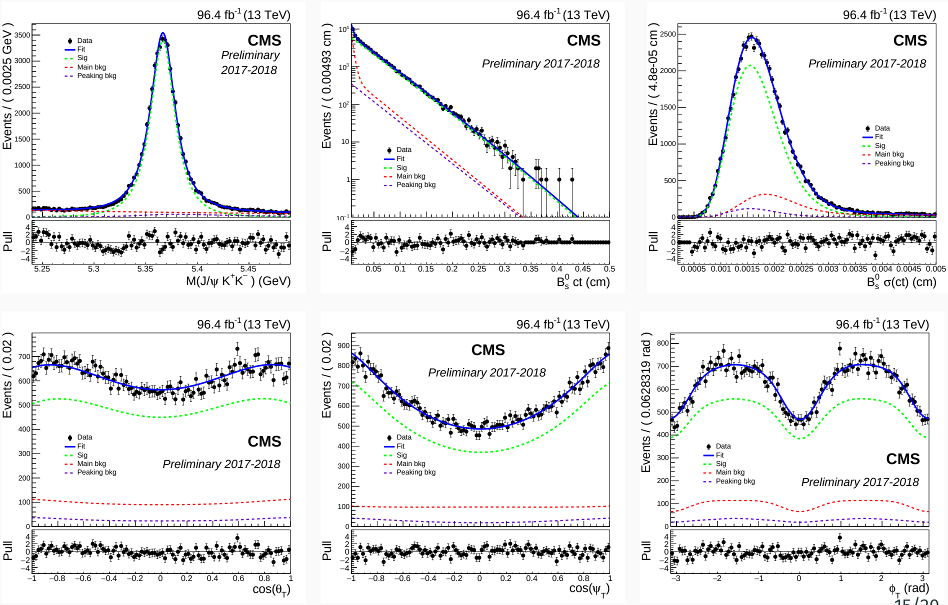
$$\boxed{P_{bkg} = P_{bkg}(\cos\theta_T, \phi_T)\, P_{bkg}(\cos\psi_T)\, P_{bkg}(ct)\, P_{bkg}(m_{B_s^0})\, P_{bkg}(\sigma_{ct})\, P_{bkg}(\xi)}$$

- $P_{bkg}(\cos\theta_T, \phi_T)$, $P_{bkg}(\cos\psi_T)$, $P_{bkg}(ct)$: background angular and lifetime PDFs

$$\boxed{P_{peak} = P_{peak}(\cos\theta_T, \phi_T)\, P_{peak}(\cos\psi_T)\, P_{peak}(ct)\, P_{peak}(m_{B_s^0})\, P_{peak}(\sigma_{ct})\, P_{peak}(\xi)}$$

- P_{peak} models the **peaking background** from $B^0 \to J/\psi\, K^{*0} \to \mu^+\mu^-\, K^+\pi^-$ where the pion is misidentified as a kaon
- Peaking background from $\Lambda_b \to J/\psi\, Kp$ estimated to be negligible

Fit 1D projections



Systematic uncertainties

	$\|A_0\|^2$	$\|A_\perp\|^2$	$\|A_S\|^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]	Γ_s [ps^{-1}]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ ps^{-1}]	$\|\lambda\|$	ϕ_s [mrad]
Model bias	0.0002	0.0012	0.0008	0.020	0.016	0.006	0.0005	0.0019	—	0.0035	**7.9**
Angular efficiency	0.0008	0.0010	0.0015	0.006	0.015	0.015	0.0002	0.0006	0.007	**0.0057**	**3.8**
Lifetime efficiency	0.0014	0.0023	0.0007	0.001	0.002	0.002	**0.0022**	**0.0062**	0.001	0.0002	0.3
Lifetime resolution	0.0007	0.0009	0.0065	0.006	0.025	0.022	0.0005	0.0008	**0.015**	0.0009	2.5
Data-MC mismatch	0.0044	0.0029	0.0065	0.007	0.007	0.028	0.0003	0.0008	0.004	0.0003	0.6
Flavour tagging	0.0003	$< 10^{-4}$	$< 10^{-4}$	0.001	0.003	0.001	$< 10^{-4}$	$< 10^{-4}$	0.001	0.0002	0.1
Unfitted ω_{evt} dist.	—	0.0008	—	—	—	0.006	0.0005	—	—	—	3.0
Model assumptions	—	0.0013	0.0012	0.017	0.019	0.011	0.0003	—	—	**0.0046**	—
Peaking background	0.0005	0.0002	0.0025	0.005	0.007	0.011	0.0002	0.0008	**0.011**	$< 10^{-4}$	0.3
Total syst.	0.0048	0.0044	0.0097	0.028	0.040	0.043	0.0024	0.0066	0.020	0.0082	9.6

Leading systematic uncertainties for the most interesting parameters

- $\phi_s \to$ model bias and angular efficiency
- $\Delta\Gamma_s \to$ lifetime efficiency
- $\Gamma_s \to$ lifetime efficiency
- $\Delta m_s \to$ lifetime resolution and peaking background model
- $|\lambda| \to$ angular efficiency and model assumptions

Results

Parameter	Value	Stat.	Syst.
ϕ_s [mrad]	**-11**	**± 50**	**± 10**
$\Delta\Gamma_s$ [ps^{-1}]	**0.114**	**± 0.014**	**± 0.007**
Γ_s [ps^{-1}]	**0.6531**	**± 0.0042**	**± 0.0024**
Δm_s [$\hbar$ ps^{-1}]	**17.51**	**$^{+0.10}_{-0.09}$**	**± 0.02**
$\lvert\lambda\rvert$	**0.972**	**± 0.026**	**± 0.008**
$\lvert A_0\rvert^2$	0.5350	± 0.0047	± 0.0048
$\lvert A_\perp\rvert^2$	0.2337	± 0.0063	± 0.0044
$\lvert A_S\rvert^2$	0.022	$^{+0.008}_{-0.007}$	± 0.010
$\delta_\parallel$ [rad]	3.18	± 0.12	± 0.03
$\delta_\perp$ [rad]	2.77	± 0.16	± 0.04
$\delta_{S\perp}$ [rad]	0.221	$^{+0.083}_{-0.070}$	± 0.043

- **ϕ_s and $\Delta\Gamma_s$ are in agreement with the SM:**

$$\phi_s^{SM} = -36.96^{+0.84}_{-0.72}\ \text{mrad}$$
$$\Delta\Gamma_s^{SM} = 0.087 \pm 0.021\ \text{ps}^{-1}$$

- **Γ_s is consistent with the world average:**

$$\Gamma_s^{WA} = 0.6623 \pm 0.0018\ \text{ps}^{-1}$$

- **Δm_s is consistent with the world average:**

$$\Delta m_s^{WA} = 17.757 \pm 0.021\ \hbar\text{ps}^{-1}$$

- **$\lvert\lambda\rvert$ is consistent with no direct CPV ($\lambda = 1$)**
- **This is the first measurement by CMS of Δm_s and $\lvert\lambda\rvert$**

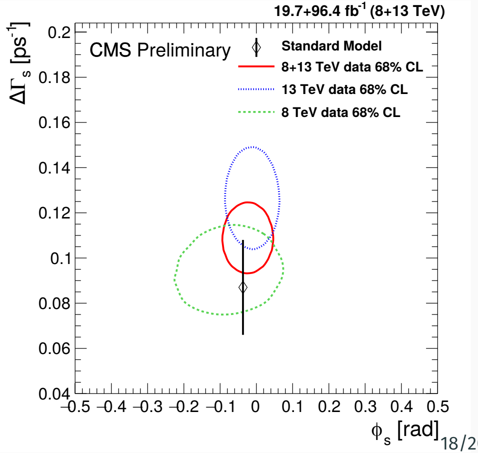
Combination with 8 TeV results

- **The results of this analysis are in agreement with the ones obtained by CMS at $\sqrt{s} = 8$ TeV [Phys.Lett.B757(2016)97] and therefore combined**
- **All systematic uncertainties are considered uncorrelated**
- **The results are in agreement with the SM predictions**

$$\phi_s = -21 \pm 45\,\text{mrad}$$

$$\Delta\Gamma_s = 0.1074 \pm 0.0097\,\text{ps}^{-1}$$

- The new trigger strategy, which trades number of events for tagging power, **pays off** for ϕ_s while **does not improve** $\Delta\Gamma_s$, which sensitivity is driven by statistics



Conclusions

Summary

- The **CPV phase ϕ_s** and the **decay width difference $\Delta\Gamma_s$** are measured using 48 500 $B_s^0 \to J/\psi\,\phi$ candidates collected at $\sqrt{s} = 13$ TeV, corresponding to $\mathcal{L}_{int} = 96.4\ \text{fb}^{-1}$
- Events are selected using a **non displaced trigger** that required an **additional muon**, which is exploited to infer the flavor of the B_s^0
  - This strategy paid off in terms of **tagging performance**, leading to a significant reduction of the ϕ_s uncertainty
  - However, the **limited number of selected events** prevented improvements on $\Delta\Gamma_s$
- A **novel** opposite-side muon tagger based on **Deep Neural Network** has been developed to **directly predict** mistag probability on per-event basis, achieving $P_{tag} \sim 10\,\%$
- Results from this analysis are **combined** with those obtained at $\sqrt{s} = 8$ TeV yielding

$$\boldsymbol{\phi_s = -21 \pm 45\ \text{mrad}}$$
$$\boldsymbol{\Delta\Gamma_s = 0.1074 \pm 0.0097\ \text{ps}^{-1}}$$

- Results are consistent with the **Standard Model** predictions

$$\phi_s^{SM} = -36.96^{+0.84}_{-0.72}\ \text{mrad} \qquad \Delta\Gamma_s^{SM} = 0.087 \pm 0.021\ \text{ps}^{-1}$$

Outlook

Comparison with other LHC experiments in the $B_s^0 \to J/\psi\, K^+K^-$ channel

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Reference
CMS	-21 ± 45	0.1074 ± 0.0097	CMS-PAS-BPH-20-001
ATLAS	-87 ± 42	0.0640 ± 0.0048	CERN-EP-2019-218
LHCb	-81 ± 32	0.0777 ± 0.0062	EUR.PHYS.J.C79(2019)706
SM	$-36.96^{+0.84}_{-0.72}$	0.087 ± 0.021	CKMfitter, 1102.4274

· All of the above are combination of Run-1 and partial Run-2 results

· Uncertanties are presented as the stat.+syst. squared sum

· LHCb results refer to the combination of measurements around the $\phi(1020)$ resonance

· New $\Delta\Gamma_s$ prediction with smaller uncertainties available: $\Delta\Gamma_s^{SM} = 0.091 \pm 0.013$ ps^{-1} [1912.07621]

- $\Delta\Gamma_s$ shows **tensions** between experiments
- Full Run-2 measurements will clarify the situation

Future plans

- CMS plans to analyze the **full Run-2 dataset**, adding a **complementary trigger** that requires a displaced J/ψ plus two charged tracks
  - Electron and jet flavour tagging algorithms will be used
- Effective statistics $N(B_s^0) \cdot P_{tag}$ expected to improve by a factor $1.5 \sim 2.0$

Thanks for your attention!

Decay rate model

$$\boxed{\frac{d^4\Gamma(B_s^0(t))}{d\Theta dt} = \sum_{i=1}^{10} \mathcal{O}_i(\alpha, t) \cdot g_i(\Theta)}$$

$$\mathcal{O}_i = N_i e^{-\Gamma_s t}\left[a_i \cosh\left(\frac{1}{2}\Delta\Gamma_s t\right) + b_i \sinh\left(\frac{1}{2}\Delta\Gamma_s t\right) + c_i \xi(1-2\omega)\cos\left(\Delta m_s t\right) + d_i \xi(1-2\omega)\sin\left(\Delta m_s t\right)\right]$$

i	$g_i(\theta_T, \psi_T, \varphi_T)$	N_i	a_i	b_i	c_i	d_i
1	$2\cos^2\psi_T(1-\sin^2\theta_T\cos^2\varphi_T)$	$\|A_0\|^2$	1	D	C	$-$S
2	$\sin^2\psi_T(1-\sin^2\theta_T\sin^2\varphi_T)$	$\|A_\parallel\|^2$	1	D	C	$-$S
3	$\sin^2\psi_T\sin^2\theta_T$	$\|A_\perp\|^2$	1	$-$D	C	S
4	$-\sin^2\psi_T\sin 2\theta_T\sin\varphi_T$	$\|A_\parallel\|\|A_\perp\|$	$C\sin(\delta_\perp-\delta_\parallel)$	$S\cos(\delta_\perp-\delta_\parallel)$	$\sin(\delta_\perp-\delta_\parallel)$	$D\cos(\delta_\perp-\delta_\parallel)$
5	$\frac{1}{\sqrt{2}}\sin 2\psi_T\sin^2\theta_T\sin 2\varphi_T$	$\|A_0\|\|A_\parallel\|$	$\cos(\delta_\parallel-\delta_0)$	$D\cos(\delta_\parallel-\delta_0)$	$C\cos(\delta_\parallel-\delta_0)$	$-S\cos(\delta_\parallel-\delta_0)$
6	$\frac{1}{\sqrt{2}}\sin 2\psi_T\sin 2\theta_T\cos\varphi_T$	$\|A_0\|\|A_\perp\|$	$C\sin(\delta_\perp-\delta_0)$	$S\cos(\delta_\perp-\delta_0)$	$\sin(\delta_\perp-\delta_0)$	$D\cos(\delta_\perp-\delta_0)$
7	$\frac{2}{3}(1-\sin^2\theta_T\cos^2\varphi_T)$	$\|A_S\|^2$	1	$-$D	C	S
8	$\frac{1}{3}\sqrt{6}\sin\psi_T\sin^2\theta_T\sin 2\varphi_T$	$\|A_S\|\|A_\parallel\|$	$C\cos(\delta_\parallel-\delta_S)$	$S\sin(\delta_\parallel-\delta_S)$	$\cos(\delta_\parallel-\delta_S)$	$D\sin(\delta_\parallel-\delta_S)$
9	$\frac{1}{3}\sqrt{6}\sin\psi_T\sin 2\theta_T\cos\varphi_T$	$\|A_S\|\|A_\perp\|$	$\sin(\delta_\perp-\delta_S)$	$-D\sin(\delta_\perp-\delta_S)$	$C\sin(\delta_\perp-\delta_S)$	$S\sin(\delta_\perp-\delta_S)$
10	$\frac{4}{3}\sqrt{3}\cos\psi_T(1-\sin^2\theta_T\cos^2\varphi_T)$	$\|A_S\|\|A_0\|$	$C\cos(\delta_0-\delta_S)$	$S\sin(\delta_0-\delta_S)$	$\cos(\delta_0-\delta_S)$	$D\sin(\delta_0-\delta_S)$

$$C = \frac{1-|\lambda|^2}{1+|\lambda|^2} \rightarrow \textbf{Sensitive to direct CPV}$$

$$S = -\frac{2|\lambda|\sin\phi_s}{1+|\lambda|^2} \rightarrow \textbf{Sensitive to small } \phi_s$$

$$D = -\frac{2|\lambda|\cos\phi_s}{1+|\lambda|^2}$$

Angular efficiency

Computed separately for 2017 and 2018 using the **“projection” method**

1. **Construct efficiency histograms**
   - Numerator: 3D angular RECO histograms from $\Delta\Gamma_s = 0$ MC samples
   - Denominator: 3D angular GEN histograms from GEN only sample
   - Binning: 70 bins for $\cos\theta_T$ and $\cos\psi_T$, and 30 for ϕ_T
2. **Project on Legendre orthogonal basis**

$$b_{l,k,m}(\Theta) = P_l^m(\cos\theta_T) \cdot P_k^m(\cos\psi_T) \cdot \begin{cases} \sin(m\,\phi_T) & \text{if } m < 0 \\ \cos(m\,\phi_T) & \text{if } m > 0 \\ 1/2 & \text{if } m = 0 \end{cases}$$

   - up to order 6
3. **Construct angular efficiency as**

$$\boxed{\epsilon(\Theta) = \sum_{l,k,m} c_{l,k,m} \cdot b_{l,k,m}(\Theta)}$$

   - $c_{l,k,m}$ are the projection coefficients

Deep neural network for flavour tagging

- **Training features**
  - Muon variables: p_T, η, d_{xy}, $\sigma(d_{xy})$, d_z, $\sigma(d_z)$, $\Delta R(\mu, B_s^0)$, DNN vs hadron fakes score
  - Cone variables: Iso_μ, Q_{cone}, $p_{T,rel}$, $p_{T,cone}$, $\Delta R(\mu, cone)$, E_μ/E_{cone}
- **Architecture: fully connected**
  - 3 layers of 200 neurons
  - ReLU activation
  - 40% dropout probability
- **Loss**: categorical crossentropy
- **Optimizer**: Adam