CPV with $B_s^0 \to J/\psi\phi(1020)$ decay

BEAUTY2020 - Kavli, IPMU - Tokyo, Japan

Giacomo Fedi (IC) on behalf of the CMS collaboration

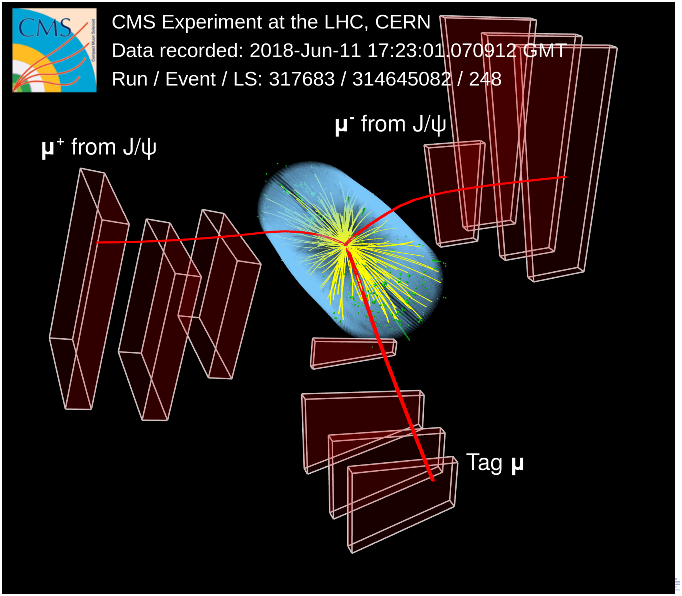
giacomo.fedi@cern.ch

Imperial College London

September 23, 2020

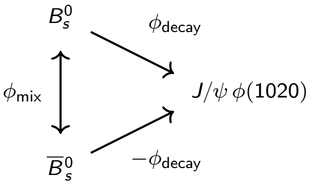
Introduction

Event display - $B_s^0 \to J/\psi\phi(1020)$ candidate



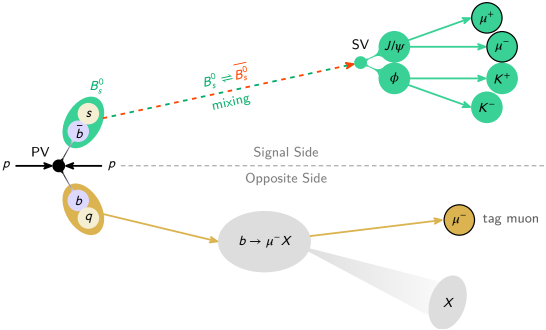
Motivational aspects

- ϕ_s is a CPV phase which stems from the interference between B_s^0 decays proceeding directly and through B_s^0-$\overline{B}_s^0$ mixing to a CP final state
- SM prediction pretty small and precisely predicted: $\phi_s \simeq -2\beta_s = arg(-V_{ts}V_{tb}^*/V_{cs}V_{cb}^*) = -36.96^{+0.72}_{-0.84}$ mrad [CKMfitter]
- New Physics can change the value of ϕ_s up to $\sim 10\%$ via new particles contributing to the B_s^0-$\overline{B}_s^0$ mixing [JHEP04(2010)031]
- Measured reconstructing the $\mathrm{B_s^0} \rightarrow \mathrm{J}/\psi\phi(1020)$ decay channel
- Other measured observables: Γ_s, $\Delta\Gamma_s$, $|\lambda|$, Δm_s, CP amplitudes and strong phases
  - $\lambda = \frac{q}{p}\frac{\overline{A}_{f.s.}}{A_{f.s.}}$, $|B_{L,H}\rangle = p|B_s^0\rangle \pm q|\overline{B}_s^0\rangle$



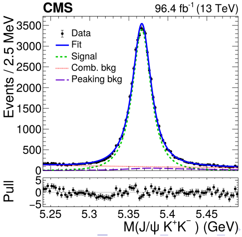
$$\phi_s = \phi_{\mathrm{mix}} - 2\phi_{\mathrm{decay}}$$

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



Offline selection

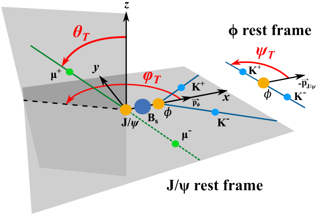
$p_T(\mu)$	$\geqslant$ 3.5 GeV
$\lvert\eta(\mu)\rvert$	$\leqslant$ 2.4
$p_T(K)$	$\geqslant$ 1.2 GeV
$\lvert\eta(K)\rvert$	$\leqslant$ 2.5
$\lvert m(\mu^+\mu^-) - m^{PDG}_{J/\psi}\rvert$	$<$ 150 MeV
$\lvert m(K^+K^-) - m^{PDG}_{\phi(1020)}\rvert$	$<$ 10 MeV
$p_T(B_s^0)$	$\geqslant$ 11 GeV
$ct(B_s^0)$	$\geqslant$ 70 μm
$B_s^0 \to J/\psi\ \phi$ Vtx prob	$\geqslant$ 0.1%
$m(\mu^+\mu^-K^+K^-)$	[5.24, 5.49] GeV

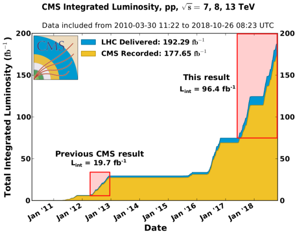


- Tagging power $P_{\text{tag}} = \epsilon_{\text{tag}}\,\mathcal{D}^2_{\text{tag}}$
- $\mathcal{L}_{int} = 96.4\text{fb}^{-1}$ collected in 2017 and 2018
- Number of signal $B_s^0 = 48\,500$

Analysis strategy

- The B_s meson decays into a final state $J/\psi\phi(1020)$, which is a mixture of two CP eigenstates (odd/even)
- Need to disentangle the two states using an angular analysis
- Time dependent analysis, the proper decay time of $\mathrm{B_s^0}$ is reconstructed
- Trigger: $J/\psi \rightarrow \mu^+\mu^-$ candidate plus an additional muon
  - no lifetime cuts and it enhances the tagging power
- Datasets: 2017 and 2018 CMS data taking
  - datasets with the highest statistics of LHC Run2 and with a new inner tracker installed (increased proper time resolution)





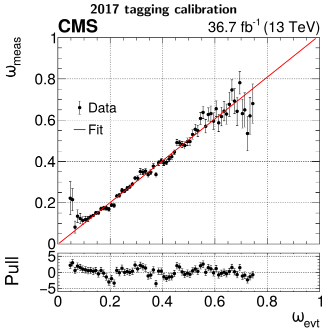
Flavour tagging

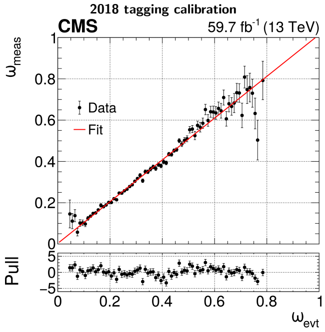
Opposite side flavour tagging

- The charge of opposite side muons were used to tag the flavour of the signal B_s at the production time
- Tagging algorithm tuned using simulations of the signal and calibrated in data using $B^+ \to J/\psi K^+$ self-tagging decays
- High tagging efficiency since the events are selected by the trigger with an extra muon
- The figure of merit is the tagging power $P_{\text{tag}} = \epsilon_{\text{tag}}\,\mathcal{D}^2_{\text{tag}} = \epsilon_{\text{tag}}\,(1 - 2\,\omega_{\text{tag}})^2$
  - tagging efficiency $\epsilon_{\text{tag}} = N_{\text{tag}}/N_{\text{tot}}$
  - mistag fraction $\omega_{\text{tag}} = N_{\text{mistag}}/N_{\text{tag}}$
- the mistag fraction ω_{tag} is used in the fit and it is measured per event using a Deep Neural Network (DNN)

Tagging algorithm and calibration

- In each event the tagging muon is chosen among those not belonging to the J/ψ, in case of multiple tagging muons (3%) the highest p_T is selected
- A fully connected DNN is used to evaluate per-event mistag probability
  - Multiple input variables, mainly related to the kinematics of the muon and its geometrical position w.r.t. signal
- ω_evt is calibrated in data with self-tagging $\mathrm{B}^+ \rightarrow \mathrm{J}/\psi\mathrm{K}^+$ decays with a linear function





Tagging performance

- Tagging performance evaluated using $\mathrm{B}^+ \to \mathrm{J}/\psi \mathrm{K}^+$ channel in data
- The use of the DNN and the dedicated trigger improved the power of tagging w.r.t. Run1
- Systematics on flavour tagging found to be quite small

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	$(50.9 \pm 0.1)\,\%$	$(27.3 \pm 0.1)\,\%$	$(\mathbf{10.5} \pm 0.1)\,\%$

Data fit

Signal model

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

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

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

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

$$S = -\frac{2|\lambda|\sin\phi_s}{1+|\lambda|^2}$$ → **Sensitive to small ϕ_s**

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

Maximum Likelihood fit

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

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

- $\epsilon(ct)\,\epsilon(\Theta)$: 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

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

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

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

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

Detector and acceptance unfolding

- Angular and proper decay time efficiencies are obtained using high statistics samples of simulated signal events for each year of data-taking
- Simulated kinematic distributions are in good agreement with the data, in case of mismatch, the systematics effects have been estimated
- Angular efficiency:
  - evaluated in bins of $\cos\theta_T$, $\cos\psi_T$ and ϕ_T and accounting for correlations
  - efficiency function is parameterised with spherical harmonics and Legendre polynomials up to order six
- Proper decay time efficiency:
  - efficiency function is parametrised with an exponential multiplying a 4th-degree Chebychev function
  - the method has been validated on data in each data taking periods using the $B^+ \to J/\psi K^+$ channel

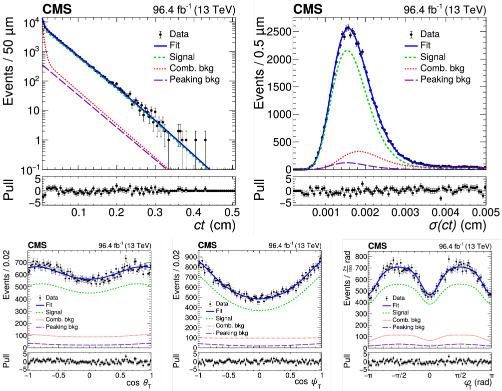
Systematic uncertainties

	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_{\parallel}$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]	Γ_s [ps^{-1}]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ ps^{-1}]	$\lvert\lambda\rvert$	ϕ_s [mrad]
Model bias	0.0002	0.0012	0.001	0.020	0.016	0.006	0.0005	**0.0019**	—	0.0035	**7.9**
Angular efficiency	0.0008	0.0010	0.002	0.006	0.015	0.015	0.0002	**0.0006**	0.007	0.0057	**3.8**
Lifetime efficiency	0.0014	0.0023	0.001	0.001	0.002	0.002	0.0022	**0.0062**	0.001	0.0002	**0.3**
Lifetime resolution	0.0007	0.0009	0.007	0.006	0.025	0.022	0.0005	**0.0008**	0.015	0.0009	**2.5**
Data-MC mismatch	0.0044	0.0029	0.007	0.007	0.007	0.028	0.0003	**0.0008**	0.004	0.0003	**0.6**
Flavour tagging	0.0003	$< 10^{-4}$	$< 10^{-3}$	0.001	0.003	0.001	$< 10^{-4}$	$< \mathbf{10^{-4}}$	0.001	0.0002	**0.1**
Unfitted ω_{evt} dist.	—	0.0008	—	—	—	0.006	0.0005	—	—	—	**3.0**
Model assumptions	—	0.0013	0.001	0.017	0.019	0.011	0.0003	**0.0008**	—	0.0046	—
Peaking background	0.0005	0.0002	0.003	0.005	0.007	0.011	0.0002	**0.0008**	0.011	$< 10^{-4}$	**0.3**
S-P wave interference	0.0005	—	0.013	—	0.019	0.019	0.0005	**0.0010**	0.019	—	—
Total syst.	0.0048	0.0044	0.016	0.028	0.045	0.047	0.0024	**0.0067**	0.028	0.0082	**9.6**

- Leading systematic effects for ϕ_s are the model bias and the angular efficiencies
- Leading systematic for $\Delta\Gamma_s$ is the lifetime efficiency

Results

Parameter	Value	Stat.	Syst.
ϕ_s [mrad]	**-11**	**$\pm$ 50**	**$\pm$ 10**
$\Delta\Gamma_s$ [ps^{-1}]	**0.114**	**$\pm$0.014**	**$\pm$ 0.007**
Γ_s [ps^{-1}]	0.6531	$\pm$ 0.0042	$\pm$ 0.0024
Δm_s [$\hbar$ ps^{-1}]	17.51	$+$ 0.10 $-$ 0.09	$\pm$ 0.02
$\lvert\lambda\rvert$	0.972	$\pm$ 0.026	$\pm$ 0.008
$\lvert A_0\rvert^2$	0.5350	$\pm$ 0.0047	$\pm$ 0.0048
$\lvert A_\perp\rvert^2$	0.2337	$\pm$ 0.0063	$\pm$ 0.0044
$\lvert A_S\rvert^2$	0.022	$+$ 0.008 $-$ 0.007	$\pm$ 0.016
$\delta_\parallel$ [rad]	3.18	$\pm$ 0.12	$\pm$ 0.03
$\delta_\perp$ [rad]	2.77	$\pm$ 0.16	$\pm$ 0.04
$\delta_{S\perp}$ [rad]	0.221	$+$ 0.083 $-$ 0.070	$\pm$ 0.048



- Simultaneous fit on 2017 and 2018 datasets
- ϕ_s and $\Delta\Gamma_s$ are in agreement with the SM expectations
- $\lvert\lambda\rvert$ compatible with no direct CPV

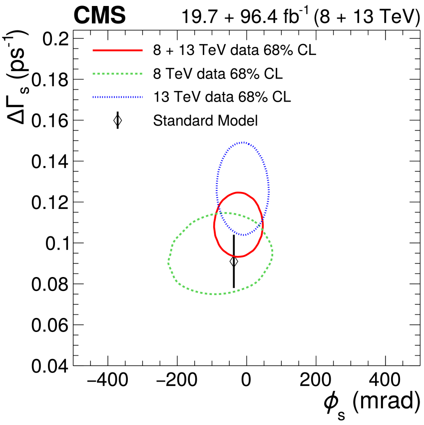
Combination with the 8 TeV results

- Presented results and Run1 8 TeV results have been combined, taking into account the statistical correlation between the fitted observable
- Result of the combination:

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

$$\Delta\Gamma_s = 0.1073 \pm 0.0097\,\text{ps}^{-1}$$

- Good agreement with the SM expectations



Summary & Outlook

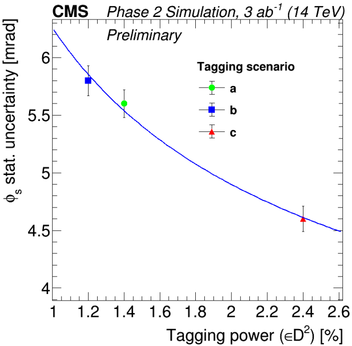
Summary

- The CPV weak phase ϕ_s and the difference of decay rates $\Delta\Gamma_s$ have been measured in 2017 and 2018 CMS data for a total of $\mathcal{L}_{int} = 96.4\ \mathrm{fb}^{-1}$
- For the measurement 48,500 candidates of B_s^0 were reconstructed using the $B_s^0 \rightarrow J/\psi\phi(1020)$ channel
- Old and present measurement of CMS have been combined obtaining: $\phi_s = -21 \pm 45$ mrad and $\Delta\Gamma_s = 0.1073 \pm 0.0097\ \mathrm{ps}^{-1}$
- The measurement is in agreement with the SM expectation; $\phi_s^{SM} = -36.96^{+0.72}_{-0.84}$ mrad and $\Delta\Gamma_s^{SM} = 0.087 \pm 0.021\ \mathrm{ps}^{-1}$

Outlook

- The statistical uncertainty still dominates over the systematic one
- The measured uncertainty on ϕ_s is one order of magnitude bigger than possible NP effect
- Some tensions are present on some observables in the LHC experiments (ATLAS and LHCb)
- Additional statistics is needed, we are looking forward to Run3 and Phase2 LHC collisions

[CMS-PAS-FTR-18-041]



Additional slides

Deep Neural Network

- 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}, $p_{T,rel}$, $p_{T,\text{cone}}$, $\Delta R(\mu, \text{cone})$, E_μ / E_{cone}
- DNN specs
  - Architecture: fully connected
  - 3 layers of 200 neurons
  - ReLU activation
  - 40% dropout probability
  - Loss: categorical crossentropy
  - Optimizer: Adam

Muon loose selections

Variable	Selection
p_T	$\geqslant$ 2.0 GeV
$\lvert\eta\rvert$	$\leqslant$ 2.4
IP_z w.r.t. PV	$\leqslant$ 1.0 cm
$\Delta R_{\eta,\phi}$ wrt	$\geqslant$ 0.4
DNN vs fakes from hadrons	Loose WP