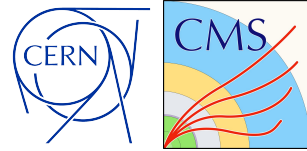




Imperial College
London





Advancing CP Violation measurements: new approaches from the CMS experiment

Alberto Bragagnolo (CERN)

Imperial College, London
13 November 2024


Introduction

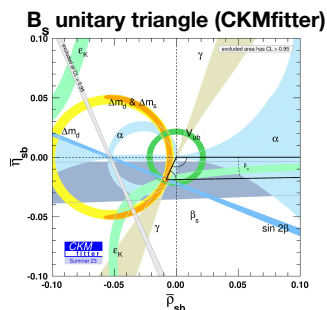
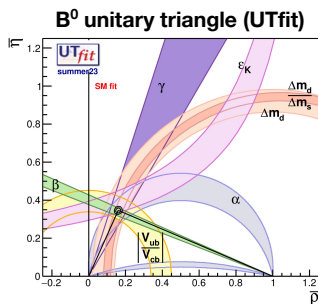
Why we need CP violation?

- **Baryon asymmetry remains one of the great mysteries of modern physics**
- Half a century ago, Andrei **Sakharov** proposed three necessary **conditions** for a baryon-generating process:
 1. Baryon number violation
 2. C and **CP violation (CPV)**
 3. Non thermal equilibrium
- In the Standard Model (SM) CP is **conserved** by the strong and electromagnetic interactions, but it is **violated** by the weak force
 - CPV was first observed in 1964 by Fitch and Cronin using neutral kaons 
 - P violation was proposed by Lee and Yang (1956)  and experimentally observed by Wu (1957)
- CPV is **allowed** in the SM, but the amount is **insufficient** to account for the observed baryon asymmetry of the universe
 - Sources of CPV beyond the SM have to exist
 - CPV observables are often precisely predicted and very sensitive to new physics



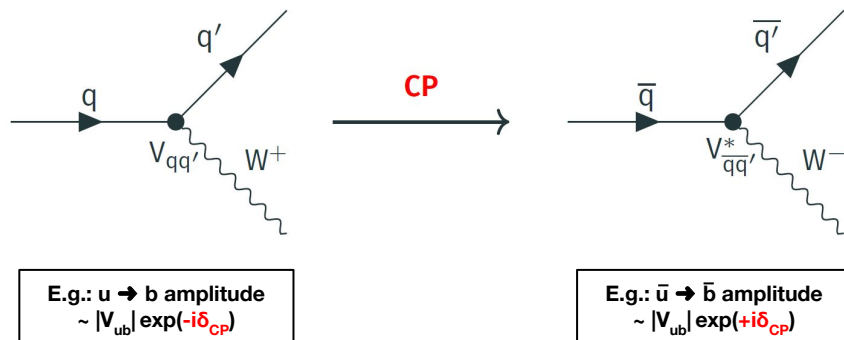
CP violation in the SM

- In the SM quark transitions are possible through flavor-changing weak interactions
- Information about the strength of the transition is contained in the Cabibbo-Kobayashi-Maskawa (CKM) matrix** 
 - Parameters: 3 angles + 1 complex phase
- The single complex phase allows for CP violation**
- In the SM, the CKM matrix is **unitary**
 - Unitary conditions can be represented by “unitary triangles”



$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{ud} & V_{us} & V_{ub} \\ V_{cd} & \mathbf{V}_{cs} & V_{cb} \\ V_{td} & V_{ts} & \mathbf{V}_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

Flavour CKM Mass



Interlude: flavor mixing

- Neutral K, D, and B mesons are subject to **flavor mixing**, that is oscillations between their C-conjugate states before decay
- They propagate as light and heavy mass eigenstates which are described by a superposition of flavor states:

$$|M_{L,H}^0\rangle = p|M^0\rangle \pm q|\bar{M}^0\rangle \quad (|q|^2 + |p|^2 = 1)$$

- The system is characterized by the parameters

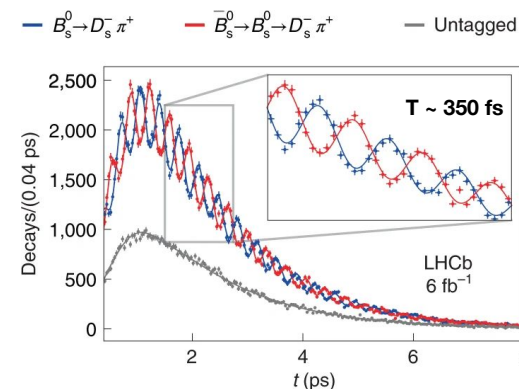
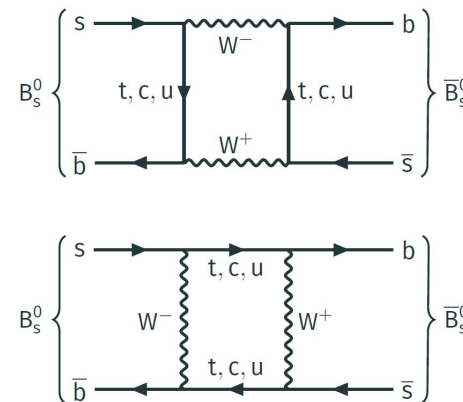
$$m \equiv \frac{m_H + m_L}{2} \quad \Gamma \equiv \frac{\Gamma_H + \Gamma_L}{2}$$

$$\Delta m \equiv m_H - m_L \quad \Delta\Gamma \equiv \Gamma_L - \Gamma_H$$

Definition may vary

- The flavor eigenstates oscillate with period $T = 2\pi/\Delta m$
- CPV in mixing**, i.e. $Pr(M^0 \rightarrow \bar{M}^0) \neq Pr(\bar{M}^0 \rightarrow M^0)$, implies $|q/p| \neq 1$

Example: B_s meson mixing



Ref: LHCb Collab. [Nat.Phys.18\(2022\)1-5](https://arxiv.org/abs/2201.09919)

Types of CP violation

- Observable CP violation in weak interaction can be classified into three different types

1. Direct CPV in **decays**

- Observed in K, B, and D mesons

$$Pr(M \rightarrow f) \neq Pr(\bar{M} \rightarrow \bar{f})$$

2. Indirect CPV in **mixing**

- Observed in K^0 oscillations

$$Pr(M^0 \rightarrow \bar{M}^0) \neq Pr(\bar{M}^0 \rightarrow M^0)$$

3. CPV in decay/mixing **interference**

- Observed in K^0 , B^0 and B_s mesons

$$Pr(M^0_{(\rightsquigarrow \bar{M}^0)} \rightarrow f_{CP}) \neq Pr(\bar{M}^0_{(\rightsquigarrow M^0)} \rightarrow f_{CP})$$

- The CPV information can be coded in the **rephasing invariant complex parameter λ** (A_f being the $M \rightarrow f$ amplitude)

$$\lambda \equiv \frac{q}{p} \frac{\bar{A}_{\bar{f}}}{A_f} \begin{cases} |\bar{A}_{\bar{f}}/A_f| \neq 1 & \rightarrow \text{direct CPV} \\ |q/p| \neq 1 & \rightarrow \text{indirect CPV} \\ |\lambda| = 1, \text{Im}(\lambda) \neq 0 & \rightarrow \text{interference CPV} \end{cases}$$



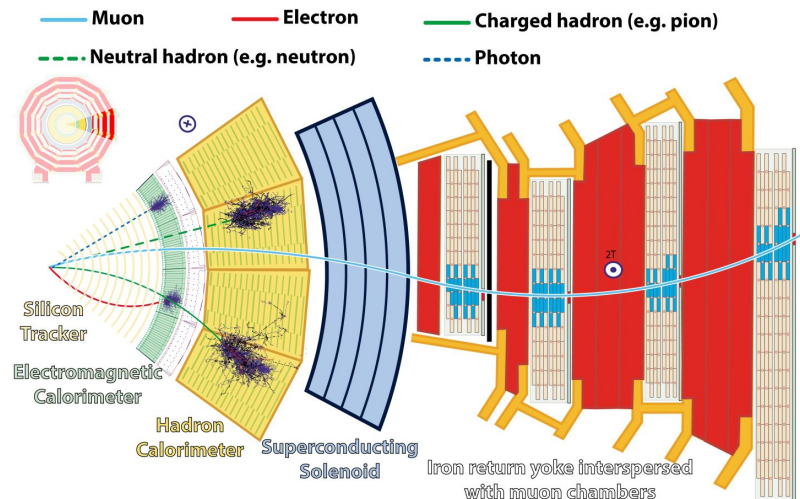
The CMS detector

CMS is a general purpose detector able to perform a vast range of physics studies, including flavor physics

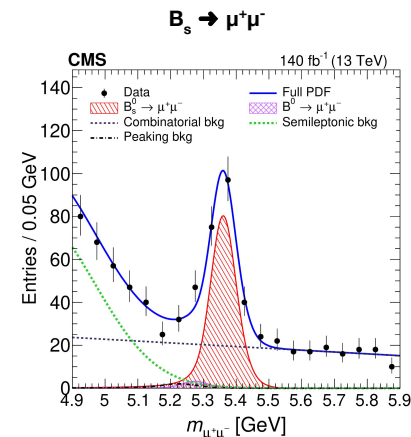
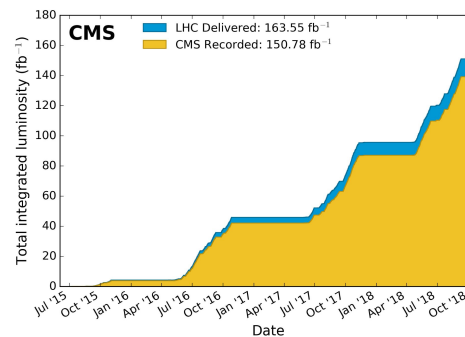
- + **Excellent tracking system** able to reconstruct vertices with high decay time resolution (e.g., $\sigma_t \sim 65$ fs for $B_s \rightarrow J/\psi \phi$) up to $|\eta| < 2.5$
 - Complementary to LHCb ($2 < |\eta| < 5$)
- + **Enormous amount of data collected**
 - $\sim 7.5 \cdot 10^{13}$ bb pairs produced at Point 5 during Run 2 (not considering geometric acceptance)
- High pile up $N_{PV} \sim 40$ (in Run 2)
- No reliable hadronic particle identification available

Some **CMS flavor physics highlights** from recent years

- $B_s \rightarrow \mu^+ \mu^-$ (world's most precise) [[PLB842\(2023\)137955](#)]
- $\eta \rightarrow \mu^+ \mu^- \mu^+ \mu^-$ observation [[PRL131\(2023\)091903](#)]
- f_s/f_u measurements [[PRL131\(2023\)121901](#)]
- Triple J/ψ production observation [[Nat.Phys.19\(2023\)338](#)]
- $R(K)$ and $R(J/\psi)$ LFU tests [[ROPP87\(2024\)077802](#)][[BPH-22-012](#)]
- This work



CMS luminosity in Run-2

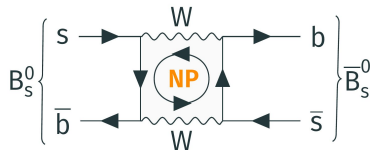


Evidence of CPV in $B_s \rightarrow J/\psi \phi(1020)$

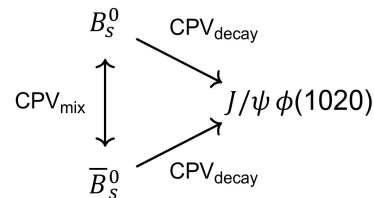
[CMS PAS BPH-23-004](#)

Motivations

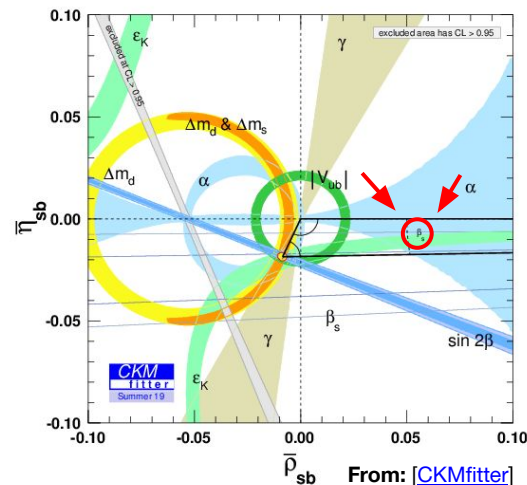
- B_s mesons decays allow us to study the time-dependent CP violation generated by the **interference** between direct decays and flavor mixing
 - CPV is possible even if there is no CPV in decay and mixing
- The weak phase ϕ_s is the main CPV observable
 - Predicted by the SM to be $\phi_s \approx -2\beta_s$ ($\beta_s \rightarrow$ angle of the B_s unit. triangle)
 - Neglecting contributions from higher-order diagrams ($\Delta\phi_s^{\text{loop}} \approx 3 \pm 10$ mrad)
 - β_s determined by CKM global fits to be $-2\beta_s = -37 \pm 1$ mrad [CKMfitter, UTfit]
- **New physics** can change the value of ϕ_s up to $\sim 100\%$ via new particles contributing to the flavor oscillations [RMP88(2016)045002]



- This seminar presents the latest CMS results with the *golden* channel $B_s \rightarrow J/\psi \phi(1020) \rightarrow \mu^+\mu^- K^+K^-$

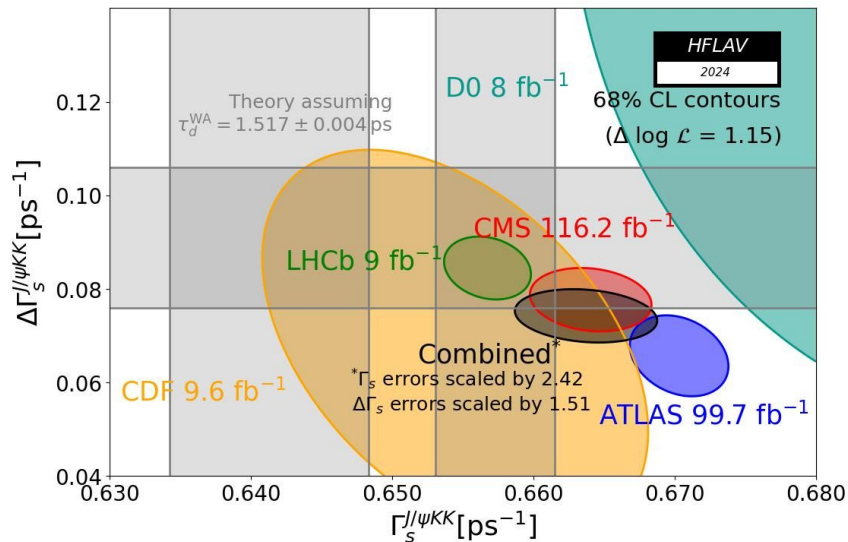
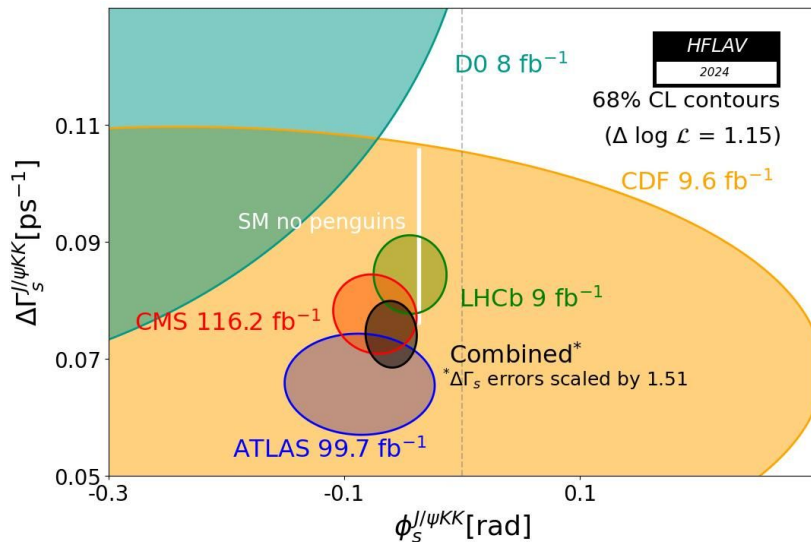


$$\Gamma(B_s^0 \rightarrow f)(t) \stackrel{?}{\neq} \Gamma(\bar{B}_s^0 \rightarrow f)(t)$$



A long history: flagship CPV analysis at hadron colliders

- ϕ_s has been **first measured** by the **Tevatron** experiments D0 and CDF
- At LHC it has been measured several times by ATLAS, LHCb, and CMS
 - LHCb has measured ϕ_s in several other channels, such as $B_s \rightarrow J/\psi \pi^+ \pi^-$, $B_s \rightarrow J/\psi (e^+ e^-) K^+ K^-$, $B_s \rightarrow \psi K^+ K^-$, $B_s \rightarrow D_s^+ D_s^-$, ...
- World-average: $\phi_s^{J/\psi KK} = -60 \pm 14 \text{ mrad}$ [\[HFLAV2024\]](#)



A time-, flavor- and angular-dependent measurement

Diagram illustrating the components of the CP violation measurement:

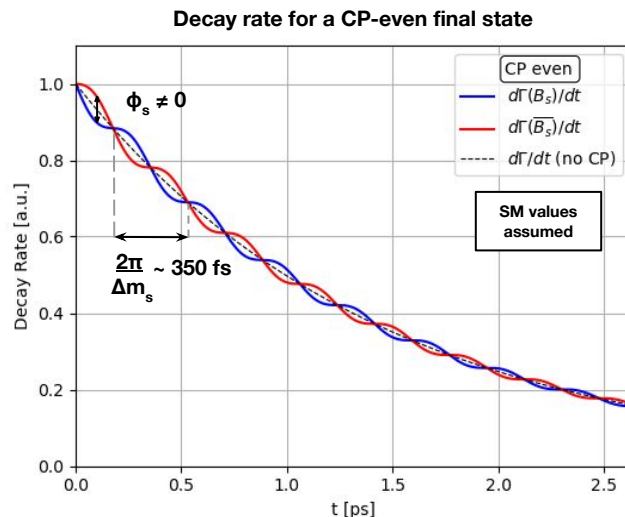
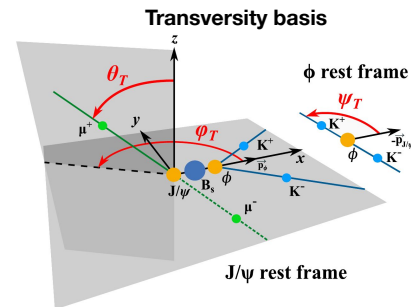
- final-state CP eigenvalue** (blue box) points to $\sin(\phi_s)$ in the numerator.
- CP violation** (red box) points to the $\sin(\phi_s)$ term.
- flavor oscillations** (green box) points to $\sin(\Delta m_s t)$ in the numerator.

$$a_{CP}(t) = \frac{-\eta_{fs} \sin(\phi_s) \sin(\Delta m_s t)}{\cosh(\frac{1}{2} \Delta \Gamma_s t) - \eta_{fs} \cos(\phi_s) \sinh(\frac{1}{2} \Delta \Gamma_s t)}$$

Core ingredients

- **Time-dependent angular analysis** to separate the final state CP eigenstates (mixture of CP-even and CP-odd)
- **Time-dependent flavor analysis** to resolve the B_s mixing oscillations ($T \sim 350$ fs)

$$\text{sensitivity} \propto \sqrt{\frac{\epsilon_{\text{tag}} \mathcal{D}_{\text{tag}}^2 N_{\text{sig}}}{2}} \sqrt{\frac{N_{\text{sig}}}{N_{\text{sig}} + N_{\text{bkg}}}} e^{-\frac{\sigma_t^2 \Delta m_s^2}{2}}$$



Decay rate model

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

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

Flavor tag decision (flips c_i and d_i signs)

Mistag probability

Decay time

Angular variables

Most sensitive terms for small ϕ_s

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(0) ^2$	1	D	C	$-S$
2	$\sin^2 \psi_T (1 - \sin^2 \theta_T \sin^2 \varphi_T)$	$ A_{\parallel}(0) ^2$	1	D	C	$-S$
3	$\sin^2 \psi_T \sin^2 \theta_T$	$ A_{\perp}(0) ^2$	1	$-D$	C	S
4	$-\sin^2 \psi_T \sin 2\theta_T \sin \varphi_T$	$ A_{\parallel}(0) A_{\perp}(0) $	$C \sin(\delta_{\perp} - \delta_{\parallel})$	$S \cos(\delta_{\perp} - \delta_{\parallel})$	$\sin(\delta_{\perp} - \delta_{\parallel})$	$D \cos(\delta_{\perp} - \delta_{\parallel})$
5	$\frac{1}{\sqrt{2}} \sin 2\psi_T \sin^2 \theta_T \sin 2\varphi_T$	$ A_0(0) A_{\parallel}(0) $	$\cos(\delta_{\parallel} - \delta_0)$	$D \cos(\delta_{\parallel} - \delta_0)$	$C \cos(\delta_{\parallel} - \delta_0)$	$-S \cos(\delta_{\parallel} - \delta_0)$
6	$\frac{1}{\sqrt{2}} \sin 2\psi_T \sin 2\theta_T \cos \varphi_T$	$ A_0(0) A_{\perp}(0) $	$C \sin(\delta_{\perp} - \delta_0)$	$S \cos(\delta_{\perp} - \delta_0)$	$\sin(\delta_{\perp} - \delta_0)$	$D \cos(\delta_{\perp} - \delta_0)$
7	$\frac{2}{3} (1 - \sin^2 \theta_T \cos^2 \varphi_T)$	$ A_S(0) ^2$	1	$-D$	C	S
8	$\frac{1}{3} \sqrt{6} \sin \psi_T \sin^2 \theta_T \sin 2\varphi_T$	$k_{SP} A_S(0) A_{\parallel}(0) $	$C \cos(\delta_{\parallel} - \delta_S)$	$S \sin(\delta_{\parallel} - \delta_S)$	$\cos(\delta_{\parallel} - \delta_S)$	$D \sin(\delta_{\parallel} - \delta_S)$
9	$\frac{1}{3} \sqrt{6} \sin \psi_T \sin 2\theta_T \cos \varphi_T$	$k_{SP} A_S(0) A_{\perp}(0) $	$\sin(\delta_{\perp} - \delta_S)$	$-D \sin(\delta_{\perp} - \delta_S)$	$C \sin(\delta_{\perp} - \delta_S)$	$S \sin(\delta_{\perp} - \delta_S)$
10	$\frac{4}{3} \sqrt{3} \cos \psi_T (1 - \sin^2 \theta_T \cos^2 \varphi_T)$	$k_{SP} A_S(0) A_0(0) $	$C \cos(\delta_0 - \delta_S)$	$S \sin(\delta_0 - \delta_S)$	$\cos(\delta_0 - \delta_S)$	$D \sin(\delta_0 - \delta_S)$

$$C = \frac{1 - |\lambda|^2}{1 + |\lambda|^2}$$

$$S = -\frac{2|\lambda| \sin \phi_s}{1 + |\lambda|^2}$$

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

Sensitive to
direct CPV

Sensitive to
 $\phi_s \sim 0$

Sensitive to
 $\phi_s \sim \pi/2$

Conventions

- $|A_{\parallel}|^2 = |A_0|^2 - |A_{\perp}|^2$
- $\delta_0 = 0$
- $\delta_{S\perp} = \delta_S - \delta_{\perp}$
- $\Delta\Gamma_s > 0$

Physics parameters

- $\phi_s, |\lambda|$
- $\Delta\Gamma_s, \Gamma_s, \Delta m_s$
- $|A_0|^2, |A_{\perp}|^2, |A_S|^2$
- $\delta_{\parallel}, \delta_{\perp}, \delta_{S\perp}$

S-P wave effective coupling

$k_{SP} \approx 0.54$

- Introduced since $m(K^+K^-)$ is not fitted
- Evaluated from the S- and P-wave lineshapes

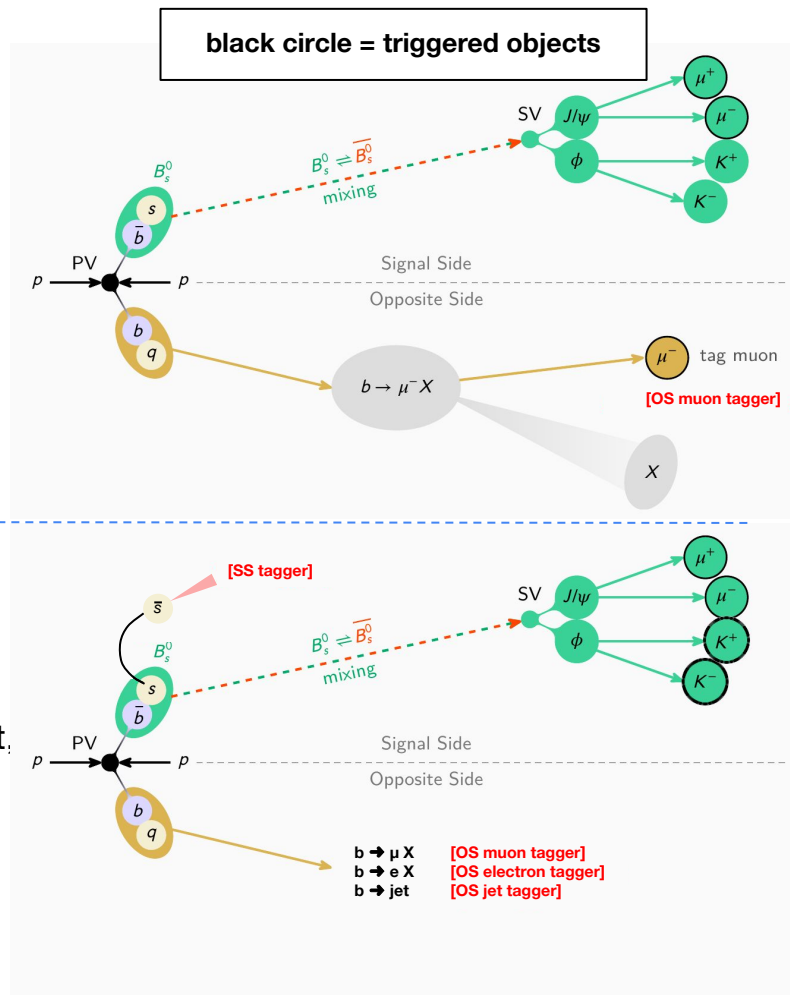
Trigger strategy

Muon-tagging trigger

- $J/\psi \rightarrow \mu^+\mu^-$ candidate plus an additional muon (for tagging)
- $\approx 50\,000$ signal candidates
- Used for time resolution modeling
- Tagging algorithms deployed: OS-muon
 - $P_{\text{tag}} \sim 10\%$ (trigger muon enhances fraction of tagged events)

Standard trigger

- Displaced $J/\psi \rightarrow \mu^+\mu^-$ candidate + $\phi(1020) \rightarrow K^+K^-$
- $\approx 450\,000$ signal candidates
- Tagging algorithms deployed: OS-muon, OS-electron, OS-jet, Same Side
 - $P_{\text{tag}} \sim 5\%$



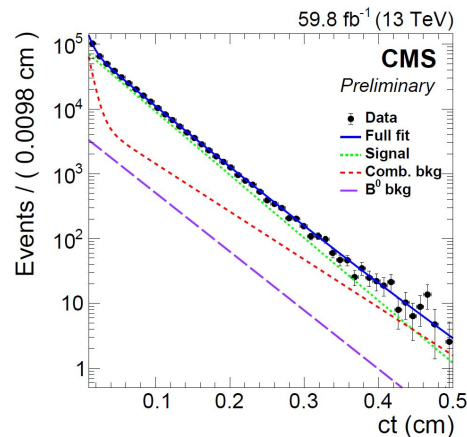
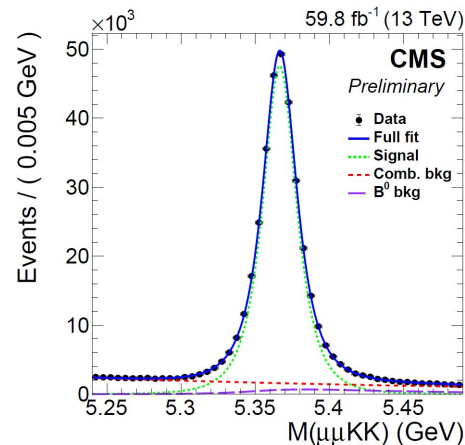
Dataset and selection

- **Dataset:** $L_{\text{int}} = 96 \text{ fb}^{-1}$ collected in 2017-2018
 - *Why no 2016 data?* Very different data set (old inner tracker detector with worse time resolution and different trigger menu)
- **Signal candidates:** $491\,270 \pm 950$
- Notable selection requirements:

Variable	Requirement
ct (<i>muon-tagging</i> HLT)	$> 60 \mu\text{m}$
ct (<i>standard</i> HLT)	$> 100 \mu\text{m}$
ct/σ_{ct} (<i>standard</i> HLT)	> 3
$ m(K^+K^-) - m_{\phi(1020)} $	$< 10 \text{ MeV}$
$ m(\mu^+\mu^-) - m_{J/\psi} $	$< 150 \text{ MeV}$

- To avoid **overlaps**, events that pass both trigger category selections are placed only in the *muon-tagging* one
 - This depletes the *standard* trigger category of OS muons
- The PV of choice is the closest in 3D to the line that passes through the SV and parallel to the B_s momentum

Invariant mass and proper decay length distributions for the *standard* trigger (2018)



Decay time and its resolution

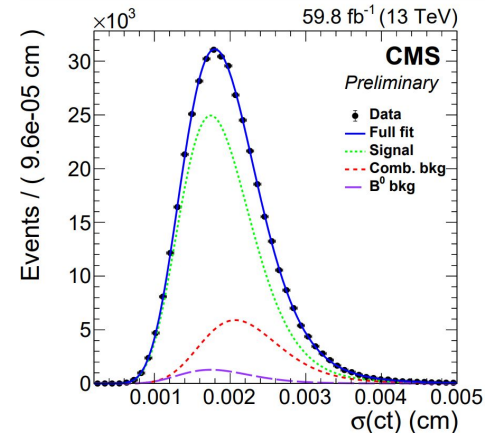
- The time dependence of the decay rate is parametrized with the **proper decay time** ct , measured in the transverse plane as

$$ct = c \cdot \frac{m_{Bs}^{w.a.} \cdot L_{xy}}{p_T} \quad \text{with} \quad L_{xy} \equiv ||\vec{r}_{xy}(SV) - \vec{r}_{xy}(PV)||$$

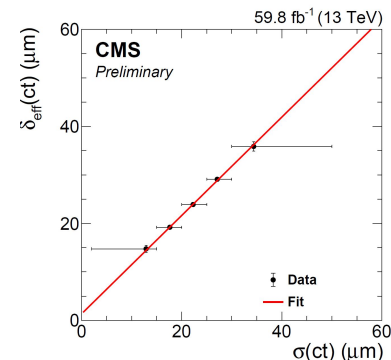
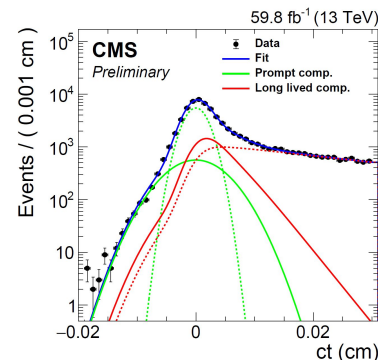
- Its **uncertainty** is obtained by fully propagating the uncertainties in L_{xy} and p_T
 - The uncertainty on L_{xy} dominates for most of the ct spectrum, with $\sigma(p_T)$ taking over at high values ($ct \gtrsim 3$ mm)
- The ct uncertainty is calibrated in a prompt data sample** of $B_s \rightarrow J/\psi \phi$, obtained by removing the displacement requirement in the *muon-tagging* data sets
 - Modeled with two gaussians to obtain the effective dilution and resolution

$$\delta_{\text{eff}} = \sqrt{\frac{-2 \ln \mathcal{D}}{\Delta m_s^2}} \quad \text{with} \quad \mathcal{D} = \sum_{i=1}^2 f_i \exp\left(-\frac{\sigma_i \Delta m_s^2}{2}\right)$$

- Excellent agreement** found, with corrections $\sim 5\%$



Time resolution calibration for 2018 data



Acceptance and efficiency effects

- The efficiency in selecting and reconstructing the B_s candidates is **not** independent of the decay time and angular observables
 - To properly fit the decay rate model an efficiency parametrization is needed

Time efficiency

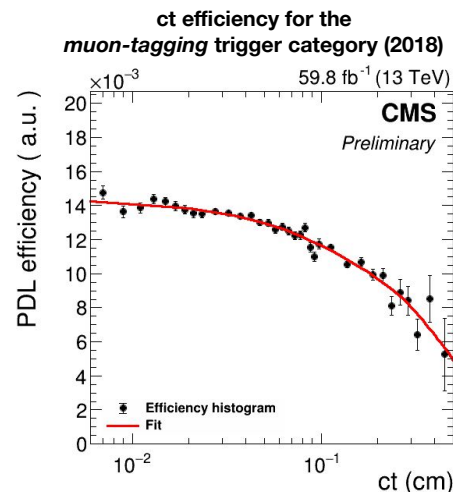
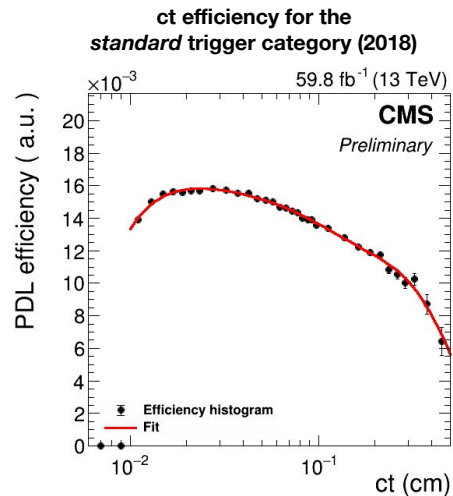
- Modeled in the $B^0 \rightarrow J/\psi K^{*0}$ data control channel with corrections from simulations
- Ultimately parametrized with Bernstein's polynomials

$$\varepsilon_{B^0}^{\text{data}}(ct) = \frac{N_{B^0}(ct)}{e^{-\Gamma_d^{\text{w.a.}}} \otimes P_{B^0}(\sigma_{ct})}$$

$$\varepsilon_{B_s}^{\text{data}}(ct) = \varepsilon_{B^0}^{\text{data}}(ct) \cdot \frac{\varepsilon_{B_s}^{\text{MC}}(ct)}{\varepsilon_{B^0}^{\text{MC}}(ct)}$$

Angular efficiency

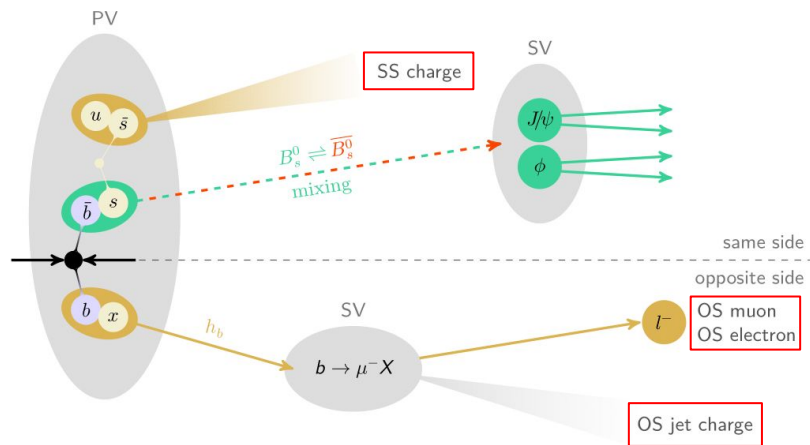
- Estimated with KDE distributions in simulated events
- The simulated data samples are corrected to match the data
 - An iterative procedure is used to simultaneously correct the kinematics of the final state particles and the differences in the physics parameters set in the MC with respect to what measured in the data



Flavor tagging overview

- A **cutting-edge flavor tagging framework** has been engineered to extract the best possible results from data
- **Four DNN-based algorithms are used**, divided into two main categories
 - **Same side (SS)**: exploits the B_s fragmentation
 1. **SS tagger**: leverages charge asymmetries near the B_s production
 - **Opposite side (OS)**: exploits decay products of the other B hadron in the event
 2. **OS muon**: leverages $b \rightarrow \mu^- X$ decays
 3. **OS electron**: leverages $b \rightarrow e^- X$ decays
 4. **OS jet**: capitalizes on charge asymmetries in the OS b -jet
- Only the OS-muon tagger is applied in the *muon-tagging* trigger category
 - The OS-electron, OS-jet and SS are applied only to the *standard* trigger category

Schematic representation of a generic event



Useful definitions

$$\xi_{tag} = \begin{cases} +1 & \text{for } B_s \\ -1 & \text{for } \bar{B}_s \\ 0 & \text{if no tagging decision is made} \end{cases}$$

$$\epsilon_{tag} = \frac{N_{tag}}{N_{tot}}, \quad \omega_{tag} = \frac{N_{mistag}}{N_{tag}}, \quad \mathcal{D}_{tag} = 1 - 2\omega_{tag}, \quad P_{tag} = \epsilon_{tag} \mathcal{D}_{tag}^2$$

Flavor, neural networks, and probabilities

- The **tagging inference logic** differs between techniques

- Lepton taggers** (OS muon, OS electron)

- Lepton charge $\rightarrow \xi_{\text{tag}}$; DNN score $\rightarrow \omega_{\text{tag}}$ *(DNN trained for correct-tag vs mistag)*

$$\begin{array}{l} \text{OS } \ell^- \rightarrow \text{OS } b \xrightarrow{\text{tag}} \text{signal } B_s \\ \text{OS } \ell^+ \rightarrow \text{OS } \bar{b} \xrightarrow{\text{tag}} \text{signal } \bar{B}_s \end{array}$$

$$\omega_{\text{tag}} = 1 - S_{\text{DNN}}$$

DNN score

- Charge-based taggers** (OS jet, SS)

- DNN score $\rightarrow \text{Prob}(B_s) \rightarrow \xi_{\text{tag}}, \omega_{\text{tag}}$

(DNN trained for B_s vs \bar{B}_s)

$$\begin{array}{ll} S_{\text{DNN}} > 0.5 + \epsilon \xrightarrow{\text{tag}} \text{signal } B_s & \text{with } \omega_{\text{tag}} = 1 - S_{\text{DNN}} \\ S_{\text{DNN}} < 0.5 - \epsilon \xrightarrow{\text{tag}} \text{signal } \bar{B}_s & \text{with } \omega_{\text{tag}} = S_{\text{DNN}} \end{array}$$

- ϵ is used to remove events with $\omega_{\text{tag}} \sim 50\%$

- The algorithms are optimized and trained in simulated events and calibrated in data with **self-tagging $B^+ \rightarrow J/\psi K^+$ decays**

- The calibration is performed by comparing the ω_{tag} predicted by the DNN and the one measured in data

Calibration strategy (and other tricks)

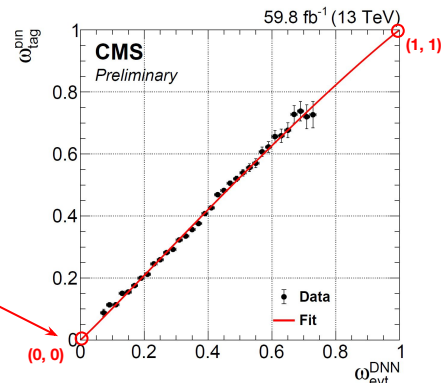
- A **multi-pronged strategy** has been devised to improve the ω_{tag} estimation and suppress systematic effects
 1. All models are constructed from the start as *probability estimators*, i.e. $\text{score} \sim \omega_{\text{tag}}$
 - Loss function: *cross-entropy*, which is the likelihood for the probability $P(\text{true class} | \text{score})$
 - Output layer: *Sigmoid* function, which normalizes the output to a probability distribution
 2. All DNNs are calibrated with the *Platt scaling*, which ensures that the calibrated score is still a probability
 - The Platt scaling is a linear calibration of the score before the last sigmoid layer
 3. In calibrating the charge-based taggers (which provide a probability for B_s vs \bar{B}_s):
 - A. The output is *symmetrized* due to the initial LHC charge imbalance

$$s_{DNN}^{\text{sym}}(x) = \frac{s_{DNN}(x) + [1 - s_{DNN}(\bar{x})]}{2}$$

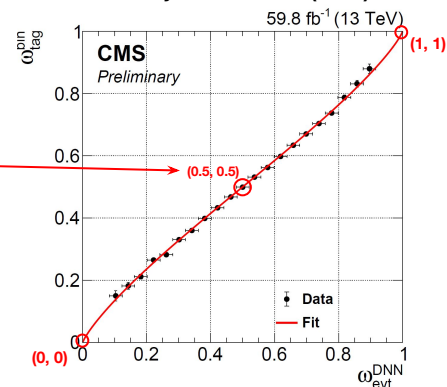
- B. The symmetry is explicitly forced in the calibration function by removing the constant term

This strategy **cancels** almost all the systematic effects associated with flavor tagging

OS-Muon calibration
(muon-tagging trigger 2018)



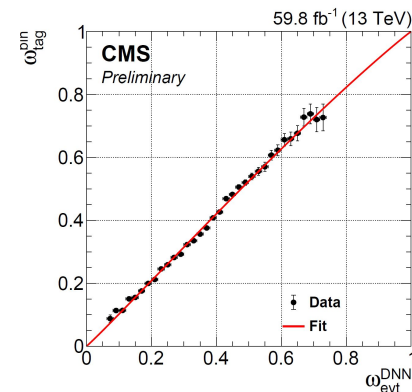
OS-jet calibration (2018)



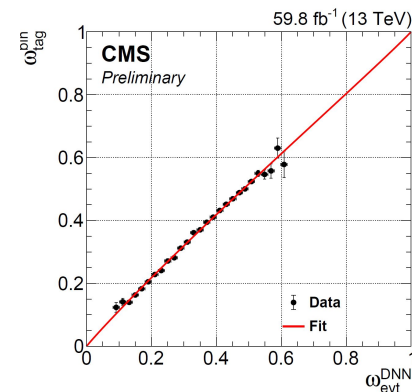
OS-lepton tagging

- OS-lepton tagging techniques search for $b \rightarrow \ell X$ decays of the other B hadron in the event
- The **charge** of the lepton is used as tagging feature and a fully connected DNN is used to estimate the mistag probability
- **Lepton selection**
 - Loose kinematic cuts
 - Separated from the signal B meson
 - MVA discriminator against fakes
 - OS-electrons are searched only if no OS-muon is found in the event (explicit orthogonality)
- **Mistag estimation**
 - Fully connected DNN with ReLU activation and dropout
 - Inputs: lepton kinematics and surrounding activity
- **Trained on simulated $B_s \rightarrow J/\psi \phi(1020)$ events and calibrated in $B^+ \rightarrow J/\psi K^+$ data**

OS-euon calibration
(muon-tagging trigger 2018)



OS-electron calibration (2018)

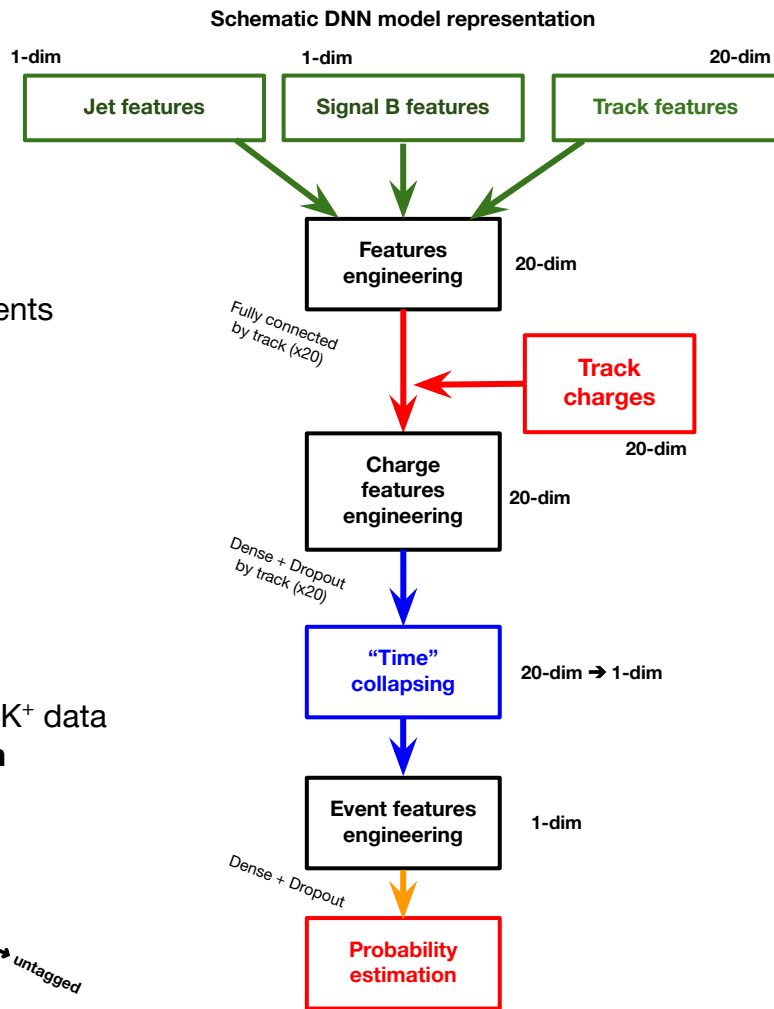


OS-jet tagging

- The OS-jet algorithm exploits charge asymmetries in the jet structure and is based on a DNN called **DeepJetCharge**
 - Inputs: features from signal B meson, OS jet and its constituents
 - Note: the only flavor asymmetry is in the charges
 - Based on the DeepSets architecture [\[ref\]](#)
- **Jet selection**
 - No OS-lepton candidate
 - At least 2 tracks with $|\text{IP}_z| < 1 \text{ cm}$
 - Separated from the signal B meson
 - jet b-tagging discriminator
- Additional nearby tracks are used due to the poor jet clustering performance in the kinematic region of interest ($p_T < 20 \text{ GeV}$)
- Trained on simulated $B_s \rightarrow J/\psi \phi$ events and calibrated in $B^+ \rightarrow J/\psi K^+$ data
- **The trained network produces the probability of signal B meson containing a \bar{b} quark (i.e. being a B_s)**
- The score is finally used to compute both ξ_{tag} and ω_{tag}

$$\begin{aligned}
 s_{DNN} > 0.52 &\xrightarrow{\text{tag}} \text{signal } B_s \text{ with } \omega_{\text{tag}} = 1 - s_{DNN} \\
 s_{DNN} < 0.48 &\xrightarrow{\text{tag}} \text{signal } \overline{B}_s \text{ with } \omega_{\text{tag}} = s_{DNN}
 \end{aligned}$$

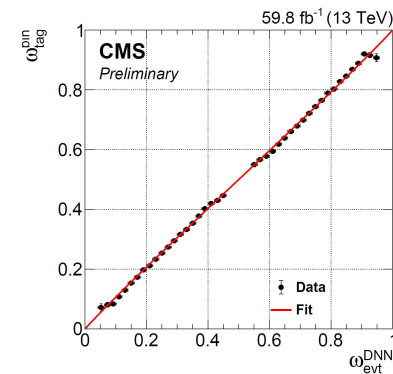
$\omega_{\text{tag}} > 0.48 \rightarrow \text{untagged}$



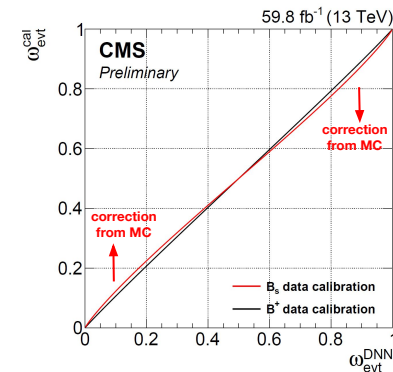
SS tagger

- The SS tagger consists of a DNN (**DeepSSTagger**), derived from **DeepJetCharge**, able to probe the fragmentation products of a B meson and exploit tracks with high flavor correlation
- **DeepSSTagger** uses the kinematic information from up to 20 tracks (ordered by $|\text{IP}_z|$) around the reconstructed B meson
- **Track selection**
 - $\Delta R(\text{trk}, B) < 0.8$, $|\text{IP}_z(\text{PV})| < 0.4 \text{ cm}$, $|\text{IP}_{xy}(\text{PV})|/\sigma_{\text{dxy}} < 1$
 - Overlap with signal and OS is carefully avoided with geometrical cuts and vetos
- **Trained on an equal-weight mixture** of $B_s \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ to make the model invariant for $B_s \leftrightarrow B^+$ for calibration purposes
 - Calibration directly in B_s was found to be not feasible in CMS
 - Tested: $B_s \rightarrow D_s \pi^+$ (not enough stat.) and $B_s^{**} \rightarrow B^{+(*)} K^-$ (too much uncer. from B^{0**} bkg)
 - The trained network produces the probability of signal B meson containing a negatively charged quark alongside the b quark (i.e., being a B_s or B^-)
- **Calibration**
 - The SS is calibrated $B^+ \rightarrow J/\psi K^+$ data, with residual differences $\sim 10\%$ corrected with simulations
 - Events with $\omega_{\text{tag}} > 0.46$ are removed before the calibration and assumed untagged

Same-side tagger calibration
(B^+ data 2018)



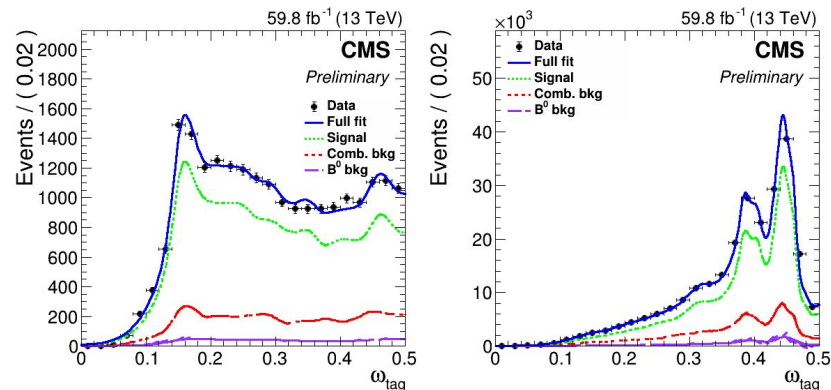
Comparison between Same-side
tagger B^+ and B_s calibrations (2018)



Flavor tagging performance

- The SS and any one of the OS algorithms overlap in about 20% of the events
 - In these cases, the information is combined to improve the tagging inference
- The combined flavor tagging framework achieves a tagging power of $P_{\text{tag}} = 5.6\%$** when applied to the B_s data sample
 - Among the highest ever recorded at LHC
 - x3~4 improvement with respect to prev. CMS results
- This is the first CMS implementation of the OS jet and same-side tagging techniques**
 - SS accounts for half of the performance
- Largest ever effective statistics $N_{B_s} \cdot P_{\text{tag}}$ ($490k \cdot 5.6\% \approx 27.5k$) for a single ϕ_s measurement
- The flavor tagging framework is validated in the $B^0 \rightarrow J/\psi K^{*0}$ data control channel with flavor mixing measurements, both integrated and time-dependent

ω_{tag} distribution in the *muon-tagging* trigger category (left) and the *standard* one (right) for 2018 data



Flavor tagging performance (mutually exclusive categories)

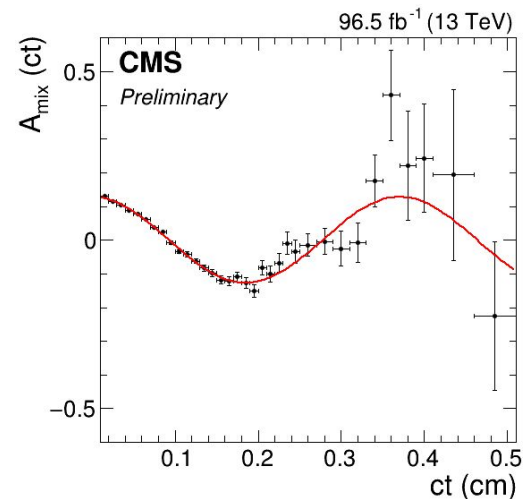
Category	$\epsilon_{\text{tag}} [\%]$	D_{eff}^2	$P_{\text{tag}} [\%]$
Only OS muon	6.07 ± 0.05	0.212	1.29 ± 0.07
Only OS electron	2.72 ± 0.02	0.079	0.214 ± 0.004
Only OS jet	5.16 ± 0.03	0.045	0.235 ± 0.003
Only SS	33.12 ± 0.07	0.080	2.64 ± 0.01
SS + OS muon	0.62 ± 0.01	0.202	0.125 ± 0.003
SS + OS electron	2.77 ± 0.02	0.150	0.416 ± 0.005
SS + OS jet	5.40 ± 0.03	0.124	0.671 ± 0.006
Total	55.9 ± 0.1	0.100	5.59 ± 0.02

Tagging validation with B^0 events

- The flavor tagging framework is validated in the $B^0 \rightarrow J/\psi K^{*0}$ control channel (~2M events)
- The time-dependent **mixing asymmetry** is measured to extract the flavor mixing oscillation frequency Δm_d with a precision of ~1% (comparable with BaBar and Belle)
 - Excellent agreement with world-averages is observed
→ **No bias** in mixing frequency measurements
- Study performed also in each tagging category (see backup)
- The **time-integrated mixing** is also measured for each tagger and their dependency on the expected tagging dilution is compared
 - The dependency between the measured A_{mix} and the estimated D_{tag} is found to be well described by a linear relationship, indicating that all four techniques behave in the same predictable way

B^0 flavor mixing asymmetry

$$A_{\text{mix}}(\text{ct}) = \frac{N_{\text{unmix}}(\text{ct}) - N_{\text{mix}}(\text{ct})}{N_{\text{unmix}}(\text{ct}) + N_{\text{mix}}(\text{ct})}$$



Fit model

- The physics parameters are extracted with **unbinned multidimensional extended maximum-likelihood (UML) fit** performed simultaneously on **12 data sets** (2 trig. cat. x 2 years x 3 ξ_{tag} values)
 - Physics parameters:* ϕ_s , $|\lambda|$, $\Delta\Gamma_s$, Γ_s , Δm_s , $|A_0|^2$, $|A_\perp|^2$, $|A_s|^2$, δ_\parallel , δ_\perp , $\delta_{S\perp}$
 - Observables:* m_{B_s} , ct , σ_{ct} , $\cos\theta_T$, $\cos\psi_T$, ϕ_T , ω_{tag}
- Fit model**

$$P = f_{\text{sig}} P_{\text{sig}} + f_{\text{bkg}} P_{\text{bkg}} + f_{\text{bkg } B^0} P_{\text{bkg } B^0}$$

Observables pdfs

SIGNAL P_{sig}

$\xrightarrow[\text{Angular eff}]{B_s \rightarrow J/\psi \phi \text{ decay rate}}$

$\varepsilon(\Theta) [f(\Theta, ct | \alpha, \xi_{\text{tag}}, \omega_{\text{tag}}) \otimes G(ct, \sigma_{ct})] P_{\text{sig}}(m_{B_s}) P_{\text{sig}}(\sigma_{ct}) P_{\text{sig}}(\omega_{\text{tag}})$

COMBINATORIAL BKG P_{bkg}

$[P_{\text{bkg}}(ct) \otimes G(ct, \sigma_{ct})]$

$P_{\text{bkg}}(\Theta) P_{\text{bkg}}(m_{B_s}) P_{\text{bkg}}(\sigma_{ct}) P_{\text{bkg}}(\omega_{\text{tag}})$

$B^0 \rightarrow J/\psi K^{0*}$ BKG $P_{\text{bkg } B^0}$

$\xrightarrow[\text{Bkg time pdf}]{\text{Time resolution convolution}}$

$[P_{\text{bkg } B^0}(ct) \otimes G(ct, \sigma_{ct})] P_{\text{bkg } B^0}(\Theta) P_{\text{bkg } B^0}(m_{B_s}) P_{\text{bkg } B^0}(\sigma_{ct}) P_{\text{bkg } B^0}(\omega_{\text{tag}})$

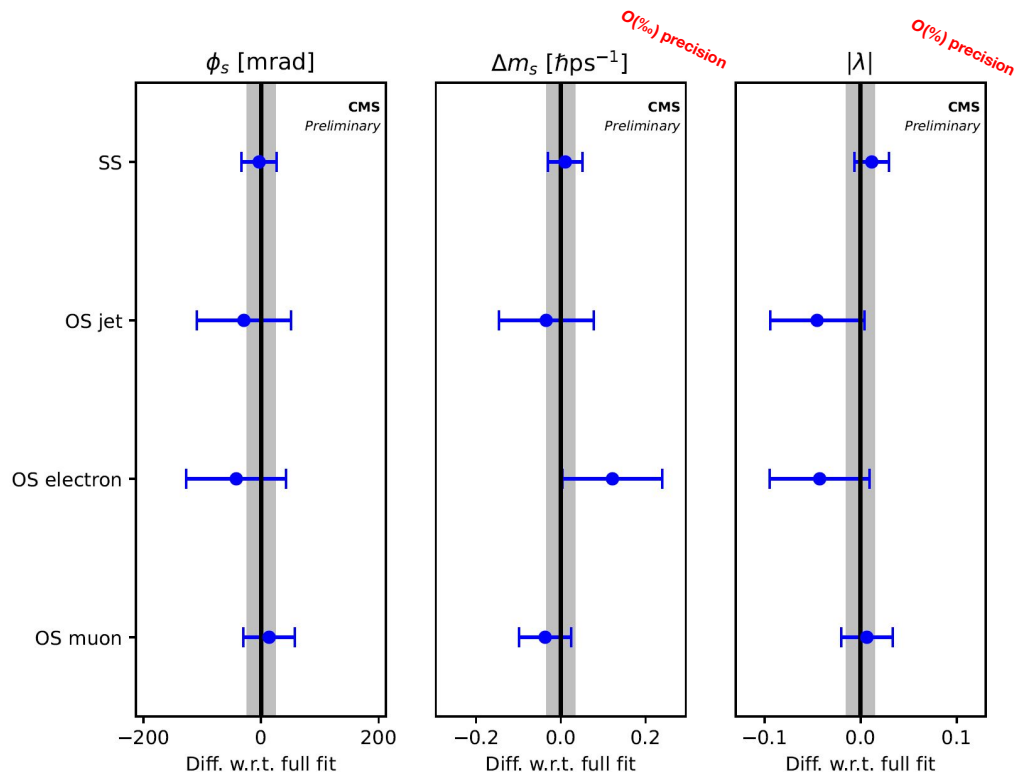
- The time efficiency is implemented as a *re-weighting* of the data events to drastically improve fit time
- The statistical uncertainties and fit bias are estimated with **1300 bootstrap distributions**
- The yield for the **$B^0 \rightarrow J/\psi K^{0*}$** is estimated **directly in data** with a 2D fit to the B_s invariant mass and its B^0 reflection
- The background from $\Lambda_b \rightarrow J/\psi K^- p^+$ is **found to be negligible** and is treated as a systematic uncertainty

Systematic uncertainty overview

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps ⁻¹]	Γ_s [ps ⁻¹]	Δm_s [ħps ⁻¹]	$ \lambda $	$ A_0 ^2$	$ A_\perp ^2$	$ A_S ^2$	δ_\parallel [rad]	δ_\perp [rad]	$\delta_{S\perp}$ [rad]
Statistical uncertainty	23	0.0043	0.0015	0.035	0.014	0.0016	0.0021	0.0033	0.074	0.089	0.15
Model bias	4	0.0011	0.0002	0.004	0.006	0.0012	0.0022	0.0006	0.015	0.017	0.03
Flavor tagging	4	< 10 ⁻⁴	0.0005	0.007	0.002	< 10 ⁻⁴	< 10 ⁻⁴	0.0006	0.012	0.016	0.03
Angular efficiency	4	0.0002	< 10 ⁻⁴	0.015	0.011	0.0042	0.0019	0.0001	0.017	0.044	0.02
Time efficiency	< 1	0.0014	0.0026	< 10 ⁻³	< 10 ⁻³	0.0004	0.0005	< 10 ⁻⁴	0.001	0.002	< 10 ⁻²
Time resolution	< 1	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻³	< 10 ⁻³	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻³	0.001	< 10 ⁻³
Model assumptions	—	0.0005	0.0006	—	—	—	—	—	—	—	—
B ⁰ background	< 1	0.0002	0.0003	< 10 ⁻³	< 10 ⁻³	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻³	< 10 ⁻³	< 10 ⁻²
Λ _b ⁰ background	—	—	0.0004	—	—	0.0004	0.0003	—	—	—	—
S-P wave interference	< 1	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻³	< 10 ⁻³	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻⁴	< 10 ⁻³	< 10 ⁻³	< 10 ⁻²
P(σ _{ct}) uncertainty	< 1	0.0002	0.0003	< 10 ⁻³	< 10 ⁻³	0.0001	0.0001	< 10 ⁻⁴	< 10 ⁻³	< 10 ⁻³	< 10 ⁻²
Total systematic uncertainty	7	0.0019	0.0028	0.017	0.012	0.0044	0.0030	0.0009	0.025	0.050	0.05

- **Model bias, flavor tagging, and angular efficiency are found to be the leading systematic sources for ϕ_s**
- The measurement is still heavily statistically limited for ϕ_s and will be until the end of HL-LHC

Validation: fit with individual tagging techniques



- To check the consistency and stability of the tagging framework, the fit to data is repeated with only one tagging algorithm deployed at a time
 - The grey area represents the result and statistical uncertainty of the full fit
 - Only flavor-sensitive parameters are presented
- **Excellent** agreement between the various tagging techniques

Results

Fit results

Parameter	Fit value	Stat. unc.	Syst. unc.
ϕ_s [mrad]	-73	± 23	± 7
$\Delta\Gamma_s$ [ps^{-1}]	0.0761	± 0.0043	± 0.0019
Γ_s [ps^{-1}]	0.6613	± 0.0015	± 0.0028
Δm_s [$\hbar\text{ps}^{-1}$]	17.757	± 0.035	± 0.017
$ \lambda $	1.011	± 0.014	± 0.012
$ A_0 ^2$	0.5300	± 0.0016	± 0.0044
$ A_\perp ^2$	0.2409	± 0.0021	± 0.0030
$ A_S ^2$	0.0067	± 0.0033	± 0.0009
δ_\parallel	3.145	± 0.074	± 0.025
δ_\perp	2.931	± 0.089	± 0.050
$\delta_{S\perp}$	0.48	± 0.15	± 0.05

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

$$\phi_s^{SM} \simeq -37 \pm 1 \text{ mrad} \quad \Delta\Gamma_s^{SM} = 0.091 \pm 0.013 \text{ ps}^{-1}$$

- Γ_s and Δm_s are **consistent** with the latest world averages

$$\Gamma_s^{WA} = 0.6573 \pm 0.0023 \text{ ps}^{-1} \quad \Delta m_s^{WA} = 17.765 \pm 0.006 \hbar\text{ps}^{-1}$$

- $|\lambda|$ is **consistent** with no direct CPV ($|\lambda| = 1$)

- This measurement utilizes the **largest ever** effective statistics $N_{B_s} \cdot P_{\text{tag}}$ for a single ϕ_s measurement
 - The precision on ϕ_s is comparable with the world's most precise single measurement by LHCb ($\phi_s = -39 \pm 22$ (stat) ± 6 (syst) mrad) [PRL132\(2024\)051802](https://arxiv.org/abs/2405.1802)
 - This is the most precise single measurement of $\Delta\Gamma_s$ to date in this channel

Combination with 8 TeV results

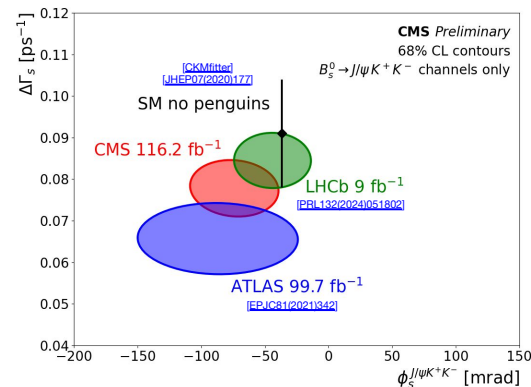
- These results supersede [PLB816\(2021\)136188](#) and are further combined with those obtained CMS at 8 TeV [\[PLB757\(2016\)97\]](#), yielding

$$\phi_s = -74 \pm 23 \text{ [mrad]}$$

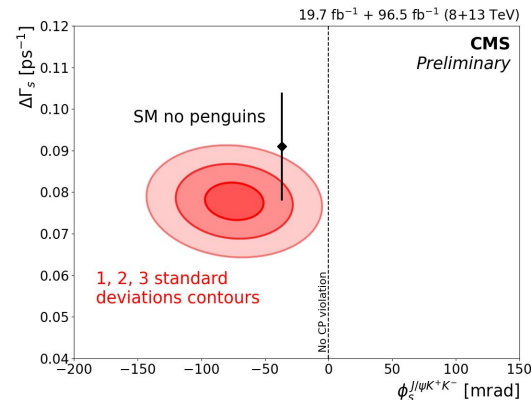
$$\Delta\Gamma_s = 0.0780 \pm 0.0045 \text{ [ps}^{-1}\text{]}$$

- Due to the high difference in statistical power between the two results the sensitivity gain is small
- The combined value for the weak phase ϕ_s is consistent with the SM prediction, the latest world average, and with zero (i.e. no CPV) at 3.2 s.d.**
 - This is the **first** evidence of CPV in $B_s \rightarrow J/\psi K^+ K^-$ decays
- These results helps to further constrain possible BSM effects in the B_s system

Comparison with other LHC experiments



1, 2, 3 standard deviations contours



Outlook

Summary and outlook

- This seminar presented the first **evidence** of time-dependent CP violation in $B_s \rightarrow J/\psi \phi$
-
- CMS recent contributions in flavor physics prove that it can be one of the leading actors in several key areas of study, such as rare decays and CP violation
- Advancements in trigger strategies and flavor tagging techniques allows CMS to compete in measurements for which the detector was not designed
- **Run 3 will provide unique opportunities thanks of a revamped trigger strategy, which will lead to the collection of an unprecedented amount of data suitable for flavor physics studies**

Stay tuned in the future for other exciting CMS results!

Thanks for the attention

Backup

Penguin contributions

We
measure
this

$$\begin{aligned} \phi_s &= \phi_s^{\text{tree}} + \Delta\phi_s^{\text{penguin}} + \Delta\phi_s^{\text{NP}} \\ \sin(2\beta) &= \sin(2\beta^{\text{tree}} + \Delta\phi_d^{\text{penguin}} + \Delta\phi_d^{\text{NP}}) \end{aligned}$$

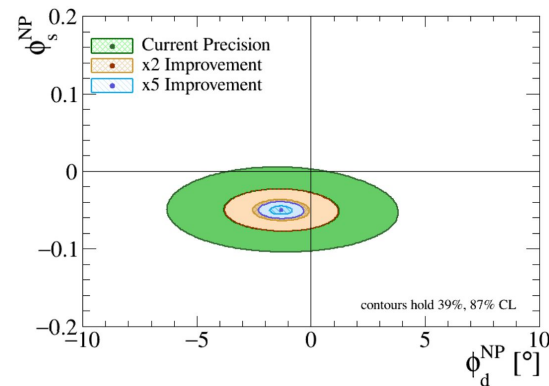
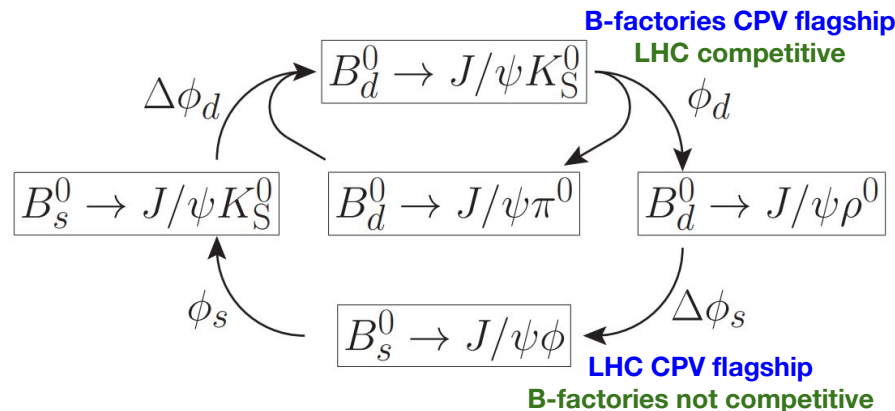
Assuming
this is
negligible

Trying to
probe this

- Penguin pollutions are expected to be small for B_s , but they are not well constrained

$$\Delta\phi_s^{\text{penguin}} \approx 3 \pm 10 \text{ mrad}$$

- Analysis of penguin and NP contributions is possible using Cabibbo-favored control channels



Offline selection

Requirements common between the two HLTs

- $5.24 < m(\mu\mu KK) < 5.49$ GeV
- $p_T(B_s) > 9.5$ GeV
- Vertex probability $> 2\%$
- $\sigma(ct) < 50$ μm
- $|\eta(\mu)| < 2.4$
- $|\eta(K)| < 2.5$
- $|m(\mu\mu) - m(J/\psi^{\text{PDG}})| < 150$ MeV
- $|m(KK) - m(\phi(1020)^{\text{PDG}})| < 10$ MeV

Requirements specific to the *muon-tagging* HLT

- $p_T(\mu) > 3.5$ GeV
- $p_T(K) > 1.15$ GeV
- $ct > 60$ μm

Requirements specific to the *standard* HLT

- *muon-tagging* trigger vetoed
- $p_T(\mu) > 4$ GeV
- $p_T(K) > 0.9$ GeV
- $p_T(\mu\mu) > 6.9$ GeV
- $ct > 100$ μm , $ct/\sigma(ct) > 3$

- Selection requirement **optimized** with the a genetic algorithm to maximize $S/\sqrt{(S + B)}$
- To avoid **overlaps**, the *muon-tagging* trigger is vetoed in the *standard* trigger category
- The **PV** of choice is the closest in 3D to the line that passes through the SV and parallel to the B_s momentum

OS-lepton taggers selection

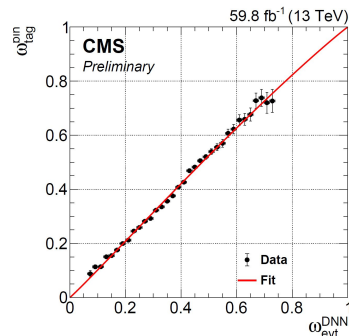
OS Muon

- **Requirements**
 - $p_T > 2 \text{ GeV}$
 - $|\eta| < 2.4$
 - $|d_z(\text{PV})| < 1 \text{ cm}$
 - $\Delta R(B_s) > 0.4$
 - Discriminators vs fakes
- Deployed in **both trigger categories**
- Dense DNN for ω_{tag} estimation
 - Inputs: kinematics, IP, surrounding activity

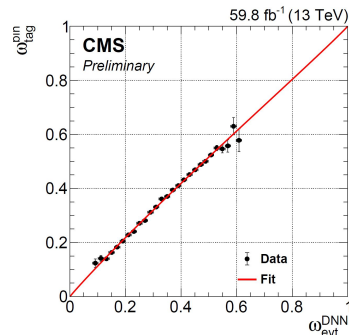
OS electron

- **Requirements**
 - No OS muon selected in the event
 - $p_T > 2.5 \text{ GeV}$
 - $|\eta| < 2.4$
 - $|d_z(\text{PV})| < 0.2 \text{ cm}$
 - $|d_{xy}(\text{PV})| < 0.08 \text{ cm}$
 - $\Delta R(B_s) > 0.4$
 - Discriminators vs fakes
- Deployed **only** in the **standard trigger category**
- Dense DNN for ω_{tag} estimation
 - Inputs: kinematics, IP, surrounding activity

OS-Muon calibration
(muon-tagging HLT 2018)



OS-Electron calibration (2018)



Taggers combination

- **Overlap logic**

Overlap	OS muon	OS electron	OS jet	SS
OS muon		X	X	✓
OS electron	X		X	✓
OS jet	X	X		✓
SS	✓	✓	✓	

- **Tag decision combination**

$$\xi(\xi_1, \xi_2, \omega_1, \omega_2) = \begin{cases} \xi_1 & \text{if } \omega_1 < \omega_2 \\ \xi_2 & \text{if } \omega_2 < \omega_1 \end{cases}$$

- **Mistag combination**

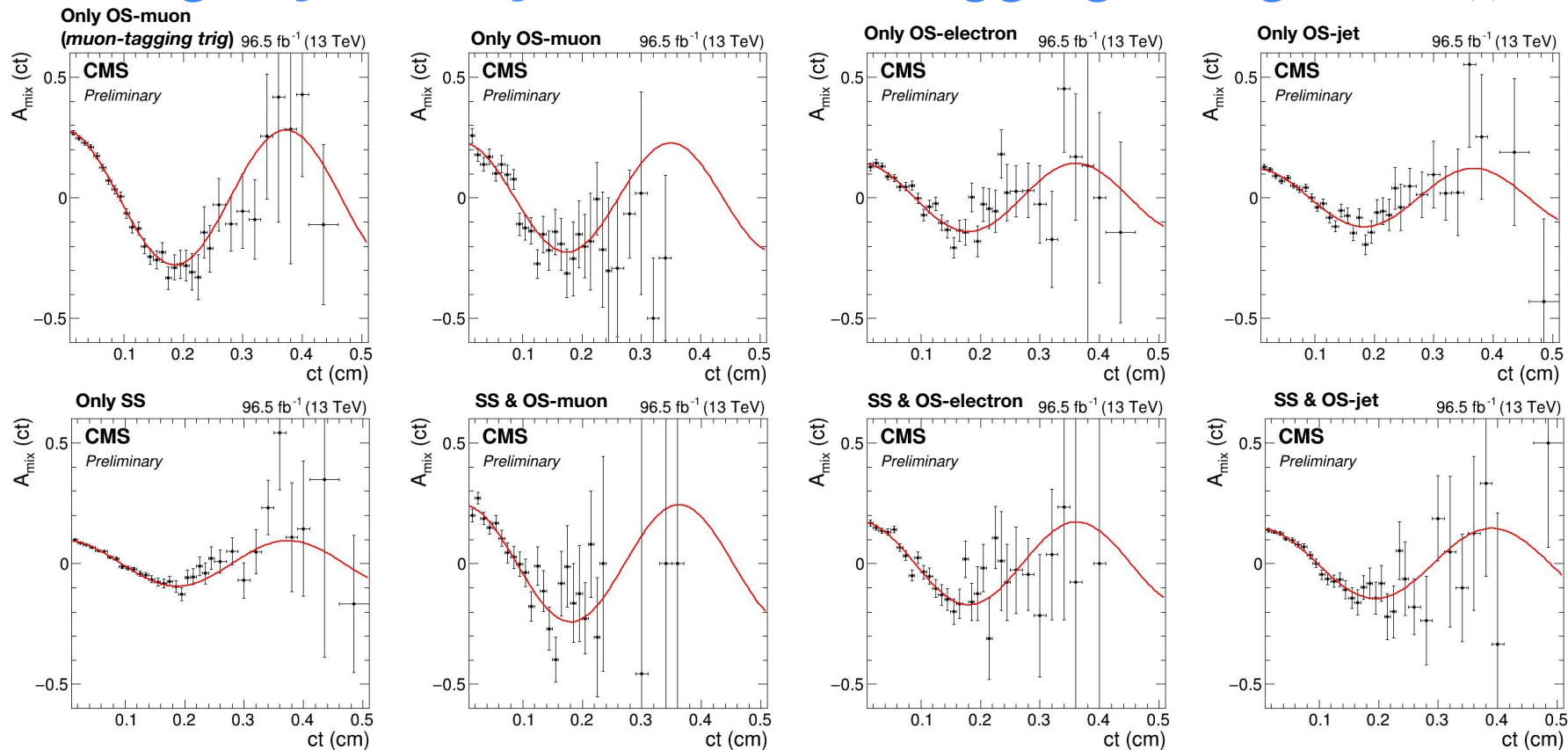
$$p(\bar{b}) = \prod_{i=1}^2 \left(\frac{1 - \xi_i}{2} + \xi_i(1 - \omega_i) \right) \quad p(b) = \prod_{i=1}^2 \left(\frac{1 + \xi_i}{2} - \xi_i(1 - \omega_i) \right)$$

$$P(\bar{b}) = \frac{p(\bar{b})}{p(\bar{b}) + p(b)} \quad P(b) = \frac{p(b)}{p(\bar{b}) + p(b)}$$

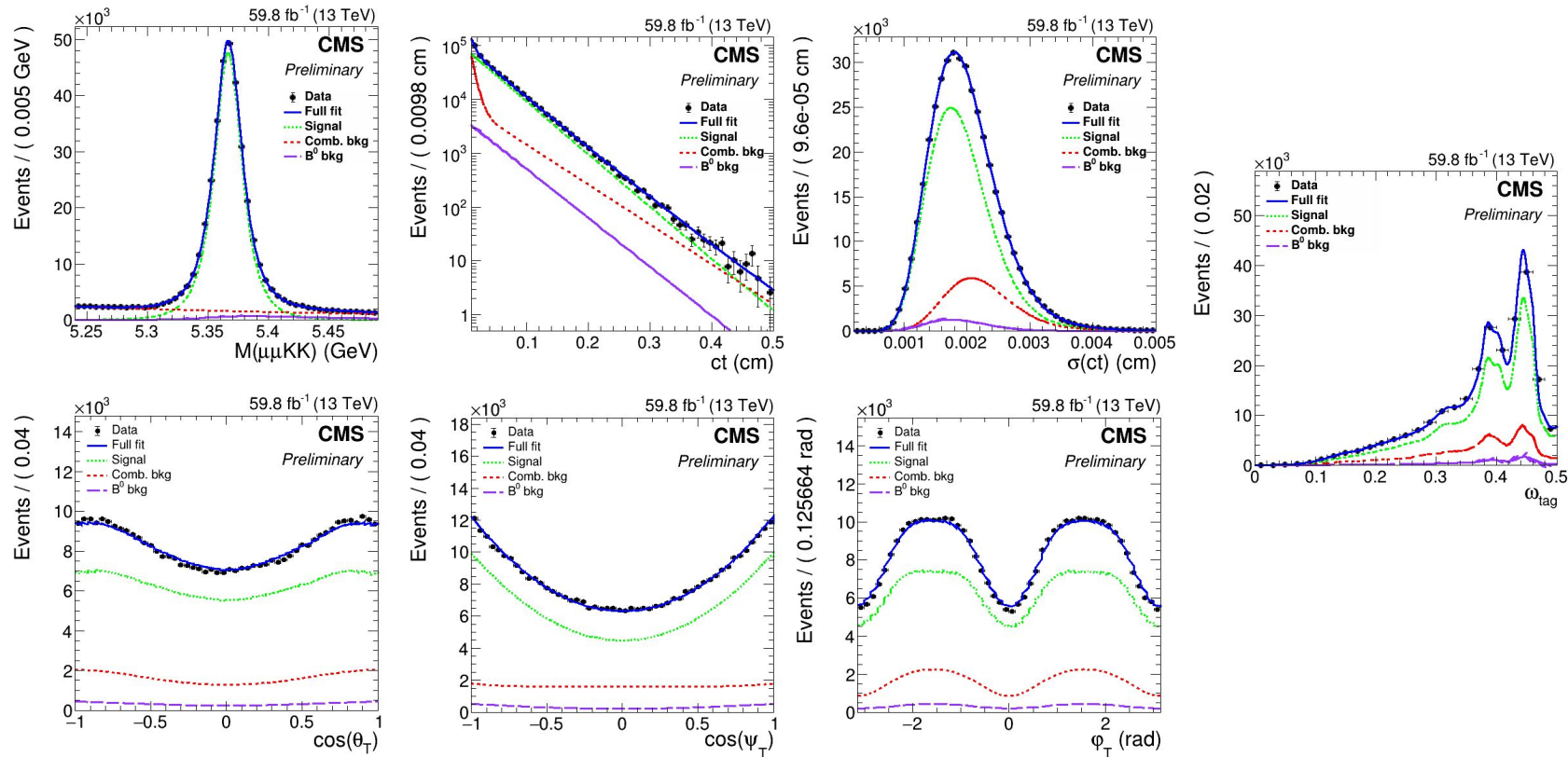
All categories are
mutually exclusive

All, but the first, refers to
the *standard trigger*
category

Mixing asymmetry for different tagging categories



Fit projections (standard trigger category 2018)



Systematic uncertainty classification

- **Type-I: unaccounted uncertainties**
 - Account for the finite statistics of simulated/control samples and uncertainties in calibrations and efficiency
 - **Always** propagated to the final results
 - Evaluated with two procedures
 1. Type-I full: obtained by sampling the samples/parameters of interest ~ 100 times, repeating the fit each time, and taking the RMS of the results as uncertainty
 2. Type-I simple: obtained by sampling a parameter only two times at $\pm 1\sigma_{\text{stat}}$
- **Type-II: method and model assumptions**
 - Account for **possible** bias induced by the assumptions made in the fit model and the analysis methods
 - Evaluated only if a **significant** bias is observed while testing an alternative (good) hypothesis
 - A significant bias for a parameter V is defined as a difference Δ in the fit results of **more than 20% of its σ_{stat}**
 - In these cases, **half of the bias** is taken as uncertainty, assuming that the *true* bias is uniformly distributed between 0 and Δ

The fit bias does **not** fall into either of these two categories

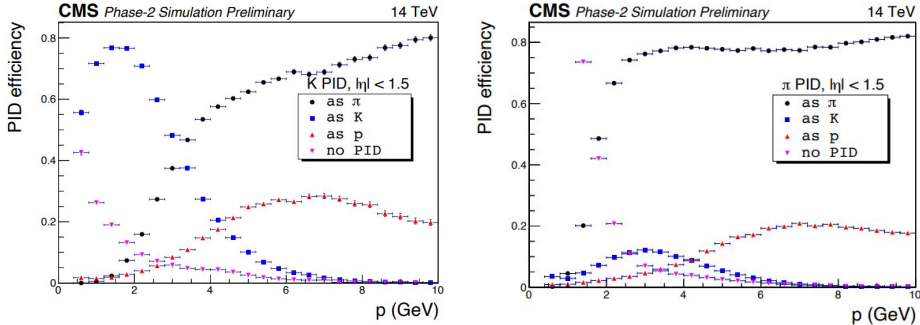
Comparison with theory and world averages

Parameter	Measured value	World-average value	Theory prediction	
ϕ_s [mrad]	-73 ± 24	-49 ± 19	-37 ± 1	[CKMfitter, UTfit]
$\Delta\Gamma_s$ [ps^{-1}]	0.0761 ± 0.0047	0.084 ± 0.005	0.091 ± 0.013	[Lenz & Tetlalmatzi-Xolocotzi]
Γ_s [ps^{-1}]	0.6613 ± 0.0032	0.6573 ± 0.0023	—	
Δm_s [$\hbar\text{ps}^{-1}$]	17.757 ± 0.039	17.765 ± 0.006	18.77 ± 0.86	[Lenz & Tetlalmatzi-Xolocotzi]
$ \lambda $	1.011 ± 0.018	1.001 ± 0.018	1	
$ A_0 ^2$	0.5300 ± 0.0047	0.520 ± 0.003	—	
$ A_\perp ^2$	0.2409 ± 0.0037	0.253 ± 0.006	—	
$ A_S ^2$	0.0067 ± 0.0034	0.030 ± 0.005	—	
δ_\parallel	3.145 ± 0.078	3.18 ± 0.06	—	
δ_\perp	2.931 ± 0.102	3.08 ± 0.12	—	
$\delta_{S\perp}$	0.48 ± 0.16	0.23 ± 0.05	—	

Flavor tagging in Phase-2 with MTD

- The MTD (Mip Timing Detector) provides time information of charged tracks at its surface
- The reconstruction algorithm utilizes compatible times of tracks from a vertex to offer time-of-flight based particle identification (PID) as a natural byproduct
- Same-side tagging could utilize charge correlation between the s-quark in the B_s and a nearby soft kaon for flavor tagging
- The PID from MTD, when integrated in the Phase-2 extrapolation of this analysis, shows a significant improvement of the tagging performances

Simulated PID efficiencies



Relative gain in P_{tag} (only SS)

PID scenario	Gains in P_{tag}
MC truth (perfect PID < 3 GeV)	+66%
PID with $\sigma_{\text{BTL}} = 40$ ps	+24%
PID with $\sigma_{\text{BTL}} = 70$ ps	+14%