Measurement of the CP-violating phase ϕ_s with CMS: present and future

Alberto Bragagnolo[a], on behalf of the CMS Collaboration
[a] University & INFN, Padova

CKM 2023, Santiago de Compostela (ES)
18-22 September 2023

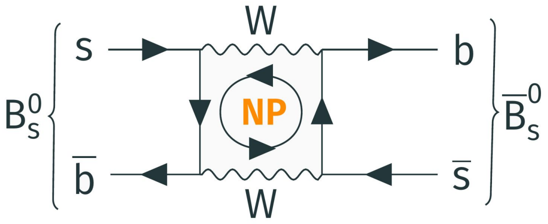
Outline

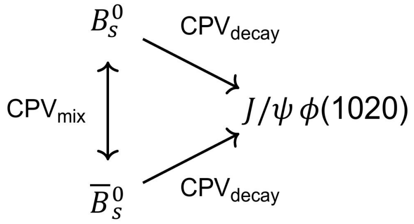
1. Introduction
2. Measuring ϕ_s with the CMS detector
3. Analysis and results
4. Future prospects
5. Conclusions

What? Why?
Who? Where?
How?
What now?

Motivations

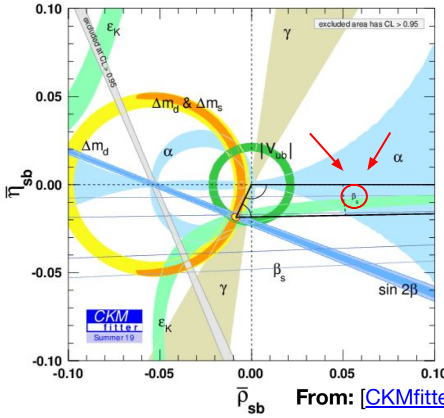
- **Decays of B_s mesons allow to study the time-dependent CP violation generated by the interference between direct decays and flavour mixing**
  - CPV in the interference is possible even if no CPV in decay and mixing
  - *Golden channel:* $B_s \rightarrow J/\psi\, \phi(1020) \rightarrow \mu^+\mu^- K^+K^-$
- **The weak phase ϕ_s is the main CPV observable**
  - Precisely predicted by the SM to be $\boldsymbol{\phi_s \approx -2\beta_s \approx -37 \pm 1}$ **mrad**, where β_s is one of the angles of the B_s unitary triangle (determined very accurately by CKM global fits) [CKMfitter, UTfit]
- New physics can change the value of ϕ_s up to ~100% via new particles contributing to the flavour oscillations [RMP88(2016)045002]





$$\Gamma\left(B^0_{s\,(\rightsquigarrow \overline{B}^0_s)} \to f\right)(t) \overset{?}{\neq} \Gamma\left(\overline{B}^0_{s\,(\rightsquigarrow B^0_s)} \to f\right)(t)$$

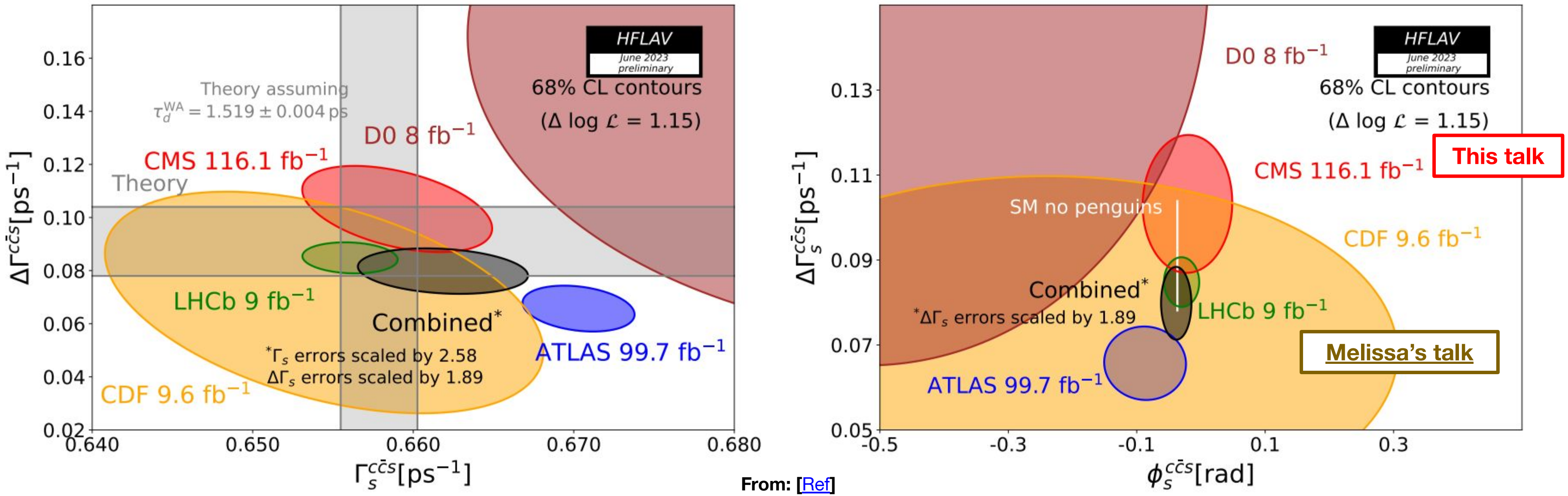
$$a_{CP}(t) \propto \Gamma_{\overline{B}_s \to f}(t) - \Gamma_{B_s \to f}(t) \propto -\eta_{fs} \sin(\phi_s) \sin(\Delta m_s t)$$



From: [CKMfitter]

State of the art (w. latest preliminary results from LHCb)

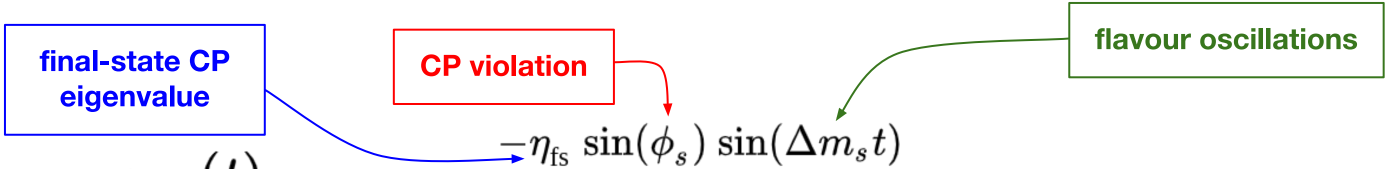
- Measurement **statistically limited** ➜ long-term commitment by multiple experimental collaborations
- Very active theoretical community (NP limits, penguin pollutions, predictions, …)
- Precision on ϕ_s **close** to 3 s.d. sensitivity for CPV in decay/mixing interference
  - $\sigma^{WA}(\phi_s) \approx 15$ mrad (40% relative uncertainty)



From: [Ref]

Measurement ingredients

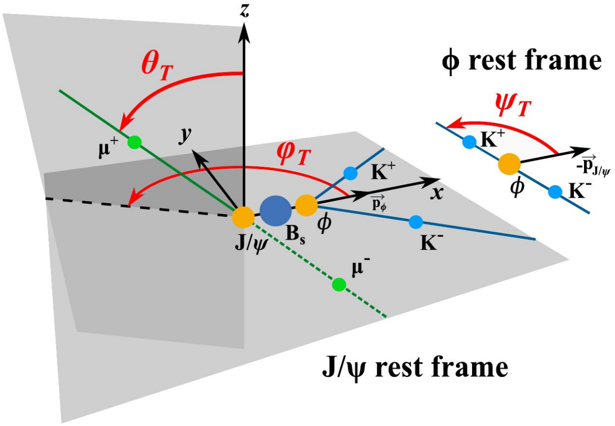
- **Time-dependent flavour asymmetry**



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

- **Essential ingredients**
  - Time-dependent **angular analysis** to separate the different CP eigenstate
  - Excellent **time resolution** and **flavour tagging** to see the B_s flavour oscillations (T ~ 350 fs)
  - Time and angular **efficiencies**

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

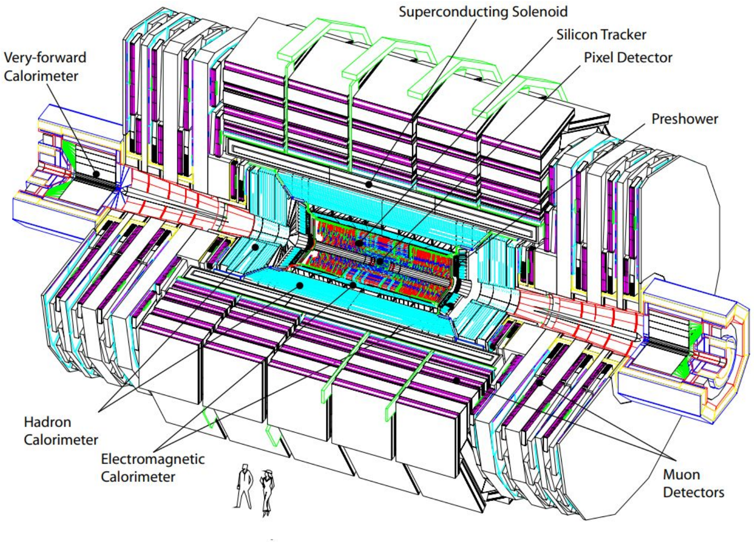


From: [PLB816(2021)136188]

Why a CMS measurement?

CMS is a general-purpose detector well suited for studying $B_s \rightarrow J/\psi\ \phi(1020) \rightarrow \mu^+\mu^-\ K^+K^-$

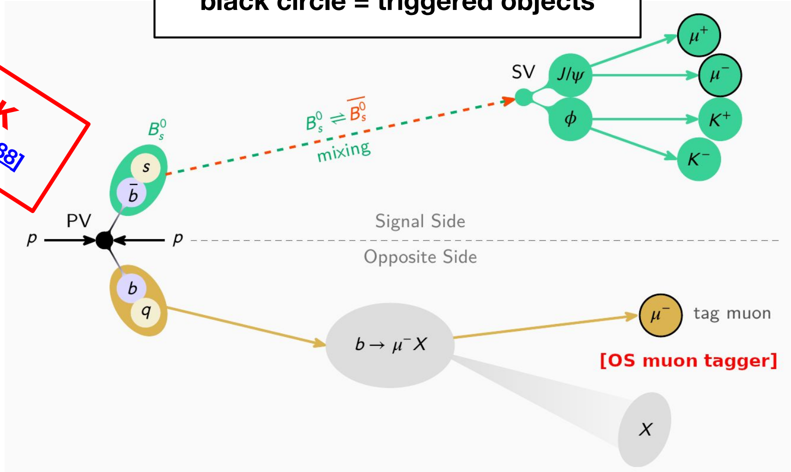
- **Silicon tracking system**
  - Excellent decay time resolution (σ_t ~ 60 fs)
  - Large pseudorapidity range up to $|\eta| = 2.5$
- **Superconducting solenoid**
  - High momentum resolution for charged tracks
- **Muon system**
  - High efficiency in triggering/reconstructing $J/\psi \rightarrow \mu^+\mu^-$
  - $\sigma(p_T)/p_T$ ~ O(1%)
- **Enormous amount of data** collected at $\sqrt{s}$ = 13 TeV
  - O(1M) of $B_s \rightarrow J/\psi\ \phi(1020) \rightarrow \mu^+\mu^-\ K^+K^-$ candidates



Trigger strategy for Run2

THIS TALK
[PLB816(2021)136188]

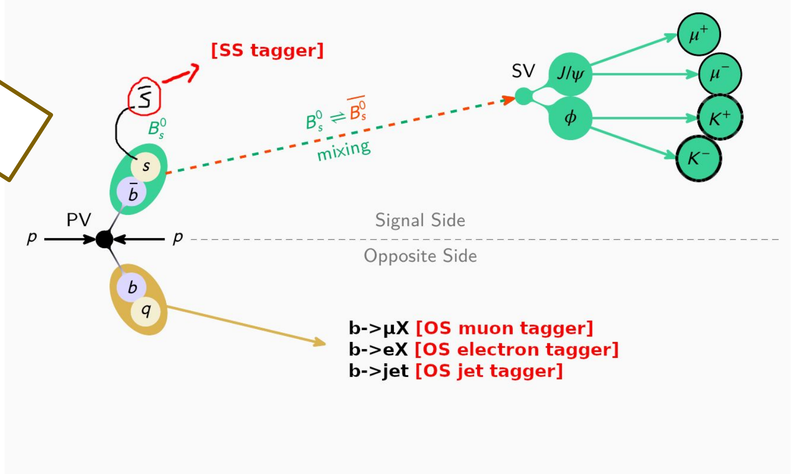
black circle = triggered objects



HLT_JPsiMuon

- J/ψ ➔ μμ candidate plus an additional muon (for tagging)
- Around 50k signal candidates in the 2017-2018 period
- “Easy” to work with (no displacement at trigger level)
- Tagging algorithms applied: OS-muon
  - P_{tag} ~ 10% (when evaluated in, and only in, this dataset)

FUTURE



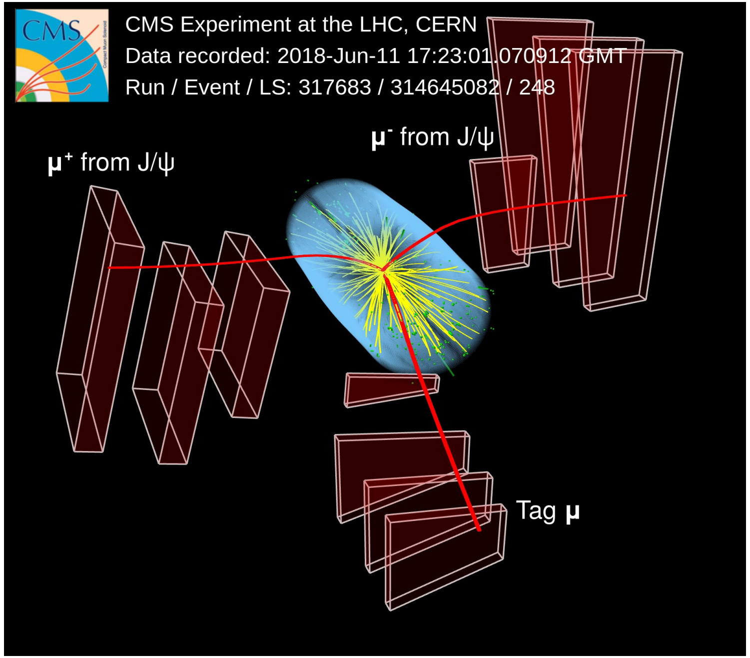
HLT_JPsiTrkTrkDisplaced

- Displaced J/ψ ➔ $\mu^+\mu^-$ candidate + two charged tracks near the ϕ(1020) resonance
- Higher statistics than *HLT_JPsiMuon* (x 8~10)
- ⚠️ Displaced (lifetime turn-on efficiency to model)
- Possible tagging strategies: OS-muon, OS-electron, OS-jet, Same Side
  - P_{tag} ~ ??

Latest CMS results overview

- **Reference:** Phys. Lett. B 816 (2021) 136188
- **Dataset**: 2017-2018 (L_{int} = 96 fb^{-1})
- **Trigger**: $J/\psi \rightarrow \mu^+\mu^-$ candidate plus an additional muon
- **Decay length cut**: >70 μm (to reduce prompt bkg.)
- **$m(K^+K^-)$ interval**: $m(\phi(1020)) \pm 10$ MeV
- **Number of signal candidates**: 48500 ± 250
- **Flavour tagging**: opposite-side muon
  - $\varepsilon_{tag} \approx 50\%$, $D_{tag} \approx 0.2$, $P_{tag} \approx 10\%$

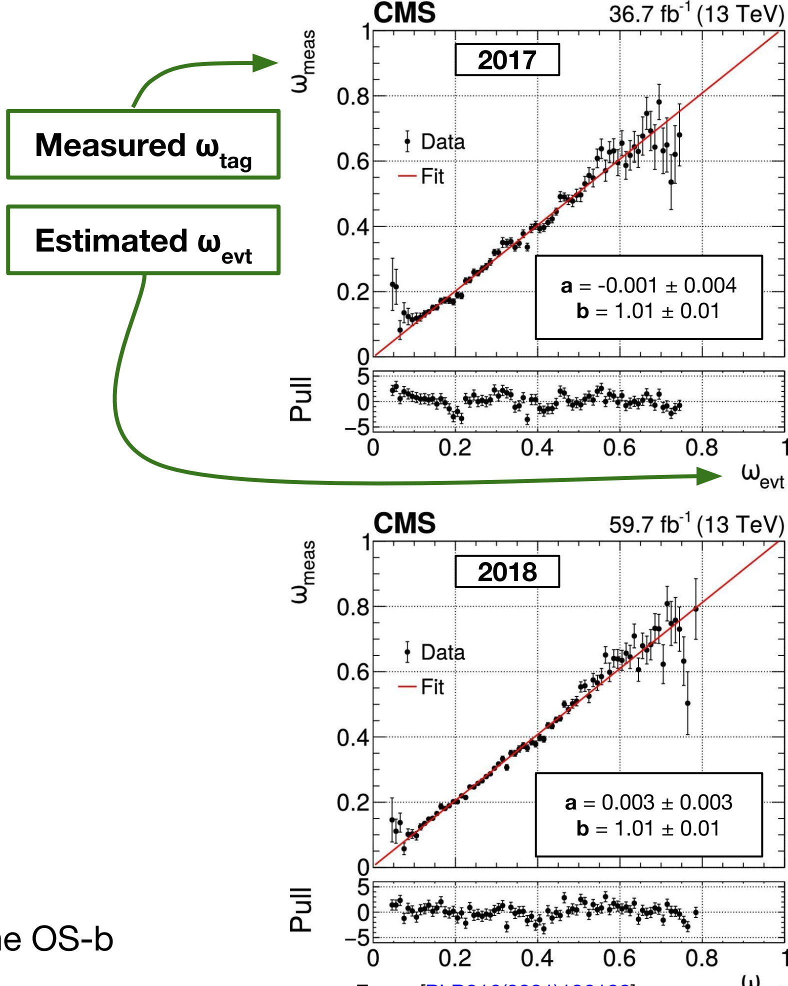
Fit: unbinned multidimensional extended maximum-likelihood
- **Input observables:** m_{Bs}, ct, σ_{ct}, θ_T, ψ_T, φ_T, ξ_{tag}, ω_{tag}
- **Fitted parameters**
  - *CPV observables*: ϕ_s, $|\lambda|$
  - *B_s system properties:* $\Delta\Gamma_s$, Γ_s, Δm_s
  - *Decay polarization*: $|A_0|^2$, $|A_\perp|^2$, $|A_S|^2$, $\delta_{//}$, $\delta_\perp$, $\delta_{S\perp}$
- **Bkg sources:** combinatorial, $B^0 \rightarrow J/\psi K^{*0} \rightarrow \mu^+\mu^- K^+\pi^-$



OS-muon tagging

1. **OS-muon selection** (very loose)
   - $p_T > 2$ GeV, $|\eta| < 2.5$, $IP_z(\mu, PV) < 1$ cm, $\Delta R_{\eta,\phi}(B_s) > 0.4$
2. **Tagging decision** (assuming $b \rightarrow \mu^- X$)
   - $\mu^- \rightarrow$ OS b $\rightarrow$ signal $\bar{b}$ (B_s)
   - $\mu^+ \rightarrow$ OS $\bar{b}$ $\rightarrow$ signal b ($\bar{B}_s$)
3. **Mistag probability evaluation**
   - Calibrated DNN trained on B_s MC and fine-tuned on self-tagging $B^+ \rightarrow J/\psi K^+$ data
   - Trained to discriminate *right* tags from *wrong* ones
   - **The output score s_{DNN} can be interpreted as a probability with the DNN trained to reproduce**

$$\textbf{Prob(right tag)} = s_{DNN} = 1 - \omega_{evt}$$



From: [PLB816(2021)136188]

Dilution sources: fakes, pileup, cascade decays, mixing of the OS-b

Systematic uncertainties

	ϕ_s [mrad]	$\Delta\Gamma_s$ [ps^{-1}]	Δm_s [$\hbar$ ps^{-1}]	$\lvert\lambda\rvert$	Γ_s [ps^{-1}]	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$ [rad]	$\delta_\perp$ [rad]	$\delta_{S\perp}$ [rad]
Statistical uncertainty	50	0.014	0.10	0.026	0.0042	0.0047	0.0063	0.0077	0.12	0.16	0.083
Model bias	7.9	0.0019	–	0.0035	0.0005	0.0002	0.0012	0.001	0.020	0.016	0.006
Model assumptions	–	–	–	0.0046	0.0003	–	0.0013	0.001	0.017	0.019	0.011
Angular efficiency	3.8	0.0006	0.007	0.0057	0.0002	0.0008	0.0010	0.002	0.006	0.015	0.015
Proper decay length efficiency	0.3	0.0062	0.001	0.0002	0.0022	0.0014	0.0023	0.001	0.001	0.002	0.002
Proper decay length resolution	3.5	0.0009	0.021	0.0015	0.0006	0.0007	0.0009	0.007	0.006	0.025	0.022
Data/simulation difference	0.6	0.0008	0.004	0.0003	0.0003	0.0044	0.0029	0.007	0.007	0.007	0.028
Flavor tagging	0.5	$<10^{-4}$	0.006	0.0002	$<10^{-4}$	0.0003	$<10^{-4}$	$<10^{-3}$	0.001	0.007	0.001
Sig./bkg. ω_{evt} difference	3.0	–	–	–	0.0005	–	0.0008	–	–	–	0.006
Peaking background	0.3	0.0008	0.011	$<10^{-4}$	0.0002	0.0005	0.0002	0.003	0.005	0.007	0.011
S-P wave interference	–	0.0010	0.019	–	0.0005	0.0005	–	0.013	–	0.019	0.019
$P(\sigma_{ct})$ uncertainty	$<10^{-1}$	0.0019	0.028	0.0004	0.0008	0.0006	0.0008	0.001	0.001	0.002	0.005
Total systematic uncertainty	10.0	0.0070	0.032	0.0083	0.0026	0.0049	0.0045	0.016	0.028	0.045	0.048

STAT LIMITED

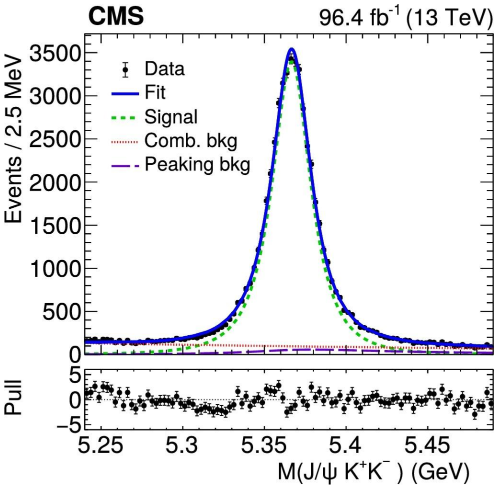
Leading systematic uncertainties

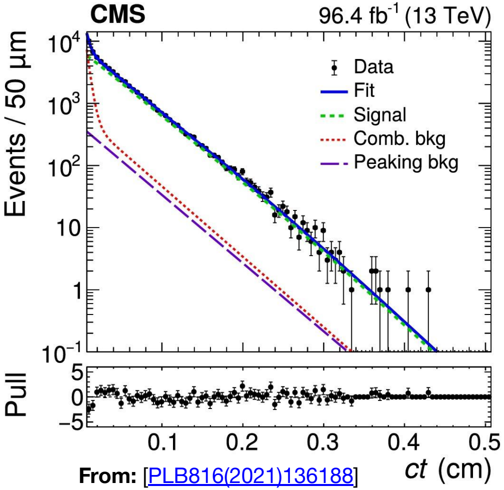
- ϕ_s ➜ model bias
- $\Delta\Gamma_s$ and Γ_s ➜ lifetime efficiency
- Δm_s ➜ lifetime uncertainty
- $|\lambda|$ ➜ angular efficiency

Results

Parameter	Fit value	Stat. uncer.	Syst. uncer.
ϕ_s [mrad]	-11	± 50	± 10
$\Delta\Gamma_s$ [ps^{-1}]	0.114	± 0.014	± 0.007
Δm_s [$\hbar$ ps^{-1}]	17.51	$^{+0.10}_{-0.09}$	± 0.03
$\lvert\lambda\rvert$	0.972	± 0.026	± 0.008
Γ_s [ps^{-1}]	0.6531	± 0.0042	± 0.0026
$\lvert A_0\rvert^2$	0.5350	± 0.0047	± 0.0049
$\lvert A_\perp\rvert^2$	0.2337	± 0.0063	± 0.0045
$\lvert A_S\rvert^2$	0.022	$^{+0.008}_{-0.007}$	± 0.016
$\delta_\parallel$ [rad]	3.18	± 0.12	± 0.03
$\delta_\perp$ [rad]	2.77	± 0.16	± 0.05
$\delta_{S\perp}$ [rad]	0.221	$^{+0.083}_{-0.070}$	± 0.048

- **Good agreement with SM predictions**
  - ϕ_s^{SM} = -37 ± 1 mrad [CKMfitter, UTfit]
  - $\Delta\Gamma_s^{SM}$ = 0.091 ± 0.013 ps^{-1} [Lenz & Tetlalmatzi-Xolocotzi]
  - $|\lambda|^{SM}$ = 1 (no direct CPV)
  - Δm_s^{SM} =18.77 ± 0.86 ħps^{-1} [Lenz & Tetlalmatzi-Xolocotzi]
- **First measurement by CMS of Δm_s and $|\lambda|$**

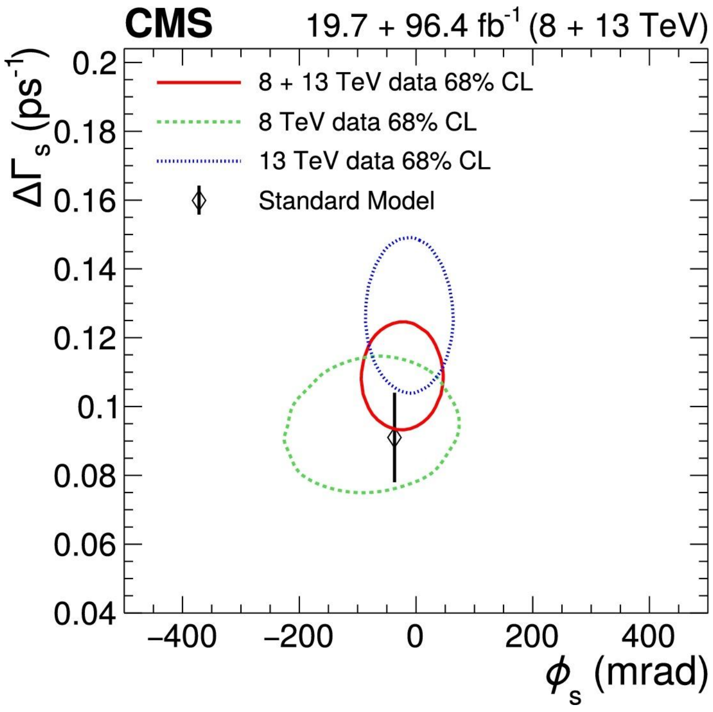




From: [PLB816(2021)136188]

Combination with Run1

- **The results of this analysis are combined with the ones obtained by CMS at $\sqrt{s}$ = 8 TeV[a]**
    - $\phi_s = -21 \pm 44\ (\text{stat}) \pm 10\ (\text{syst})$ mrad
    - $\Delta\Gamma_s = 0.1032 \pm 0.0095\ (\text{stat}) \pm 0.0048\ (\text{syst})\ \text{ps}^{-1}$
- **Results in agreement with the SM predictions**
- The new trigger strategy, which trades the number of events for tagging power, pays off for ϕ_s while does not improve $\Delta\Gamma_s$, which sensitivity is driven mainly by statistics



From: [PLB816(2021)136188]

[a] [PLB757(2016)9]

Future prospects: precision measurement

CMS is currently working on a **precision** **measurement of ϕ_s with the Run2 dataset by using all available triggers** (see SL7)

- **Statistics:** expected to increase the number of signal candidates by a factor of **8~10**
- **Flavour tagging**: muon, electron, jet and same-side (first implementation without hadronic PID)
  - Large enhancement of the effective statistics N_{Bs} x P_{tag}
- **Methodology:** various refinements to deal with the peculiarities of the new dataset
  - Efficiency modelization, background estimation, lifetime resolution, simulation corrections, …
  - *Not just a simple statistical scaling!*

- **Large improvements are expected for all physics parameters**
  - Reminder: *sensitivity*$(\phi_s, \Delta m_s) \propto \sqrt{(P_{tag} N_{Bs})}$ and *sensitivity*$(\Delta\Gamma_s, \Gamma_s) \propto \sqrt{(N_{Bs})}$
- This measurement will be the benchmark of several new analysis techniques, laying the foundations for future CMS works in the field CP violation

Conclusions

- **The CPV phase ϕ_s and the decay width difference $\Delta\Gamma_s$ have been measured using 48 500 $B_s \rightarrow J/\psi\ \phi(1020)$ signal candidates collected at $\sqrt{s}$ = 13 TeV, corresponding to L_{int} = 96.4 fb^{-1}**
- Events are selected using a trigger that requires an additional muon, which is exploited to infer the flavour of the B_s meson at production time, achieving $P_{tag} \approx 10\%$ with small associated systematic uncertainties
- Results from this measurement are combined with those obtained at $\sqrt{s}$ = 8 TeV, yielding

$$\phi_s = -21 \pm 44\ (\text{stat}) \pm 10\ (\text{syst})\ \text{mrad}$$

$$\Delta\Gamma_s = 0.1032 \pm 0.0095\ (\text{stat}) \pm 0.0048\ (\text{syst})\ \text{ps}^{-1}$$

- **Results are found to be consistent with the Standard Model predictions, allowing to further constrain possible contributions from new physics in the B_s meson decay and mixing**
- With the increase in statistics and the development of new techniques, the future for ϕ_s at LHC looks promising and challenging
- **CMS is actively working to release an update of the Run2 measurement, adding new data sets and tagging strategies**
  - **Stay tuned in the next conference seasons!**

Thanks for the attention

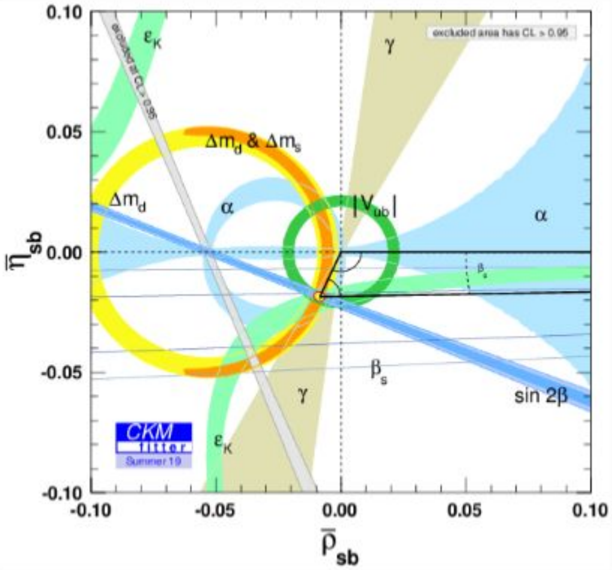
Backup

Unitary triangles

- The **unitary** condition of the CKM matrix leads to the following set of constrains

$$\sum_i |V_{ik}|^2 = \sum_k |V_{ik}|^2 = 1 \Longrightarrow \text{weak universality}$$

$$\sum_i V_{ij}V_{ik}^* = 0 \qquad \sum_k V_{ik}V_{jk}^* = 0 \Longrightarrow \text{six triangles in the complex plane ("unitary triangles")}$$



- Of particular interest for this work is the so-called "B_s^0 unitary triangle":

$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0$$

with angles: $\alpha_s = \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{us}V_{ub}^*}\right), \quad \beta_s = \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right), \quad \gamma_s = \arg\left(-\frac{V_{us}V_{ub}^*}{V_{cs}V_{cb}^*}\right).$

B_s meson mixing

- B_s^0 mesons are subject to **flavour mixing**, that is oscillations between their C-conjugate states before decay
- The light and heavy mass eigenstates are described by a **superposition** of flavour states, as

$$\left|B_s^{L,H}\right\rangle = p\left|B_s^0\right\rangle \pm q\left|\overline{B}_s^0\right\rangle \quad \text{with } |q|^2 + |p|^2 = 1$$

- The B_s^0 system is characterized by the parameters

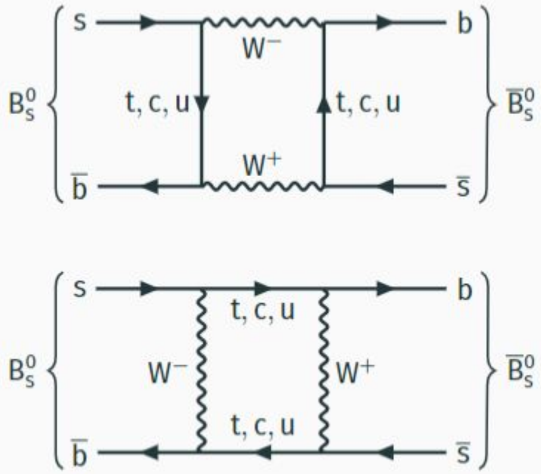
$$m_s \equiv \frac{m_H + m_L}{2}, \quad \Gamma_s \equiv \frac{\Gamma_H + \Gamma_L}{2}$$
$$\Delta m_s \equiv m_H - m_L, \quad \Delta\Gamma_s \equiv \Gamma_L - \Gamma_H$$

- For the B_s^0 system $|q/p| \simeq 1$ is observed[1], so that **the ratio q/p can be expressed in terms of a complex phase:**

$$\frac{q}{p} \equiv e^{-i\phi_M} \simeq \frac{V_{ts}V_{tb}^*}{V_{ts}^*V_{tb}}$$

[1] World-average value: $|q/p| = 1.0003 \pm 0.0014$ [HFLAV]

LO diagrams for $B_s^0 \leftrightarrow \overline{B}_s^0$ mixing



- **The flavour eigenstates oscillate with a period of**

$$T = \frac{2\pi}{\Delta m_s} \sim 350 \text{ fs}$$

CPV in mesons

- **Observable CP violation is generated by interference between amplitudes**
- **Three different types of CP violation are possible**
  1. "Direct" CPV in **decays** — $\mathcal{P}(P \to f) \neq \mathcal{P}(\bar{P} \to \bar{f})$
     - Observed in kaons, B and D mesons[1]
  2. "Indirect" CPV in **mixing** — $\mathcal{P}(P^0 \to \bar{P}^0) \neq \mathcal{P}(\bar{P}^0 \to P^0)$
     - Observed in K^0 oscillations[2]
  3. CPV in the **interference** of decays and mixing — $\mathcal{P}(P^0 \to f) \neq \mathcal{P}(P^0 \to \bar{P}^0 \to f)$
     - Observed in K^0 and B^0 mesons[3]
- Defining A_f as the $P \to f$ amplitude, **CPV information is encoded in the rephasing-invariant complex parameter λ:**

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

CPV in $B_s \to J/\psi\ \phi(1020) \to \mu^+\mu^- K^+K^-$

- **No direct CPV:** $\left|\frac{\bar{A}_{\bar{f}}}{A_f}\right| = \left|\frac{V_{cb}V_{cs}^*}{V_{cb}^*V_{cs}}\right| = 1$
- **No CPV in mixing:** $\left|\frac{q}{p}\right| = \left|\frac{V_{tb}^*V_{ts}}{V_{tb}V_{ts}^*}\right| \simeq 1$
- **CPV in the interference** (neglecting penguin contributions)[1]:

$$\lambda = \eta_f \underbrace{\left(\frac{V_{cb}V_{cs}^*}{V_{cb}^*V_{cs}}\right)}_{=\bar{A}_f/A_f} \underbrace{\left(\frac{V_{tb}^*V_{ts}}{V_{tb}V_{ts}^*}\right)}_{=q/p} \overset{|\lambda|=1}{=} \eta_f\, e^{-i\phi_s}$$

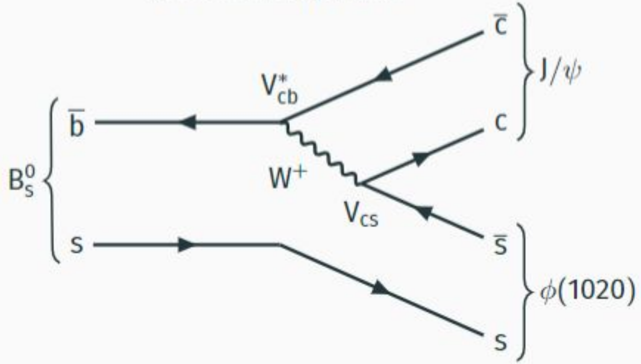
$$\boxed{\phi_s = -2\arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right) = -2\beta_s}$$

where η_f is the CP eigenvalue of the final state and β_s is one of the angles of the B_s^0 unitary triangle

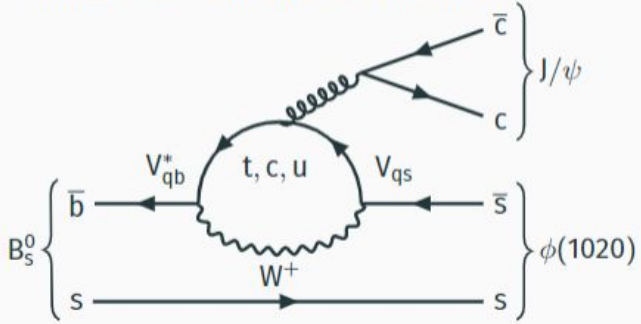
- β_s can be determined very precisely with global CKM fits

[1] Penguin transitions are predicted to change the value of ϕ_s by about $\sim$1 mrad, almost two order of magnitudes smaller than the current experimental sensitivity ($\sim$30–50 mrad)

Tree-level amplitude

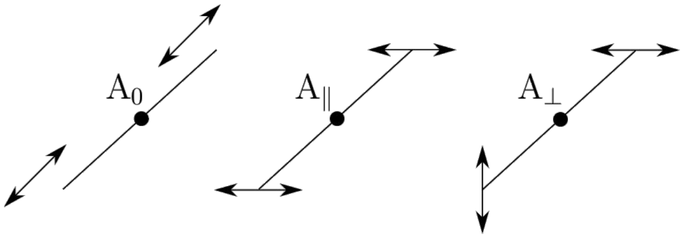


Higher-order loop amplitude ("penguin")



The $B_s \to J/\psi\ \phi(1020) \to \mu^+\mu^-\ K^+K^-$ decay

- **$B_s \to J/\psi\ \phi(1020) \to \mu^+\mu^-\ K^+K^-$ is the golden channel for measuring ϕ_s**
  1. The final state can be reconstructed with high S/B ratio
  2. $J/\psi \to \mu^+\mu^-$ is easy to trigger
  3. Only one CP-violating phase (“golden mode”) if neglecting penguin contributions
  4. SM predicts no direct CPV
- **The final state is a mixture of CP-even and CP-odd eigenstates**
  - Spin-0 pseudo-scalar meson (B_s) decaying into two spin-1 vector mesons ($J/\psi\ \phi(1020)$)
  - The CP eigenvalue of the final state depends on the value of the orbital momentum, as $\eta_f = (-1)^l$
- **The $B_s \to J/\psi\ \phi(1020) \to \mu^+\mu^-\ K^+K^-$ decay amplitude can be decomposed into three polarization states**
  - A_0 : $l = 0 \to$ CP-even
  - $A_\perp$: $l = 1 \to$ CP-odd
  - $A_\parallel$: $l = 2 \to$ CP-even



- **Additional contribution (“S-wave”) from non-resonant $B_s \to J/\psi\ K^+K^-$ and $B_s \to J/\psi\ f^0(980)$ is assumed**
  - A_S : $l = 0 \to$ CP-odd

Decay rate model

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

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

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	$1/\sqrt{2}\sin 2\psi_T\sin^2\theta_T\sin 2\varphi_T$	$\lvert A_0\rvert\lvert A_\parallel\rvert$	$\cos(\delta_\parallel-\delta_0)$	$D\cos(\delta_\parallel-\delta_0)$	$C\cos(\delta_\parallel-\delta_0)$	$-S\cos(\delta_\parallel-\delta_0)$
6	$1/\sqrt{2}\sin 2\psi_T\sin 2\theta_T\cos\varphi_T$	$\lvert A_0\rvert\lvert A_\perp\rvert$	$C\sin(\delta_\perp-\delta_0)$	$S\cos(\delta_\perp-\delta_0)$	$\sin(\delta_\perp-\delta_0)$	$D\cos(\delta_\perp-\delta_0)$
7	$2/3(1-\sin^2\theta_T\cos^2\varphi_T)$	$\lvert A_S\rvert^2$	1	−D	C	S
8	$1/3\sqrt{6}\sin\psi_T\sin^2\theta_T\sin 2\varphi_T$	$\lvert A_S\rvert\lvert A_\parallel\rvert$	$C\cos(\delta_\parallel-\delta_S)$	$S\sin(\delta_\parallel-\delta_S)$	$\cos(\delta_\parallel-\delta_S)$	$D\sin(\delta_\parallel-\delta_S)$
9	$1/3\sqrt{6}\sin\psi_T\sin 2\theta_T\cos\varphi_T$	$\lvert A_S\rvert\lvert A_\perp\rvert$	$\sin(\delta_\perp-\delta_S)$	$-D\sin(\delta_\perp-\delta_S)$	$C\sin(\delta_\perp-\delta_S)$	$S\sin(\delta_\perp-\delta_S)$
10	$4/3\sqrt{3}\cos\psi_T(1-\sin^2\theta_T\cos^2\varphi_T)$	$\lvert A_S\rvert\lvert A_0\rvert$	$C\cos(\delta_0-\delta_S)$	$S\sin(\delta_0-\delta_S)$	$\cos(\delta_0-\delta_S)$	$D\sin(\delta_0-\delta_S)$

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

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

$D = -\frac{2|\lambda|\cos\phi_s}{1+|\lambda|^2}$ ⇒ **Sensitive to $\phi_s \sim \pi/2$**

$\xi_{tag} = \pm 1$ ⇒ **flavour tagging decision**
$\omega_{tag} \in [0, 1]$ ⇒ **mistag probability**
$|\lambda| = |q/p|\left|\bar{A}_f/A_f\right|$ ⇒ **direct/mixing CPV if ≠1**

Angular efficiency

- **Detector acceptance and event selection lead to non-uniform angular efficiency**
- 3D angular efficiency functions are evaluated in bins of $\cos\theta_T$, $\cos\psi_T$ and φ_T, separately for 2017 and 2018, using simulated samples
- The efficiency functions are obtained with a projection of the 3D angular efficiency histograms on an orthogonal basis:
  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**

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

     - 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

DNNs as probability estimators

- **A DNN can be (naturally) engineered to predict the probability of a given input example to belong to one of the classes for which the network has been trained**
  - e. g.: what is the probability that the flavour inference for a given event is correct?
- To this end, the **softmax activation function** can be used in the last layer to normalize the output of the network to a probability distribution consisting of K probabilities in the interval [0, 1] that add up to 1:[1]

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (i = 1, \ldots, K)$$

  - A DNN is called "**calibrated**" if the output probability of the predicted class reflects its true posterior probability (the softmax function does not ensure calibrated networks)
- To improve probability calibration, the **cross-entropy loss function** can be used in the training process

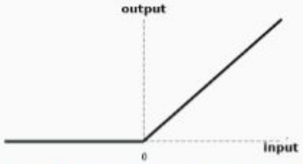
$$\mathcal{L}_{CE} = -\sum_{i=1}^{K} t_i \log(s_i) \quad \text{where} \quad \begin{aligned} s_i &= \text{network scores} \\ t_i &= \text{one-hot encoded truth labels}^2 \end{aligned}$$

  - $\mathcal{L}_{CE}$ can be interpreted as the negative log-likelihood for the conditional probability $P(\vec{t}\,|\,\vec{s})$

[1] Based on the Luce's choice axiom. Ref: R. D. Luce, "Individual choice behavior: a theoretical analysis", Wiley, New York (1959)
[2] Every entry of $\vec{t}$ is equal to 0 but the one corresponding to the true class, which is equal to 1

Deep Neural Network for flavour tagging

- **Training features**
  - Muon variables
    - p_T, η, IP_{xy}, $\sigma_{IP_{xy}}$, IP_z, σ_{IP_z}, $\Delta R(B^0_s)$, ...
  - Surrounding activity variables ("muon cone")
    - Constructed from tracks around the muon direction
    - Iso_μ, Q_{cone}, $p_{T,cone}$, E_μ / E_{cone}, ...
- **Architecture: fully connected**
  - 3 layers of 200 neurons
  - Rectified Linear Unit activation: $f(x) = \max(0, x)$
  - 40% dropout probability
    - Probability of temporarily removing a neuron in each training iteration, used to reduce overtraining



- **Output:** softmax
- **Loss**: binary cross-entropy
- **Optimizer**: Adam
  - Adaptive optimization algorithm specifically designed for training DNNs

Fit model

$$P = \frac{N_{sgn}}{N_{tot}} P_{sgn} + \frac{N_{bkg}}{N_{tot}} P_{bkg}$$

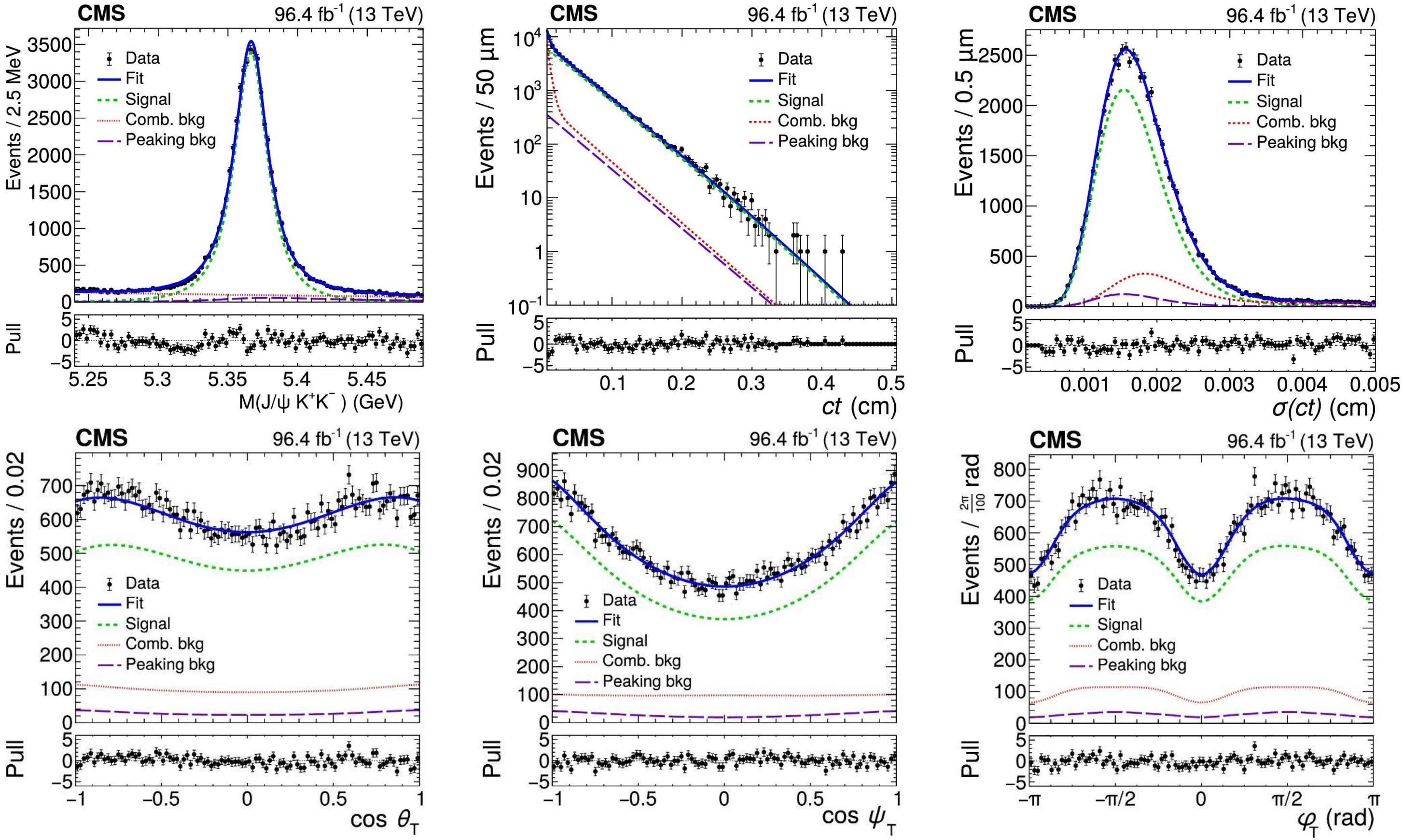
$$P_{sgn} = \epsilon(ct)\,\epsilon(\Theta)\;[f(\Theta, ct\,|\,\alpha) \otimes G(ct, \sigma_{ct})]\;P_{sgn}(m_{B^0_s})\,P_{sgn}(\sigma_{ct})\,P_{sgn}(\xi_{tag})$$

- $\epsilon(ct)\,\epsilon(\Theta)$: efficiency functions
- $f(\Theta, ct\,|\,\alpha)$: differential decay rate pdf
- $G(ct, \sigma_{ct})$: Gaussian resolution function
- $P_{sgn}(m_{B^0_s})$: mass pdf
- $P_{sgn}(\sigma_{ct})$: proper decay length uncertainty pdf
- $P_{sgn}(\xi_{tag})$: tag decision pdf
  - $\Theta = (\cos\theta_T, \cos\psi_T, \varphi_T)$
  - $\alpha = (\phi_s, \Gamma_s, \Delta\Gamma_s, \Delta m_s, |\lambda|, A_0, A_\perp, A_S, \delta_\parallel, \delta_\perp, \delta_{S\perp})$

$$P_{bkg} = P_{bkg}(\cos\theta_T, \varphi_T)\,P_{bkg}(\cos\psi_T)\,P_{bkg}(ct)\,P_{bkg}(m_{B^0_s})\,P_{bkg}(\sigma_{ct})\,P_{bkg}(\xi_{tag})$$

- $P_{bkg}(\cos\theta_T, \varphi_T)$, $P_{bkg}(\cos\psi_T)$, $P_{bkg}(ct)$: background angular and proper decay length pdfs
- P_{bkg} contains a dedicated term to model the **peaking background** from $B^0 \to J/\psi\, K^*(892)^0 \to \mu^+\mu^-\, K^+\pi^-$ where the pion is misidentified as a kaon
  - The peaking background from $\Lambda^0_b \to J/\psi\, K^- p \to \mu^+\mu^-\, K^- p$ is estimated to be negligible

Fit results


CMS
96.4 fb-1 (13 TeV)
Events / 2.5 MeV
Data
Fit
Signal
Comb. bkg
Peaking bkg
Pull
M(J/ψ K+K−) (GeV)
Events / 50 μm
ct (cm)
Events / 0.5 μm
σ(ct) (cm)
Events / 0.02
cos θT
cos ψT
Events / 2π/100 rad
φT (rad)

Correlations in the 13 TeV results

Table 1: Statistical correlation matrix between the physics parameters as obtained from the ML fit to the 13 TeV data.

	ϕ_s	$\Delta\Gamma_s$	Δm_s	$\|\lambda\|$	Γ_s	$\|A_0\|^2$	$\|A_\perp\|^2$	$\|A_S\|^2$	$\delta_\parallel$	$\delta_\perp$	$\delta_{S\perp}$
ϕ_s	+1.00	−0.02	−0.19	+0.22	0.00	−0.01	+0.01	−0.01	−0.02	−0.09	+0.03
$\Delta\Gamma_s$		+1.00	−0.02	0.00	−0.48	+0.63	−0.71	0.00	+0.01	−0.01	−0.04
Δm_s			+1.00	−0.14	+0.03	−0.01	+0.02	+0.03	+0.01	+0.68	−0.05
$\|\lambda\|$				+1.00	−0.02	0.00	−0.01	−0.03	−0.06	−0.18	+0.05
Γ_s					+1.00	−0.31	+0.42	+0.15	−0.02	+0.02	−0.05
$\|A_0\|^2$						+1.00	−0.61	+0.15	−0.01	−0.01	−0.09
$\|A_\perp\|^2$							+1.00	−0.11	−0.06	0.00	+0.06
$\|A_S\|^2$								+1.00	−0.07	+0.02	−0.44
$\delta_\parallel$									+1.00	+0.27	−0.02
$\delta_\perp$										+1.00	−0.10
$\delta_{S\perp}$											+1.00

Full combination results

CMS 8 TeV results

Parameter	Value	Stat.	Syst.
ϕ_s [mrad]	-75	± 97	± 31
$\Delta\Gamma_s$ [ps^{-1}]	0.095	± 0.013	± 0.007
Γ_s [ps^{-1}]	0.6704	± 0.0043	± 0.0055
$\lvert A_0\rvert^2$	0.510	± 0.005	± 0.011
$\lvert A_\perp\rvert^2$	0.243	± 0.008	± 0.012
$\lvert A_S\rvert^2$	0.012	± 0.009	± 0.022
$\delta_\parallel$ [rad]	3.48	± 0.09	± 0.68
$\delta_\perp$ [rad]	2.98	± 0.36	± 0.66
$\delta_{S\perp}$ [rad]	0.37	± 0.12	± 0.18

CMS 13 TeV results

Parameter	Value	Stat.	Syst.
ϕ_s [mrad]	-11	± 50	± 10
$\Delta\Gamma_s$ [ps^{-1}]	0.114	± 0.014	± 0.007
Δm_s [$\hbar$ps^{-1}]	17.51	± 0.10	± 0.03
$\lvert\lambda\rvert$	0.972	± 0.026	± 0.008
Γ_s [ps^{-1}]	0.6531	± 0.0042	± 0.0026
$\lvert A_0\rvert^2$	0.5350	± 0.0047	± 0.0049
$\lvert A_\perp\rvert^2$	0.2337	± 0.0063	± 0.0045
$\lvert A_S\rvert^2$	0.022	± 0.008	± 0.016
$\delta_\parallel$ [rad]	3.18	± 0.12	± 0.03
$\delta_\perp$ [rad]	2.77	± 0.16	± 0.05
$\delta_{S\perp}$ [rad]	0.221	± 0.083	± 0.048

CMS 8+13 TeV combined results

Parameter	Fit value	Stat.	Syst.
ϕ_s [mrad]	-21	± 44	± 10
$\Delta\Gamma_s$ [ps^{-1}]	0.1032	± 0.0095	± 0.0048
Γ_s [ps^{-1}]	0.6590	± 0.0032	± 0.0023
$\lvert A_0\rvert^2$	0.5289	± 0.0038	± 0.0041
$\lvert A_\perp\rvert^2$	0.2393	± 0.0050	± 0.0037
$\lvert A_S\rvert^2$	0.016	± 0.006	± 0.013
$\delta_\parallel$ [rad]	3.19	± 0.12	± 0.04
$\delta_\perp$ [rad]	2.78	± 0.15	± 0.06
$\delta_{S\perp}$ [rad]	0.238	± 0.078	± 0.046

Correlation in the combination

Table 3: Correlations between the physics parameters as obtained from the combination between the CMS 8 TeV and 13 TeV results. Correlations are both statistical and systematic.

	ϕ_s	$\Delta\Gamma_s$	Γ_s	$\lvert A_0\rvert^2$	$\lvert A_\perp\rvert^2$	$\lvert A_S\rvert^2$	$\delta_\parallel$	$\delta_\perp$	$\delta_{S\perp}$
ϕ_s	+1.00	+0.02	−0.03	+0.01	−0.01	+0.01	−0.01	−0.08	+0.03
$\Delta\Gamma_s$		+1.00	−0.45	+0.43	−0.57	+0.01	+0.01	0.00	−0.01
Γ_s			+1.00	−0.17	+0.30	+0.06	−0.03	0.00	−0.08
$\lvert A_0\rvert^2$				+1.00	−0.56	+0.25	−0.03	+0.01	−0.18
$\lvert A_\perp\rvert^2$					+1.00	−0.08	−0.03	+0.01	+0.14
$\lvert A_S\rvert^2$						+1.00	−0.02	+0.02	−0.20
$\delta_\parallel$							+1.00	+0.26	0.00
$\delta_\perp$								+1.00	−0.05
$\delta_{S\perp}$									+1.00