LHCb-ANA-20XX-YYY
May 8, 2017

Measurement of the CKM angle γ using $B_s^0 \to D_s K\pi\pi$ decays

P. d'Argent[1], E. Gersabeck[1], M. Kecke[1], M. Schiller[2]

[1] *Physikalisches Institut, Ruprecht-Karls-Universitt Heidelberg, Heidelberg, Germany*
[2] *School of Physics and Astronomy, University of Glasgow, Glasgow, United Kingdom*

Abstract

Guidelines for the preparation of LHCb documents are given. This is a "living" document, that should reflect our current practice. It is expected that these guidelines are implemented for papers already before they go into the first collaboration wide review. Please contact the Editorial Board chair if you have suggestions for modifications.

1 Introduction

The weak phase γ is the least well known angle of the CKM unitary triangle. A key channel to measure γ is the time-dependent analysis of $B_s^0 \to D_s K$ decays [1], [2].
The $B_s^0 \to D_s K\pi\pi$ proceeds at tree level via the transitions shown in Fig. 1 a) and b).

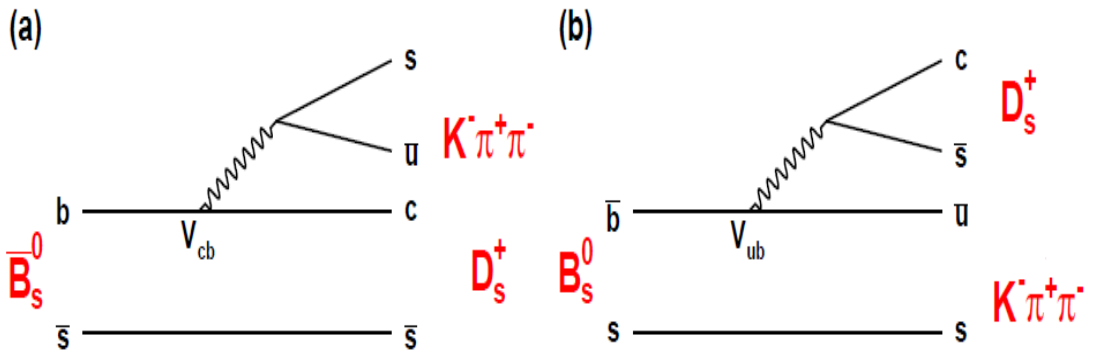

Figure 1: Feynman diagram of the $B_s^0 \to D_s K\pi\pi$ decay, proceeding via a) $b \to c$ transitions or b) $b \to u$ transitions.

To measure the weak CKM phase $\gamma \equiv arg[-(V_{ud}V_{ub}^*)/(V_{cd}V_{cb}^*)]$, a decay with interference between $b \to c$ and $b \to u$ transitions at tree level is needed [1]. As illustrated in Fig. 1, this is the case for the presented decay mode. A measurement of γ using $B_s^0 \to D_s K\pi\pi$ decays, where the $K\pi\pi$ subsystem is dominated by excited kaon states such as the $K_1(1270)$ and $K_1(1400)$ resonances, will succeed the branching ratio measurement presented in this note. It is complementary to the above mentioned analysis of $B_s^0 \to D_s K$, making use of a fully charged final state, where every track is detected in the vertex locator. To account for the non-constant strong phase across the Dalitz plot, one can either develop a time-dependent amplitude model or select a suitable phase-space region and introduce a coherence factor as additional hadronic parameter to the fit.
This analysis is based on the first observation of the $B_s^0 \to D_s K\pi\pi$ decay presented in [3] and [4], where its branching ratio is measured relative to $B_s^0 \to D_s\pi\pi\pi$. The result obtained by the previous analysis is $0.052 \pm 0.005\ 0.003$, where the uncertainties are statistical and systematical, respectively. The branching ratio measurement is updated, exploiting the full Run 1 data sample, corresponding to 3 fb^{-1} of integrated luminosity.

2 Selection

For the presented analysis, we reconstruct the $B_s^0 \to D_s K\pi\pi$ decay through three different final states of the D_s meson, $D_s \to KK\pi$, $D_s \to K\pi\pi$ and $D_s \to \pi\pi\pi$. Of those three final states $D_s \to KK\pi$ is the most prominent one, while $\mathcal{BR}(D_s \to \pi\pi\pi) \approx 0.2 \cdot \mathcal{BR}(D_s \to KK\pi)$ and $\mathcal{BR}(D_s \to K\pi\pi) \approx 0.1 \cdot \mathcal{BR}(D_s \to KK\pi)$ holds for the other two.
A two-fold approach is used to isolate the $B_s^0 \to D_s K\pi\pi$ candidates from data passing the stripping line. First, further one-dimensional cuts are applied to reduce the level of combinatorial background and to veto some specific physical background. This stage is

specific to the respective final state in which the D_s meson is reconstructed, since different physical backgrounds, depending on the respective final state, have to be taken into account. After that, a multivariate classifier is trained which combines the information of several input variables, including their correlation, into one powerful discriminator between signal and combinatorial background. For this stage, all possible D_s final states are treated equally.

2.1 Cut-based selection

In order to minimize the contribution of combinatorial background to our samples, we apply the following cuts to the b hadron:

- DIRA > 0.99994
- min IP $\chi^2 < 20$ to any PV,
- FD $\chi^2 > 100$ to any PV,
- Vertex $\chi^2/\text{nDoF} < 8$,
- $(Z_{D_s} - Z_{B_s^0}) > 0$, where Z_M is the z-component of the position $\vec{x}$ of the decay vertex for the B_s^0/D_s meson.

Additionally, we veto various physical backgrounds, which have either the same final state as our signal decay, or can contribute via a single misidentification of $K \to \pi$ or $K \to p$. In the following, the vetoes are ordered by the reconstructed D_s final state they apply to:

1. All:
   (a) $B_s^0 \to D_s^+ D_s^-$: $|M(K\pi\pi) - m_{D_s}| > 20$ MeV/c^2.
   (b) $B_s^0 \to {D_s}^- K^+ K^- \pi^+$: possible with single missID of $K^- \to \pi^-$, rejected by requiring π^- to fulfill $\text{DLL}_{K\pi} < 5$.
2. $D_s \to KK\pi$
   (a) $B^0 \to D^+(\to K^+\pi^-\pi^+)K\pi\pi$: possible with single missID of $\pi^+ \to K^+$, vetoed by changing particle hypothesis and recompute $|M(K^+\pi^-\pi^+) - m_{Dp}| > 30$ MeV/c^2, or the K^+ has to fulfill $\text{DLL}_{K\pi} > 10$.
   (b) $\Lambda_b^0 \to \Lambda_c^+(\to pK^-\pi^+)K\pi\pi$: possible with single missID of $p \to K^+$, vetoed by changing particle hypothesis and recompute $M(pK^-\pi^+) - m_{\Lambda_c^+} > 30$ MeV/c^2, or the K^+ has to fulfill ($\text{DLL}_{K\pi}$ - $\text{DLL}_{p\pi}$) > 5.
   (c) $D^0 \to KK$: D^0 combined with a random π can fake a $D_s \to KK\pi$ decay and be a background to our signal, vetoed by requiring $M(KK) < 1840\,\text{MeV}/c^2$.
3. $D_s \to K\pi\pi$

(a) $D^0 \to \pi^+K^-$: D^0 combined with a random π^- can fake a ${D_s}^- \to K^-\pi^+\pi^-$ decay and be a background to our signal, vetoed by requiring $M(\pi^+K^-) < 1750\,\mathrm{MeV}/c^2$.

(b) $\Lambda_b^0 \to \Lambda_c^+(\to p\pi^-\pi^+)K\pi\pi$: possible with single missID of $p \to K^+$, vetoed by changing particle hypothesis and recompute $M(p\pi^-\pi^+) - m_{\Lambda_c^+} > 30$ MeV/c^2, or the K^+ has to fulfill ($\mathrm{DLL}_{K\pi}$ - $\mathrm{DLL}_{p\pi}$) > 5.

4. $D_s \to \pi\pi\pi$

(a) $D^0 \to \pi\pi$: combined with a random π can fake a $D_s \to \pi\pi\pi$ decay and be a background to our signal, vetoed by requiring both possible combinations to have $M(\pi\pi) < 1700\,\mathrm{MeV}/c^2$.

The most prominent final state used in this analysis is $B_s^0 \to D_s(\to KK\pi)K\pi\pi$, where the D_s decay can either proceed via the narrow ϕ resonance, the broader K^{*0} resonance, or non resonant. Depending on the decay process being resonant or not, we apply additional PID requirements on this final state:

- resonant case:
  - $D_s^+ \to \phi\pi^+$, with $|M(K^+K^-) - m_\phi| < 20$ MeV/c^2 : no additional requirements, since ϕ is narrow and almost pure K^+K^-.
  - $D_s^+ \to \overline{K}^{*0}K^+$, with $|M(K^-\pi^+) - m_{K^{*0}}| < 75$ MeV/c^2 : $\mathrm{DLL}_{K\pi} > 0$ for kaons, since this resonance is more than ten times broader than ϕ.
- non resonant case: $\mathrm{DLL}_{K\pi} > 5$ for kaons, since the non resonant category has significant charmless contributions.

For the other two final states, we apply global PID requirements:

- $D_s \to K\pi\pi$
  - $\mathrm{DLL}_{K\pi} > 10$ for kaons, since we expect significantly higher charmless background contribution in this channel.
  - $\mathrm{DLL}_{K\pi} < 5$ for pions.
- $D_s \to \pi\pi\pi$
  - $\mathrm{DLL}_{K\pi} < 10$ for all pions.
  - $\mathrm{DLL}_{p\pi} < 10$ for all pions.

2.2 Multivariate stage

We use TMVA [5] to train a multivariate discriminator, which is used to further improve the signal to background ratio. The 17 variables used for the training are:

- max(ghostProb) over all tracks
- cone(p_{T}) asymmetry of every track, which is defined to be the difference between the p_{T} of the π/K and the sum of all other p_{T} in a cone of radius $r = \sqrt{(\Delta\Phi)^2 + (\Delta\eta)^2}$ < 1 rad around the signal π/K track.
- min(IPχ^2) over the X_s daughters
- max(DOCA) over all pairs of X_s daughters
- min(IPχ^2) over the D_s daughters
- D_s and B_s^0 DIRA
- D_s FD significance
- max(cos($D_s h_i$)), where cos($D_s h_i$) is the cosine of the angle between the D_s and another track i in the plane transverse to the beam
- B_s^0 IPχ^2, FDχ^2 and Vertex χ^2

Various classifiers were investigated in order to select the best performing discriminator. Consequently, a boosted decision tree with gradient boost (BDTG) is chosen as nominal classifier. We use truth-matched MC as signal input. Simulated signal candidates are required to pass the same trigger, stripping and preselection requirements, that were used to select the data samples. For the background we use events from the high mass sideband ($m_{B_s^0 candidate} > 5600$ MeV/c^2) of our data samples. As shown in Fig. 2, this mass region is sufficiently far away from signal structures and is expected to be dominantly composed of combinatorial background.

The distributions of the input variables for signal and background are shown in Fig. 3.

The relative importance of the input variables for the BDTG training is summarized in Table 2.

The BDTG output distribution for test and training samples is shown in Fig 4. No sign of overtraining is observed.

We determine the optimal cut value by maximizing the figure of merit $S/\sqrt{S+B}$ where S is the signal yield and B the background yield in the signal region, defined to be within ± 50 MeV/c^2 of the nominal B_s^0 mass. To avoid a bias in the determination of the branching fraction, we determine S and B using our normalization channel. All trigger, stripping and additional selection criteria described in this and the previous chapter are applied to the $B_s^0 \to D_s\pi\pi\pi$ data samples. After that, we perform a simplified version of the fit to the invariant mass distribution of $B_s^0 \to D_s\pi\pi\pi$ candidates described in Sec. **??**. Here, a Gaussian function to model the signal and an exponential function to model combinatorial background is used. From this fit we estimate the number of signal events in our normalization channel. Multiplying that number with the PDG branching fraction

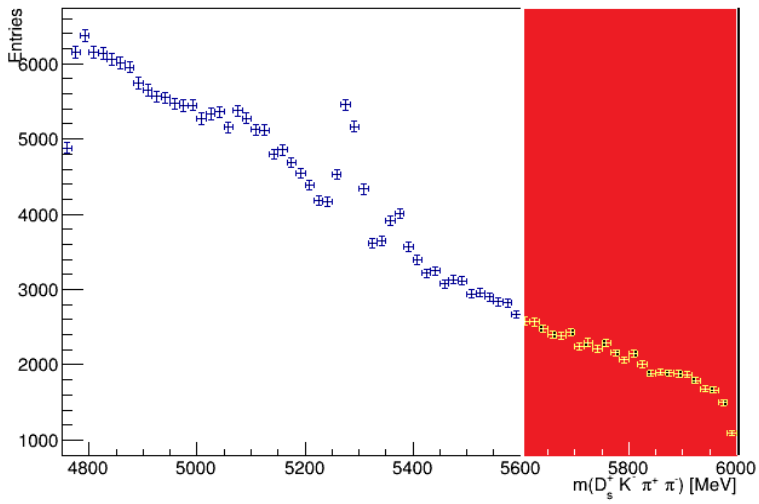

Figure 2: Invariant mass distribution of preselected $B_s^0 \to D_s K\pi\pi$ candidates. The red coloured region with $m_{B_s^0 candidate} > 5600$ MeV/c^2 is used as background input for the boosted decision tree.

Variable	relative importance [%]
pi_minus_ptasy_1.00	7.32
log_Ds_FDCHI2_ORIVX	7.23
K_plus_ptasy_1.00	7.17
log_Ds_DIRA	6.96
Bs_ENDVERTEX_CHI2	6.82
max_ghostProb	6.76
pi_plus_ptasy_1.00	6.57
log_DsDaughters_min_IPCHI2	6.21
log_Bs_DIRA	6.15
K_plus_fromDs_ptasy_1.00	6.10
log_XsDaughters_min_IPCHI2	5.87
K_minus_fromDs_ptasy_1.00	5.62
cos(Ds h)	5.58
log_Bs_IPCHI2_OWNPV	5.08
log_Bs_FDCHI2_OWNPV	4.04
Xs_max_DOCA	3.98
pi_minus_fromDs_ptasy_1.00	2.59

Table 1: Summary of the relative importance of each variable in the training of the BDTG.

of $\frac{\mathcal{B}(B_s^0 \to D_s K\pi\pi)}{\mathcal{B}(B_s^0 \to D_s \pi\pi\pi)}$ and the ratio of efficiencies discussed in Sec. **??** allows us to estimate the expected number of $B_s^0 \to D_s K\pi\pi$ signal decays. The number of background events can then be computed as

$$N_{bkg} = N_{all} - N_{sig}|_{m_{B_s^0} \pm 50\,\mathrm{MeV}/c^2}. \quad (1)$$

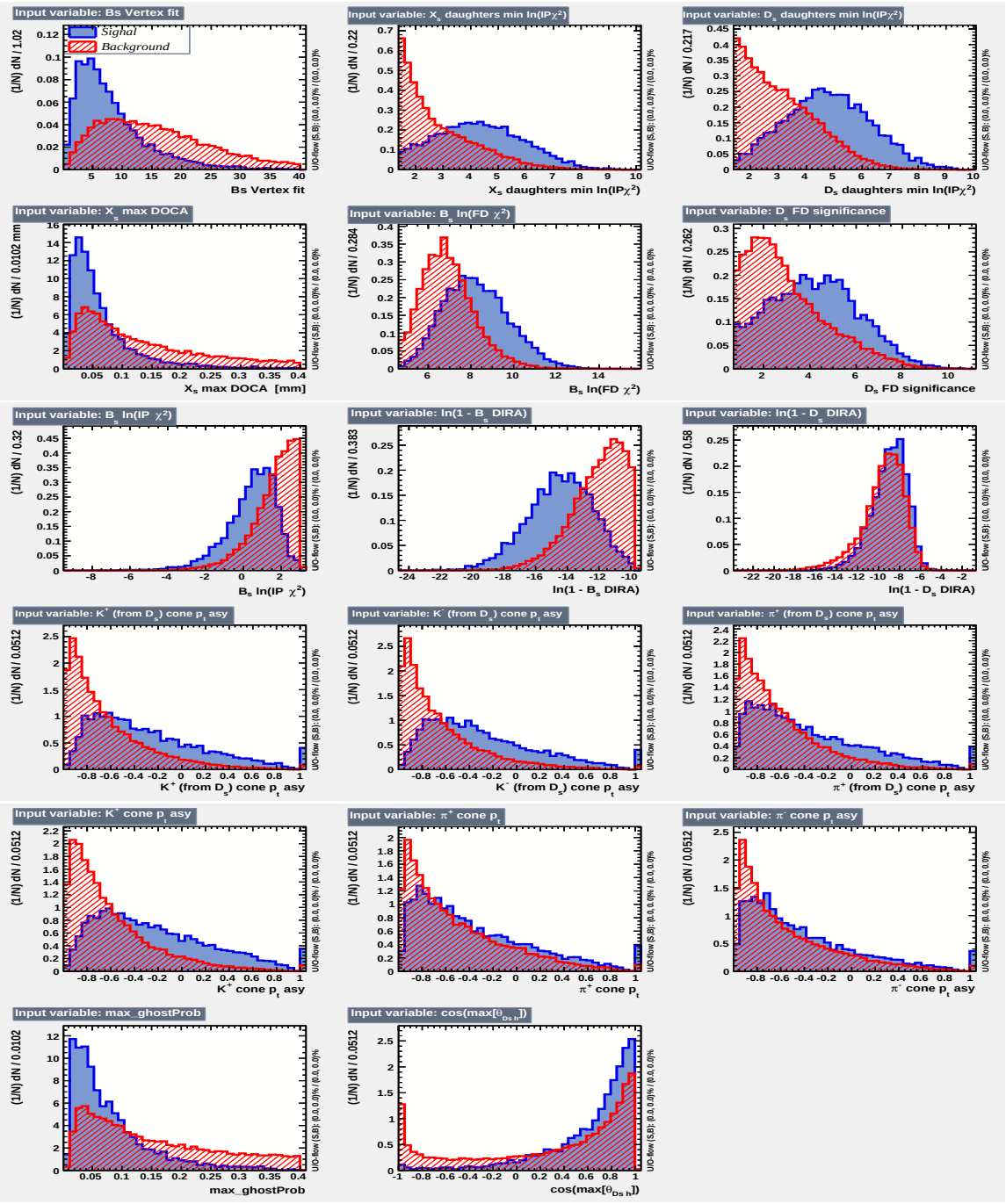

Figure 3: Distributions of the input variables used in the BDTG training. The background is shown as red hatched, while the signal is depicted solid blue.

The efficiency curves as a function of the cut value are shown in Fig. 5. The optimal cut value is found to be BDTG > 0.7012. At this working point the signal efficiency is estimated to be 72.47 %, while the background rejection in the signal region is 97.38 %.

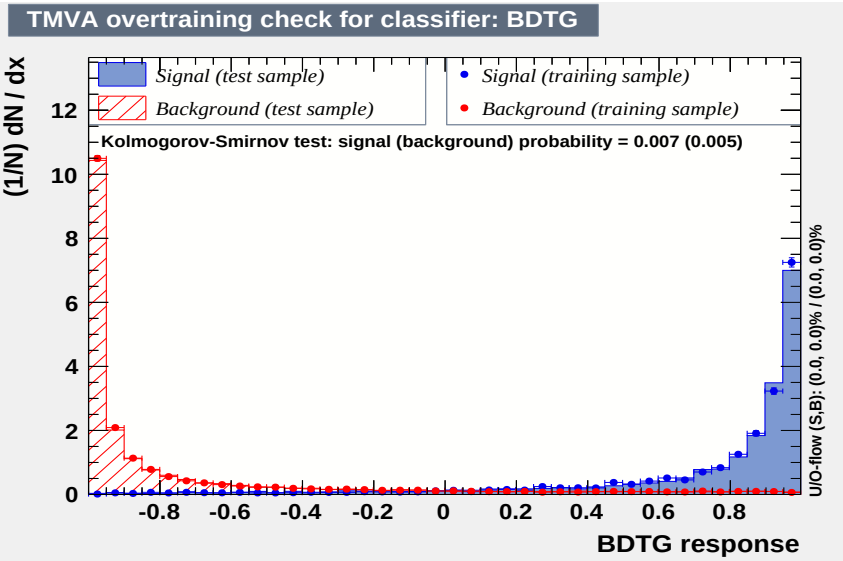

Figure 4: BDTG output classifier distribution for (blue) signal and (red) background. The response of an independent test sample (dots) is overlaid.

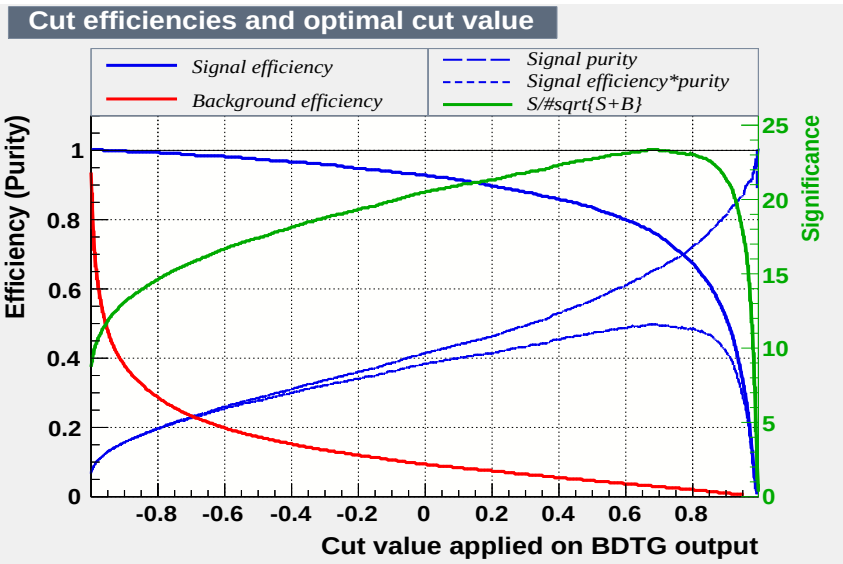

Figure 5: Efficiency and purity curves for (blue) signal, (red) background and the (green) FoM curve, as a function of the chosen cut value.

3 Decay-time Acceptance

The decay-time distribution of the B_s^0 mesons is sculpted due to the geometry of the LHCb detector and the applied selection cuts, which are described in Section 2. In particular, any requirement on the flight distance (FD), the impact parameter (IP) or the direction angle (DIRA) of the B_s^0 mesons, as well as the direct cut on the lifetime, will lead to a decay-time dependent efficiency $a(t)$. This efficiency will distort the theoretically expected, time-dependent decay rate

$$\frac{\Gamma(t)^{observed}}{dt} = \frac{\Gamma(t)^{theory}}{dt} \cdot a(t), \tag{2}$$

and has to be modelled correctly, in order to describe the observed decay rate. We use our control channel for this measurement, because for $B_s^0 \to D_s K\pi\pi$ decays the decay-time acceptance is correlated with the CP-observables which we aim to measure. Therefore, floating the CP-observables and the acceptance shape at the same time is not possible. Hence, a fit to the decay-time distribution of $B_s^0 \to D_s\pi\pi\pi$ candidates is performed and the obtained acceptance shape is corrected by the difference in shape found for the $B_s^0 \to D_s K\pi\pi$ and $B_s^0 \to D_s\pi\pi\pi$ MC.

A PDF of the form

$$\mathcal{P}(t', \vec{\lambda}) = \left[(e^{\Gamma_s t} \cdot cosh(\frac{\Delta\Gamma_s t}{2}) \times \mathcal{R}(t - t') \right] \cdot \epsilon(t', \vec{\lambda}), \tag{3}$$

is fit to the decay time distribution of $B_s^0 \to D_s\pi\pi\pi$ candidates in data. Since the fit is performed untagged, the PDF shown in Eq. 3 contains no terms proportional to Δm_s. The values for Γ_s and $\Delta\Gamma_s$ are fixed to the latest HFAG results [6]. The decay-time acceptance $\epsilon(t', \vec{\lambda})$ is modelled using the sum of cubic polynomials $v_i(t)$, so called Splines [7]. The polynomials are parametrised by so-called knots which determine their boundaries. Knots can be set across the fitted distribution to account for local changes in the acceptance shape. Using more knots is equivalent to using more base splines which are defined on a smaller sub-range. In total, $n + 2$ base splines $v_i(t)$ are needed to describe an acceptance shape which is parametrised using n knots.

For fits shown in the following, the knots have been placed at $t = [0.5, 1.0, 1.5, 2.0, 3.0, 9.5]ps$. To accomodate these 6 knot positions, 8 basic splines v_i, $i = [1, ..., 8]$ are used. Since a rapid change of the decay time acceptance at low decay times due to the turn-on effect generated by the lifetime and other selection cuts is expected, more knots are placed in that regime. At higher decay times we expect linear behaviour, with a possible small effect due to the VELO reconstruction. Therefore fewer knots are used. Furthermore, v_7 is fixed to 1 in order to normalize the overall acceptance function. To stabilise the last spline, v_8 is fixed by a linear extrapolation from the two previous splines:

$$v_N = v_{N-1} + \frac{v_{N-2} - v_{N-1}}{t_{N-2} - t_{N-1}} \cdot (t_N - t_{N-1}). \tag{4}$$

Here, $N = 8$ and t_{N-1} corresponds to the knot position associated with v_{N-1}. The nominal fit to $B_s^0 \to D_s\pi\pi\pi$ data using this configuration is shown in Figure 6. Note that the normalization of the splines in the following figures is not in scale.

The fits to $B_s^0 \to D_s\pi\pi\pi$ and $B_s^0 \to D_s K\pi\pi$ simulation are shown in Figure 7.

The fit parameters obtained from the described fits to data and simulation are summarised in Tab. xXx.

References

[1] R. Fleischer, *New strategies to obtain insights into CP violation through B(s) —¿ D(s)+- K-+, D(s)*+- K-+, ... and B(d) —¿ D+- pi-+, D*+- pi-+, ... decays*, Nucl. Phys. **B671** (2003) 459, arXiv:hep-ph/0304027.

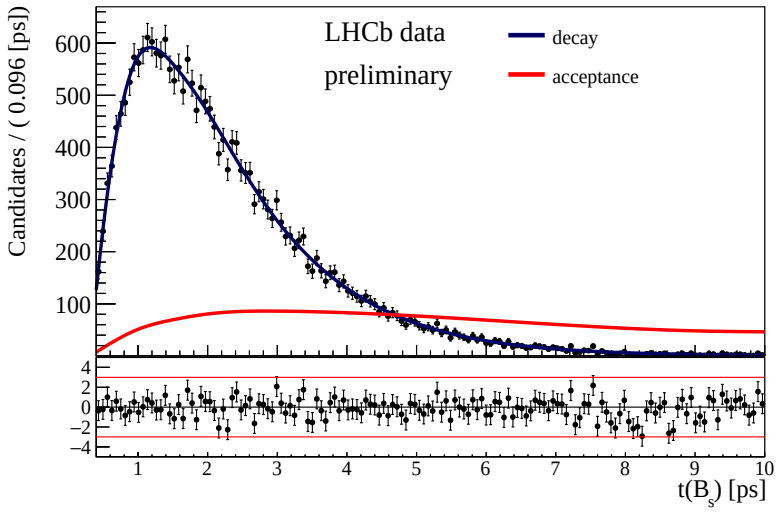

Figure 6: Decay-time distribution of $B_s^0 \to D_s\pi\pi\pi$ candidates for the Run 1 data sample. The fit described in the text is overlaid. The red line shows the spline function describing the acceptance and the blue line depicts the total fit function.

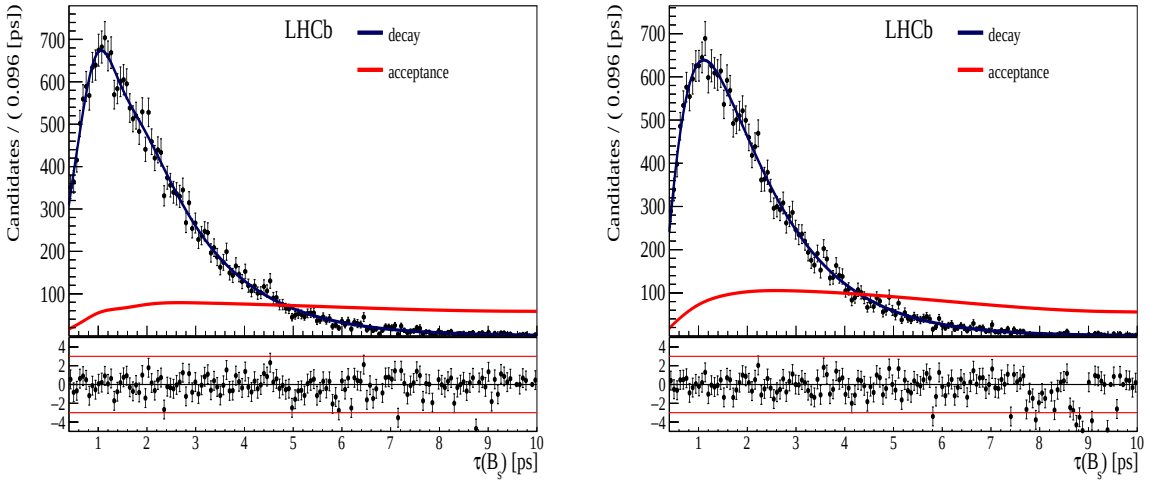

Figure 7: Decay-time distribution of (left) $B_s^0 \to D_s\pi\pi\pi$ and (right) $B_s^0 \to D_sK\pi\pi$ candidates in MC using truth information. The fit described in the text is overlaid. The red line shows the spline function describing the acceptance and the blue line depicts the total fit function.

[2] K. De Bruyn *et al.*, *Exploring $B_s \to D_s^{(*)\pm}K^{\mp}$ Decays in the Presence of a Sizable Width Difference $\Delta\Gamma_s$*, Nucl. Phys. **B868** (2013) 351, arXiv:1208.6463.

[3] S. Blusk, *First observations and measurements of the branching fractions for the decays $\bar{B}_s^0 \to D_s^+K^-\pi^+\pi^-$ and $\bar{B}^0 \to D_s^+K^-\pi^+\pi^-$*, .

[4] LHCb, S. Blusk, *Measurement of the CP observables in $\bar{B}_s^0 \to D_s^+K^-$ and first observation of $\bar{B}_{(s)}^0 \to D_s^+K^-\pi^+\pi^-$ and $\bar{B}_s^0 \to D_{s1}(2536)^+\pi^-$*, 2012. arXiv:1212.4180.

Parameter	Fit to $B_s^0 \to D_s\pi\pi\pi$ data	Fit to $B_s^0 \to D_s\pi\pi\pi$ MC	Fit to $B_s^0 \to D_sK\pi\pi$ MC
v_1			
v_2			
v_3			
v_4			
v_5			
v_6			
v_7	fixed	fixed	fixed

Table 2: Summary of the obtained parameters from the acceptance fits described above.

[5] A. Hoecker *et al.*, *TMVA: Toolkit for Multivariate Data Analysis*, PoS **ACAT** (2007) 040, arXiv:physics/0703039.

[6] Heavy Flavor Averaging Group, Y. Amhis *et al.*, *Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2014*, arXiv:1412.7515, updated results and plots available at http://www.slac.stanford.edu/xorg/hfag/.

[7] T. M. Karbach, G. Raven, and M. Schiller, *Decay time integrals in neutral meson mixing and their efficient evaluation*, arXiv:1407.0748.

The author list for journal publications is generated from the Membership Database shortly after 'approval to go to paper' has been given. It will be sent to you by email shortly after a paper number has been assigned. The author list should be included already at first circulation, to allow new members of the collaboration to verify that they have been included correctly. Occasionally a misspelled name is corrected, or associated institutions become full members. Therefore an updated author list will be sent to you after the final EB review of the paper. In case line numbering doesn't work well after including the authorlist, try moving the `\bigskip` after the last author to a separate line.

The authorship for Conference Reports should be "The LHCb collaboration", with a footnote giving the name(s) of the contact author(s), but without the full list of collaboration names.