Determination of the CKM angle γ in $B^{\pm} \to D(\to K^+K^-\pi^+\pi^-)h^{\pm}$ decays

Martin Duy Tat

4th June 2021

Abstract

A model-independent measurement of the CKM angle γ is performed in $B^{\pm} \to Dh^{\pm}(h = K, \pi)$ decays, where $D \to K^+K^-\pi^+\pi^-$, using the LHCb Run 1+2 dataset. The measurement is based on an optimal binning scheme using strong-phase info determined from quantum-correlated $D^0\bar{D}^0$ pairs at BESIII.

1 Introduction

In the Standard Model, CP-violation is studied by measuring the lengths and angles of the Unitary Triangle of the CKM matrix [1]. The angle $\gamma = \arg(-V_{ud}V_{ub}^*/V_{cd}V_{cb}^*)$, in particular, is the only angle accessible at tree level, with negligible theoretical uncertainties. Thus, a precise determination of γ is a good Standard Model benchmark.

γ is sensitive to $b \to c\bar{u}s$ and $b \to u\bar{c}s$ interference, such as in $B^{\pm} \to DK^{\pm}$ decays reconstructed at LHCb. Fig. 1 shows the $B^- \to D^0K^-$ mode and the colour suppressed $B^- \to \bar{D}^0K^-$ channel, and they interfere when D^0 and $\bar{D}^0$ decay to a common state.

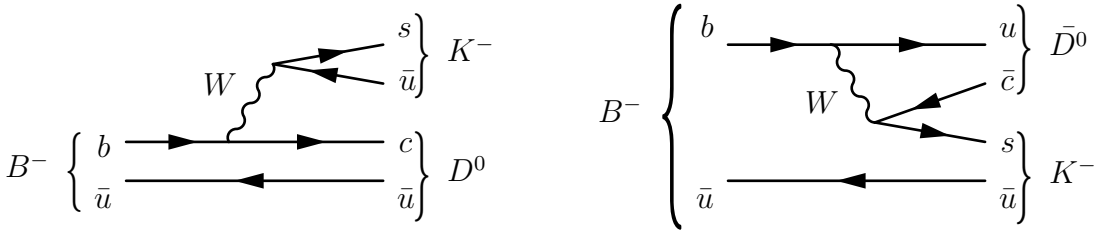


Figure 1: Feynman diagrams of $B^- \to DK^-$ decays at tree level.

A wide range of subsequent D decays has been studied, where $\gamma = (68.7^{+5.2}_{-5.1})^\circ$ is the single most precise measurement using $D \to K^0_S h^+h^-(h = \pi, K)$[2]. In the following analysis, the decay $D \to K^+K^-\pi^+\pi^-$ is considered, where the challenge is the non-trivial variation of D^0 and $\bar{D}^0$ strong-phases across the 5D phase space. An amplitude model may be used to predict the strong-phase variation, but at the cost of a model uncertainty.

In this analysis, a model-independent approach is chosen. Strong-phases will be independently measured, in phase space bins, at the BESIII charm factory. The current BESIII 2010-2011 dataset is insufficient, but significantly more data are expected during 2022 and 2023. An amplitude model for $D \to K^+K^-\pi^+\pi^-$, which is referred to as the LHCb model [3], will be used to study strong-phase variations and develop an effective binning scheme. A poor binning scheme may decrease the statistical sensitivity, but the result remains unbiased. Thus, with a model-independent approach there is no associated systematic uncertainty due to modelling.

2 Formalism

2.1 γ sensitity through $B^\pm$ decays

The amplitude of $B^\pm \to DK^\pm$ is a coherent sum of the diagrams in Fig. 1,

$$\mathcal{A}(B^- \to DK^-) = \mathcal{A}_D(\Phi) + r_B^{DK} e^{i(\delta_B^{DK} - \gamma)} \mathcal{A}_{\bar{D}}(\Phi), \tag{1}$$

$$\mathcal{A}(B^+ \to DK^+) = \mathcal{A}_{\bar{D}}(\Phi) + r_B^{DK} e^{i(\delta_B^{DK} + \gamma)} \mathcal{A}_D(\Phi). \tag{2}$$

r_B is the relative magnitude of the diagrams and $A_{D,\bar{D}}$ are D decay amplitudes. Under CP, the $B^\pm$ decay strong-phase δ_B^{DK} is invariant while the weak phase γ swaps sign.

The $B^\pm \to DK^\pm$ decay is considered in $2\times N$ phase space bins, labelled $i = -N, ..., N$, excluding zero. Bin i is related to $-i$ by a CP transformation. When integrating the square of Eqs. (1)-(2) over phase space Φ, the $B^\mp \to DK^\mp$ yields in bin $\pm i$ are

$$N_i^- = h_{B^-}\Big[F_i + \Big((x_-^{DK})^2 + (y_-^{DK})^2\Big)\bar{F}_i + 2\sqrt{F_i\bar{F}_i}\Big(x_-^{DK}c_i + y_-^{DK}s_i\Big)\Big], \tag{3}$$

$$N_{-i}^+ = h_{B^+}\Big[F_i + \Big((x_+^{DK})^2 + (y_+^{DK})^2\Big)\bar{F}_i + 2\sqrt{F_i\bar{F}_i}\Big(x_+^{DK}c_i + y_+^{DK}s_i\Big)\Big], \tag{4}$$

$$\begin{pmatrix} c_i \\ s_i \end{pmatrix} = \frac{\int_i \mathrm{d}\Phi\, |\mathcal{A}_D||\mathcal{A}_{\bar{D}}| \begin{pmatrix} \cos(\Delta\delta_D) \\ \sin(\Delta\delta_D) \end{pmatrix}}{\sqrt{\int_i \mathrm{d}\Phi\, |\mathcal{A}_D|^2 \int_i \mathrm{d}\Phi\, |\mathcal{A}_{\bar{D}}|^2}}, \quad F_i = \frac{\int_i \mathrm{d}\Phi\, |\mathcal{A}_D|^2}{\sum_i \int_i \mathrm{d}\Phi\, |\mathcal{A}_D|^2}, \quad \bar{F}_i = \frac{\int_i \mathrm{d}\Phi\, |\mathcal{A}_{\bar{D}}|^2}{\sum_i \int_i \mathrm{d}\Phi\, |\mathcal{A}_{\bar{D}}|^2}. \tag{5}$$

$h_{B^\pm}$ are a normalization constants. c_i (s_i) is the cosine (sine) of the strong-phase difference $\Delta\delta_D$ between the D^0 and $\bar{D}^0$ decays, amplitude-averaged over bin i. F_i is the fractional yield of $B^- \to D^0K^-$ in bin i. Assuming CP conservation in D decays, $\bar{F}_i = F_{-i}$. Furthermore, the CP observables in Eqs. (1)-(2) are

$$x_\pm^{DK} = r_B^{DK}\cos\Big(\delta_B^{DK} \pm \gamma\Big), \quad y_\pm^{DK} = r_B^{DK}\sin\Big(\delta_B^{DK} \pm \gamma\Big). \tag{6}$$

From the binned yields of $B^\pm \to DK^\pm$, one can perform a Maximum Likelihood (ML) fit of Eqs. (3)-(4), with external inputs of c_i and s_i from BESIII, to obtain the CP observables $x_\pm^{DK}$ and $y_\pm^{DK}$. These are interpreted in terms of γ, δ_B^{DK} and r_B^{DK}.

To constrain F_i, which are free parameters, the decay mode $B^\pm \to D\pi^\pm$ is included in the fit. This adds another set of Eqs. (1)-(2) with $K \to \pi$. The common topology means $B^\pm \to DK^\pm$ and $D\pi^\pm$ will share the same F_i, but $B^\pm \to D\pi^\pm$ has smaller CP-violation because $r_B^{D\pi} \ll r_B^{DK}$. The fit stability is improved by introducing

$$x_\xi^{D\pi} = \mathrm{Re}(\xi^{D\pi}), \quad y_\xi^{D\pi} = \mathrm{Im}(\xi^{D\pi}), \quad \xi^{D\pi} = \frac{r_B^{D\pi}}{r_B^{DK}} e^{i(\delta_B^{D\pi} - \delta_B^{DK})}.$$

Therefore, the six CP observables in the ML fit are $x_\pm^{DK}$, $y_\pm^{DK}$, $x_\xi^{D\pi}$ and $y_\xi^{D\pi}$.

2.2 Strong-phase data from quantum-correlated $D^0\bar{D}^0$ decays

In $\psi(3770) \to D^0\bar{D}^0$ decays, $D^0\bar{D}^0$ is a correlated antisymmetric state, since $\psi(3770)$ has charge conjugation $\mathcal{C} = -1$. Thus, c_i and s_i can be measured using a double-tag method.

The number of events where only one D meson is reconstructed as $D \to f$ is the single tag yield of f. If both D mesons are reconstructed in the signal mode $D \to K^+K^-\pi^+\pi^-$ and tag mode $D \to f$, one obtains the double-tag yield.

c_i is measured in events where the tag mode is a CP eigenstate or a mode with known CP-even fraction F_+. The yield of CP tagged $D \to K^+K^-\pi^+\pi^-$ events in bin i is

$$M_i^{\pm} = \frac{S_{\pm}}{2S_f}\Big[K_i - 2c_i(2F_+ - 1)\sqrt{K_iK_{-i}} + K_{-i}\Big]. \tag{7}$$

K_i are the double-tag yields, where the tag is a flavour tag, such as $D^0 \to K^-\pi^+$. $S_{\pm}$ and S_f are single tag yields of the CP and flavour tag modes used, respectively. For CP even (odd) modes, $F_+ = 1$ (0). Table 1 shows the tag modes in this analysis. To obtain s_i, the tag mode phase space must also be binned, labelled by j. The analogous expression is

$$M_{ij} = \frac{N_{D\bar{D}}}{2S_fS'_f}\Big[K_iK'_{-j} + 2\sqrt{K_iK'_{-j}K_{-i}K'_j}(c'_ic_j + s'_is_j) + K_{-i}K'_j\Big]. \tag{8}$$

$N_{D\bar{D}}$ is the total number of $D^0\bar{D}^0$ pairs. If the tag is $K^0_S\pi^+\pi^-$, then c'_i and s'_i are well known [4]. If the tag is identical to the signal mode $K^+K^-\pi^+\pi^-$, then $c'_i = c_i$ and $s'_i = s_i$.

Table 1: List of tag modes in the double-tag analysis. Flavour mode conjugates are implied.

Flavour	CP even	CP odd	Self conjugate
$K^-\pi^+$, $K^-\pi^+\pi^0$, $K^-\pi^+\pi^-\pi^+$, $K^-e^+\nu_e$	K^+K^-, $\pi^+\pi^-$, $\pi^+\pi^-\pi^0$, $K^0_S\pi^0\pi^0$, $K^0_L\pi^0$, $K^0_L\eta$, $K^0_L\omega$	$K^0_S\pi^0$, $K^0_S\phi$, $K^0_S\eta$, $K^0_S\eta'$, $K^0_S\omega$	$K^0_S\pi^+\pi^-$, $K^+K^-\pi^+\pi^-$

3 LHCb and BESIII detectors

LHCb [5] is a single arm forward spectrometer designed to study beauty and charm hadrons in pp collisions. The components important for this analysis are the tracking system and the Ring Imaging Cherencov counters (RICH1 and RICH2). The tracking system includes the Vertex Locator (VELO), which provide high precision tracking and identification of displaced secondary vertices. A dipole magnet and the tracking stations measure charged particle momenta. The RICH detectors provides identification of kaons and pions. A $9\,\text{fb}^{-1}$ dataset at $\sqrt{s} = 7$, 8 and 13 TeV will be used in this analysis.

BESIII [6] is a general purpose solenoidal detector. The parts most relevant to this analysis are the Helium gas drift chamber for measuring the momenta and $\mathrm{d}E\,/\,\mathrm{d}x$ of charged particles, a time of flight system for particle ID and an electromagnetic calorimeter to measure shower energies. The dataset currently under consideration has $2.9\,\text{fb}^{-1}$ of $e^+e^- \to \psi(3770)$ events, but this will grow to $20\,\text{fb}^{-1}$ over the next two years.

4 Binning scheme

In Ref. [2], the 2D phase space, visualized on a Dalitz plot, was separated into bins of similar strong-phase to avoid diluting the amplitude-averaged c_i and s_i. Cabbibo-favoured and suppressed resonances were assigned bin numbers with opposite sign to enhance the interference terms in Eqs. (3)-(4), and thus enhance the sensitivity to CP observables.

The 5D phase space of $D \to K^+K^-\pi^+\pi^-$ is not easily visualized. Instead, the LHCb model, implemented in AmpGen [7], is used to calculate the amplitude $\mathcal{A}(D)$ from the D daughter momenta. Defining $\mathcal{A}(D^0)/\mathcal{A}(\bar{D}^0) \equiv r_De^{i\Delta\delta_D}$, where $\Delta\delta_D$ and r_D are the strong-phase difference and relative magnitude, $\Delta\delta_D$ is first uniformly separated into bins.

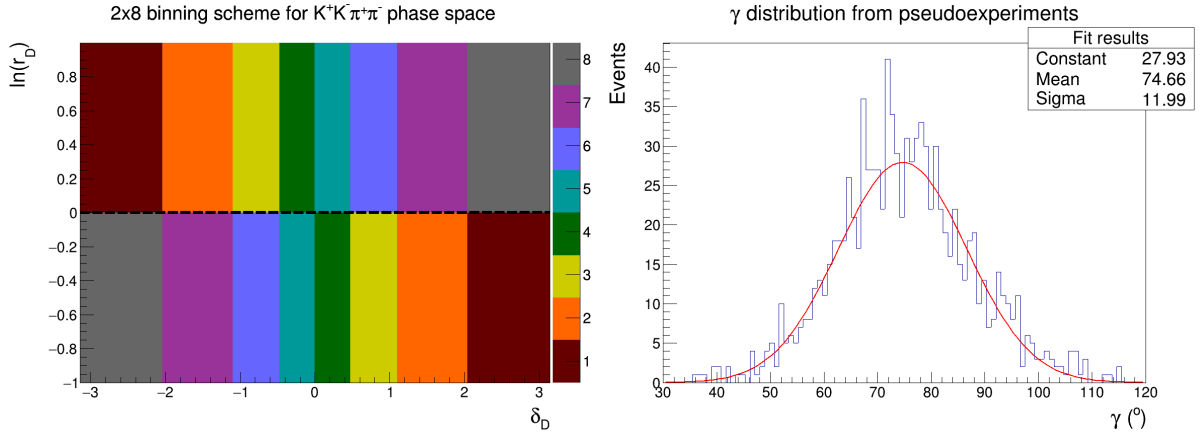
Under CP, $\Delta\delta_D \to -\Delta\delta_D$ and $\ln(r_D) \to -\ln(r_D)$. For $\ln(r_D) > 0$ (< 0), the $\bar{D}^0$ decay is suppressed (favoured), relative to the D^0 decay. Therefore, the interference terms in Eqs. (3)-(4) may be enhanced if bins with $i > 0$ and $i < 0$ are separated along $\ln(r_D) = 0$.

To optimize the binning, the bin boundaries are moved symmetrically around $\Delta\delta_D = 0$ to maximize Q, defined as the $x_\pm$ and $y_\pm$ sensitivity in a binned fit divided by that of an unbinned fit. It can be shown, with $N_i^\pm$ from Eqs. (3)-(4), that

$$Q^2 = \frac{1}{2}\left(Q_+^2 + Q_-^2\right), \quad Q_\pm^2 = 1 - \sum_i \frac{F_i F_{-i}\left(1 - c_i^2 - s_i^2\right)}{N_i^\pm}.$$

A minimal estimate of the signal yield, based on related studies, is 2000 $B^\pm$ candidates from LHCb Run 1+2. To assess the achievable γ precision and the sensitivity loss due to binning, 1000 toy experiments, each with 2000 $B^\pm$ candidates, were generated with the LHCb model in AmpGen, using $\gamma = 75°$, $\delta_B = 130°$ and $r_B = 0.1$. An unbinned ML fit was performed to establish a benchmark of $\Delta\gamma = 11°$.

With 2×8 bins, Fig. 2a shows the optimal binning scheme, with $Q = 0.90$, indicating that 10% sensitivity is lost due to binning. Using the binning in Fig. 2a, a ML fit was performed on each toy experiment using Eqs. (3)-(4) to extract $x_\pm^{DK}$ and $y_\pm^{DK}$. c_i, s_i and F_i were obtained from Monte Carlo integration of Eqs. (5) with $\mathcal{A}(D)$ from the LHCb model. Finally, γ, δ_B and r_B were obtained from $x_\pm^{DK}$ and $y_\pm^{DK}$ using Eqs. (6).



(a) Binning scheme definition for 2×8 bins **(b)** Distribution of γ from toy study

Figure 2

Fig. 2b shows the γ distribution in the toy experiments. The binned γ precision is $\Delta\gamma = (12.0 \pm 0.4)°$, which is consistent with the unbinned benchmark and $Q = 0.90$. The pull distributions of $x_\pm^{DK}$, $y_\pm^{DK}$, γ, δ_B and r_B have mean and standard deviations consistent with zero and one, respectively, except for r_B, which has a small bias.

5 $B^\pm$ candidate selection

The $B^\pm \to (K^+K^-\pi^+\pi^-)_D h^\pm$ candidates in the LHCb dataset are reconstructed using standard requirements from Ref. [2]. The D invariant mass must be within 25 MeV of the D^0 mass. The tracks are refitted with the D invariant mass constrained and its momentum pointing towards $B^\pm$. A cut on χ^2 from this fit removes the D mass sidebands.

$B^\pm \to K^+K^-\pi^+\pi^-K^\pm$ candidates form a charmless background in the $B^\pm$ mass spectrum, but does not peak in the D mass spectrum. To estimate the contamination, the χ^2 cut was removed to preserve the D mass sidebands. Inside the mass region $m(K^+K^-\pi^+\pi^-) \in [1770, 1820]$MeV, a fit to the $B^\pm$ mass spectrum gives a yield of 2605 ± 57. A flight significance cut, defined as the flight distance divided by its error, reduces this yield to 110 ± 19. No charmless background was found for $B^\pm \to D\pi^\pm$.

A significant mis-ID background is $B^\pm \to Dh^\pm$, $D \to K^\pm\pi^\mp\pi^\pm\pi^\mp$, where a pion is identified as a kaon. A Monte Carlo (MC) simulation sample indicated that 7.2% of $B^\pm$ candidates inside the signal region are $D \to K^\pm\pi^\mp\pi^\pm\pi^\mp$. A tighter particle identification requirement reduces this to 1.8% while keeping 93% of $D \to K^+K^-\pi^+\pi^-$ candidates.

For combinatorial background, a Boosted Decision Tree (BDT) was trained using MC samples as signal and the region $m(Dh^\pm) \in [5800, 7000]$MeV in data as background. 99.4% of the combinatorial background was removed and 93.0% of the signal remained.

6 $B^\pm$ global mass fit and binned CP fit

A global ML fit of the $B^\pm$ mass spectrum, shown in Fig. 3, was performed to determine global yields and shape parameters, with shape parameterisations from Ref. [2]. The final analysis requires minor adjustments. Currently, only Run 2 data have been processed.

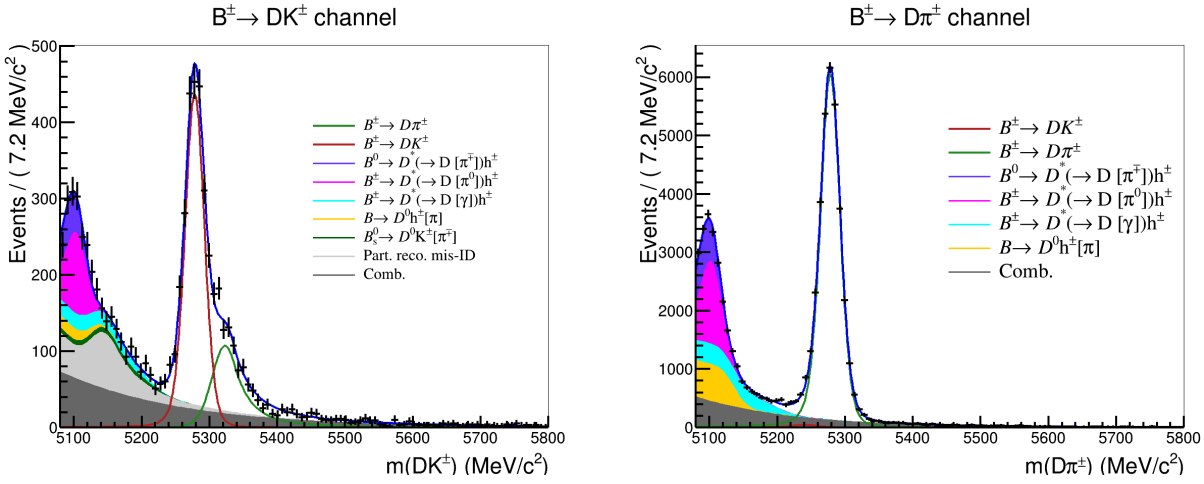


Figure 3: Global mass fit of $B^\pm \to DK^\pm$ (left) and $B^\pm \to D\pi^\pm$ (right) in LHCb data.

The combinatorial background is parameterised by an exponential function. The signal is a sum of a Gaussian $f_{\rm G}(m|m_B, \sigma)$ and a modified Gaussian,

$$f_{\rm MG}(m|m_B, \sigma, \alpha_L, \alpha_R, \beta) \propto \begin{cases} \exp\left(\frac{-\Delta m^2(1+\beta\Delta m^2)}{2\sigma^2+\alpha_L\Delta m^2}\right), & \Delta m = m - m_B < 0 \\ \exp\left(\frac{-\Delta m^2(1+\beta\Delta m^2)}{2\sigma^2+\alpha_R\Delta m^2}\right), & \Delta m = m - m_B > 0, \end{cases} \tag{9}$$

which accounts for the radiated tail. To the left of the signal are partially reconstructed $B_{(s)}$ decays where a $\pi^\pm$ or γ is missed, indicated by $[\pi/\gamma]$. For $B^\pm \to DK^\pm$, there is a mis-ID component of $B^\pm \to D\pi^\pm$, and a partially reconstructed mis-ID component, where $\pi^\pm$ is misidentified as $K^\pm$. The $B^\pm \to DK^\pm$ signal yield, from Run 2 alone, is 2290 ± 59, which is higher than the original estimate. The $B^\pm \to D\pi^\pm$ yield is $33\,113 \pm 211$.

Finally, $B^{\pm}$ candidates are split by charge and bins, using the binning in Fig. 2a. The current study uses c_i and s_i from the LHCb model. A fit is performed with shape parameters fixed from the global fit, but the yields are floated. Fig. 4 shows the fitted CP observables. The geometrical angle between $(x_{\pm}, y_{\pm})$ is 2γ, but its exact value is blinded.

To check the fit robustness, 1000 toy datasets are generated, for both global and CP fits, using the fitted parameters. Each toy is run through the fitting procedure. All CP observables have pulls consistent with zero mean and unit standard deviation.

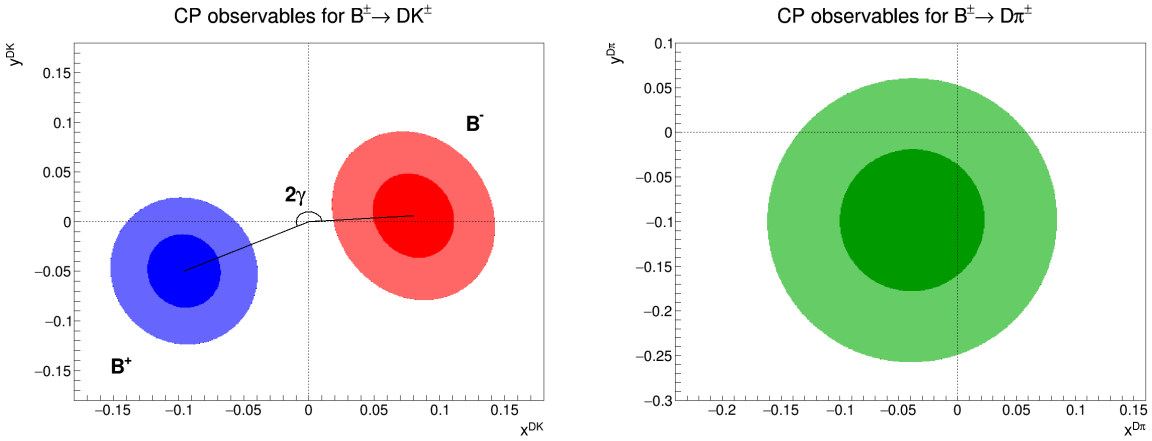


Figure 4: Confidence levels at 68.2% and 95.5% of $(x_{\pm}^{DK}, y_{\pm}^{DK})$ (left) and $(x_{\xi}^{D\pi}, y_{\xi}^{D\pi})$ (right).

7 External strong-phase input from BESIII

Event reconstruction at BESIII follows Ref. [4]. The single tag yield is obtained from a ML fit to the beam constrained D mass $M_{\rm BC} = (E^2_{\rm beam} - |\sum_i \mathbf{p}_i|^2)^{1/2}$, shown on the left in Fig. 5 for $D \to K^+K^-\pi^+\pi^-$. Peaking backgrounds are fixed from inclusive MC. The main component, in green, is $D \to K^0_S K^+ K^-$. The signal shape is a MC sample convolved with a Gaussian. An Argus shape parameterises the combinatorial background.

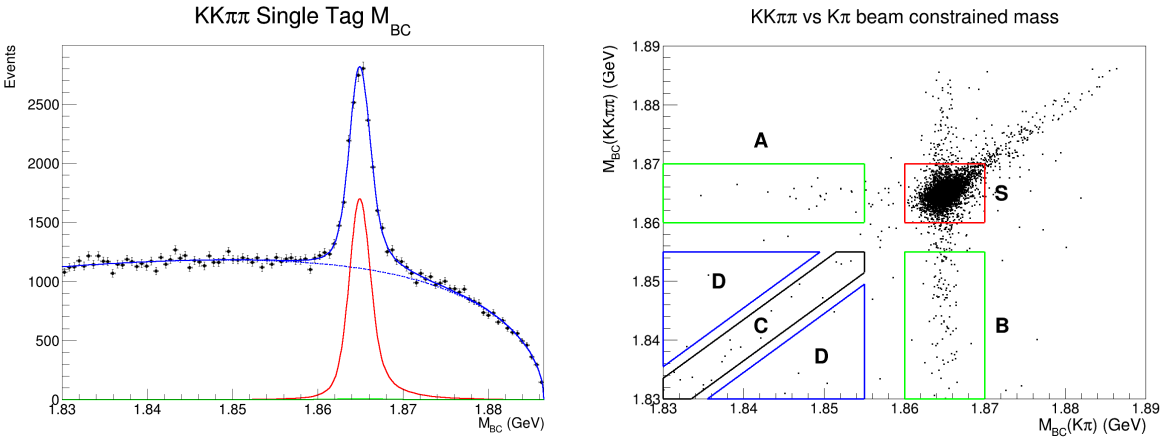


Figure 5: $M_{\rm BC}$ of single tag $K^+K^-\pi^+\pi^-$ (left) and double-tag $K^+K^-\pi^+\pi^-$ vs $K^{\mp}\pi^{\pm}$ (right).

For double-tag yields, a sideband subtraction technique is used. Fig. 5, on the right, shows $D \to K^+K^-\pi^+\pi^-$ versus $D \to K^{\pm}\pi^{\mp}$ events, where the $M_{\rm BC}$ for each D meson is plotted in a 2D scatter plot. Region S is the signal region, while A and B contain only

one real D meson. In region C, daughters have been swapped between the D mesons. Region D contains non-charm background. The background is estimated using

$$B = P + \frac{a_S}{a_D} Y_D + \sum_{i=A,B,C} \frac{a_S}{a_i} \left(Y_i - \frac{a_i}{a_D} Y_D \right),$$

where P is the peaking background, and Y_i and a_i are the yield and area of region i. The single and double-tag yields will be used in a fit of Eqs. (7)-(8) to obtain c_i and s_i.

Fig. 6 shows, on the left, $D \to K^{\pm}\pi^{\mp}$ double-tag yields with the LHCb model prediction, using 2×4 bins. On the right are the double-tag yields with CP odd and even tags $K^0_S\pi^0$ and K^+K^-, normalized by $K^{\pm}\pi^{\mp}$ yields. The agreement is satisfactory, but the yields are not corrected for efficiencies yet, so perfect agreement is not expected.

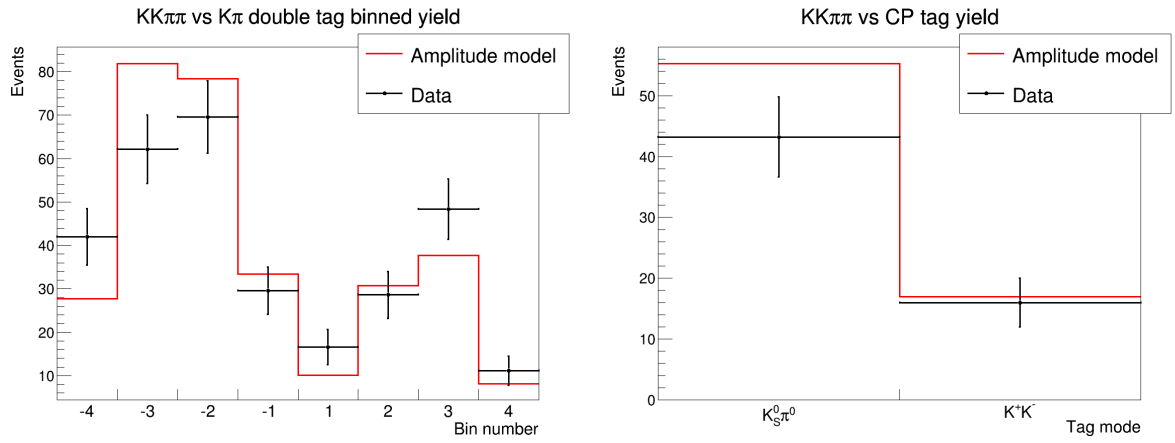


Figure 6: Double-tag yield of $K^{\pm}\pi^{\mp}$ (left), and K^+K^- and $K^0_S\pi^0$ (right).

8 Conclusion and future work

The current results at both LHCb and BESIII are consistent with expectations. The next steps in the LHCb analysis are cut optimizations and systematics studies. For BESIII, all tag mode selections must be finalized and the yields determined on the current and future datasets. The analyses will be combined into a first γ determination in this channel.

References

[1] M. Kobayashi and T. Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49:652–657, 02 1973.

[2] R. Aaij et al. Measurement of the CKM angle γ in $B^{\pm} \to DK^{\pm}$ and $B^{\pm} \to D\pi^{\pm}$ decays with $D \to K^0_S h^+ h^-$. *JHEP*, 02:169, 2021.

[3] R. Aaij et al. Search for CP violation through an amplitude analysis of $D^0 \to K^+K^-\pi^+\pi^-$ decays. *JHEP*, 02:126, 2019.

[4] M. Ablikim et al. Improved model-independent determination of the strong-phase difference between D^0, $\bar{D}^0 \to K^0_{S,L}K^+K^-$ decays. *Phys. Rev. D*, 102(5):052008, 2020.

[5] A. Augusto Alves, Jr. et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[6] M. Ablikim et al. Design and Construction of the BESIII Detector. *Nucl. Instrum. Meth. A*, 614:345–399, 2010.

[7] T. D. Evans. AmpGen. https://github.com/GooFit/AmpGen.

DPhil thesis plan

During my 1st year, I started with developing the binning scheme, and selecting data at both LHCb and BESIII. The LHCb analysis has now developed to a stage where γ can be determined, and the plan for 2nd year is to perform extensive background and systematics studies, before going through the review process. The data selection at BESIII is mainly finished and during 2nd year I will finalize the yields and backgrounds of all tag modes. I will run the same selection on the new 2022-2023 dataset. From Spring 2022, I will be working on testbeam studies of TORCH at CERN, which will continue into my 3rd year.

Below is an outline of this work, which will be included in my DPhil thesis:

1. Introduction and theory
    - Introduction to the Standard Model of Particle Physics
    - Motivation for studying CP-violation
2. LHCb and BESIII detectors
    - Description of the LHCb and BESIII sub-detectors
    - Overview of data and simulation samples
3. Time Of internally Reflected CHerenkov light (TORCH) sub-detector
    - Testbeam evaluation of the prototype module
4. Formalism and optimization of binning scheme
    - GGSZ method for model-independent γ measurement
    - Motivations for an optimal binning of $B^{\pm} \to (K^+K^-\pi^+\pi^-)_D h^{\pm}$ phase space
5. BESIII event selection
    - Selection of single and double-tagged events
    - Identification of peaking backgrounds
6. Determination of yields and strong-phases
    - Fit procedure to obtain c_i and s_i from single and double-tag yields
    - Evaluation of systematic uncertainties
7. LHCb $B^{\pm} \to (K^+K^-\pi^+\pi^-)_D h^{\pm}$ candidate selection
    - Offline selection with rectangular cuts and BDT
    - Treatment of charmless, mis-ID and other backgrounds
8. Global $B^{\pm}$ mass spectrum fit
    - Signal shape and partially reconstructed background parameterisation
9. Binned CP fit and determination of γ
    - Fit of CP observables and interpretation in terms of γ
    - Study of systematic uncertainties
10. Conclusion and outlook