PROCEEDINGS OF SCIENCE

Flavour Tagging in the LHCb experiment

Davide Fazzini*, on behalf of the LHCb collaboration
Universitá degli studi & INFN Milano-Bicocca, Italy
E-mail: davide.fazzini@cern.ch

The *CP* violation measurements represent one of the key topic covered by the LHCb collaboration. The time-dependent *CP* violation measurements require the knowledge of the flavour at production of the *B* signal candidate, thus the flavour tagging tool represents a fundamental ingredient for this kind of analyses. In this contribution the most recent results in the flavour tagging achieved at the LHCb experiment are highlighted.

Sixth Annual Conference on Large Hadron Collider Physics (LHCP2018)
4-9 June 2018
Bologna, Italy

*Speaker.

PoS(LHCP2018)230

1. Introduction

The LHCb detector has been designed to perform precision measurement of b- and c- hadron decays [1, 2]. In particular the study of the *CP* violation represents one of the most important topic covered by the LHCb collaboration. A fundamental information, required by the measurements of *CP* violation, is the knowledge of flavour at production of the B signal candidate. In the $B^0_{(d,s)}$ decays this information is difficult to determine since the final states are often *CP* eigenstates and because of the presence of the neutral flavour oscillations. On the other hand, for all the other hadrons it is possible to identify the initial B flavour looking at the flavour at decay, which is defined by the electric charge of the decay products in the final state. The Flavour Tagging (FT) technique is an essential tool that allows to fulfil this task by looking at the charge correlation between the signal B and the other particles generated in the event. For this reason it represents a key ingredient in all the time-dependent *CP* violation measurements.

2. Flavour tagging algorithms

The Flavour Tagging tool comprises various different algorithms looking for a specific type of particle generated in the event, which has high possibility to be correlated in charge with the flavour of the signal B meson. These algorithms, so-called taggers, are divided in two categories: the "Opposite Side" (OS) taggers, whose target particles come from the decay of the opposite B, and "Same Side" (SS) algorithms, whose particle is generated from the remnants of the signal b fragmentation. While the OS algorithms [3] are able to tag both the B^0_d and B^0_s mesons indifferently, the usage of the SS taggers depends on the quark content of the signal B meson. The OS algorithms available so far have been optimised in order to look for the correlation between the signal B flavour with the electric charge of a kaon (OSK) coming from the $b \to c \to s$ decay chain, of a muon (OSμ) or an electron (OSe) coming from a semileptonic b decay and of a reconstructed secondary charm hadron (OSc) [4]. In addition, there is the OSVtx tagger which has been developed in order to exploit the charge of the inclusive secondary vertex reconstructed from the opposite b-hadron decay products. On the other hand the SS taggers can exploit the information related to a pion (SSπ) or a proton (SSp) in case of a B^0_d meson [5], and to a kaon (SSK) if the signal of interest is a B^0_s meson [6]. A schematic representation of the taggers available within the LHCb collaboration is shown in Fig. 1.

Each taggers is based on the output of one or more multivariate classifiers, taking as input both geometrical and kinematic information. The classifiers are trained on flavour specific decays, where the flavour at decay is uniquely defined by the flavour of the decay products. If more than one algorithm is able to provide a tagging decision for the initial flavour of the B meson candidate, a combination of their information is computed in order to improve the overall mistag rate. Examples of combinations are represented by the OScomb and SScomb taggers which are the combination of all the OS taggers available and the combination of the SSπ and SSp tagging algorithms, respectively.

3. Flavour tagging performance

For each reconstructed signal candidate, the flavour tagging algorithms can provide a tag deci-

PoS(LHCP2018)230

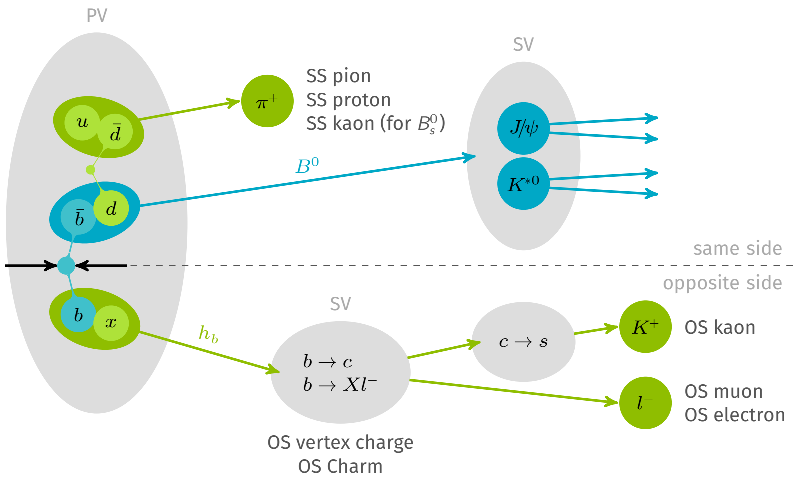

Figure 1: Schematic representation of the FT algorithms available at LHCb.

sion (d) equal to 1 if the signal candidate is a B meson, equal to -1 if the candidate is an antimeson and null if the algorithm is not able to assign a decision on the initial flavour. The tagging decisions are based on the charge of the tagging particle, correlated to the signal B meson flavour. However the flavour tagging algorithms are not perfect tools and their performance can be estimated by means of three different quantities: the mistag rate, the tagging efficiency and the tagging power.

The tagging efficiency represents the fraction of B candidates for which the tagging algorithm is able to provide a tagging decision and a mistag rate. It is defined as:

$$\varepsilon_{tag} = \frac{N_R + N_W}{N_R + N_W + N_U} \tag{3.1}$$

where N_W and N_R and N_U are the numbers of events wrongly tagged, rightly tagged and for which the algorithm in not able to give a response, respectively.

In addition to the tagging decision, each tagger provides also an estimation of the probability (ω) for the tag decision to be wrong. The mistag rate is a continuous variable in the range [0, 0.5] and can be defined as:

$$\omega = \frac{N_W}{N_R + N_W}. \tag{3.2}$$

The mistag rate can be measured only on flavour specific decays. In particular the formula in Eq. (3.2) is relevant only for the charged B mesons where it is possible to compare directly the flavour of the reconstructed meson with the flavour tagging decision. The estimation of the mistag rate is more complicated when neutral B mesons are involved, since they are affected by neutral flavour oscillations. In this case the mistag rate has to be extracted by means of a time-dependent fit on the B flavour oscillations as a function of the proper decay-time. Finally, when the flavour tagging algorithms are applied to non-flavour specific decays, it is not possible to measure directly the mistag rate but it has to be estimated, as described in Sec. 4.

The mistag rate and the tagging efficiency allow a determination of the sensitivity to the *CP* asymmetry. The measured time-dependent *CP* asymmetry (A_{CP}^{meas}) related to the tagged events is reduced by a dilution factor depending on the mistag with respect to the true asymmetry (A_{CP}):

$$A_{CP}^{meas}(t) = \frac{N(\overline{B}^0 \to f)(t) - N(B^0 \to f)(t)}{N(\overline{B}^0 \to f)(t) + N(B^0 \to f)(t)} = (1-2\omega)A_{CP}(t) = DA_{CP}(t). \quad (3.3)$$

Thus the true *CP* asymmetry and its statistical error can be evaluated as:

$$A_{CP} = \frac{A_{CP}^{meas}}{1-2\omega}, \qquad \sigma_{A_{CP}} \propto \frac{1}{\sqrt{\varepsilon_{tag}N}(1-2\omega)} \quad (3.4)$$

which is inversely proportional to the quantity, named tagging power, defined as:

$$\varepsilon_{eff} = \varepsilon_{tag}(1-2\omega)^2 = \varepsilon_{tag}D^2. \quad (3.5)$$

Thus, the tagging power (ε_{eff}) is used as figure of merit to be maximized during the training and development of the flavour tagging algorithms [7].

4. Development of a tagging algorithm

The main steps constituting the development of a flavour tagging algorithm is common to all the existing taggers. The various steps comprise the selection of the tagging tracks' candidates, the training of the tagging algorithm, the conversion from the classifier output into a mistag rate predicted by the algorithm and the calibration of such mistag. The tagging candidate selection represents the first stage of the development of a tagging algorithm and is performed to enhance the purity of the tagging candidates, removing most of the background tracks. The amount of this contamination is often orders of magnitude greater than the amount of tagging tracks, thus the candidate selection covers a crucial role for the optimisation of the tagging algorithm response. Generally, the candidate selection is performed by applying a set of rectangular cuts on the most sensitive variables, including information related to the signal B candidate, the reconstructed particles, which will be selected as tagging candidates, and to the whole event.

The tagging candidates passing the selection are used to train one or more multivariate classifiers. The aim of the multivariate algorithm is to choose the best tagging candidate, whose electric charge will be used to infer the flavour of the B meson at production. The classifier takes as input both kinematic and geometrical information related to the signal B meson and the tagging candidates. The output of the multivariate algorithm is then transformed into a predicted mistag rate (η).

The final step is represented by the calibration of the predicted mistag rate. Since the tagging performance depends on the kinematic of the decay of interest and on the requirements applied selecting the signal, it is important to calibrate the predicted mistag rate to obtain the true mistag rate (ω). The calibration of the mistag rate is therefore performed on flavour specific decay channels, so-called "control samples", which shares the same kinematic properties of the decay of interest.

5. Flavour Tagging in Run 2

As mentioned in the previous section, the flavour tagging performance depends on the kinematic of the decay of interest. In addition to that the flavour tagging performance is also sensitive to

PoS(LHCP2018)230

the center of mass energy, the trigger efficiency, the tracks multiplicity and the number of primary vertices reconstructed in the event. Because of this dependence a variation in the flavour tagging performance is expected passing from Run 1 to Run 2 data taking conditions. However the trend of the variations is not expected to be the same for all the tagging algorithms available in LHCb. Indeed, while the SS taggers show a small natural improvement ($\sim 10\%$ with respect to those in Run 1), mainly due to the higher transverse momentum of the B mesons, the OS algorithms turn out to have a loss in the flavour tagging performance ($\sim 30\%$ with respect to those in Run 1).

In order to regain the tagging power loss, related to the OS taggers, and to increase the overall flavour tagging performance with the Run 2 data taking conditions, a wide re-optimisation campaign has been performed. This campaign has consisted in a retuning or redesigning of the flavour tagging algorithms using the new Run 2 data. In particular the SSK, the OSe, the OSμ and the OSK tagging algorithms have been completely revisited and optimized, while the other taggers are remained untouched, since their performance were compatible or greater with respect to those obtained with Run 1 data.

The reoptimisation of the OSe, OSμ and OSK algorithms consists of two steps. Each of these steps is performed on an independent subsample of events taken from the $B^+ \to J/\psi K^+$ Run 2 data sample. Firstly a tagging candidate selection is performed using various kinematic, geometrical and PID information. A numerical optimisation of the candidate selection has been performed by means of gradiant boosted regression trees as a function of the applied requirements, maximising the average tagging power defined as:

$$\langle \varepsilon_{eff} \rangle = f(\hat{\theta} > \hat{x}) \tag{5.1}$$

where $\hat{\theta}$ is the set of information used in the candidate selection and $\hat{x}$ is the best set of values determined by the optimisation. At each step, the tagging track candidate with the highest transverse momentum is taken in order to evaluate the average tagging power. The second step consists in the training of the multivariate classifier. The aim of the training lies in the discrimination between the signal, represented by the tracks correctly correlated to the B meson flavour, and the background, comprising the tracks wrongly correlated to the B meson flavour. Since the B^+ meson is not affected by the flavour oscillations, the rightly and wrongly tagged B candidates are easily identified, since the true flavour is determined by the B charge. Also, in this case, both kinematic and geometrical information are used as input to the algorithm. Finally the multivariate output is converted into a predicted mistag rate. The OS tagging performance has been evaluated on an independent sample of $B^0 \to D^-\pi^+$ Run 2 data, after having properly calibrated the predicted mistag rate. The tagging performance is reported in Tab. 1 and is compatible with those obtained in Run 1. Thus the initial loss in the tagging power has been recovered thanks to the tagging re-optimisation.

The redesign of the SSK tagging algorithm is performed on fully simulated events of $B_s^0 \to D_s^-\pi^+$ decay mode, since the fast oscillations of the B_s^0 meson makes impossible the classifier training on data. After a loose pre-selection applied to the tagging tracks in order to reduce the background contamination, the optimisation strategy consists of two classifiers. The first multivariate algorithm is trained to discriminate between the true tagging tracks, coming from the fragmentation of the signal B_s^0 meson, and underlying tracks, originating from soft QCD processes and uncorrelated to the signal B_s^0 meson flavour. For each B_s^0 candidate, the three tagging tracks' candidates with highest multivariate score are used for the training of the second classifier. This

Table 1: Summary of the performance of the tagging algorithms after the re-optimisation campaign on the $B^0 \to D^- \pi^+$ decay channel ($B_s^0 \to D_s^- \pi^+$ for the SSK).

Tagger	ε [%]	ω [%]	$\varepsilon \langle D^2 \rangle = \varepsilon(1-2\omega)^2$ [%]
OSμ	0.915 ± 0.053	30.713 ± 0.434	1.361 ± 0.062
OSe	4.451 ± 0.038	34.038 ± 0.604	0.454 ± 0.035
OSK	19.600 ± 0.073	37.557 ± 0.315	1.214 ± 0.061
OSVtx	20.834 ± 0.075	36.994 ± 0.308	1.410 ± 0.067
OSc	5.025 ± 0.040	34.062 ± 0.620	0.511 ± 0.040
OScomb	40.154 ± 0.090	35.123 ± 0.211	3.555 ± 0.101
SSK	68.190 ± 0.177	39.667 ± 0.507	2.912 ± 0.286
SSπ	83.486 ± 0.068	42.561 ± 0.145	1.848 ± 0.072
SSp	37.767 ± 0.089	43.645 ± 0.221	0.610 ± 0.042
SScomb	87.590 ± 0.061	41.787 ± 0.142	2.364 ± 0.081

algorithm is trained with the aim to distinguish the B_s^0 from the $\overline{B_s^0}$ mesons and providing a tagging decision and a predicted mistag rate. The SSK tagging performance, reported in Tab. 1, have been obtained on a Run 2 data sample of $B_s^0 \to D_s^- \pi^+$ decays.

The tagging power results to be about 45% higher with respect to those available in Run 1. For sake of completeness also the tagging performance of the algorithms that did not go through a reoptimisation process are reported in Tab. 1.

6. Inclusive Tagger algorithm

The Run 2 re-optimisation campaign increases noticeably the overall tagging power provided by the various flavour tagging algorithms. However, further improvements are necessary in order to tackle the more challenging LHCb environment in the future. Indeed, improving the single classical tagging algorithm is becoming harder and harder and the LHCb is working on a new concept of flavour tagging algorithm. The idea consists in developing an algorithm based not anymore on a specific physical process, but using the entire event information to infer the signal B meson flavour. Preliminary developments of such an algorithm are currently conducted based on a Recursive Neural Network (RNN), which represents a natural approach for handling variables sized (tracks and vertices).

These preliminary studies are performed on fully simulated events of $B^+ \to J/\psi K^+$ decays. The RNN is trained using as input kinematic, topological, tracking and PID information from all tracks in the event with the aim to distinguish between B^+ and B^- mesons. Even if the development is still in a preliminary stage, the results obtained so far are very promising, as shown in Fig. 2. The preliminary performance of the inclusive tagger is compared with the one obtained combining all the available classical tagging algorithms, and appear to be significantly better.

7. Conclusions

The precision measurements on the CP violating asymmetries performed by LHCb, have been possible thanks to the flavour tagging tool, which represents a key ingredient of these time-

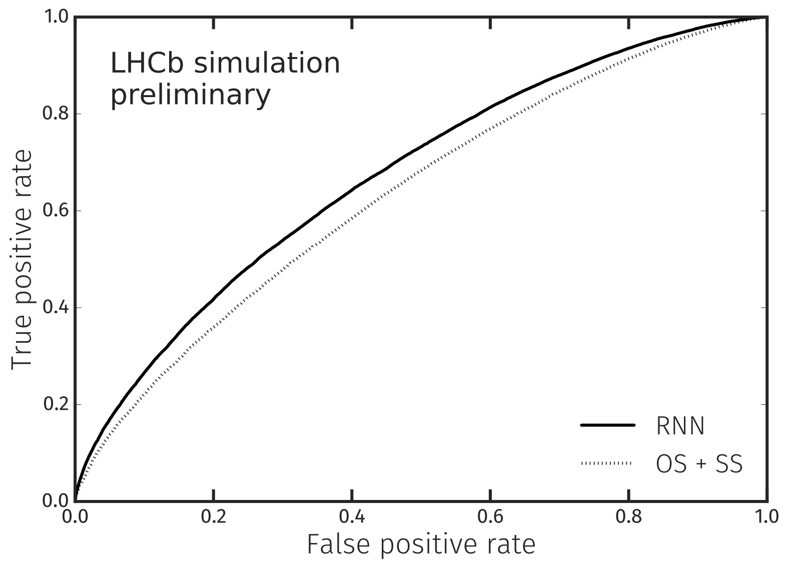

Figure 2: Comparison of the inclusive tagger performance with respect to the one obtained from the combination of the classical algorithms.

dependent analyses. Since the flavour tagging performance depends on the kinematic of the signal event and on the data taking conditions, a re-optimisation of the tagging algorithms was performed leading to higher combined tagging performance on Run 2 data with respect those available on Run 1 data. In addition, a completely new algorithm, namely the "Inclusive tagger", is under development. Even if the tagger is still in an early stage of development, the corresponding preliminary results seem very promising. These improvements will allow the LHCb experiment to obtain even more precise results regarding the *CP* violation in the *b*-quark sector in the future.

References

[1] LHCb collaboration, A. A. Alves, Jr. et al., *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005.

[2] LHCb collaboration, R. Aaij et al., *LHCb Detector Performance*, *Int. J. Mod. Phys.* **A30** (2015) 1530022 [1412.6352].

[3] LHCb collaboration, R. Aaij et al., *Opposite-side flavour tagging of B mesons at the LHCb experiment*, *Eur. Phys. J.* **C72** (2012) 2022 [1202.4979].

[4] LHCb collaboration, R. Aaij et al., *B flavour tagging using charm decays at the LHCb experiment*, *JINST* **10** (2015) P10005 [1507.07892].

[5] LHCb COLLABORATION collaboration, R. Aaij et al., *New algorithms for identifying the flavour of* B^0 *mesons using pions and protons*, *Eur. Phys. J. C* **77** (2016) 238. 23 p.

[6] R. Aaij et al., *A new algorithm for identifying the flavour of b 0 s mesons at lhcb*, *Journal of Instrumentation* **11** (2016) P05010.

[7] LHCb COLLABORATION collaboration, R. Aaij et al., *Optimization and calibration of the LHCb flavour tagging performance using 2010 data*, .