◇ Monograph Excerpt ◇

Matter Antimatter Fluctuations

search, discovery and analysis of Bs flavor oscillations

Nuno Leonardo

Complete work published as:
Analysis of B_s oscillations at CDF, MIT Thesis (2006)
Matter antimatter fluctuations, Monograph, LAP Lambert (2011)

Chapter 9

Same-side kaon tagging algorithm development

The basic ideas behind the method of same-side tagging have been exposed in Chapter 6. The specific algorithm presented therein was applied, along with the opposite-side tagging methods, to samples of fully reconstructed B^+ and B^0 decays, in Chapter 7. The algorithm's dilution was measured, for the charged modes, as the observed overall flavor asymmetry and, in the case of the neutral modes, from the analysis of the time-dependent flavor asymmetry. A similar evaluation of the method's performance in samples of B_s mesons was not provided then, as it is forbidden by the rapid mixing oscillations which characterize the system, for the available data samples.

The goal presently is to develop an optimized same-side tagging method, and estimate its performance so it can be applied in the analysis of B_s flavor oscillations. The strategy in algorithm development is to validate the Monte Carlo description of the relevant processes and observables, and subsequently extract the algorithm's performance from Monte Carlo samples. The cornerstone of such validation is to be provided by the successful performance description, verified against data, for the B^+ and B^0 mesons.

The generation of large Monte Carlo samples further offers the possibility of identifying and parameterizing dilution dependencies on relevant quantities. This will result in increased tagging power. For the B_s mesons, as illustrated in Figure 6.6, the leading fragmentation particle is expected to be a kaon. The performance of the method can thus be enhanced through the use of particle identification techniques for identifying kaons among the track candidates. The use of such information when available will accordingly be explored, and the corresponding algorithm is sometimes referred to as same-side *kaon* tagger.

Unlike the study of the tagging methods presented in Chapter 6, where a *binned* approach was employed throughout, the proposed SST algorithms' study benefits presently from the more thorough *unbinned* likelihood technique and description of the samples developed in previous chapters.

9.1 Data and Monte Carlo samples

Samples of fully reconstructed decays of the B^+, B^0 and B_s meson species are studied.

9.1.1 Monte Carlo generation

The Monte Carlo samples ought to contain a complete description of the hadronization process, and of the properties of the resulting fragmentation particles accompanying the B mesons.

Events are generated with `Pythia` [80] (version 6.216); the program's parameter `msel` is set to 1. All of the following QCD high-$p_\perp$ processes are generated: $f_i f_j \to f_i f_j$, $f_i \bar{f}_i \to f_k \bar{f}_k$, $f_i \bar{f}_i \to gg$, $f_i g \to f_i g$, $gg \to f_k \bar{f}_k$ and $gg \to gg$, where $f_{i,j,k}$ stand for fermions and g for gluons. The following $b\bar{b}$ production mechanisms [81] are correspondingly included. The leading process, denoted *flavor creation*, corresponds to the production of a $b\bar{b}$ pair by gluon fusion or by the annihilation of light quarks via the 2-to-2 parton subprocesses $q\bar{q} \to b\bar{b}$ and $gg \to b\bar{b}$. Another production process, referred to as *flavor excitation*, occurs when a virtual heavy quark from the parton distribution of a beam particle is put on mass shell through scattering by a light quark or gluon, via the subprocesses $gb \to gb$ and $qb \to qb$. A third source, denoted *gluon splitting*, comes from reactions with only gluons and light quarks participating in the 2-to-2 hard parton scattering subprocess, and where the $b\bar{b}$ pair is produced in a gluon shower either by a final or initial state gluon. All such processes need to be included in Monte Carlo generation for obtaining an accurate description of data.

The default `Pythia` Lund fragmentation model was used to describe the production and the properties of particles formed in the hadronization process. The energy taken away from the b string by the B meson is described by a symmetric Lund function, whose parameterization has been taken from fits to LEP data.

The default values for B^{**} masses and widths were updated with recent measurements [1]. A fraction of 20% of the B^+ and B^0 mesons was set to have originated from B^{**} decays.

In order to efficiently produce large Monte Carlo samples, the underlying processes from an original sample are kept, while the heavy flavor particles are re-decayed, using the program `EvtGen` [70]. No neutral B meson mixing is included in the simulation.

9.1.2 Selection and fitting model

The criteria for signal reconstruction follow the selection presented in Chapter 4. The unbinned likelihood based fitting model developed in Chapters 5 and 7 is employed.

For selecting B decays in the $J/\psi K$ samples so far, criteria have been employed to ensure that no bias in proper decay time is introduced. Presently, however, a minimum decay distance significance threshold ($L_{xy}/\sigma_{L_{xy}} \gtrsim 4.5$) is imposed on the B candidates in order to further suppress the background contamination. This is relevant for obtaining meaningful comparisons of signal distributions in data to those obtained from Monte Carlo, where mass-sideband subtraction is performed for the former. The proper decay time bias introduced by this selection condition is exactly described in the likelihood model, as derived in Section 5.2 and expressed in (5.17). Figures **??** and **??** show the likelihood projection in mass and proper time spaces for $J/\psi K$ sample decays. The decrease in the combinatorial background level, when comparing to Figure 5.7, is apparent from the mass distributions, as it is from the suppression of the prompt peak in proper decay time.

The general level of agreement of Monte Carlo and data is illustrated in Figure **??** for selected B meson candidates' distributions.

9.2 Pre-selection of track candidates

For being considered as potential tagging candidates tracks must fulfill certain criteria. These have been motivated and outlined in Section 6.2. The specific requirements imposed on the track candidates are the following:

- $\Delta R(\text{track}, B) \leq 0.7$,
- $p_T \geq 450$ MeV/c,
- $|d_0/\sigma_{d_0}| \leq 4$,
- $|\eta| \leq 1$,
- $|\Delta z_0(\text{track}, B)| \leq 1.2$ cm,
- number of hits in Silicon (ϕ) ≥ 3 and in COT ≥ 1,
- electrons and muons are excluded,
- B daughter tracks are excluded.

Figure **??** shows data and Monte Carlo distributions compared for some of the involved quantities, where all cuts but that on the corresponding observable have been applied. In general, a satisfactory level of agreement is observed.

The set of tagging candidates is identified with those tracks which satisfy the above pre-selection criteria. The transverse momentum p_T of the selected candidates is shown in Figure **??** for data and Monte Carlo simulation. The partial distributions for the generator level identified particles are also represented. The origin of the pion and kaon track candidates is found from generator level information and is shown in Figure **??**; right (wrong) sign indicate whether the track has the correct (incorrect) charge correlation with the B flavor expected by the SST.

The set of tagging candidates may contain none, single or multiple elements. Figure **??** shows the tagging candidates multiplicity for the selected decays. When one or multiple tagging candidates have been accepted the corresponding event is assigned to one of the following classes:

Agreeing class: if only one single track has been selected, or the charges of the selected tracks coincide;

Disagreeing class: if not all of the selected tracks have a common charge sign.

Rather roughly, about 60% of the time there's at least one track candidate, and approximately 40% of the time the track candidates have a common charge. For events belonging to the former class, the SST decision is unambiguously given by the identified track(s) charge. For the latter class, a decision needs to be made about which track should be selected as the tagging track. Several algorithm implementations are explored in the coming sections for accomplishing such purpose.

9.3 Candidate algorithms

The aim of the SST method is to identify the charge signal of the leading track, expected to be found *nearest* in phase space to the B meson. When several track candidates are available for a given event, a specific definition of *nearest* track needs to be adopted for selecting the tag. The following implementation possibilities are explored:

1. p_T^{rel}, selects the track with the smallest relative transverse momentum to the direction given by the sum of the track and B momenta,

2. p_L^{rel}, selects the track with the largest relative longitudinal momentum to the direction given by the sum of the track and B momenta,

3. p_T, selects the track with the largest transverse momentum,

4. $M_{B\pi}$, selects the track with the smallest invariant mass with the B,

5. ΔR, selects the track with the smallest ΔR relative to the B,

6. $p_L^{rel}Q$, the decision is given by the sum of the charges of the tag candidate tracks, or, in case this gives a null decision, by that of the p_L^{rel} algorithm above.

The first criterion listed corresponds to that employed in Section 6.2. The definition of the quantities p_T^{rel} and p_L^{rel} is illustrated in Figure 6.7. The corresponding distributions are shown in Figure **??**.

Each of the various listed criteria, which are more or less strongly inter-correlated, offers in principle some discriminating power. In the future, these may therefore be combined along with additional discriminating information through an adequate multivariate mechanism, such as an artificial neural network. Currently nevertheless we are interested in adopting a simpler selection criteria which will be used in data and in Monte Carlo simulation to further explore the tagging related processes associated to the various B meson species.

Figure **??** provides a comparison of the tagging performance of the listed candidate algorithms. These are applied solely to events that contain multiple track candidates with differing charges. For these cases, a decision is always achieved, implying that all algorithms deliver identical tagging efficiencies.

The p_L^{rel} algorithm shows a relatively good performance, and it is the one we decide to elect among the list of candidates above. In particular, we conclude that it performs better than the p_T^{rel} algorithm which has formerly been adopted.

9.4 Parameterized dilution

The parameterization of potential dilution dependences offers several advantages over using an overall, average dilution value. In general and foremost it results in a relative increase of the algorithm's tagging performance. In addition it accounts for variations of the parameterized quantity which may exist between samples, and, as it is here most important, between data and Monte Carlo simulation. If for example less high momentum events would be found in data than in Monte Carlo (for instance due to a prescaling of the trigger not implemented in the simulation) predicted dilutions depending on that quantity would automatically assign to the data sample a smaller (as it would be the case) overall dilution relative to the Monte Carlo sample, while the shape of the predicted (*i.e.* parameterized) dilution wouldn't be affected.

The classification of tagged events into the classes of agreeing and disagreeing charges was introduced anticipating differences in expected characteristic dilutions. Accordingly, dilutions are evaluated and assigned separately to each of those classes. As expected a better performance is found for the agreeing case compared to the disagreeing case.

The SST dilution reveals an expected increase with the transverse momentum p_T of the tagging track. This dependency is shown in Figure **??** for the adopted p_L^{rel} algorithm, for the agreeing and disagreeing classes. It is parameterized as

$$\mathcal{D}(p_T) \quad = \quad \alpha_0 - \alpha_1 \cdot e^{-\alpha_2 \cdot p_T} \; . \tag{9.1}$$

This shape is in general suitable, as the projections in Figure **??** indicate. However, a deviant behavior is observed for the B^0 mesons in the class of disagreeing tracks. This is understood from the generator level information displayed in Figure **??**. In effect, at higher momentum kaons tend to be more often selected as the tagging track, and these contribute negatively to the dilution in the case of the B^0 mesons. We mention in passing that the referred deviant behavior is observed for the p_L^{rel} and p_T algorithms, but not for others, such as the p_T^{rel} implementation. For the disagreeing class in B^0 modes only an average dilution is used, as the available sub-samples size is too small for allowing an accurate description of the resulting shape.

9.5 Kaon identification

The tagging tracks contribute distinctively to the SST dilution depending on the identity of the associated particles. This can already be inferred from the schematic in Figure 6.7, and it is demonstrated by the generator level information displayed in Figures **??** and **??**. In particular, in the case of B_s mesons the correct SST charge correlation is expected to be provided by kaons, while no contribution is expected on average from either pions or protons. The performance of the SST method is in this way expected to benefit considerably from the use of particle identification information which may be available for the tracks.

Particle identification (PID) information for charged tracks is provided at CDF by two means: energy loss dE/dx in the tracking chamber COT, and time of flight measured by the TOF detector. The two measurements are complementary in that the latter is more effective for low momentum tracks and the former is more effective for intermediate momenta. An optimized separation power is obtained by combining the corresponding information,

$$L(\text{particle}) = P_{TOF}(\text{particle}) \cdot P_{dE/dx}(\text{particle}) \qquad \text{with} \quad \text{particle} = K, p, \pi \; ,$$

where P_{TOF} and $P_{dE/dx}$ are the probability density functions which have been determined on pure particle samples. We use the following combined likelihood ratio, defined for the kaon hypothesis,

$$\log(LH(PID)) = \log\left(\frac{P_{TOF}(K)\cdot P_{dE/dx}(K)}{f_p\cdot P_{TOF}(p)\cdot P_{dE/dx}(p) + f_\pi\cdot P_{TOF}(\pi)\cdot P_{dE/dx}(\pi)}\right), \quad (9.2)$$

where $f_p = 0.1$ and $f_\pi = 0.9$ are the prior probabilities for background composition. In case no TOF or no dE/dx information is available, the corresponding probabilities are taken as 1. The dE/dx, TOF, and combined likelihoods are illustrated in Figures **??** and **??**.

With particle identification information available for the track candidates, an alternative criterion for selecting the tagging track among multiple candidates is proposed: select the track with the highest probability for being a kaon. This will be referred to as the PID algorithm, and it is proposed having specifically in mind the B_s case. Although it is clearly not optimal for the B^+ and B^0 cases, we extend its study to these high statistics B meson samples to further test the Monte Carlo description.

The algorithms performance is studied as a function of the value of the PID variable for the tagging track. The observed dilution dependences are shown in Figure **??** for the classes of agreeing and disagreeing track candidates. For the B^+ we see high dilution both for more pion like and more kaon like tracks. For the B^0, the pions are likely to provide the expected SST charge-flavor correlation (positive dilution), while kaons are more likely anti-correlated (negative dilution). The pattern is somewhat inverted for the B_s, where kaons are potentially good tagging tracks; pions do not carry in this case any information on the B production flavor (zero dilution). The decrease of predicted dilution at very hight kaon probability values is related to the fact that both particle ID systems (dE/dx and TOF) give best kaon-pion separation for low momentum tracks. Thus the high kaon probability tracks are most likely true low momentum kaons which are coming from the underlying event. However leading fragmentation tracks are more likely to have higher momenta.

9.6 Tagging performance

The tagging performance is evaluated using the unbinned likelihood framework developed in previous chapters, applied to the exclusive decay samples indicated above.

The results are summarized in Tables 9.1 and 9.2 for the p_L^{rel} and PID algorithms, respectively. The tagging efficiency is denoted by ϵ, and the average dilution by $\mathcal{D}$, which are floating parameters of the fit.

The evaluation of the tagging performance with the parameterized dilution proceeds as follows. The value of the dilution is predicted for each event according to the classification and dependencies which were derived in the previous sections and obtained from the corresponding Monte Carlo samples. The predicted dilution values are provided as input to the fit. Correspondingly, dilution templates for signal and background are derived and used in the likelihood model. The procedure is identical to that used in Chapter 7 for the purpose of OST calibration. A dilution scale factor $S_{\mathcal{D}}$ is introduced as a floating parameter of the fit, which directly multiplies the dilution in the proper decay time PDF for flavor tagged events. Finally, the effective dilution $\mathcal{D}_{\rm eff}$ is obtained

$$\mathcal{D}_{\rm eff} = S_{\mathcal{D}}\sqrt{\langle\mathcal{D}^2\rangle}\ , \tag{9.3}$$

from the fitted scale factor value, and the dilution squared average over signal events.

p_L^{rel} algorithm [%]		B^+ $J/\psi K^+$	B^+ $\bar{D}^0\pi^+$	B^0 $J/\psi K^{*0}$	B^0 $D^-\pi^+$	B_s $J/\psi\phi$	B_s $D_s^-\pi^+$
MC	ϵ	55.2±0.1	55.9±0.1	54.4±0.1	56.6±0.1	49.5±0.3	52.1±0.3
	$\mathcal{D}$	25.3±0.2	25.7±0.3	14.9±0.3	15.1±0.4	16.2±0.8	19.0±0.8
	$S_{\mathcal{D}}$	100.3±0.7	100.3±0.9	101.1±1.8	98.0±2.1	99.7±4.4	100.5±3.3
	$\mathcal{D}_{\rm eff}$	28.5±0.2	29.0±0.3	16.8±0.3	17.2±0.4	18.6±0.8	22.8±0.7
	$\epsilon\mathcal{D}^2_{\rm eff}$	4.5±0.1	4.7±0.1	1.5±0.1	1.7±0.1	1.7±0.2	2.7±0.2
data	ϵ	60.0±0.7	58.4±0.5	57.3±1.0	57.2±0.6	48.1±2.9	49.3±2.3
	$\mathcal{D}$	22.3±1.9	25.9±1.4	10.7±4.6	13.3±2.9	—	—
	$S_{\mathcal{D}}$	91.8±6.3	103.6±4.7	85.9±29.1	110.0±16.0	—	—
	$\mathcal{D}_{\rm eff}$	26.4±2.1	30.4±1.3	13.6±5.4	19.0±2.5	18.6±1.0	23.7±1.0
	$\epsilon\mathcal{D}^2_{\rm eff}$	4.2±0.6	5.4±0.5	1.0±0.7	2.1±0.6	1.7±0.2	2.8±0.3

Table 9.1: Tagging performance of the p_L^{rel} algorithm in Monte Carlo and data; the quoted uncertainties are statistical only.

PID algorithm [%]		B^+ $J/\psi K^+$	B^+ $\bar{D}^0\pi^+$	B^0 $J/\psi K^{*0}$	B^0 $D^-\pi^+$	B_s $J/\psi\phi$	B_s $D_s^-\pi^+$
MC	ϵ	55.2±0.1	55.9±0.1	54.4±0.1	56.6±0.1	49.5±0.3	52.1±0.3
	$\mathcal{D}$	24.2±0.3	24.5±0.3	12.7±0.3	12.9±0.4	18.9±0.8	21.8±0.8
	$S_{\mathcal{D}}$	100.0±0.7	100.0±0.9	98.0±1.7	98.5±2.1	98.8±3.6	98.8±2.5
	$\mathcal{D}_{\rm eff}$	27.0±0.2	27.5±0.2	17.9±0.3	17.4±0.4	22.3±0.8	28.5±0.7
	$\epsilon\mathcal{D}^2_{\rm eff}$	4.0±0.1	4.2±0.1	1.7±0.1	1.7±0.1	2.5±0.2	4.2±0.2
data	ϵ	60.0±0.7	58.4±0.5	57.3±1.0	57.2±0.6	48.1±2.9	49.3±2.3
	$\mathcal{D}$	21.7±1.9	25.4±1.4	10.7±4.6	14.2±2.9	—	—
	$S_{\mathcal{D}}$	87.6±6.7	95.1±4.7	78.8±25.1	104.7±14.4	—	—
	$\mathcal{D}_{\rm eff}$	23.9±2.2	25.7±1.3	13.7±5.5	17.6±2.3	22.1±1.0	28.3±1.1
	$\epsilon\mathcal{D}^2_{\rm eff}$	3.4±0.5	3.9±0.4	1.1±0.7	1.8±0.5	2.4±0.3	4.0±0.4

Table 9.2: Tagging performance of the PID algorithm in Monte Carlo and data; the quoted uncertainties are statistical only.

A relatively better performance is in general obtained with the PID algorithm for the B_s mesons, and with the p_L^{rel} algorithm for the $B^{+,0}$ samples. It should also be clear that their combination, along with the use of additional tracks' information, will potentially result in further improvements. The introduction of dilution parameterizations induces an absolute gain of about 2-5% in dilution. The observation that the dilution scale factors are close to unity further indicates that those parameterizations provide adequate descriptions in both Monte Carlo and data.

The tagging dilution is not *measured* for the B_s *data* samples. The reason, to re-state it, being that the rapid flavor oscillations do not allow for a direct measurement of the unknown oscillation frequency, as it is done for the B^0 system, given current sample sizes, resolutions and flavor taggers. The strategy therefore is to use the Monte Carlo predictions as estimates of the tagger performance in data. In Tables 9.1 and 9.2, the dilution and tagging power quoted for the B_s samples use obtained accordingly employing the predicted dilution and scale factor from simulation. For the hadronic B_s mixing data sample the tagging power of the same side kaon (PID) algorithm implemented is

$$\epsilon\mathcal{D}^2 = 4.0 \pm 0.4 \ .$$

The evaluated uncertainty corresponds to statistical effects. Systematic variations associated to fragmentation and simulation processes in the Monte Carlo are necessary for assessing the full dilution uncertainty. The latter is particularly relevant in the absence of a signal in the mixing sample, for the purpose of evaluating frequency exclusion conditions.

9.7 Systematic uncertainties

An appropriate understanding of the systematic effects associated to the Monte Carlo predictions is crucial for employing the same side tagger in the analysis of flavor oscillations in the B_s data samples. The systematics studies are performed for all B meson species, for both average dilution and parameterized dilution algorithm implementations. For the latter case, in which the dilution is predicted for each event from the Monte Carlo parameterizations achieved in earlier sections, the systematic uncertainties are evaluated for a corresponding overall dilution scale factor, $S_{\mathcal{D}}$. Both of these quantities, namely the dilution parameterizations and the dilution scale factors, constitute the SST information to be provided to the B_s mixing analysis.

Systematic studies related to variations of underlying physics processes have been performed by either re-weighting or filtering out events from the original Monte Carlo samples. Systematic variations to the particle ID response in the detector or multiple interactions have been studied by re-simulating those aspects of the original Monte Carlo events.

The following sources of systematic uncertainties are investigated.

Production mechanisms: The contributions of the $b\bar{b}$ pair production processes – flavor creation (FC), flavor excitation (FE), gluon splitting (GS) – contain uncertainties which can have influence in flavor tagging. The fractions of reconstructed B events in our nominal Monte Carlo which come from FC, FE and GS are about 26%, 55% and 19%, respectively. The most distinguishing variable for those processes is the angular difference $\Delta\phi$ between the signal and opposite-side B directions. While FC and FE mainly produce B hadrons back-to-back, B hadrons from GS processes point more often in the same direction. In the context of same-side tagging, opposite-side B daughters and fragmentation tracks are more likely to disturb the tagger for GS events than the two other processes. Opposite-side jets for this purpose are found with the jet charge tagger mechanism (Section 6.1.2) for identifying secondary vertices inside jets (JVX algorithm). By fitting Monte Carlo $\Delta\phi$ distributions of the different processes to the $\Delta\phi$ distribution in data, the following range for systematic variations has been determined: GS fraction within [-68%,+46%], FE and FC fractions within [-50%,+50%] relative to their nominal appearance.

Fragmentation function: The Lund string fragmentation model has been chosen to produce the nominal default Monte Carlo samples. This has been tuned by many experiments, and describes well the LEP data, as well as our own to the extent we have tested it. Variations are implemented to the fragmentation function,

$$f_L(z) \;\propto\; \frac{1}{z}\,(1-z)^a\, e^{-\frac{B}{z}} \qquad \text{with} \quad z = \frac{E^B + p_L^B}{E^b + p_L^b}\ ,$$

where a and B are free parameters of the model, and z relates the energy E and the longitudinal momentum p_L of the B meson and the original b quark. Nominally, the chosen parameter values for the pair (a, B) are $(1.68, 15.6)$. From a simultaneous fit to several distributions in data and Monte Carlo which are most sensitive to the fragmentation function, such as track multiplicity, transverse momentum of the B and of the fragmentation tracks, the allowed parameter space for the parameters of the Lund function f_L has been determined. The chosen systematic variations correspond to the (a, B) pair values of $(1, 10)$, $(3, 22)$ and $(9, 55)$, and are represented in Figure **??**.

Variations within data statistics: Fragmentation determines the formation of hadrons out of the string and therefore event properties such as the multiplicity, flavor and momenta of the tracks around the B meson, along with the latter's momentum. By varying the Monte Carlo sample within the ranges allowed by the statistical uncertainties on corresponding distributions in data, a model independent estimation of the associated description is obtained. Systematic uncertainties are assigned by re-weighting the Monte Carlo events according to the corresponding variations for the following selected observables: transverse momentum of the B candidate, number of tagging track candidates, p_L^{rel} and PID variable of the selected tagging track. This estimation encompasses effects not only from fragmentation but also from other sources potentially inducing variations in those distributions.

Particle content around the B meson: The fragmentation process determines the particle species, and respective rates, produced in association with the B meson. A measurement of the content of stable charged particles around B mesons has been carried out in the high statistics lepton and displaced track trigger sample [62] . The observed rates of kaons, pions, and protons agree well between data and Monte Carlo for B^+ and B^0; no systematic variation related to this effect is thus performed for these mesons. For the B_s the fraction of kaons we find in data is $20.2 \pm 1.4\%$ compared to $23.6 \pm 0.2\%$ in the Monte Carlo sample. As a systematic variation the kaon fraction is effectively reduced to 19.5%. This is done firstly by re-weighting the events with a kaon as the selected tagging track. An additional systematic contribution is computed by re-weighting only those events for which the selected kaon originates directly from the b string; in this case, the assigned uncertainty corresponds to half of the variation. In addition to the stable charged particles, we checked for resonances and vector particles such as ϕ, K_S and K^*. Their content in the Monte Carlo samples is varied by factors of 1.5 and 0.5; the largest deviations in tagging performance are assigned as an additional systematics contribution.

Rate of B^{**}: The decay of excited B-mesons to B^+ is accompanied by charged pions and kaons with the correct SST flavor correlation. Also, B^0 mesons that come from the decay of B^{**} can be accompanied by pions, again with the correct information of the b-quark flavor at the time of production. In the case of B_s mesons coming from B^{**} no accompanying charged particles are produced in the decay. Therefore, tagging on the kaon or pion from B^{**} decays increases the tagging performance, both efficiency and dilution, for B^+ and B^0. The size of this increase is potentially reduced if, instead of tagging on tracks coming from the B^{**} decay, the associated B^{**} fragmentation tracks are selected, as these provide the opposite charge correlation. The fraction of B^+ and B^0 from B^{**} decays in the nominal Monte Carlo samples has been chosen to be 20% according to recent LEP measurements. For evaluating the systematics associated with the B^{**} contributions, rate variations within 16-35% are considered, based upon the largest and smallest single experiment measurements ($\pm 1\sigma$) available from CDF Run I and LEP. The B^{**} contribution may be further explored for tagging purposes, for example employing the observable $M(B\pi) - M(B) - M(\pi)$ to check for the narrow B^{**} states.

Multiple interactions: The required criteria on impact parameter significance and $\Delta z_0(B, track)$ imposed for pre-selecting tagging candidates lead to the rejection of most tracks coming from multiple interactions, rendering their effect on the tagging performance relatively small. The overall remaining fraction of such tracks is estimated to be about

0.22%. This rate has been further determined for low and high luminosity events, obtaining fractions of about 0.18% and 0.40%, respectively. Such rates of additional potential tagging tracks have been added to the Monte Carlo sample for estimating the associated systematic uncertainty.

Particle identification, dE/dx & TOF: Particle identification is based on specific ionization and flight time measurements, respectively in the drift chamber and the TOF. Both detector subsystems have been calibrated in data, and PDFs of their particle identification response have been modelled using input from data. We varied the TOF and dE/dx resolution parameterizations and efficiencies in the simulation according to measurement uncertainties obtained in the data. The distributions of the tagging tracks multiplicity and of the combined PID variable including systematic variations are displayed in Figure **??**.

The evaluated systematic uncertainties are summarized in Table 9.4 for the tagging efficiency and in Table 9.5 for the tagging dilution and associated scale factor. The dominant dilution uncertainties for the B^+ and B^0 mesons are associated to the B^{**} rates, while for the B_s meson these come from variations of the particles content found in its vicinity.

A very good agreement is verified for the dilution values measured in data and in Monte Carlo, for the B^+ and B^0 mesons, within the combined statistical and systematic uncertainties. As this statement cannot be made more precise than the uncertainties on the data and Monte Carlo samples, the combined uncertainties from the B^+ and B^0 cases, averaged over the four modes studied, have been added to the final result for the B_s meson. These additionally assigned uncertainty contributions are listed under the column "data/MC agreement" in Table 9.3, which summarizes the results of the predicted dilutions for the B_s meson obtained from $D_s\pi$ Monte Carlo samples.

The final SST performance results are summarized for the studied algorithms in Table 9.6, for both Monte Carlo and data, and for all three B meson species. Results relative to Monte Carlo (MC) include statistical and systematic effects; for the B_s case, "data/MC agreement" uncertainties for the dilution are also included. Results relative to data include statistical uncertainties; for the B_s case, the values of $\mathcal{D}$ and $S_{\mathcal{D}}$ are taken from MC, along with the respective combined uncertainties, while ϵ and $\langle\mathcal{D}^2\rangle$ (9.3) are obtained from data.

SST dilution for B_s [%]		fitted value	statistical MC	systematic MC	agreement data/MC	uncertainty total
p_L^{rel}	average, $\mathcal{D}$	19.0	± 0.8	$^{+1.9}_{-4.4}$	± 2.1	$^{+2.9}_{-4.9}$
	parameterized, $S_\mathcal{D}$	100.5	± 3.3	$^{+8.6}_{-17.7}$	± 5.5	$^{+10.7}_{-18.8}$
PID	average, $\mathcal{D}$	21.8	± 0.8	$^{+2.7}_{-5.2}$	± 1.6	$^{+3.2}_{-5.5}$
	parameterized, $S_\mathcal{D}$	98.8	± 2.5	$^{+7.1}_{-11.8}$	± 7.6	$^{+10.7}_{-14.3}$

Table 9.3: Summary of SST dilution predicted from Monte Carlo for the B_s meson.

SST efficiency systematics	B^+		B^0		B_s	
[%]	$J/\psi K^+$	$\bar{D}^0\pi^+$	$J/\psi K^{*0}$	$D^-\pi^+$	$J/\psi\phi$	$D_s^-\pi^+$
fragmentation function	+2.5	+2.7	+2.6	+3.0	+2.3	+3.3
production mechanisms	$^{+0.7}_{-0.7}$	$^{+0.6}_{-0.5}$	$^{+0.7}_{-0.7}$	$^{+0.7}_{-0.5}$	$^{+0.6}_{-0.7}$	$^{+0.8}_{-0.7}$
multiple interactions	+0.1	+0.1	+0.1	+0.1	±0.2	±0.1
rate of B^{**}	$^{+1.8}_{-0.5}$	$^{+1.8}_{-0.5}$	$^{+2.0}_{-0.6}$	$^{+2.0}_{-0.5}$	–	–
variation within data statistics	$^{+4.3}_{-0.2}$	$^{+3.4}_{-0.0}$	$^{+3.4}_{-0.1}$	$^{+3.7}_{-0.0}$	$^{+7.9}_{-0.3}$	$^{+2.0}_{-4.4}$
total systematic uncertainty	+5.3 -0.9	+4.8 -0.7	+4.8 -0.9	+5.2 -0.7	+8.3 -0.8	+3.9 -4.4

Table 9.4: Systematic uncertainties of the SST algorithms' efficiency.

SST dilution systematics [%]	$J/\psi K^+$	$\bar{D}^0\pi^+$	$J/\psi K^{*0}$	$D^-\pi^+$	$J/\psi\phi$	$D_s^-\pi^+$
p_L^{rel} algorithm, average dilution $\mathcal{D}$						
production mechanisms	$^{+0.0}_{-0.2}$	$^{+0.2}_{-0.3}$	$^{+0.3}_{-0.5}$	$^{+0.3}_{-0.5}$	$^{+0.3}_{-0.7}$	$^{+0.3}_{-0.3}$
fragmentation function	$^{+0.5}_{-0.0}$	$^{+0.7}_{-0.0}$	$^{+0.0}_{-0.1}$	$^{+0.1}_{-0.1}$	$^{+0.7}_{-0.0}$	$^{+1.7}_{-0.0}$
multiple interactions	$^{+0.0}_{-0.1}$	$^{+0.1}_{-0.0}$	$^{+0.1}_{-0.1}$	$^{+0.0}_{-0.1}$	$^{+0.0}_{-0.2}$	$^{+0.1}_{-0.0}$
rate of B^{**}	$^{+2.2}_{-0.6}$	$^{+2.3}_{-0.6}$	$^{+2.6}_{-0.7}$	$^{+2.5}_{-0.6}$	–	–
particle content around B_s	–	–	–	–	$^{+0.1}_{-3.1}$	$^{+0.4}_{-3.6}$
variation within data statistics	$^{+1.1}_{-0.6}$	$^{+1.2}_{-0.4}$	$^{+0.6}_{-0.7}$	$^{+0.4}_{-0.4}$	$^{+0.5}_{-4.0}$	$^{+0.7}_{-2.4}$
total systematic uncertainty	$^{+2.5}_{-0.9}$	$^{+2.7}_{-0.8}$	$^{+2.7}_{-1.1}$	$^{+2.5}_{-0.9}$	$^{+0.9}_{-5.1}$	$^{+1.9}_{-4.4}$
p_L^{rel} algorithm, dilution scale factor $S_{\mathcal{D}}$						
production mechanisms	$^{+0.0}_{-0.9}$	$^{+0.6}_{-1.1}$	$^{+0.4}_{-2.9}$	$^{+1.3}_{-2.9}$	$^{+1.1}_{-2.5}$	$^{+0.0}_{-2.2}$
fragmentation function	$^{+1.8}_{-0.5}$	$^{+2.2}_{-0.1}$	$^{+0.2}_{-0.8}$	$^{+0.3}_{-0.6}$	$^{+6.0}_{-0.0}$	$^{+6.3}_{-0.0}$
multiple interactions	$^{+0.0}_{-0.4}$	$^{+0.2}_{-0.1}$	$^{+2.8}_{-0.0}$	$^{+0.1}_{-0.9}$	$^{+0.0}_{-2.9}$	$^{+0.3}_{-0.0}$
rate of B^{**}	$^{+7.7}_{-1.8}$	$^{+10.2}_{-1.7}$	$^{+17.4}_{-4.8}$	$^{+14.8}_{-4.3}$	–	–
particle content around B_s	–	–	–	–	$^{+5.0}_{-18.3}$	$^{+5.4}_{-16.6}$
variation within data statistics	$^{+1.0}_{-0.1}$	$^{+1.4}_{-0.0}$	$^{+1.7}_{-2.0}$	$^{+1.0}_{-1.9}$	$^{+1.7}_{-11.0}$	$^{+2.2}_{-5.9}$
total systematic uncertainty	$^{+0.8}_{2.1}$	$^{+10.5}_{-2.0}$	$^{+17.7}_{-6.0}$	$^{+14.9}_{-5.6}$	$^{+8.0}_{-21.6}$	$^{+8.6}_{-17.7}$
PID algorithm, average dilution $\mathcal{D}$						
production mechanisms	$^{+0.0}_{-0.3}$	$^{+0.1}_{-0.5}$	$^{+0.2}_{-0.6}$	$^{+0.4}_{-0.5}$	$^{+0.5}_{-0.6}$	$^{+0.0}_{-0.4}$
fragmentation function	$^{+0.4}_{-0.7}$	$^{+0.9}_{-0.2}$	$^{+0.0}_{-0.2}$	$^{+0.0}_{-0.2}$	$^{+0.9}_{-0.0}$	$^{+1.4}_{-0.0}$
multiple interactions	$^{+0.1}_{-0.1}$	$\pm\, 0.0$	$^{+0.0}_{-0.1}$	$^{+0.1}_{-0.0}$	$^{+0.0}_{-0.1}$	$^{+0.1}_{-0.0}$
rate of B^{**}	$^{+2.1}_{-0.7}$	$^{+2.1}_{-0.6}$	$^{+2.4}_{-0.7}$	$^{+2.5}_{-0.6}$	–	–
particle content around B_s	–	–	–	–	$^{+1.2}_{-3.1}$	$^{+1.2}_{-4.7}$
variation within data statistics	$^{+0.8}_{-1.1}$	$^{-0.7}_{+1.1}$	$^{+0.2}_{-2.6}$	$^{+0.4}_{-0.7}$	$^{+2.4}_{-3.6}$	$^{+2.0}_{-2.3}$
PID simulation and resolution	$^{+0.1}_{-0.4}$	$^{+0.1}_{-0.4}$	$^{+0.1}_{-0.1}$	$^{+0.1}_{-0.1}$	$^{+0.0}_{-0.5}$	$^{+0.3}_{-0.2}$
total systematic uncertainty	$^{+2.3}_{-1.6}$	$^{+2.5}_{-1.1}$	$^{+2.4}_{-1.9}$	$^{+2.6}_{-1.1}$	$^{+2.9}_{-4.8}$	$^{+2.7}_{-5.2}$
PID algorithm, dilution scale factor $S_{\mathcal{D}}$						
production mechanisms	$^{+0.0}_{-0.8}$	$^{+0.6}_{-1.2}$	$^{+0.1}_{-2.0}$	$^{+0.3}_{-1.8}$	$^{+2.0}_{-3.2}$	$^{+0.6}_{-1.5}$
fragmentation function	$^{+2.1}_{-0.1}$	$^{+2.5}_{-0.0}$	$^{+0.3}_{-0.4}$	$^{+2.4}_{-0.6}$	$^{+3.0}_{-0.0}$	$^{+4.4}_{-0.0}$
multiple interactions	$^{+0.5}_{-0.6}$	$^{+0.2}_{-0.1}$	$^{+0.1}_{-0.7}$	$^{+0.1}_{-0.5}$	$^{+0.2}_{-0.8}$	$^{+0.2}_{-0.2}$
rate of B^{**}	$^{+9.1}_{-2.5}$	$^{+9.5}_{-2.4}$	$^{+12.1}_{-3.9}$	$^{+14.6}_{-4.1}$	–	–
particle content around B_s	–	–	–	–	$^{+5.8}_{-14.1}$	$^{+4.0}_{-9.6}$
variation within data statistics	$^{+2.3}_{-0.6}$	$^{+2.7}_{-0.0}$	$^{+1.4}_{-2.0}$	$^{+2.4}_{-2.2}$	$^{+3,1}_{-11.4}$	$^{+1.6}_{-4.3}$
PID simulation and resolution	$^{+1.6}_{-3.1}$	$^{+1.7}_{-3.4}$	$^{+1.8}_{-4.1}$	$^{+2.2}_{-3.1}$	$^{+2.0}_{+6.1}$	$^{+3.6}_{-4.8}$
total systematic uncertainty	$^{+9.8}_{-4.1}$	$^{+10.0}_{-4.3}$	$^{+12.3}_{-4.8}$	$^{+15.2}_{-5.9}$	$^{+7.7}_{-19.4}$	$^{+7.1}_{-11.8}$

Table 9.5: Systematic uncertainties of the SST algorithms' dilution.

SST [%]		B^+ $J/\psi K^+$	B^+ $\bar{D}^0\pi^+$	B^0 $J/\psi K^{*0}$	B^0 $D^-\pi^+$	B_s $J/\psi\phi$	B_s $D_s^-\pi^+$
MC	ϵ	$55.2^{+5.3}_{-0.9}$	$55.9^{+4.8}_{-0.7}$	$54.4^{+4.8}_{-0.9}$	$56.6^{+5.2}_{-0.7}$	$49.5^{+8.3}_{-0.9}$	$52.1^{+3.9}_{-4.4}$
data	ϵ	60.0±0.7	58.4±0.5	57.3±1.0	57.2±0.6	48.1±2.9	49.3±2.3

p_L^{rel} algorithm, average dilution

MC	$\mathcal{D}$	$25.3^{+2.5}_{-0.9}$	$25.7^{+2.7}_{-0.9}$	$14.9^{+2.7}_{-1.1}$	$15.1^{+2.5}_{-1.0}$	$16.2^{+2.4}_{-5.6}$	$19.0^{+2.9}_{-4.9}$
	$\epsilon\mathcal{D}^2$	$3.5^{+0.8}_{-0.3}$	$3.7^{+0.8}_{-0.3}$	$1.2^{+0.5}_{-0.2}$	$1.3^{+0.4}_{-0.2}$	$1.3^{+0.4}_{-0.9}$	$1.9^{+0.6}_{-1.0}$
data	$\mathcal{D}$	22.3±1.9	25.9±1.4	10.7±4.6	13.3±2.9	–	–
	$\epsilon\mathcal{D}^2$	3.0±0.5	3.9±0.4	0.7±0.6	1.0±0.4	$1.3^{+0.4}_{-0.9}$	1.8±0.5

p_L^{rel} algorithm, parameterized dilution

MC	$S_{\mathcal{D}}$	$100.3^{+8.0}_{-2.2}$	$100.3^{+10.5}_{-2.2}$	$101.1^{+17.8}_{-6.3}$	$98.0^{+15.0}_{-6.0}$	$99.7^{+10.6}_{-22.7}$	$100.5^{+18.8}_{-14.3}$
	$\mathcal{D}_{\text{eff}}$	$28.5^{+2.3}_{-0.6}$	$29.0^{+3.0}_{-0.6}$	$16.8^{+3.0}_{-1.0}$	$17.3^{+2.6}_{-1.1}$	$18.6^{+2.0}_{-4.2}$	$22.8^{+2.4}_{-4.3}$
	$\epsilon\mathcal{D}^2$	$4.5^{+0.8}_{-0.2}$	$4.7^{+1.1}_{-0.2}$	$1.5^{+0.6}_{-0.2}$	$1.7^{+0.5}_{-0.2}$	$1.7^{+0.5}_{-0.8}$	$2.7^{+0.6}_{-1.0}$
data	$S_{\mathcal{D}}$	91.8±6.3	103.6±4.7	85.9±29.1	110.0±16.0	–	–
	$\mathcal{D}_{\text{eff}}$	26.4±1.8	30.5±1.4	13.7±4.6	19.3±2.8	$18.6^{+2.1}_{-4.3}$	$23.7^{+2.6}_{-4.5}$
	$\epsilon\mathcal{D}^2$	4.2±0.6	5.4±0.5	1.0±0.7	2.1±0.6	$1.7^{+0.4}_{-0.8}$	$2.8^{+0.6}_{-0.8}$

PID algorithm, average dilution

MC	$\mathcal{D}$	$24.2^{+2.3}_{-1.6}$	$24.5^{+2.5}_{-1.1}$	$12.7^{+2.4}_{-1.9}$	$12.9^{+2.6}_{-1.2}$	$18.9^{+3.4}_{-5.2}$	$21.8^{+3.2}_{-5.5}$
	$\epsilon\mathcal{D}^2$	$3.2^{+0.7}_{-0.4}$	$3.4^{+0.7}_{-0.3}$	$0.9^{+0.3}_{-0.3}$	$0.9^{+0.4}_{-0.2}$	$1.8^{+0.7}_{-1.0}$	$2.5^{+0.8}_{-1.3}$
data	$\mathcal{D}$	21.7±1.9	25.4±1.4	10.7±4.6	14.2±2.9	–	–
	$\epsilon\mathcal{D}^2$	2.8±0.5	3.8±0.4	0.7±0.6	1.2±0.5	$1.7^{+0.6}_{-1.0}$	$2.3^{+0.7}_{-1.2}$

PID algorithm, parameterized dilution

MC	$S_{\mathcal{D}}$	$100.0^{+9.8}_{-4.2}$	$100.0^{+10.0}_{-4.4}$	$98.0^{+12.4}_{-5.1}$	$98.5^{+15.3}_{-6.3}$	$98.8^{+11.4}_{-21.1}$	$98.8^{+10.7}_{-14.3}$
	$\mathcal{D}_{\text{eff}}$	$27.1^{+2.7}_{-1.1}$	$27.5^{+2.8}_{-1.2}$	$17.9^{+2.3}_{-0.9}$	$17.1^{+2.7}_{-1.1}$	$22.3^{+2.6}_{-4.8}$	$28.5^{+3.1}_{-4.1}$
	$\epsilon\mathcal{D}^2$	$4.0^{+0.9}_{-0.3}$	$4.2^{+0.9}_{-0.4}$	$1.7^{+0.5}_{-0.2}$	$1.7^{+0.5}_{-0.2}$	$2.5^{+0.7}_{-1.1}$	$4.2^{+1.0}_{-1.3}$
data	$S_{\mathcal{D}}$	87.6±6.7	95.1±4.7	78.8±25.1	104.7±14.4	–	–
	$\mathcal{D}_{\text{eff}}$	23.9±1.8	25.7±1.3	13.7±4.4	17.6±2.4	$22.2^{+2.6}_{-4.8}$	$28.3^{+3.2}_{-4.2}$
	$\epsilon\mathcal{D}^2$	3.4±0.5	3.9±0.4	1.1±0.7	1.8±0.5	$2.4^{+0.6}_{-1.0}$	$4.0^{+0.9}_{-1.2}$

Table 9.6: Summary of SST performance with total uncertainties, in units of percent.

9.8 Résumé

The same-side tagging method has been developed, optimized, and calibrated. Particle identification techniques, based on energy loss and time of flight are explored as part of the algorithms.

The SST method is based on flavor charge correlations between the B candidate and tracks found in its vicinity. It further differs from the opposite-side tagging methods in that the tagging performance depends on the B meson species. Unlike those OST methods, the SST performance cannot thus be measured on data samples of B^+ and B^0 decays and transferred directly as input to the B_s analysis. The strategy adopted consists of performing a thorough performance estimation based on Monte Carlo samples. The procedure involves the validation of the Monte Carlo simulation, through comparison against data of various relevant distributions. Systematic variations of fragmentation and other simulation processes should be also implemented. The successful description of the tagging properties achieved in Monte Carlo is ultimately verified in data samples of B^+ and B^0 mesons.

Various algorithms are initially explored. Two such implementations are elected and their performance fully evaluated. One of these, denoted p_L^{rel} algorithm, is based on pure kinematical quantities, and is the chosen method for B^+ and B^0 meson samples. The other, denoted PID algorithm, further uses particle identification information for selecting kaon tracks expected to be associated, from fragmentation, with B_s mesons. The tagging performance is summarized as:

algorithm		tagging power, $\epsilon\mathcal{D}^2$ [%]
p_L^{rel}	B^+	5.4 ± 0.5
	B^0	2.1 ± 0.6
PID	B_s	$4.0 \,^{+\,0.9}_{-\,1.2}$

The SST performance achieved is considerably superior to that of the OST methods. A combined tagging power of above 5% for B_s is reached. This determines a dramatic increase in the sensitivity of the samples, which is fully explored in the next chapter.