CDF/ANAL/BOTTOM/CDFR/11148
Version 1.0
December 5, 2014

Measurement of the branching fraction ratio of $\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})/\mathcal{B}(B_s^0 \to D_s^- \pi^+)$

P. Marino, M. J. Morello
INFN Pisa and Scuola Normale Superiore

G. Punzi
INFN Pisa and University of Pisa

F. Ruffini
INFN Pisa and University of Siena

Abstract

This note describes the measurement of the ratio of branching fractions of $\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})/\mathcal{B}(B_s^0 \to D_s^- \pi^+)$ with the full data sample corresponding to an integrate luminosity of 9.3fb^{-1}. The value obtained for the ratio of branching fractions $\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})/\mathcal{B}(B_s^0 \to D_s^- \pi^+) = 0.0777 \pm 0.0079\,(\text{stat}) \pm 0.0026\,(\text{sys})$ is compatible with the state-of-art measurements and with the theoretical prediction.

Contents

1 Introduction

The decay mode $B_s^0 \to D_s^{\pm} K^{\mp}$ is interesting because it can be used to extract the CKM-matrix γ angle, with a time-dependent, flavor-tagged measurement. The relative weak phase between $B_s^0 \to D_s^- K^+$ and $B_s^0 \to D_s^+ K^-$ is the CKM γ angle, or rather $\arg(-V_{cs}V_{ub}^*/V_{us}V_{cb}^*)$ as shown in figure 1. The precise value of CKM γ angle is one of the remaining open question in flavour physics.

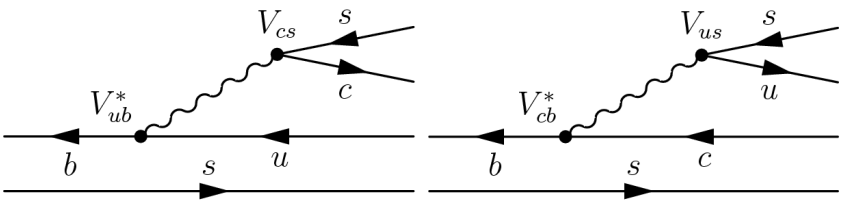

Figure 1: Diagram contributing to $B_s^0 \to D_s^{\pm} K^{\mp}$. The relative phase between the two diagrams is $\gamma = \arg(-V_{cs}V_{ub}^*/V_{us}V_{cb}^*)$.

In this note we describe a measurement of $\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})/\mathcal{B}(B_s^0 \to D_s^- \pi^+)$. An analysis of the $B_s^0 \to D_s^{\pm} K^{\mp}$ decays has been performed at CDF in 2007 with a sample of integrated luminosity correspoding to 1.2fb^{-1} [3].

2 Data sample

The analysis uses the data collected by the `B_CHARM` trigger between December 2004 (run 190697) to September 2011 (run 312510), corresponding to the follow period range P1-P30.[1] We produced BSt-ntuple using the official BottomMods executable v6.1.4.m and we selected the candidates from the `Bs-DsPi-PhiPi` block of BSt-ntuple. We run the executable over the following SAM datasets: `xbhdih, xbhdii, xbhdij, xbhdik, xbhdfm, xbhdfn` and `xbhdfp`. We used the standard good run list following the prescription of the B-Group (Good Run list V45, `goodrun_b_bs_nocal_nomu.list`). The integrated luminosity of the sample is about 9.3 fb^{-1}.

The selection cuts, listed in table 1, have been applied to the data sample.

Table 1: Preselection cuts used applied to the candidates.

$B_s^0 \to D_s^- \pi^+$	unit	selection
$\chi^2_{2D}(B_s^0)$	–	< 40
$\chi^2_{3D}(B_s^0)$	–	< 100
$L_{xy}(B_s^0)$	μm	> 0
$p_T(B_s^0)$	GeV/c	> 5
$\lvert d_0(B_s^0)\rvert$	μm	< 100
$L_{xy}(D_s)_{B_s^0}$	μm	> 0
$\Delta R(D_s, \pi_{B_s^0})$	–	< 2
$p_T(\pi_{B_s^0})$	GeV/c	> 2
$\lvert\eta(\text{all tracks})\rvert$	–	< 1
$m(KK)$	GeV/c^2	$[1.012, 1.027]$
$m(\phi K)$	GeV/c^2	$[1.949, 1.989]$

The variables in the table 1 characterize the decay of the B_s^0. In particular $\chi^2_{2D}(B_s^0)$ and $\chi^2_{3D}(B_s^0)$ are, respectively, the χ^2 of the fit on the 2-dimensional and 3-dimensional vertex of tracks. The $L_{xy}(B_s^0)$ is the decay length of B_s^0 in the transverse plane; similar $L_{xy}(D_s)_{B_s^0}$ is the decay length of D_s but it refers to the B_s^0 decay vertex. p_T is the transverse momentum, d_0 is the impact parameter that is the minimum distance of the track from the beam line and η is the pseudo-rapidity of the track. ΔR is the angular distance defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$

Figure 2 shows the distribution of invariant mass of the B_s^0 candidates (in the $D_s\pi$ hypothesis), after the cuts listed in 1. We can observe the $B_s^0 \to D_s^- \pi^+$ peak at the B_s^0 nominal mass of 5.3663 GeV/c^2. At higher mass of the $B_s^0 \to D_s^- \pi^+$ peak we expected the combinatorial background, which is mostly composed of random charged particles, displaced from the beam-line, accidentally satisfying the

[1] We excluded the period zero because the dE/dx response is not calibrated on this period.

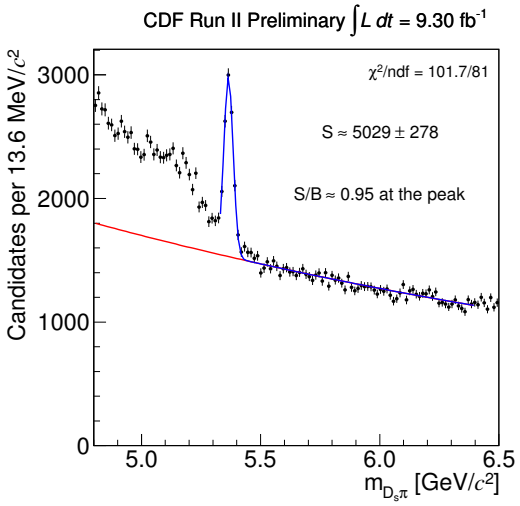

Figure 2: Invariant mass distribution of the pair $D_s^-\pi^+$ after the baseline cuts summarized in table 1.

selection requirements. In addition, in lower mass region, we expect background due to partially-reconstructed heavy-flavor decays (referred as "physics background") interpreted as mis-reconstructed b-hadron decays. We expect, also, the $B_s^0 \to D_s^{\pm}K^{\mp}$ signal shifted with respect to the $B_s^0 \to D_s^-\pi^+$peak.

To give a rough estimate of the signal yield and of the purity of the sample, a simple χ^2-binned fits was performed. A Gaussian shape was assumed to parametrize the "signal" peak, while an exponential distribution was used to parametrize the combinatorial background. The region mass of the physics background was excluded from this fit. The figure 2 shows the invariant mass distribution of the B_s^0 candidates with the fit overlaid.

We estimate a yield of 5029 ± 278 $B_s^0 \to D_s^-\pi^+$ events with a standard deviation $\sigma = (19.7 \pm 0.6)$ MeV/c^2, and a purity $S/B \approx 0.95$ at the peak. The one-dimensional binned fit is only a qualitative tool to estimate the yield and the purity of the final optimized samples in a quick way.

3 Monte Carlo

Monte Carlo samples for $B_s \to D_s X$ modes have been generated following the standard prescription of the B Monte Carlo Group that takes into account the many changes of detector and trigger configurations during the data-taking: changes of silicon coverage, of the XFT and the SVT configurations, as well as of the Two-Track Trigger trigger selection. In the Monte Carlo, the beamline, the detector and the trigger performance corresponding to the configuration of each run are simulated.

We generated single B^0, B_s^0 and Λ_b^0 meson using BGenerator: fragmentation was turned off and rapidity and $p_T(B)$ distributions were taken from an external histogram containing a smooth fit to the data published in CDF Run II measurement [1], according to the standard prescription of the B Monte Carlo Group.

The generated sample was processed with the realistic trigger and GEANT

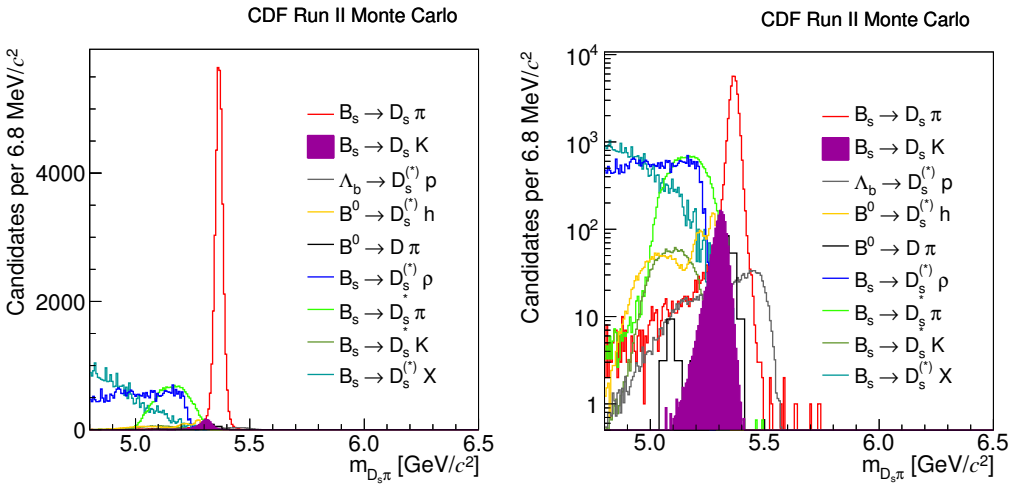

Figure 3: Invariant $D_s\pi$-mass distribution of the simulated B_s^0 Λ_b^0 B^0 and B^-. Relative fractions are weighted using measured branching ratios for the for the observed modes.

simulation, and then reconstructed using the release 6.1.4.m of ProductionExe. Radiative tails (Final State Radiation) have been included through the `PHOTOS` package embedded in the full CDF Simulation. $B - \bar{B}$ oscillations were inhibited ($\Delta m_d = \Delta m_s = 0$), and the lifetime difference in the B_s^0 system was set to zero ($\Delta\Gamma_s/\Gamma_s = 0$). Figure 3 shows the expected $D_s\pi$-mass distribution of the data sample, resulting from the Monte Carlo simulation of each mode normalized using the branching fractions derived from the current experimental knowledge [19]. As expected, in the $D_s\pi$ mass assignment, we have a high peak due to the $B_s^0 \to D_s^-\pi^+$ decay at the nominal B_s^0 mass, while the $B_s^0 \to D_s^\pm K^\mp$ peak is shifted at lower masses with an asymmetric tail, due to the wrong mass assignment. The different contributions of mis-reconstructed decays can be seen in the figure 3.

3.1 Monte Carlo validation

In order to evaluate whether the Monte Carlo simulation describes the data reliably, we performed a comparison between their kinematics distributions. To extract the signal distributions we must accurately subtract the background. However this is not possible with an high accuracy from the sample selected by the our baseline cuts (see tab. 1), since the amount of background in the sample is sizeable (see figure 2). We then decided to use a tighter selection, inspired to the work done in ref. [3]. Therefore we used different requirement reported in the table 2, and the $D_s\pi$-mass distribution obtained is reported in figure 4. The background level is now acceptable for our purpose.

3.1.1 Kinematic composition of the sample

The data sample is collected using three different trigger paths: `B_CHARM_LOWPT`, `B_CHARM`, and `B_CHARM_HIGHPT`. These trigger paths have different momentum thresholds, thus our final data sample is composed by a mixture of different kinematics.

Table 2: Cuts applied to the data sample to perform the comparison between simulated and real events.

$B_s^0 \to D_s^- h^+$	unit	selection
$\chi^2_{2D}(B_s^0)$	–	< 10
$\chi^2_{3D}(D_s)$	–	< 15
$L_{xy}(B_s^0)$	µm	> 300
$p_T(B_s^0)$	GeV/c	> 5.5
$\lvert d_0(B_s^0)\rvert$	µm	< 60
$L_{xy}(D_s)_{B_s^0}$	µm	> 0
$\Delta R(D_s, h)$	–	< 1.5
$p_T(h)$	GeV/c	> 2
$\lvert\eta\rvert$	–	< 1
$m(KK)$	GeV/c^2	$[1.013, 1.028]$
$m(\phi h)$	GeV/c^2	$[1.948, 1.988]$

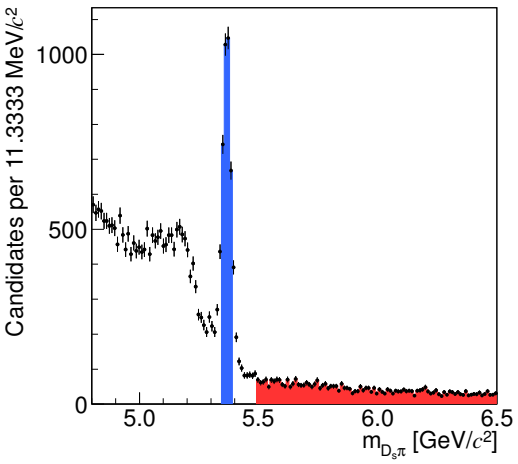

Figure 4: $D_s\pi$-mass distribution after cuts reported in table 2.

Table 3: Kinematic composition of the data sample.

sub-sample	kinematic	fraction
L	B_CHARM_LOWPT	27.2%
M	B_CHARM	46.5%
H	B_CHARM_HIGHPT	26.3%

Since our trigger system uses dynamical prescales, even the coefficient of the three kinematics are not trivial to determine. In fact we can have, for example, some events triggered by B_CHARM_HIGHPT and not by B_CHARM, although the B_CHARM kinematics includes the B_CHARM_HIGHPT. Therefore, in order perform a correct comparison between data and Monte Carlo we have to take into account this effect.

The simulation does not have access to the database trigger dynamical prescales information, therefore we have to separately generate the different trigger scenarios, and combine them with the relative fraction observed in data. To extract these fractions we divided the data sample in three independent sub-samples with well-defined kinematics. The three sub-samples were defined as follows:

Sample L where events triggered B_CHARM_LOWPT;

Sample M where events triggered B_CHARM, but did not trigger B_CHARM_LOWPT

Sample H where events triggered B_CHARM_HIGHTPT, but triggered neither B_CHARM nor B_CHARM_LOWPT.

The composition of the sample is reported in table 3. The $D_s\pi$ mass distribution is shown in the figure 5 for the three different sub-samples.

3.1.2 Primary Vertex

The primary vertex requires a special treatment. In fact, there is a difference between the calculation of the primary vertex in the data and in the simulation. In data the spatial position of the primary vertex is measured using algorithms exploiting the global event information, in which the candidate of interest is removed. In our Monte Carlo, we generate only the B_s^0 candidates so we have no access to the distributions of the large number of additional tracks in the detector coming from the $p\bar{p}$ collision. This means that in the Monte Carlo we cannot use the same procedure of data. Thus to estimate the simulated primary vertex, we get the point of the beamline at the minimum distance from the straight line which has the momentum of the B_s^0 candidate as direction and passes through its decay vertex. The figure 6 shows the comparison of the distributions of the coordinate of the primary vertex (PV_x, PV_y, PV_z); the agreement is satisfactory. Incidentally, this confirms that the offline algorithm does a good job in determining the vertex from the global event information.

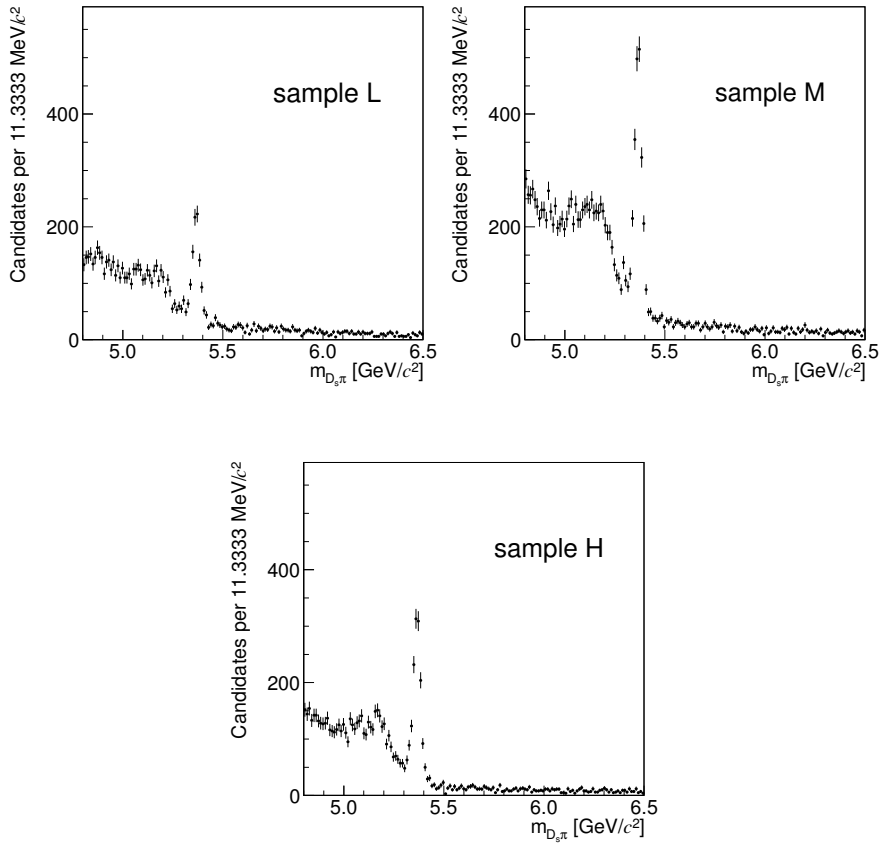

Figure 5: Invariant mass distribution of the pair $D_s\pi$ in the three samples described in the text.

3.1.3 Transverse decay-length error

We notice a discrepancy between data and Monte Carlo for the distribution of the transverse decay-length error σ_{L_T} as reported in figure 7. Since probing the sources of this discrepancy requires a large amount of work, which is clearly out of scope of this work, we decide do not use σ_{L_T} as variable in any part of the selection procedure (or a related observable such as the transverse decay-length significance L_T/σ_{L_T}).

3.1.4 $p_T(B)$ reweighting of the Monte Carlo

We also observe a very small discrepancy at lower value in the $p_T(B)$ distribution between simulated b-hadrons candidates and real data (see fig. 8). Therefore, to obtain a better match, we reweight the $p_T(B)$ spectrum of the simulation to the $p_T(B)$ spectrum observed in data.

We compare the distributions of several observables for the $B_s^0 \to D_s^-\pi^+$ mode, for which we can easily extract a sizeable and clean signal from data (see fig. 4). The comparisons between the distribution of several observables of the sideband subtracted data and the simulation are shown in the fig 9, 10 and 11, respectively

for the sub-samples L, M and H. The agreement between simulation and data is satisfactory for all the observables. We expect the same level of agreement to occur for the $B_s^0 \to D_s^{\pm} K^{\mp}$ mode, which is kinematically very similar to the $B_s^0 \to D_s^- \pi^+$.

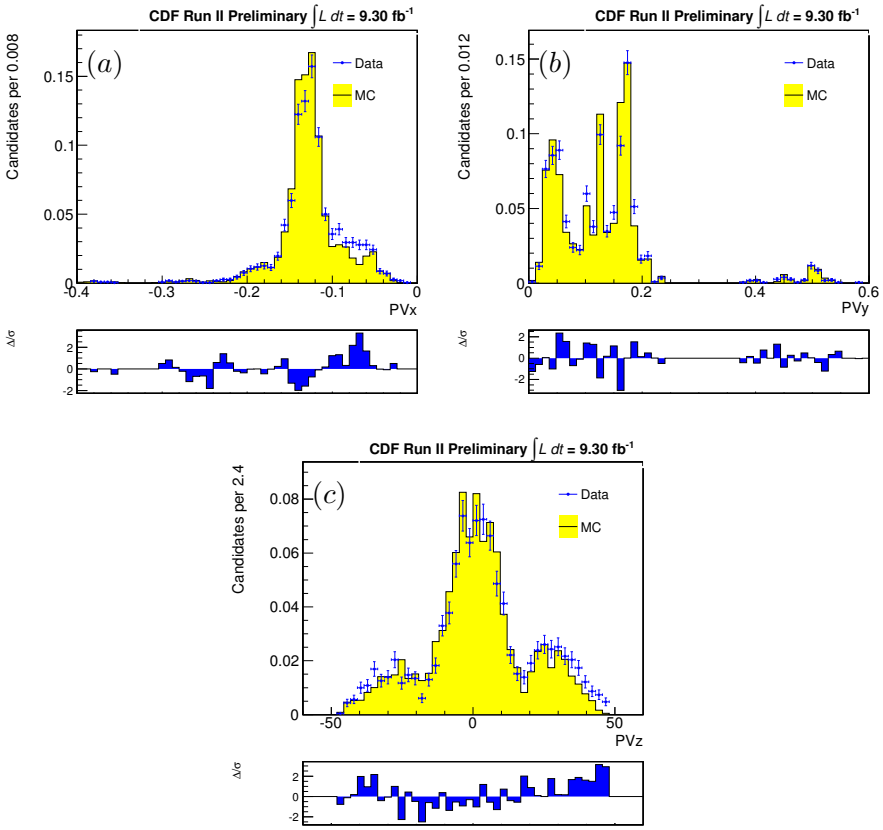

Figure 6: Distribution of the primary vertex coordinate in data (points with error bars) and in Monte Carlo (filled histogram): x-coordinate (a), y-coordinate (b), z-coordinate (c).

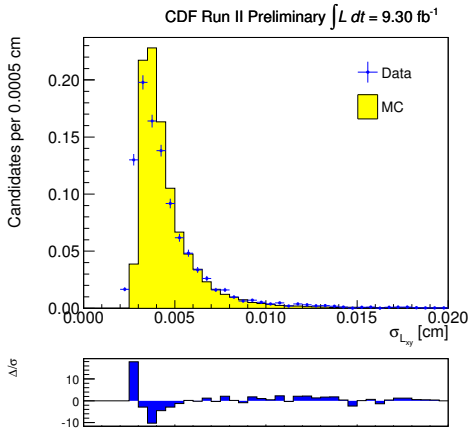

Figure 7: Distribution of the error on the transverse decay-length in data (points with error bars) and in Monte Carlo (filled histogram).

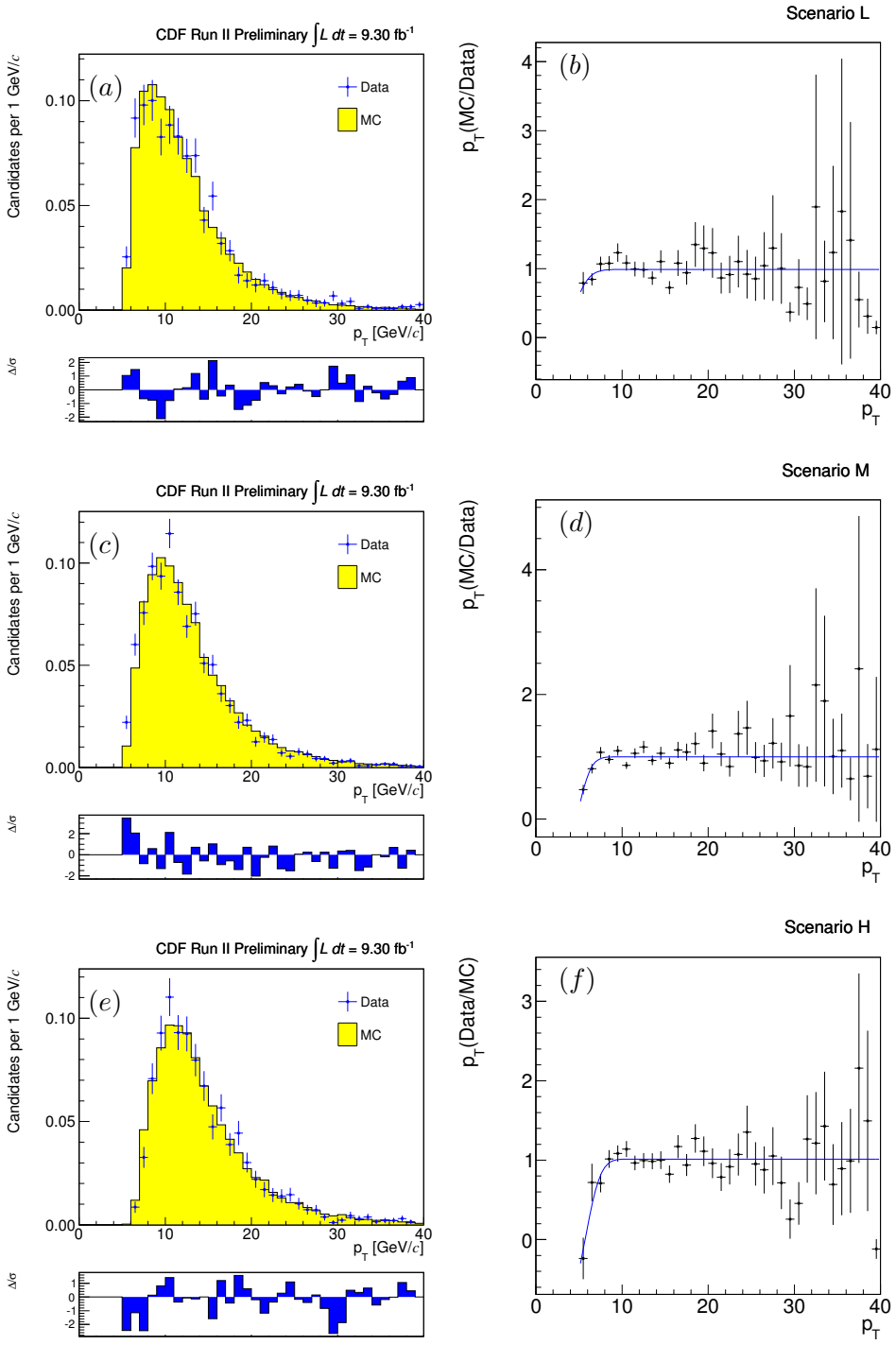

Figure 8: Data-simulation comparison of the $p_T(B_s^0)$ distribution in the sub-samples L (a), M (c) and H (e) (filled histograms are the Monte Carlo, dots with errors are data). Ratio between data histogram and Monte Carlo histogram for sub-samples L (b), M (d) and H (f).

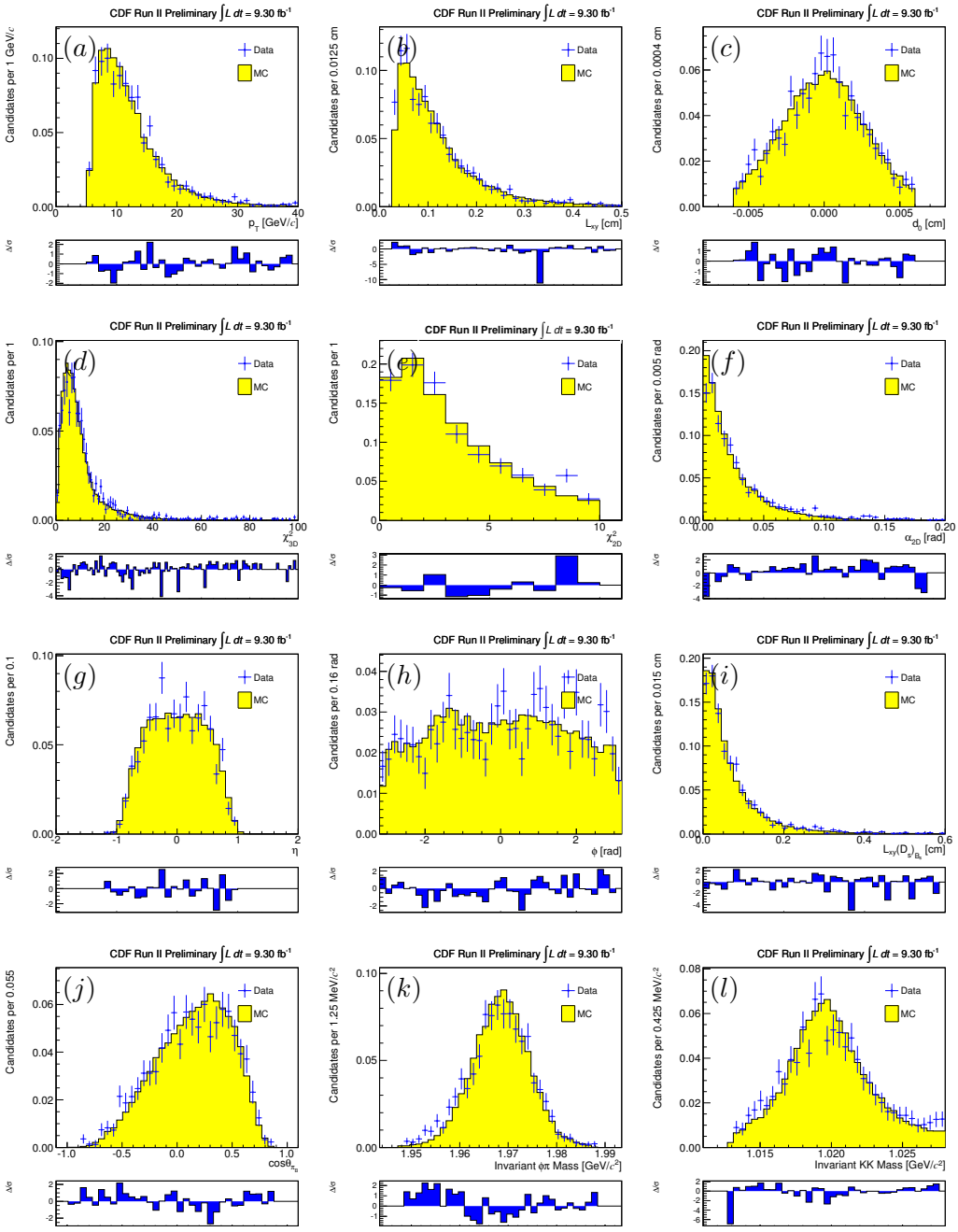

Figure 9: Comparison of background-subtracted distributions in the $B_s^0 \to D_s^- \pi^+$ decay and equivalent Monte Carlo distributions for sub-sample L: $p_T(B_s^0)$ (a), $L_T(B_s^0)$ (b), $d_0(B_s^0)$ (c), $\chi^2_{3D}(B_s^0)$ (d), $\chi^2_{2D}(B_s^0)$ (e), α_T (f), $\eta(B_s^0)$ (g), $\varphi(B_s^0)$ (h), $L_T(D_s)_{B_s^0}$ (i), $\cos\theta^*(D_s)$ (j), $m_{\phi\pi}$ (k), m_{KK} (l). Data (points with error bars) are compared with reweighted Monte Carlo simulation (filled histogram).

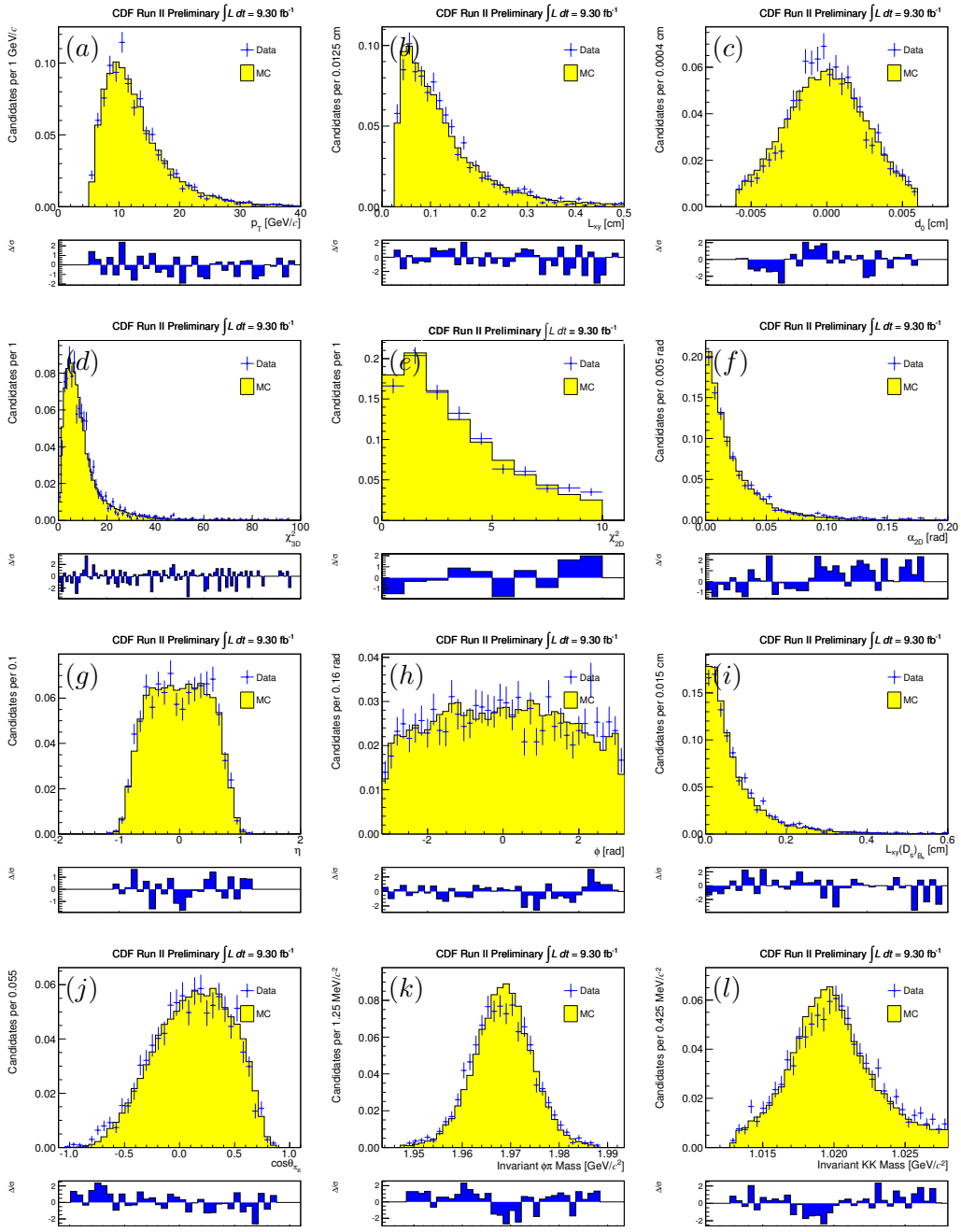

Figure 10: Comparison of background-subtracted distributions in the $B_s^0 \to D_s^- \pi^+$ decay and equivalent Monte Carlo distributions for sub-sample M: $p_T(B_s^0)$ (a), $L_T(B_s^0)$ (b), $d_0(B_s^0)$ (c), $\chi^2_{3D}(B_s^0)$ (d), $\chi^2_{2D}(B_s^0)$ (e), α_T (f), $\eta(B_s^0)$ (g), $\varphi(B_s^0)$ (h), $L_T(D_s)_{B_s^0}$ (i), $\cos\theta^*(D_s)$ (j), $m_{\phi\pi}$ (k), m_{KK} (l). Data (points with error bars) are compared with reweighted Monte Carlo simulation (filled histogram).

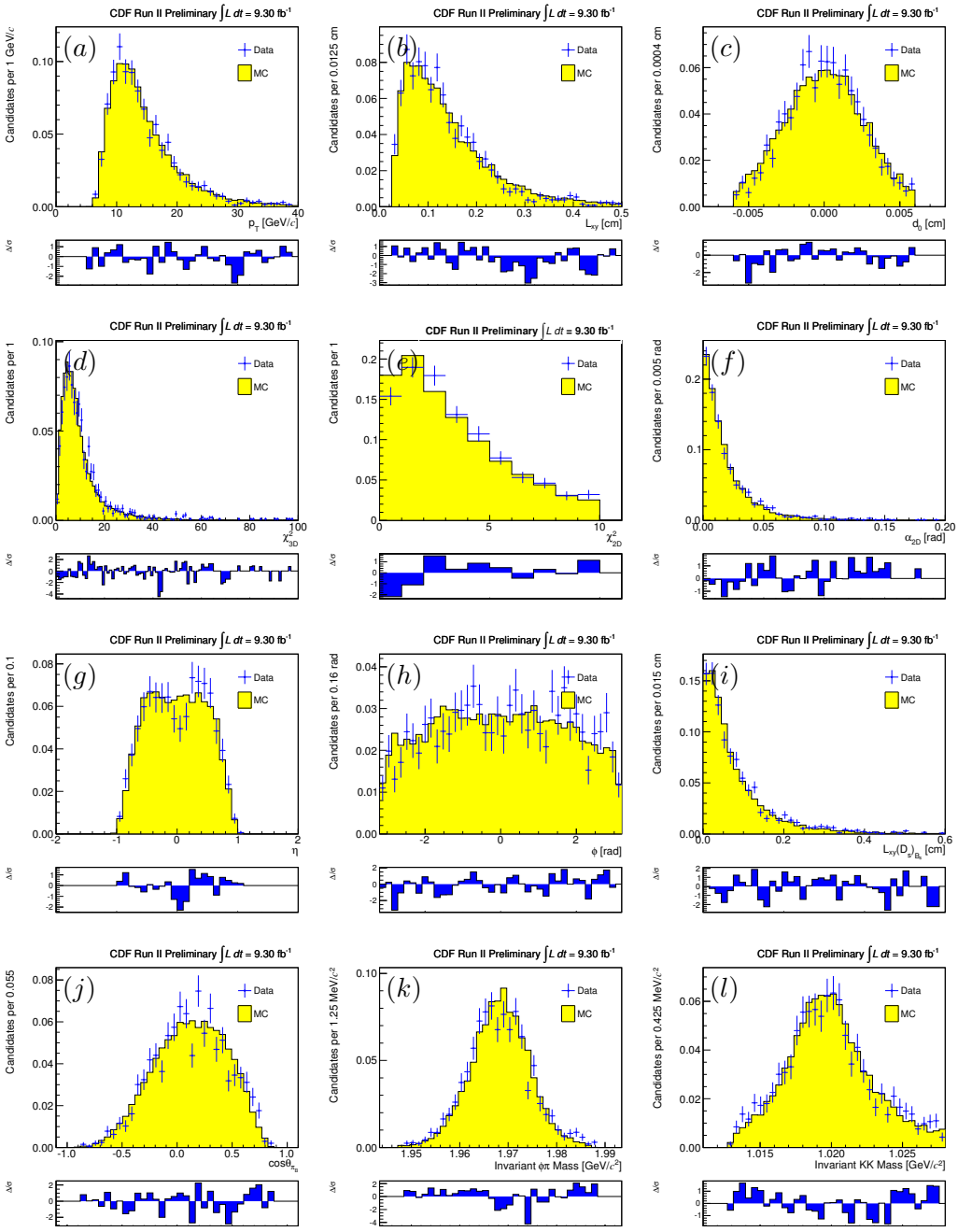

Figure 11: Comparison of background-subtracted distributions in the $B_s^0 \to D_s^- \pi^+$ decay and equivalent Monte Carlo distributions for sub-sample H: $p_T(B_s^0)$ (a), $L_T(B_s^0)$ (b), $d_0(B_s^0)$ (c), $\chi^2_{3D}(B_s^0)$ (d), $\chi^2_{2D}(B_s^0)$ (e), α_T (f), $\eta(B_s^0)$ (g), $\varphi(B_s^0)$ (h), $L_T(D_s)_{B_s^0}$ (i), $\cos\theta^*(D_s)$ (j), $m_{\phi\pi}$ (k), m_{KK} (l). Data (points with error bars) are compared with reweighted Monte Carlo simulation (filled histogram).

4 Neural Network optimization

The procedure used to optimize the selection of the data sample is based on an Artificial Neural Network (NN). The NN is implemented in the software packages Toolkit for Multivariate Analysis (TMVA) [30].

4.1 Cuts optimization

We chose as input of the Neural Network the following variables:

$\boldsymbol{p_T(B_s^0)}-$ transverse momentum of the B_s^0 candidates;

$\boldsymbol{L_T(B_s^0)}-$ transverse decay-length of the B_s^0 candidates;

$\boldsymbol{\chi^2_{3D}(B_s^0)}-$ the χ^2 of the 3-dimensional fit to the B_s^0 candidates decay vertex;

$\boldsymbol{d_0(B_s^0)}-$ impact parameter of the B_s^0 candidates;

$\boldsymbol{\alpha_T(B_s^0)}-$ pointing angle in the transverse plane of the B_s^0 candidates;

$\boldsymbol{L_T(D_s)_{B_s^0}}-$ transverse decay-length for the D_s candidates with respect to the decay vertex of the B_s^0 candidates;

$\boldsymbol{\cos(\theta^*_{D_s})}-$ angular distribution of the D_s candidates in the center of mass frame of the B_s^0.

The signal input distributions are taken from the Monte Carlo of the $B_s^0 \to D_s^{\pm}K^{\mp}$, as described in section 3, while for the background sample we chose the higher mass sideband, i.e. events with masses in the range $[5.5, 6.5]\,\mathrm{GeV}/c^2$. The distribution of the training variables, for signal and background, are shown in figure 12. These variables have been chosen because of their discriminant power which is satisfactory for the scope of this work. Adding other variables would not improve greatly the separation power, at the price of an they will increasing of the level of complexity of the Neural Network. In fact using too many variables may be very powerful but also very dangerous. The supervised learning approach is based on the assumption that the simulation perfectly reproduces real data. Since this is clearly wrong, a NN which uses a very large numbers of input variables may use "small" discrepancies between data and Monte Carlo to reject signal events, leading to a non optimal (maybe biased) selection. Fortunately in our data sample we can a posteriori verifies that NN works fine looking at the reference $B_s^0 \to D_s^-\pi^+$ peak. Anyway, the variables of our choice are both expected, and experimentally verified, to be well-reproduced by our simulations with all their correlations, being of essentially kinematical nature. A separate discussion is worthwhile for the specific case of the pointing angle ($\alpha_T(B_s^0)$).

4.1.1 Pointing angle

As explained in the section 3.1.2, the position of the primary vertex (PV) is estimated in the Monte Carlo using a different algorithm than data. We used the information from candidates and the beamline to extract the primary vertex coordinates. Although the agreement between data and Monte Carlo is satisfactory

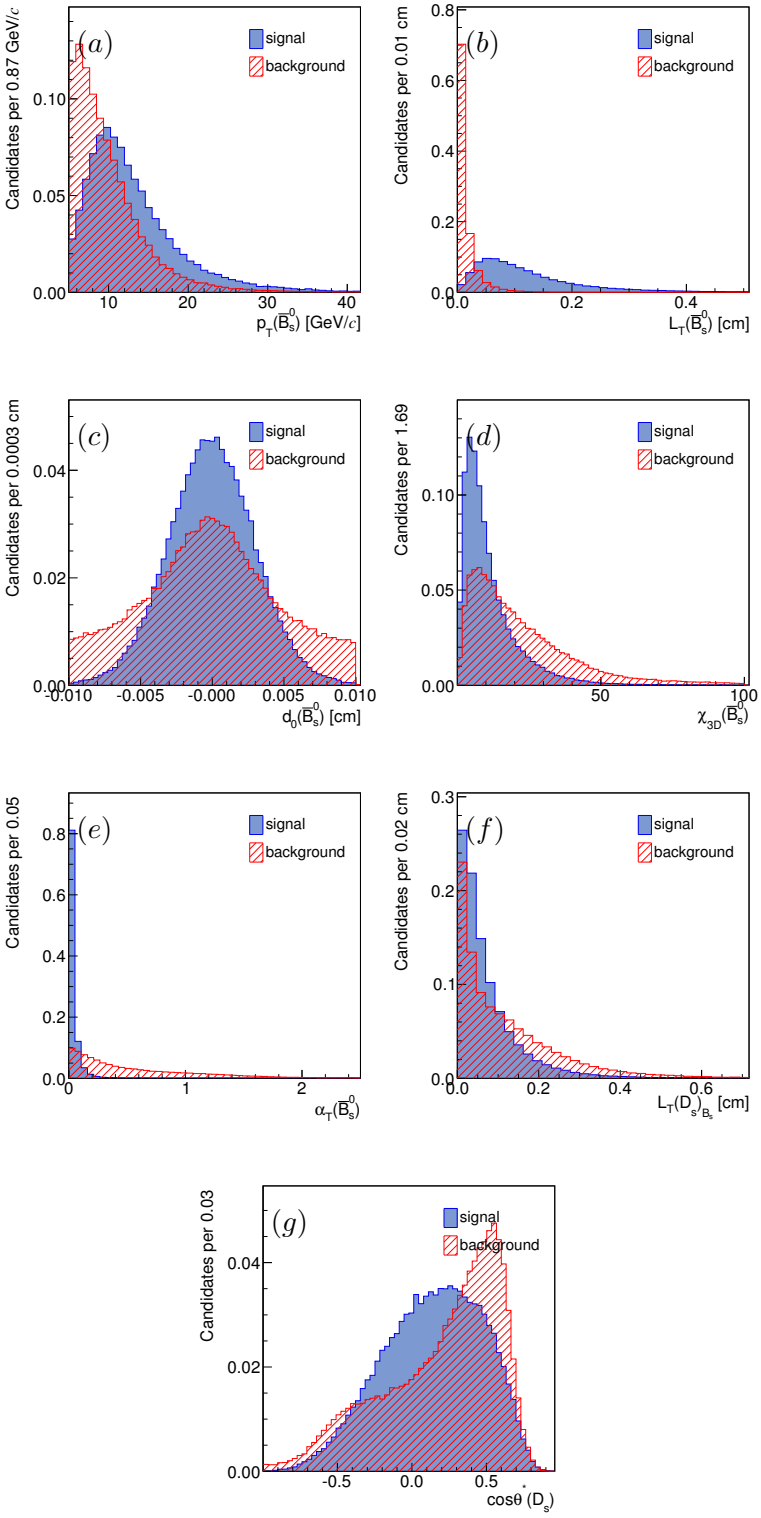

Figure 12: Input variables to the Neural Network: $p_T(B_s^0)$ (a), $L_T(B_s^0)$ (b), $d_0(B_s^0)$ (c), $\chi^2_{3D}(B_s^0)$ (d), α_T (e), $L_T(D_s)_{B_s^0}$ (f), $\cos\theta^*(D_s)$ (g). The signal is the filled histogram (in blue), while the background is the hatched one (in red). The histograms are normalized to one.

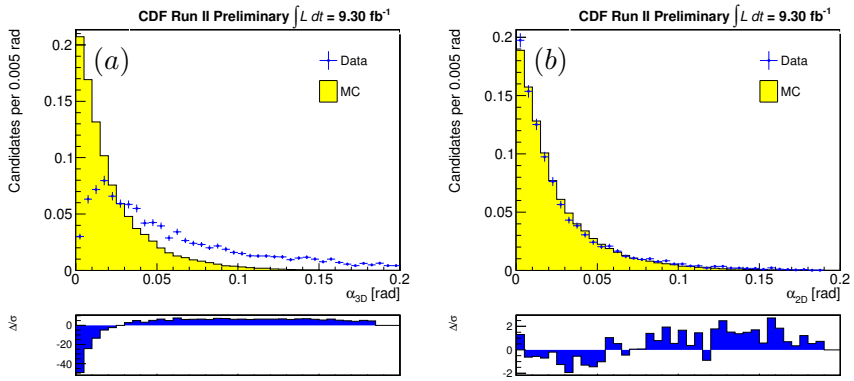

Figure 13: Comparison of the pointing angle distribution in data and in the simulation (where the coordinates of the primary vertex are estimated with the algorithm explained in the text). (*a*) The 3-dimensional pointing angle, (*b*) the transverse pointing angle.

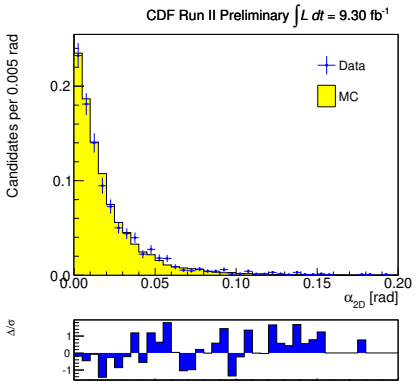

Figure 14: Comparison of the transverse pointing angle distribution in data and in simulation. The pointing angle is calculated for both data and simulation with the primary vertex estimate with the algorithm used in the simulation.

(see fig. 6) the strategy used introduces an issue in the estimate of the pointing angle (see fig. 13a). In fact the PV position extracted in such a way, is highly correlated to the direction of the momentum of the B_s^0 candidates, above all in the z-coordinate, where we have a large uncertainty. Unfortunately the $p_T(B_s^0)$ and the position of the PV directly enter in the definition of the three-dimensional pointing angle α_{3D}.

Since the problem affects especially the z-direction (see fig. 13b), we decide to use the projection of the pointing angle onto the transverse plane. To avoid introducing a bias in the Neural Network, we decided to use own algorithm for determination the primary vertex both in the simulation and real data, for both the signal and the background sample, and the agreement is satisfactory (see fig. 14 and fig. 15).

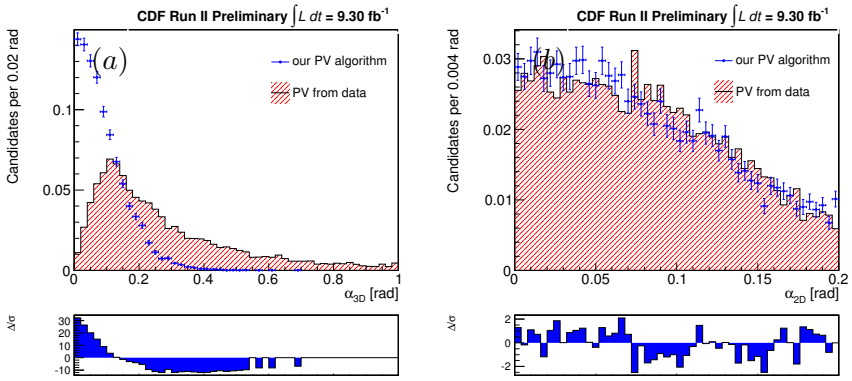

Figure 15: Comparison of pointing angle distribution, in the combinatorial background mass region $[5.5, 6.5]\,\mathrm{GeV}/c^2$, calculate with the algorithm introduced in Monte Carlo simulation for calculation of the primary vertex (points) and with the standard algorithm (hatched histogram). (a) 3-dimensional pointing angle, (b) transverse pointing angle.

4.2 Final selection

Finally we trained the Neural Network and we obtain as output the distributions shown in figure 16. Figure 16a shows the distribution of the output variable of the NN for the signal (in blue), peaked at 1, and for the background (in red). The separation between the two distributions is the equivalent of 3.1 standard deviations between Gaussians. Figure 16 reports the correlation matrix of the variables used in the NN training, for the signal sample (see fig. 16c) and for the background sample (fig. 16d).

The cut on the NN output response is chosen by maximizing the score function

$$\text{score function} = \frac{S}{\sqrt{S+B}}, \tag{1}$$

where S is the number of $B_s^0 \to D_s^{\pm} K^{\mp}$ events estimated by the simulation, while B is the number of background events taken by fitting, with an exponential function, the high mass sideband in the data. The score function in (1) is a good choice for a typical "counting experiment" being inversely proportional to the statistical uncertainty of the measurement of a signal yield. The figure 17 shows the score function for several cuts on the NN variable. We choose as our final selection the cut $NN > 0.9$.

The $D_s\pi$-mass distribution is shown in the figure 18b. For comparison in the figure 18a is also reported the mass distribution with the baseline selection of table 1. A strong reduction of background is apparent, with little reduction of the $B_s^0 \to D_s^- \pi^+$ peak height, as expected from the good separation shown in fig. 16. Background reduction is particularly strong in the region above the $B_s^0 \to D_s^- \pi^+$ peak, which is essentially pure combinatorial.

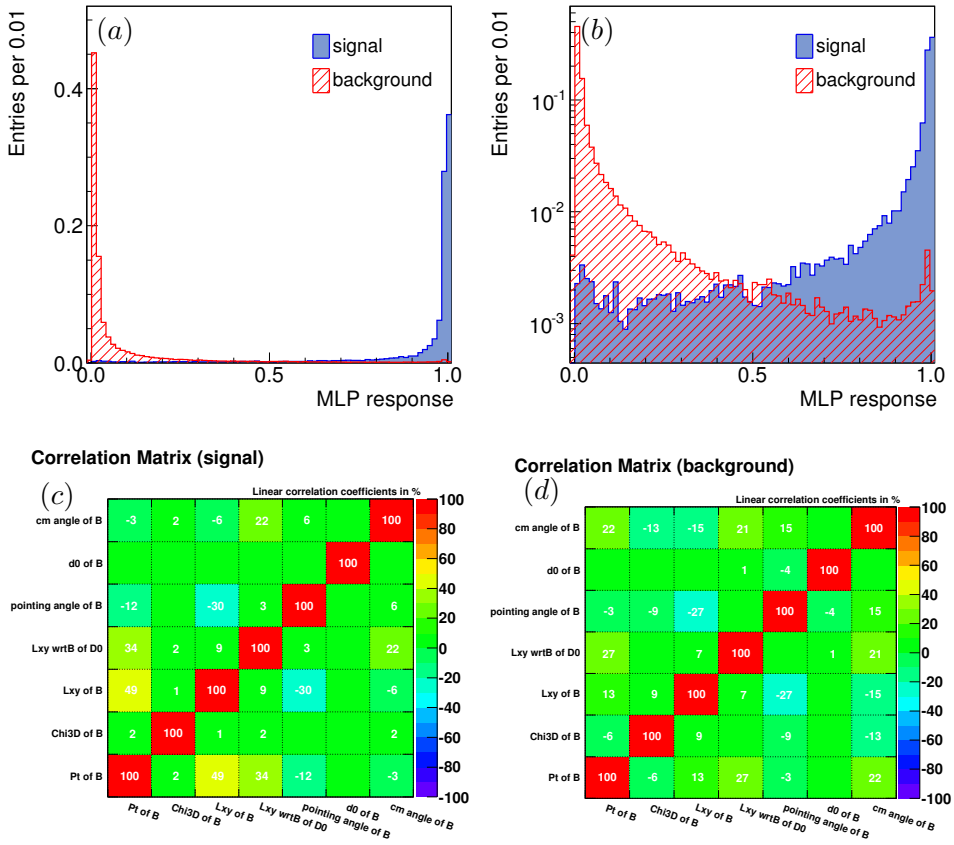

Figure 16: Output response of the Neural Network (a); output response of the Neural Network in logarithmic scale (b). The histograms are normalized to one. Also it is reported the linear correlation matrix for the signal (c) and for the background (d)

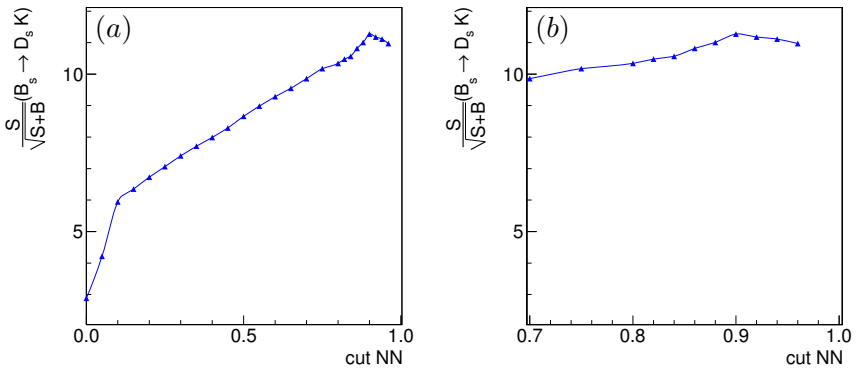

Figure 17: Score function as function of the Neural Network response (a), and its zoom (b).

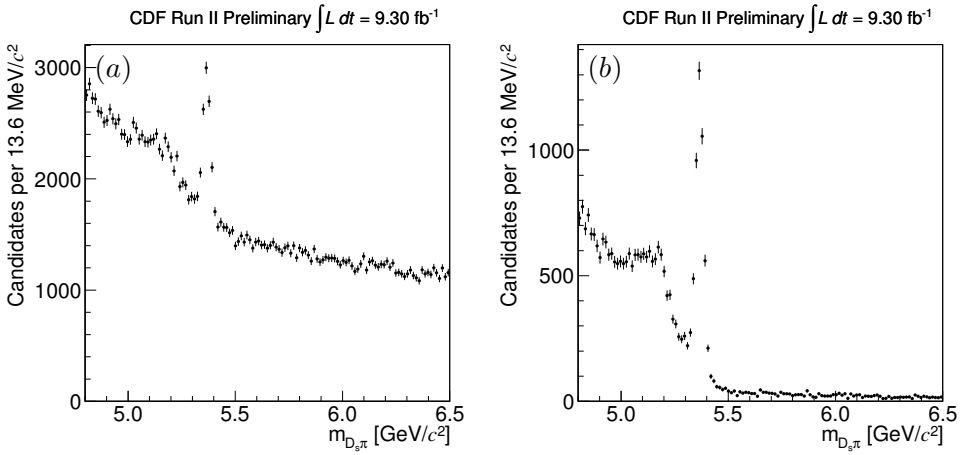

Figure 18: Invariant mass distribution of the $D_s\pi$ pair after the baseline selection (a), and after the final selection cut (b).

4.3 Procedure validation

To validate the optimization procedure we check that the simulation and real data have the same response for different NN requirements, using the reference $B_s^0 \to D_s^-\pi^+$ peak. Table 4 reports the efficiency for different NN requirements, as resulting from simulation and real data. The $B_s^0 \to D_s^-\pi^+$ yield is roughly estimated with the same strategy described in section 2, through a χ^2-binned fit.

Although the efficiency in data is systematically higher than of that observed in the simulation, the agreement is satisfactory, confirming that the NN gives a similar response if applied to data or simulation. The observed systematic effect may be due to the trivial technique used to extract the $B_s^0 \to D_s^-\pi^+$ number of events.

Table 4: Comparison between the Monte Carlo efficiency and the efficiency of the $B_s^0 \to D_s^-\pi^+$ decay mode.

NN cut	$\epsilon_{\rm MC}$	$\epsilon_{\rm Data}$
0.9	0.796	0.861 ± 0.078
0.8	0.856	0.905 ± 0.082
0.7	0.892	0.929 ± 0.085
0.6	0.914	0.953 ± 0.087
0.5	0.934	0.972 ± 0.089
0.4	0.948	0.992 ± 0.092
0.3	0.961	0.993 ± 0.093

5 Particle identification

Individual hadron identification is difficult with the CDF II detector, since the detector was designed for high-p_T physics measurements. The TOF is the

only detector entirely devoted to do this function, but its performance is marginal for particles of interest for this analysis, having transverse momenta greater than $2\,\mathrm{GeV}/c$. For charged particles with $p_T \gtrsim 2\,\mathrm{GeV}/c$, a reasonably effective separation can be obtained from the rate of energy loss through ionization (dE/dx) in the gas that fills the active volume of the drift chamber. The dE/dx calibration is based on a large samples of $D^0 \to K^-\pi^+$ and $\Lambda \to p\pi^-$ decays taken with the displaced track trigger. In this work we use the official CDF dE/dx universal curves and the official templates for the different mass hypothesis [33].

5.1 dE/dx residual

The dE/dx residual (in m_A mass hypothesis) of a charged particle, with momentum p and observed specific energy-loss $dE/dx|_{\mathrm{obs}}$, is defined as

$$\mathrm{res}_A = \left.\frac{dE}{dx}\right|_{\mathrm{obs}} - \left.\frac{dE}{dx}\right|_A , \tag{2}$$

where the $dE/dx|_A$ is the expected dE/dx,evaluated at $\beta\gamma = p/m_A$. The official CDF II parameterization of dE/dx residual distributions with the correct mass hypothesis (see ref. [33]) is made with analytical functions (convolution of several Gaussian terms), currently available in a stand-alone `C` code. Figure 19 shows the dE/dx residual distribution observed for pions, kaons and protons, in the m_π, m_K and m_p mass hypothesis, respectively, and their official parameterizations. We will indicate with $\mathcal{P}_{\pi^+}(\mathrm{res}_{\pi^+})$ the parameterization of the positive pion dE/dx residual, and with $\mathcal{P}_{K^+}(\mathrm{res}_{K^+})$, $\mathcal{P}_p(\mathrm{res}_p)$ the parameterization of the positive kaon and proton residual, respectively. The anti-particles have similar notation.

In order to have just one single dE/dx observable, to be used in the fit of composition (see section 6), we modified these templates to account for the momentum dependence, as explained in the next subsection.

5.2 PID observable

In this analysis we follow a statistical approach that combines information from PID and kinematics into a fit of composition as we will describe in section 6. We use the PID information on the daughter track of the B_s^0 candidates, and the information is summarized in a single observable, the dE/dx residual in the pion hypothesis:

$$\mathrm{res}_\pi = \left.\frac{dE}{dx}\right|_{\mathrm{obs}} - \left.\frac{dE}{dx}\right|_\pi . \tag{3}$$

The res_π is a momentum-dependent observable, as it can be seen from figure 21, if the daughter particle of the B_s^0 candidates is not a pion. We then have to account for this momentum dependence in the Likelihood terms [34]. Therefore we decide to integrate over the momentum dependence to avoid a complex momentum-dependent parametrization of the mis-reconstructed decay modes. If the daughter particle of the B_s^0 candidates is a pion, the template is exactly the $\mathcal{P}_\pi(\mathrm{res}_\pi)$ function, from ref. [33], as reported in figures 19a and 19b. Instead if it is a kaon, a proton or an electron, some adjustments are necessary. For instance the probability density function (p.d.f.) of the dE/dx residual in the positive pion mass hypothesis of a

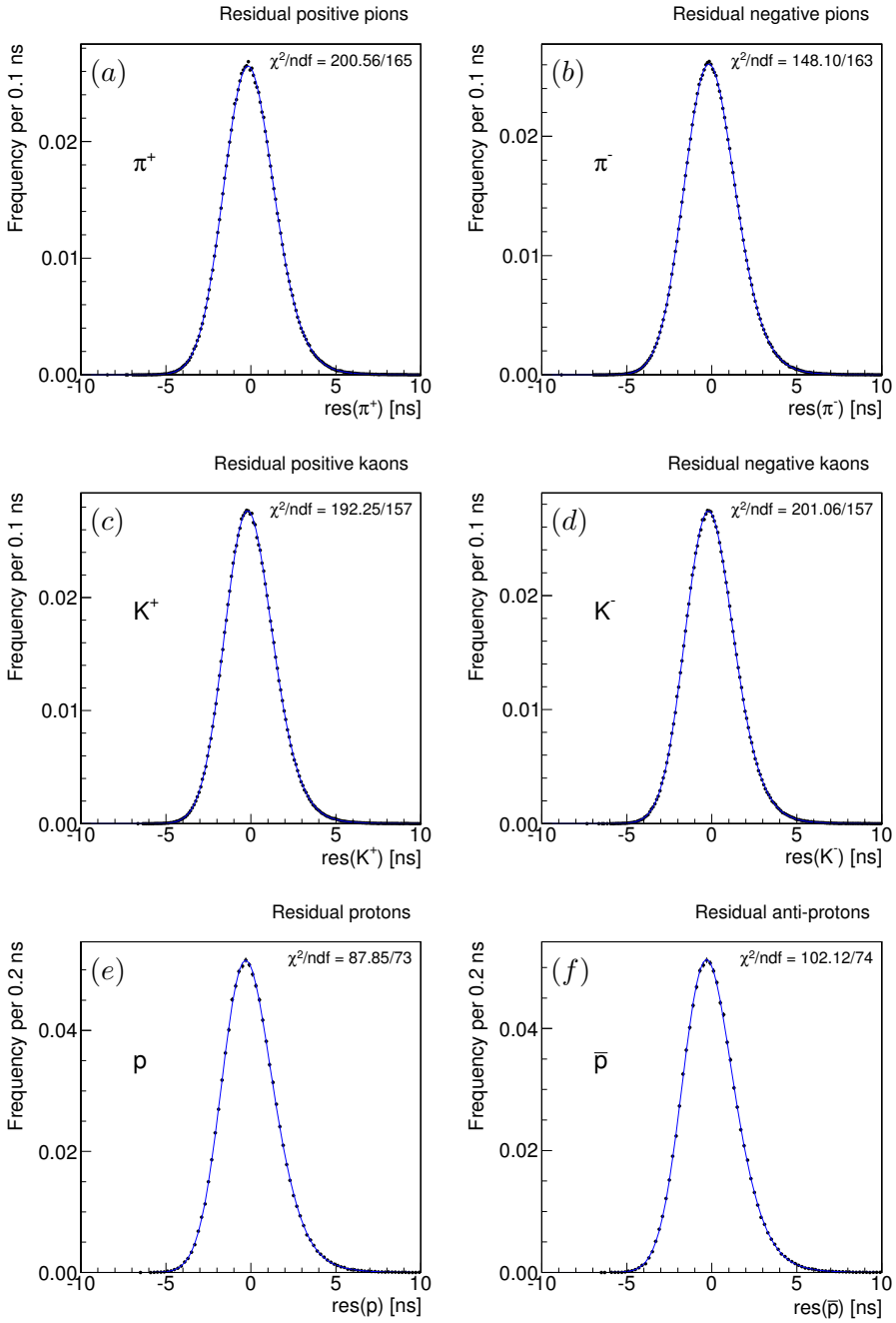

Figure 19: Distribution of observed dE/dx residual, for pions (with pion mass hypothesis) (a, b), for kaons (with kaon mass hypothesis) (c, d) and for protons (with proton mass hypothesis) (e, f).

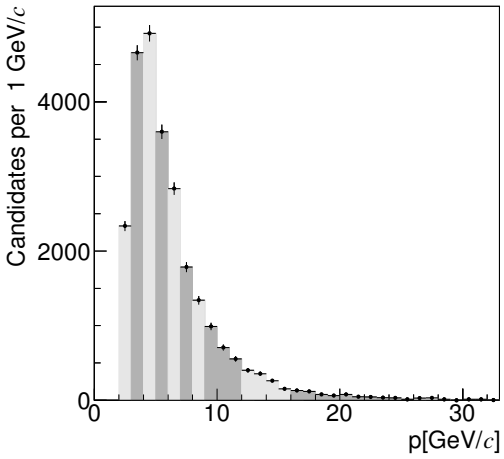

Figure 20: Monte Carlo momentum distribution for K from B_s^0 candidate in the decay $B_s^0 \to D_s^{\pm} K^{\mp}$. The fill region show the section used to make the template of the PID observable.

generic particle A^+ ($A^+ = K^+, p, e^+$), that we can indicate with $\wp_A(\mathrm{res}_{\pi^+})$, can be extracted by marginalizing the momentum dependence:

$$\begin{aligned}
\wp_{\pi^+}(\mathrm{res}_{\pi^+}) &= \mathcal{P}_{\pi^+}(\mathrm{res}_{\pi^+}), \\
\wp_{A^+}(\mathrm{res}_{\pi^+}) &= \int \mathcal{P}_{A^+}(\mathrm{res}_{\pi^+} + \delta^{\pi^+ A^+}(p))\ f_{A^+}(p)\, dp \\
&\simeq \sum_i \mathcal{P}_{A^+}(\mathrm{res}_{\pi^+} + \delta^{\pi^+ A^+}(p_i))\ f_{A^+}(p_i)\, \Delta p_i,
\end{aligned}$$

where $f_{A^+}(p)$ is the particle momentum distribution and $\delta^{\pi^+ A^+}(p)$ is the difference between expected dE/dx in π^+ and A^+ mass hypothesis, according to the universal curves of fig. 21:

$$\delta^{\pi^+ A^+}(p) = \left.\frac{dE}{dx}\right|_{\pi^+} - \left.\frac{dE}{dx}\right|_{A^+}.$$

To simplify calculations, we performed a numerical integration by dividing the momentum distribution in ten bins, where Δp_i is the width of each bin, as shown in fig. 20. We will assess a systematic uncertainty due to the binning of the momentum in the section 7. The p.d.f.s of negatively-charged particles are extracted in the same way.

The result of the procedure is reported in figure 22a, where we compare dE/dx response for positively-charged pions, kaons (from $B_s^0 \to D_s^- K^+$ decay), protons (from $\Lambda_b^0 \to D_s^- p$ decay) and electrons (from $B_s^0 \to D_s^- e^+ \nu_e$ decay). The separation power between pions and kaons is $\approx 1.4\sigma$, between pions and protons is $\approx 1.6\sigma$, and between pions and electrons is $\approx 0.9\sigma$. The PID distributions for negatively-charged particles have similar separation power, and are shown in figure 22b. Distributions of daughter particle of the other decays are similar.

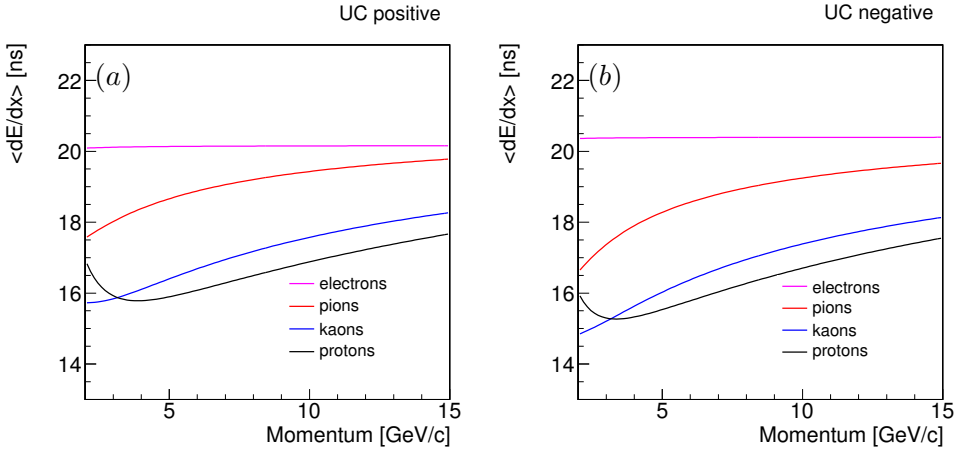

Figure 21: Universal curves as a function of particle momentum for positive (a) and negative (b) particle.

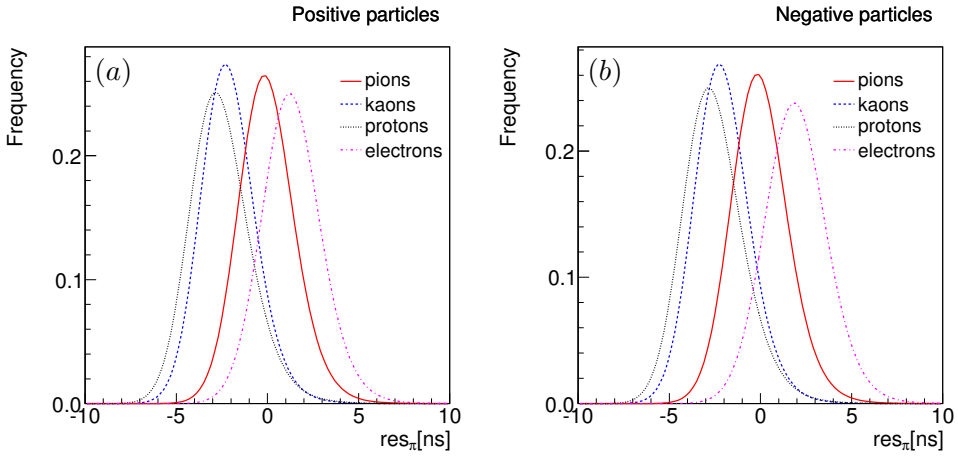

Figure 22: Comparison of the PID distribution for pions, kaons, protons and electrons. In (a) positive particles PID-distributions, (b) negative particles.

6 Fit of composition

To disentangle all the components of the data sample we decide to perform an extended unbinned Maximum Likelihood fit. A good choice of the discriminating observables is crucial to fully exploit the available information. The goal is to obtain most of the available information using the minimum number of observables. In addition, the independence of variables simplifies the modeling of the probability density, since it factorizes the joint probability density.

We represent the kinematic and PID information using two discriminating observables:

$\boldsymbol{m_{D_s\pi}}-$ invariant mass of the final state particles with the $D_s\pi$ mass assignments;

$\mathbf{res_\pi}-$ dE/dx residual in the pion mass hypothesis.

Particle identification information is summarized with one observable, the residual res_π, defined as

$$\mathrm{res}_\pi = \left.\frac{dE}{dx}\right|_{\mathrm{obs}} - \left.\frac{dE}{dx}\right|_{\pi}, \tag{4}$$

where $dE/dx|_{\mathrm{obs}}$ indicates the observed energy-loss, while $dE/dx|_\pi$ indicates the expected energy-loss in pion mass hypothesis.

6.1 Likelihood function

The Likelihood function $\mathscr{L}$ is the product of the Likelihoods $\mathscr{L}_i$ of all events:

$$\mathscr{L}(\nu, \boldsymbol{\theta}) = \frac{\nu^N}{N!} e^{-\nu} \cdot \prod_{i=1}^{N} \mathscr{L}_i(\boldsymbol{\theta}|\boldsymbol{x}_i) \tag{5}$$

where the index i runs over the events. N is the total number of events passing the final selection, $\boldsymbol{\theta}$ is the vector of parameters that we want to estimate, $\boldsymbol{x}$ is the vector of the discriminating observables $\boldsymbol{x}_i = \{m_{D_s\pi}, \mathrm{res}_\pi\}_i$. The Poisson term in eq. (5) takes into account the uncertainty due to the finite size of the total sample, where ν is the mean number of events.

The Likelihood of each event is written as the sum of a "decays" term and a combinatorial background term:

$$\mathscr{L}_i = f_{\mathrm{bkg}} \cdot \mathscr{L}_i^{\mathrm{bkg}} + (1 - f_{\mathrm{bkg}}) \cdot \mathscr{L}_i^{\mathrm{dec}}. \tag{6}$$

The index bkg (dec) labels the part of the function that describes the combinatorial background (all the decays) term; f_{bkg} is the fraction of combinatorial background events and $1 - f_{\mathrm{bkg}}$ is the fraction of all the decays in the data sample. We conventionally label as $\wp^m$ the term that describes the invariant-mass distributions ("mass term"), and $\wp^{\mathrm{PID}}$ the term that models the dE/dx density ("PID term"). The Likelihood of each individual decay mode is factorized as a product of two p.d.f.s:

$$\mathscr{L}^{\mathrm{dec}} = \sum_{j=1}^{s} f_j \cdot \wp_j^m(m_{D_s\pi}) \cdot \wp_j^{\mathrm{PID}}(\mathrm{res}_\pi), \tag{7}$$

in which the index j runs over the thirteen expected components: $B_s^0 \to D_s^- \pi^+$, $B_s^0 \to D_s^\pm K^\mp$, $B_s^0 \to D_s^{\star -} \pi^+$, $B_s^0 \to D_s^{\star \pm} K^\mp$, $B^0 \to D_s^{(\star)+} \pi^-$, $B^0 \to D_s^{(\star)-} K^+$, $B_s^0 \to D_s^- \rho^+$, $B_s^0 \to D_s^- \pi^+ \pi^0$, $B_s^0 \to D_s^- e^+ \nu_e$, $B_s^0 \to D_s^- \mu^+ \nu_\mu$, $\Lambda_b^0 \to D_s^- p$, $\Lambda_b^0 \to D_s^{\star -} p$ and $B^0 \to D^- \pi^+$. The parameters f_j are their fractions and are determined by the fit. From the $(s-1)$ independent fractions resulting by the normalization condition,

$$f_s = \sum_{j=1}^{s-1} f_j, \tag{8}$$

we determine the yield of each mode. The fractions f_j are the same for a decay mode and its CP conjugate, therefore the parameter we measure in our fit is the CP-averaged branching ratio $\mathcal{B}(B_s^0 \to D_s^\pm K^\mp) \equiv [\mathcal{B}(B_s^0 \to D_s^+ K^-) + \mathcal{B}(B_s^0 \to D_s^- K^+) + \mathcal{B}(\bar{B}_s^0 \to D_s^+ K^-) + \mathcal{B}(\bar{B}_s^0 \to D_s^- K^+)]/2$.
The Likelihood of the background term factorizes as the decays term:

$$\mathscr{L}^{\text{bkg}} = \wp^m(m_{D_s\pi}) \cdot \wp^{\text{PID}}(\text{res}_\pi). \tag{9}$$

In equations (6)-(9) the functional dependence on the vector $\boldsymbol{\theta}$ was omitted, since in the equations we explicitly wrote some terms of this vector as f_{bkg} and f_j.

6.2 Mass probability density function

Signal and full reconstructed decays

We extract the p.d.f. mass templates of various decays mode from Monte Carlo sample of $B \to DX$ described in the section 3. The mass line shape of the full reconstructed decay modes, in particular the signal modes $B_s^0 \to D_s^\pm h^\mp$ (see fig. 23), is parameterized using the following p.d.f.:

$$\begin{aligned} \wp^m(m;\boldsymbol{\alpha}) = & f_{\text{bulk}} \Big[f_1 \mathscr{G}(m;\mu_1,\sigma_1) + (1-f_1)\mathscr{G}(m;\mu_2,\sigma_2) \Big] \\ & + (1-f_{\text{bulk}})\mathscr{T}(m;b,c,\mu_1), \end{aligned} \tag{10}$$

where

$$\begin{aligned} \mathscr{G}(m;\mu,\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{1}{2}\Big(\frac{m-\mu}{\sigma}\Big)^2 \Big), \\ \mathscr{T}(m;b,c,\mu) &= \frac{1}{K} \exp\big(b(m-\mu)\big) \cdot \text{erfc}\big(c(m-\mu)\big), \\ K &= \int_{m_1}^{m_2} \exp\big(b(m-\mu)\big) \cdot \text{erfc}\big(c(m-\mu)\big)\, dm, \\ \text{erfc}(x) &= 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-t^2}\, dt. \end{aligned}$$

We use a sum of two Gaussians to parameterize the bulk of the distribution, while the long lower-mass tail is parameterized with the function $\mathscr{T}(m;b,c,\mu)$. f_{bulk} is the relative fraction of the double Gaussian bulk with respect to the total (bulk plus tail), while $1 - f_{\text{bulk}}$ is the fraction of the tail term.

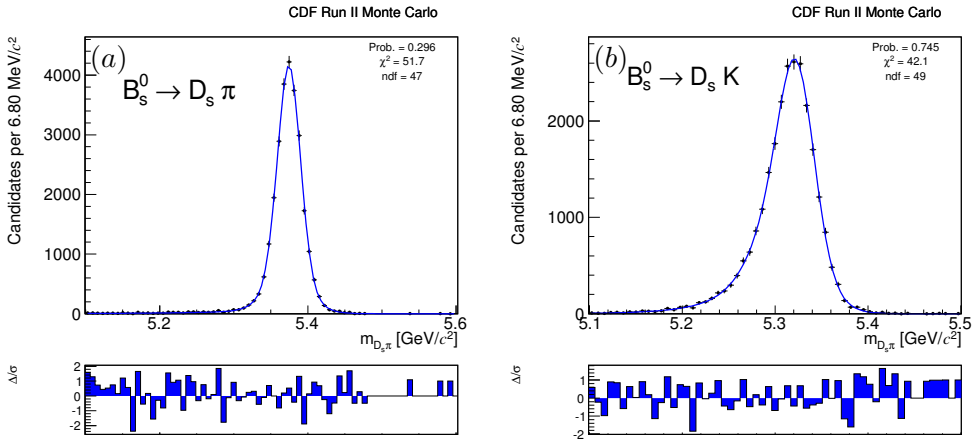

Figure 23: Mass p.d.f. for the $B_s^0 \to D_s^{\pm} h^{\mp}$ decays. (*a*) $B_s^0 \to D_s^- \pi^+$, (*b*) $B_s^0 \to D_s^{\pm} K^{\mp}$.

Mis-reconstructed

The $m_{D_s\pi}$ distribution for the mis-reconstructed decay modes is modeled with the convolution of a resolution function, a Gaussian, and a so-called "Argus function" [14]:

$$\wp^m(m;\boldsymbol{\beta}) = \mathscr{G}(m;0,\sigma) * \mathscr{A}(m;m_A,c_A),$$

$$\mathscr{A}(m;m_A,c_A) = \begin{cases} \dfrac{1}{K_A}\left[m \cdot \sqrt{1-\left(\dfrac{m}{m_A}\right)^2} \cdot \exp\left(-c_A\left(\dfrac{m}{m_A}\right)^2\right)\right] & \text{if } m < m_A, \\ 0 & \text{if } m \geq m_A, \end{cases}$$

where the normalization K_A is:

$$K_A = \int_{m_1}^{m_A} m \cdot \sqrt{1-\left(\frac{m}{m_A}\right)^2} \cdot \exp\left(-c_A\left(\frac{m}{m_A}\right)^2\right) dm \qquad (m_1 < m_A),$$

where m_A and c_A are the Argus function parameters, while σ is the resolution parameter. Some mass templates are reported in figure 25.

The semi-leptonic decays enter in the category of mis-reconstructed decays, since the neutrino is not detected. The p.d.f mass templates of those decay modes (see figures 25f (g)) are parameterized with an exponential function:

$$\begin{aligned} \mathscr{E}(m;a) &= \frac{1}{K}\exp(-am), \\ K &= \int_{m_1}^{m_2} \exp(-am)\, dm. \end{aligned} \qquad (11)$$

Background mass term

The mass shape of the combinatorial background is extracted from real data, using the higher mass side-band. The mass region $[5.6, 6.5]\,\mathrm{GeV}/c^2$, excluded from

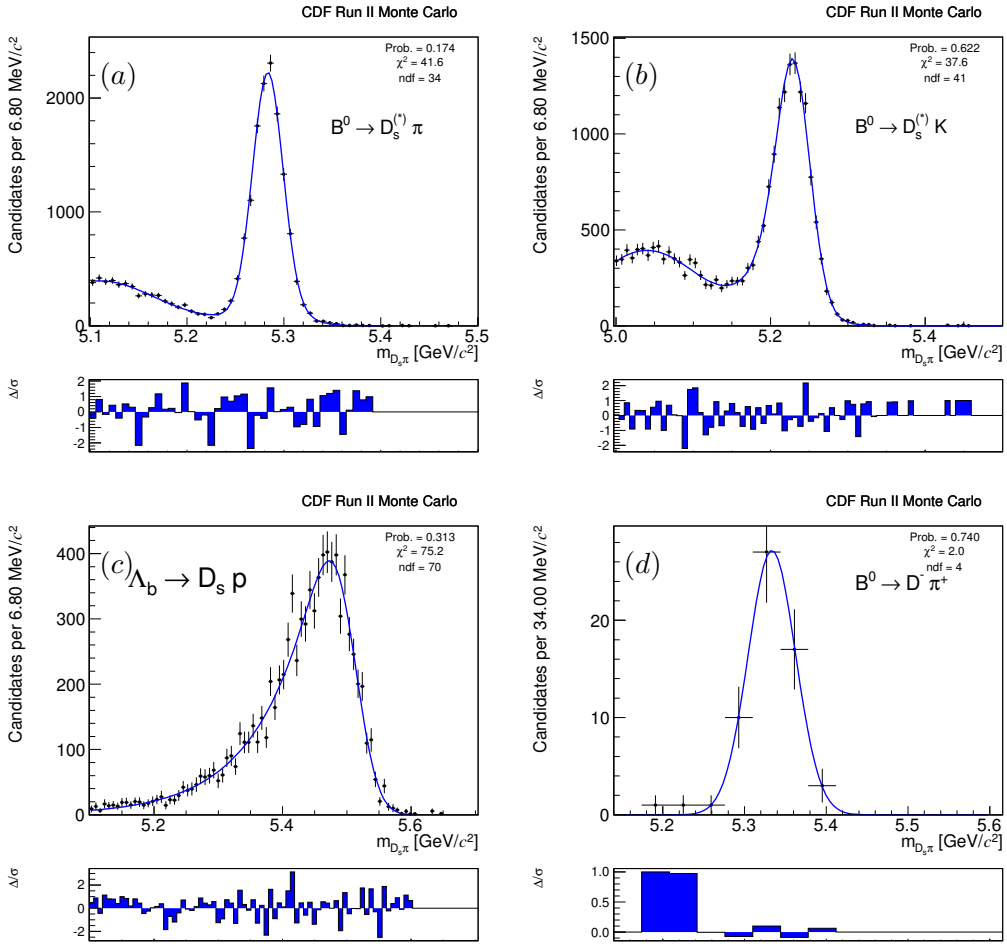

Figure 24: Mass template of various fully-reconstructed decay modes: $B^0 \to D_s^{(\star)+}\pi^-$ (a), $B^0 \to D_s^{(\star)-}K^+$ (b), $\Lambda_b^0 \to D_s^- p$ (c), and $B^0 \to D^-\pi^+$ (d).

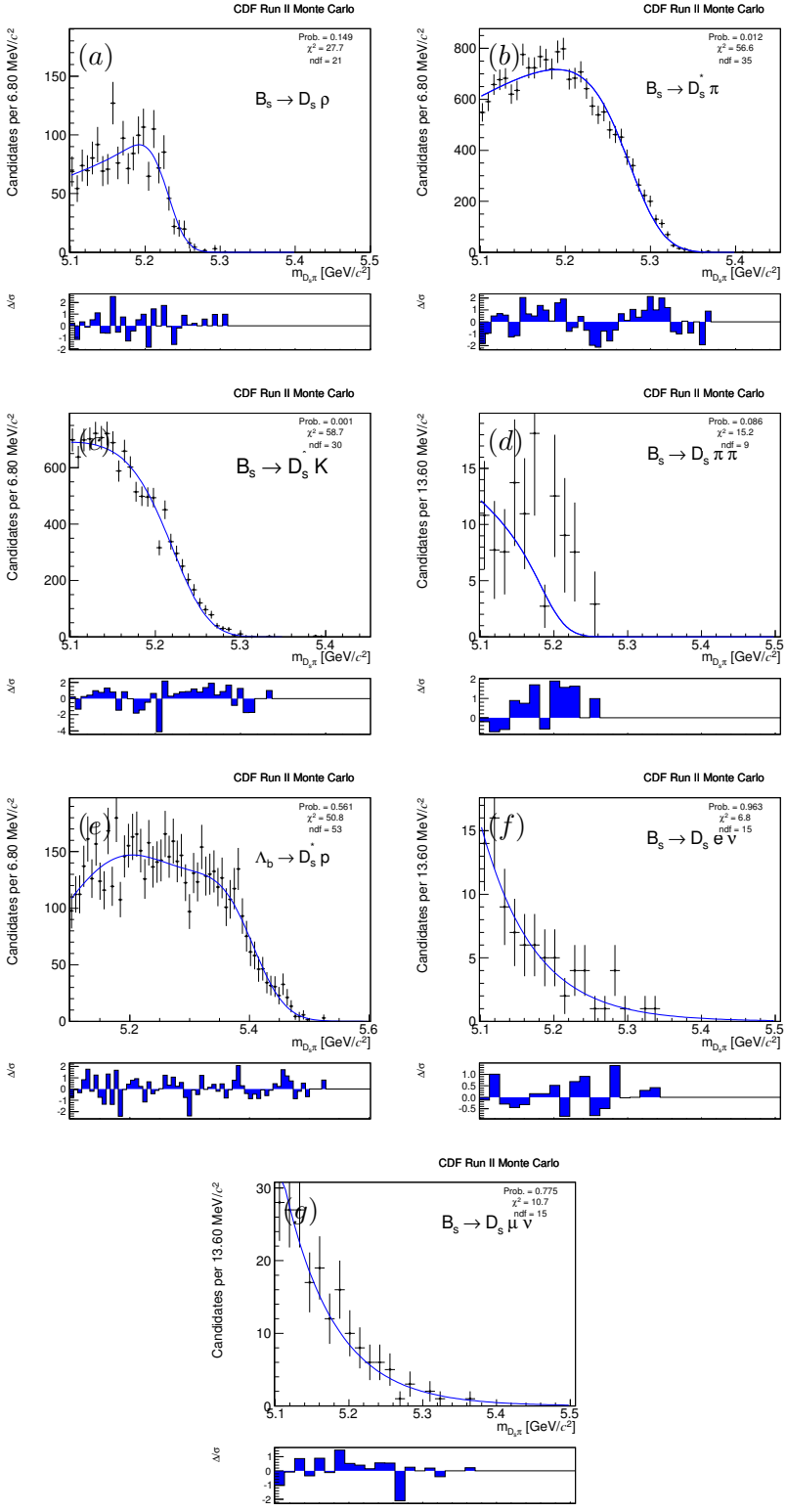

Figure 25: Mass template of various mis-reconstructed decay modes: $B_s^0 \to D_s^- \rho^+$ (a), $B_s^0 \to D_s^{\star -} \pi^+$ (b), $B_s^0 \to D_s^{\star \pm} K^{\mp}$ (c), $B_s^0 \to D_s^{(\star)-} \pi^+ \pi^0$ (d), $\Lambda_b^0 \to D_s^{\star -} p$ (e), $B_s^0 \to D_s^- e^+ \nu_e$ (f) and $B_s^0 \to D_s^- \mu^+ \nu_\mu$ (g).

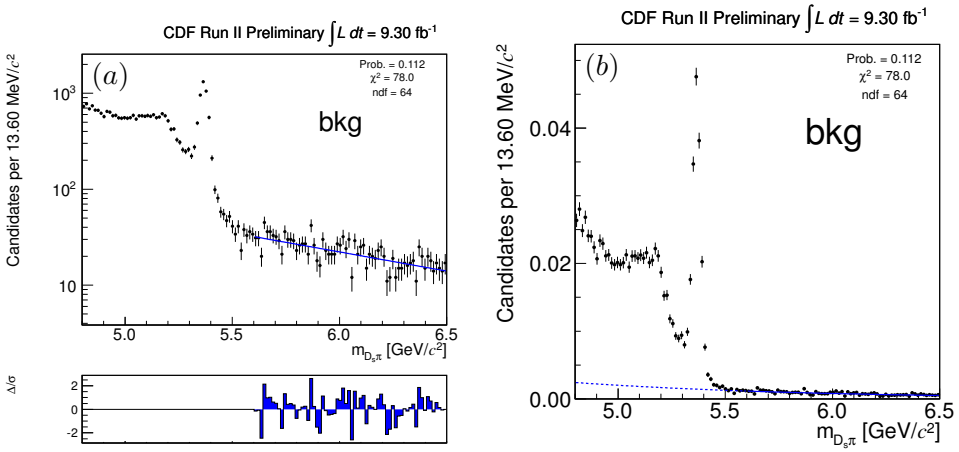

Figure 26: Mass template of the background term. (a) mass distribution of the pair $D_s\pi$ in log-scale. (b) mass distribution of the pair $D_s\pi$ in linear-scale, where the background p.d.f. is extrapolated in the lower mass region.

the central fit, is fitted with an exponential function (see eq. (11)), as shown in Figure 26. The value of the slope of the exponential function obtained is $a_{\rm bkg} = (0.98 \pm 0.10)\,({\rm GeV}/c^2)^{-1}$, and it is a fixed parameter in the central fit. A systematic uncertainty on the level of knowledge of this slope will be assessed, details are reported in section 7.

6.3 PID probability density function

Here we report p.d.f.s for kaons, protons and electrons for some decays (see figure 27). The p.d.f.s of the other decays are similar. We assume that the background is mostly composed by pions and kaons. Therefore the p.d.f. of the background can be written as:

$$\wp^{\rm PID}_{\rm bkg}({\rm res}_{\pi^+}) = f^{\pi}_{\rm bkg} \cdot \wp_{\pi^+}({\rm res}_{\pi^+}) + (1 - f^{\pi}_{\rm bkg}) \cdot \wp_{K^+}({\rm res}_{\pi^+}), \quad (12)$$

$$\wp^{\rm PID}_{\rm bkg}({\rm res}_{\pi^-}) = f^{\pi}_{\rm bkg} \cdot \wp_{\pi^-}({\rm res}_{\pi^-}) + (1 - f^{\pi}_{\rm bkg}) \cdot \wp_{K^-}({\rm res}_{\pi^-}), \quad (13)$$

where $\wp_{\pi^\pm}({\rm res}_{\pi^\pm}), \wp_{K^\pm}({\rm res}_{\pi^\pm})$ are the p.d.f.s for positively- and negatively-charged pions and kaons in the pion mass hypothesis, respectively. $f^{\pi}_{\rm bkg}(1 - f^{\pi}_{\rm bkg})$ is the inclusive charge-averaged fraction of pions(kaons) in the background.

6.4 Fit implementation

The fit of composition is performed on the D_sh data sample. We use only candidates whose discriminating observables satisfy the following conditions: the invariant $D_s\pi$ mass within $[5.1, 5.6]\,{\rm GeV}/c^2$, and $|{\rm res}_\pi| < 6$. The requirement $|{\rm res}_\pi| > 6$ excludes candidates with unlikely values of observed dE/dx to reduce a small contamination from tracks with corrupted dE/dx information. The total number of fitted candidates is $N = 12453$.

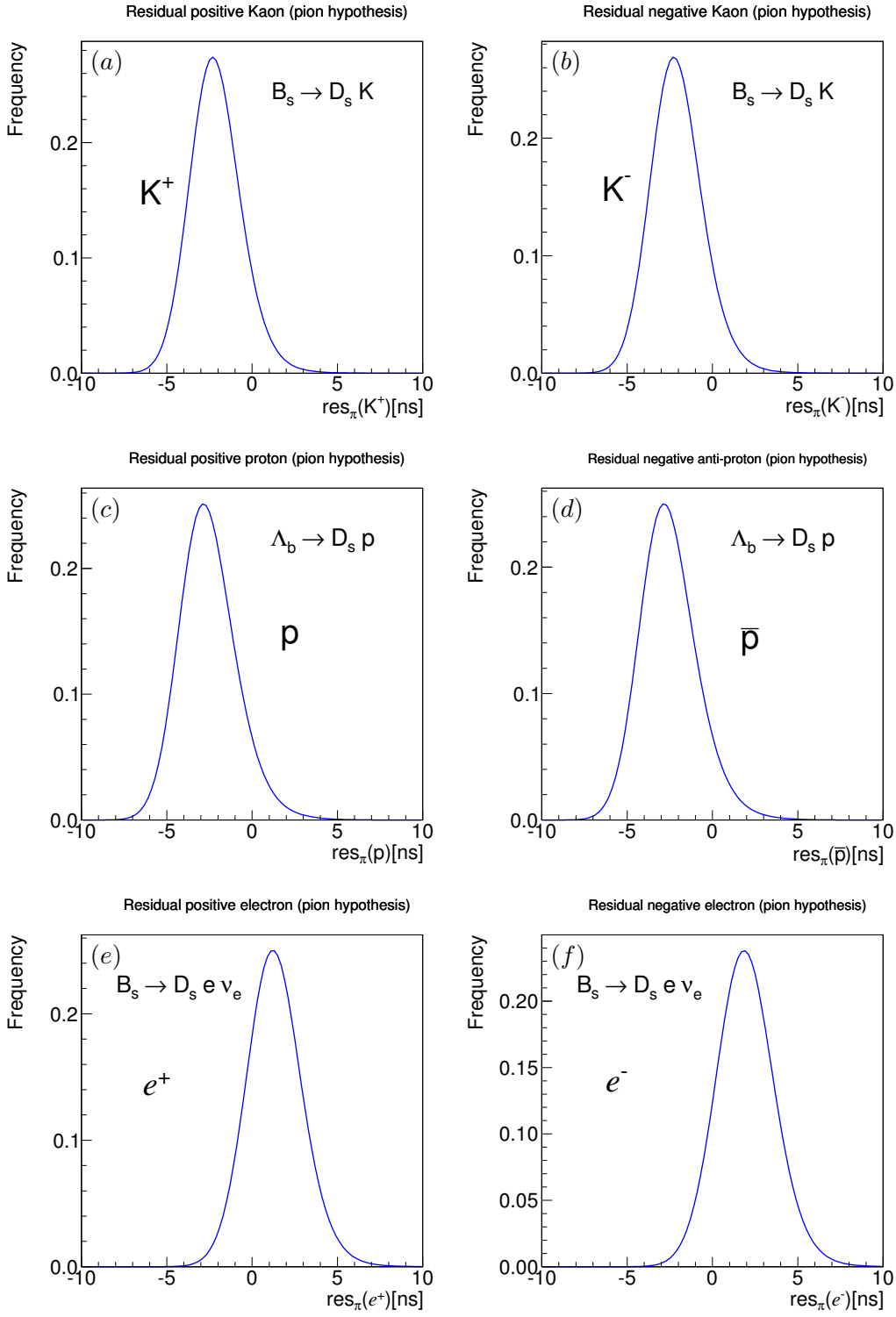

Figure 27: PID templates of the daughter tracks of some decays. (a, b) kaons from $B_s^0 \to D_s^\pm K^\mp$, (c, d) protons from $\Lambda_b^0 \to D_s^- p$, (e, f) electrons from $B_s^0 \to D_s^{(\star)-} e^+ \nu_e$.

Table 5: Table of the parameter constrained in the fit of composition and their values.

parameter (λ_i)	branching fractions [**?**]	constrain ($\mu_i \pm \sigma_i$)
	$\mathcal{B}(B_s^0 \to D_s^- \pi^+) = (3.2 \pm 0.4) \times 10^{-3}$	–
$\frac{f_{B^0 \to D_s^{(\star)+} \pi^-}}{f_{B_s^0 \to D_s^- \pi^+}}$	$\frac{\mathcal{B}(B^0 \to D_s^{(\star)+} \pi^-)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = \frac{(2.16 \pm 0.26) \times 10^{-5}}{(3.2 \pm 0.4) \times 10^{-3}}$	$(3.3 \pm 0.5) \times 10^{-2}$
$\frac{f_{B^0 \to D_s^{(\star)-} K^+}}{f_{B_s^0 \to D_s^- \pi^+}}$	$\frac{\mathcal{B}(B^0 \to D_s^{(\star)-} K^+)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = \frac{(2.2 \pm 0.5) \times 10^{-5}}{(3.2 \pm 0.4) \times 10^{-3}}$	$(2.6 \pm 0.7) \times 10^{-2}$
$\frac{f_{B_s^0 \to D_s^{\star -} \pi^+}}{f_{B_s^0 \to D_s^- \pi^+}}$	$\frac{\mathcal{B}(B_s^0 \to D_s^{\star -} \pi^+)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = 0.65^{+0.15}_{-0.13} \pm 0.07$	0.45 ± 0.11
$\frac{f_{B_s^0 \to D_s^- \rho^+}}{f_{B_s^0 \to D_s^- \pi^+}}$	$\frac{\mathcal{B}(B_s^0 \to D_s^- \rho^+)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = 2.3 \pm 0.4 \pm 0.2$	0.31 ± 0.08

6.4.1 Gaussian constrains

Branching fractions of several mis-reconstructed decay modes are known. Therefore we added a Gaussian constraint to the fit for each known mode to help the convergence. For each of them the Likelihood function $\mathscr{L}$ multiplies a Gaussian term $\mathscr{G}(\lambda_i; \mu_i, \sigma_i)$, where λ_i is the parameter to be constrained, μ_i is the expected value of such a parameter and σ_i is its total uncertainty. For instance, if we consider a generic $B \to DX$ decays, and its branching fraction relative to the $B_s^0 \to D_s^- \pi^+$ decay mode is known, the $\lambda_{B \to DX}$ parameter can be written as:

$$\lambda_{B \to DX} = \frac{f_{B \to DX}}{f_{B_s^0 \to D_s^- \pi^+}}. \tag{14}$$

$f_{B \to DX}$ and $f_{B_s^0 \to D_s^- \pi^+}$ are the observed relative fraction in our data sample. However, to translate the information from a ratio of branching fractions to a ratio of relative fractions we need to account for several different factors, as the reconstruction efficiency correction ϵ (see section 6.6 for the definition). In fact we obtain that

$$\frac{f_{B \to DX}}{f_{B_s^0 \to D_s^- \pi^+}} = \frac{\mathcal{B}(B \to DX)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} \cdot \frac{f_q}{f_s} \cdot \frac{\epsilon(B \to DX)}{\epsilon(B_s^0 \to D_s^- \pi^+)} \cdot \frac{\mathcal{B}(D \to Y)}{\mathcal{B}(D_s \to \phi\pi)}, \tag{15}$$

where $\epsilon(B_s^0 \to D_s^- \pi^+)$ and $\epsilon(B \to DX)$ are respectively the reconstruction efficiencies for the $B_s^0 \to D_s^- \pi^+$ and for $B \to DX$. Y is the final state of the D decay and f_q ($q = d, s$) is the probability that a b-quark hadronized in a B meson with a q-quark. If the $\mathcal{B}(B \to DX)$ is known, we can easily calculate the value of $\lambda_{B \to DX}$, and its uncertainty using the equation (15). All input branching fractions, f_s, f_q come from PDG [19], while efficiency corrections from CDF simulation. Table 5 reports the constrained parameters in the fit, the known values of branching fractions ratios and the applied constraints (third column).

6.4.2 Mass shift and resolution

By the comparison between simulated and real $B_s^0 \to D_s^- \pi^+$ decays we observe that simulation does not accurately reproduce data. This is a known feature of the CDF simulation. In particular the simulated invariant mass distribution is shifted by few MeV with respect to data, and the mass resolution is smaller, about 10%. Table 6 reports the values obtained for data and simulation.

To account for these differences we added to the fit of composition two free parameters. One is a global mass shift, assumed the same for all the decays, to allow the mass scale to be determined by the real data. The other one is a mass resolution scale factor. This was applied only to the $B_s^0 \to D_s^- \pi^+$ decay p.d.f., which is reconstructed with a correct mass assignments. In particular σ_1, σ_2 of eq. (10) are re-defined as $\sigma_1 \to s\sigma_1$ and $\sigma_2 \to s\sigma_2$, where s is a free parameter of the fit, close to the unit. Instead for all other mis-reconstructed decays, including also $B_s^0 \to D_s^\pm K^\mp$ mode, we did not apply any mass resolution scale factor, because the shapes of their mass distributions is much wider and mainly determined by the wrong mass assignment, which dominates over the 10% effect on resolution. We assess systematic uncertainty due to neglecting the resolution corrections on the other decays in section 7.

Table 6: Mass mean values and widths for data and simulation for the $B_s^0 \to D_s^- \pi^+$ decay.

parameter	Data	Simulation
$\mu\,[\mathrm{GeV}/c^2]$	5.36574 ± 0.00036	5.37431 ± 0.00018
$\sigma\,[\mathrm{GeV}/c^2]$	$(2.001 \pm 0.036) \times 10^{-2}$	$(1.7861 \pm 0.0015) \times 10^{-2}$

6.4.3 Simultaneous fit of $D_s^- h^+$ and $D_s^+ h^-$ samples

The probability density function of the fit of composition we wrote so far, does not distinguish between $D_s^- h^+$ and $D_s^+ h^-$ final sample. However since the PID response is separately parameterized for negatively- and positively-charged particles (see section 6.3), we must account for that in the fit. Therefore we perform a simultaneous fit of these two sub-samples and we can write the total Likelihood function $\mathscr{L}$ of all events:

$$\mathscr{L}(\boldsymbol{\theta}) = \prod_{i=1}^{N_+} \mathscr{L}_i^+(\boldsymbol{\theta}|\boldsymbol{x}_i^+) \cdot \prod_{i=1}^{N_-} \mathscr{L}_i^-(\boldsymbol{\theta}|\boldsymbol{x}_i^-)$$

where the index i runs over the events. N_+ is the number of events of the $D_s^- h^+$ sample, N_- is the number of events of $D_s^+ h^-$ sample, and where $N = N_+ + N_-$ is the total number of events. $\boldsymbol{\theta}$ is the vector of parameters that we want to estimate, $\boldsymbol{x}^+$ ($\boldsymbol{x}^-$) is the vector of the discriminating observables $\boldsymbol{x}_i^+ = \{m_{D_s\pi}, \mathrm{res}_{\pi^+}\}_i$ ($\boldsymbol{x}_i^- = \{m_{D_s\pi}, \mathrm{res}_{\pi^-}\}_i$). As in section 6.1 we have to account for the poissonian uncertainty due to the finite size of the total sample N, plus a term for the binomial uncertainty due to the fact we splitted the sample in two sub-samples $N = N_+ + N_-$. This means that we have to consider N as a Poisson variable with mean ν and and that

N_+ and N_- are binomially distributed, with a probability p to have N_+, and $1-p$ to have N_- events, when the sum is constrained to be, in our specific case, equal to N. Then the new extended Likelihood function can be written as:

$$\mathscr{L}(\nu, p, \boldsymbol{\theta}) = \frac{\nu^N}{N!} e^{-\nu} \cdot \frac{N!}{N_+!(N-N_+)!} p^{N_+} (1-p)^{N-N_+} \cdot \prod_{i=1}^{N_+} \mathscr{L}_i^+(\boldsymbol{\theta}|\boldsymbol{x}_i^+) \cdot \prod_{i=1}^{N_-} \mathscr{L}_i^-(\boldsymbol{\theta}|\boldsymbol{x}_i^-).$$

All the parameters $\boldsymbol{\theta}$ remain unchanged, as described in the previous sections. They do not double since they are in common (charge averaged) between the two sub-samples during the minimization.

6.5 Fit results

The fit of composition was performed by minimizing the quantity $-2\ln(\mathscr{L})$, as defined in equation (6), using the MINUIT numerical minimization package [31]. Table 7 reports the results, while the corresponding correlation matrix is discussed in subsection 6.5.1. Table 8 reports the yields returned from the fit.

As expected from the current known branching fractions involved, we have a sizable ($\approx 80\%$ of the sample) contribution from the three modes $B_s^0 \to D_s^- \pi^+$, $B_s^0 \to D_s^- \rho^+$ and $B_s^0 \to D_s^{\star -} \pi^+$. All the other decay modes have comparable fractions and share about the 20% of the sample. The global shift and mass resolution scale factor parameters are in agreement with the simple estimate done in subsection 6.4.2 (see table 6). The values returned from the fit are consistent with nominal values of the b-hadrons masses, and are consistent with what we observe in other similar analyses in CDF [4, 10]. Table 7 also reports the uncorrected value of the ratio of branching fractions $f(B_s^0 \to D_s^{\pm} K^{\mp}) / f(B_s^0 \to D_s^- \pi^+)$.

6.5.1 Correlation matrix

The correlation matrix corresponding to the fit of composition is shown in table 9. The correlation coefficients are defined as $\rho_{ij} = \mathrm{Cov}(\theta_i, \theta_j)/\sigma_{\theta_i}\sigma_{\theta_j}$, where $\mathrm{Cov}(\theta_i, \theta_j)$ is the off-diagonal element of the estimated covariance matrix of the fit. The large correlation coefficients, related to the relative fraction of the $B_s^0 \to D_s^- \pi^+$ and the $B_s^0 \to D_s^{\pm} K^{\mp}$ decay modes are: $\rho_{6,4} \approx -50\%$ and $\rho_{6,5} \approx -34\%$, $\rho_{13,4} \approx -31\%$ and $\rho_{13,5} \approx -34\%$. These are due to the limited separation power of the fit to disentangle $\Lambda_b^0 \to D_s^{(\star)-} p$ and $B^0 \to D^- \pi^+$ decay modes from signals. In fact they lay down exactly under the $B_s^0 \to D_s h$ modes, and the PID is not helping too much. In fact the pions of the $B^0 \to D^- \pi^+$ are not distinguishable from pions of $B_s^0 \to D_s^- \pi^+$, and protons of $\Lambda_b^0 \to D_s^{(\star)-} p$ are very similar to kaons of $B_s^0 \to D_s^{\mp} K^{\pm}$. Other considerable correlations are $\rho_{12,2} \approx -57\%$ and $\rho_{6,1} \approx -48\%$. $\rho_{12,2}$. This can be explained considering that the decay $B_s^0 \to D_s^- \pi^+ \pi^0$ is at the low edge of the fitted mass region and a global shift influence its fraction. Instead, the amount of background can vary the fraction of the mode $\Lambda_b^0 \to D_s^{(\star)-} p$, which generates $\rho_{6,1}$.

Table 7: Results of the fit of composition. The last row report the legend to convert the parameter into physics quantity for interpreting the correlation matrix. C-conjugate modes are implied.

parameter	value	parameter #
$f_{\rm bkg}$	0.1246 ± 0.0069	1
global shift [GeV/c^2]	0.00768 ± 0.00002	2
scale factor ($B_s^0 \to D_s^- \pi^+$)	1.094 ± 0.017	3
$f_{B_s^0 \to D_s^- \pi^+}$	0.4127 ± 0.0057	4
$f_{B_s^0 \to D_s^\pm K^\mp}$	0.0307 ± 0.0031	5
$f_{\Lambda_b^0 \to D_s^- p}$	0.0104 ± 0.0026	6
$f_{\Lambda_b^0 \to D_s^{\star -} p}$	fixed to $f_{\Lambda_b^0 \to D_s^- p}$	-
$f_{B^0 \to D_s^{(\star)+} \pi^-}$	0.0148 ± 0.0020	7
$f_{B^0 \to D_s^{(\star)-} K^+}$	0.0096 ± 0.0025	8
$f_{B_s^0 \to D_s^- \rho^+}$	0.136 ± 0.013	9
$f_{B_s^0 \to D_s^{\star \pm} K^\mp}$	0.0426 ± 0.0059	10
$f_{B_s^0 \to D_s^- e^+ \nu_e}$	0.0375 ± 0.0051	11
$f_{B_s^0 \to D_s^- \mu^+ \nu_\mu}$	fixed to $f_{B_s^0 \to D_s^- e^+ \nu_e}/\epsilon_{\rm rel}$	-
$f_{B_s^0 \to D_s^- \pi^+ \pi^0}$	0.0141 ± 0.0097	12
$f_{B^0 \to D^- \pi^+}$	0.0157 ± 0.0037	13
$f_{B_s^0 \to D_s^{\star -} \pi^+}$	0.248 ± 0.015	$1 - \sum_{j=1}^{s-1} f_j$
$f_{\rm bkg}^{\pi}$	0.738 ± 0.041	14
ν	12453 ± 112	15
p	0.508 ± 0.004	16
$\dfrac{f_{B_s^0 \to D_s^\pm K^\mp}}{f_{B_s^0 \to D_s^- \pi^+}}$	0.0744 ± 0.0076	-

Table 8: Yields returned from the fit of composition. C-conjugate modes are implied.

mode	number of events
$\mathcal{N}(B_s^0 \to D_s^- \pi^+)$	4498 ± 138
$\mathcal{N}(B_s^0 \to D_s^\pm K^\mp)$	335 ± 40
$\mathcal{N}(\Lambda_b^0 \to D_s^- p) + \mathcal{N}(\Lambda_b^0 \to D_s^{\star -} p)$	114 ± 31
$\mathcal{N}(B^0 \to D_s^+ \pi^-) + \mathcal{N}(B^0 \to D_s^{\star +} \pi^-)$	162 ± 24
$\mathcal{N}(B^0 \to D_s^- K^+) + \mathcal{N}(B^0 \to D_s^{\star -} K^+)$	104 ± 29
$\mathcal{N}(B_s^0 \to D_s^- \rho^+)$	1480 ± 170
$\mathcal{N}(B_s^0 \to D_s^{\star \pm} K^\mp)$	464 ± 71
$\mathcal{N}(B_s^0 \to D_s^- e^+ \nu_e)$	409 ± 63
$\mathcal{N}(B_s^0 \to D_s^- \mu^+ \nu_\mu)$	188 ± 29
$\mathcal{N}(B_s^0 \to D_s^- \pi^+ \pi^0)$	153 ± 108
$\mathcal{N}(B^0 \to D^- \pi^+)$	172 ± 43
$\mathcal{N}(B_s^0 \to D_s^{\star -} \pi^+)$	2709 ± 209

Table 9: Correlation matrix returned by the fit of composition. The legend for the fit parameters is reported in the third column of table 7.

PARAMETER NO.	GLOBAL	CORRELATION COEFFICIENTS 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.54906	1.000															
2	0.77032	0.044	1.000														
3	0.27440	-0.001	0.038	1.000													
4	0.60180	0.393	0.119	0.115	1.000												
5	0.53644	0.112	0.196	0.053	0.053	1.000											
6	0.77022	-0.482	-0.128	-0.180	-0.507	-0.340	1.000										
7	0.25389	0.052	-0.015	-0.055	0.043	-0.026	-0.072	1.000									
8	0.45339	0.022	-0.176	-0.004	0.018	-0.147	-0.002	-0.023	1.000								
9	0.45529	0.022	0.149	0.008	-0.010	0.042	0.027	0.129	-0.148	1.000							
10	0.62024	-0.023	0.331	0.065	-0.021	0.202	-0.173	0.060	-0.321	0.054	1.000						
11	0.54671	-0.035	0.262	-0.023	-0.090	0.047	0.039	-0.022	-0.090	-0.023	0.252	1.000					
12	0.76518	0.102	-0.573	0.016	0.007	-0.032	-0.049	0.053	0.174	-0.315	-0.267	-0.459	1.000				
13	0.54971	-0.207	-0.273	-0.154	-0.307	-0.337	0.286	-0.094	0.127	0.040	-0.196	-0.189	0.162	1.000			
14	0.65481	-0.309	-0.077	-0.134	-0.312	-0.096	0.495	-0.012	-0.061	0.163	0.167	0.076	-0.167	0.154	1.000		
15	0.00028	0.000	0.000	-0.000	0.000	0.000	-0.000	0.000	-0.000	0.000	0.000	-0.000	0.000	-0.000	0.000	1.000	
16	0.01640	0.000	0.008	-0.008	0.000	0.002	-0.004	0.001	-0.000	0.001	0.006	-0.004	-0.004	-0.002	-0.003	-0.004	1.000

6.5.2 Fit projections

In order to test the goodness of our fit, we compare the distributions of data with the joint p.d.f. corresponding to the Likelihood function with the maximized set of parameters $\bar{\boldsymbol{\theta}}$. If $\boldsymbol{x} = x_1, \ldots, x_n$ is a generic vector of observables and $\wp(\boldsymbol{x}, \bar{\boldsymbol{\theta}})$ is the probability density function of the observables $\boldsymbol{x}$, we can define the *projection* onto the observable x_i as the following one-dimensional function:

$$\wp(x_i, \bar{\boldsymbol{\theta}}) = \int \wp(\boldsymbol{x}, \bar{\boldsymbol{\theta}})\, dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n, \tag{16}$$

which is the predicted distribution for x_i under the assumed values for the fit parameters, and can be overlaid to the experimental data. This allows a way of detecting possible discrepancy between the observed distributions and the model. Distributions of the discriminating observables with the fit projection overlaid are shown in figures 28-29. The distributions of individual components are also shown. To better visualize the agreement between the PID discriminating observable and the data we complemented the projection of res_π, with the projection of its mean value ($\langle \mathrm{res}_\pi \rangle$) as function of the invariant mass (see fig 30). The agreement between data and fit projections is satisfactory.

6.6 Efficiency correction

In order to translate the parameters returned from the fit of composition into a physics measurement we need to apply the correction for the different reconstruction efficiency. In general, the efficiency for each mode is defined as the ratio between the number of events passing the final selection (N_{passing}) and the number of real events produced (N_{produced}):

$$\epsilon = \frac{N_{\mathrm{passing}}}{N_{\mathrm{produced}}}. \tag{17}$$

This term accounts for all the acceptance effects. It includes the trigger efficiency and the efficiency of the offline reconstruction and selection. For our measurement we extract the efficiency from the CDF Monte Carlo simulation. Any geometric acceptance effect is properly taken into account, since the simulation reproduces the kinematic distributions of the decays and it includes an accurate description of the detector geometry. Any possible discrepancy between real data and the simulation vanishes in the efficiency ratio between two different modes.

Thus for the $B_s^0 \to D_s^\pm K^\mp$ we can write

$$\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = \frac{f_{B_s^0 \to D_s^\pm K^\mp}}{f_{B_s^0 \to D_s^- \pi^+}} \cdot \frac{\epsilon(B_s^0 \to D_s^- \pi^+)}{\epsilon(B_s^0 \to D_s^\pm K^\mp)}, \tag{18}$$

where $\epsilon(B_s^0 \to D_s^- \pi^+)$ and $\epsilon(B_s^0 \to D_s^\pm K^\mp)$ are respectively the reconstruction efficiencies for the $B_s^0 \to D_s^- \pi^+$ and for $B_s^0 \to D_s^\pm K^\mp$. The efficiency correction extracted from simulation is

$$\frac{\epsilon(B_s^0 \to D_s^- \pi^+)}{\epsilon(B_s^0 \to D_s^\pm K^\mp)} = 1.044 \pm 0.007, \tag{19}$$

where the uncertainty is due to the finite statistics of the simulated sample. It corresponds to the Poisson fluctuation of the number of events passing the selection.

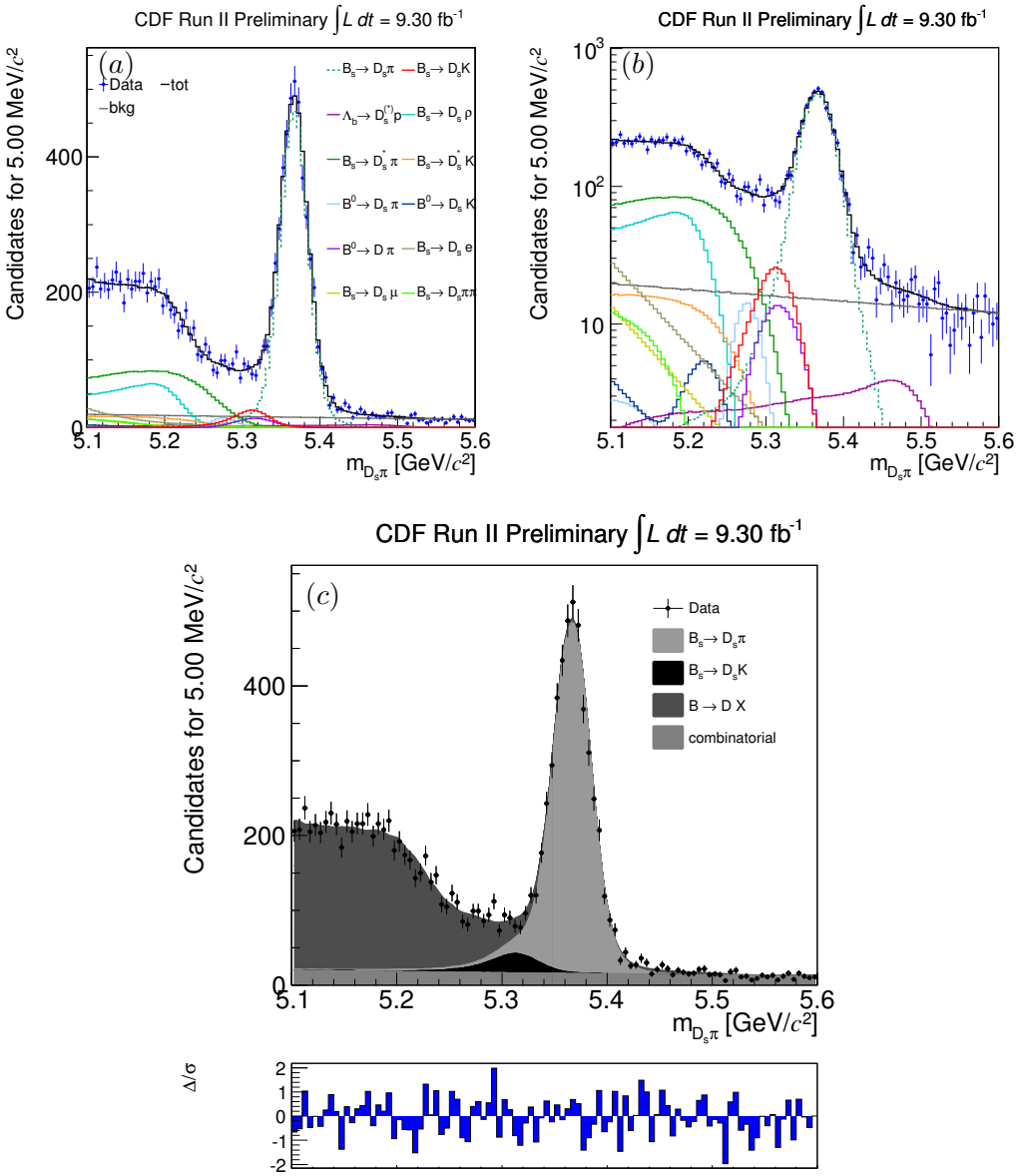

Figure 28: $m_{D_s\pi}$ distribution with the fit projection overlaid: (a) the fit projection with all the fit components (logarithmic scale (b)), (c) the components are grouped for clarity. The residual plot at the bottom of the figure (c) shows the number of σ discrepancy (data minus projection).

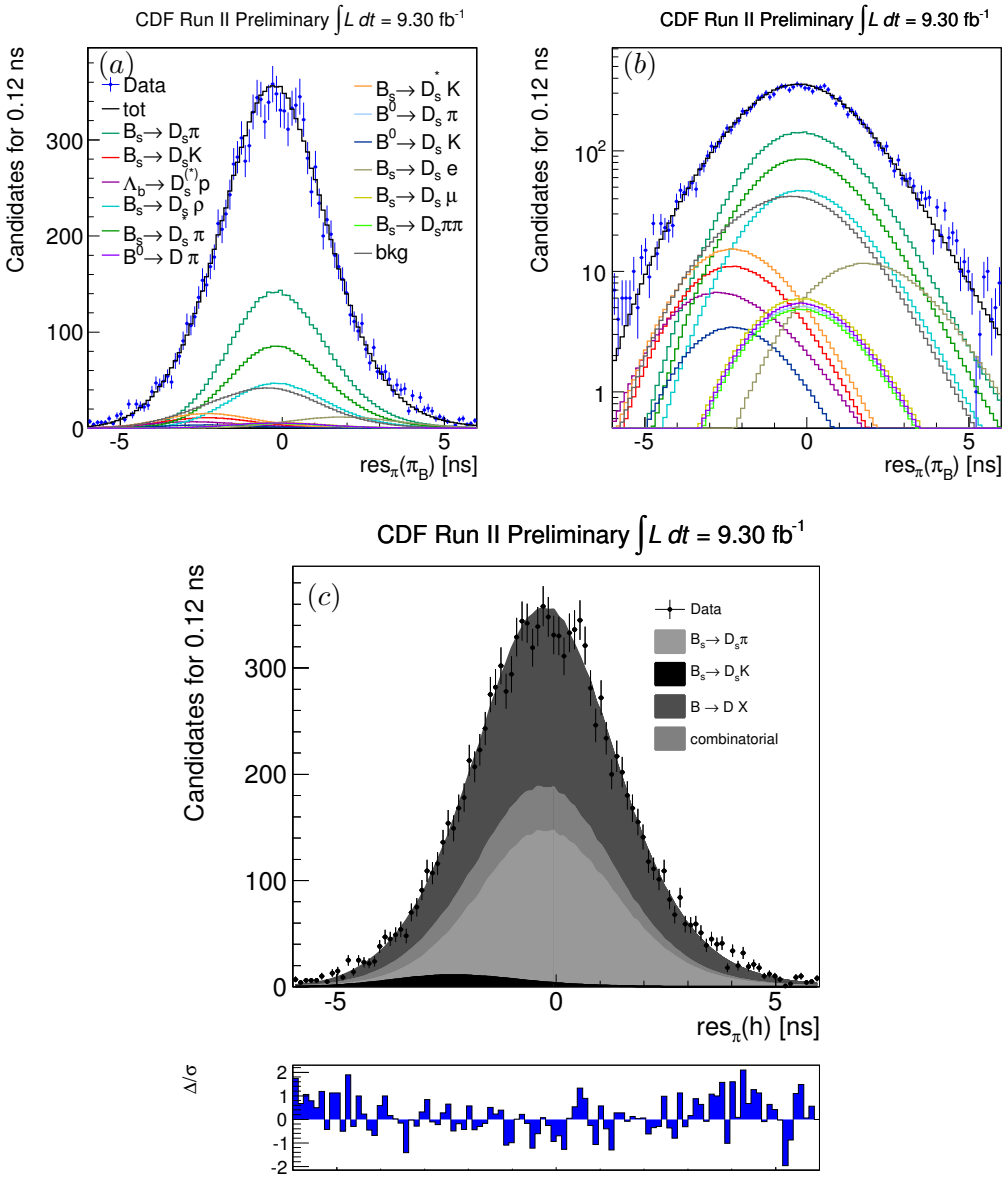

Figure 29: res_π distribution with the fit projection overlaid: (a) the fit projection with all the fit components (logarithmic scale (b)), (c) the components are grouped for clarity. The residual plot at the bottom of the figure (c) shows the number of σ discrepancy (data minus projection).

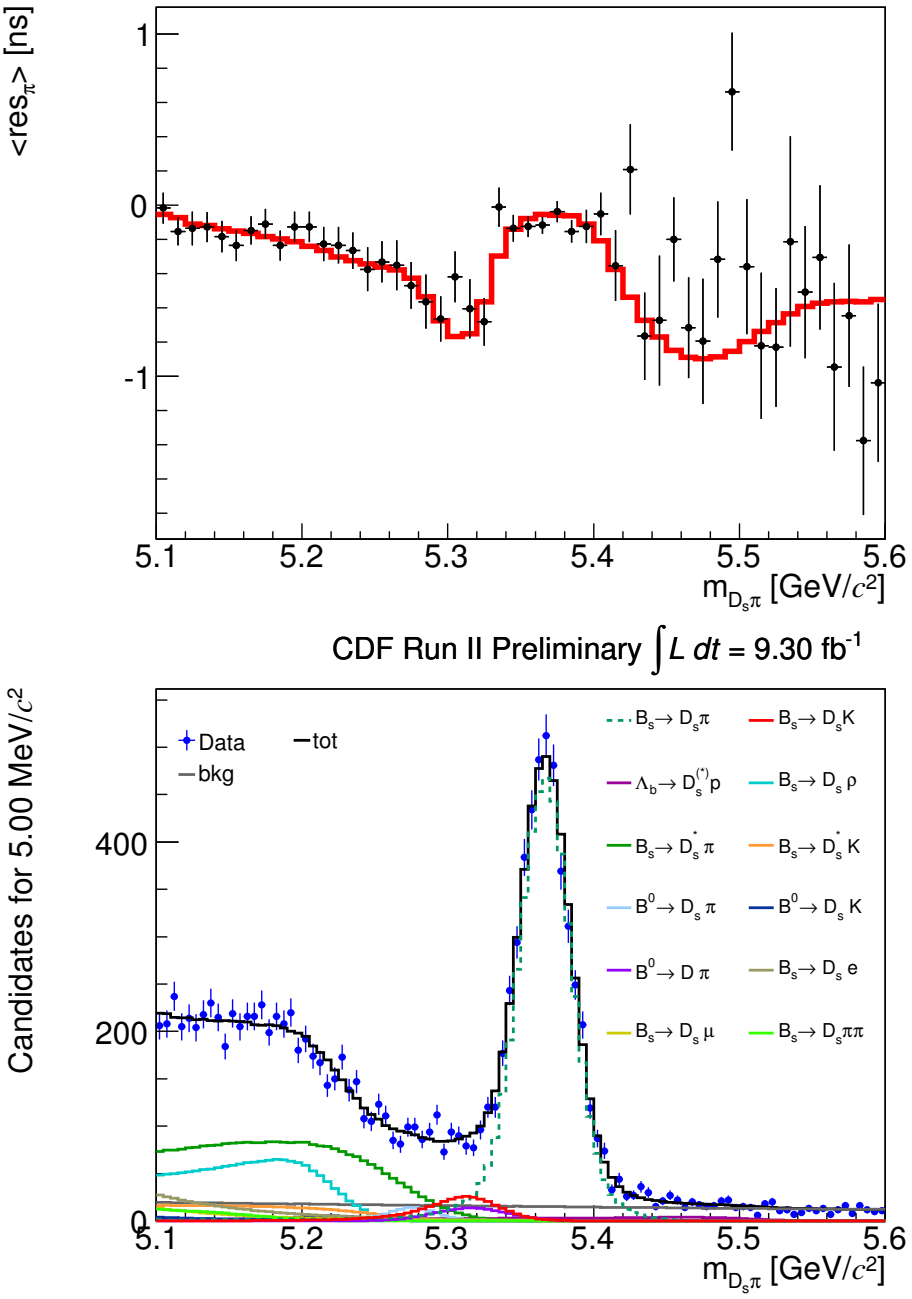

Figure 30: Fit projection onto the variable $\langle \mathrm{res}_\pi \rangle$ as function of the $m_{D_s\pi}$ (top figure).

6.7 Corrected result

In summary, using the ratio of relative fractions returned from the fit of composition

$$\frac{f_{B_s^0 \to D_s^\pm K^\mp}}{f_{B_s^0 \to D_s^- \pi^+}} = 0.0744 \pm 0.0076 \tag{20}$$

and the efficiency ratio returned from CDF simulation

$$\frac{\epsilon(B_s^0 \to D_s^- \pi^+)}{\epsilon(B_s^0 \to D_s^\pm K^\mp)} = 1.044 \pm 0.007, \tag{21}$$

we obtain the measurement of the following ratio of branching fractions:

$$\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = 0.0777 \pm 0.0079\,(\text{stat}), \tag{22}$$

where the uncertainty is only statistical.

7 Evaluation of systematic uncertainties

The measurement described in this note focuses on a ratio of branching fractions of kinematically similar decay modes. We expect that most systematic effects related to the individual modes, e.g., the uncertainty on the integrated luminosity of the sample, will cancel out in the ratio.

To evaluate the systematic uncertainty we used the following method. For each source of systematic effects, s, we varied the value of s within a range of $\pm 1\sigma_s$, where σ_s is the statistical uncertainty on the parameter s. The resulting systematic uncertainty associated to s is the largest difference between the results of the analysis of the samples with alternative configurations, and the results of the sample with the nominal configuration.

7.1 Uncertainty on the nominal b-hadron masses (nominal masses)

The B^0, B_s^0, and Λ_b^0 masses are external inputs of the Monte Carlo simulation and therefore inputs to our p.d.f.s. To evaluate the systematic uncertainty associated to our limited experimental knowledge of nominal input masses we repeated our fit after shifting independently the B^0, B_s^0, and Λ_b^0 input masses within $\pm 1\sigma$ uncertainty. We fitted the eight possible combinations of B^0, B_s^0, and Λ_b^0 masses by independently increasing (decreasing) by one statistical standard deviation world-average mass values [**?**]: $m_{B^0} = (5279.58 \pm 0.17)\,\text{MeV}/c^2$, $m_{B_s^0} = (5366.77 \pm 0.24)\,\text{MeV}/c^2$ and $m_{B^0} = (5619.4 \pm 0.7)\,\text{MeV}/c^2$. The largest discrepancy between the results of the analysis with alternative masses configuration and the results of the sample with the nominal configuration was taken as the systematic uncertainty.

Since we added to the fit a global mass shift parameter, as explained in the subsection 6.4.2, we do not have any systematic uncertainty associated to the global mass scale uncertainty. The uncertainty of the knowledge of the absolute mass scale is already included in the statistical error returned from the fit of composition.

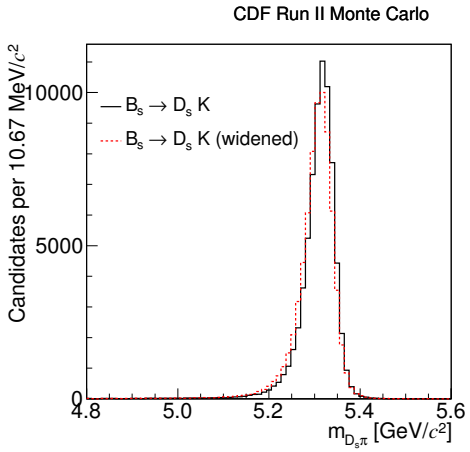

Figure 31: Comparison of the invariant mass distribution of the $B_s^0 \to D_s^{\pm} K^{\mp}$ before and after the transformation in eq. (23).

7.2 Uncertainty on mass resolution (mass resolution)

As explained in subsection 6.4.2, we added to the fit a scale factor to inflate the width of the p.d.f. of the $B_s^0 \to D_s^- \pi^+$ extracted from the Monte Carlo, since we observe a smaller mass resolution in it with respect to data. This has been done only for the $B_s^0 \to D_s^- \pi^+$ decay, because we assumed negligible the net effect of the discrepancy between real data and simulation for the mis-reconstructed decays, since the enlargement is mainly due to the wrong mass assignment. We assess a systematic uncertainty on these assumptions.

From the central fit we exactly know the size of the scale factor $s = 1.094 \pm 0.017$ of the $B_s^0 \to D_s^- \pi^+$ which is fully reconstructed with the correct mass assignment. However we cannot use this factor to enlarge the mis-reconstructed modes since their mass invariant spectrum is sculpted by the fact that we assigned wrong masses to the particles in the final state, and the final effect is much smaller of what we observe for the $B_s^0 \to D_s^- \pi^+$ decays.

For instance to estimate the scaling factor of the "mis-reconstructed" $B_s^0 \to D_s^{\mp} K^{\pm}$ decays (and $\Lambda_b^0 \to D_s^- p$, $B^0 \to D_s^- K^+$) we scaled our simulation event-by-event with the following transformation:

$$m^i_{D_s\pi} \to \frac{m^i_{D_s\pi} - \mu(p^i_{D_s}, p^i_{\pi})}{s} + \mu(p^i_{D_s}, p^i_{\pi}), \tag{23}$$

where $\mu(p^i_{D_s}, p^i_{\pi})$ is the expected $D_s\pi$-mass $\langle m_{D_s\pi} \rangle$ given the momenta p_{D_s} and p_{π}, when we assign the pion mass to the kaon (or to the proton) in the final state, and s is the scaling factor obtained from the fit of composition for the $B_s^0 \to D_s^- \pi^+$ decay. Figure 31 reports the comparison of the invariant mass distribution of the $B_s^0 \to D_s^{\pm} K^{\mp}$ before and after the transformation in eq. (23). As expected, the widening of the distribution is small, but not completely negligible. In conclusion to assess the systematic uncertainty we re-adapted the fit of composition where: 1) the mass distribution of the $B_s^0 \to D_s^{\pm} K^{\mp}$, $\Lambda_b^0 \to D_s^- p$ and $B^0 \to D_s^- K^+$ is scaled using the transformation of equation 23; 2) the mass distribution of the $B^0 \to D_s^- \pi^+$ is scaled as the $B_s^0 \to D_s^- \pi^+$, 3) no scaling is applied to the mass distribution of the

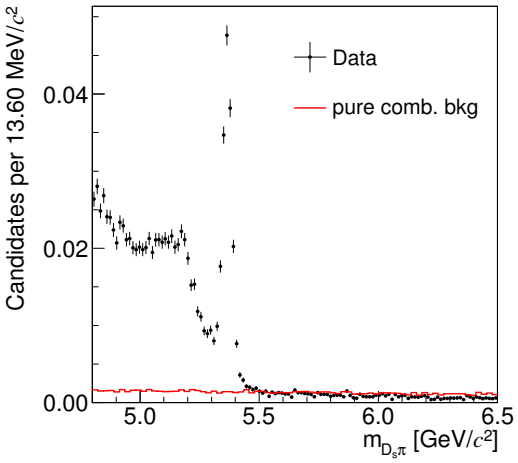

Figure 32: Comparison between the pure combinatorial background sample and the data-background.

decays in which some particles escape from the detection, as a neutrino, a photon or a neutral pion, because in this case the mass invariant shape is mainly sculpted by the wrong mass assignments and by the fact that part of the energy of the decay is lost; 4) we add a free parameter for the scale of the $B^0 \to D^-\pi^+$, since we do not know calculate the scaling factor for this decay mode in a simple way. The difference between this fit and the central fit is our systematic uncertainty.

7.3 Uncertainty on the combinatorial background mass term

Since our central analysis assume an empirical mass model for the combinatorial background we assess a systematic uncertainty due to our limited knowledge of the real distribution. We use an exponential shape, where the slope $a_{\rm bkg} = (0.98 \pm 0.10)$ $(\mathrm{GeV}/c^2)^{-1}$ is extracted from the higher mass side-band (see section 6.2)). To assess a systematic uncertainty we repeat the fit of composition varying the slope of the background distribution within $\pm 1\sigma$ range, and the largest difference between these two fits and the central one is quoted as systematic uncertainty (bkg p.d.f.).

In supporting to our background mass model we generated an alternative "pure" combinatorial background data sample. This is done by combining a real D_s^- decay of the $i^{\rm th}$ event with an independent pion, that is the pion associated to the B_s^0 candidate in the $(i+1)^{\rm th}$ event. The resulting invariant mass is reported in figure 32 superimposed to the invariant mass distribution of real decays. It confirms that our background mass model is reasonable. The two backgrounds samples seem very similar in the higher mass region, however the pure combinatorial background sample has a lower slope $a_{\rm pure\ comb.} = (0.27 \pm 0.05)\,(\mathrm{GeV}/c^2)^{-1}$, with respect to the slope extracted from real data. Since this artificial background is very realistic, we repeat the fit of composition using $a_{\rm pure\ comb.}$ as the slope of the exponential of the combinatorial p.d.f., and the difference with the central fit is taken as an additional systematic uncertainty (pure comb. bkg).

7.4 dE/dx related systematic

Section 6.3 summarize how the fit of composition exploit the PID information in separating the different signal modes and backgrounds. The model used to introduce this information in the Likelihood is sophisticated, need a large number of parameters (see ref. [33]), extracted using high statistics and very pure samples of charged pions and kaons (from the decay $D^0 \to K^-\pi^+$), and protons and antiprotons (from $\Lambda \to p\pi^-$).

The systematic uncertainty related to the statistical uncertainty on the determination of PID probability density functions is assessed following a standard CDF procedure (see ref. [4]), by repeating the fit of composition in which all PID parameters are randomly varied in a 1σ-radius multidimensional sphere, keeping into account all the statistical correlations among parameters. In order to statistically sample a sufficient number of directions in this large dimensions space, we repeat the analysis for various (500) seed values. For each seed value the PID functions change in a different way and we can obtain a measurement of the effect of systematic uncertainties on the analysis results. The systematic uncertainty on the physics observables associated to the statistical uncertainty of the templates parameterization is given by the 3×r.m.s. of the distribution of the observables returned from the fits of composition performed with different seeds (dE/dx).

In addition to the systematic uncertainty due to the limited knowledge of PID templates, we have also to account for the approximated procedure used to marginalize the momentum dependence, as discussed in section 5. The associated systematic uncertainty is assessed by re-binning the momentum distribution by a factor 2 and by a factor 1/2 and extracting alternative PID templates (binning dE/dx mom.). We repeat the fit of composition in the two cases and the largest difference between the values obtained and central value is taken as systematic uncertainty.

7.5 Uncertainty related to the efficiency correction (MC stat.)

The relative efficiency ratio (see sec. 6.6) used to convert the ratio of event yields in ratio of branching fractions, is determined with $\mathcal{O}(0.6\%)$ statistical uncertainty (see eq. (19)). The ratio of branching fractions is reevaluated by using acceptance correction shifted by one standard deviation in either direction. The largest difference between the resulting ratios of branching fractions and the central value is taken as systematic uncertainty.

7.6 Total systematic uncertainties

All systematic uncertainties are summarized in the table 10. The total systematic uncertainty on the measurement is determined as the sum in quadrature of the individual systematic uncertainties.

8 Final results

Using the raw fit results and the efficiency correction from section 6 and the systematic uncertainty from section 7 we obtain the measurement of branching

Table 10: Summary of the systematic uncertainties for the observable measured in this work.

source	$\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)}$
nominal masses	0.0002
mass resolution	0.0021
bkg p.d.f.	0.0002
pure comb. bkg	0.0009
dE/dx	0.0005
binning *dE/dx* mom.	0.0009
MC stat.	0.0005
Total	**0.0026**

fraction of the $B_s^0 \to D_s^\pm K^\mp$ decay mode relative to the $B_s^0 \to D_s^- \pi^+$ decay mode at CDF with $9.3\,\text{fb}^{-1}$ of data. From the observed yields of $N(B_s^0 \to D_s^\pm K^\mp) = 335 \pm 40$ and $N(B_s^0 \to D_s^- \pi^+) = 4498 \pm 138$ we measure the following ratio of branching fractions:

$$\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = 0.0777 \pm 0.0079\,(\text{stat}) \pm 0.0026\,(\text{sys}), \tag{24}$$

where the first uncertainty is statistical and second one is systematic. This result is compatible with the other existing measurements:

$$\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = \begin{cases} 0.097 \pm 0.018\,(\text{stat}) \pm 0.009\,(\text{sys}) & \text{CDF (2008) [3]}, \\ 0.065^{+0.035}_{-0.029}\,(\text{stat}) & \text{Belle (2008) [32]}, \\ 0.0646 \pm 0.0043\,(\text{stat}) \pm 0.0025\,(\text{sys}) & \text{LHCb (2012) [2]}. \end{cases}$$

The final statistical uncertainty of our measurement is better than by a factor 2.3 with respect to the previous CDF result published in 2008, it is much better than Belle result, and it is worse than by a factor 1.8 with respect to the very recent LHCb measurement which is today the world's best result[2]. On the other hand the systematic uncertainty is at the same level of LHCb measurement and it is better than previous CDF measurement by a factor 3.5.

Our result is in agreement with a very recent theoretical expectation from ref. [23]:

$$\left.\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)}\right|_{\text{theory},\, SU(3)} = 0.0864^{+0.0087}_{-0.0072}, \tag{25}$$

where the $SU(3)$ flavor symmetry was assumed. Without the $SU(3)$ assumption, the estimated lower bound on the ratio of branching fractions is [23]

$$\left.\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)}\right|_{\text{theory}} \geq 0.080 \pm 0.007, \tag{26}$$

[2]LHCb measurement is performed using an inclusive data sample of reconstructed $D_s^- \to K^+K^-\pi^-$ decays, including also $D_s^- \to \phi\pi^-$ decays, which is the only decay mode used in our measurement. Therefore a more appropriate comparison between current LHCb and CDF performances would require a measurement with the same set of D_s^- decay modes.

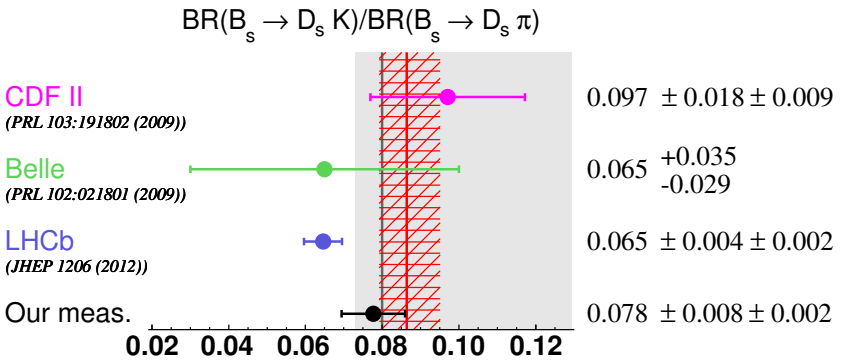

Figure 33: Current knowledge of the ratio of branching fractions $\mathcal{B}(B_s^0 \to D_s^{\pm}K^{\mp})/\mathcal{B}(B_s^0 \to D_s^-\pi^+)$, including this thesis measurement. The hatched region is the uncertainty of theoretical predicted value (central line in the hatched region) assuming the $SU(3)$ flavor symmetry. Filled region represents the permitted region theoretical predicted.

and our result is on the lower bound of the allowed region. A smaller value of the branching fractions ratio would imply a not real value for the hadronic parameter $x_s \propto R_b\, A_f/\bar{A}_f$ which quantifies the strength of the interference effects between the $B_s^0 \to D_s^+K^-$ and $\bar{B}_s^0 \to D_s^+K^-$ decay processes induced through the $B_s^0 - \bar{B}_s^0$ mixing; the parameter $R_b \propto |V_{ub}/(\lambda V_{cb})|$ measures one side of the unitary triangle. Figure 33 summarize experimental values obtained for the ratio of branching fractions (including our result). Also it reports the allowed region and the theoretical expectation.

8.1 Absolute branching fraction

Using the world average value $\mathcal{B}(B_s^0 \to D_s^-\pi^+) = (3.2 \pm 0.4) \times 10^{-3}$ from PDG 2012 [19] we can extracted the following absolute branching fraction:

$$\mathcal{B}(B_s^0 \to D_s^{\pm}K^{\mp}) = (2.49 \pm 0.25\,(\text{stat}) \pm 0.08\,(\text{syst}) \pm 0.31\,(\text{br})) \times 10^{-4}, \qquad (27)$$

where the last uncertainty is dominated by the uncertainty on the value of $\mathcal{B}(B_s^0 \to D_s^-\pi^+)$. This agrees with the world average $\mathcal{B}(B_s^0 \to D_s^{\pm}K^{\mp}) = (2.9 \pm 0.6) \times 10^{-4}$ reported in the PDG 2012 [19]. However LHCb, very recently, measured $\mathcal{B}(B_s^0 \to D_s^-\pi^+) = (2.95 \pm 0.05 \pm 0.17^{+0.18}_{-0.22}) \times 10^{-3}$ [2] (the third uncertainty is the uncertainty from the f_s/f_d measurement), which is more precise than PDG 2012 [19]. Thus if we use the LHCb measurement as input, instead of PDG 2012, we obtain the following absolute branching fraction

$$\mathcal{B}(B_s^0 \to D_s^{\pm}K^{\mp}) = (2.29 \pm 0.23\,(\text{stat}) \pm 0.08\,(\text{syst}) \pm 0.21\,(\text{br})) \times 10^{-4}, \qquad (28)$$

which agrees with $\mathcal{B}(B_s^0 \to D_s^{\pm}K^{\mp}) = (1.90 \pm 0.12 \pm 0.13^{+0.12}_{-0.14}) \times 10^{-4}$ from LHCb [2].

References

[1] CDF Internal Note. Relative transverse momentum distributions of bottom hadrons in 1.96 TeV $p\bar{p}$ collision. 9338, 2008.

[2] Aaij, R. et al. Measurements of the branching fractions of the decays $B_s^0 \to D_s^{\mp} K^{\pm}$ and $B_s^0 \to D_s^- \pi^+$. *JHEP*, 1206:115, 2012.

[3] Aaltonen, T. et al. First observation of $\bar{B}_s^0 \to D_s^{\pm} K^{\mp}$ and measurement of the ratio of branching fractions $\mathcal{B}(\bar{B}_s^0 \to D_s^{\pm} K^{\mp})/\mathcal{B}(\bar{B}_s^0 \to D_s^+ \pi^-)$. *Phys. Rev. Lett.*, 103:191802, Nov 2009.

[4] Aaltonen, T. et al. Measurements of branching fraction ratios and cp-asymmetries in suppressed $B^- \to D(\to K^+\pi^-)K^-$ and $B^- \to D(\to K^+\pi^-)\pi^-$ decays. *Phys. Rev. D*, 84:091504, Nov 2011.

[5] Aaltonen, T. et al. Evidence for the charmless annihilation decay mode $B_s^0 \to \pi^+\pi^-$. *Phys. Rev. Lett.*, 108:211803, May 2012.

[6] Aaltonen, T. et al. Precise measurement of the W-boson mass with the CDF II detector. *Phys.Rev.Lett.*, 108:151803, 2012.

[7] Abe, F. et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys.Rev.Lett.*, 74:2626–2631, 1995.

[8] Abulencia, A. et al. Measurement of the $B_s^0 - \bar{B}_s^0$ Oscillation Frequency. *Phys.Rev.Lett.*, 97:062003, 2006.

[9] Abulencia, A. et al. Observation of $B_s^0 - \bar{B}_s^0$ Oscillations. *Phys.Rev.Lett.*, 97:242003, 2006.

[10] Abulencia, A. et al. Observation of $B_s^0 \to K^+K^-$ and measurements of branching fractions of charmless two-body decays of B^0 and B_s^0 mesons in $\bar{p}p$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys. Rev. Lett.*, 97:211802, Nov 2006.

[11] Acosta, D. et al. The performance of the CDF luminosity monitor. *Nucl. Instrum. Meth.*, A494:57, 2002.

[12] Affolder, A. et al. Status report of the intermediate silicon layers detector at CDFII. *Nucl. Instrum. Meth.*, A485:6–9, 2002.

[13] Affolder, T. et al. CDF Central Outer Tracker. *Nucl. Instrum. Meth.*, A526:249, 2004.

[14] Albrecht, H. et al. Search for hadronic $b \to u$ decays. *Phys.Lett.*, B241:278–282, 1990.

[15] Ashmanskas, B. et al. The CDF Silicon Vertex Trigger. *Nucl. Instrum. Meth.*, A518:532, 2004.

[16] Atwood, D. et al. Enhanced *CP* violation with $B \to KD^0(\bar{D}^0)$ modes and extraction of the Cabibbo-Kobayashi-Maskawa angle γ. *Phys. Rev. Lett.*, 78:3257–3260, Apr 1997.

[17] Balka, L. et al. The CDF central electromagnetic calorimeter. *Nucl. Instrum. Meth.*, A267:272, 1988.

[18] Belforte, S. et al. Silicon Vertex Trigger Technical Design Report. *CDF Internal Note 3108*, (unpublished), 1995.

[19] Beringer, J. et al. Particle Data Group. *Phys. Rev.*, D86:010001, 2012.

[20] Blair, R. et al. The CDF-II detector: Technical design report. *FERMILAB-Pub*, 96/390-E, 1996.

[21] Bona, M. et al. UTfit Collaboration. *JHEP*, 0507:28, 2005, update results and plots available at http://www.utfit.org/UTfit.

[22] Charles, J. et al. CKMfitter Group. *Eur. Phys. J.*, C41:1–131, 2005, update results and plots available at http://ckmfitter.in2p3.fr.

[23] De Bruyn, K. et al. Exploring $B_s \to D_s^{(*)\pm} K^{\mp}$ Decays in the Presence of a Sizable Width Difference $\Delta\Gamma_s$. *ArXiv Physics e-prints*, [physics/1208.6463], 2012.

[24] Dell'Orso, M. The CDF Silicon Vertex Trigger. *Nucl. Phys. B*, 156:139, 2006.

[25] Fleischer, R. New strategies to obtain insights into CP violation through $B_s \to D_s^{\pm} K^{\mp}$, $D_s^{*\pm} K^{\mp}$, ... and $B_d \to D^{\pm}\pi^{\mp}, D^{*\pm}\pi^{\mp}$, ... decays. *Nucl.Phys.*, B671:459–482, 2003.

[26] Fleischer, R. et al. In Pursuit of New Physics with $B_s^0 \to K^+K^-$. *Eur. Phys. J.*, C71:1532, 2011.

[27] Gronau, M. et al. How to determine all the angles of the unitarity triangle from $B_d^0 \to DK_s$ and $B_s^0 \to D\phi$. *Phys. Lett. B*, 253(3–4):483 – 488, 1991.

[28] Gronau, M. et al. On determining a weak phase from charged B decay asymmetries. *Phys. Lett. B*, 265(1–2):172 – 176, 1991.

[29] Hill, C. S. Initial experience with the CDF layer ØØ silicon detector. *Nucl. Instrum. Meth.*, A511:118–120, 2003.

[30] Hoecker, A. et al. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007.

[31] James F. et al. 'Minuit' A system for function minimization and analysis of the parameter errors and correlations. *Comput. Phys. Commun.*, 10:343, 1975.

[32] Louvot, R. et al. Measurement of the Decay $B_s^0 \to D_s^- \pi^+$ and Evidence for $B_s^0 \to D_s^{\pm} K^{\pm}$ in e^+e^- Annihilation at $\sqrt{s} = 10.87$ 6GeV. *Phys.Rev.Lett.*, 102:021801, 2009.

[33] Morello, M. J. et al. dE/dx for pions, kaons and protons for 9/fb analyses. *CDF Internal Note 10756*, (unpublished), 2012.

[34] Punzi, G. Comments on Likelihood fits with variable resolution. *ArXiv Physics e-prints*, [physics/0401045], 2004.

[35] Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1958.

[36] Sill, A. et al. CDF Run II silicon tracking projects. *Nucl. Instrum. Meth.*, A447:1, 2000.

[37] Sphicas, S. A $b\bar{b}$ Monte Carlo Generator. *CDF Internal Note 2655*, (unpublished), 1994.

[38] Yosef, N. CP violation in meson decays. *ArXiv Physics e-prints*, [physics/0510413v1], 2005.