Università degli Studi di Pisa

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea in Fisica

Tesi di laurea magistrale

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

Candidato
Pietro Marino

Relatore
Prof. Giovanni Punzi

Correlatore
Dott. Michael J. Morello

Anno Accademico 2011–2012

Contents

List of Figures

List of Tables

Introduction

The CP violation has played a central role in particle physics since 1964, year of its discovery in the kaon system. The large asymmetries predicted, and later observed, in the B meson system have confirmed the Cabibbo–Kobayashi–Maskawa (CKM) picture, but opened further questions and stimulated an increase of precision of the measurements in the search for new physics sources of CP violation. While the CKM scheme can accommodate a certain amount of CP violation, it does not require it, nor it provides any fundamental motivation for its existence. In fact, the "cosmological question" of the dominance of matter over antimatter in our universe is a strong hint for the presence of CP-violating effects of a much larger strength than the standard CKM mechanism. Experimental measurements allow to determine the angles and the sides of the so-called "unitarity triangle", formed by CKM matrix parameters, using a variety of neutral and charged B decays, with and without mixing. Today a good level of precision has been reached for two of the angles, but the resolution on the third one (angle γ) is still rather poor, and limited by the size of the available data samples. Therefore a precise measurement of γ using "theoretically clean" decays, where it appears at tree-level (most notably the family of $B \to DK$ decay modes), is important not only as a measurement of a fundamental parameter of the theory, but also as a reference point for decays where the presence of significant loop contribution may exhibit additional CP-violating effects due to processes beyond the Standard Model.

One of the most important decay to probe such a field is the Cabibbo-suppressed mode $B_s^0 \to D_s^{\pm} K^{\mp}$, which can have a large CP-violating interference via $B_s^0 - \bar{B}_s^0$ mixing. This decay is still unexplored, because it cannot be studied in e^+e^- B-factories. In fact, it would be the first B_s^0 decay where CP violation is measured. Therefore isolating a clean sample of such decays is a crucial step toward a precise measurement of γ and this thesis deals this experimental challenge. Selecting a enriched sample of $B_s^0 \to D_s^{\pm} K^{\mp}$ decays, keeping high the signal efficiency and rejecting the huge amount background, which overwhelms the signal of interest by several order of magnitude, is crucial ingredient. It is also essential to disentangle $B_s^0 \to D_s^{\pm} K^{\mp}$ decays from kinematically similar decays ($B_s^0 \to D_s^- \pi^+, B^0 \to D_s^{(\star)+} \pi^-, B_s^0 \to D_s^{\star\pm} K^{\mp}, B^0 \to D_s^{(\star)-} K^+, B^0 \to D^- \pi^+$, etc.) which lay down in the same mass region of the signal.

In particular the thesis describes the analysis of the $B_s^0 \to D_s^{\pm} K^{\mp}$ decays collected by the CDF II experiment at the Tevatron collider, using a specialized trigger on displaced tracks. These modes have been measured for the first time by the CDF collaboration with $1.2\,\mathrm{fb}^{-1}$, we update the measurement of the ratio of branching fractions $\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp}) / \mathcal{B}(B_s^0 \to D_s^- \pi^+)$ to the full data sample corresponding

to about $10\,\mathrm{fb}^{-1}$ of integrated luminosity. This quantity, besides being an overall test of the analysis procedure, also provides useful information in the determination of γ itself.

To improve the signal to background ratio the selection of the $B_s^0 \to D_s^{\pm} K^{\mp}$ decays is optimized using an Artificial Neural Network, allowing an optimal use of the multidimensional information contained in the input variables. These are chosen with the aim to maximize the available information and to reduce their correlation. As far as the presence of similar and larger physics backgrounds, which are more favored than the signal decays, a careful determination of the backgrounds features and a precision calibration of the particle identification observables is done. For physics backgrounds we used the detailed CDF simulation, while for energy losses in the drift chamber we used real data. Thus all information, coming from kinematics and particle identification of the final state particles, is combined in an unbinned maximum likelihood fit to disentangle relative fractions of all contributions.

As a result, we measure the branching ratio of the Cabibbo-suppressed $B_s^0 \to D_s^{\pm} K^{\mp}$ decay. With respect to the first iteration of this analysis by CDF the resolution on the branching fraction is improved by a factor 1.4, because of analysis techniques developed in this thesis. Considering also the increasing of the statistics our final measurement is better than previous one by a factor 2.3.

1 Theory and motivations

In this chapter we focus our attention on the theoretical frame. In particular we discuss the connections between CP violation in the Standard Model and the $B_s^0 \to D_s^{\pm} K^{\mp}$ decay, which is a golden mode to extract the CKM angle γ, and consequently for searching New Physics.

1.1 CP violation in the Standard Model

The model of elementary particles and their interactions is defined by the symmetries of the Lagrangian and the representation of fermions and scalar with respect to the group of symmetry. The group of symmetry of the Standard Model (SM) is

$$G_{\rm SM} = SU(3)_{\rm C} \otimes SU(2)_{\rm L} \otimes U(1)_{\rm Y}, \tag{1.1}$$

where $SU(3)_{\rm C}$ is the group of symmetry of the strong interaction, while $SU(2)_{\rm L} \otimes U(1)_{\rm Y}$ is the electroweak interaction group of symmetry as introduced by the theory of Glashow-Weinberg-Salam. There are three fermion generations, each consisting of following five representations of $G_{\rm SM}$:

$$Q^I_{Li}(3,2)_{+1/6}, \quad u^I_{Ri}(3,1)_{+2/3}, \quad d^I_{Ri}(3,1)_{-1/3}, \quad L^I_{Li}(1,2)_{-1/2}, \quad \ell^I_{Ri}(1,1)_{-1}, \tag{1.2}$$

where the notation mean that, for example, left-handed quarks Q^I_{Li} are triplets of $SU(3)_{\rm C}$, doublets of $SU(2)_{\rm L}$ and carry hypercharge $Y = +1/6$. The super-index I denotes interaction eigenstates. The sub-index $i = 1, 2, 3$ is the flavor (or generation) index. There is a single scalar representation,

$$\phi(1,2)_{+1/2}, \tag{1.3}$$

which assumes a vacuum expectation value,

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix},$$

so that the gauge group is spontaneously broken,

$$G_{\rm SM} \to SU(3)_{\rm C} \otimes U(1)_{\rm Y}.$$

The Standard Model Lagrangian, $\mathcal{L}_{\rm SM}$, is the most general renormalizable Lagrangian that is consistent with the gauge symmetry of equation (1.1) and the particle content of eqs. (1.2) and (1.3). It can be divided into three parts:

$$\mathcal{L}_{\rm SM} = \mathcal{L}_{\rm kinetic} + \mathcal{L}_{\rm Higgs} + \mathcal{L}_{\rm Yukawa}.$$

As concerns the kinetic terms, to maintain gauge invariance, one has to replace the derivative with a covariant derivative:

$$D^\mu = \partial^\mu + ig_s G_a^\mu L_a + igW_b^\mu T_b + ig' B^\mu Y. \tag{1.4}$$

Here G_a^μ are the eight gluon fields, W_b^μ the three weak interaction bosons and B^μ the single hypercharge boson. The L_a's are $SU(3)_{\rm C}$ generators (the 3×3 Gell-Mann matrices $\lambda_a/2$ for triplets, 0 for singlets), the T_b's are $SU(2)_{\rm L}$ generators (the 2×2 Pauli matrices $\tau_b/2$ for doublets, 0 for singlets), and the Y's are the $U(1)_{\rm Y}$ charges. For example, for the left-handed quarks Q_L^I, we have

$$\mathcal{L}_{\rm kinetic}(Q_L^I) = i\overline{Q}_{Li}^I \gamma_\mu \left(\partial^\mu + \frac{i}{2} g_s G_a^\mu \lambda_a + \frac{i}{2} g W_b^\mu \tau_b + \frac{i}{6} g' B^\mu\right) Q_{Li}^I, \tag{1.5}$$

while for the left-handed leptons L_L^I, we have

$$\mathcal{L}_{\rm kinetic}(L_L^I) = i\overline{L}_{Li}^I \gamma_\mu \left(\partial^\mu + \frac{i}{2} g W_b^\mu \tau_b - \frac{i}{2} g' B^\mu\right) L_{Li}^I,$$

The $\mathcal{L}_{\rm kinetic}$ part is always CP conserving.

The Higgs potential, which describes the scalar self interactions, is given by:

$$\mathcal{L}_{\rm Higgs} = \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2.$$

This part of the Lagrangian is also CP conserving.

The quark Yukawa interactions are given by

$$-\mathcal{L}_{\rm Yukawa} = Y_{ij}^d \overline{Q}_{Li}^I \phi d_{Rj}^I + Y_{ij}^u \overline{Q}_{Li}^I \tilde{\phi} u_{Rj}^I + Y_{ij}^l \overline{L}_{Li}^I \phi \ell_{Rj}^I + \text{h.c.}, \tag{1.6}$$

where $\tilde{\phi} = i\tau_2 \phi^\dagger$. This part of the Lagrangian is, in general, CP violating as we are going to explain now.

After the breaking of symmetry, with the replacement $\mathcal{R}e(\phi^0) \to (v + H^0)/\sqrt{2}$, the Yukawa interactions (equation (1.6)) give rise to mass terms:

$$-\mathcal{L}_M = (M_d)_{ij} \overline{d}_{Li}^I d_{Rj}^I + (M_u)_{ij} \overline{u}_{Li}^I u_{Rj}^I + (M_\ell)_{ij} \overline{\ell}_{Li}^I \ell_{Rj}^I + \text{h.c.},$$

where $M_{q,\ell} = vY^{q,\ell}/\sqrt{2}$ $(q = u, d)$ and the $SU(2)_{\rm L}$ doublets was decomposed into their components:

$$Q_{Li}^I = \begin{pmatrix} u_{Li}^I \\ \\ d_{Li}^I \end{pmatrix} \qquad L_{Li}^I = \begin{pmatrix} \nu_{Li}^I \\ \\ \ell_{Li}^I \end{pmatrix}.$$

The mass basis corresponds, by definition, to diagonal mass matrices. We can always find unitary matrices V_{qL} and V_{qR} such that

$$V_{qL}\, M_q\, V_{qR}^\dagger = M_q^{\rm diag} \quad (q = u, d),$$

with $M_q^{\rm diag}$ diagonal and real. The quark mass eigenstates are identified as

$$q_{Li} = (V_{qL})_{ij}\, q_{Lj}^I, \qquad q_{Ri} = (V_{qR})_{ij}\, q_{Rj}^I \quad (q = u, d). \tag{1.7}$$

The charged current interactions for quarks (that is the interactions of the charged $SU(2)_{\rm L}$ gauge bosons $W^{\pm}_{\mu} = (W^1_{\mu} \mp iW^2_{\mu})/\sqrt{2}$), which in interaction basis are described by eq. (1.5), have a complicated form in the mass basis:

$$-\mathcal{L}^q_{W^{\pm}} = \frac{g}{\sqrt{2}} \overline{u}_{Li} \gamma^{\mu} \left(V_{uL} V^{\dagger}_{dL}\right)_{ij} d_{Lj} W^{+}_{\mu} + \text{h.c.}.$$

The unitary 3×3 matrix,

$$V_{\rm CKM} = V_{uL} V^{\dagger}_{dL} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{1.8}$$

is the Cabibbo-Kobayashi-Maskawa (CKM) mixing matrix for quarks. A $n \times n$ unitary matrix contains n^2 independent real parameters, $2n-1$ of these can be eliminated redefining the phases of the n up-quarks and n down-quarks (an overall phase is irrelevant, if all quark phases are changed in the same way $V_{\rm CKM}$ is unaffected); hence there are $(n-1)^2$ parameters left. A unitary matrix is also orthogonal, and as such it contains $n(n-1)/2$ parameters corresponding to the independent rotation angles between the n basis vectors; thus the remaining $(n-1)(n-2)/2$ parameters must be the complex phases. In the case of $n = 3$, we have four physical parameters: three rotation angles (corresponding to the Euler angles) and one complex phase. The latter is the CKM phase, which is the single source of CP violation in the quark sector of the Standard Model. As result of the fact that $V_{\rm CKM}$ is not diagonal, the $W^{\pm}$ gauge bosons couple to quark (mass eigenstates) of different generations. Within the Standard Model, this is the only source of *flavor changing* interactions.

The fact that there are only three real and one imaginary physical parameters in $V_{\rm CKM}$, can be made manifest by choosing an explicit parametrization. For example, the standard parametrization, used by the Particle Data Group (PDG) [18], is given by

$$V_{\rm CKM} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix},$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$. The three $\sin\theta_{ij}$ are the three real mixing parameters while δ is the CKM phase. Another parametrization is the Wolfenstein parametrization, where the four mixing parameters are (λ, A, ρ, η), with $\lambda = |V_{us}| \approx 0.22$ (the Cabibbo angle) playing the role of an expansion parameter and η representing the CP violating phase. Up to $\mathcal{O}(\lambda^3)$ the parametrization is given by

$$V_{\rm CKM} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3[1 - (\rho + i\eta)] & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4).$$

The CKM matrix is very close to a unit matrix with off diagonal terms that are small. The order of magnitude of each element can be read from the power of λ in the Wolfenstein parametrization.

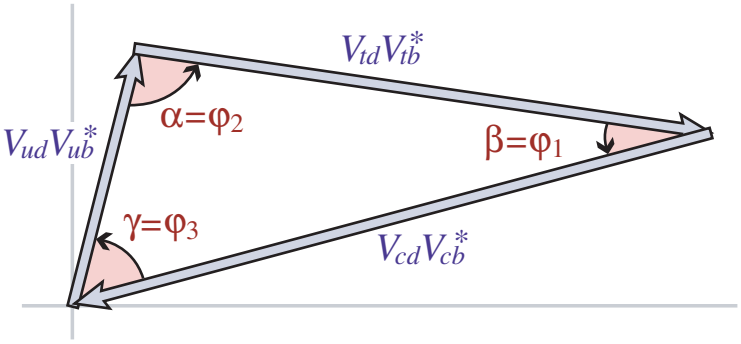


Figure 1.1 Graphical representation of the unitarity constraint $V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0$ as a triangle in the complex plane.

The current knowledge of the CKM matrix elements moduli, as obtained from ref. [21], is the following

$$|V_{\mathrm{CKM}}| = \begin{pmatrix} 0.97425^{+0.00022}_{-0.00014} & 0.22543^{+0.00059}_{-0.00095} & 0.00355^{+0.00015}_{-0.00012} \\ 0.22529^{+0.00060}_{-0.00094} & 0.97342^{+0.00022}_{-0.00015} & 0.04126^{+0.00060}_{-0.00104} \\ 0.00857^{+0.00033}_{-0.00030} & 0.04051^{+0.00060}_{-0.00104} & 0.999142^{+0.000043}_{-0.000025} \end{pmatrix} \tag{1.9}$$

1.2 The unitarity triangles

The unitarity of the CKM matrix ($V_{\mathrm{CKM}}V_{\mathrm{CKM}}^\dagger = \mathbf{1}$) leads to various relations among the matrix elements, i. e.

$$\sum_{k\in\{u,c,t\}} V_{ki}V_{kj}^* = \delta_{ij} \qquad (i,j \in \{d,s,b\}).$$

There are six of these relations that require the sum of three complex quantities to vanish. Therefore, they can be geometrically represented in the complex plane as a triangle and they are called *unitarity triangles*. One of these triangles has sides roughly the same length and it corresponds to the relation

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0; \tag{1.10}$$

for these reasons, the term "the unitarity triangle" is reserved for the equation (1.10). The unitarity triangle is depicted in the figure 1.1. We further define the rescaled unitarity triangle. It is derived from (1.10) by choosing a phase convention such that $(V_{cd}V_{cb}^*)$ is real and dividing the lengths of all sides by $|V_{cd}V_{cb}^*|$. The rescaled unitarity triangle is similar to the unitarity triangle. Two vertices of the rescaled unitarity triangle are fixed at $(0,0)$ and $(1,0)$. The coordinates of the remaining vertex correspond to the Wolfenstein parameters (ρ,η).

The three angles of the unitarity triangle are defined as follows:

$$\alpha \equiv \arg\left[-\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right], \quad \beta \equiv \arg\left[-\frac{V_{cd}V_{cb}^*}{V_{td}V_{tb}^*}\right], \quad \gamma \equiv \arg\left[-\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right]. \tag{1.11}$$

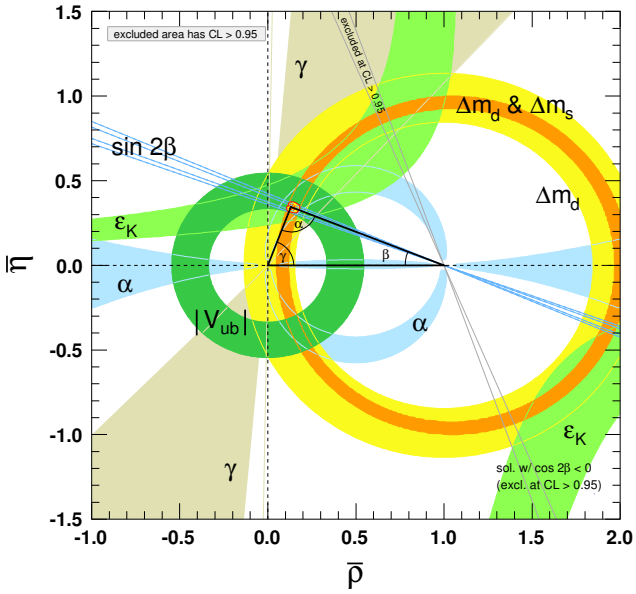


Figure 1.2 Current knowledge of the unitarity triangle parameters from the CKM-Fitter global fit [21].

All the measurements sensitive to the unitarity triangle are regularly collected and compiled into a global fit [20, 21]. The current knowledge of the unitarity triangle is reported in the figure 1.2, from [21]. From the figure, we can see that the γ angle is determined with the worse precision than all the other sides and angles.

1.3 B_s mixing

The phenomenology of CP violation in neutral flavored meson decays is enriched by the possibility to have $B_s^0 \leftrightarrow \bar{B}_s^0$ transitions, also known as flavor mixing or oscillations. Here we describe the $B_s^0 - \bar{B}_s^0$ system and the phenomenology of mixing. The phenomenology of the flavor mixing is superficially different in K, D, B and B_s decays. However, as a matter of practical convenience, we will refer specially to the B_s. Our treatment will follow very closely the one presented in ref. [37].

We define decay amplitudes of a B_s meson and its CP conjugate $\bar{B}_s$ to a multi-particle final state f and its CP conjugate $\bar{f}$ as

$$A_f = \langle f | \mathcal{H} | B_s \rangle , \quad \bar{A}_f = \langle f | \mathcal{H} | \bar{B}_s \rangle , \quad A_{\bar{f}} = \langle \bar{f} | \mathcal{H} | B_s \rangle , \quad \bar{A}_{\bar{f}} = \langle \bar{f} | \mathcal{H} | \bar{B}_s \rangle ,$$

where $\mathcal{H}$ is the decay Hamiltonian. The particle-antiparticle mixing phenomenon causes an initial (at time $t = 0$), pure B_s^0 meson state to evolve in time to a linear combination of B_s^0 and $\bar{B}_s^0$ states. If the times t, in which we are interested, are much larger than the typical strong interaction scale, then the time evolution can be described by an effective Hamiltonian and the time evolution equation can be described as follow

$$i \frac{d}{dt} \begin{pmatrix} B_s^0(t) \\ \bar{B}_s^0(t) \end{pmatrix} = \left[\boldsymbol{M} - \frac{i}{2} \boldsymbol{\Gamma} \right] \begin{pmatrix} B_s^0(t) \\ \bar{B}_s^0(t) \end{pmatrix} ,$$

where $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ are 2×2 Hermitian matrices,

$$\boldsymbol{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{12}^* & M_{22} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Gamma} = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma_{22} \end{pmatrix}.$$

Diagonal elements of the $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ are associated with the flavor-conserving transitions, $B_s^0 \to B_s^0$ and $\bar{B}_s^0 \to \bar{B}_s^0$, while off-diagonal elements are associated with flavor-changing transitions $B_s^0 \leftrightarrow \bar{B}_s^0$. Since $\mathcal{H}_{\text{eff}}$ is not diagonal, B_s^0 and $\bar{B}_s^0$ are not mass eigenstates, and thus do not have well defined masses and widths. We denote the light and heavy eigenstates as B_s^L and B_s^H respectively, with $m_H > m_L$. The matrix element of $\boldsymbol{M}$ and $\boldsymbol{\Gamma}$ must satisfy $M_{11} = M_{22}$ and $\Gamma_{11} = \Gamma_{22}$ in order to be consistent with the CPT invariance.

The eigenstates of the effective Hamiltonian are

$$|B_s^{L,H}\rangle = p\,|B_s^0\rangle \pm q\,|\bar{B}_s^0\rangle\,,$$

with the p and q complex coefficients, satisfying $|p|^2 + |q|^2 = 1$, and

$$\frac{q}{p} = \sqrt{\frac{M_{12}^* - i\Gamma_{12}^*/2}{M_{12} - i\Gamma_{12}/2}}.$$

The real and imaginary parts of the eigenvalues of $\mathcal{H}_{\text{eff}}$ corresponding to $|B_s^{L,H}\rangle$ represent their masses and decay-widths, respectively.

The time evolution of a state $|B_s^0(t)\rangle$ ($|\bar{B}_s^0(t)\rangle$), which is a pure $|B_s^0\rangle$ ($|\bar{B}_s^0\rangle$) state at the time $t = 0$, is the given by

$$\begin{aligned} |B_s^0(t)\rangle &= g_+(t)\,|B_s^0\rangle + \frac{q}{p} g_-(t)\,|\bar{B}_s^0\rangle\,, \\ |\bar{B}_s^0(t)\rangle &= \frac{p}{q} g_-(t)\,|B_s^0\rangle + g_+(t)\,|\bar{B}_s^0\rangle\,, \end{aligned}$$

where

$$g_\pm(t) \equiv \frac{1}{2}\Big(\exp\Big(-im_H t - \frac{\Gamma_H}{2} t\Big) \pm \exp\Big(-im_L t - \frac{\Gamma_L}{2} t\Big)\Big).$$

Defining the dimensionless mixing parameter

$$x_s = \frac{\Delta m_s}{\Gamma_s}, \qquad y_s = \frac{\Delta \Gamma_s}{2\Gamma_s},$$

with $\Gamma_s \equiv (\Gamma_L + \Gamma_H)/2 = 1/\tau_s$, the time-dependent decay rate for an initial pure B_s^0 state is then

$$\begin{aligned} \frac{d\Gamma}{dt}(B_s^0(t) \to f) \propto |A_f|^2 \Big[&(1 - |\lambda_f|^2)\cos(x_s t/\tau_s) + (1 + |\lambda_f|^2)\cosh(y_s t/\tau_s) \\ &- 2\Im m(\lambda_f)\sin(x_s t/\tau_s) + 2\Re e(\lambda_f)\sinh(y_s t/\tau_s)\Big], \end{aligned} \tag{1.12}$$

with

$$\lambda_f = \frac{q}{p}\frac{\bar{A}_f}{A_f},$$

while the time-dependent decay rates for an initial pure $\bar{B}_s^0$ state, is similar:

$$\frac{d\Gamma}{dt}(\bar{B}_s^0(t) \to f) \propto |\bar{A}_f|^2 \Big[(1 - |\lambda_f^{-1}|^2) \cos(x_s t/\tau_s) + (1 + |\lambda_f^{-1}|^2) \cosh(y_s t/\tau_s) \\ - 2\Im m(\lambda_f^{-1}) \sin(x_s t/\tau_s) + 2\Re e(\lambda_f^{-1}) \sinh(y_s t/\tau_s) \Big]. \tag{1.13}$$

Decay rates to the CP-conjugate final state $\bar{f}$ are obtained analogously, with the substitutions $A_f \to A_{\bar{f}}$ and $\bar{A}_f \to \bar{A}_{\bar{f}}$ in eqs. (1.12)-(1.13). Terms proportional to $|A_f|^2$ or $|\bar{A}_f|^2$ are associated with decays that occur without any net $B_s^0 \leftrightarrow \bar{B}_s^0$ oscillation, while terms proportional to $|(q/p)\bar{A}_f|^2$ or $|(p/q)A_f|^2$ are associated with decays following a net oscillation. The $\sinh(y_s t/\tau_s)$ and $\sin(x_s t/\tau_s)$ terms of eqs. (1.12)-(1.13) are associated with the interference between these two cases.

The $B_s^0 - \bar{B}_s^0$ oscillation was observed by CDF II [8], and the value of Δm_s and $\Delta\Gamma_s$ were measured [7]. The current values of these quantities as reported in ref. [18] are:

$$\Delta m_s = (17.69 \pm 0.08) \times 10^{12}\, \hbar\, \mathrm{s}^{-1},$$
$$\Delta\Gamma_s = (0.100 \pm 0.013) \times 10^{12}\, \mathrm{s}^{-1}.$$

1.4 γ measurement from $B_s^0 \to D_s^{\pm} K^{\mp}$ decays

A good knowledge of the γ angle allows both testing the Standard Model and probing New Physics (NP) scenarios. While α and β have been determined to a good level of precision, the resolution on γ is still rather poor. The method to extract γ from the $B_s^0 \to D_s^{\pm} K^{\mp}$ decay modes uses interference from B_s mixing: a B_s^0 can reach the final state $D_s^+ K^-$ through the decay $B_s^0 \to D_s^+ K^-$, or it can become a $\bar{B}_s^0$, by the $B_s^0 - \bar{B}_s^0$ mixing, and then decay through $\bar{B}_s^0 \to D_s^+ K^-$ channel. The interfering diagrams lead to an observable effect related to γ through the time-dependent evolution of the following decay rates:

$$\begin{aligned}
\frac{d\Gamma}{dt}(B_s^0 \to D_s^- K^+) =& \frac{|A_f|^2 e^{-\Gamma_s t}}{2}(1 + |\lambda_f|^2)\Big[\cosh(y_s t/\tau_s) + C_f \cos(x_s t/\tau_s) \\
&+ D_f \sinh(y_s t/\tau_s) - S_f \sin(x_s t/\tau_s) \Big], \\
\frac{d\Gamma}{dt}(\bar{B}_s^0 \to D_s^- K^+) =& \left|\frac{p}{q}\right|^2 \frac{|A_f|^2 e^{-\Gamma_s t}}{2}(1 + |\lambda_f|^2)\Big[\cosh(y_s t/\tau_s) - C_f \cos(x_s t/\tau_s) \\
&+ D_f \sinh(y_s t/\tau_s) + S_f \sin(x_s t/\tau_s) \Big], \\
\frac{d\Gamma}{dt}(\bar{B}_s^0 \to D_s^+ K^-) =& \frac{|A_{\bar{f}}|^2 e^{-\Gamma_s t}}{2}(1 + |\lambda_f|^2)\Big[\cosh(y_s t/\tau_s) + C_f \cos(x_s t/\tau_s) \\
&+ D_{\bar{f}} \sinh(y_s t/\tau_s) - S_{\bar{f}} \sin(x_s t/\tau_s) \Big], \\
\frac{d\Gamma}{dt}(B_s^0 \to D_s^+ K^-) =& \left|\frac{q}{p}\right|^2 \frac{|A_{\bar{f}}|^2 e^{-\Gamma_s t}}{2}(1 + |\lambda_f|^2)\Big[\cosh(y_s t/\tau_s) - C_f \cos(x_s t/\tau_s) \\
&+ D_{\bar{f}} \sinh(y_s t/\tau_s) + S_{\bar{f}} \sin(x_s t/\tau_s) \Big],
\end{aligned} \tag{1.14}$$

where we abbreviated f with $D_s^- K^+$ and $\bar{f}$ with $D_s^+ K^-$, and

$$\begin{aligned} C_f &= \frac{1-|\lambda_f|^2}{1+|\lambda_f|^2}, \\ D_{f(\bar{f})} &= \frac{2|\lambda_f|}{1+|\lambda_f|^2}\cos(\delta \mp (\gamma - 2\beta_s)), \\ S_{f(\bar{f})} &= \frac{2|\lambda_f|}{1+|\lambda_f|^2}\sin(\delta \mp (\gamma - 2\beta_s)). \end{aligned} \tag{1.15}$$

γ and δ are the weak and strong phases[1], respectively, and β_s is the mixing phase on the $B_s^0 - \bar{B}_s^0$ system. The five observables in equation (1.15) are related to the three physics parameters $\lambda_f, \delta, \gamma - 2\beta_s$. Therefore with a simultaneous fit the relation in (1.14) can be extracted the value of $\gamma - 2\beta_s$, and using the value of β_s from other measurement[2] can be extracted the value of γ.

1.5 State of art

The current measurement of γ uses the interference of the $B^- \to D^0 K^- \to [f]K^-$ and $B^- \to \bar{D}^0 K^- \to [f]K^-$ decays, using the method called ADS [15] if $f = K^+\pi^-$, and called GLW [27, 26] if $f = \pi^+\pi^-$ of $f = K^+K^-$. As an example, figure 1.3 reports the combination of the current results, where the value estimated for γ is:

$$\begin{cases} \gamma = (67.1 \pm 4.3)^\circ & \text{CKMFitter [21]}, \\ \gamma = (68.5 \pm 3.1)^\circ & \text{UTFit [20]}. \end{cases}$$

However these measurements are not model independent. In fact they result from a global fit of the all available experimental information, in which several Standard Model assumptions are used, that, in principle, one may want to independently verify. On the other hand a model-independent measurement was done only very recently [25]. The decay $B_s^0 \to K^+K^-$ is related to the $B^0 \to \pi^+\pi^-$ channel through the U-spin flavor symmetry of strong interactions, and allows a determination of the angle γ:

$$\gamma = (68 \pm 7)^\circ \text{ [25]}.$$

However the $B_s^0 \to K^+K^-$ decay mode is governed by QCD penguin topologies and has a doubly Cabibbo-suppressed tree contribution in the Standard Model, therefore this measurement may include contributions from new particles.

Therefore a precise direct measurement of γ, free from penguin pollution, is needed to clarify our knowledge of the theory, and the $B_s^0 \to D_s^\pm K^\mp$ decay mode is a golden channel to achieve this result. Collecting a sizeable data sample of these decays is then the first and most important step toward such a measurement. Thus the main goal of this thesis is focused on how to select and disentangle the largest as possible data sample of $B_s^0 \to D_s^\pm K^\mp$ decays in an hadronic environment. This

[1] The origin of the strong phase is the possible contribution from intermediate on-shell states in the decay process, that is an absorptive part of an amplitudes that has contributions from coupled channels.

[2] The SM prediction for β_s is very small with respect to γ.

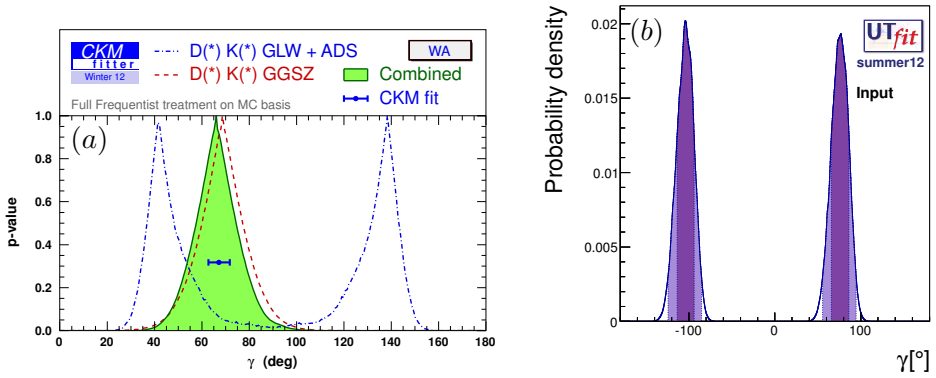


Figure 1.3 Combination of the current results on γ, (a) from the CKMFitter Collaboration [21], (b) form UTFit Collaboration [20].

is a real challenge, since they are overwhelmed from a huge amount of physics and combinatorial backgrounds.

The $B_s^0 \to D_s^{\pm} K^{\mp}$ was observed for the first time at CDF II in 2008, and later from Belle and LHCb. The LHCb measurement is the world's best and it was performed very recently this year. The current available measurements of its relative branching fraction are:

$$\frac{\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})}{\mathcal{B}(B_s^0 \to D_s^{-} \pi^{+})} = \begin{cases} 0.097 \pm 0.018\,(\text{stat}) \pm 0.009\,(\text{sys}) & \text{CDF (2008) [2]}, \\ 0.065^{+0.035}_{-0.029}\,(\text{stat}) & \text{Belle (2008) [31]}, \\ 0.0646 \pm 0.0043\,(\text{stat}) \pm 0.0025\,(\text{sys}) & \text{LHCb (2012) [1]}. \end{cases}$$

2 CDF II detector at Tevatron Collider

This chapter briefly describes the Tevatron collider accelerator and CDF II detector, focusing on the subsystems most important for the analysis presented, such as the trigger and tracking systems. For a detailed description refer to Ref. [19].

2.1 The Tevatron Collider

The Tevatron is a proton synchrotron of 1 km in radius located at Fermi National Accelerator Laboratory (a.k.a FNAL or Fermilab), about 50 km at West of Chicago (IL) in the US. The Tevatron was shut-down in 30 September 2011 after about 25 years of activity. In its period of activity the Tevatron accumulate about $10\,\text{fb}^{-1}$ of integrated luminosity. The physics results of this 25 years of research cover different kind of very important measurements, from the discovery of the quark top to the world's best measurement of W boson mass, through the first observation of the B_s^0 mixing [6, 5, 8].

In the Tevatron circulating "bunches"[1] of protons and antiprotons both at energy of 980 GeV. The available energy in the center-of-mass is $\sqrt{s} = 1.96\,\text{TeV}$.

The performance of a collider are evaluated by two key parameters: the energy of the center-of-mass, $\sqrt{s}$, and the luminosity, $\mathcal{L}$. The former defines accessible phase-space for the production of resonances in the final states. The latter is the proportional coefficient between the rate, $\mathcal{R}$, of the events of a given process and its cross section, σ:

$$\mathcal{R}\,[\text{events s}^{-1}] = \mathcal{L}\,[\text{cm}^{-2}\ \text{s}^{-1}] \cdot \sigma\,[\text{cm}^2].$$

The time-integral of luminosity (integrated luminosity) is therefore a measure of the expected number of events, n, produced in a time T:

$$n(T) = \int_0^T \mathcal{L}\sigma\, dt.$$

Assuming an ideal head-on $p\bar{p}$ collision the instantaneous luminosity is defined as

$$\mathcal{L} = 10^{-5}\frac{N_p N_{\bar{p}} B f \beta\gamma}{2\pi\beta^\star\sqrt{(\epsilon_p+\epsilon_{\bar{p}})_x(\epsilon_p+\epsilon_{\bar{p}})_y}}\mathcal{H}(\sigma_z/\beta^\star) \qquad [10^{-30}\,\text{cm}^{-2}\ \text{s}^{-1}]. \tag{2.1}$$

The luminosity depends on the following Tevatron parameters: the number of circulating bunches in the ring ($B = 36$), the revolution frequency ($f = 47.713\,\text{Hz}$),

[1] A bunch is a collection of particles with the same energy.

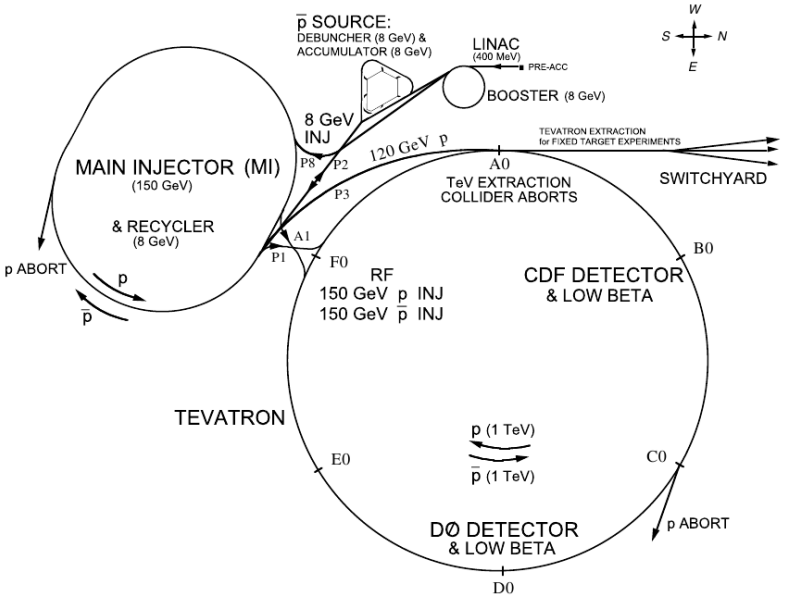


Figure 2.1 Illustration of the Fermilab Tevatron collider.

the Lorentz relativistic factor ($\beta\gamma = 1045.8$ at 980 GeV), the average number of protons ($N_p \approx 250 \times 10^9$) and antiprotons ($N_{\bar{p}} \approx 25 \times 10^9$) in a bunch, an empirical factor ($\mathcal{H} = 0.6 \div 0.7$), which is function of the ratio between the longitudinal r.m.s. width of the bunch ($\sigma_z \approx 60$ cm) and the "beta function" calculated in the interaction point ($\beta^\star \approx 31$ cm), and the 95% normalized emittances of the beam ($\epsilon_p \approx 18$ mm mrad and $\epsilon_{\bar{p}} \approx 13$ mm mrad after injection).[2]

At the Tevatron the dominant limiting factor of the luminosity is the availability of antiprotons because it is difficult to produce and compact them into bunches and to transfer them efficiently through the subsequent accelerator stages.

The Tevatron provides beams for experiments in different modes (fixed-target, collider, etc.). For the purpose of the present analysis, we will describe the procedure for obtaining a continuous period of collider operation using the same collection of protons and antiprotons, called *store*.

2.1.1 Proton beam

The process of production of the protons starts with the H^- ions produced by ionization of gaseous hydrogen. The H^- ions are boosted to 750 keV by a commercial Cockroft-Walton accelerator, then they are injected in linear accelerator (*Linac*, see fig. 2.1) which increases their energy to 400 MeV. A carbon foil is used to strip the electrons from the H^-. The resulting protons are injected to the *Booster* (see fig. 2.1) a synchrotron of 75.5 m in radius that accelerates protons up to 8 GeV and compacts

[2]The empirical factor $\mathcal{H}$ is a parametrization of the longitudinal profile of the beams in the collision region, which assumes the shape of an horizontal "hourglass" in the central region. The beta function is a parameter convenient for solving the equation of motion of a particle through an arbitrary beam transport system. The emittance ϵ measures the phase-space occupied by the particles of the beam. The quantity $\sqrt{\beta\epsilon}$ is proportional to the r.m.s. width of the beam in the corresponding phase plane.

them into bunches. The protons are then transferred to a synchrotron, called the *Main Injector*, which brings their energy to 150 GeV. The last stage of the process is the transfer to the Tevatron, a circular synchrotron with 5.7 T superconductive magnets, that accelerates the protons to their final energy of 980 GeV.

2.1.2 Antiproton beam

While the energy of the protons bunches circulating in the Main Injector reaches 120 GeV they are slammed to a 7 cm thick nickel or copper target. Spatially widespread antiprotons are produced and focused into a beam via a cylindrical lithium lens which separates $\bar{p}$ from other charged interaction products. The emerging antiprotons are stored in a *Debuncher* (see fig. 2.1). It is a storage ring where the momentum spread of the $\bar{p}$ is reduced while maintaining a constant energy of 8 GeV, via stochastic cooling stations. Many cycles of Debuncher cause the destruction of the bunch structure which results in a continuous beam of antiprotons. At the end of the process the monochromatic antiprotons are stored in the *Accumulator* where they are further cooled. When a current sufficient to create 36 bunches with the required density is available, the $\bar{p}$ are injected into the Main Injector where they are accelerates up to 150 GeV. Then they are transferred to the Tevatron.

2.1.3 The collision

When 36 bunches of both protons and antiprotons are circulating in the Tevatron, the energy of the machine is increased in about 10 seconds from 150 to 980 GeV and the collisions begin at the two interaction points: DØ (where the homonym detector is located) and BØ (home of CDF II). Special quadrupole magnets located at both extremities of the detectors along the beam pipe "squeeze" the beam to maximize luminosity inside the detectors. A roughly Gaussian distribution of the interaction region along the beam axis ($\sigma_z \approx 28\,\text{cm}$) and also in transverse plane ($\sigma_T \approx 30\,\mu\text{m}$) is achieved.

Next the injection take place a procedure called "scraping", which shapes the beam transverse profile in order to avoid detector damages due to the tails of the $p(\bar{p})$ distributions entering the active volumes. The scraping is done by moving iron plates which act as collimators in the transverse plane.

When the beam profile is narrow enough and the conditions are safely stable the detectors are powered and the data taking starts.

The inter-bunch crossing is 396 ns and this defines an overall time constant which influences the whole detector design: on this parameter depends the choice of the active parts, the design of the readout electronics, the structure of the trigger etc.. The number of overlapping interactions N for each bunch crossing is a Poisson-distributed variable dependent on instantaneous luminosity and on the number of colliding bunches. At Tevatron peak luminosities of $\mathcal{L} \approx 1 \times 10^{-32}\,\text{cm}^{-2}\,\text{s}^{-1}$ $\overline{N}$ is approximately 2.

2.2 CDF II Detector

The CDF II detector is a large multi-purpose solenoidal magnetic spectrometer surrounded by projective calorimeters and fine-grained muon detectors. It is installed at the BØ interaction point of the Tevatron (see fig. 2.1) to determine energy, momentum and, whenever possible, the identity of a broad range of particles produced in 1.96 TeV $p\bar{p}$ collisions. After its construction (1985) several upgrades modified original design. The most extensive upgrade (starts in 1995) led to the current detector whose operation is generally referred to as Run II.

CDF II is approximately a 5000-ton cylinder assembly of sub-detectors, $\approx 15\,\mathrm{m}$ in length, $\approx 15\,\mathrm{m}$ in diameter. The innermost part of the detector is the tracking system, consisting of three silicon subdetectors and a large outer drift-chamber, all containing in a superconducting solenoid, 1.5 m in radius and 4.8 m in length, which generated a 1.4 T magnetic field along the beam axis. The tracking system allows to measure the momentum and the track of the charged particles and allows to reconstruct both the primary vertex of the $p\bar{p}$ collision and the secondary decay vertex of the long-live particles. This system is surrounded by a time of flight detector which identified particles of low momentum. A system of projective tower calorimetry measure the energy of hadrons, photons and electrons, outside the solenoid. At last a system of proportional chamber and scintillator counter reveals the muons.

2.2.1 Coordinates and notation

CDF II employs a right-handed Cartesian coordinates system with the origin in the BØ interaction point, assumed coincident with the center of the drift chamber. The positive z-axis lies along the nominal beam-line pointing toward the proton direction (east). The (x, y) plan is therefore perpendicular to either beams, with positive y-axis pointing vertically upward and positive x-axis in the horizontal plane of the Tevatron, pointing radially outward with respect to the center of the ring.

Since the colliding beams of the Tevatron are unpolarized, the resulting physical observations are invariant under rotations around the beam line axis. Thus, a cylindrical (r, ϕ, z) coordinates system is particular convenient to describe the detector geometry. From now on, *longitudinal* means parallel to the proton beam direction (i.e., to the z-axis), and *transverse* means perpendicular to the protons direction, i.e., in the $(x, y) \equiv (r, \phi)$ plane.

Since protons and antiprotons are composite particles, the actual interaction occurs between the individuals partons (valence or sea quarks and gluons) contained within them. Each parton carries an unknown fraction of the (anti)proton momentum. As a consequence of the possible imbalance in the longitudinal components of the momenta of interacting partons, possible large velocities along $\hat{z}$ for the center-of-mass of the interacting partons may occur. Thus, an invariant under $\hat{z}$ boosts variable is used instead of the polar angle θ. This variable is the *rapidity* defined as

$$Y = \frac{1}{2} \ln \left[\frac{E + p\cos(\theta)}{E - p\cos(\theta)} \right],$$

where $(E, \vec{p})$ is the energy-momentum four-vector of the particle. A measurement of rapidity still requires a detector with accurate identification capabilities because

of the mass term entering E. Thus, for practical reasons, it is often preferred to substitute Y with its approximate expression η in the ultra-relativistic limit ($p \gg m$):

$$Y \to \eta + \mathcal{O}(m^2/p^2),$$

where the *pseudo-rapidity* $\eta \equiv -\ln[\tan(\theta/2)]$ is only function of the momenta. As the event-by-event longitudinal position of the actual interaction is distributed around the nominal interaction point with 30 cm r.m.s. width, it is useful to distinguish the *detector pseudo-rapidity*, η_{det}, respect to the $(0, 0, 0)$ nominal interaction point, from the *particle pseudo-rapidity*, η, measured with respect to the z_0 position of the real vertex where the particle originated.

An other convenient variable is the transverse momentum with respect to the beam axis

$$\boldsymbol{p}_T \equiv (p_x, p_y) \to p_T \equiv p \sin(\theta).$$

Since the magnetic field, the trajectory of charged particles produced with non-zero initial velocity in the bending plane of the magnet is described by an helix. It is parametrized using five parameters:

- curvature C, defined as $C \equiv \text{sign}(Q)/2R$, where R is the radius of helix and Q is the charge of the particle;
- signed impact parameter d_0, i.e., the distance of the closest approach to the z-axis, defined as $d_0 \equiv Q(\sqrt{x_c^2 + y_c^2} - R)$, where (x_c, y_c) are the coordinates of the center-guide;
- $\cot(\theta) \equiv p_z/p_T$, where θ is the polar direction of particle at the point of its closest approach to the z-axis;
- φ azimuthal angle, i.e., φ direction of the particle at the point of closest approach to the z-axis;
- z_0, the z coordinate of the point of closest approach to the z-axis.

2.2.2 Tracking system

Three-dimensional charged particle tracking is achieved through an integrated system consisting of three silicon inner subdetectors and a large outer drift-chamber, all contained in a superconducting solenoid of 1.41 T.

Layer ØØ (LØØ)

Layer ØØ is the innermost layer of the microvertex silicon detector [28]. It consist of one layer of microstrip silicon sensors which covers the beryllium beam pipe along 80 cm longitudinally. The state-of-the-art silicon sensors of LØØ can be biased to very high ($\mathcal{O}(500\,\text{V})$) voltages allowing to maintain a good signal-to-noise ratio even after high integrated radiation dose ($\mathcal{O}(5\,\text{Mrad})$). The radiation hardness of such sensors allowed their installation at radius of about 1.5 cm to the beam axis. This feature is a crucial parameter to ensure a good resolution of the decay vertices.

The LØØ strips are parallel to the beam axis allowing the first sampling of the track within the r-φ plane, the inter-strip pitch is 25 µm and the read-out pitch is 50 µm

The signal of the 13,824 channels are fed via special optical fiber cables to the front-end electronic which is placed in a region separated from the sensors and less exposed to the radiation. The operation temperature of this device is around 0 °C maintained by sophisticated cooling system. LØØ contributes to the radial thickness of about $0.6\%X_0$ for only sensors and increases to $1\%X_0$ where pass the cooling system.[3]

Silicon VerteX detector II (SVXII)

The Silicon VerteX detector II [35] is a fine resolution vertex detector which provides five 3D sampling of a track between about 2.5 and 10 cm of radial distance from the beam (see fig. 2.2). Its cylindrical geometry coaxial with the beam is segmented along z into three *barrels* for a total length of 96 cm which provides complete geometrical coverage within $|\eta_{\text{det}}| < 2$ (see fig. 2.2a). Each barrel consists of twelve azimuthal wedges each of which subtends approximately 30°. One wedge of a given barrel comprises five concentric layers of silicon sensors installed to a radial distance from the beam between 2.5 to 10.6 cm as shown in fig 2.2b. Sensors in a layer are arranged into independent readout units, called *ladders*. The ladder components are two double sided strip silicon sensors and a multilayer board where all front-end electronics, biasing circuits etc. are allocated. There are three different possible sort of strip orientations in each sensor's side: r-φ (axial) strip oriented parallel to the beam axis, small angle stereo (SAS) strips whose orientation is tilted by 1.2° with respect to the beam axis and the 90° stereo strips which lie in the transverse plane. All the five layers have axial strips on one side, three of the other sides have 90° stereo and two SAS strips. Resolution into r-z plane is about 70 µm while r-φ resolution is 11 µm. The latter provides an impact parameter resolution of about 20 µm for tracks with high p_T while is only 35 µm for tracks of 2 GeV/c of transverse momentum.

The read-out electronics are arranged at the end of the SVXII along z-axis. The average signal-to-noise ratio is $S/N \geq 10$, with a single hit efficiency grater than 99%. To reduce thermal noise and to prevent thermal expansion SVXII is held roughly constant temperature of 10-15 °C. The average material of SVXII corresponds to $5\%X_0$.

Intermediate Silicon Layer (ISL)

The Intermediate Silicon Layer [11] detector is a silicon tracker placed at intermediate radial distance between the SVXII and the drift chamber (see fig 2.3). The polar coverage extends to $|\eta_{\text{det}}| < 2$ and the coverage along z-axis is 2 m which gives a total active surface of about 3.5 m^2. In the central region ISL consist of a single layer of silicon installed over a cylindrical barrel at radius of 22 cm. In the forward region, two layers of silicon are placed on concentric barrels at radii of 20 and 28 cm. Each silicon layer is azimuthally divided into a 30° wedge structure matching that

[3] The symbol X_0 indicates the radiation length.

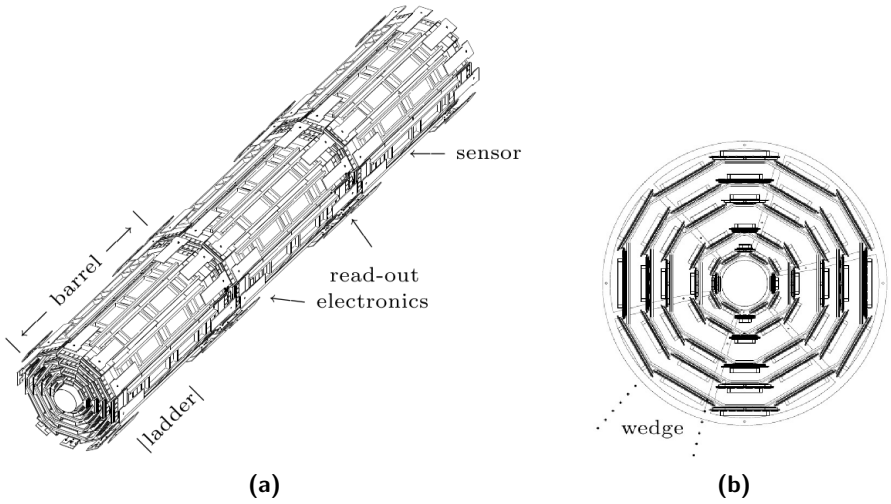


Figure 2.2 Schematic illustration of three instrumented mechanical barrels of SVXII (a) and of the cross-section of a SVXII barrel in the (r, φ) plane (b).

of SVXII. The basic readout unit is the ISL *ladder* which is similar to the SVXII ladder but consists of three, instead od two, sensors wirebonded in series. Average mass of the detector is $2\%X_0$.

Central Outer Tracker (COT)

The outermost tracking volume of CDF II is a large open cell drift chamber called the Central Outer Tracker.[12]

The COT has a coaxial bi-cylindrical geometry and extends, within the central region, from 44 to 132 cm radially from the beam axis. The volume of the COT is filled with Argon (50%) and Ethane (50%) gas mixture. The chamber consists of 8 *superlayers* (see fig 2.4a), each of them contains 12 radial layers of 40 µm diameter gold-plated tungsten sense wires (anode). Four superlayers have their constituent sense wires oriented parallel to the beam axis in order to measure the hit coordinates in the r-φ plane. These are radially interleaved with four *stereo* superlayers having wires canted at angle of either $+3°$ or $-3°$ with respect to the beam-line. Combined readout of stereo and axial superlayers allows the measurement of the r-z hit coordinates. Each superlayer is azimuthally segmented into open drift cells. A drift cell contains a row of 12 sense wires alternating with 13 gold-plated tungsten potential wires which optimizing the electric field of 2.5 kV/cm. The cathode of the detection circuit is the *field panel* which closes the cell along the azimuthal direction. It is made of gold on a 0.25 mm thick Mylar sheet. Innermost and outermost radial extremities of a cell are closed both mechanically and electrostatically by the *shaper panels*, which are Mylar strips carrying field-shaping wires attached.

Wire planes are azimuthal tilted by 35° in order that charges drift approximately perpendicular to radial direction $\hat{r}$, as result of the combined effect of crossed electrical and magnetic field. Drift electrons follow approximately azimuthal trajectories at speed $v \approx 100\,\mu m/ns$. The resulting maximum drift time is about 100 ns, well smaller than the inter-bunch spacing 396 ns, providing the read-out and processing

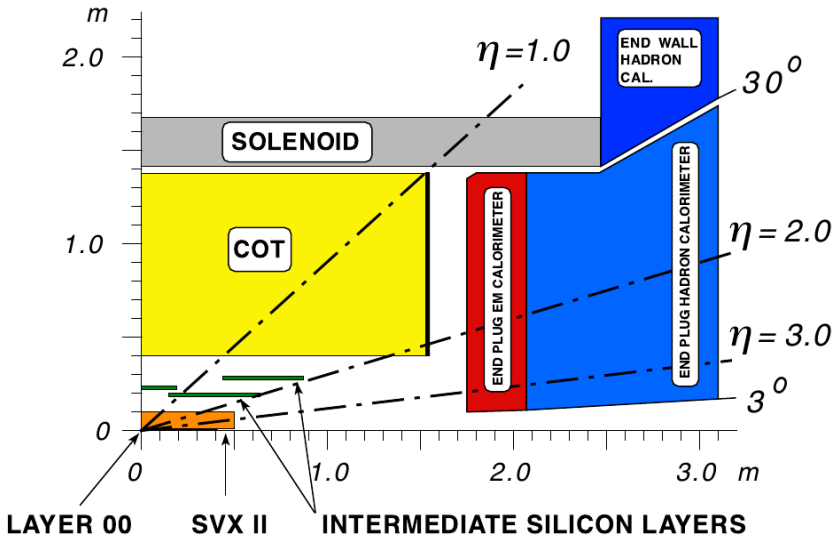


Figure 2.3 Schematic illustration of the CDF II tracking system in the (r, z) plane.

of the COT data available for the Level 1 trigger.

The COT efficiency is grater then 99% for tracks of $p_T > 20\,\mathrm{GeV}/c$ with a single hit resolution $\sigma_{hit} \approx 175\,\mu\mathrm{m}$ and p_T resolution $\sigma_{p_T}/p_T^2 \sim 0.15\%(\mathrm{GeV}/c)^{-1}$.

The COT also provides a measurement of charge particles ionization energy loss (dE/dx). Information about the mass of particle can be obtained combining the values of dE/dx of the particles and their momentum. From information of dE/dx a separation of 1.5σ is obtained between π and K with a momentum greater than $2\,\mathrm{GeV}/c$. COT average material corresponds to $1.6\%X_0$.

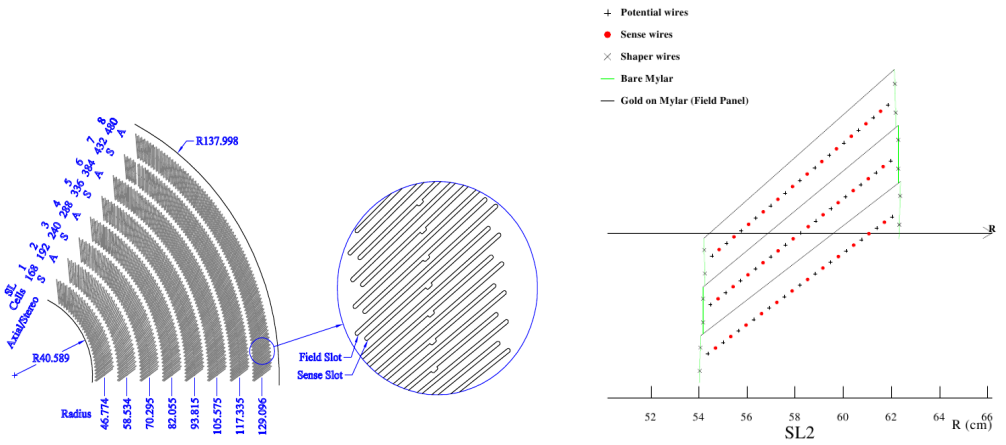
2.2.3 Other CDF II subdetectors

In this subsection are briefly discussed the subdetectors not used in this analysis.

Time Of Flight detector (TOF)

The Time of Flight detector (TOF) is a cylindrical array made of 216 scintillating bars and it is located between the external surface of the COT and the cryostat containing the superconducting solenoid. Bars of 280 cm long and oriented along the beam axis all around the inner cryostat surface at an average radial distance of 138 cm. Both longitudinal sides of the bars collect the light pulse into PMT (photomultiplier) and measure accurately the timing of the two pulses. The time between the bunch crossing and the scintillation signal in these bars defines the β of the charged particle while the momentum is provided by the tracking. Information about the mass of the particle is available through the combination of TOF information and tracking measurements.

The separation between charged pions and kaons with $p_T \lesssim 1.5\,\mathrm{GeV}/c$ is equivalent to 2σ. Considering the p_T selection trigger of this analysis ($p_T > 2\,\mathrm{GeV}/c$) the TOF information is irrelevant.



(a) (b)

Figure 2.4 A 1/6 section of the COT (a). Sketch of three cells in the second superlayer (b) .

Calorimeters

Outside the solenoid, scintillator-based calorimetry covers the region $|\eta_{\text{det}}| \leq 3.6$, and is devoted to the measurement of the net energy deposition of photons, electrons and hadrons using the *shower sampling* technique.

The basic structure consists of alternating layers of passive absorber, which allows the shower process, and plastic scintillator, which measures the energy of the shower. The CDF II calorimeters are finely segmented in solid angle around the nominal collision point, and the angular segmentation is organized in projective *towers* in the η-φ plane.

The central electromagnetic calorimeter (CEM) [16] consists of twenty four 15° wedge in φ. A wedge consists of 31 layers of scintillator interleaved with aluminum-clad lead sheets, divided along η_{det} in ten tower ($\Delta\eta_{\text{det}} \simeq 0.1$). To maintain a constant thickness in X_0, independent from the polar angle, some lead layers are replaced with increasing amounts of acrylic as a function of η_{det}. Light from each tower is shifted in wavelength and guided to the PMT, as shown in fig 2.5a. At a radial depth of about $6X_0$, which is approximately the depth corresponding to the peak of shower development, the central strip multi-wire proportional chambers (CES) measure the transverse shower shape improves the efficiency of separation of $e^{\pm}$ and γ from π^0 and single hadrons.

The total thickness of the electromagnetic section corresponds to approximately $19X_0$ and $1\lambda_{\text{int}}$ (absorption length). The relative energy resolution is $\sigma_E/E = 13.5\%/\sqrt{E}$ plus a constant term of 2% due to calorimeter non-uniformities and to the uncertainty on the calibrations.

Around the electromagnetic calorimeter it is located the central hadronic calorimeter (CHA) which consists of alternating layers of plastic scintillator and iron (passive absorber). The total thickness of the hadron section corresponds to approximately $4.5\lambda_{\text{int}}$, for an energy resolution of $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$.

The lateral region up to $|\eta_{\text{det}}| < 3.6$ it is covered by the *plug* calorimeter (see

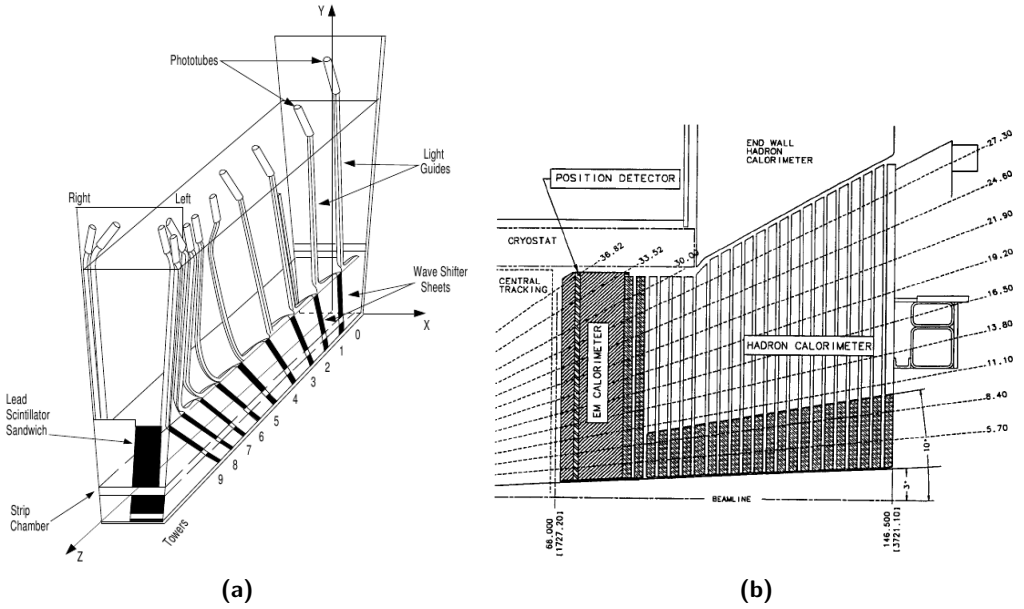


Figure 2.5 Schematic illustration of an azimuthal sector of the central electromagnetic calorimeter (a). Elevation view of one quarter of the plug calorimeter (b).

.

fig 2.5b). The absorber in the plug electromagnetic calorimeter (PEM) consists in twenty three 4.5 mm thick lead plates sandwiched between two 0.5 mm thick stainless-steel sheets. Between the absorber plates are inserted the 4 mm thick scintillator *tiles*. The signal of each tile is collected independently by embedded wavelength-shifter fibers which guide it to the photomultipliers. A preshower detector consist of a thicker (10 mm) amount of scintillator installed in the first layer of PEM.

The plug hadronic calorimeter (PHA) consists of 23 layers of 5 cm thick iron absorber alternated with 6 mm scintillator layers. Within each sampling layer the scintillator is arranged in tiles similar to those used in the PEM.

The total thickness of the electromagnetic section of the plug corresponds to approximately $21X_0$, for an energy resolution of $\sigma_E/E = 16\%/\sqrt{E} \oplus 1\%$. The total thickness of the hadronic section corresponds to approximately $7\lambda_{\text{int}}$, for an energy resolution of $\sigma_E/E = 74\%/\sqrt{E} \oplus 4\%$.

Muon System

CDF II is equipped with scintillating counters and drift tubes installed at various radial distances from the beam to detect muons and shielded by the iron structure of the inner detector. Scintillators serve as trigger and vetoes while the drift chambers measure the φ and z coordinates. These systems cover the whole range of pseudorapidity $|\eta_{\text{det}}| < 2$. The shield is constituted by the iron of the calorimeter, the return yoke and further steel walls intended to filter out the punch-through of hadrons. The amount of material between the interaction point and the innermost muons detector corresponds approximately to $5.4\lambda_{\text{int}}$ for orthogonal incident π.

Different muon sub-systems cover different geometrical regions. In the $|\eta_{\text{det}}| \lesssim$

0.6 region moving outward from the beam we encounter the inner CMU chambers at radial distance of 3.5 m. In the same $\eta_{\rm det}$ region, but separated radially from the CMU by 60 cm thick wall of steel, lie the CMP chambers. The wall of steel is useful to recognize and discard high energy hadrons unchecked by the calorimeter. The muon coverage in the $0.6 \lesssim |\eta_{\rm det}| \lesssim 1$ region is ensured by the CMX chambers, embedded in scintillator counters and placed at radius of 3.5 m. The IMU detectors are instead drift tubes covering the pseudorapidity range $1 \lesssim |\eta_{\rm det}| \lesssim 2$

CDF II triggers on muons only emerging at $|\eta_{\rm det}| \leq 1.5$, where the muon coverage is segmented with sufficient granularity to survive high occupancies, and $p_T \gtrsim 1.5\,{\rm GeV}/c$. The muon emerging at $|\eta_{\rm det}| > 1.5$ can be used only in offline analysis.

Cherenkov Luminosity Counters

The luminosity ($\mathcal{L}$) is inferred from the average number of inelastic interactions ($\overline{N}$) according to

$$\overline{N} \cdot f_{\rm bc} = \sigma^{\rm in}_{p\bar{p}} \cdot \varepsilon \cdot \mathcal{L},$$

where $f_{\rm bc}$ is the bunch crossing frequency, $\sigma^{\rm in}_{p\bar{p}} = (59.3 \pm 2.3)\,{\rm mb}$ is the inelastic $p\bar{p}$ cross-section and ε is the efficiency to detect an inelastic scattering.

The number of interactions is measured by the Cherenkov Luminosity Counters (CLC) [10]. They consist of two separate modules, covering $3.7 \lesssim |\eta_{\rm det}| \lesssim 4.7$ range symmetrically in the forward and backward regions. Each module consists of 48 thin 110–180 cm long, conical, isobutane-filled Cherenkov counters.[4] They are arranged around the beam-pipe in three concentric layers and point to the nominal interaction region. Cherenkov-light is collected into a 2.5 cm diameter long PMTs. With a Cherenkov angle $\theta_C = 3.4°$, the momentum thresholds for light emission are $9.3\,{\rm MeV}/c$ for electrons and $2.6\,{\rm MeV}/c$ for charged pions.

Prompt charged particles from the $p\bar{p}$ interactions are likely to traverse the full counter length, thus generating large signals and allowing discrimination from the smaller signals of particles emitted at the same angle due to the beam halo or to secondary interactions. In addition, the signal amplitude distribution shows distinct peaks for different particle multiplicities entering the counters. This allows a measurement of $\overline{N}$ with 4.4% relative uncertainty in the luminosity range $10^{31} \lesssim \mathcal{L} \lesssim 10^{32}\,{\rm cm}^{-2}\,{\rm s}^{-1}$. This accuracy, combined with the 4% relative uncertainty on the inelastic $p\bar{p}$ cross-section, results in an instantaneous luminosity measured with 5.9% relative uncertainty. This uncertainty does not affect the results of this analysis since ratios of branching fractions, instead of absolute branching fractions, are measured.

2.3 Trigger system

Since the interaction rate at the Tevatron collider is well beyond the current maximum storage rate, the task of separating the great majority of background events from the tiny fraction of interesting events is of crucial importance. This goal is achieved by the trigger system which evaluates the partial information provided by the detector in real time and discards the uninteresting events.

[4] Isobutane guarantees high refraction index and good transparency for ultraviolet photons.

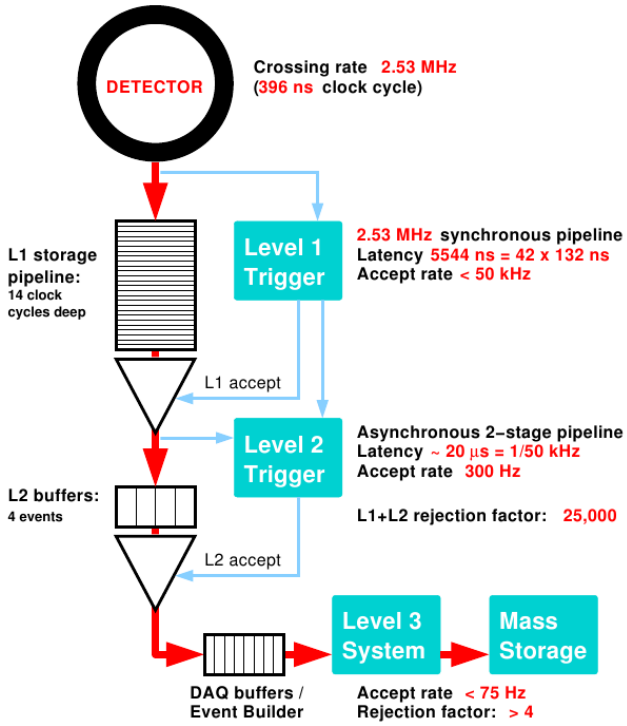


Figure 2.6 Functional block diagram of the CDF II trigger and data acquisition system.

The Tevatron running at 396 ns of interbunch spacing has a collision rate of about 2.53 MHz. The writing of events on permanent memories cannot proceed faster than 100 Hz. For practical reasons the CDF II trigger has been designed as a multi-stage system in order to reduce the acquisition rate allowing to record only the events with a physical interest. Its architecture is modular and divided into three levels, represented in fig 2.6. Each level receives the data event from the previous one and, provided with more accurate detector information and more time for processing, chooses to discard it or to send it to the next level. The Level-1 and Level-2 consist of custom-designed hardware components while the Level-3 consists of commercial CPU running software algorithms. Level-1 receives the data directly from the detector front end electronics. Events passing the Level-3 are stored to permanent memory.

Level 1

In order to avoid deadtime caused by the trigger processing time, the Level-1 has to sustain the clock of the Tevatron. In a complex detector as CDF II, it is inconceivable an effective trigger architecture able to process data and make a decision in less than 396 ns. The *impasse* is overcome with a fully pipelined front-end electronics for the whole detector. The signal of each channel is stored, every 396 ns, in a buffer of a 42-cell long pipeline. This means that the Level-1 has $396 \times 42\,\text{ns} \simeq 16\,\mu\text{s}$ to make its decision before the content of the buffer is deleted. The actual latency of the Level-1 is 5.5 μs.

The Level-1 processes and reconstructs coarse information from the COT (two-

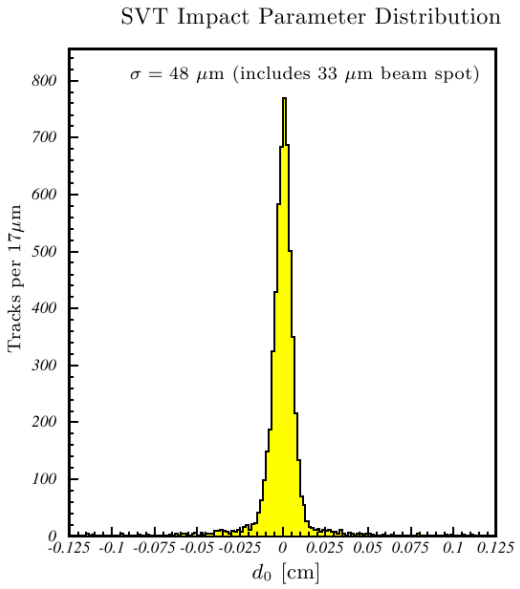


Figure 2.7 Impact parameter distribution as measured by the SVT.

dimensional tracks in the transverse plane), the calorimeter (total energy and presence of single towers over threshold) and the muon system (muon stubs in the chambers).

The eXtremely Fast Tracker (XFT) is a custom processor that identifies two-dimensional tracks in the (r, φ) view of the COT with a momentum greater than $1.5\,\mathrm{GeV}/c$ in the time with the Level-1 decision. It uses pattern matching to first identify short segments of tracks and then to link them into full-length tracks. The output rate of the Level-1 is 18 kHz.

Level 2

The Level-2 trigger combines the information of the calorimeters, the muons chambers ed the tracking system to better identify the interesting physical objected (e, μ, γ, missing transverse energy). Calorimetric information is used to perform clustering and identification of hadronic jets.

The Silicon Vertex Trigger (SVT)[14, 23] is the principal component of the Level-2. It is a processor dedicated to the identification of the events with a secondary vertex, especially that associated to decays of heavy quarks. The SVX combines XFT track information with the SVXII hits and reconstructs two dimensional tracks with $p_T > 2\,\mathrm{GeV}/c$ with an impact parameter resolution of about 35 µm. The impact parameter distribution of the real data is shown in fig 2.7.

The maximum latency of Level-2 is about 20 µs for each event and the output rate is 300 Hz.

Level 3

This stage of the trigger is implemented by software running in parallel on commercial processors. Level-3 reconstruct the event provided by Level-2 at full

detector resolution. Level-3 codes are very similar to the offline reconstruction codes. The Level-3 decision to write on tape happens after the full reconstruction of the event is completed and the integrity of its data is checked in less than 10 ms. Typical size for an event is 150 kB. The available output rate passed from 75 Hz.

3 Data sample

In this chapter we describe the data sample used in this work and the procedure to extract the $B_s^0 \to D_s^{\pm} h^{\mp}$ (where $h = \pi$ or K) decays (referred to as "signal") from other events. We also describe the Monte Carlo simulation, which is a crucial ingredient of this analysis.

3.1 $B_s^0 \to D_s^{\pm} h^{\mp}$ at CDF II

Data collected at CDF II is organized in acquisition periods called *runs*; a run is a period of continuous operation of the CDF II Data Acquisition system. To ensure the quality of data requested for the physics measurements each run has to be approved by the on-line and off-line operators and by the detectors experts. The list of the approved runs is called the *GoodRunList*.

The analysis described in this thesis uses data collected between December 2004 (run 190697) and September 2011 (run 312510). After the application of the standard CDF data-quality requirements on triggers and of the GoodRunList the data sample corresponds to an integrated luminosity of $\int \mathcal{L}\, dt \simeq 9.3\,\text{fb}^{-1}$.

To discuss in detail the trigger and the off-line selection, it is useful to define some relevant quantities used in the analysis. They are calculated in the laboratory frame, and are illustrated in figure 3.1.

Transverse plane: the plane perpendicular to the proton beam direction, in which the profile region of the beam is approximately Gaussian with r.m.s. $\sigma_T \approx 30\,\mu\text{m}$.

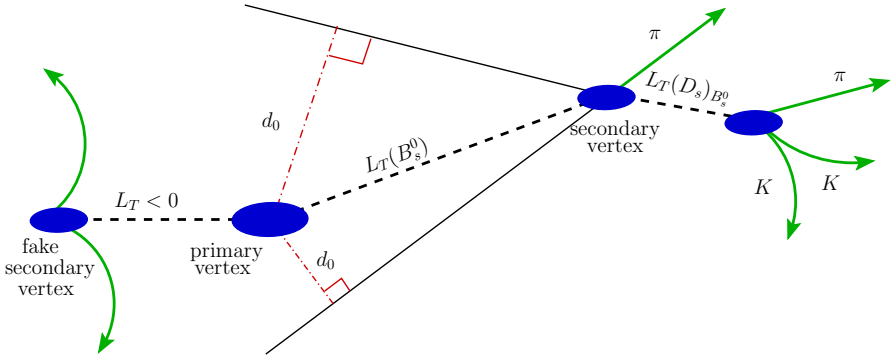


Figure 3.1 Schematic illustration of a $p\bar{p}$ event containing a $B_s^0 \to D_s^{\pm} h^{\mp}$ decay. Ellipses indicate vertices, arrows indicate tracks of charge particle. Nothing are in scale.

Transverse momentum $\boldsymbol{p_T}$: the projection of the momentum vector onto the transverse plane.

Primary vertex: the space-point of the reconstructed primary $p\bar{p}$ interaction, where b-quark, once produced, quickly hadronizes.

Secondary vertex: the space point in which the decay of a long-lived particle occurs. The components of its displacement with respect to the primary vertex in the transverse plane are indicated by the vector $\boldsymbol{x}_v = \boldsymbol{\beta}_T \gamma ct = (\boldsymbol{p}_T/m)ct$, for a particle of mass m and momentum p that decays at time t after its production.

Transverse decay-length ($\boldsymbol{L_T}$ or $\boldsymbol{L_{xy}}$): the displacement of a secondary vertex with respect to the primary one, projected onto the transverse momentum vector of the decaying particle. The transverse displacement of the secondary vertex ($\boldsymbol{x}_v$) may not be collinear with ($\boldsymbol{p}_T$) because of the measurement uncertainties. Thus, the transverse decay-length,

$$L_T \equiv \frac{\boldsymbol{p}_T \cdot \boldsymbol{x}_v}{p_T}, \tag{3.1}$$

is usually preferred to $\boldsymbol{x}_v$ as an estimator of the transverse decay-length traveled before decay. L_T is typically positive for a true long-lived decays, while it is negative or positive with almost equal probability for decays from fake vertex or for combinations of prompt tracks.

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

Pointing angle ($\boldsymbol{\alpha_T}$ or $\boldsymbol{\alpha_{xy}}$): the angle, in the transverse plane, between the momentum of the B_s^0 candidate and the displacement vector of the secondary vertex ($\boldsymbol{x}_v$) with respect to the primary one.

Impact parameter ($\boldsymbol{d_0}$): the component of the distance of the closest approach between a track and the primary vertex in the transverse plane. This is a signed quantity defined as

$$d_0 \equiv \frac{(\boldsymbol{p}_T \times \boldsymbol{x}_v) \cdot \hat{\boldsymbol{z}}}{p_T}. \tag{3.2}$$

The impact parameter is typically different from zero for products of long-lived decays, while is comparable with the convolution of its resolution and the transverse size of the beam for particles produced in the vicinity of the primary vertex (prompt background).

Azimuthal opening angle ($\boldsymbol{\Delta\varphi}$): the opening angle between two outgoing particles projected in the transverse plane.

3.2 Online data selection

The data have been collected with the Displaced-Tracks Trigger. The Displaced-Tracks Trigger is composed by several trigger *paths*. A "trigger path" is a well defined sequence of Level-1, Level-2 and Level-3 requirements. The CDF standards require each data sample used for physics analysis to have been selected by a trigger path which is appropriate for that particular analysis. This means that no use is made of the so-called "volunteers" (potentially valid events randomly picked up from trigger paths devoted to a different purpose). The specific trigger paths used for this analysis can be grouped in three main categories having similar kinematics requirements: `B_CHARM`, `B_CHARM_LOWPT` and `B_CHARM_HIGHPT`. In particular the `B_CHARM_LOWPT` and `B_CHARM_HIGHPT` paths have different thresholds on the transverse momenta with respect to the `B_CHARM`, lower the first and higher the second.

The purpose of having several paths with different transverse momenta requirements is to keep the trigger as efficient as possible at higher instantaneous luminosities. In fact, during a store the luminosity decreases as a function of the time approximately as an exponential. Since a set of trigger selections is optimized for an average reference luminosity, during the store, it can be either too loose, suffering high trigger accept-rates and inducing dead-time in the trigger decision, or too tight, leaving a fraction of trigger bandwidth unused. Therefore CDF uses simultaneously different trigger selections (having different p_T-thresholds) and self-adjusting prescale factors ("dynamic prescale", changing on a scale of milliseconds) to optimally exploit the available bandwidth in a single store.[1]

Displaced-Tracks Trigger is a generic trigger for the flavor physics. It selects a large variety of decays, as $B^0 \to h^+h'^-$, $B^0_{(s)} \to D^-_{(s)}\pi^+$, $D^{\star+} \to D^0\pi^+$, $B^0 \to D^+_s D^-_s$, etc., and that is possible because it uses criteria which are common to all flavored decays, for instance a "long" lifetime. In particular it requires a pair of tracks satisfying some requirements onto typical variables in flavor physics, as the impact parameter.

Our data sample $B^0_s \to D^\pm_s K^\mp$ has a four-tracks final state which may have more than one pair of tracks satisfying the trigger requirements. If we indicate with 1 and 2 the tracks of a trigger-pair, the requirements of the `B_CHARM` path are:

Level-1 Two XFT tracks having $p_T(1,2) > 2\,\mathrm{GeV}/c$ and $p_T(1) + p_T(2) > 5\,\mathrm{GeV}/c$ with opposite curvature are required. Also it is required an angular opening in the transverse plane $0^\circ < \Delta\varphi < 135^\circ$, to reduce the fraction of events with light-quark background, which has azimuthally-opposed direction.

Level-2 Trigger requires a pair of oppositely-curved SVT tracks that satisfy a minimal linearized fit quality requirement: $\chi^2_{\mathrm{SVT}} < 25$ [17]. The requirements of the first level are confirmed, but the azimuthal opening-angle requirement is further tightened with respect to the previous trigger stage to $2^\circ < \Delta\varphi < 90^\circ$. A lower threshold in impact parameter d_0 is required to enrich the sample in b-hadrons. An upper threshold is dictated by hardware constrains, then

[1] A trigger prescaled of a factor N (acceptance fraction) will only accept every randomly-chosen Nth event that satisfies the trigger requirements. This reduces the trigger accept-rate by a factor N, leaving unchanged the signal-to-background ratio in the sample.

Table 3.1 Kinematic requirements of `B_CHARM` path and its variants.

trigger path	requirement
`B_CHARM_LOWPT`	$p_T(1,2) > 2\,\mathrm{GeV}/c$ $p_T(1) + p_T(2) > 4\,\mathrm{GeV}/c$
`B_CHARM`	$p_T(1,2) > 2\,\mathrm{GeV}/c$ $p_T(1) + p_T(2) > 5\,\mathrm{GeV}/c$
`B_CHARM_HIGHPT`	$p_T(1,2) > 2.5\,\mathrm{GeV}/c$ $p_T(1) + p_T(2) > 6.5\,\mathrm{GeV}/c$

the trigger requirement is $120\,\mu\mathrm{m} < |d_0(1,2)| < 1\,\mathrm{mm}$. The spatial resolution of SVT ($35\,\mu\mathrm{m}$ for tracks with $p_T > 2\,\mathrm{GeV}/c$) in identifying secondary vertices is further exploited: positive decay-length of the tracks are required, $L_T > 200\,\mu\mathrm{m}$.

Level-3 The Level-1 and Level-2 criteria are reapplied on Level-3 tracks. In addition, a requirement on the longitudinal separation between the two tracks at the point of their minimum distance from the beam is applied: $|\Delta z| < 5\,\mathrm{cm}$. This significantly reduces the fraction of combinations of two tracks descending from particles produced in distinct primary vertex. A $|\eta| < 1.2$ requirement on tracks excludes events with particles outside the XFT fiducial acceptance.

The `B_CHARM_LOWPT` trigger path applies the same requirement of the `B_CHARM`, but loosing few of them. At Level-1 charge-opposed tracks are not required. The threshold on transverse momenta is $p_T(1,2) > 2\,\mathrm{GeV}/c$.

The `B_CHARM_HIGHPT` trigger path requirements are the same of the `B_CHARM`, but with higher thresholds on transverse momentum and scalar sum of transverse momenta, respectively $p_T(1,2) > 2.5\,\mathrm{GeV}/c$ and $p_T(1) + p_T(2) > 6.5\,\mathrm{GeV}/c$. The momenta thresholds for the three trigger paths are summarize in the table 3.1.

3.3 Extraction of the $B_s^0 \to D_s^{\pm} h^{\mp}$ signal

The first step of the off-line analysis consists in applying a baseline selection to the events collected by the `B_CHARM`, `B_CHARM_LOWPT` and `B_CHARM_HIGHPT` trigger paths. The reconstruction of $B_s^0 \to D_s^{\pm} h^{\mp}$ candidates is solely based on tracking, neglecting any particle identification information. The particle identification (PID) in CDF is made observing the loss of energy of the track in the COT. Since the available separation is about 1.5σ (for tracks with a momentum greater than $2\,\mathrm{GeV}/c$), event-by-event separation is not possible.[2]

For each candidates, the invariant mass of all possible pairs of oppositely-curved tracks is computed, requiring an invariant mass in the range $[1.005, 1.035]\,\mathrm{GeV}/c^2$ to form the ϕ meson. A third track is associated to these two tracks to form the

[2] In fact in this work we follow a statistical approach that combines information from PID and kinematics into a fit of composition as described in chapter 6.

Table 3.2 Baseline selection for $B_s^0 \to D_s^- h^+$ data sample.

$B_s^0 \to D_s^- h^+$	unit	selection
$\chi^2_{2D}(B_s^0)$	–	< 40
$\chi^2_{3D}(B_s^0)$	–	< 100
$L_T(B_s^0)$	µm	> 0
$p_T(B_s^0)$	GeV/c	> 5
$\lvert d_0(B_s^0)\rvert$	µm	< 100
$L_T(D_s)_{B_s^0}$	µm	> 0
$\Delta R(D_s, h)$	–	< 2
$p_T(h)$	GeV/c	> 2
$\lvert\eta\rvert$	–	< 1
$m(KK)$	GeV/c^2	$[1.012, 1.027]$
$m(\phi h)$	GeV/c^2	$[1.9485, 1.9885]$

D_s candidate in a mass range $[1.7, 2.2]\,\mathrm{GeV}/c^2$. An additional track is associated to the D_s to reconstruct the B_s^0 candidates. The invariant mass is calculated in the $D_s\pi$ hypothesis and a mass range of $[4.4, 6.6]\,\mathrm{GeV}/c^2$ is required.

In a hadronic collider, considering the high number of random tracks that accidentally form an invariant mass in the B_s^0 window, a simple invariant mass selection is not enough to reject the high amount of events not coming from a B_s^0. To reduce such a background, additional selection requirements are made on the following B_s^0 and D_s candidates quantities: transverse impact parameter $|d_0(B_s^0)|$, transverse momentum $p_T(B_s^0)$, transverse decay-length $L_T(B_s^0)$ and $L_T(D_s)_{B_s^0}$. We, also, required the $D_s^- - h^+$ pair to satisfy a two- and three-dimensional fit to the B_s^0 decay vertex returning in two parameters measuring the quality of these fits, respectively $\chi^2_{2D}(B_s^0)$ and $\chi^2_{3D}(B_s^0)$. In addition we require a pseudo rapidity $|\eta|$ range for all four tracks and $\Delta R(D_s^-, h^+)$ selection, where $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ is the angular distance between the D_s^- candidate and the h^- track. Furthermore we choose a tighter window in the mass region both for the ϕ and for D_s^- candidates. The selection requirements are reported in table 3.2.

The invariant mass distribution in $D_s^-\pi^+$ mass assignment is reported in figure 3.2. A clear peak at the nominal B_s^0 mass it observed. At higher masses region we expected background events due to random tracks combinations ("combinatorial background"). Instead in the lower mass region we expected also the contribution of the mis-reconstructed events ("physics background"). To get a rough estimate of the signal yield a simple χ^2-binned fit was performed. A Gaussian shape was assumed to parameterize the "signal" peak, while an exponential distribution was used to parameterize the contribution given by combinatorial background. We do not consider the mis-reconstructed region in this rough fit because it is complex to parameterize that region. The physics background distributions will be studied in detail in the chapter 6. We estimate a yield of 5029 ± 278 $B_s^0 \to D_s^-\pi^+$ events forming a signal peak with a width $\sigma = (19.7 \pm 0.6)\,\mathrm{MeV}/c^2$.

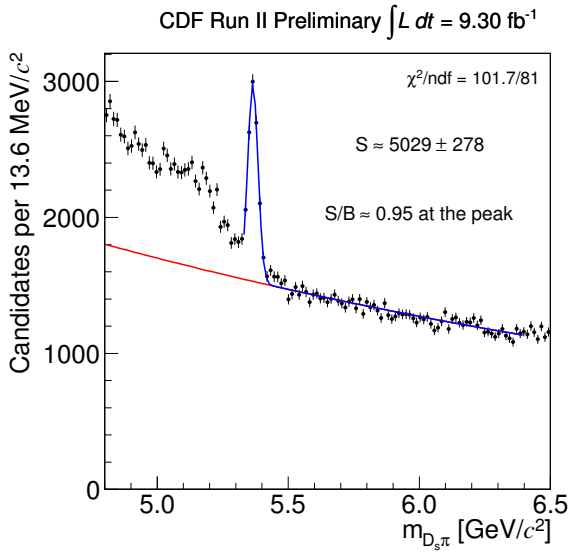


Figure 3.2 Invariant mass distribution of the pair $D_s^- \pi^+$ after the baseline cuts summarized in the table 3.2.

3.4 Monte Carlo simulation

We used the official CDF II simulation to study the general features of the $B_s^0 \to D_s^{\pm} h^{\mp}$ decays and mis-reconstructed decays. The simulation is used in several parts of this analysis: it is a fundamental tool for the optimization procedure of selection (chap. 4) and for the extraction of the mass and momentum templates (chap. 6). In the optimization procedure we use an Artificial Neural Network, trained with the Monte Carlo of the $B_s^0 \to D_s^{\pm} K^{\mp}$, then a well tuned simulated sample is crucial. Here, we summarize only the general features of the standard CDF II simulation, without technical details.

3.4.1 CDF II simulation

We used the BGENERATOR package to generate large samples of b-hadron decays [36]. BGENERATOR simulates the production and the decay of b-hadrons only: no fragmentation products, collision remnants, pile-up[3] events or information about QCD background and fragmentation are present in the simulated data. These information are superfluous for this work since we used the Monte Carlo simulation just to study the features of the decays of interest (signal and physics background), while we extracted the information on background from the data.

The simulation takes into account how the detector and trigger configurations change during data-taking period: changes of the silicon coverage, of the XFT and the SVT configurations, as well as of the Displaced-Track Trigger trigger selection. For convenience the $B - \bar{B}$ 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 3.3 shows the expected $D_s\pi$-mass distribution of the data sample, resulting from the Monte

[3]The overlapping interactions for each bunch crossing.

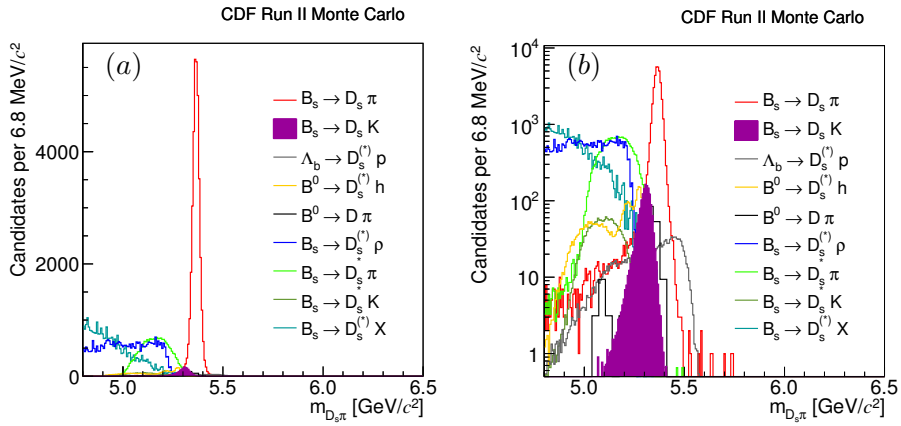


Figure 3.3 $D_s\pi$-mass distribution of the simulated events (a) (log scale (b)).

Carlo simulation of each mode normalized using the branching fractions derived from the current experimental knowledge [18]. We generated about 1.6×10^7 events for the signal decays: $B_s^0 \to D_s^- \pi^+$ and $B_s^0 \to D_s^\pm K^\mp$. As shown in fig 3.3 we expected, in the $D_s\pi$ mass assignment, a high peak due to the $B_s^0 \to D_s^- \pi^+$ decay at the B_s^0 nominal mass, while the $B_s^0 \to D_s^\pm K^\mp$ peak is shifted at lower masses, and it has an asymmetric tail at lower masses. The different contributions of mis-reconstructed decays can be seen in figure 3.3. We generated about 8×10^6 events for each of the following mis-reconstructed decays :

- $B_s^0 \to D_s^{\star -} \pi^+$ and $B^0 \to D_s^{(\star)+} \pi^-$, where the γ/π^0 from the $D_s^\star$ is not reconstructed;
- $B_s^0 \to D_s^{(\star)-} \rho^+$, where the γ/π^0 from the $D_s^\star$ is not reconstructed, and the same happens to the π^0 from $\rho^+ (\to \pi^+ \pi^0)$;
- $B_s^0 \to D_s^{(\star)-} \pi^+ \pi^0$, similar to the decay described above;
- $B_s^0 \to D_s^{\star\pm} K^\mp$ and $B^0 \to D_s^{(\star)-} K^+$, where the $D_s^{\star -}$ is not reconstructed, in addition the mass assignment is wrong because a K is mistaken for a π;
- $B_s^0 \to D_s^{(\star)-} e^+ \nu_e$ and $B_s^0 \to D_s^{(\star)-} \mu^+ \nu_\mu$, as above we have a wrong mass assignment and further $\nu_{e,\mu}$ not reconstructed;
- $\Lambda_b^0 \to D_s^{(\star)-} p$, where it has a mis-assigned mass;
- $B^0 \to D^- [\to K^+ 2\pi^-] \pi^+$, where there is a double-wrong mass assignment.

3.4.2 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 signal distribution we must accurately subtract background. However this is not possible with an high accuracy from the sample selected by the our baseline cuts, since the amount of background in the sample is sizeable (see figure 3.2).

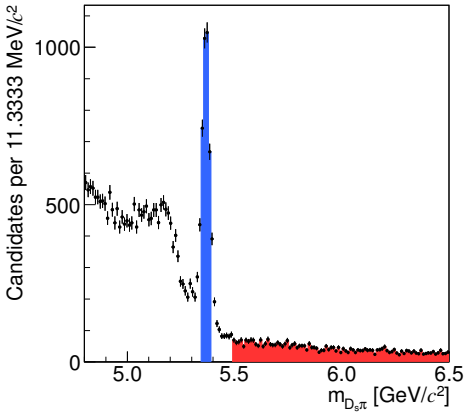


Figure 3.4 $D_s\pi$-mass distribution after cuts reported in table 3.3.

We then decided to use a tighter selection, inspired to the work done in ref. [2]. Therefore we used different requirement reported in the table 3.3, and the $D_s\pi$-mass distribution obtained is reported in figure 3.4. The background level is now acceptable for our purpose.

To remove the effect of the background we applied a procedure called *sideband subtraction*: for each quantity, we subtracted the distributions of *background* candidates (red region in figure 3.4) from the distributions of *signal plus background* candidates (blue region). The signal plus background candidates are defined as those found in the invariant-mass range $|m_{D_s\pi} - \mu| < 2\sigma$ where $\mu = 5.3663\,\mathrm{GeV}/c^2$ and $\sigma = 20\,\mathrm{MeV}/c^2$ are respectively the parameters returned from a simple χ^2-binned fit performed on the mass distribution in figure 3.4. For the background candidates, we assumed that their contribution below the signal peak is dominated by random tracks satisfying the selection requirements (combinatorial background). We sampled this component using candidates at higher masses with respect to the signal peak; we therefore used, as background candidates, those in the mass range $[5.5, 6.5]\,\mathrm{GeV}/c^2$. Assuming an exponential shape for the combinatorial background, we rescaled the distribution of background candidates to the number of background expected below the signal.

Kinematic composition of the sample

The data sample is collected using three different trigger paths: `B_CHARM_LOWPT`, `B_CHARM`, and `B_CHARM_HIGHPT`. As explained in section 3.2, these trigger paths have different momentum thresholds, thus our final data sample is composed by a mixture of different kinematics. 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 CDF simulation does not have access to the database trigger dynamical

Table 3.3 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
$L_T(B_s^0)$	µm	> 300
$p_T(B_s^0)$	GeV/c	> 5.5
$\lvert d_0(B_s^0)\rvert$	µm	< 60
$L_T(D_s)_{B_s^0}$	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	$[1.948, 1.988]$

Table 3.4 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%

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 3.4. The $D_s\pi$ mass distribution is shown in the figure 3.5 for the three different sub-samples.

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

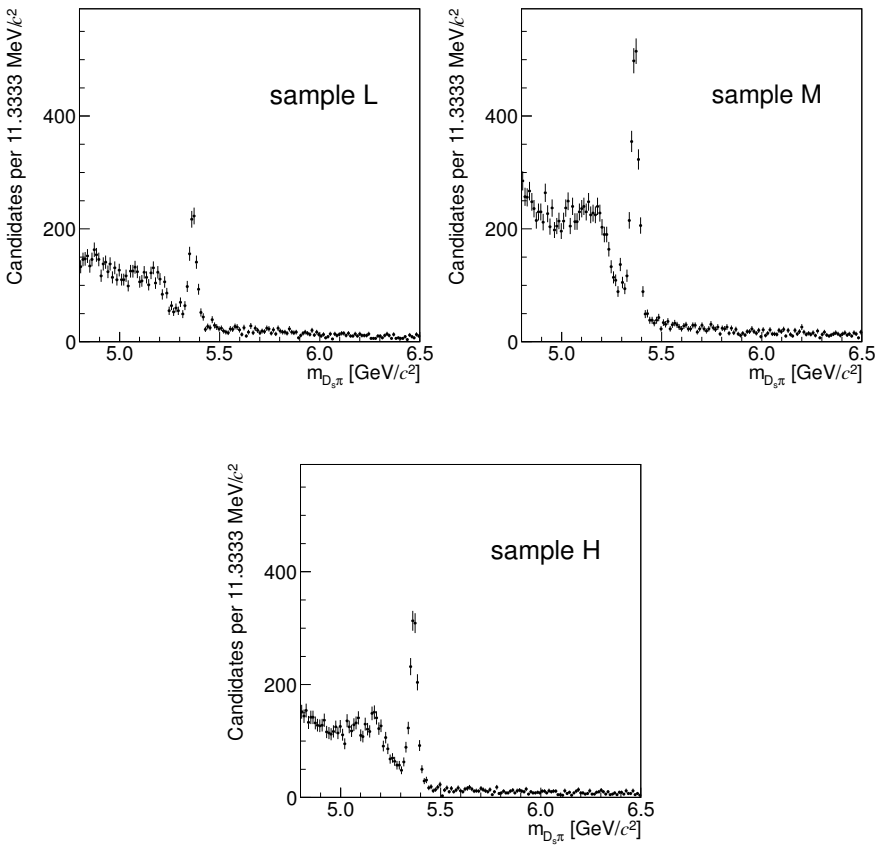


Figure 3.5 Invariant mass distribution of the pair $D_s\pi$ in the three samples described in the text.

our Monte Carlo, we generate only the B_s^0 candidates (see section 3.4) 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 3.6 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.

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 3.7. 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}).

$p_T(B)$ 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. 3.8). This is not necessarily an indication of a problem in the way our detector simulation emulates the reconstruction of transverse momentum. Most likely, this is simply due to the production spectra of B_s^0 mesons not being exactly in agreement with the nominal spectrum assumed in BGENERATOR.

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. This is completely equivalent to correcting the input generator spectrum, but is simpler to implement in practice. We parameterize the histogram ratio with the following function $\mathcal{F}$ (see fig 3.8b, 3.8d and 3.8f):

$$\mathcal{F}(p_T(B); a_0, a_1, a_2, a_3) = a_0 \cdot \mathrm{erf}(a_1(p_T(B) - a_2)) + a_3,$$

where a_0, a_1, a_2, a_3 are free parameters in the fit and $\mathrm{erf}(x) = 2\int_x^\infty \exp(-t^2)dt/\sqrt{\pi}$. 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. The comparisons are shown in the fig 3.9, 3.10 and 3.11, 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 $B_s^0 \to D_s^-\pi^+$.

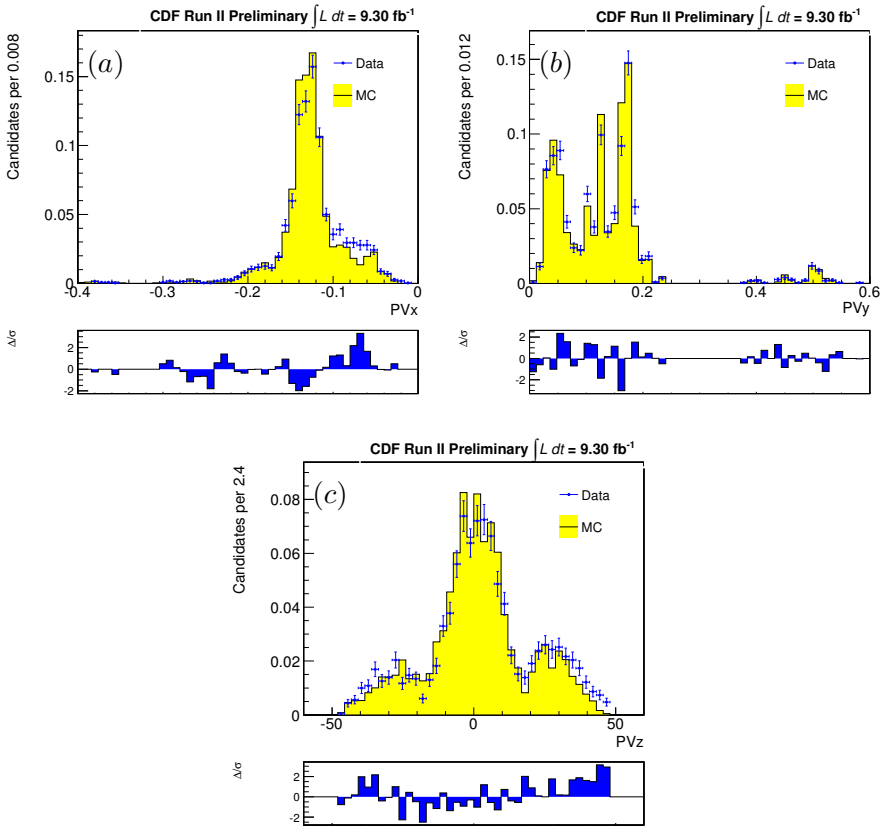


Figure 3.6 Distribution of the primary vertex coordinate in data (points with error bars) and in Monte Carlo (filled histogram): x-coordinate (a), y-coordinate (b), z-coordinate (c).

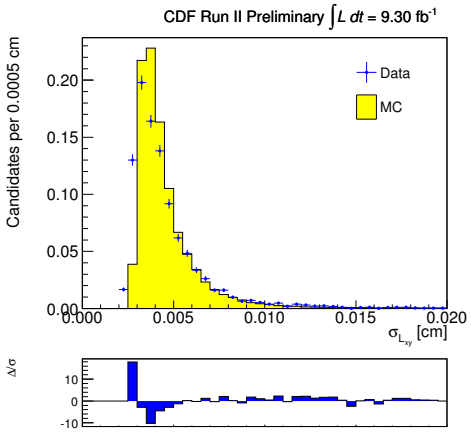


Figure 3.7 Distribution of the error on the transverse decay-length in data (points with error bars) and in Monte Carlo (filled histogram).

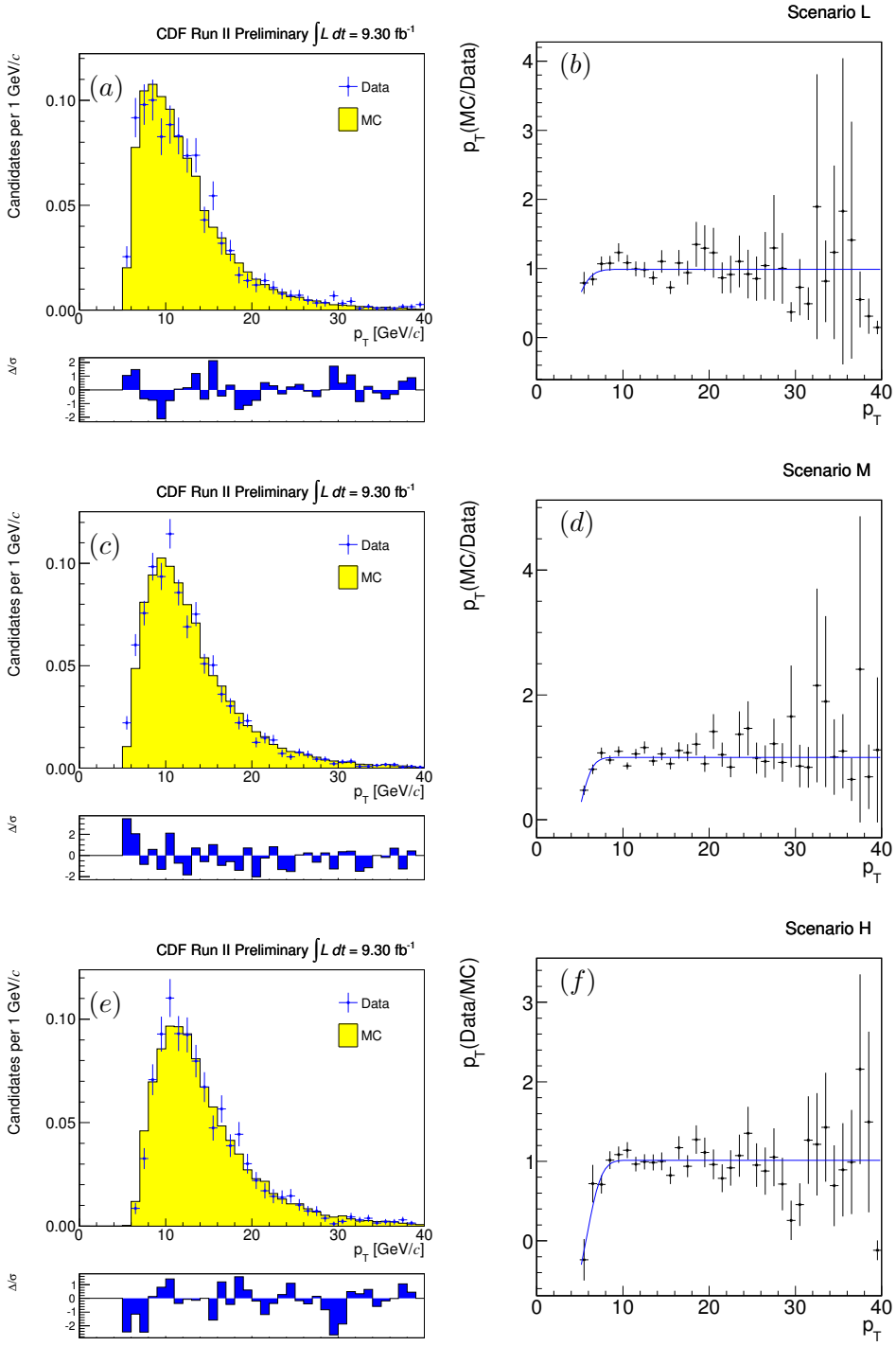


Figure 3.8 Data-simulation comparison of the $p_T(B_s^0)$ distribution in the sub-samples L (a), M (c) and H (e) (filled histograms are the Monte Carlo, dots with errors are data). Ratio between data histogram and Monte Carlo histogram for sub-samples L (b), M (d) and H (f).

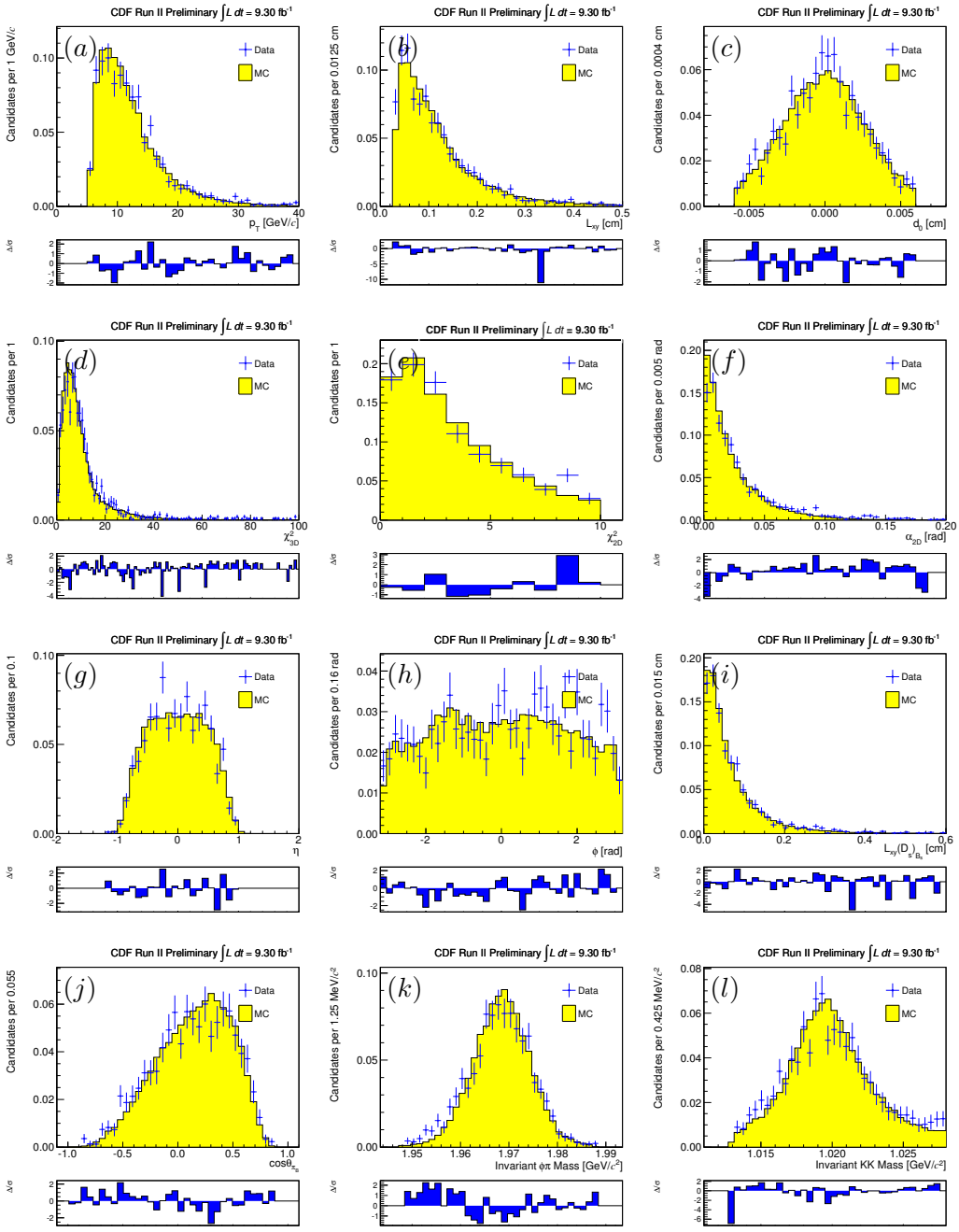


Figure 3.9 Comparison of background-subtracted distributions in the $B_s^0 \to D_s^- \pi^+$ decay and equivalent Monte Carlo distributions for sub-sample L: $p_T(B_s^0)$ (a), $L_T(B_s^0)$ (b), $d_0(B_s^0)$ (c), $\chi^2_{3D}(B_s^0)$ (d), $\chi^2_{2D}(B_s^0)$ (e), α_T (f), $\eta(B_s^0)$ (g), $\varphi(B_s^0)$ (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).

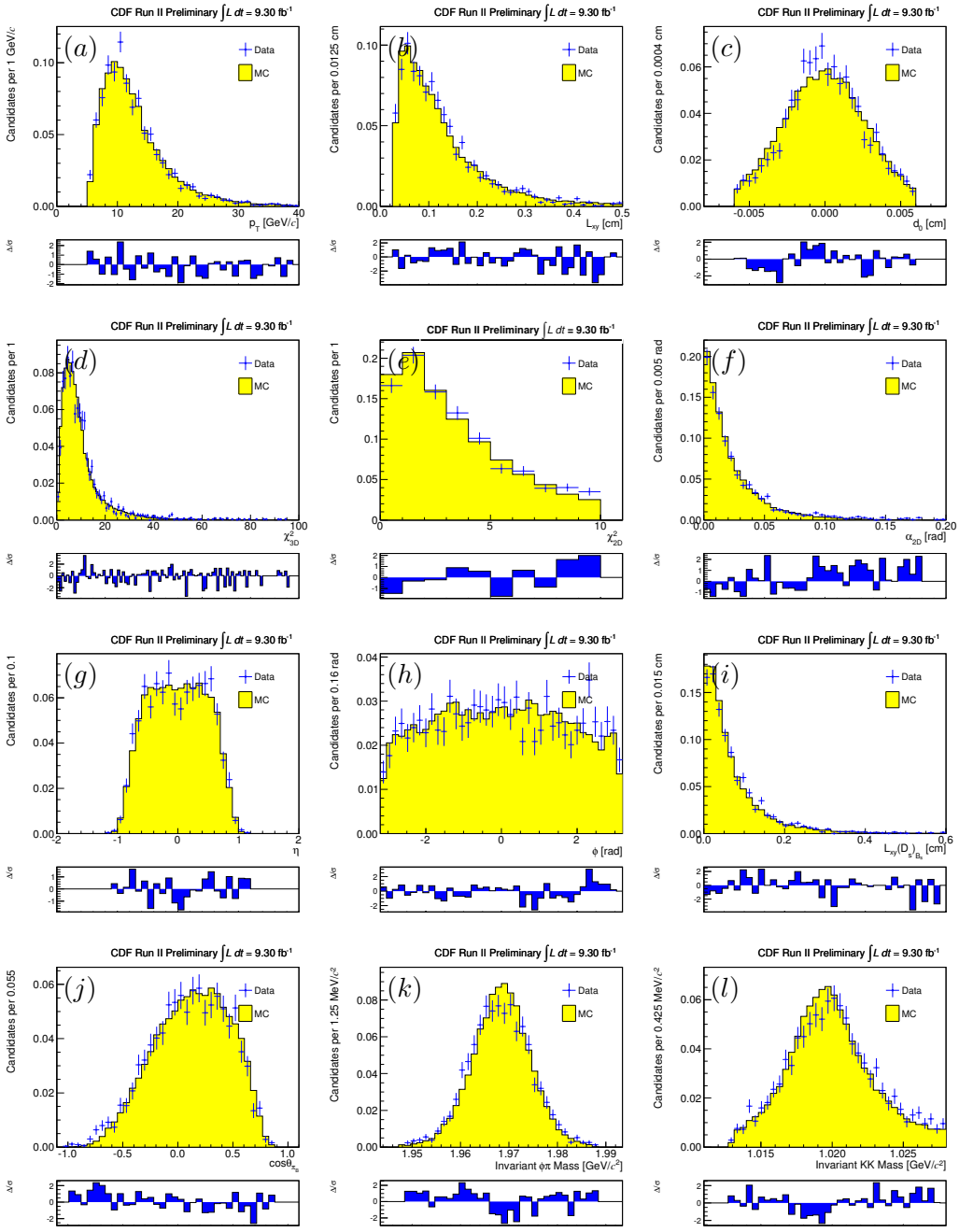


Figure 3.10 Comparison of background-subtracted distributions in the $B_s^0 \to D_s^- \pi^+$ decay and equivalent Monte Carlo distributions for sub-sample M: $p_T(B_s^0)$ (a), $L_T(B_s^0)$ (b), $d_0(B_s^0)$ (c), $\chi^2_{3D}(B_s^0)$ (d), $\chi^2_{2D}(B_s^0)$ (e), α_T (f), $\eta(B_s^0)$ (g), $\varphi(B_s^0)$ (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).

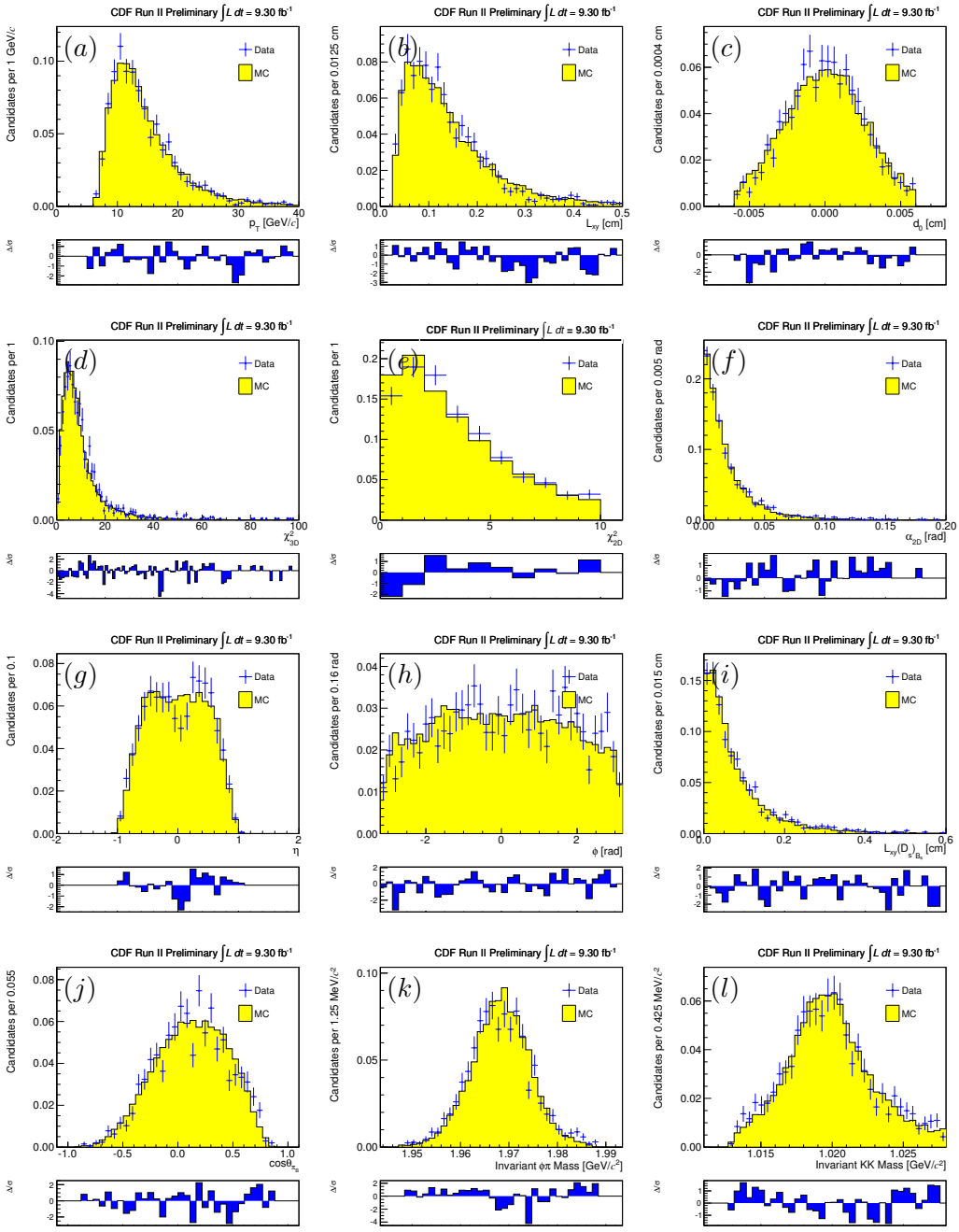


Figure 3.11 Comparison of background-subtracted distributions in the $B_s^0 \to D_s^- \pi^+$ decay and equivalent Monte Carlo distributions for sub-sample H: $p_T(B_s^0)$ (a), $L_T(B_s^0)$ (b), $d_0(B_s^0)$ (c), $\chi^2_{3D}(B_s^0)$ (d), $\chi^2_{2D}(B_s^0)$ (e), α_T (f), $\eta(B_s^0)$ (g), $\varphi(B_s^0)$ (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-based cuts optimization

> I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.
>
> Alan Turing

In this chapter we describe the procedure used to optimize the selection of the data sample, which is based on an Artificial Neural Network. The final configuration is chosen maximizing the score function $S/\sqrt{S+B}$ *for the* $B_s^0 \to D_s^{\pm} K^{\mp}$ *decay.*

4.1 Artificial Neural Network

An Artificial Neural Network (NN) [34] is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. ANNs, like people, learn by examples. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurones. This is true of ANNs as well.

Neural Networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained NN can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions.

There are three major learning paradigms, each corresponding to a particular abstract learning task. These are supervised learning, unsupervised learning and reinforcement learning. Here we do not discuss the unsupervised learning and the reinforcement learning, we will talk only about the supervised learning, the learning paradigm used in this work and in the majority of the HEP applications.

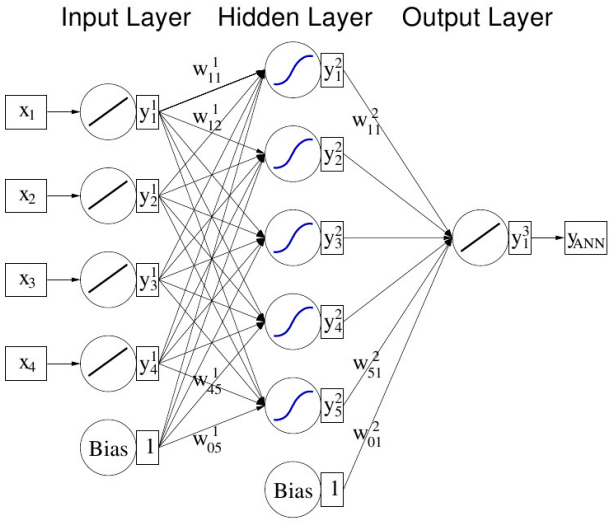


Figure 4.1 Multilayer perceptron with one hidden layer.

In the supervised learning there is an external teacher, so that each output unit is told what its desired response to input signals ought to be. During the learning process global information may be required. An important issue concerning supervised learning is the problem of error convergence, i.e. the minimization of error between the desired and computed unit values. The aim is to determine a set of weights which minimizes the error. The most common algorithm for adjusting the weights that optimise the classification performance of a neural network is the so-called *back propagation*. In the back propagation algorithm the error is backward propagating to the weights.

4.2 Neural Network implementation

The Neural Network used in this work is implemented in the package TMVA [29], a toolkit for multivariate analysis. Here, we describe with more details the algorithm of the NN used in this analysis.

An Artificial Neural Network can be considered as a simulated collection of interconnected neurons, with each neuron producing a certain response at a given set of input signals. By applying an external signal to some (input) neurons the network is put into a defined state that can be measured from the response of one or several (output) neurons. One can therefore view the neural network as a mapping from space of input variables $x_1, \ldots, x_{n_{\text{var}}}$ onto one-dimensional output variable (e.g. in case of a signal-vs-background discriminant problem). The mapping is nonlinear if at least one neuron has a nonlinear response to its input.

While in principle a neural network with n neuron can have n^2 directional connections, the complexity can be reduced by the organising the neurons in layers and only allowing direct connections from a given layer to the following layer (see fig. 4.1). This kind of NN is termed *multilayer perceptron*. The first layer of a multilayer perceptron is the input layer, the last one the output layer, and all others are *hidden* layers. The first layer consists of n_{var} neurons that hold the input

values, $x_1, \ldots, x_{n_{\text{var}}}$, where n_{var} is the number of variables for the problem. While the output layer consists of one neuron that holds the output variable, the NN estimator y_{ANN}. A weight is associated to each directional connection between the output of one neuron and the input of another neuron. The output of a NN (here for simplicity assumed to have a single hidden layer) is given by

$$\begin{gathered} y_{\text{ANN}} = \rho(y_1^{(2)}, \ldots, y_{n_{\text{h}}}^{(2)} | w_{11}^{(2)}, \ldots, w_{n_{\text{h}}1}^{(2)}) = \\ \rho\Big(\rho\Big(x_1 | w_{11}^{(1)}, \ldots, w_{n_{\text{var}}1}^{(1)}\Big), \ldots, \rho\Big(x_{n_{\text{var}}} | w_{11}^{(1)}, \ldots, w_{n_{\text{var}}1}^{(1)}\Big) | w_{11}^{(2)}, \ldots, w_{n_{\text{h}}1}^{(2)}\Big), \end{gathered} \tag{4.1}$$

where n_{var} and n_{h} are the number of neurons in the input layer and in the hidden layer, respectively, $w_{ij}^{(1)}$ is the weight between input-layer neuron i and hidden-layer neuron j, and $w_{j1}^{(2)}$ is the weight between the hidden-layer neuron j and the output neuron. The function ρ is the *neuron response function* ($\mathcal{R}^n \mapsto \mathcal{R}$). The response function often it can be separate into a $\mathcal{R}^n \mapsto \mathcal{R}$ *synapse function* κ, and a $\mathcal{R} \mapsto \mathcal{R}$ *neuron activation function* α, so that $\rho = \alpha \circ \kappa$.[1]

The supervised learning consists of getting as input of the Neural Network N training events $\boldsymbol{x}_a = (x_1, \ldots, x_{n_{\text{var}}})_a$, $a = 1, \ldots, N$, where we know the desired output for every input event. For each training event a the NN output $y_{\text{ANN,a}}$ is computed and compare with the desired output $\hat{y}_a \in \{1, 0\}$ (1 for signal events and 0 for the background events). An *error function* E, measuring the agreement of the network response with the desired one, is defined by

$$E(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N | \boldsymbol{w}) = \sum_{a=1}^{N} E_a(\boldsymbol{x}_a | \boldsymbol{w}) = \sum_{a=1}^{N} \frac{1}{2}(y_{\text{ANN,a}} - \hat{y}_a)^2, \tag{4.2}$$

where $\boldsymbol{w}$ denotes the ensemble of adjustable weights in the Network. The set of weights that minimises the error function can be found using the method of *gradient descent*. Starting form a random set of weights $\boldsymbol{w}^{(\rho)}$ the weights are updated by moving a small distance in $\boldsymbol{w}$-space into the direction $-\boldsymbol{\nabla}_{\boldsymbol{w}} E$ where E decreases most rapidly

$$\boldsymbol{w}^{(\rho+1)} = \boldsymbol{w}^{(\rho)} - \eta \boldsymbol{\nabla}_{\boldsymbol{w}} E, \tag{4.3}$$

where the positive number η is the *learning rate*.

4.3 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;

[1] The synapse function used in this work is the sum of the weights $\sum_{i=1}^{n} y_i^{(l)} w_{ij}^{(l)}$, while the neuron activation function is $\tanh(x)$.

$\boldsymbol{\alpha_T(B_s^0)}-$ 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 chapter 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 4.2. These variables have been chosen because of their discriminant power (see sec. 4.4) which is satisfactory for the scope of this work. Adding other variables would not improve greatly the separation power, at the price of an 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 ($\alpha_T(B_s^0)$).

Pointing angle

As explained in the section 3.4.2, 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 (see fig. 3.6) the strategy used introduces an issue in the estimate of the pointing angle (see fig. 4.3a). 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. 4.3b), 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. 4.4 and fig. 4.5).

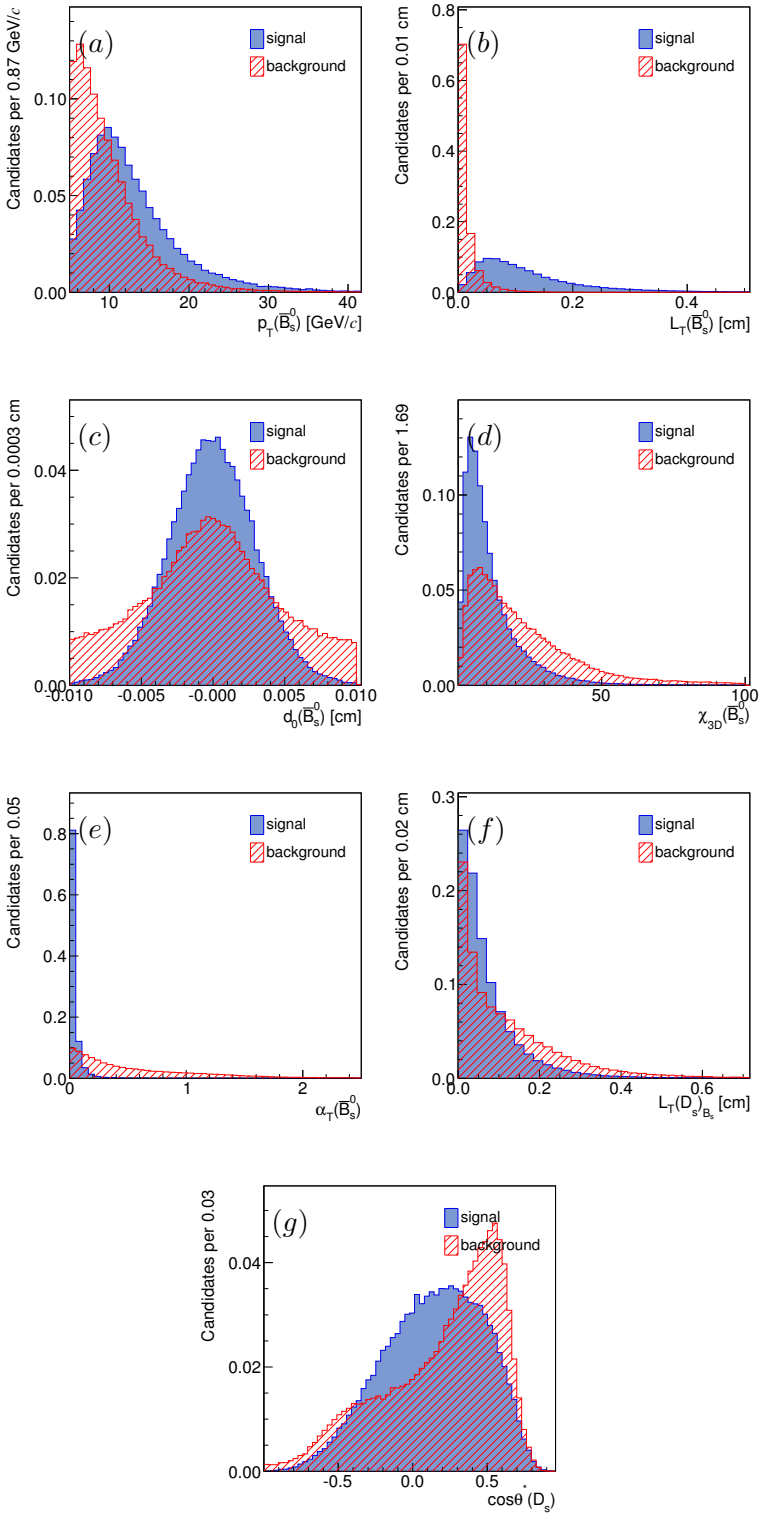


Figure 4.2 Input variables to the Neural Network: $p_T(B_s^0)$ (a), $L_T(B_s^0)$ (b), $d_0(B_s^0)$ (c), $\chi^2_{3D}(B_s^0)$ (d), α_T (e), $L_T(D_s)_{B_s^0}$ (f), $\cos\theta^*(D_s)$ (g). The signal is the filled histogram (in blue), while the background is the hatched one (in red). The histograms are normalized to one.

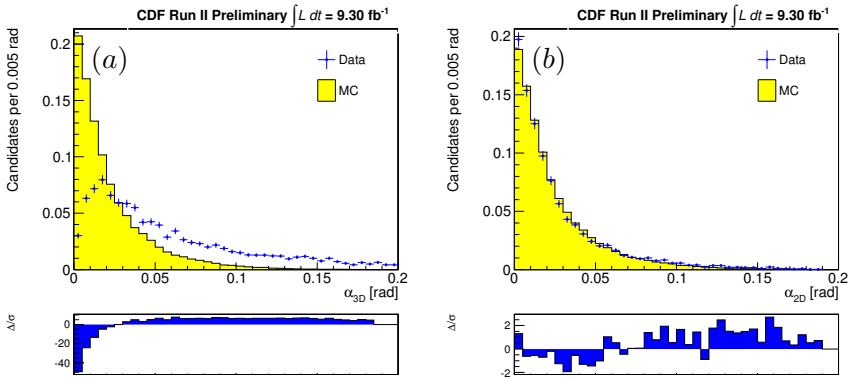


Figure 4.3 Comparison of the pointing angle distribution in data and in the simulation (where the coordinates of the primary vertex are estimated with the algorithm explained in the text). (a) The 3-dimensional pointing angle, (b) the transverse pointing angle.

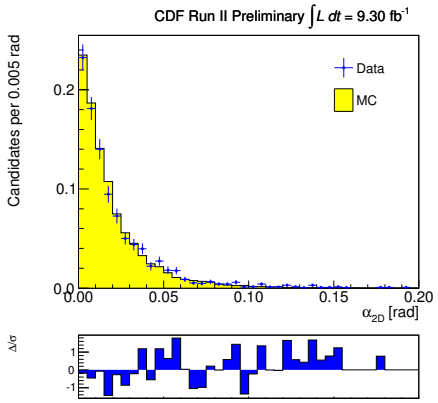


Figure 4.4 Comparison of the transverse pointing angle distribution in data and in simulation. The pointing angle is calculated for both data and simulation with the primary vertex estimate with the algorithm used in the simulation.

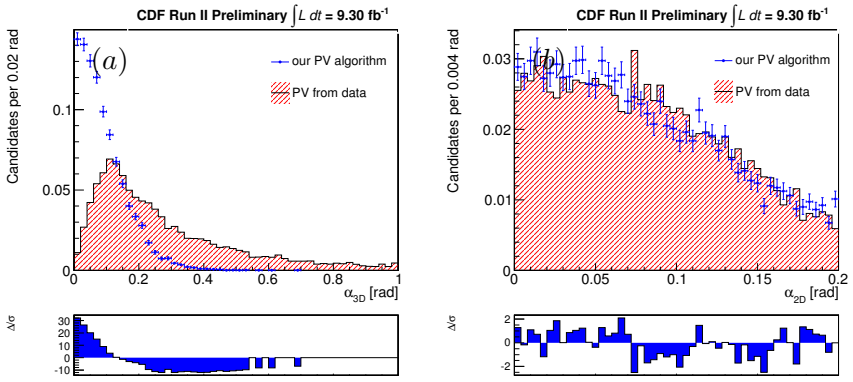


Figure 4.5 Comparison of pointing angle distribution, in the combinatorial background mass region $[5.5, 6.5]\,\mathrm{GeV}/c^2$, calculate with the algorithm introduced in Monte Carlo simulation for calculation of the primary vertex (points) and with the standard algorithm (hatched histogram). (a) 3-dimensional pointing angle, (b) transverse pointing angle.

4.3.1 Isolation of the B_s^0 candidate

In the previous iteration of this analysis, as described in ref. [2], a requirement on the isolation of the B_s^0 candidate was optimized and used in the final selection. In fact the isolation is one of the most used variables in the B-physics analyses, since it has a large discriminant power between signal and background. In the most of cases it provides information orthogonal to the standard kinematic variables. However we decide to drop it from our analysis. The improvement of the statistical uncertainty because its use in the final selection would not justify the large increasing of the level of sophistication of the fit of composition of the data sample. This would cause a larger systematic uncertainty, that would totally compensate the statistical improvement. The definition of the isolation and the summary of the studies performed on this are reported in appendix A.

4.4 Final selection

Finally we trained the Neural Network and we obtain as output the distributions shown in figure 4.6. Figure 4.6a 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 4.6 reports the correlation matrix of the variables used in the NN training, for the signal sample (see fig. 4.6c) and for the background sample (fig. 4.6d).

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

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

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 (4.4) 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 4.7 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 4.8b. For comparison in the figure 4.8a is also reported the mass distribution with the baseline selection of table 3.2. 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. 4.6. Background reduction is particularly strong in the region above the $B_s^0 \to D_s^- \pi^+$ peak, which is essentially pure combinatorial.

4.5 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 4.1 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

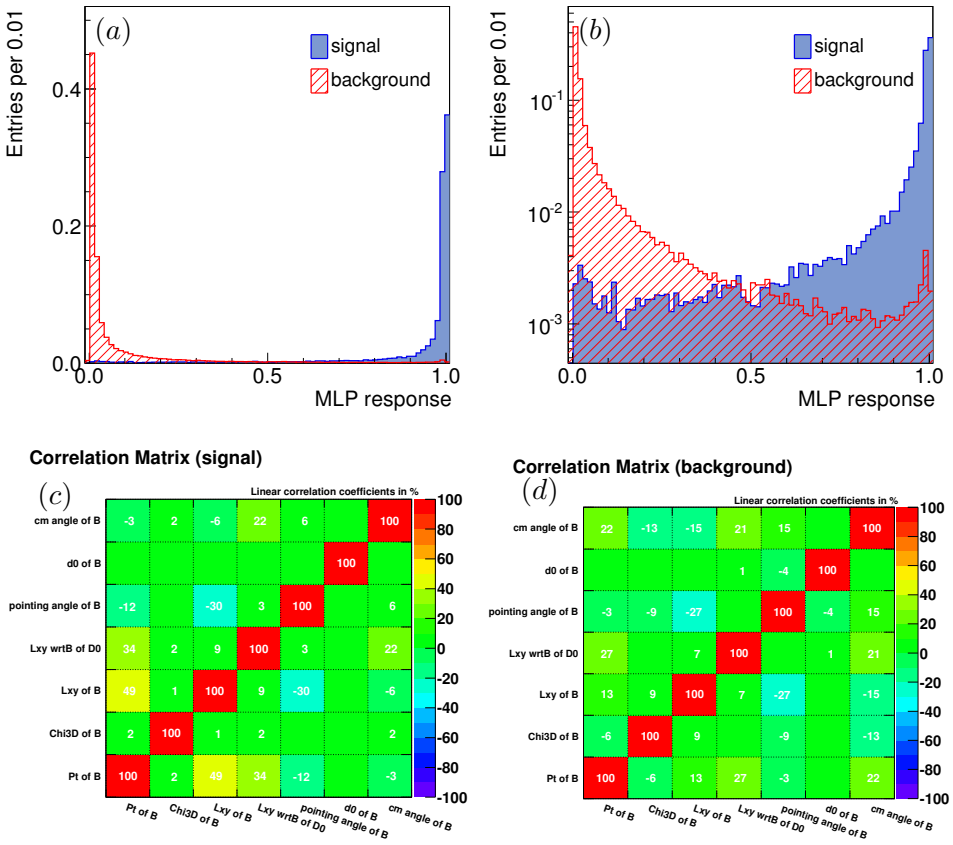


Figure 4.6 Output response of the Neural Network (a); output response of the Neural Network in logarithmic scale (b). The histograms are normalized to one. Also it is reported the linear correlation matrix for the signal (c) and for the background (d)

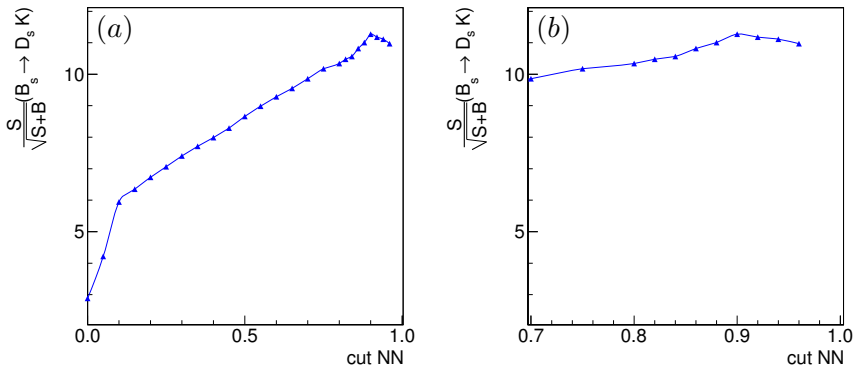


Figure 4.7 Score function as function of the Neural Network response (a), and its zoom (b).

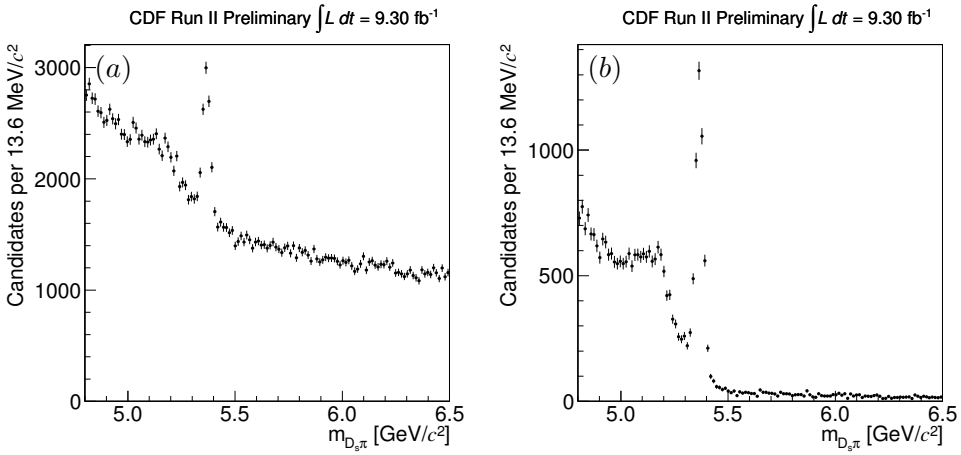


Figure 4.8 Invariant mass distribution of the $D_s\pi$ pair after the baseline selection (a), and after the final selection cut (b).

estimated with the same strategy described in section 3.3, 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 4.1 Comparison between the Monte Carlo efficiency and the efficiency of the $B_s^0 \to D_s^-\pi^+$ decay mode.

NN cut	ϵ_{MC}	ϵ_{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

This chapter is devoted to the description of the Particle Identification observable used in the fit of composition.

5.1 Particle Identification (PID) at CDF II

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 average total energy-loss per unit length of a particle (heavier than electron) of charge q, traversing a gas volume with velocity $c\beta$, is approximated by the Bethe-Bloch formula [18]

$$\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi N e^4}{m_e c^2 \beta^2} q^2 \left[\ln\left(\frac{2 m_e c^2 \beta^2 \gamma^2}{I} \right) - \beta^2 - \frac{\delta(\beta)}{2} \right], \tag{5.1}$$

where N is the electron density in the medium, m_e (e) is the electron mass (charge), I is the mean excitation energy of the medium atoms, and $\delta(\beta)$ is the correction that accounts for the density effect at high velocities. To a good approximation, the most probable dE/dx value of a charged particle is a function just of its velocity. If the momentum of the particle is measured, the mass can also be determined. In the COT, the signal induced on each sense-wired depends on the amount of ionization charge produced by the passage of the charged particle near the wire. It is measured in nanoseconds because it is encoded as the digital pulse-width between the leading and the trailing-edge time of the hit. Multiple samplings along the trajectory of the charged particle allow a more reliable estimation of dE/dx. The COT samples a maximum of 96 dE/dx measurements per track, from which a 80% truncated mean is calculated to avoid the adverse effect of long positive tails in the estimation of the average dE/dx.

The empirical equation, that better models the COT average energy-loss as a function of velocity, is the following variant of Bethe-Bloch curve:

$$\left\langle \frac{dE}{dx} \right\rangle = \frac{1}{\beta^2} \left[c_1 \ln\left(\frac{\beta\gamma}{b + \beta\gamma} \right) + c_0 \right] + a_1(\beta - 1) + a_2(\beta - 1)^2 + C, \tag{5.2}$$

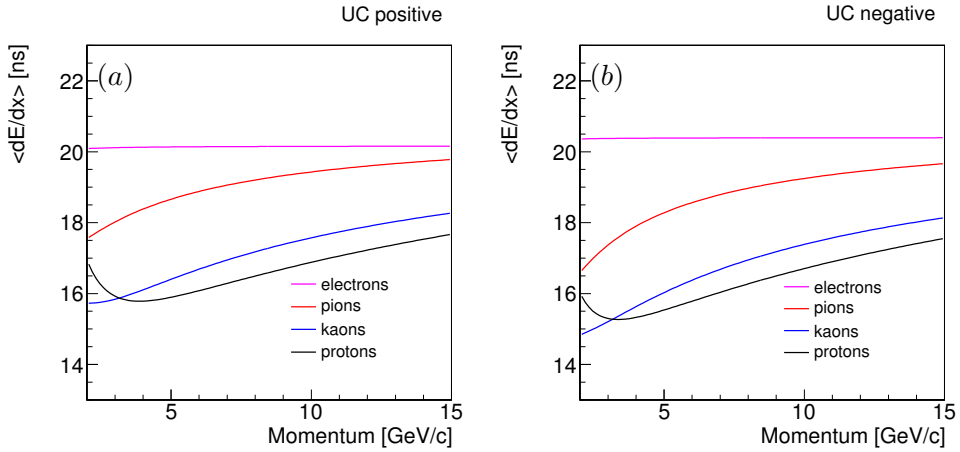


Figure 5.1 Universal curves as a function of particle momentum for positive (a) and negative (b) particle.

where a_i, b, c_j and C are parameters extracted from data. The function in the equation (5.2), has all the features that are present in the Bethe-Bloch curve (eq. (5.1)). The parameters c_0 and c_1 represent the intensities of the $1/\beta^2$ fall and of the relativistic rise. The parameter b is associated with the COT gas properties, e.g., mean excitation energy of the gas atoms, etc.. The parameters a_1 and a_2 provide a further adjustment, especially in the low $\beta\gamma$ region. Figure 5.1 shows the universal curves as a function of the momentum for pions, kaons, protons and electrons (positive charge 5.1a, negative 5.1b). Muons are indistinguishable from pions: the ≈ 1.5 ns dE/dx resolution is insufficient to resolve the difference between their ionization rates, which is inappreciable because of the small mass difference $m_{\pi^\pm} - m_{\mu^\pm} \simeq 34\,\mathrm{MeV}/c$. Electron curve is approximately a straight line at about 20 ns as shown in the figure 5.1.

The individual charge collections output by the COT are subject to several corrections to eliminate a number of detector related conditions as high-voltage correction, angle and drift distance correction, pressure correction, etc.. In addition, an accurate calibration of the uniformity of the dE/dx response in time is required. 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 [32].

5.2 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

$$\mathrm{res}_A = \left.\frac{dE}{dx}\right|_{\mathrm{obs}} - \left.\frac{dE}{dx}\right|_A , \tag{5.3}$$

where the $dE/dx|_A$ is the expected dE/dx, determined from the function (5.2) evaluated at $\beta\gamma = p/m_A$. The official CDF II parameterization of dE/dx residual

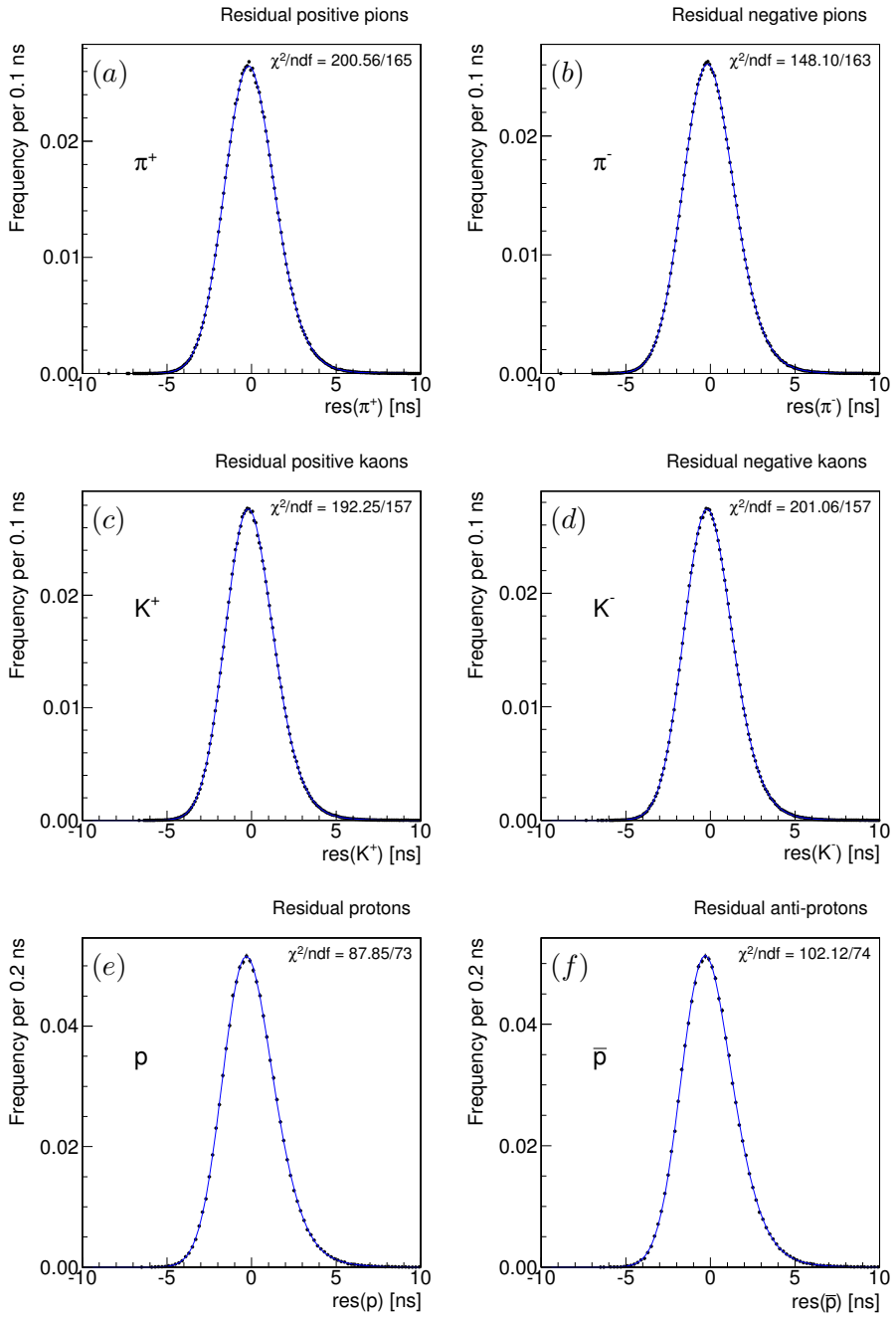


Figure 5.2 Distribution of observed dE/dx residual, for pions (with pion mass hypothesis) (a, b), for kaons (with kaon mass hypothesis) (c, d) and for protons (with proton mass hypothesis) (e, f).

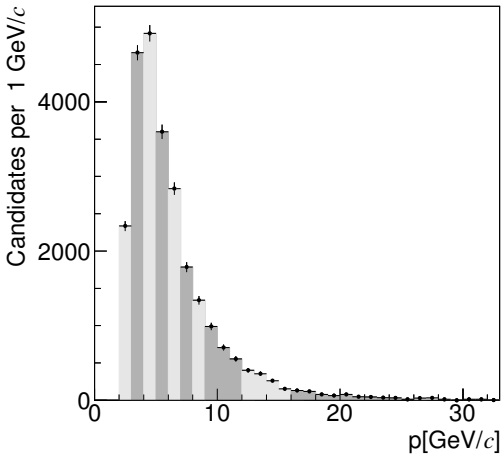


Figure 5.3 Monte Carlo momentum distribution for K from B_s^0 candidate in the decay $B_s^0 \to D_s^{\pm} K^{\mp}$. The fill region show the section used to make the template of the PID observable.

distributions with the correct mass hypothesis (see ref. [32]) is made with analytical functions (convolution of several Gaussian terms), currently available in a stand-alone `C` code. Figure 5.2 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 $\mathcal{P}_{\pi^+}(\text{res}_{\pi^+})$ the parameterization of the positive pion dE/dx residual, and with $\mathcal{P}_{K^+}(\text{res}_{K^+})$, $\mathcal{P}_p(\text{res}_p)$ 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 section.

5.3 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 chapter 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:

$$\text{res}_\pi = \left.\frac{dE}{dx}\right|_{\text{obs}} - \left.\frac{dE}{dx}\right|_{\pi}. \tag{5.4}$$

The res_π is a momentum-dependent observable, as it can be seen from figure 5.1, 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 [33]. 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 $\mathcal{P}_\pi(\text{res}_\pi)$ function, from ref. [32], as reported in figures 5.2a and 5.2b. Instead if

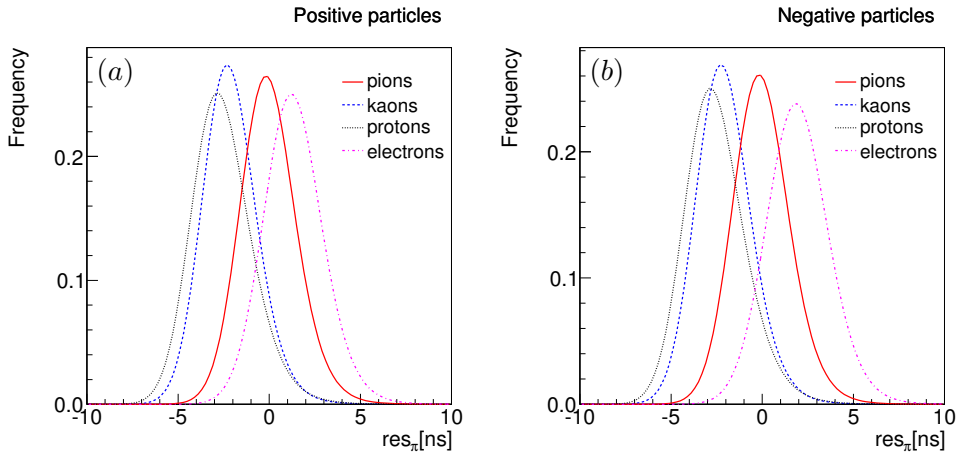


Figure 5.4 Comparison of the PID distribution for pions, kaons, protons and electrons. In (a) positive particles PID-distributions, (b) negative particles.

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

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

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. 5.1:

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

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. 5.3. We will assess a systematic uncertainty due to the binning of the momentum in the chapter 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 5.4a, 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 5.4b. Distributions of daughter particle of the other decays are similar.

6 Fit of composition

This chapter describes the fit of composition of the $B_s^0 \to D_s^{\pm} h^{\mp}$ data sample.

6.1 Discriminating observables

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;

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

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

$$\mathrm{res}_\pi = \left.\frac{dE}{dx}\right|_{\mathrm{obs}} - \left.\frac{dE}{dx}\right|_{\pi}, \tag{6.1}$$

where $dE/dx|_{\mathrm{obs}}$ indicates the observed energy-loss, while $dE/dx|_\pi$ indicates the expected energy-loss in pion mass hypothesis, as discussed in chapter 5.

6.2 Likelihood function

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

$$\mathscr{L}(\nu, \boldsymbol{\theta}) = \frac{\nu^N}{N!} e^{-\nu} \cdot \prod_{i=1}^{N} \mathscr{L}_i(\boldsymbol{\theta}|\boldsymbol{x}_i) \tag{6.2}$$

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. (6.2) 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:

$$\mathscr{L}_i = f_{\rm bkg} \cdot \mathscr{L}_i^{\rm bkg} + (1 - f_{\rm bkg}) \cdot \mathscr{L}_i^{\rm dec}. \tag{6.3}$$

The index bkg (dec) labels the part of the function that describes the combinatorial background (all the decays) term; $f_{\rm bkg}$ is the fraction of combinatorial background events and $1 - f_{\rm bkg}$ 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 $\wp^{\rm PID}$ 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:

$$\mathscr{L}^{\rm dec} = \sum_{j=1}^{s} f_j \cdot \wp_j^m(m_{D_s\pi}) \cdot \wp_j^{\rm PID}({\rm res}_\pi), \tag{6.4}$$

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,

$$f_s = \sum_{j=1}^{s-1} f_j, \tag{6.5}$$

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:

$$\mathscr{L}^{\rm bkg} = \wp^m(m_{D_s\pi}) \cdot \wp^{\rm PID}({\rm res}_\pi). \tag{6.6}$$

In equations (6.3)-(6.6) the functional dependence on the vector $\boldsymbol{\theta}$ was omitted, since in the equations we explicitly wrote some terms of this vector as $f_{\rm bkg}$ and f_j.

6.3 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 3.4. 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. 6.1), is parameterized using the following p.d.f.:

$$\begin{aligned} \wp^m(m;\boldsymbol{\alpha}) =& f_{\rm bulk}\Big[f_1 \mathscr{G}(m;\mu_1,\sigma_1) + (1-f_1)\mathscr{G}(m;\mu_2,\sigma_2)\Big] \\ &+ (1 - f_{\rm bulk})\mathscr{T}(m;b,c,\mu_1), \end{aligned} \tag{6.7}$$

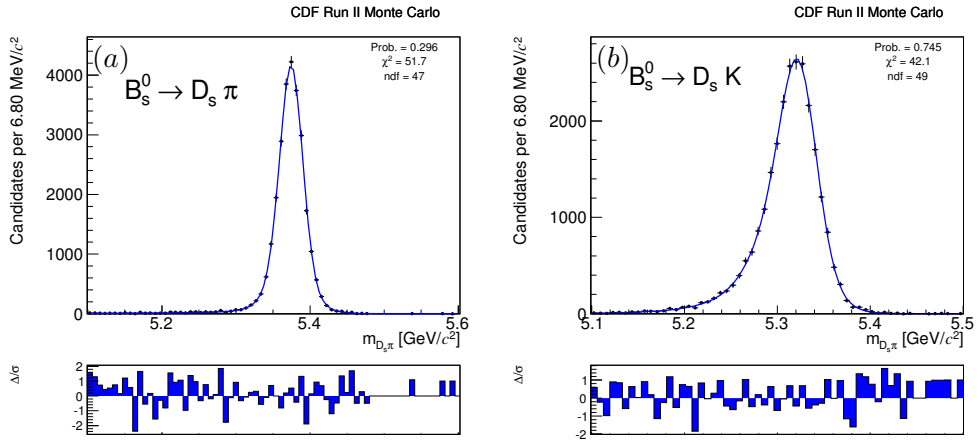


Figure 6.1 Mass p.d.f. for the $B_s^0 \to D_s^{\pm} h^{\mp}$ decays. (a) $B_s^0 \to D_s^- \pi^+$, (b) $B_s^0 \to D_s^{\pm} K^{\mp}$.

where

$$\mathscr{G}(m;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2\right),$$

$$\mathscr{T}(m;b,c,\mu) = \frac{1}{K} \exp\left(b(m-\mu)\right) \cdot \operatorname{erfc}\left(c(m-\mu)\right),$$

$$K = \int_{m_1}^{m_2} \exp\left(b(m-\mu)\right) \cdot \operatorname{erfc}\left(c(m-\mu)\right) dm,$$

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{+\infty} e^{-t^2}\, dt.$$

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.

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" [13]:

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

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

where the normalization K_A is:

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

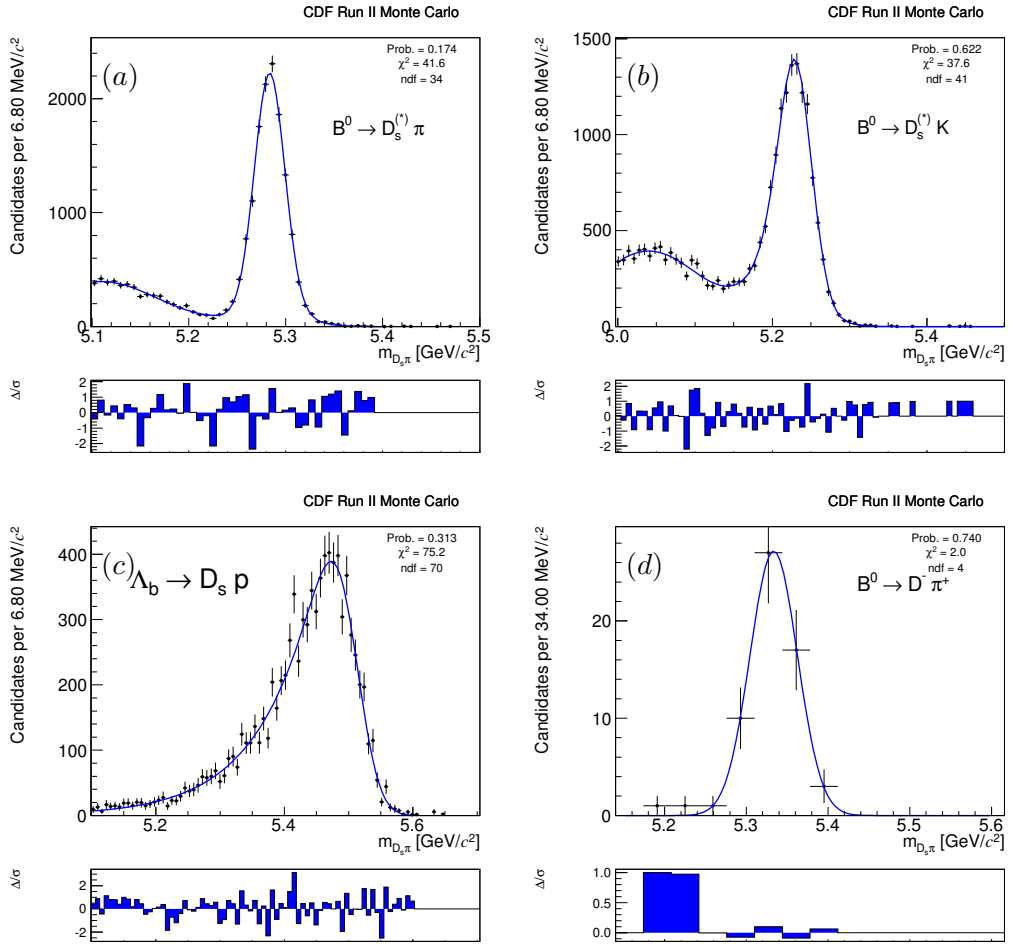


Figure 6.2 Mass template of various fully-reconstructed decay modes: $B^0 \to D_s^{(\star)+}\pi^-$ (a), $B^0 \to D_s^{(\star)-}K^+$ (b), $\Lambda_b^0 \to D_s^- p$ (c), and $B^0 \to D^-\pi^+$ (d).

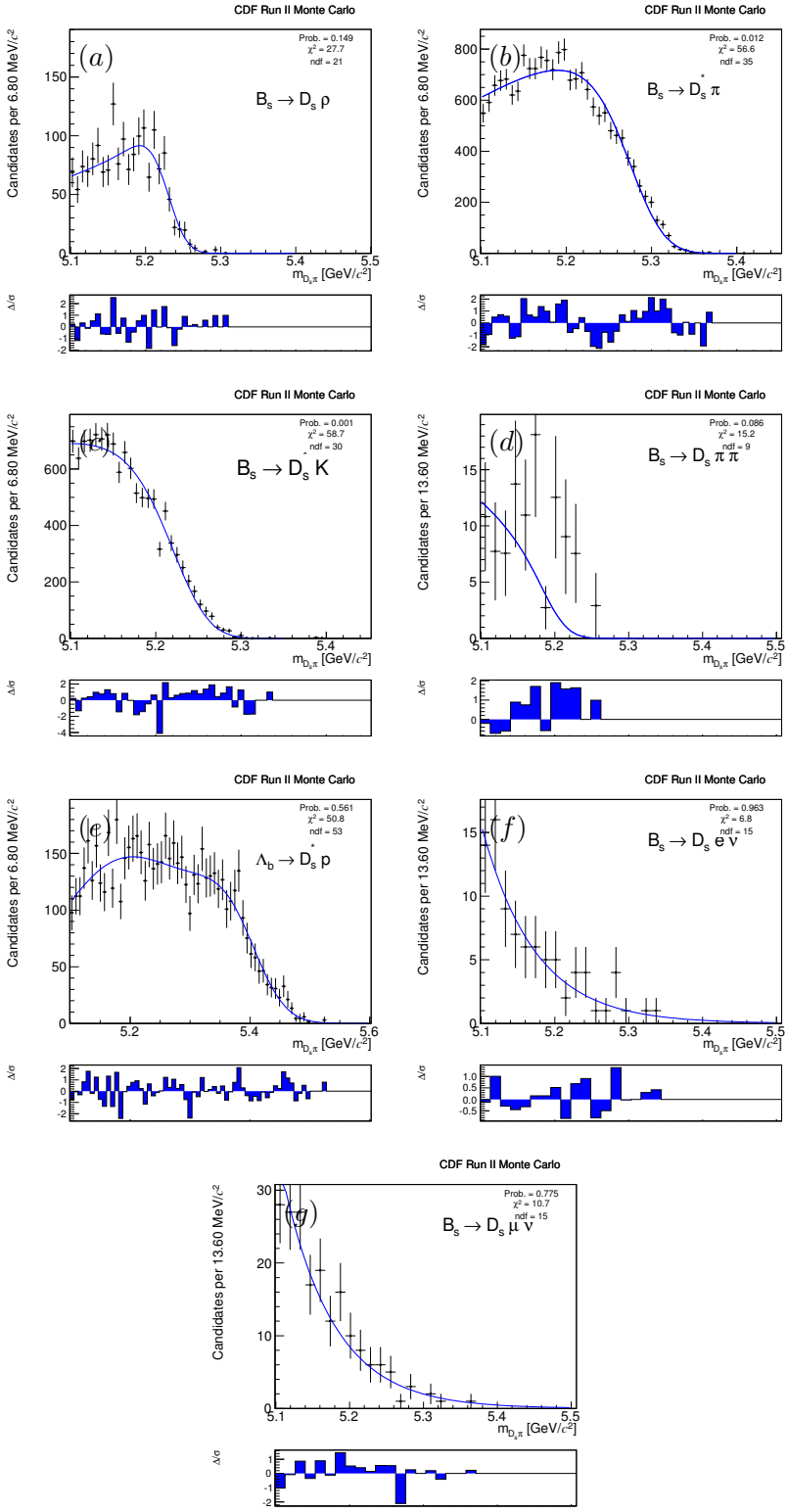


Figure 6.3 Mass template of various mis-reconstructed decay modes: $B_s^0 \to D_s^- \rho^+$ (a), $B_s^0 \to D_s^{\star -} \pi^+$ (b), $B_s^0 \to D_s^{\star \pm} K^{\mp}$ (c), $B_s^0 \to D_s^{(\star) -} \pi^+ \pi^0$ (d), $\Lambda_b^0 \to D_s^{\star -} p$ (e), $B_s^0 \to D_s^- e^+ \nu_e$ (f) and $B_s^0 \to D_s^- \mu^+ \nu_\mu$ (g).

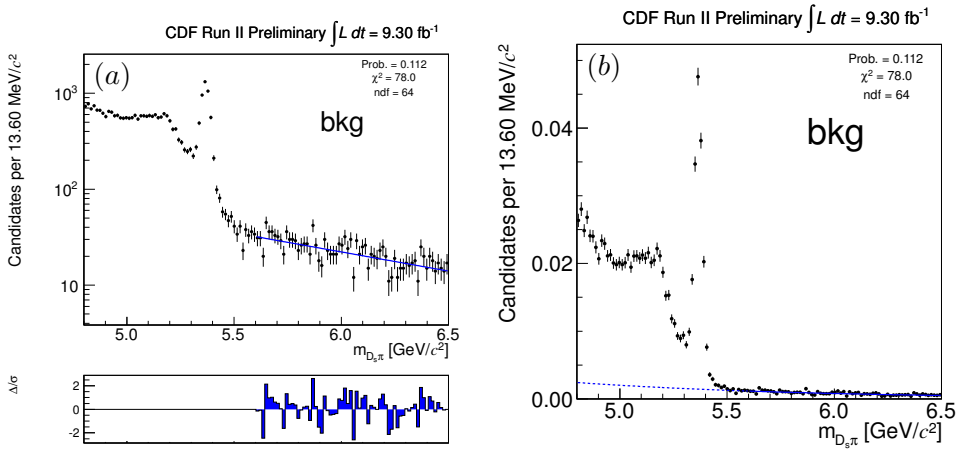


Figure 6.4 Mass template of the background term. (a) mass distribution of the pair $D_s\pi$ in log-scale. (b) mass distribution of the pair $D_s\pi$ in linear-scale, where the background p.d.f. is extrapolated in the lower mass region.

where m_A and c_A are the Argus function parameters, while σ is the resolution parameter. Some mass templates are reported in figure 6.3.

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 6.3f (g)) are parameterized with an exponential function:

$$\begin{aligned} \mathscr{E}(m;a) &= \frac{1}{K}\exp(-am), \\ K &= \int_{m_1}^{m_2} \exp(-am)\, dm. \end{aligned} \tag{6.8}$$

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 the central fit, is fitted with an exponential function (see eq. (6.8)), as shown in Figure 6.4. The value of the slope of the exponential function obtained is $a_{\mathrm{bkg}} = (0.98 \pm 0.10)\,(\mathrm{GeV}/c^2)^{-1}$, 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 chapter 7.

6.4 PID probability density function

The p.d.f.s of the dE/dx residual for the various components are extracted using the procedure illustrated in chapter 5. Here we just show p.d.f.s for kaons, protons and electrons for some decays (see figure 6.5). The p.d.f.s of the other decays are similar. We assume that the background is mostly composed by pions and kaons.

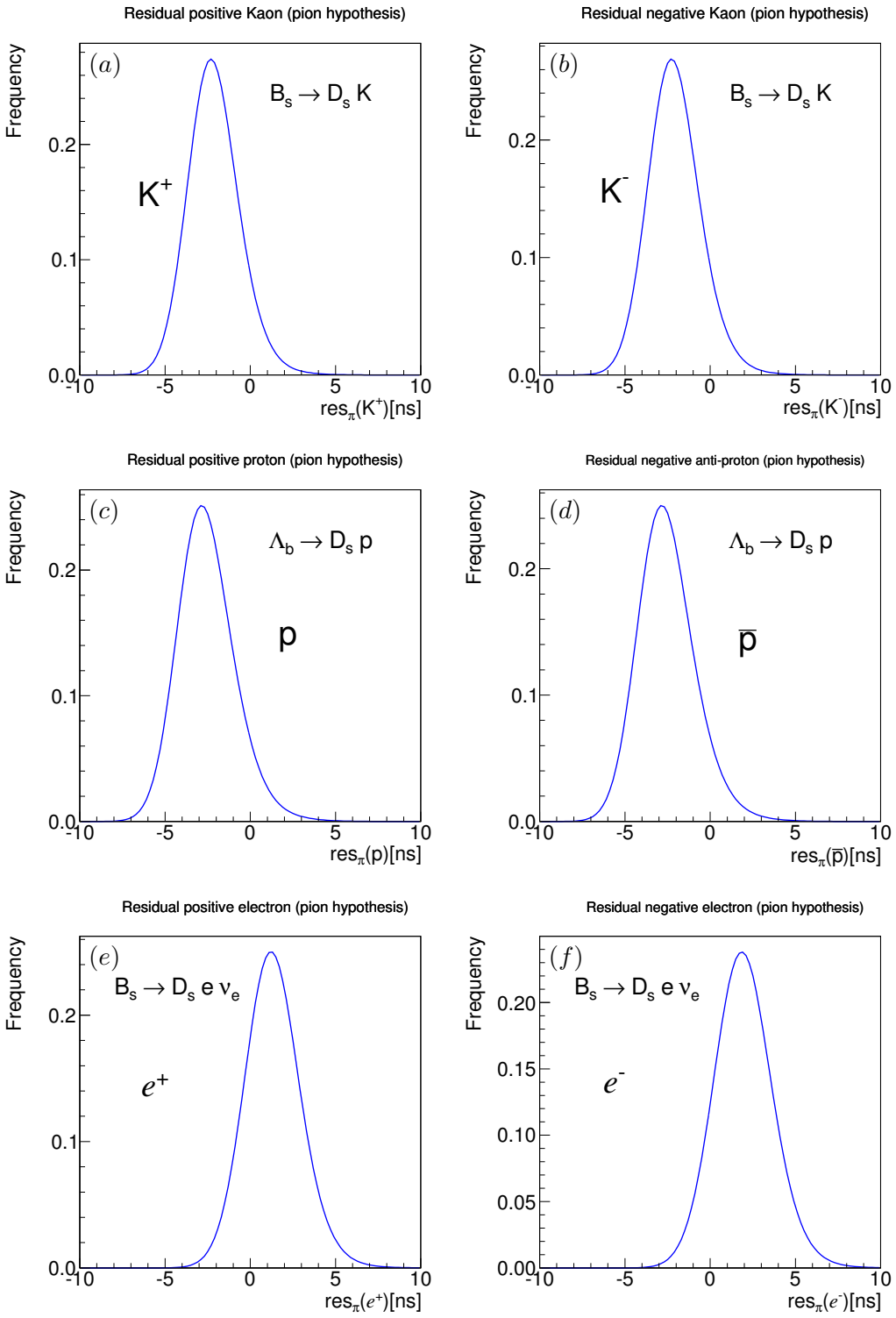


Figure 6.5 PID templates of the daughter tracks of some decays. (a, b) kaons from $B_s^0 \to D_s^\pm K^\mp$, (c, d) protons from $\Lambda_b^0 \to D_s^- p$, (e, f) electrons from $B_s^0 \to D_s^{(\star)-} e^+ \nu_e$.

Therefore the p.d.f. of the background can be written as:

$$\wp^{\mathrm{PID}}_{\mathrm{bkg}}(\mathrm{res}_{\pi^+}) = f^{\pi}_{\mathrm{bkg}} \cdot \wp_{\pi^+}(\mathrm{res}_{\pi^+}) + (1 - f^{\pi}_{\mathrm{bkg}}) \cdot \wp_{K^+}(\mathrm{res}_{\pi^+}), \quad (6.9)$$

$$\wp^{\mathrm{PID}}_{\mathrm{bkg}}(\mathrm{res}_{\pi^-}) = f^{\pi}_{\mathrm{bkg}} \cdot \wp_{\pi^-}(\mathrm{res}_{\pi^-}) + (1 - f^{\pi}_{\mathrm{bkg}}) \cdot \wp_{K^-}(\mathrm{res}_{\pi^-}), \quad (6.10)$$

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

6.5 Fit implementation

The fit of composition is performed on the $D_s h$ data sample. We use only candidates whose discriminating observables satisfy the following conditions: the invariant $D_s\pi$ mass within $[5.1, 5.6]\,\mathrm{GeV}/c^2$, and $|\mathrm{res}_\pi| < 6$. The requirement $|\mathrm{res}_\pi| > 6$ 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 $N = 12453$.

6.5.1 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 $B^0_s \to D^-_s\pi^+$ decay mode is known, the $\lambda_{B\to DX}$ parameter can be written as:

$$\lambda_{B\to DX} = \frac{f_{B\to DX}}{f_{B^0_s\to D^-_s\pi^+}}. \quad (6.11)$$

$f_{B\to DX}$ and $f_{B^0_s\to D^-_s\pi^+}$ 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 6.7 for the definition). In fact we obtain that

$$\frac{f_{B\to DX}}{f_{B^0_s\to D^-_s\pi^+}} = \frac{\mathcal{B}(B \to DX)}{\mathcal{B}(B^0_s \to D^-_s\pi^+)} \cdot \frac{f_q}{f_s} \cdot \frac{\epsilon(B \to DX)}{\epsilon(B^0_s \to D^-_s\pi^+)} \cdot \frac{\mathcal{B}(D \to Y)}{\mathcal{B}(D_s \to \phi\pi)}, \quad (6.12)$$

where $\epsilon(B^0_s \to D^-_s\pi^+)$ and $\epsilon(B \to DX)$ are respectively the reconstruction efficiencies for the $B^0_s \to D^-_s\pi^+$ 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 (6.12). All input branching fractions, f_s, f_q come from PDG [18], while efficiency corrections from CDF simulation. Table 6.1 reports the constrained parameters in the fit, the known values of branching fractions ratios and the applied constraints (third column).

Table 6.1 Table of the parameter constrained in the fit of composition and their values.

parameter (λ_i)	branching fractions [18]	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

6.5.2 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 6.2 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. (6.7) 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 chapter 7.

Table 6.2 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
$\sigma\,[\mathrm{GeV}/c^2]$	$(2.001 \pm 0.036) \times 10^{-2}$	$(1.7861 \pm 0.0015) \times 10^{-2}$

6.5.3 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 6.4), 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 6.2 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.

6.6 Fit results

The fit of composition was performed by minimizing the quantity $-2\ln(\mathscr{L})$, as defined in equation (6.3), using the MINUIT numerical minimization package [30]. Table 6.3 reports the results, while the corresponding correlation matrix is discussed in subsection 6.6.1. Table 6.4 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 6.5.2 (see table 6.2). 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 [3, 9]. Table 6.3 also reports the uncorrected value of the ratio of branching fractions $f(B_s^0 \to D_s^{\pm} K^{\mp}) / f(B_s^0 \to D_s^- \pi^+)$.

6.6.1 Correlation matrix

The correlation matrix corresponding to the fit of composition is shown in table 6.5. 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

Table 6.3 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_{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 $f_{B_s^0 \to D_s^- e^+ \nu_e}/\epsilon_{\text{rel}}$	-
$f_{B_s^0 \to D_s^- \pi^+ \pi^0}$	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_{bkg}^{π}	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 6.4 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

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 6.5 Correlation matrix returned by the fit of composition. The legend for the fit parameters is reported in the third column of table 6.3.

PARAMETER NO.	CORRELATION COEFFICIENTS GLOBAL	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

6.6.2 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:

$$\wp(x_i, \bar{\boldsymbol{\theta}}) = \int \wp(\boldsymbol{x}, \bar{\boldsymbol{\theta}})\, dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n, \tag{6.13}$$

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 6.6-6.7. 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 6.8). The agreement between data and fit projections is satisfactory.

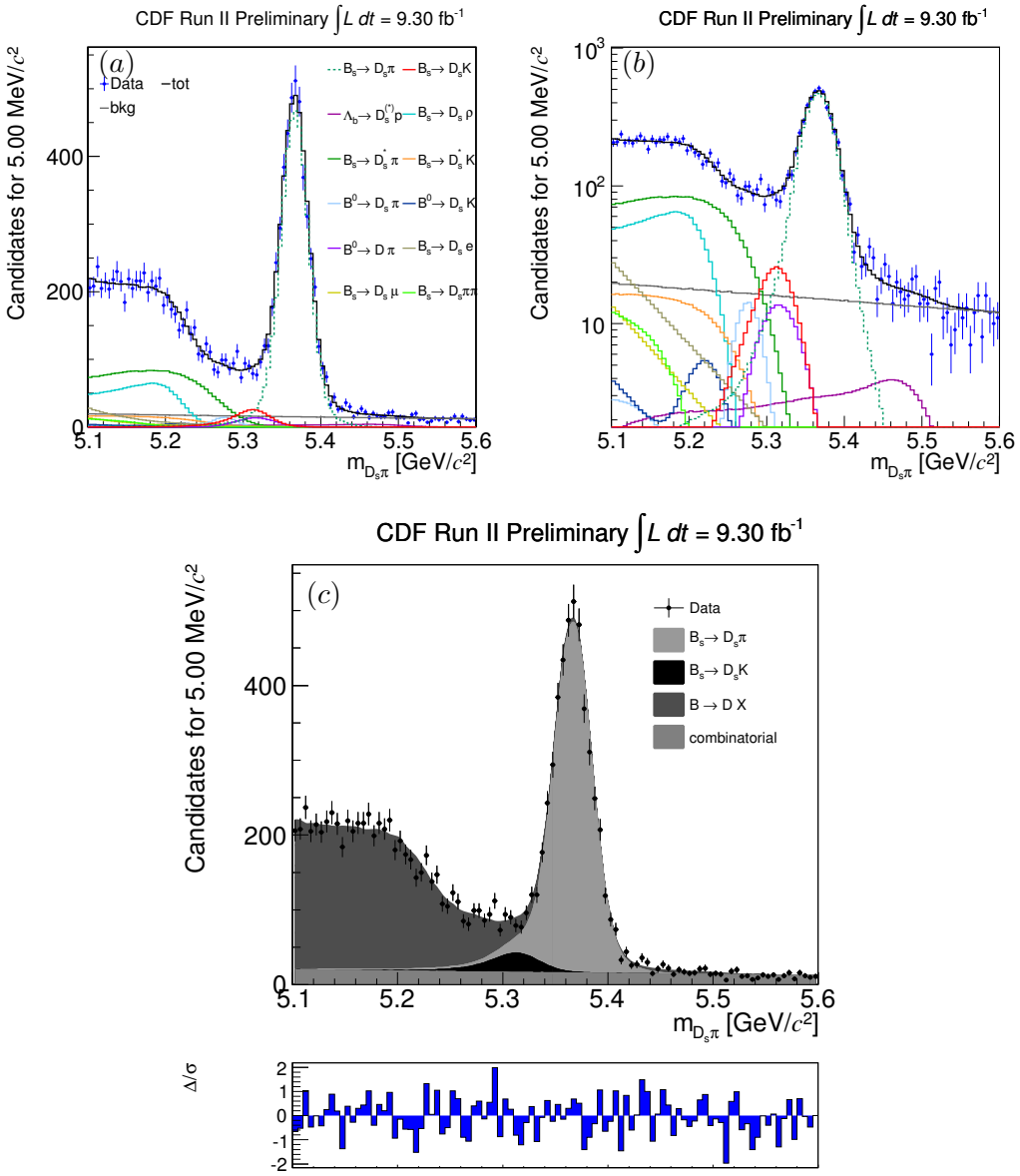


Figure 6.6 $m_{D_s\pi}$ distribution with the fit projection overlaid: (a) the fit projection with all the fit components (logarithmic scale (b)), (c) the components are grouped for clarity. The residual plot at the bottom of the figure (c) shows the number of σ discrepancy (data minus projection).

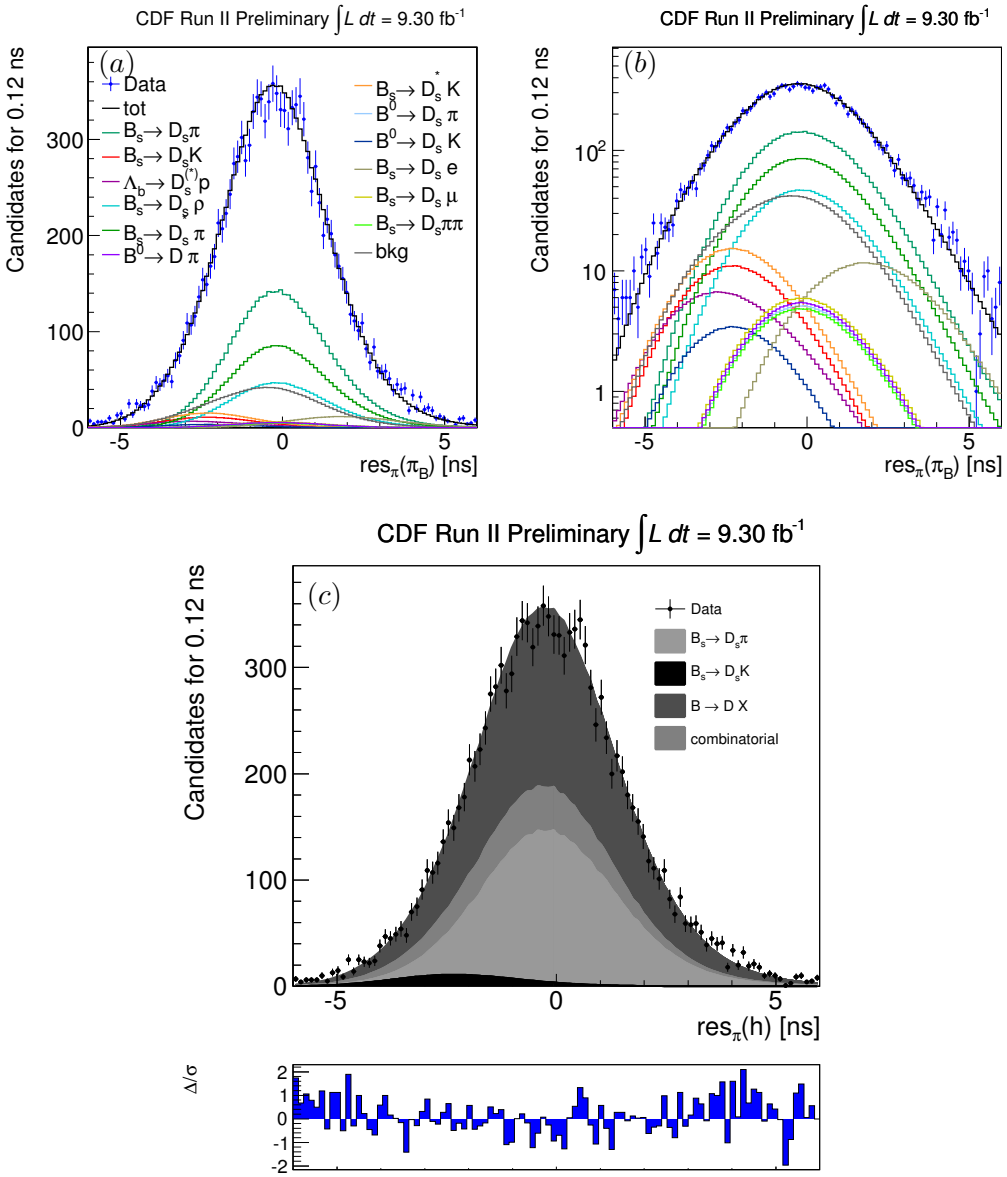


Figure 6.7 res_π distribution with the fit projection overlaid: (a) the fit projection with all the fit components (logarithmic scale (b)), (c) the components are grouped for clarity. The residual plot at the bottom of the figure (c) shows the number of σ discrepancy (data minus projection).

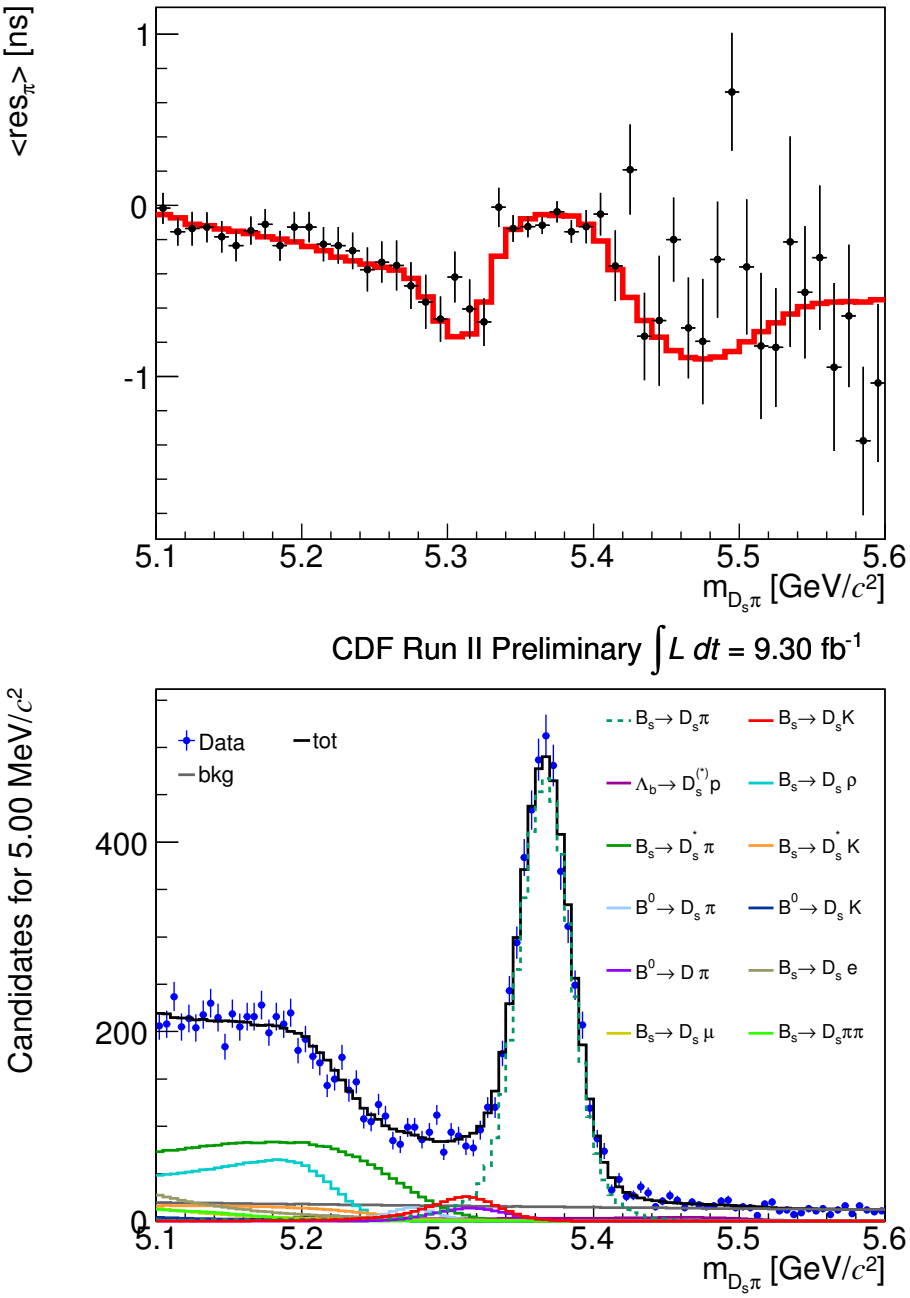


Figure 6.8 Fit projection onto the variable $\langle \mathrm{res}_\pi \rangle$ as function of the $m_{D_s\pi}$ (top figure).

6.7 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}):

$$\epsilon = \frac{N_{\text{passing}}}{N_{\text{produced}}}. \tag{6.14}$$

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

$$\frac{\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = \frac{f_{B_s^0 \to D_s^{\pm} K^{\mp}}}{f_{B_s^0 \to D_s^- \pi^+}} \cdot \frac{\epsilon(B_s^0 \to D_s^- \pi^+)}{\epsilon(B_s^0 \to D_s^{\pm} K^{\mp})}, \tag{6.15}$$

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

$$\frac{\epsilon(B_s^0 \to D_s^- \pi^+)}{\epsilon(B_s^0 \to D_s^{\pm} K^{\mp})} = 1.044 \pm 0.007, \tag{6.16}$$

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.

6.8 Corrected result

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

$$\frac{f_{B_s^0 \to D_s^{\pm} K^{\mp}}}{f_{B_s^0 \to D_s^- \pi^+}} = 0.0744 \pm 0.0076 \tag{6.17}$$

and the efficiency ratio returned from CDF simulation

$$\frac{\epsilon(B_s^0 \to D_s^- \pi^+)}{\epsilon(B_s^0 \to D_s^{\pm} K^{\mp})} = 1.044 \pm 0.007, \tag{6.18}$$

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

$$\frac{\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = 0.0777 \pm 0.0079\,(\text{stat}), \tag{6.19}$$

where the uncertainty is only statistical.

7 Systematic uncertainties

This chapter describes the main sources of systematic uncertainty for the measurement of the ratio of branching fractions $\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})/\mathcal{B}(B_s^0 \to D_s^- \pi^+)$.

7.1 Evaluation of systematic uncertainties

The measurement described in this thesis 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.

7.2 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 [18]: $m_{B^0} = (5279.58 \pm 0.17)\,\mathrm{MeV}/c^2$, $m_{B_s^0} = (5366.77 \pm 0.24)\,\mathrm{MeV}/c^2$ and $m_{B^0} = (5619.4 \pm 0.7)\,\mathrm{MeV}/c^2$. 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 6.5.2, 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.

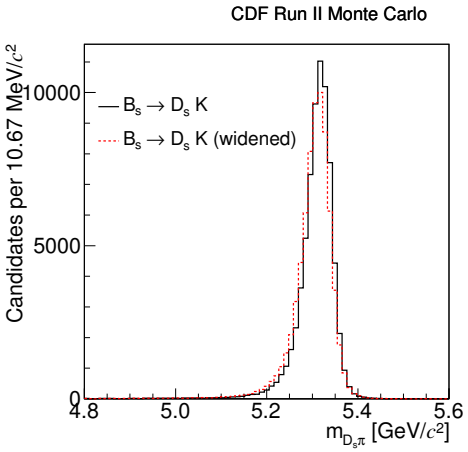


Figure 7.1 Comparison of the invariant mass distribution of the $B_s^0 \to D_s^{\pm} K^{\mp}$ before and after the transformation in eq. (7.1).

7.3 Uncertainty on mass resolution (mass resolution)

As explained in subsection 6.5.2, 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:

$$m^i_{D_s\pi} \to \frac{m^i_{D_s\pi} - \mu(p^i_{D_s}, p^i_{\pi})}{s} + \mu(p^i_{D_s}, p^i_{\pi}), \tag{7.1}$$

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 7.1 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. (7.1). 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 7.1; 2) the mass distribution of the $B^0 \to D_s^- \pi^+$ is

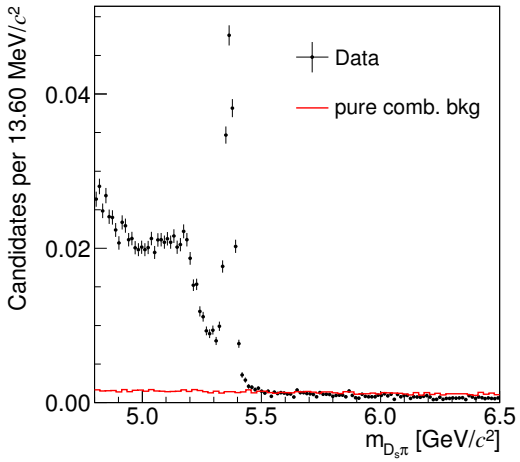


Figure 7.2 Comparison between the pure combinatorial background sample and the data-background.

scaled as the $B_s^0 \to D_s^- \pi^+$, 3) no scaling is applied to the mass distribution of the 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.

7.4 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 $a_{\text{bkg}} = (0.98 \pm 0.10)$ $(\text{GeV}/c^2)^{-1}$ is extracted from the higher mass side-band (see section 6.3)). 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^{th} event with an independent pion, that is the pion associated to the B_s^0 candidate in the $(i+1)^{\text{th}}$ event. The resulting invariant mass is reported in figure 7.2 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 $a_{\text{pure comb.}} = (0.27 \pm 0.05)\,(\text{GeV}/c^2)^{-1}$, with respect to the slope extracted from real data. Since this artificial background is very realistic, we repeat the fit of composition using $a_{\text{pure comb.}}$ 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).

7.5 dE/dx related systematic

Chapter 5 and section 6.4 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. [32]), 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. [3]), 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 chapter 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.

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

The relative efficiency ratio (see sec. 6.7) used to convert the ratio of event yields in ratio of branching fractions, is determined with $\mathcal{O}(0.6\%)$ statistical uncertainty (see eq. (6.16)). 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.

Table 7.1 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
$\boldsymbol{dE/dx}$	0.0005
binning $\boldsymbol{dE/dx}$ **mom.**	0.0009
MC stat.	0.0005
Total	**0.0026**

7.7 Total systematic uncertainties

All systematic uncertainties are summarized in the table 7.1. The total systematic uncertainty on the measurement is determined as the sum in quadrature of the individual systematic uncertainties.

8 Results and conclusions

This chapter presents the final result of this thesis and a discussion of its impact in the current experimental and theoretical picture.

8.1 Final results

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

$$\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = 0.0777 \pm 0.0079\,(\text{stat}) \pm 0.0026\,(\text{sys}), \tag{8.1}$$

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

$$\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)} = \begin{cases} 0.097 \pm 0.018\,(\text{stat}) \pm 0.009\,(\text{sys}) & \text{CDF (2008) [2]}, \\ 0.065^{+0.035}_{-0.029}\,(\text{stat}) & \text{Belle (2008) [31]}, \\ 0.0646 \pm 0.0043\,(\text{stat}) \pm 0.0025\,(\text{sys}) & \text{LHCb (2012) [1]}. \end{cases}$$

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[1]. 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. [22]:

$$\left.\frac{\mathcal{B}(B_s^0 \to D_s^\pm K^\mp)}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)}\right|_{\text{theory}, SU(3)} = 0.0864^{+0.0087}_{-0.0072}, \tag{8.2}$$

[1] 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.

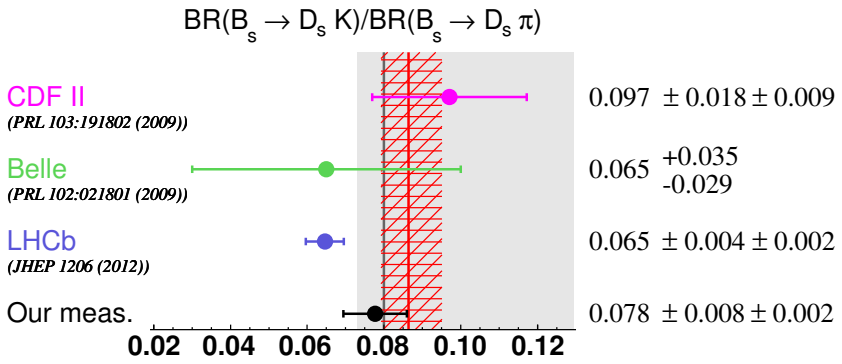


Figure 8.1 Current knowledge of the ratio of branching fractions $\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})/\mathcal{B}(B_s^0 \to D_s^- \pi^+)$, including this thesis measurement. The hatched region is the uncertainty of the-oretical predicted value (central line in the hatched region) assuming the $SU(3)$ flavor symmetry. Filled region represents the permitted region theoretical predicted.

where the $SU(3)$ flavor symmetry was assumed. Without the $SU(3)$ assumption, the estimated lower bound on the ratio of branching fractions is [22]

$$\left.\frac{\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp})}{\mathcal{B}(B_s^0 \to D_s^- \pi^+)}\right|_{\text{theory}} \geq 0.080 \pm 0.007, \tag{8.3}$$

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 8.1 summarize experimental values obtained for the ratio of branching fractions (including our result). Also it reports the allowed region and the theoretical expectation.

8.1.1 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 [18] we can extracted the following absolute branching fraction:

$$\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp}) = (2.49 \pm 0.25\,(\text{stat}) \pm 0.08\,(\text{syst}) \pm 0.31\,(\text{br})) \times 10^{-4}, \tag{8.4}$$

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 [18]. However LHCb, very recently, measured $\mathcal{B}(B_s^0 \to D_s^- \pi^+) = (2.95 \pm 0.05 \pm 0.17^{+0.18}_{-0.22}) \times 10^{-3}$ [1] (the third uncertainty is the uncertainty from the f_s/f_d measurement), which is more precise than PDG 2012 [18]. Thus if we use the LHCb measurement as input, instead of PDG 2012, we obtain the following absolute branching fraction

$$\mathcal{B}(B_s^0 \to D_s^{\pm} K^{\mp}) = (2.29 \pm 0.23\,(\text{stat}) \pm 0.08\,(\text{syst}) \pm 0.21\,(\text{br})) \times 10^{-4}, \tag{8.5}$$

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 [1].

8.2 Final remarks and conclusions

With respect to the previous CDF measurement [2] the improvement of statistical and systematic uncertainty is remarkable. If we do not consider the increasing of statistics of the CDF data sample (the data sample analyzed in this thesis is larger by a factor 3.3 than the previous iteration) we get an improvement of the statistical resolution of about 40%. This is completely due to the analysis improvements studied and developed in this thesis:

- the new selection, which plays a key role, it was optimized through an Artificial Neural Network to optimally use the multidimensional information of the input variables with the aim to minimize the uncertainty on the measurement of the branching fraction. Neural Network input variables were chosen to exploit the largest as possible amount of information from a limited and poor correlated physics observables, to get a reliable response from the simulation;

- a huge amount of full CDF simulation was produced to accurately study the features of signal mode and of mis-reconstructed b-hadron decays which lay down in the same mass region of the $B_s^0 \to D_s^{\pm} K^{\mp}$. The abundant fully reconstructed $B_s^0 \to D_s^- \pi^+$ decay mode was used to adjust the observed discrepancies between real data and simulation;

- the PID response for different particles provided by the official CDF tool was specifically adapted for this analysis to correctly account for the different momentum dependence of the particles in the final state. For each decay mode and background a different PID template was extracted;

- all information from invariant mass and PID was combined in a multidimensional extended unbinned likelihood fit to disentangle all components and then to measure the relative fraction of $B_s^0 \to D_s^{\pm} K^{\mp}$ decay mode.

As a consequence of a more sophisticated analysis we also drastically reduced the systematic uncertainty, which is improved by a factor 3.5. This is mainly due to a deeper knowledge of data gained from the studies performed in this thesis, exploiting all the experience developed in CDF during the last ten years of data taking, which has produced several world's best measurements in this field.

In conclusion, the measurement described in this thesis, in addition to its intrinsic value, is a necessary milestone on the road to the measurement of γ angle. In fact, even a more complex tagged time-dependent analysis requires a powerful and efficient selection as a starting point, an excellent understanding of the mass spectrum and consequently of the backgrounds, an excellent knowledge of PID response from the experimental apparatus. All these experimental techniques were perfected in this thesis. Although a full tagged time-depended analysis in CDF seems not doable because of the limited statistics, the work done has a general worthiness, and may be very useful in the next generation experiments, as LHCb, with similar condition as CDF.

A Isolation cut

One of the most powerful variables to improve the purity in the off-line selection, in B physics analyses, is the "isolation". Given the hard fragmentation, b-hadrons tend to carry a larger fraction of the transverse momentum than the particles produced in such a process. Thus the "isolation" is an estimator of the fraction of momentum, available from the b-quark fragmentation, carried by the b-hadron candidate:

$$I|_{R=1}(B) = \frac{p_T(B)}{p_T(B) + \sum\limits_{i \neq j : B \to j}^{R} p_T(i)}, \tag{A.1}$$

where the sum in the right-hand term of the denominator runs over all tracks (other than those of the b-hadron candidate decay-chain) satisfying standard quality requirements and found in a local region around the flight direction of the candidate. Such a region is a cone in the $(\eta - \phi)$ space, unitary in radius ($R = \sqrt{\phi^2 + \eta^2} = 1$), whose apex is the primary vertex and the axis is collinear with $p(B)$. Candidates with large values of the isolation are more likely to be a real b-hadrons than candidates with a low isolation.

However the introduction of the isolation adds further complexity to the analyses: its distribution depends on the mechanism of hadronization of the b-quark, which is not described by the signal-only simulation discussed in chapter 3. It depends on the multiplicity and on the momenta of the charged-particles not belonging to the b-hadron decay-chain, and on the transverse momentum of the b-hadron. In other words it depends on the production mechanism of the b-hadrons, but it is independent of the decay mode. We therefore expect different isolation distributions for the B^0,B_s^0, B^+ and Λ_b^0 hadrons, which are produced through different fragmentation processes.

From the isolation definition it results that we can heavily affect our analysis, if a requirement is applied on it. The reconstruction efficiency for the different signals and physics backgrounds depends on the isolation requirement, therefore all the Gaussian constraints of the fit of composition on the mis-reconstructed decays would need an accurate knowledge of this relative efficiency to be correctly applied. In addition a requirement on the isolation would sculpt the mass line shape of all mis-reconstrucetd decays, ($B_s^0 \to D_s^{\mp} K^{\pm}$ decays are included in this category) in a strange way, since the isolation is strongly momentum dependent. All these effects must be studied using real data, since typical $p\bar{p}$ collision simulators, as PYTHIA, cannot reproduce reliably the details of the b-quark fragmentation.

In supporting to our statements above, we studied the effect of a standard isolation requirement ($I > 0.5$) to our final data sample. If we define the relative

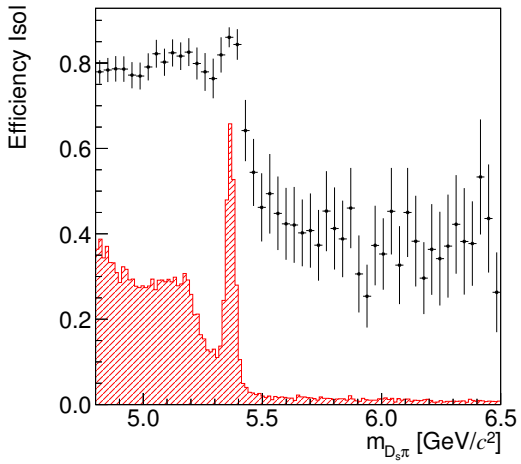


Figure A.1 Isolation efficiency as function of the invariant $D_s\pi$-mass. The hatched region is the distribution of the invariant $D_s\pi$-mass reported for clarity.

efficiency of the isolation requirement as:

$$\epsilon(\mathrm{I}) = \frac{N^{\mathrm{NN+Isol}}_{\mathrm{passing}}}{N^{\mathrm{NN}}_{\mathrm{passing}}},$$

where $N^{\mathrm{NN}}_{\mathrm{passing}}$ is the number of the events passing the NN requirement (see chapter 4) and $N^{\mathrm{NN+Isol}}_{\mathrm{passing}}$ is the number of events passing the NN and the isolation requirements, we can plot this quantity as a function of the invariant $m_{D_s\pi}$-mass, as reported in figure A.1. Some structures are clearly visible for mass value less than 5.4 GeV/c^2. In particular we have an hole in the $B^0_s \to D^{\mp}_s K^{\pm}$ region, where several contributions overlap. Therefore we decided to do not use any isolation requirement in the final version of the analysis.

Bibliography

[1] 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. 11, 79, 80

[2] 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. 11, 33, 48, 79, 81

[3] 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. 65, 77

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

[5] Aaltonen, T. et al. Precise measurement of the W-boson mass with the CDF II detector. *Phys.Rev.Lett.*, 108:151803, 2012. 12

[6] Abe, F. et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys.Rev.Lett.*, 74:2626–2631, 1995. 12

[7] Abulencia, A. et al. Measurement of the $B_s^0 - \bar{B}_s^0$ Oscillation Frequency. *Phys.Rev.Lett.*, 97:062003, 2006. 9

[8] Abulencia, A. et al. Observation of $B_s^0 - \bar{B}_s^0$ Oscillations. *Phys.Rev.Lett.*, 97:242003, 2006. 9, 12

[9] 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. 65

[10] Acosta, D. et al. The performance of the CDF luminosity monitor. *Nucl. Instrum. Meth.*, A494:57, 2002. 22

[11] Affolder, A. et al. Status report of the intermediate silicon layers detector at CDFII. *Nucl. Instrum. Meth.*, A485:6–9, 2002. 17

[12] Affolder, T. et al. CDF Central Outer Tracker. *Nucl. Instrum. Meth.*, A526:249, 2004. 18

[13] Albrecht, H. et al. Search for hadronic $b \to u$ decays. *Phys.Lett.*, B241:278–282, 1990. 58

[14] Ashmanskas, B. et al. The CDF Silicon Vertex Trigger. *Nucl. Instrum. Meth.*, A518:532, 2004. 24

[15] 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. 10

[16] Balka, L. et al. The CDF central electromagnetic calorimeter. *Nucl. Instrum. Meth.*, A267:272, 1988. 20

[17] Belforte, S. et al. Silicon Vertex Trigger Technical Design Report. *CDF Internal Note 3108*, (unpublished), 1995. 28

[18] Beringer, J. et al. Particle Data Group. *Phys. Rev.*, D86:010001, 2012. 5, 9, 32, 51, 63, 64, 74, 80

[19] Blair, R. et al. The CDF-II detector: Technical design report. *FERMILAB-Pub*, 96/390-E, 1996. 12

[20] Bona, M. et al. UTfit Collaboration. *JHEP*, 0507:28, 2005, update results and plots available at http://www.utfit.org/UTfit. 7, 10, 11

[21] Charles, J. et al. CKMfitter Group. *Eur. Phys. J.*, C41:1–131, 2005, update results and plots available at http://ckmfitter.in2p3.fr. 6, 7, 10, 11

[22] 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. 79, 80

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

[24] 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.

[25] Fleischer, R. et al. In Pursuit of New Physics with $B_s^0 \to K^+K^-$. *Eur. Phys. J.*, C71:1532, 2011. 10

[26] 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. 10

[27] Gronau, M. et al. On determining a weak phase from charged B decay asymmetries. *Phys. Lett. B*, 265(1–2):172 – 176, 1991. 10

[28] Hill, C. S. Initial experience with the CDF layer ØØ silicon detector. *Nucl. Instrum. Meth.*, A511:118–120, 2003. 16

[29] Hoecker, A. et al. TMVA: Toolkit for Multivariate Data Analysis. *PoS*, ACAT:040, 2007. 43

[30] James F. et al. 'Minuit' A system for function minimization and analysis of the parameter errors and correlations. *Comput. Phys. Commun.*, 10:343, 1975. 65

[31] 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. 11, 79

[32] Morello, M. J. et al. dE/dx for pions, kaons and protons for 9/fb analyses. *CDF Internal Note 10756*, (unpublished), 2012. 52, 54, 77

[33] Punzi, G. Comments on Likelihood fits with variable resolution. *ArXiv Physics e-prints*, [physics/0401045], 2004. 54

[34] Rosenblatt, F. The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386–408, 1958. 42

[35] Sill, A. et al. CDF Run II silicon tracking projects. *Nucl. Instrum. Meth.*, A447:1, 2000. 17

[36] Sphicas, S. A $b\bar{b}$ Monte Carlo Generator. *CDF Internal Note 2655*, (unpublished), 1994. 31

[37] Yosef, N. CP violation in meson decays. *ArXiv Physics e-prints*, [physics/0510413v1], 2005. 7