UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "G. Galilei"

Corso di Laurea Magistrale in Fisica

B flavor tagging at CMS

Relatore: Prof. Franco Simonetto
Co-relatore: Dott. Jacopo Pazzini

Laureanda: Sonia Vigolo

Anno Accademico 2014/2015

Contents

Chapter 1

Introduction

1.1 B-physics

The goal of this section is giving a theoretical background for different characteristics of hadrons containing a b quark and in particular the B neutral mesons, B^0 and B_S^0.

1.1.1 the Standard Model Lagrangian

The behaviour of the subatomic particles is described in the Lagrangian of the model, interpreted using Quantum Field Theory (QFT) and perturbation theory.

The first thing that need to be considered when choosing the terms to be placed in the Lagrangian is defining the symmetries under which the Lagrangian is invariant.

Since the space is assumed to be homogeneous and isotropic and the time to be homogeneous the Lagrangian needs to be an invariant of Poincaré. To account for Quantum Chromodynamics ($SU(3)_C$) and Electroweak Interactions ($SU(2)_L \times U(1)_Y$) there is a gauge symmetry

$$G_{SM} = SU(3)_C \times SU(2)_L \times U(1)_Y \ .$$

G_{SM} does not take into account the gravitational force.

The next thing is the representation of fermions and scalars under the symmetry. There are 5 fermions representations

$$Q^I_{Li}(3,2)_{+1/6} \ , \quad U^I_{Ri}(3,1)_{+2/3} \ , \quad D^I_{Ri}(3,1)_{-1/3} \ , \quad L^I_{Li}(1,2)_{-1/2} \ , \quad E^I_{Ri}(1,1)_{-1} \ .$$

where the index I indicates that these are the gauge interaction eigenstates, L and R indicate the chirality eigenstates and $i = 1,2,3$ indicate the generation, the numbers in bracket indicate the multiplicity of the representation under $SU(3)_C$ and $SU(2)_L$ and the last number is the hypercharge. The $SU(3)_C$ triplets are fermions that have a color charge. The representations can be written it term of quarks and leptons as

$$Q_{Li} = \begin{pmatrix} u_{Li} \\ d_{Li} \end{pmatrix} \ , \quad U_{Ri} = u_{Ri} \ , \quad D_{Ri} = d_{Ri} \ , \quad L_{Li} = \begin{pmatrix} e_{Li} \\ \nu_{Li} \end{pmatrix} \ , \quad E_{Ri} = e_{RI} \ .$$

There isn't, in the SM, a field that can be interpreted as ν_{Ri} and this is the reason neutrinos don't have mass in the SM.

There is also one scalar field $\phi(1,2)_{+1/2}$, that assumes vacuum expectation

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

resulting in a spontaneous symmetry breaking

$$SU(3)_C \times SU(2)_L \times U(1)_Y \to SU(3)_C \times U(1)_E \ .$$

From this choices there are few renormalizable terms that can be placed in the Lagrangian, and it can be summarized as follows:

$$\mathcal{L} = \mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{Hig}} + \mathcal{L}_{\text{Yuk}} \ .$$

The first part, $\mathcal{L}_{\text{kin}}$, is the kinetic term of the fermion and scalar fields where the derivative is replaced by the covariant derivative:

$$D^\mu = \partial^\mu + ig_s G_a^\mu L_a + igW_b^\mu T_b + ig' B^\mu Y \ ,$$

where G_a^μ are the eight gluon fields, W_b^μ the three weak interaction fields and B^μ the hypercharge field, L_a and T_b respectively the SU(3) and SU(2) generators (of the same multiplicity of the fermion representation they are applied to) and Y is the $U(1)_Y$ charge, g_s, g and g' are parameters that can only be derived from experiments.

The second part is the Higgs potential, describing the scalar field self interaction:

$$\mathcal{L}_{\text{Hig}} = \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 \ ,$$

where $\lambda > 0$ and $\mu^2 < 0$, requesting the vacuum stability and the VEV different from zero, are other parameters to be experimentally determined.

The third part contains the Yukawa couplings between the scalar and the fermion fields:

$$\mathcal{L}_{\text{Yuk}} = -Y_{ij}^d \overline{Q_{Li}^I} \phi D_{Rj}^I - Y_{ij}^u \overline{Q_{Li}^I} \bar{\phi} U_{Rj}^I - Y_{ij}^e \overline{L_{Li}^I} \phi E_{Rj}^I + h.c.$$

Because of the scalar field VEV at the available energy scales, these terms provide the fermion fields with the mass term (with the exception of neutrinos, as previously mentioned).

1.1.2 The CKM matrix

The quark mass terms derive from the Yukawa coupling, substituting for the scalar field its VEV:

$$\mathcal{L}_{\text{mass}}^q = -(M_d)_{ij} \overline{D_{Li}^I} D_{Rj}^I - (M_u)_{ij} \overline{U_{Li}^I} U_{Rj}^I + \text{h.c.} \ , \qquad M_q = \tfrac{v}{\sqrt{2}} Y^q$$

However most of the measures use the mass eigenstates, which diagonalize the Yukawa coupling terms, and perfoerming a basis rotation becomes useful. In the quark sector the flavor changing interactions guarantee that the weak interaction eigenstates and the mass ones are different.

To diagonalize the mass matrix M_q two unitary matrices are needed, V_{qL} and V_{qR}, such that:

$$V_{qL} M_q V_{qR}^{\dagger} = M_q^{\rm diag} \ ,$$

where $M_q^{\rm diag}$ is diagonal and real. Then the quark mass eigenstates are:

$$q_{Li} = (V_{qL})_{ij} q_{Lj}^I \quad , \qquad q_{Ri} = (V_{qR})_{ij} q_{Rj}^I$$

The only Lagrangian term that is changed by this basis rotation is the charged current coupling. Defining

$$W_{\mu}^{\pm} = \frac{W_{1\mu} \pm W_{2\mu}}{\sqrt{2}} \ ,$$

after the basis rotation that term becomes:

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

The unitary matrix

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

is the Cabibbo Kobayashi Maskawa (CKM) mixing matrix, that quantifies the charge current weak interaction mixing of quark from different generations. An unitary 3×3 matrix has 9 free parameters, 3 real values and 6 phases. Only one of these phases is physical, the other 5 can be set to zero by some quark field phase redefinitions. There are different ways to parametrize the CKM matrix with the 4 physical parameters, the most used is the standard parametrization:

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

where $c_{ij} = \cos\theta_{ij}$ and $s_{ij} = \sin\theta_{ij}$. The three real parameters are the three θ_{ij} angles and δ is the physical Kobayashi-Maskawa phase. Another typical parametrization, sometimes more useful then the standard one, is the Wolfenstein parametrization. It makes use of the three parameters λ, A and ρ and the phase η. It is an expansion in power of λ and usually is written to the third order:

$$V = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{1}{2}\lambda_2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4) \ .$$

A very useful tool that allows to summarize the measure of the CKM matrix elements taking into account its unitarity is the unitary triangle. The unitarity implies that:

$$\sum_i V_{ij}V_{ij}^* = 0, \qquad j,k = d,s,b \quad j \neq k \ .$$

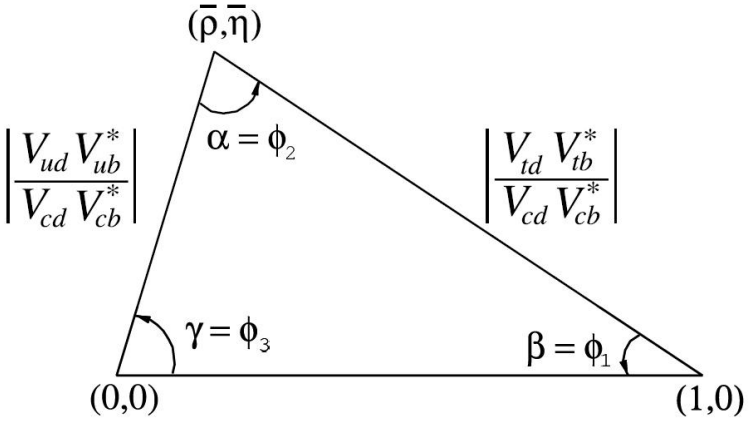


Figure 1.1: The unitary triangle for the (d,b) quark pair.

These are three independent relations between the CKM elements. Using one of these constrains it is possible to construct a triangle in the complex plane so that any term of the summation represents one side of the triangle. Now choosing the phase convention such that the "charm addend" $(V_{cj}V_{ck}^*)$ is real and rescaling the triangle such that the "charm side" is equal to 1, one obtains the unitary triangle for the (j,k) quark pair.

An example, for the (d,b) pair, is shown in figure 1.1. In this triangle the complex coordinates of the vertex are the Wolfenstein parameters ρ and η and the amplitude of the angles is related to the phases of the CKM matrix elements:

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

The γ angle is also equal to the standard parametrization phase δ_{KM}. This triangle is the easiest to probe and the most precisely measured, thanks to the fact that its side lengths have the same order of magnitude. A plot that summarizes all the measurements of the quantities related to this triangle is shown 1.2.

Also the unitary triangle for the (s,b) quark pair has been probed. It has a side two order of magnitude smaller than the other two, so it is more difficult to constrain the position of its vertex. The plot that summarizes the measures related to the elements of this triangle is shown in figure 1.3.

The third triangle is useless, being one side three order of magnitude smaller than the other sides and compatible with their error.

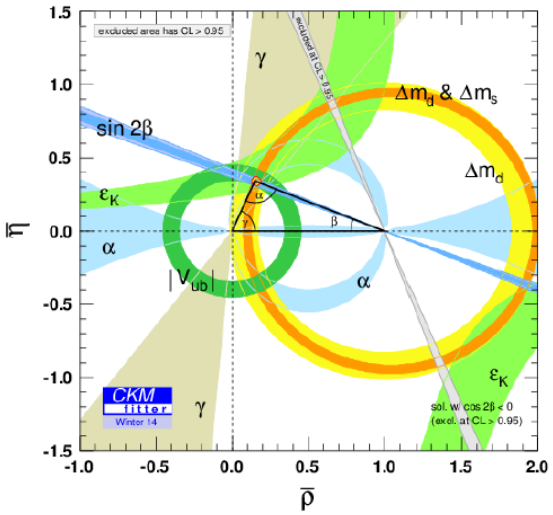


Figure 1.2: Plot of the measurement status for the (d, b) unitary triangle.

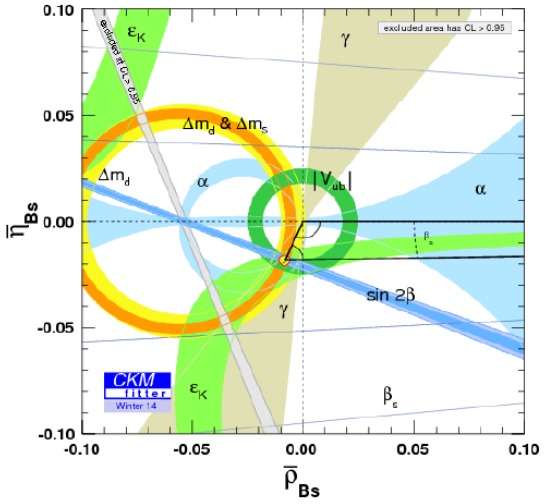


Figure 1.3: Plot of the measurement status for the (s, b) unitary triangle.

1.1.3 CP violation and mixing

In the SM there are no flavor changing neutral currents (FCNC) at tree level. This because all the neutral interactions are diagonal and universal, such that any basis rotation leaves them diagonal and then flavor conserving. Thanks to this fact, the leading order FCNC processes, such as meson mixing, are generated by one loop charge current mediated diagrams as can be seen in Fig 1.4.

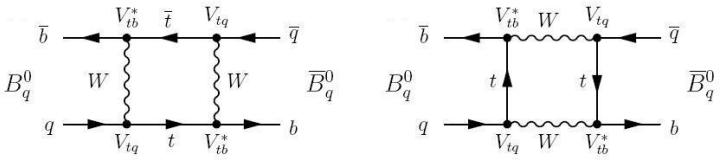


Figure 1.4: Feynmann box diagram for the oscillation of a neutral B meson

In 1955 Gell-Mann and Pais predicted the oscillations of neutral strange mesons and Lande at Brookhaven confirmed this prediction through the observation of two strange neutral particles with different life time and masses. The only hadrons that can undergo these oscillation are K^0, D^0 B^0 and B_S^0. The formalism discussed will be the one for the B mesons since they are the relevant particles in this thesis. The others, however, have a similar formalism.

CPT symmetry

There are three discrete symmetries: C, P, T. Any Lorentz invariant QFT must be invariant under CPT. This invariance guarantees that a particle and its antiparticle have the same mass and total decay rate, then a theory violates T if and only if it violates CP.

The SM is not invariant under C and under P. These transformations change the chirality of the fermion fields but left-handed and right-handed fields have different representations, so the C and P symmetries are maximally violated. This fact is true independently of the value of the Lagrangian parameters.

The CP invariance of the theory is more complicated to probe. In fact, a CP transformation applied on the Lagrangian transforms any term in its Hermitian conjugate, but leaving untouched the coefficients. So the Lagrangian is invariant under CP transformation, unless any of its parameters contains a physical phase. The only term that can contain such a phase is the Yukawa term.

The kinetic and Higgs terms are invariant under CP, the only term that can contain such a phase is the Yukawa term, in fact the Lagrangian is invariant under CP if $Y_{ij} = Y_{ij}^*$. With the rotations that was made to diagonalize the mass matrix the presence of this phase falls on the V_{CKM} matrix and measuring its parameters become essential to see CP violations.

Mixing

The strong interactions creates two stable states

$$|B\rangle = |\bar{b}q\rangle \ ,$$

$$|\overline{B}\rangle = |b\overline{q}\rangle \ ,$$

where $q = d$, s and $B = B^0$, B_S^0. These states are eigenstates of the Hamiltonian $\mathcal{H} = m_0 \mathbb{1}_2$ that needs to be diagonal for the CPT symmetry that is, to our knowledge, an unbroken symmetry.

When the weak interactions are introduced these states are not stable any more and can decay. The final states can be accessible by only one of the state or by both, if that is the case, the two discrete states are connected. Furthermore, the states are connected by a direct matrix element $\langle B|\mathcal{H}_{\mathcal{W}}|\overline{B}\rangle$ or via off-shell continuum states accessible to both. This has various effects:

- The masses of both states $|B\rangle$ and $|\overline{B}\rangle$ are shifted by δE, and become $M = m_0 + \delta E$.
- The possible interactions through off-shell continuum states or via $\mathcal{H}_{\mathcal{W}}$ produces a non diagonal elements in the real part of the Hamiltonian different from zero.
- The decays introduce an imaginary part to H that is connected with the on-shell decays accessible only by one state (Γ_{ii}) and with continuum states common to both (Γ_{12})

Again the CPT symmetry implies $\Gamma_{11} = \Gamma_{22}$. This result in a new Hamiltonian that is not Hermitian

$$\mathcal{H} = \begin{pmatrix} m_0 + \delta E & W_{12} + \delta E_{12} \\ W_{12}^* + \delta E_{12}^* & m_0 + \delta E \end{pmatrix} - \frac{i}{2} \begin{pmatrix} \Gamma & \Gamma_{12} \\ \Gamma_{12}^* & \Gamma \end{pmatrix} \ ,$$

and has its own eigenvalues and eigenstates that can be found diagonalizing the matrix.

$$|B_L\rangle = p|B\rangle + q|\overline{B}\rangle \tag{1.1}$$

$$|B_H\rangle = p|B\rangle - q|\overline{B}\rangle \tag{1.2}$$

where H and L are a reference to the different masses and stand for *Heavy* and *Light* and where

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

and the eigenvalues are

$$\lambda_{H,L} = m_{H,L} - \frac{i}{2}\Gamma_{H,L}$$

The masses and the widths of these states are

$$\begin{aligned} m_{H,L} &= M \pm \Re\left(\sqrt{|M_{12}|^2 - \frac{|\Gamma_{12}|^2}{4} - i\Re(M_{12}\Gamma_{12}^*)}\right) \\ &\equiv M \pm \Delta m/2 \\ \Gamma_{H,L} &= \Gamma \pm 2\Im\left(\sqrt{|M_{12}|^2 - \frac{|\Gamma_{12}|^2}{4} - i\Re(M_{12}\Gamma_{12}^*)}\right) \\ &\equiv \Gamma \pm \Delta\Gamma/2 \end{aligned}$$

which satisfy

$$\begin{aligned} \Delta m^2 - \frac{\Delta\Gamma}{4} &= 4\left(|M_{12}|^2 - \frac{|\Gamma_{12}|^2}{4}\right) \\ \Delta m \cdot \Delta\Gamma &= 4\Re(M_{12}\Gamma_{12}^*) \end{aligned}$$

Equation 1.1 can be rewritten as:

$$\begin{aligned} |B_L\rangle &= \frac{p+q}{2}\left[(|B\rangle + |\overline{B}\rangle) + \frac{1-q/p}{1+q/p}(|B\rangle - |\overline{B}\rangle)\right] \\ |B_H\rangle &= \frac{p+q}{2}\left[(|B\rangle - |\overline{B}\rangle) + \frac{1-q/p}{1+q/p}(|B\rangle + |\overline{B}\rangle)\right] \end{aligned}$$

where

$$\frac{1-q/p}{1+q/p} \equiv \varepsilon_B$$

is a measure of the amount by which $|B_L\rangle$ and $|B_H\rangle$ differ form CP eigenstates. In the standard model ε_B is very small .

$$\begin{aligned} \frac{\Re(\varepsilon_{B^0})}{1+|\varepsilon_{B^0}|^2} &= (-0.1 \pm 1.4)\cdot 10^{-3} \\ \frac{\Re(\varepsilon_{B_S^0})}{1+|\varepsilon_{B_S^0}|^2} &= (-0.92 \pm 2.35)\cdot 10^{-3} \end{aligned}$$

The limit of no CP violation is thus $q/p = 1$. There is another possible approximation that can be done in order to simplify the computation of the time dependencies of the states. The lifetime difference $\Delta\Gamma$ between $|B_L\rangle$ and $|B_H\rangle$ is so small that can be neglected.

A general secondary effect to this approximation is that there is no CP violation in the mixing, therefore even in presence of CP violation a simple form of the time evolution can be obtained. The lifetime difference between the heavy and light states is expected to be small, $\Delta\Gamma = 1\%$ for the B^0 and perhaps as large as 25% for the B_S^0. From Equation 4.44 follows that $\Delta\Gamma = 0$ in general only if $\Gamma_{12} = 0$. In this case

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

such that $|q/p| = 1$.

The time evolution of a state which was a pure $|B\rangle$ at $t = 0$ expressed in terms of flavor eigenstates can be obtained through the simple time dependence $|B_{L,H}(t)\rangle = e^{-i\lambda_{L,H}t}|B_{L,H}(t=0)\rangle$. This procedure implies the existence of a reference frame with both $|B_L\rangle$ and $|B_H\rangle$ components of the $|B\rangle$ at rest, and such frame does not, strictly speaking, exist.

However, setting the relative velocity of the $|B_L\rangle$ and $|B_H\rangle$ states to zero is an excellent approximation.

$$\begin{aligned}|B(t)\rangle &= \frac{|B_L(t)\rangle + |B_H(t)\rangle}{2} \\ &= \frac{1}{4}\cdot\left[e^{-i(m_L - i\frac{\Gamma_L}{2})}|B_L(0)\rangle + e^{-i(m_H - i\frac{\Gamma_H}{2})}|B_H(0)\rangle\right] \\ &= e^{-i\Delta m t - \frac{\Gamma}{2}t}\cdot\left[cos(\Delta m t)|B\rangle + i\cdot sin(\Delta m t)|\overline{B}\rangle\right]\end{aligned}$$

The probability to find a B or a $\overline{B}$ after $\Delta t = t$ from its creation is

$$P_u(t) = \frac{|\langle B|B(t)\rangle|^2}{\int_0^\infty \langle B(t)|B(t)\rangle dt} = \frac{\Gamma}{2}e^{\frac{\Gamma}{t}}\cdot\left[1 + cos^2(\Delta m t)\right]$$

$$P_m(t) = \frac{|\langle \overline{B}|B(t)\rangle|^2}{\int_0^\infty \langle B(t)|B(t)\rangle dt} = \frac{\Gamma}{2}e^{\frac{\Gamma}{t}}\cdot\left[1 - cos^2(\Delta m t)\right]$$

where the index m indicates the *mixed* probability, that is the probability to find a final state different from the initial one. In the same way the index u means *unmixed*. The time-integrated versions expresses the probability that a B decays as a $\overline{B}$.

Using the $\Delta\Gamma = 0$ approximation they can be written as follows:

$$\chi_d = \int_0^\infty P_m(t) = \frac{1}{2}\frac{x}{1+x^2} = 0.1873 \pm 0.0024$$

$$\chi_s = \int_0^\infty P_m(t) = \frac{1}{2}\frac{x}{1+x^2} = 0.49927 \pm 0.00003$$

where $x \equiv \dfrac{\Delta m}{\Gamma}$.

A common variable used in hadron colliders, where both B^0 and B_S^0 can be produced, is the average time-integrated mixing probability

$$\overline{\chi} = f_d\chi_d + f_s\chi_s$$

,

where f_d is the B^0 fraction in the b sample while f_s is the B_S^0 fraction.

Table 1.1: b-flavored hadrons fragmentation fractions

b-hadron species	Fragmentation fraction	ρ with f_{B_S}	ρ with $f_{b-\text{baryons}}$
B_s	$f_s = 0.113 \pm 0.013$		
$b-$ baryons	$f_{b-\text{baryons}} = 0.085 \pm 0.022$	-0.41	
$B^0 \text{or} B^{\pm}$	$f_d = f_u = 0.401 \pm 0.013$	-0.483	-0.855

Table 1.2: Results for ϕ_S and $\Delta\Gamma_S$ from different experiments. The first uncertainty is statistical, and the second is systematic (apart from the DØ result, for which the uncertainties are combined). The CDF confidence level (C.L.) range quoted is that consistent with other experimental measurements of ϕ_S.

Experiment	ϕ_S [rad]	$\Delta\Gamma_S$ [ps^{-1}]
LHCb	$0.06 \pm 0.12 \pm 0.06$	$0.123 \pm 0.029 \pm 0.0011$
ATLAS	$0.22 \pm 0.44 \pm 0.10$	$0.053 \pm 0.021 \pm 0.010$
CMS	-	$0.048 \pm 0.024 \pm 0.003$
DØ	$-0.55^{+0.38}_{-0.36}$	$0.163^{+0.065}_{-0.064}$
CDF	[-0.60;0.12] at 68% C.L.	$0.068 \pm 0.026 \pm 0.009$

1.1.4 Measurement of CP violation for $B_S^0 \to J/\psi\phi$

In B_S^0 decays some final states can be reached both via a direct decay amplitude and via a mixing amplitude. In the case of $B_S^0 \to J/\psi\phi$ decays, the interference between these two amplitudes allows the observation of a CP violating phase. This phase is predicted to be $\phi_S = -2\beta_S$ by the standard model, where β_S is function of the CKM matrix elements as can be seen in figure 1.1 for the B^0. The current measure of β_S gives $\phi_S = -0.036 \pm 0.0016$ which is an order of magnitude smaller of $\phi_D = -0.682 \pm 0.019$ [22], so ϕ_S is more likely to be significantly increased by other sources of CP violation.

Since the CP violating phase for the B_S^0 is small the mass eigenstates should not be much different from the CP eigenstates. The $B_S^0 \to J/\psi\phi$ decay with $\phi \to K^+K^-$ has two intermediate vector particles and the K^+K^- pair is in a P-wave configuration. The final state is a superposition of CP-even and CP-odd states depending on the relative orbital angular momentum of the J/ψ and the ϕ. The measurement of ϕ_S requires the CP-even and CP-odd components to be disentangled by analyzing the distribution of the reconstructed decay angles of the final-state particles.

Results of the measurement of ϕ_S obtained by LHCb [2], ATLAS [20], CMS [24], DØ [5] and CDF [3] are presented in table 1.2.

1.2 Flavor tagging

The flavor tagging is a procedure composed of different algorithms that is used to determine if a physical state at production is in a particle or antiparticle state. It is immediate for charged particles that are self tagging, it is not obvious for neutral particles such as B^0 or B_S^0 that can be subjected to flavor mixing and can decay in non-tagging final

states. Flavor tagging is important in measurements that require the knowledge of the particle state at one or more than one point such as flavor oscillation and CP violation. Flavor tagging has two approaches, Same Side and Opposite Side that are visualized in Fig 1.5.

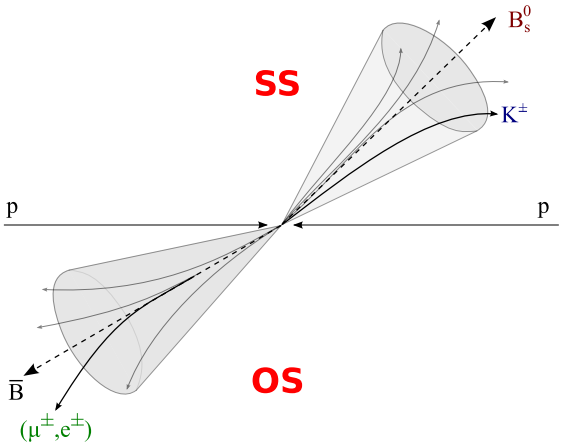


Figure 1.5: Same Side and Opposite Side tagging

1.2.1 Same side tagging

Same Side (SS) tagging uses the correlation between the state of the B meson and particles produced in the hadronization process. In the fragmentation cascade, the accompanying quark in the B meson gives rise to a quark with opposite charge conjugation, which tends to form a pion (kaon) for B^0 (B_s) with a definite charge sign. For example, if a B_s is produced in the fragmentation of a $\overline{b}$ quark, an extra $\overline{s}$ is available to form a hadron, which leads to a charged K in about 50% of the cases. [21]

1.2.2 Opposite side tagging

Opposite Side (OS) tagging algorithms rely on the pair production of b and $\overline{b}$ quarks and infer the flavor of a given B meson (signal B) from the identification of the flavor of the other b hadron (tagging B).

The algorithms use the charge of the lepton (μ, e) from semileptonic b decays, the charge of the kaon from the $b \to c \to s$ decay chain or the charge of the inclusive secondary vertex reconstructed from b-hadron decay products. All these methods have an intrinsic dilution on the tagging decision, for example due to the possibility of flavor oscillations of the tagging B.

1.2.3 Tagging power

Taggers are characterized by their effective tagging efficiencies, or tagging power that represents the effective statistical reduction of a sample due to imperfect tagging. It is

Table 1.3: The tagging power obtained by other experiments for the combined SS and OS flavor tagging algorithms.

Collider	Experiment	OS	SS
e^+e^-	BaBar	33.1 ± 0.1	
	Belle	30.1 ± 0.4	
hadronic	CDF	1.8 ± 0.1	3.5 ± 0.9
	D∅	2.48 ± 0.21	
	ATLAS	1.45 ± 0.05	
	LHCb	2.10 ± 0.25	2.42 ± 0.39

defined as:

$$\mathcal{P}_{tag} = \varepsilon_{tag}\mathcal{D}^2 = \varepsilon_{tag}(1-2\omega)^2$$

where ε_{tag}, $\mathcal{D}$ is the diluition and ω is the mistag fraction.

Defining three categories, R being the rightly tagged events, W being the wrongly tagged events and U the untagged events

$$\varepsilon_{tag} = \frac{W+R}{W+R+U} \quad \omega = \frac{W}{W+R} \ .$$

The tagging power can be improved by exploiting multiple mutually exclusive tagging categories. For each category j, a fraction of events ε and a mean dilution $\mathcal{D}_j$ can be defined. The total tagging efficiency is defined as $\varepsilon = \sum_j \varepsilon_j$, and the average dilution is described as $\mathcal{D} = \sum_j \varepsilon_j \mathcal{D}_j/\varepsilon$.

Using a categorization of the dilution, one finds a tagging power:

$$\mathcal{P}_{tag} = \sum_j \varepsilon_j \mathcal{D}_j^2 = \varepsilon\mathcal{D}^2 + \sum_j \varepsilon_j(\mathcal{D}_j - \mathcal{D})^2$$

which is larger or equal to the one obtained considering the entire sample as a single category.

The same principle applies whene there are different taggers, as long as they don't overlap.

In e^+e^- colliders values as high as 30% can be obtained, whereas the tagging power at hadron colliders reaches at most few percent. Table 1.3 reports the tagger performance obtained by different experiments at hadron colliders.

1.3 Goal of the thesis

The goal of this work is creating a lepton tagger algorithm to perform a ϕ_s measurement on the subsample $B_S^0 \to J/\psi\phi$. Hereafter, lepton means either electron or muon. The tagging algorithm should however be independent from the subsample, since the Opposite Side tagger is based on the flavor relationship between the $b\bar{b}$ pair produced and should be independent from the hadronization particles. To test this theory the tagger will be

tested on a sample of $B^+ \to J/\psi K^+$ and a sample of $B^0 \to J/\psi K^*$ both Data and MC. The B^+ is a self tagging state, so the mistag can be calcolated instantly, while the $B^0 \to J/\psi K^*$ is not and the final state is accessible to both the B^0 and $\overline{B^0}$, so a Same Side tagging algorithm will be employed as it will be discussed on chapter 4. A review of the tests made on the B^+ sample will be presented in section 3.7 since the problem was addressed by Jacopo Pazzini [27].

To develop the tagger the leptons are divided in three categories using MC-truth information: those produced in a B-hadron decay with the correct relationship between charge and flavor (labelled CC), those with opposite charge-flavor relationship (WC), and those that were not produced from a B hadron and carry random charge-flavor relation (RC).

A set of leptons variables is combined in a multivariate discriminator, trained on the simulated B_S^0 sample, in order to effectively exploit the discriminating power and the correlations of all the variables involved. In order to reduce the contribution of leptons with random charge-flavor correlation, pollution for the purpose of the tagging, a selection preliminary to the multivariate study is applied as will be discussed in section 3.2.2.

The multivariate discriminator returns a variable that should separate the CC and RC categories, and as such can be linked with the probability of the lepton being CC or WC. A subsample of events enriched in correctly tagging OS-leptons can be obtained with a simple cut on this variable.

An improved tagging performances can be obtained without reducing the number of available events by dividing the sample in bins of the MLP variable, and for each of them the tagging efficiency and the wrong tag fraction are evaluated independently. The overall tagging power is then obtained as the sum of the tagging power obtained in each MLP category, resulting in an improved perfomance compared to the single cut approach.

The result can be furtherly improved calculating an unbinned mistag fraction in order to evaluate a per-event mistag fraction. The mistag is hence defined using an analytic function depending on the value of the discriminator. In order to define the function to be used, the mistag is evaluated separately on data and MC for several bins of the discriminator response, chosen in order to have comparable tagging efficiency.

Chapter 2

LHC and CMS

The Large Hadron Collider (LHC) is an accelerator located at the European Laboratory for Particle Physics Research (CERN) in Geneva. It has been conceived to collide proton beams at a centre-of-mass energy of $\sqrt{s} = 14$ TeV and a nominal instantaneous luminosity of $\mathcal{L} = 10^{34} cm^{-2} s^{-1}$, representing a seven-fold increase in energy and a hundred-fold increase in integrated luminosity over the previous hadron collider experiments. Its main purpose is to search for rare processes like the production of Higgs or new particles with mass of 1 TeV and beyond.

Two experiments have been installed around the LHC to pursue these results: ATLAS [1] and CMS [14]. Furthermore, the LHCb [7] experiment studies the properties of charm and beauty hadrons produced with large cross sections in the forward region in collisions at the LHC, and the ALICE [4] experiment analyses the data from relativistic heavy ion collisions to study the hadronic matter in extreme temperature and density conditions (i.e. high quark-gluon density).

2.1 The Large Hadron Collider

The LHC has been installed in the same tunnel which hosted the $e^+ + e^-$ collider LEP (Large Electron-Positron collider). Accelerated electrons and positrons suffer large energy loss due to the synchrotron radiation, which is proportional to $E^4/(Rm^4)$, where E is the electron energy, m its mass and R the accelerator radius. To obtain energies of the order of TeV, at the fixed accelerator radius, only massive charged particles could have been used: protons and heavy nuclei. The energy loss is reduced by a factor $(2000)^4$ for a given fixed energy E for protons, respect to electrons.

Another important aspect of the LHC is the collision rate. To produce a sufficient number of rare processes, the collision rate needs to be very high. Beam protons are collected in packets called bunches. The collision rate is proportional to the instantaneous luminosity of the accelerator, defined as:

$$\mathcal{L} = \frac{f k n_p^2}{4\pi \sigma_x \sigma_y},$$

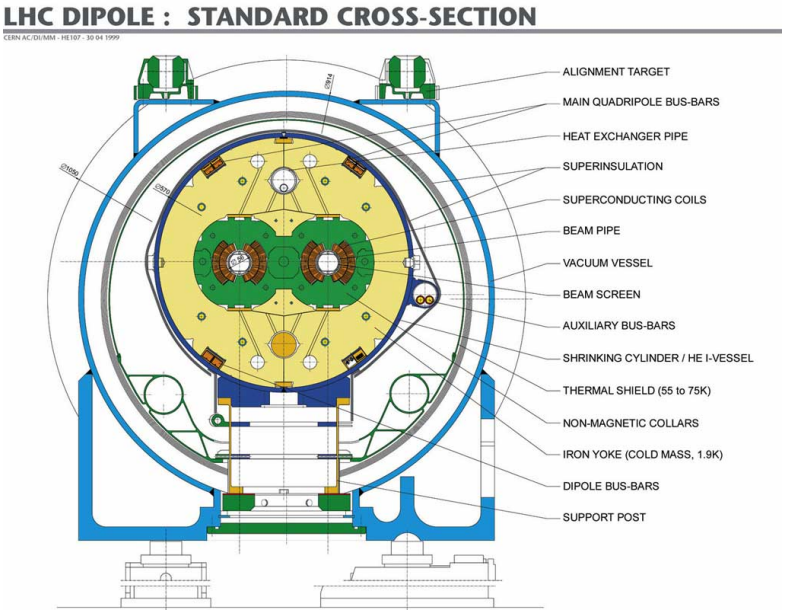


Figure 2.1: LHC dipole magnet section scheme.

where f is the bunch revolution frequency, k the number of bunches, n_p the number of protons per bunch, σ_x and σ_y their transverse dispersion along the x and y axis. At the nominal 14 TeV LHC conditions ($\mathcal{L} = 10^{34} cm^{-2} s^{-1}$) the parameter values are: $k = 2808$, $n_p = 1.5 \cdot 10^{-11}$ and $\sigma_x \sigma_y = 16.6\,\mu m^2$ (with $\sigma_z = 7.6$ cm along the beam). The integrated luminosity is defined as $L = \int \mathcal{L} dt$. For comparison we can consider the Tevatron accelerator at Fermilab, which produced proton-antiproton collisions since 1992. Its centre of mass energy was 1.8 TeV until 1998 and 1.96 TeV since 2001. To increase $\mathcal{L}$ by two orders of magnitude, protons are injected in both LHC beams. The antiprotons, in fact, are obtained by steering proton beams onto a nickel target and represent only a small fraction of the wide range of secondary particles produced in this interactions, thus have a production rate lower than the proton one.

The LHC is composed by 1232 super-conducting dipole magnets each 15 m long, providing a 8.3 T magnetic field to let the beams circulate inside their trajectories along the 27 km circumference. Two vacuum pipes are used to let beams circulate in opposite directions. A scheme representing the transverse dipole magnet section is represented in figure 2.1. More than 8000 other magnets are utilized for the beam injection, their collimation, trajectory correction, crossing. All the magnets are kept cool by superfluid helium at 1.9 K temperature. The beams are accelerated from 450 GeV (the injection energy from the SPS) to 7 TeV with 16 Radio Frequency cavities (8 per beam) which raise the beam energy by 16 MeV each round with an electric field of 5 MV/m oscillating at 400 MHz frequency.

Before the injection into the LHC, the beams are produced and accelerated by different components of the CERN accelerator complex. Being produced from ionized hy-

drogen atoms, protons are accelerated by the linear accelerator LINAC, Booster and the Proton Synchrotron (PS) up to 26 GeV energy, the bunches being separated by 25 ns each. The beams are then injected into the Super Proton Synchrotron (SPS) where they are accelerated up to 450 GeV. They are then finally transferred to the LHC and accelerated up to 7 TeV energy per beam. The CERN accelerator complex is illustrated in figure 2.2.

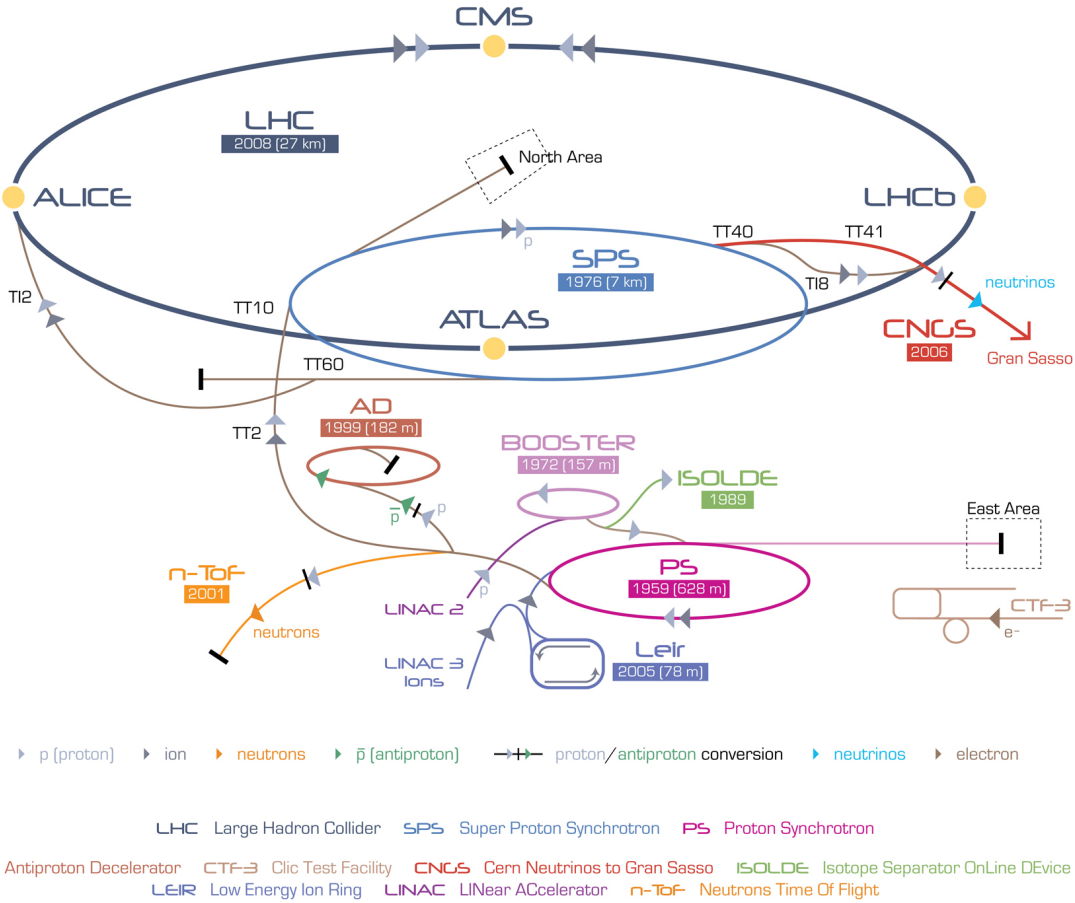
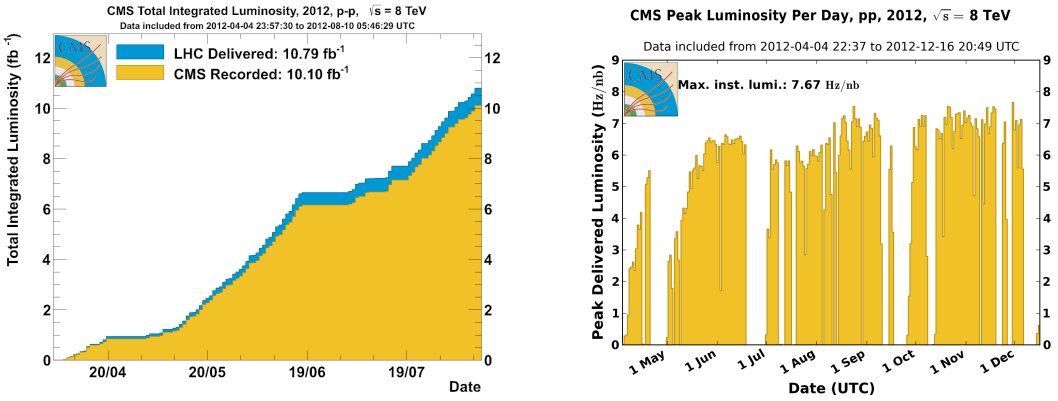


Figure 2.2: Scheme representing the CERN accelerator complex.

The LHC started its operations in December 2009 with centre of mass energy for the proton-proton collision $\sqrt{s} = 0.9$ TeV. The centre of mass energy was set to $\sqrt{s} = 7$ TeV in the 2010 and 2011 runs and raised to $\sqrt{s} = 8$ TeV in the 2012 runs. Here are reported the CMS detected peak and integrated luminosities for proton-proton runs. In 2010 the peak luminosity reached $L = 203.80 Hz/\mu b$ and the integrated luminosity has been $L = 40.76 pb^{-1}$, while during 2011 the peak luminosity increased to $L = 4.02 Hz/nb$ and the integrated luminosity has been $L = 5.55 fb^{-1}$. In the 2012 runs the peak luminosity reached $L = 7.67 Hz/nb$ and the integrated luminosity has been $L = 21.79 fb^{-1}$, as graphically summarized in figure 2.3.



(a) CMS detected integrated luminosity

(b) CMS detected peak luminosity

Figure 2.3: LHC performance in 2012.

2.2 The CMS experiment

The Compact Muon Solenoid, installed in an undergroud cave near Cessy, France, is a 22-metre-long, 15 metres wide cilindric revelator weighting 12500 tons, in charge of various aims. The majoruty of physical pocesses under exploration have small cross section, while protons' collisions products are dominated by QCD backgrounds: CMS is designed for high discirmination capacity of rare events, in particular taking advantage of channels including electrons and muons, and a great precision in the measure of secondary vertices, necessary to distinguish τ and hadrons containing heavy quarks.

The high luminosity of LHC entails a problem of pile-up: this effect can be overcame by high-granularity revelators. The occupation lowers segmenting revelators in many subgroups, at the price of attaining an excellent synchronization between them. Also, high temporal resolution is necessary due to the high frequency of interaction. Finally, high levels of radiation near the vertex require robust equipment.

In fig 2.4 a scheme of the experiment and of the main revelators groups is presented. Given the geometrical structure of CMS, the external surface of the cylinder is called "barrel", the circular basis "endcap". The "forward" revelators, placed over the endcaps, are functional to cover another geometrical region where radiatio is most intense. The silicon track revelators are often referenced to as "trackers"; the electromagnetic calorimeter is called ECAL; for the hadron calorimeter is HCAL; the first level trigger is L1; the high level trigger is HLT.

Fig 2.5 represent the typical trajectory of each kind of particle inside CMS: the muon passes through all the revelators describing a curved trajectory, leaving signal in the tracker, the calorimeters and the various muon stations; the electron leaves signal in the tracker and is absorbed by the electromagnetic calorimeter; the photon travels a straight trajectory and only showers in the electromagnetic calorimeter, without leaving signal in the tracker; the hadron leaves signal in the tracker if charged and showers in the hadronic

calorimeter.

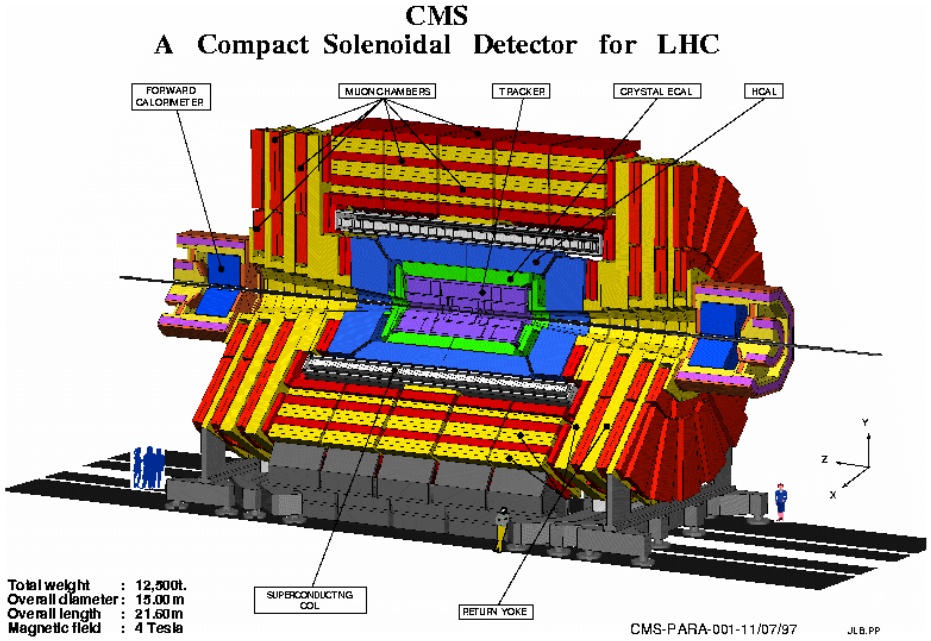


Figure 2.4: Graphical representation of the CMS revelator.

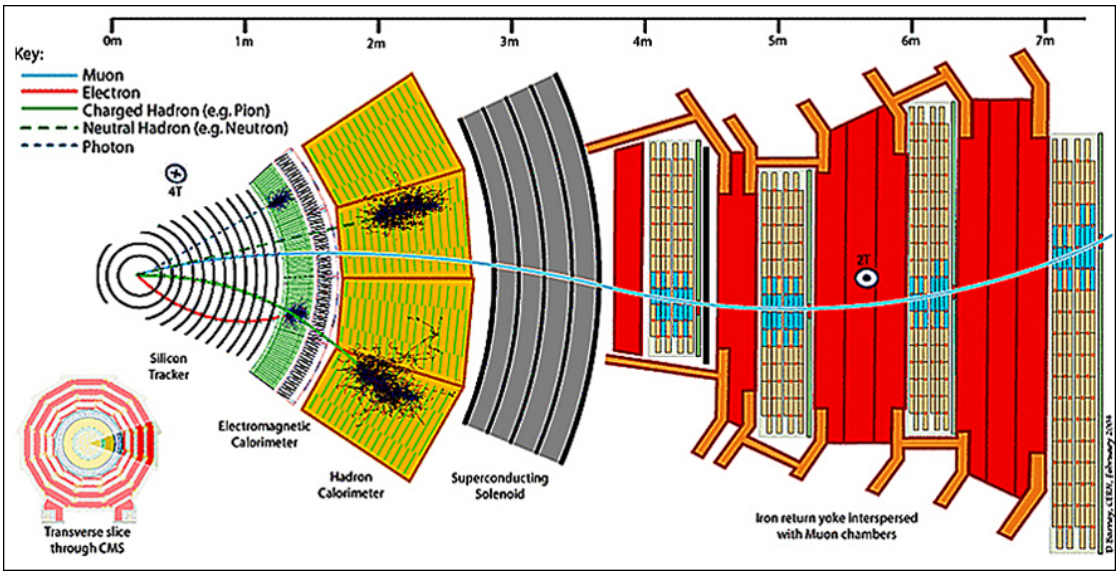


Figure 2.5: Typical mean path of each particle inside CMS.

2.2.1 The coordinate system

The reference frame used to describe the CMS detector and the collected events has its origin in the geometrical centre of the solenoid. Different types of global coordinates measured with respect to the origin (global coordinates are measured in the CMS reference frame while local coordinates are measured in the reference frame of a specific sub-detector or sensitive element) are used:

- cartesian coordinate system, $\widehat{x}$ axis points towards the centre of LHC, $\widehat{y}$ points upwards, perpendicular to LHC plane, while $\widehat{z}$ completes the right-handed reference,
- polar coordinate system, directions are defined with an azimuthal angle $\tan\phi = y/x$ and a polar angle $\tan\theta = \rho/z$, where $\rho = \sqrt{x^2+y^2}$,
- polar coordinate system, with instead of the polar angle the rapidity y and the pseudorapidity η, obtained for any particle from

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) \ ,$$
$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \ ,$$

where E is the particle energy and p_z the component of its momentum along the beam direction.

2.2.2 The magnet

The superconductive solenoid consists of a 13 metres long and 6 meters inner diameter cylinder, providing a 3.8 T constant magnetic field. Field lines of the magnet are closed by 10^4 tons of iron, the return yoke (made of barrels and endcaps) is instrumented by four layers of muon stations. A solenoidal field was chosen because it keeps the bending in the transverse plane, where the accuracy better than 20 μm is achieved in vertex poition measurement. The size of the solenoid allows efficient track recostruction up to a pseudorapidity of 2.4. The inner radius is large enough to accomodate both the silicon tracking system and the calorimeters.

2.2.3 Tracking system

The core of CMS is a Silicon Tracking System [25] with 2.5 m diameter and 5.8 m length, designed to provide a precise and efficient measurement of the trajectories of charged particles emerging from LHC collisions and reconstruction of secondary vertices. The CMS Tracking System is composed of both silicon Pixel and Strip Detectors, as shown in figure 2.6.

The Pixel Detector consists of 1440 pixel modules arranged in three barrel layers and two disks in each end-cap as in figure 2.7. The Strip Detector consists of an inner tracker

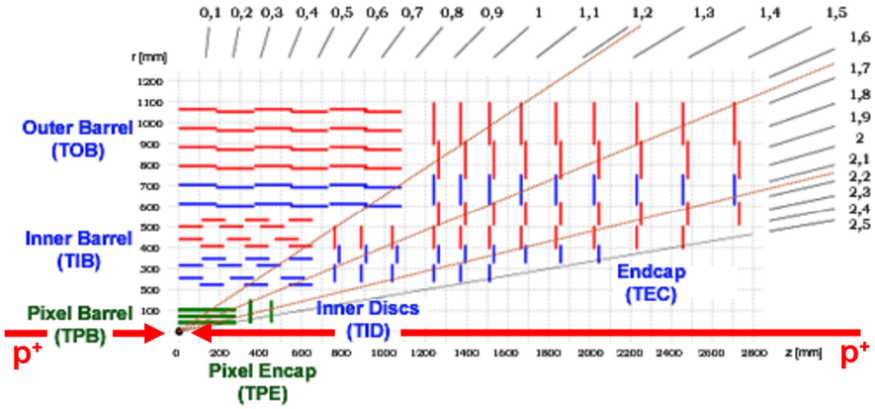


Figure 2.6: Layout of the CMS silicon tracker showing the relative position of hybrid pixels, single-sided strips and double-sided strips.

with four barrel layers and three end-cap disks and an outer tracker with six barrel layers and nine end-cap disks, housing a total amount of 15148 strip modules of both single-sided and double-sided types. Its active silicon surface of about 200 m^2 makes the CMS tracker the largest silicon tracker ever built.

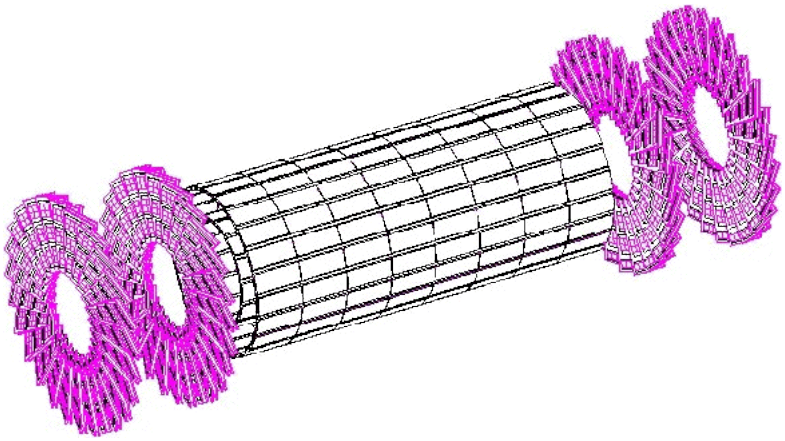


Figure 2.7: Layout of the current CMS Pixel Detector.

The LHC physics programme requires high reliability, efficiency and precision in reconstructing the trajectories of charged particles with transverse momentum larger than 1 GeV in the pseudorapidity range $|\eta| < 2.5$. Heavy quark flavors can be produced in many of the interesting channels and a precise measurement of secondary vertices is therefore needed.

The tracker completes the functionalities of ECAL and Muon System to identify electrons and muons. Also hadronic decays of tau leptons need robust tracking to be identified in both the one-prong and three-prongs topologies. Tracker information is heavily used in the High Level Trigger of CMS to help reducing the event collection rate

from the 40 MHz of bunch crossing to the 100 Hz of mass storage.

2.2.4 Silicon pixel detector

The large number of particles produced in 25 pile-up events (Events that occur in the same bunch crossing), at nominal LHC luminosity, results into a hit rate density of 1 MHz mm^{-2} at 4 cm from the beamline, decreasing down to 3 kHz mm^{-2} at a radius of 115 cm. Pixel detectors are used at radii below 10 cm to keep the occupancy below 1%. The chosen size for pixels, $0.100 \times 0 : 150\text{mm}^2$ in the transverse and longitudinal directions respectively, leads to an occupancy of the order of 10^{-4}. The layout of the Pixel Detector consists of a barrel region (BPIX), with three barrels at radii of 4.4, 7.3 and 10.2 cm, complemented by two disks on each side (FPIX), at 34.5 and 46.5 cm from the nominal interaction point. This layout provides about 66 million pixels covering a total area of about 1 m^2 and measuring three high precision points on each charged particle trajectory up to $|\eta| = 2.5$. Detectors in FPIX disks are tilted by 20° in a turbine-like geometry to induce charge sharing and achieve a spatial resolution of about 20 μm.

2.2.5 Silicon strip tracker

In the inner Strip Tracker, which is housed between radii of 20 and 55 cm, the reduced particle flux allows a typical cell size of $0.080 \times 100mm^2$, resulting in a 2% occupancy per strip at design luminosity. In the outer region, the strip pitch is increased to $0.180 \times 250mm^2$ together with the sensor thickness which scales from 0.320 mm to 0.500 mm. This choice compensates the larger capacitance of the strip and the corresponding larger noise with the possibility to achieve a larger depletion of the sensitive volume and a higher charge signal.

The Tracker Inner Barrel and Disks (TIB and TID) deliver up to 4 (r, ϕ) measurements on a trajectory using 0.320 mm thick silicon strip sensors with strips parallel to the beamline. The strip pitch is 0.080 mm in the first two layers and 0.120 mm in the other two layers, while in the TID the mean pitch varies from 0.100 mm to 0.141 mm. Single point resolution in the TIB is 0.023 mm with the finer pitch and 0.035 mm with the coarser one. The Tracker Outer Barrel (TOB) surrounds the TIB/TID and provides up to 6 $r-\phi$ measurements on a trajectory using 0.500 mm thick sensors. The strip pitch varies from 0.183 mm in the four innermost layers to 0.122 mm in the outermost two layers, corresponding to a resolution of 0.053 mm and 0.035 mm respectively. Tracker End-Caps (TEC) enclose the previous sub-detectors at 124 cm $< |z| <$ 282 cm with 9 disks carrying 7 rings of microstrips, 4 of them are 0.320 mm thick while the remaining 3 are 0.500 mm thick. TEC strips are radially oriented and their pitch varies from 0.097 mm to 0.184 mm.

As shown in figure 2.6, the first two layers and rings of TIB, TID and TOB, as well as three out of the TEC rings, carry strips on both sides with a stereo angle of 100 milliradians to measure the other coordinate: z in barrels and r in rings. This layout ensures 9 hits in the silicon Strip Tracker in the full acceptance range $|\eta| < 2.4$, and at

least four of them are two-dimensional. The total area of Strip Tracker is about 198 m^2 read out by 9.3 million channels.

2.2.6 Trajectory reconstruction

Due to the magnetic field charged particles travel through the tracking detectors on a helical trajectory which is described by 5 parameters: the curvature κ, the track azimuthal angle ϕ, the pseudorapidity η, the signed transverse impact parameter d_0 and the longitudinal impact parameter z_0. The transverse (longitudinal) impact parameter of a track is defined as the transverse (longitudinal) distance of closest approach of the track to the primary vertex.

The main standard algorithm used in CMS for track reconstruction is the Combinatorial Track Finder (CFT) algorithm [29] which uses the reconstructed positions of the passage of charged particles in the silicon detectors to determine the track parameters. The CFT algorithm proceeds in three stages: track seeding, track finding and track fitting. Track candidates are best seeded from hits in the pixel detector because of the low occupancy, the high efficiency and the unambiguous two-dimensional position information. The track finding stage is based on a standard Kalman filter pattern recognition approach which starts with the seed parameters. The trajectory is extrapolated to the next tracker layer and compatible hits are assigned to the track on the basis of the χ^2 between the predicted and measured positions. At each stage the Kalman filter updates the track parameters with the new hits.

The tracks are assigned a quality based on the χ^2 and the number of missing hits and only the best quality tracks are kept for further propagation. Ambiguities between tracks are resolved during and after track finding. In case two tracks share more than 50% of their hits, the lower quality track is discarded. For each trajectory the finding stage results in an estimate of the track parameters. However, since the full information is only available at the last hit and constraints applied during trajectory building can bias the estimate of the track parameters, all valid tracks are refitted with a standard Kalman filter and a second filter (smoother) running from the exterior towards the beam line. The expected performance of the track reconstruction is shown in figure 2.8 for muons, pions and hadrons.

The track reconstruction efficiency for high energy muons is about 99% and drops at $|\eta| > 2.1$ due to the reduced coverage of the forward pixel detector. For pions and hadrons the efficiency is in general lower because of hadronic interactions with the material in the tracker. The material budget is shown in figure 2.9 and 2.10 as a function of pseudorapidity, with the different contributions of sub-detectors and services.

The performance of the Silicon Tracker in terms of track reconstruction efficiency and resolution, of vertex and momentum measurement, are shown in figure 2.10 respectively. The first one, in particular, shows the difference in reconstruction efficiency for muons and pions, due to the larger interaction cross section of pions, which , in addition to multiple scattering, undergo hadronic interaction therefore are much more degraded by the amount of material.

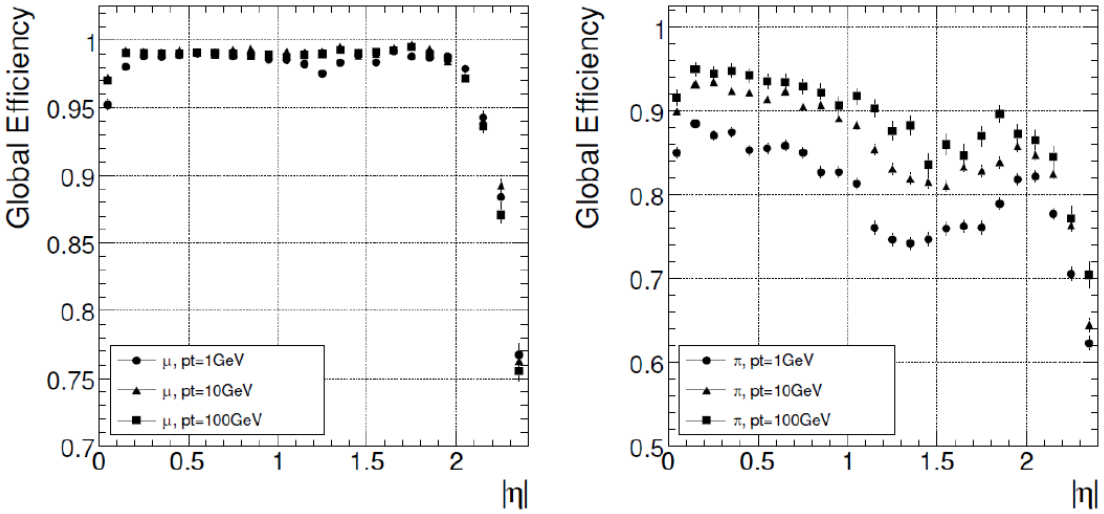


Figure 2.8: Global track reconstruction efficiency as a function of track pseudorapidity for muons (left) and pions (right) of transverse momenta of 1, 10 and 100 GeV. Figures from [14].

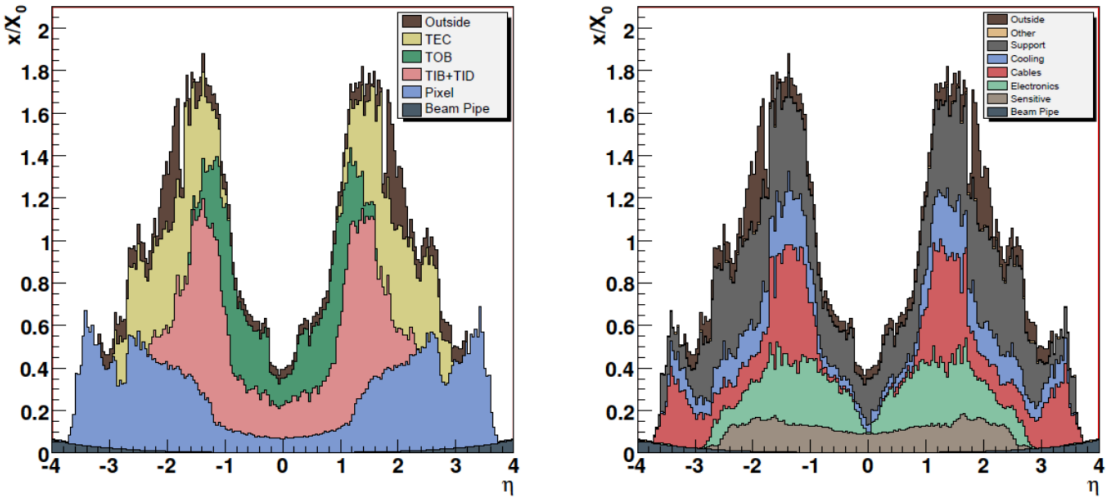


Figure 2.9: Material budget of the current CMS Tracker in units of radiation length X_0 as a function of the pseudorapidity, showing the different contribution of sub-detectors (left) and functionalities (right). Figures from [14].

2.2.7 Vertex reconstruction

The reconstruction of interaction vertices allows CMS to reject tracks coming from pile-up events. The primary vertex reconstruction is a two-step process.

Firstly the reconstructed tracks are grouped in vertex candidates and their z coordinates at the beam closest approach point are evaluated, retaining only tracks with

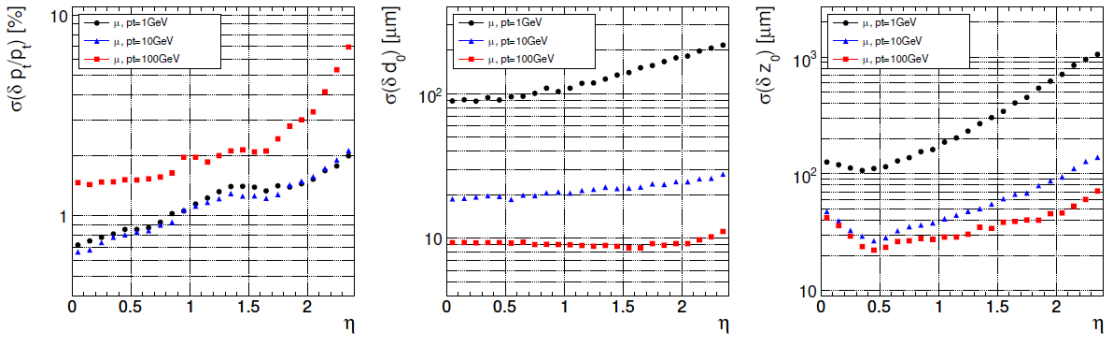


Figure 2.10: Resolution of several track parameters as a function of track pseudorapidity for single muons with transverse momenta of 1, 10 and 100 GeV: transverse momentum (left), transverse impact parameter (middle) and longitudinal impact parameter (right). Figures from [14].

impact parameter respect to the vertex candidate less than 3 cm. Vertices are then reconstructed through a recursive method for parameter estimation through a Kalman filter [30] algorithm. For a given event, the primary vertices are ordered according to the total transverse momentum of the associated tracks, $\sum p_T$. The vertex reconstruction effciency is very close to 100% and the position resolution is of the order of $\mathcal{O}(10)\mu$m in all directions.

It is also possible to reconstruct the secondary vertices, for example those from b-quark decays. The secondary vertex reconstruction uses tracks associated to jets applying further selection cuts: the transverse impact parameter of the tracks must be greater than 100 μm, to avoid tracks coming from the primary vertex, and the longitudinal impact parameter below 2 cm, to avoid tracks from pile-up events.

2.2.8 Muon spectrometer

Detection of muons at CMS exploits different technologies and is performed by a "Muon System" rather than a single detector [15]. Muons are the only particles able to reach the external muon chambers with a minimal energy loss when traversing the calorimeters, the solenoid and the magnetic field return yoke. Muons can provide strong indication of interesting signal events and are natural candidates for triggering purposes. The CMS Muon System was designed to cope with three major functions: robust and fast identification of muons, good resolution of momentum measurement and triggering.

The Muon System is composed of three types of gaseous detectors, located inside the empty volumes of the iron yoke, and therefore arranged in barrel and end-cap sections. The coverage of Muon System is shown in figure 2.11.

In the barrel region the neutron-induced background is small and the muon rate is low; moreover, the field is uniform and contained in the yoke. For these reasons, standard drift chambers with rectangular cells are used. The barrel Drift Tubes (DT) cover the $|\eta| < 1.2$ region, are divided in five wheels in the beam direction and are organized in four stations housed among the yoke layers. The first three stations contain 12 cell planes,

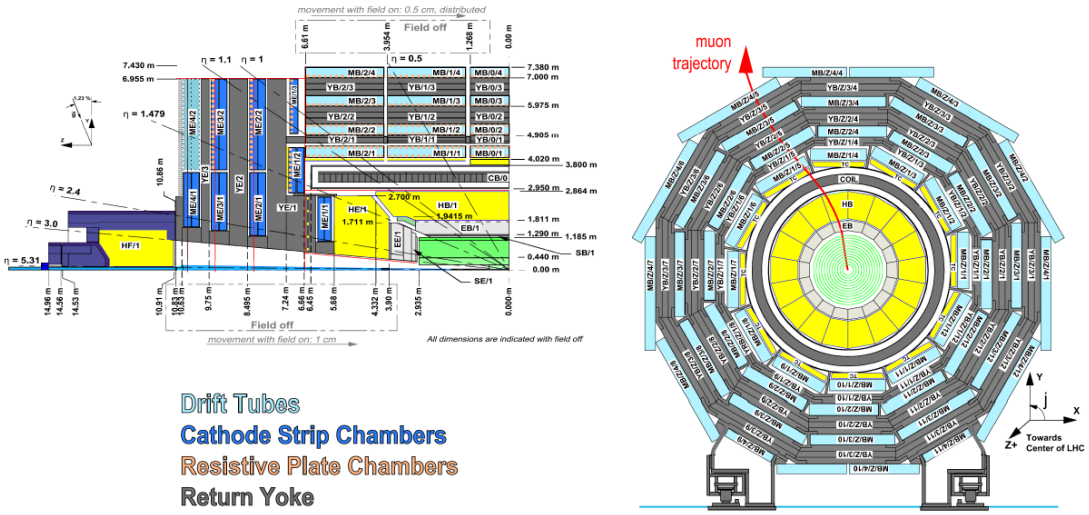


Figure 2.11: Transverse and longitudinal cross sections of the CMS detector showing the Muon System with particular emphasis on the different technologies used for detectors; the ME/4/2 CSC layers in the end-cap were included in the design but are not currently installed. Figures from [14]

arranged in two superlayers providing measurement along $r\phi$ and one superlayerlayer measuring along z, each of them containing four layers. The fourth station provides measurement only in the transverse plane.

Both the muon rates and backgrounds are high in the forward region, where the magnetic field is large and non uniform. The choice for muon detectors fell upon cathode strip chambers (CSC) because of their fast response time, fine segmentation and radiation tolerance. Each end-cap is equipped with four stations of CSCs. The CSCs cover the $0.9 < |\eta| < 2.4$ pseudorapidity range. The cathode strips are oriented radially and provide precise measurement in the bending plane, the anode wires run approximately perpendicular to the strips and are read out to measure the pseudorapidity and the beam-crossing time of a muon. The muon reconstruction efficiency is typically 95 - 99% except for the regions between two barrel DT wheels or at the transition between DTs and CSCs, where the efficiency drops.

Both the DTs and CSCs can trigger on muons with a Level 1 pT resolution of 15% and 25%, respectively. Additional trigger dedicated muon detectors were added to help measured the correct beam crossing time. These are Resistive Plate Chambers (RPC), gaseous detector operated in the avalanche mode, which can provide independent and fast trigger with high segmentation and sharp p_T threshold over a large portion of the pseudorapidity range. The overall p_T resolution on muons is shown in figure 2.12, with emphasis on the different contribution from the Muon System and the Silicon Tracker.

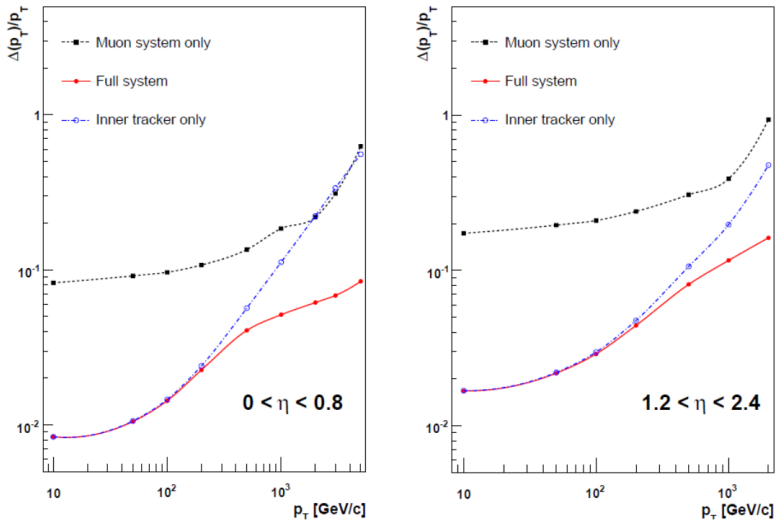


Figure 2.12: Resolution on p_T measurement of muons with the Muon System, the Silicon Tracker or both, in the barrel (left) and end-caps (right). Figures from [?].

2.2.9 Muon reconstruction

Muon detection and reconstruction play a key role in the CMS physics program, both for the discovery of New Physics and for precision measurements of SM processes. CMS has been designed for a robust detection of muons over the entire kinematic range of the LHC and in a condition of very high background. The muon system allows an efficient and pure identification of muons, while the inner tracker provides a very precise measurement of their properties. An excellent muon momentum resolution is made possible by the high-field solenoidal magnet. The steel flux return yoke provides additional bending power in the spectrometer, and serves as hadron absorber to facilitate the muon identification. Several muon reconstruction strategies are available in CMS, in order to fulfil the specific needs of different analyses. The muon reconstruction consists of three main stages:

- local reconstruction: in each muon chamber, the raw data from the detector read-out are reconstructed as individual points in space; in CSC and DT chambers, such points are then fitted to track segments;
- stand-alone reconstruction: points and segments in the muon spectrometer are collected and fitted to tracks, referred to as "stand-alone muon tracks";
- global reconstruction: stand-alone tracks are matched to compatible tracks in the inner tracker and a global fit is performed using the whole set of available measurements: the resulting tracks are called "global muon tracks".

Muon identification represents a complementary approach with respect to global reconstruction: it starts from the inner tracker tracks and flags them as muons by searching for matching segments in the muon spectrometer. The muon candidates produced with

this strategy are referred to as "tracker muons". After the completion of both algorithms, the reconstructed standalone, global and tracker muons are merged into a single software object, with the addition of further information, like the energy collected in the matching calorimeter towers. This information can be used for further identification, in order to achieve a balance between efficiency and purity of the muon sample.

2.2.10 Calorimetry

Identification of electrons, photons, and hadrons relies on accurate calorimetry, which is a destructive measurement of the energy of a particle. As in most of the particle physics experiments, a distinction is made between electromagnetic calorimetry and hadron calorimetry. Electromagnetic calorimetry is based on the production of EM showers inside a high-Z absorber, while hadron calorimetry measures the effects of hadron inelastic scattering with heavy nuclei, including production of photons from neutral pions and muons, and neutrinos from weak decays. Calorimetry must be precise and hermetic also to measure any imbalance of momenta in the transverse plane which can signal the presence of undetected particles such as high-p_T neutrinos.

The electromagnetic calorimeter of CMS, ECAL, is a homogeneous calorimeter, where the absorber material is the same as the sensitive one [17]. ECAL is composed of 61200 lead tungstate ($PbWO_4$) crystals in the barrel region and 7324 crystals in the end-caps, as shown in figure 2.13.

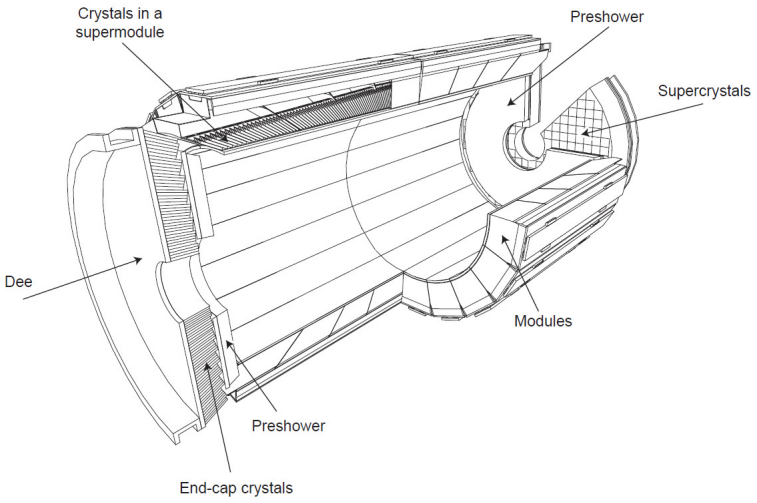


Figure 2.13: Cut-away view of the CMS ECAL showing the hierarchical structure of crystals arranged in supercystals and modules and the orientation of crystals whose major axis is always directed to the origin of the reference frame.

The crystal cross-section is $22 \times 22mm^2$ at the front face, while the length is 230 mm. End-caps are equipped with a preshower detector. Lead tungstate was chosen because of its high density, 8.28 g cm^{-3}, short radiation length, 0.89 cm, and small Molière

radius, 2.2 cm. This way, the calorimeter can be kept compact with fine granularity, while scintillation and optical properties of $PbWO_4$ make it fast and radiation tolerant. Signal transmission exploits total internal reflection. Scintillation light detection relies on two different technologies. Avalanche photodiodes (APD) are used in the barrel region, mounted in pairs on each crystals, while vacuum phototriodes (VPT) are used in the end-caps. The preshower detector is a sampling calorimeter composed of lead radiators and silicon strips detectors, and it is used to identify neutral pions in the forward region. The nominal energy resolution, measured with electron beams having momenta between 20 and 250 GeV, is

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \ ,$$

where all the energies are in GeV and the different contributions are respectively: the stochastic one (S), due to fluctuations in the lateral shower containment and in the energy released in the preshower, that due to electronics (N), digitization and pile-up, and the constant term (C), due to intercalibration errors, energy leakage from the back of the crystal and nonuniformity in light collection.

The hadron calorimeter of CMS, HCAL, is a sampling calorimeter employed for the measurement of hadron jets and neutrinos or exotic particles resulting in apparent missing transverse energy [18]. A longitudinal view of HCAL is shown in figure 2.14.

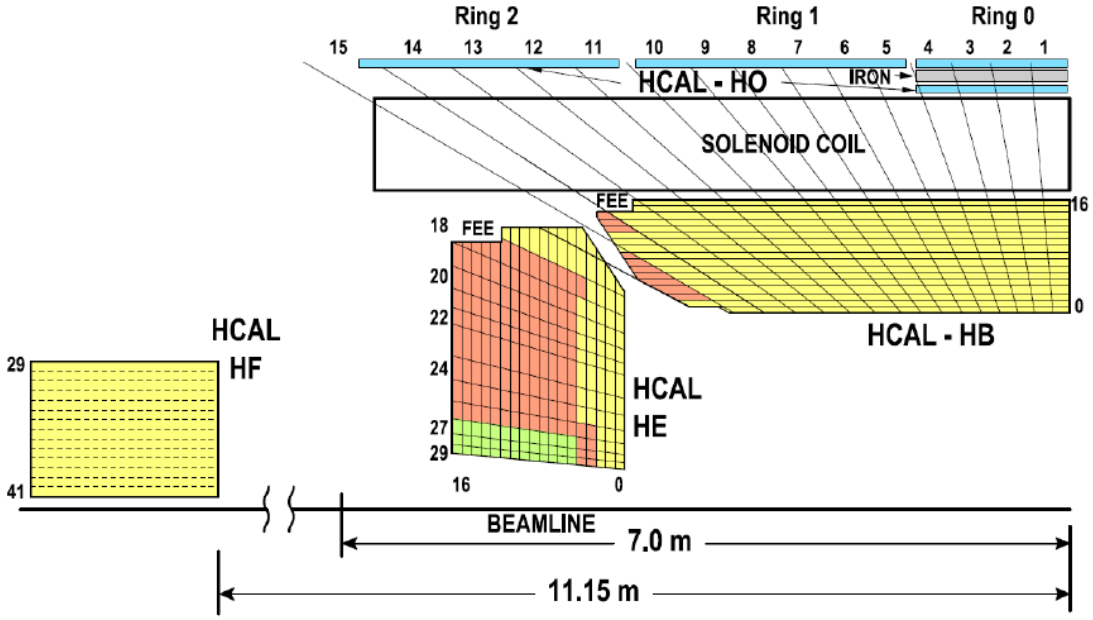


Figure 2.14: Cross section of the CMS HCAL showing the tower segmentation. Figure from [14]

The hadron calorimeter size is constrained in the barrel region, $|\eta| < 1.3$, by the maximum radius of ECAL and the inner radius of the solenoid coil. Because of this, the total amount of the absorber material is limited and an outer calorimeter layer is located

outside of the solenoid to collect the tail of the showers. The pseudorapidity coverage is extended in the $3 < |\eta| < 5.2$ by forward Cherenkov-based calorimeters. The barrel part, HB, consists of 36 wedges, segmented into 4 azimuthal sectors each, and made out of flat brass absorber layers, enclosed between two steel plates and bolted together without any dead material on the full radial extent.

There are 17 active plastic scintillator tiles interspersed between the stainless steel and brass absorber plates, segmented in pseudorapidity to provide an overall granularity of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The same segmentation is maintained in end-cap calorimeters, HE, up to $|\eta| < 1.6$, while it becomes two times larger in the complementary region. The maximum material amount in both HB and HE corresponds to approximately 10 interaction lengths λ_I . The energy resolution on single electron and hadron jets is shown in figure 2.15.

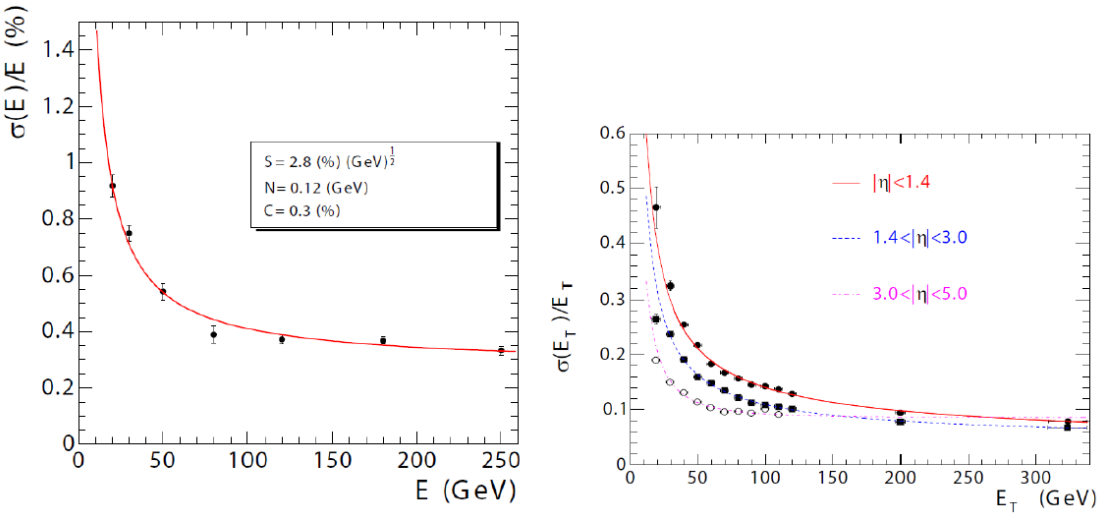


Figure 2.15: Left: ECAL energy resolution as a function of the electron energy as measured from a beam test. The energy was measured in a 3×3 crystals array with the electron impacting the central one. The stochastic, noise and constant terms are given. Right: the jet transverse energy resolution as a function of the transverse energy for barrel jets, end-cap jets and very forward jets reconstructed with an iterative cone algorithm with cone radius $R = 0.5$. Figures from [14]

2.2.11 Trigger and data acquisition

High bunch crossing rates and design luminosity at LHC correspond to approximately 20-25 superimposed events every 25 ns, for a total of 10^9 events per second. The large amount of data associated to them is impossible to store and process, therefore a dramatic rate reduction has to be achieved. This is obtained with two steps: the Level 1 Trigger [10] and the High Level Trigger, HLT [13]. The Level 1 Trigger is based on custom and programmable electronics, while HLT is a software system implemented on a ∼1000 commercial processors farm. The maximum allowed output rate for Level 1 Trigger is 100 kHz, which should be even kept lower, about 30 kHz, for safe operation.

Level 1 Trigger uses rough information from coarse segmentation of calorimeters and Muon Detectors and holds the high-resolution data in a pipeline until acceptance/rejection decision is made. HLT exploits the full amount of collected data for each bunch crossing accepted by Level 1 Trigger and is capable of complex calculations such as the off-line ones. HLT algorithms are those expected to undergo major changes in time, particularly with increasing luminosity.

Configuration and operation of the trigger components are handled by a software system called Trigger Supervisor. The Level 1 Trigger relies on local, regional and global components. The Global Calorimeter and Global Muon Triggers determine the highest-rank calorimeter and muon objects across the entire experiment and transfer them to the Global Trigger, the top entity of the Level 1 hierarchy. The latter takes the decision to reject an event or to accept it for further evaluation by the HLT. The total allowed latency time for the Level 1 Trigger is 3.2 μs. A schematic representation of the Level 1 Trigger data flow is presented in figure 2.16.

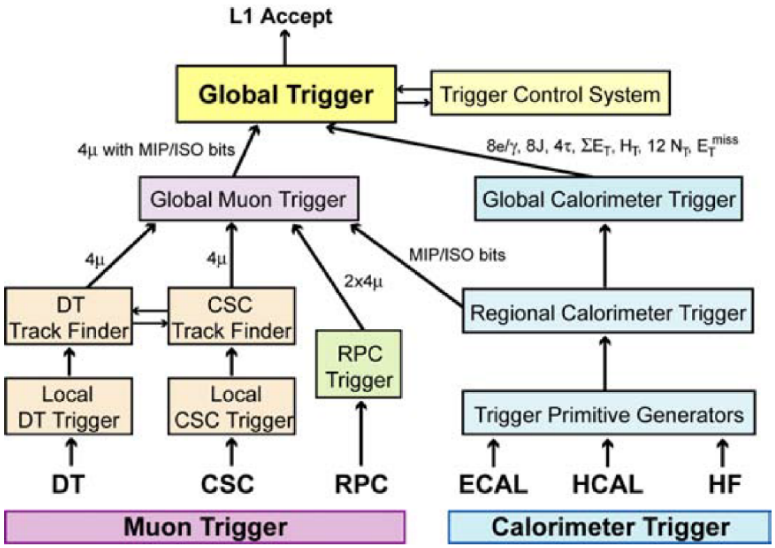


Figure 2.16: Schematic representation of the Level 1 Trigger data flow.

2.2.12 Muon trigger

All Muon Detectors (DT, CSC and RPC) contribute to the Trigger. Barrel DTs provide Local Trigger in the form of track segments in ϕ and hit patterns in η. End-cap CSCs provide 3-dimensional track segments. Both CSCs and DTs provide also timing information to identify the bunch crossing corresponding to candidate muons. The Local DT Trigger is implemented in custom electronics. BTIs, Bunch and Track Identifiers, search for coincidences of aligned hits in the four equidistant planes of staggeredn drift tubes in each chamber superlayer. From the associated hits, track segments defined by position and angular direction are determined. TRACOs, Track Correlators, attempt to correlate

track segments measured in the two ϕ superlayers of each DT chamber, enhancing the angular resolution and producing a quality hierarchy.

The requirement of robustness implies redundancy, which introduces, however, a certain amount of noise or duplicate tracks giving rise to false Triggers. Therefore the BTIs, the TRACOs and the different parts of the Local Trigger contain complex noise and ghost reduction mechanisms. The position, transverse momentum and quality of tracks are coded and transmitted to the DT regional Trigger, called the Drift Tube Track Finder (DTTF), through high-speed optical links. The Global Muon Trigger (GMT) combines the information from DTs, CSCs and RPCs, achieving an improved momentum resolution and efficiency compared to the stand-alone systems. It also reduces the Trigger rate and suppresses backgrounds by making use of the complementarity and redundancy of the three Muon Systems.

The Global Muon Trigger also exploits MIP/ISO bits (The MIP bit is set if the calorimeter energy is consistent with the passage of a minimum ionizing particle, the isolation bit is set if a certain energy threshold in the trigger towers surrounding the muon is not exceeded) from the Regional Calorimeter Trigger. A muon is considered isolated if its energy deposit in the calorimeter region from which it emerged is below a defined threshold. DT and CSC candidates are first matched with barrel and forward RPC candidates based on their spatial coordinates. If a match is possible, the kinematic parameters are merged. Several merging options are possible and can be selected individually for all track parameters, taking into account the strengths of the individual Muon Systems. Muons are back-extrapolated through the calorimeter regions to the vertex, in order to retrieve the corresponding MIP and ISO bits, which are then added to the GMT output and can be taken into account by the Global Trigger (GT). Finally, the muons are sorted by transverse momentum and quality to deliver four final candidates to the GT. The Muon Trigger is designed to cover up to $|\eta| < 2.4$.

2.2.13 Global trigger

The Global Trigger takes the decision to accept or reject an event at Level 1, based on candidate $e = \gamma$, muons, jets, as well as global quantities such as the sums of transverse energies (defined as $E_T = E\sin\theta$), the missing transverse energy and its direction, the scalar transverse energy sum of all jets above a chosen threshold (usually identified by the symbol H_T), and several threshold-dependent jet multiplicities. Objects representing particles and jets are ranked and sorted. Up to four objects are available and characterized by their p_T or E_T, direction and quality. Charge, MIP and ISO bits are also available for muons. The Global Trigger has five basic stages implemented in Field-Programmable Gate-Arrays (FPGAs): input, logic, decision, distribution and read-out. If the Level 1 Accept decision is positive, the event is sent to the Data Acquisition stage.

2.2.14 High Level Trigger and Data Acquisition

The CMS Trigger and DAQ system is designed to collect and analyse the detector information at the LHC bunch crossing frequency of 40 MHz. The DAQ system must

sustain a maximum input rate of 100 kHz, and must provide enough computing power for a software filter system, the High Level Trigger (HLT), to reduce the rate of stored events by a factor of 1000. In CMS all events that pass the Level 1 Trigger are sent to a computer farm (Event Filter) that performs physics selections, using faster versions of the offline reconstruction software, to filter events and achieve the required output rate. The various subdetector front-end systems store data continuously in 40 MHz pipelined buffers. Upon arrival of a synchronous Level 1 Trigger Accept via the Timing, Trigger and Control System (TTCS) the corresponding data are extracted from the front-end buffers and pushed into the DAQ system by the Front-End Drivers (FEDs).

The event builder assembles the event fragments belonging to the same Level 1 Trigger from all FEDs into a complete event, and transmits it to one Filter Unit (FU) in the Event Filter for further processing. The DAQ system includes back-pressure from the filter farm through the event builder to the FEDs. During operation, Trigger thresholds and pre-scales will be optimized in order to fully utilize the available DAQ and HLT throughput capacity.

Chapter 3

Opposite side tagging with Multivatiate Analysis

3.1 OS tagger

In an unbiased $b\bar{b}$ sample more than 20% of the B hadrons decay via weak interaction into an inclusive semileptonic decay $b \to \ell X$ final state, where ℓ is an electron or a muon. Given a generic hadron decay $b \to cW^-$, the lepton is mainly originated from the subsequent $W^- \to \ell^- \nu$ decay, and for this reason the charge of the lepton is a good estimator of the flavor of the related B at decay time. Assuming the hadron didn't change his flavor between production and decay, the charge of the lepton gives a good estimation of the B hadron's flavor at production time and of the flavor of the other hadron of the $b\bar{b}$ pair. In particular,

- OS-ℓ^+ tags an OS-B hadron, thus a reconstructed $\overline{B}$ at production time
- OS-ℓ^- tags an OS-$\overline{B}$ hadron, thus a reconstructed B at production time

Several background processes contribuite to the diluition of the tagging information:

- sequential decays, $b \to cW^-$ where $c \to W^+X \to \ell^+\nu X$ decay produces a lepton with the opposite sign correlation
- In about 12% of the cases the OS-B undergoes mixing, in this case $b \to \bar{b}$ and the charge-flavor relation is once again inverted for direct decays, and restored for sequential decays
- gluon splitting can result in the 2 pairs of heavy quarks and the lepton can come from a quark of the other pair, having no relation with the flavor of the reconstructed B
- hadrons or photons misidentified with leptons (due to $\gamma \to e^+e^-$ conversion, kaons or pions decaying into muons within the detector volume, hadrons absorbed in

Table 3.1: Muon provenance evaluated for the three MC samples, after the OS-muon preselection is applied. Muons coming from a B hadron decay are separated into two categories: muons with correct charge correlation with respect to the B flavor (CC), and muons contributing to the tag dilution with opposite charge-flavor relation (WC). Muons not originated from B decays have random charge-flavor correlation (RC).

	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC
N_{ev} reco-B	608843	521855	469172
N_{ev} reco-B + OS-μ	121023 (19.9%)	107577 (20.6%)	93694 (20.0%)
$B \to \mu$(CC)	15.1%	14.8%	15.4%
$B \to \mu$(WC)	5.4%	5.2%	5.4%
μ not form B (RC)	79.5%	80.0%	79.2%

Table 3.2: Electron provenance evaluated for the three MC samples, after the OS-electron preselection is applied. Electrons coming from a B hadron decay are separated into two categories: electrons with correct charge correlation with respect to the B flavor (CC), and electrons contributing to the tag dilution with opposite charge-flavor relation (WC). Electrons not originated from B decays have random charge-flavor correlation (RC).

	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC
N_{ev} reco-B	608843	521855	469172
N_{ev} reco-B + OS-e	77529 (12.7%)	71960 (13.8%)	60984 (13.0%)
$B \to e$(CC)	14.6%	13.6%	14.6%
$B \to e$(WC)	4.9%	4.7%	5.0%
e not form B (RC)	80.5%	81.7%	80.4%

ECAL or sailing through HCAL reaching the muon chambers) carry weak to none charge correlation.

The composition of muons and electrons at selection stage can be found in table 3.1 and 3.2 respectively.

Some of this effects are reduced by the triggers, but not completely. Nonetheless a tagger based on the lepton charge only was created both for the muon and the electron. Their tagging power are shown in tab 3.3 for the muon and 3.4 for the electron.

Table 3.3: Muon tagging performances evaluated on the $B_S^0 \to J/\psi\phi$, $B^+ \to J/\psi K^+$, and $B^0 \to J/\psi K^*$ simulations, and on the $B^+ \to J/\psi K^+$ channel of the 2012 Data.

[%]	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC	$B^+ \to J/\psi K^+$ 2012 Data
ε_{tag}	3.77 ± 0.03	3.88 ± 0.03	3.97 ± 0.03	4.59 ± 0.03
ω	29.1 ± 0.3	28.3 ± 0.3	28.8 ± 0.3	31.0 ± 0.4
$\mathcal{P}_{tag}$	0.66 ± 0.02	0.73 ± 0.02	0.72 ± 0.02	0.67 ± 0.03

Table 3.4: Electron tagging performances evaluated on the $B_S^0 \to J/\psi\phi$, $B^+ \to J/\psi K^+$, and $B^0 \to J/\psi K^*$ simulations, and on the $B^+ \to J/\psi K^+$ channel of the 2012 Data.

[%]	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC	$B^+ \to J/\psi K^+$ 2012 Data
ε_{tag}	3.35 ± 0.02	3.41 ± 0.03	3.47 ± 0.03	3.86 ± 0.03
ω	34.3 ± 0.3	34.8 ± 0.4	34.1 ± 0.4	35.5 ± 0.4
$\mathcal{P}_{tag}$	0.33 ± 0.01	0.32 ± 0.02	0.35 ± 0.02	0.33 ± 0.02

3.2 Data and MC selection

B-flavor tagging algorithms are developed and validated on Monte Carlo (MC) simulations. The samples of simulated data sets used are shown in Table 3.5. The primary processes are generated with `PYTHIA` [28], particle decays is managed by `EVTGEN` [26]. Final state radiation is included in `EVTGEN` through the `PHOTOS` [8], [9] package. Long-lived generated particles are traced through a detailed model of the detector with the `GEANT` [6] package.

A dedicated sample of 5658339 simulated events containing at least one $B_S^0 \to J/\psi\phi$ decay has been produced by selecting at the generator level pp collisions resulting in the production of a $b\bar{b}$ pair, with one of the two b quarks hadronizing as a B_S^0, and forced then to decay according to the following cascade process: $B_S^0 \to J/\psi\phi$, followed by $J/\psi \to \mu^+\mu^-$, $\phi(1020) \to K^+K^-$, while the other b in the event is allowed to hadronize and decay without any constraint. A sample of 4898805 $B^+ \to J/\psi K^+$ events and a sample of 6093292 $B^0 \to J/\psi K^*$ events have been produced as well.

Tagging is also applied on the data recorded during the 2012 pp run, corresponding to an integrated luminosity of 19.8 fb^{-1}. Good runs and luminosity sections are selected with the JSON file: `Cert_190456-208686_8TeV_22Jan2013ReReco_Collisions12_JSON.txt`

3.2.1 Trigger and event selection

All the candidates are required to pass one of the `HLT4_Jpsi_Displaced_v9-v12` trigger paths for the $J/\psi \to \mu^+\mu^-$ reconstruction. It requires two muons, each with $p_T > 4$GeV/c, $|\eta| < 2.1$ and $\cos\theta > 0.9$, where θ is the angle between the muon momentum and the di-muon flight direction. The two opposite charge tracks are required to have a distance of closest approach of less than 0.5 cm. The muon tracks are then fitted to a common vertex and their momenta are computed with the vertex constraint.

The dimuon invariant mass must be compatible with that of the J/ψ candidate, which must have $p_T > 7$ Gev/c and a decay lenght significance $L_{xy}/\sigma(L_{xy}) > 3$, the L_{xy} is the distance between the primary vertex of the event and the dimuon vertex in the transverse plane and $\sigma(L_{xy})$ is its measured uncertainty. This last constraint is not a simple cut and it can affect the fraction of B mesons with short lifetime.

All the other particles are formed fitting the two tracks of their decay products: a summary of the selection criteria on the invariant masses and on p_T can be found in table 3.6. Two B^0 mass hypotheses are evaluated under the $K - \pi$ ($\pi - K$) mass assumption

Table 3.5: Data and Monte Carlo simulation samples used in the tagging studies. The data are `MuOnia-AOD` *datasets. MC are* `AODSIM` *datasets.*

Data
/MuOnia/Run2012A-22Jan2013-v1/AOD
/MuOnia/Run2012B-22Jan2013-v1/AOD
/MuOnia/Run2012C-22Jan2013-v1/AOD
/MuOnia/Run2012D-22Jan2013-v1/AOD
$B^+ \to J/\psi K^+$ Monte Carlo
/BuToJPsiK_K2MuPtEtaEtaFilter_8TeV-pythia6-evtgen/...
.../Summer12_DR53X-PU_S10_START53_V7A-v2/AODSIM
$B^0 \to J/\psi K^*$ Monte Carlo
/BdToJpsiKstar_EtaPtFilter_8TeV-pythia6-evtgen/...
.../Summer12_DR53X-PU_S10_START53 V7C-v1/AODSIM
/BdToKstarJPsi EtaPtFilter 8TeV-pythia6-evtgen/...
.../Summer12 DR53X-PU RD2 START53 V19F-v1/AODSIM
$B_S^0 \to J/\psi\phi$ Monte Carlo
/BsToJPsiPhi_2K2MuPtEtaFilter_8TeV-pythia6-evtgen/...
.../Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM
/BsToJpsiPhiV2_BFilter_TuneZ2star_8TeV-pythia6-evtgen/..
.../Summer12_DR53X-PU RD__START53_V19F-v3/AODSIM

for the two tracks. At least one of the two hypothesis must pass the selection cuts for the event to be selected.

Table 3.6: Selection cuts used in the $B \to J/\psi X$ meson decay reconstruction

Variable	$B_S^0 \to J/\psi\phi$	$B^0 \to J/\psi K^*$	$B^+ \to J/\psi K^+$
vertex probability	> 0.02	> 0.02	> 0.02
M(J /psi)	3.08-3.11 GeV/c^2	3.08-3.11 GeV/c^2	3.08-3.11 GeV/c^2
M($\phi(1200)$)	1.01-1.03 GeV/c^2	-	-
M(B)	5.2-5.6 GeV/c^2	5.00-5.50 GeV/c^2	5.00-5.55 GeV/c^2
$p_T(J/\psi)$	> 7 GeV/c	> 7 GeV/c	> 7 GeV/c
$p_T(K)$	> 0.7 GeV/c	-	> 2 GeV/c
p_T(tracks)	-	> 0.7 GeV/c	-

3.2.2 Opposite Side lepton selection

Different variables are used to select the candidate muons: muon transverse momentum p_T, and pseudorapidity η, muon's inner track 3-dimensional impact parameter d_{xyz}, the separation from the fully reconstructed B meson $\Delta R(B)$, and an isolation criterion evaluated with the particle flow algorithm using a ΔR cone of 0.4 (defined in detail below).

Independent cuts on the whole number of listed muon variables have been applied and for each combination the tagging efficiency and the mistag fraction have been evaluated on the B_S^0 MC sample, and the best configuration is chosen to be the one which maximizes the tagging power.

The final selection requires:

- $p_T > 2.2$ GeV/c
- $d_{xyz} < 0.1$ cm
- $\Delta R(B) > 0.3$

After these cuts the composition of the muon sample changes from that seen in table 3.1 to 3.7

Table 3.7: Composition of the muon sample at the selection stage in the three simulated samples.

	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC
N_{ev} reco-B	608843	521855	469172
N_{ev} reco-B + OS-e sel	22981 (3.8%)	20252 (3.9%)	18620 (4.0%)
$B \to \mu$(CC)	60.1%	60.8%	60.4%
$B \to \mu$(WC)	18.5%	18.2%	18.2%
μ not form B (RC)	21.4%	21.0%	21.4%

The same process was done for the OS-electron selection. Electrons are reconstructed using the Particle Flow algorithm [16,19] and the highest transverse momentum electron is selected. To reduce the number of electrons that don't originate from the B a series of cuts were studied including the following variables: electron transverse momentum p_T, and η, electron's inner track 3-dimensional impact parameter d_{xyz}, the separation from the fully reconstructed B meson $\Delta R(B)$, an isolation criterion evaluated with the particle flow algorithm using a ΔR cone of 0.4 (defined in detail below), and an electron identification variable PFmva$(e-\pi)$ computed combining in a multivariate discriminator several electron variables and optimized to discriminate electrons from background sources, mostly pions.

The final selection requires:

- $p_T > 2.2$ G0V/c
- $d_{xyz} < 0.1$ cm
- $\Delta R(B) > 0.2$
- PFmva$(e-\pi) > 0.2$

After these cuts the composition of the electron sample changes from that seen in table 3.2 to 3.8.

Table 3.8: Composition of the electron sample at the selection stage in the three simulated samples.

	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC
N_{ev} reco-B	608843	521855	469172
N_{ev} reco-B + OS-e sel	20381 (3.4%)	17789 (3.4%)	16297 (3.5%)
$B \to e$(CC)	43.6%	43.2%	43.2%
$B \to e$(WC)	13.6%	14.3%	14.1%
e not form B (RC)	42.8%	42.5%	42.7%

3.3 The Toolkit for Multivariate Analysis (TMVA)

The Toolkit for Multivariate Analysis (TMVA) provides a ROOT-integrated environment for the processing, parallel evaluation and application of multivariate classification and multivariate regression techniques [23].

All techniques used belong to the family of the "supervised learning" algorithms. They use a set of training events that consists of training examples each with an input object (a vector of variables) and an output value. The algorithm analyses the training data and produces an inferred function that can be used to map samples. The function is then checked on a test sample, for which the output value is known.

The function, generally called *weight*, is used to obtain an estimator of the output value, in a process called application. This is useful in samples for which the output value is unknown.

In the classification process the function describes a decision threshold, while in the regression process the function is an approximation of the underlying functional behaviour defining the target value. In the following discussion we will only use the classification process.

The TMVA package include various methods, some based on cuts (binary splits) or on likelihood and some with better performances for our use. In particular we will study a Boost Decision Tree method and an Artificial Neural Network method.

3.3.1 Boost Decision Tree (BDT)

A decision tree is a binary tree structure where left/right decisions are taken on one single variable at a time until a stop criterion is fulfilled. The phase space is divided in many regions that are eventually classified as signal or background, depending on the majority of training events that fall in the region, represented as a leaf node. An example of the structure can be seen in figure 3.1.

The boosting of a decision tree is an extension of the concept from one tree to several trees to create a forest. The trees are derived from the same training ensemble by assigning different weights to the same events, that are then combined into a single classifier which is given by the average of the individual decision trees.

Boosting stabilizes the response of the decision trees with respect to fluctuations in the training sample and enhances the performance of a single tree.

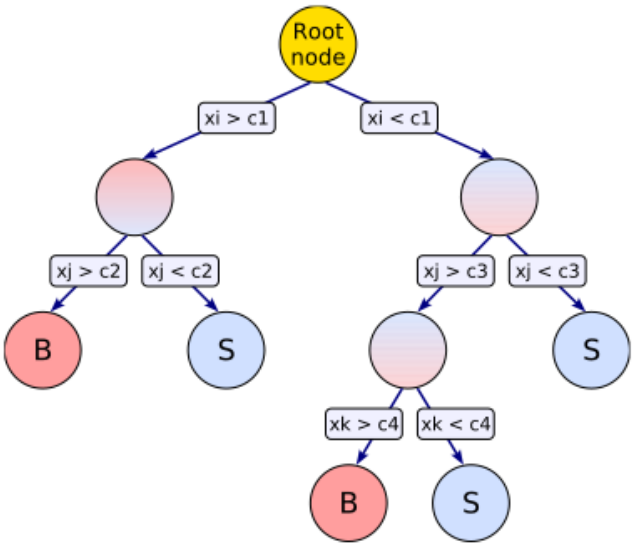


Figure 3.1: Example of decision tree structure.

3.3.2 Multilayer Perceptron (MLP)

An Artificial Neural Network is, in general terms, a simulated collection of interconnected neurons and each one of them, given a set of input signals, provides a certain response.

While a neural network with n neurons can have n^2 connections, the complexity can be reduced by organizing the neurons in layers and allowing connections only from a layer to the following one, as shown in figure 3.2. This kind of neural network is called multilayer perceptron. The first layer is called input layer and its number of neurons equals the number of the input variables. The last layer is the output layer with the number of neurons equal to the number of output variables. For the purpose of this analysis there will be only one output variable, the estimator of the neural network; all the others are hidden layers.

3.4 The TMVA for Opposite Side Tagging

The goal is to identify a set of characteristics of the opposite side muon that have informations about the the flavor of the B_S^0. Using a multivariate approach it is not only necessary to find the appropriate variables but also to choose a relation between the muon properties and the the flavor of the B_S^0 at production time. The best way is to define two classes based on the difference between the the flavor of the meson and the charge of the opposite side muon:

- the flavor of the B_S^0 opposite to the charge of the opposite side
- the flavor of the B_S^0 equal to the charge of the opposite side μ μ

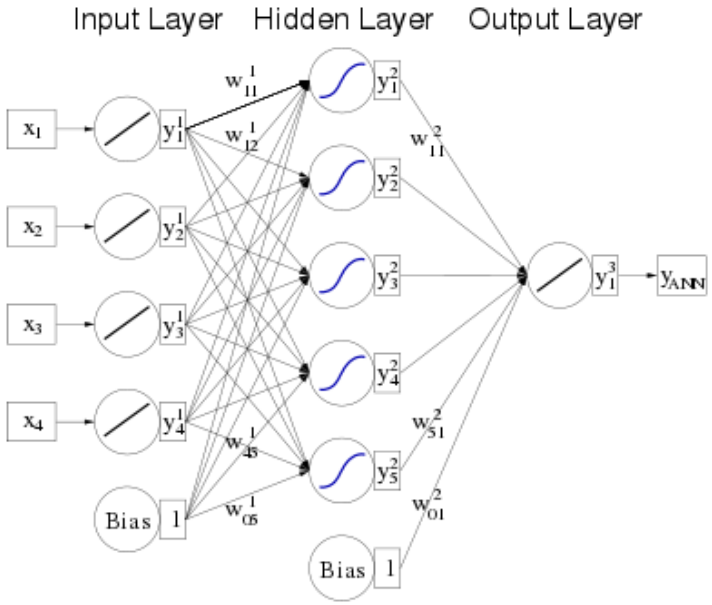


Figure 3.2: Example of multilayer perceptron structure

The TMVA is a good tool for separating two categories, but by default it considers one of the two categories to be signal and the other background. This can be misleading because in this analysis the two categories are not signal or background, and the goal is to divide them and use all the events. However for the TMVA purposes, signal is the first class (charge and flavor have opposite sign), background is the second class (charge and flavor have the same sign).

3.4.1 The choice of variables

We want to understand which of the characteristics of the opposite side muon are useful in the distinction between this cases. We have a set of variables going form the basic information of the muon (p_T, η, Impact Parameter d_{xyz} or d_{xy}, Charge) to information regarding the Jet of the muon, the isolation of the muon and the jet of highest p_T.

A simple method to decide if a choice of parameters offers a better classification than another is checking the background rejection versus signal efficiency also called ROC (Receiver Operating Characteristics) curve, of which an example is given in 3.3. In particular if the integral under the ROC curve is higher it means there is a higher separation between the two categories. This is a useful tool while comparing two results but the curve is calculated under the assumption that there is the same number of events in both the categories.

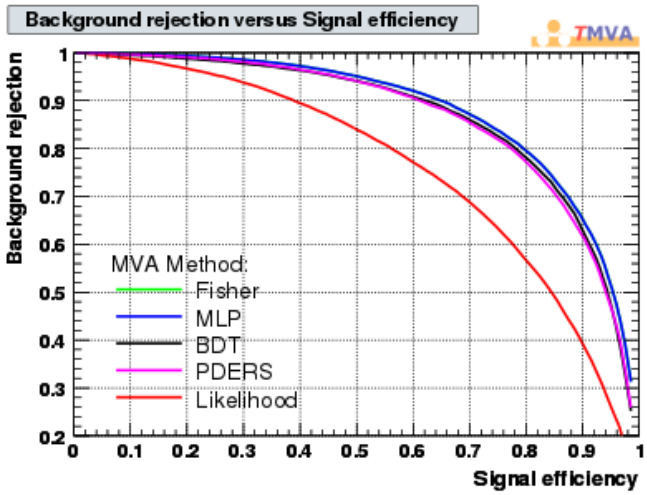


Figure 3.3: Example of ROC (Receiver Operating Characteristics) curves.

The charge

Ignoring cascade decays and the oscillation of B^0 and B_S^0 mesons, the flavor of the B and the charge of the muon would always be opposite. This is not the case, but the charge of the muon still appear to be a fair discriminating variable. Instead of using a discreet variable a continuous variable with similar meaning was opted for, the charge cone defined as:

$$\text{Charge Cone} = \frac{\sum_i q_i p_{Ti}^k}{\sum_i p_{Ti}^k} \tag{3.1}$$

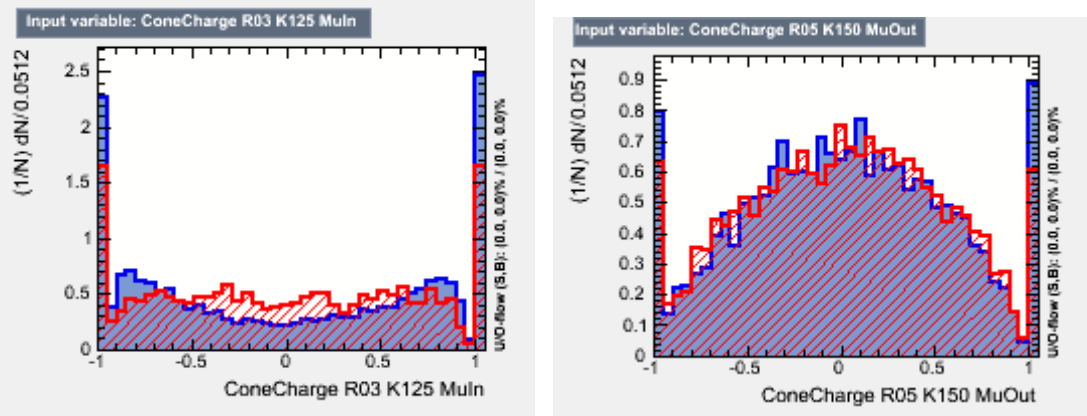
where i goes over the particles contained in a cone with ΔR (with respect to the muon) smaller than a certain value, q is the charge and p_{Ti} the transverse momentum of the particle. k is a parameter that represents the relevance of low p_T tracks in the Charge Cone. The charge cone can be calculated including or excluding the muon, as can be seen in 3.4.

While discussing the charge cones, values for ΔR and k had to be decided. It becomes clear that two charge cones, one with the muon and the other without it, are not very correlated ($\lesssim 15\%$).

These are the first two candidates for the variables to be used in the classification, and the decision of ΔR and k will be discussed in section 3.4.1.

The Impact Parameter

Searching for other variables it seemed reasonable to include the information about the muon itself. As we can see in figure 3.5 and 3.6, the variables don't appear to be very discriminating, but running trials with the TMVA methods it's clear that they're useful



(a) Charge Cone with muon

(b) Charge Cone without muon

Figure 3.4: Two examples of Charge Cones, in red the same sign class, in blu the opposite sign class.

to separate the two categories, in particular with the MLP method that gives high rating to p_T and the Impact Parameter. The only variable that won't be used in the final selection is $\Delta R(B)$ because in all the test that were done with different set of variables, adding $\Delta R(B)$ (with respect to the B) never improved the ROC integral.

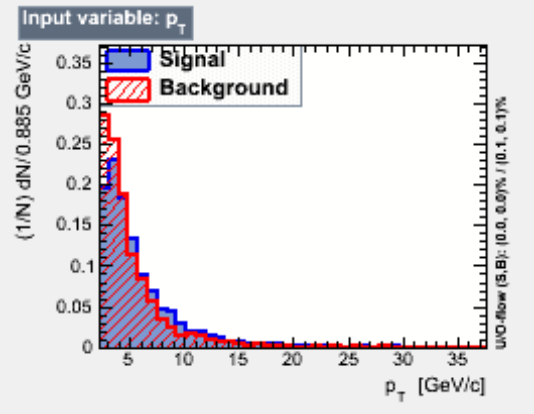
At this point a choice needs to be done: the 3D and 2D Impact Parameter (d_{xyz} and d_{xyz}) are heavily correlated. Although both BDT and MLP can deal with non linear correlations the results are clearly better when only one of them is used. Actually, instead of the impact parameter we can use the impact parameter significance (both 2D or 3D) defined as:

$$S_{xyz} = \frac{d_{xyz}}{\sigma_{d_{xyz}}} \qquad S_{xy} = \frac{d_{xy}}{\sigma_{d_{xy}}}$$

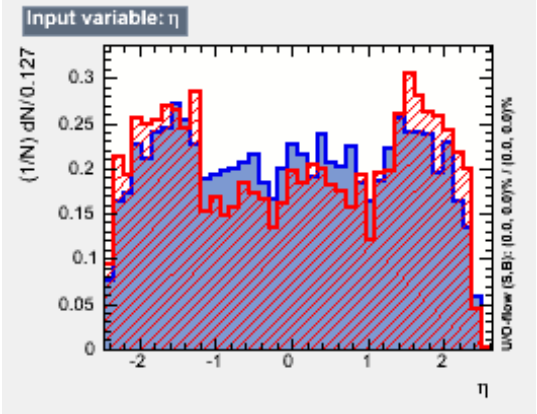
Their distribution can be seen in figure 3.6. Obviously S_{xyz} and S_{xy} are correlated between them and with the Impact Parameter, so only one of the four can be used. The correlation coefficients between 2D and 3D variables can be found in Tab 3.9. The one giving the best results in terms of separation of the categories is d_{xyz}.

Table 3.9: Linear correlation coefficients (%) between the impact parameter variables. Only the coefficients between 2D and 3D variables are displayed because only the pair with one 2D and one 3D variables were deemed interesting.

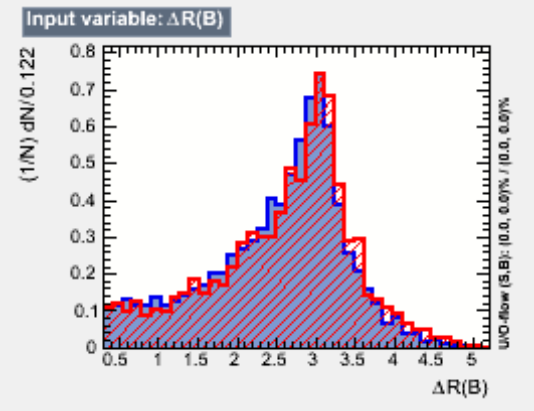
	d_{xyz}	d_{xy}	S_{xyz}	S_{xy}
d_{xyz}	/	76		64
d_{xy}	76	/	71	
S_{xyz}		71	/	84
S_{xy}	64		84	/



(a) p_T of the muon



(b) η of the muon



(c) $\Delta R(B)$ of the muon

Figure 3.5: Muon variables. In red the same sign class, in blue the opposite sign class.

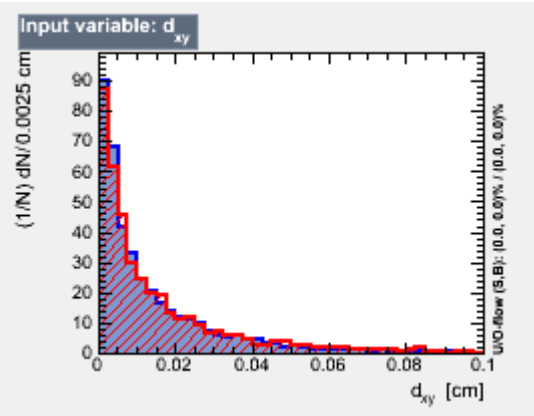
The Relative Isolation

Another interesting variable is the relative isolation of the muon reconstructed with the Particle Flow algorithm, defined as:

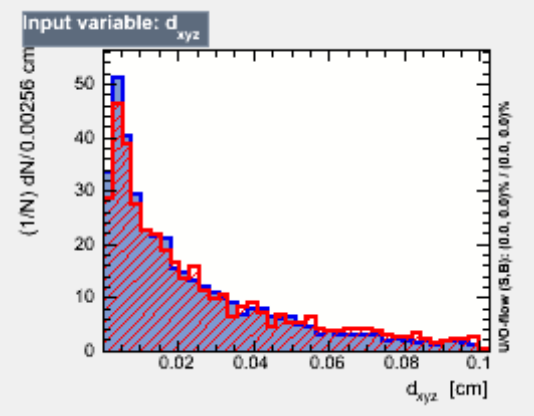
$$\text{PFIsolation} = \frac{\sum_i p_{Ti} + \sum_j E_{Tj} + \sum_k E_{Tk}}{p_T} \tag{3.2}$$

where i, j and k indicates particles with $\Delta R < 0.4$ from the muon. i runs over charged hadrons, j over neutral hadrons and k over photons, as illustrated in figure 3.7.

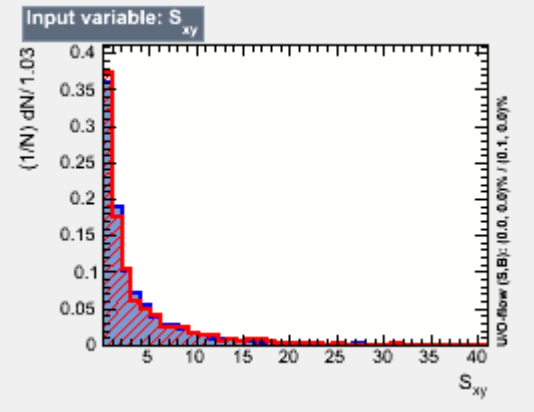
This is one of the variables with the highest ranking for the MLP categorization. The correlations between the variables introduced so far and a first result, both for BDT and MLP, is illustrated by fig 3.8 and 3.9.



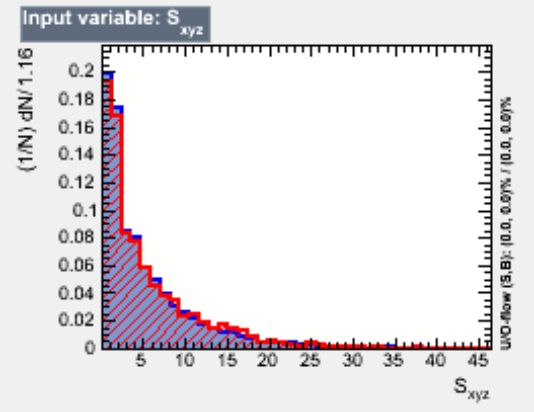
(a) 2D impact parameter



(b) 3D impact parameter



(c) S_{xy}



(d) S_{xyz}

Figure 3.6: Impact parameter variables. In red the same sign class, in blue the opposite sign class.

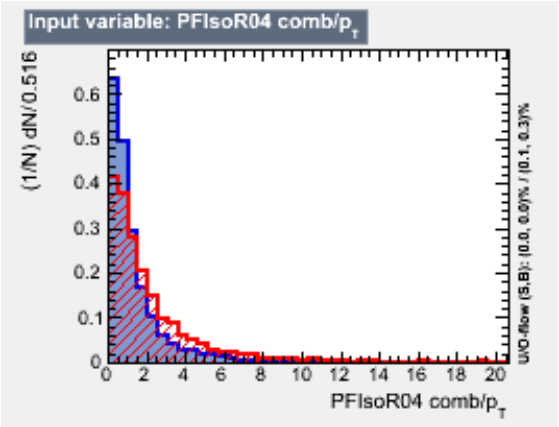
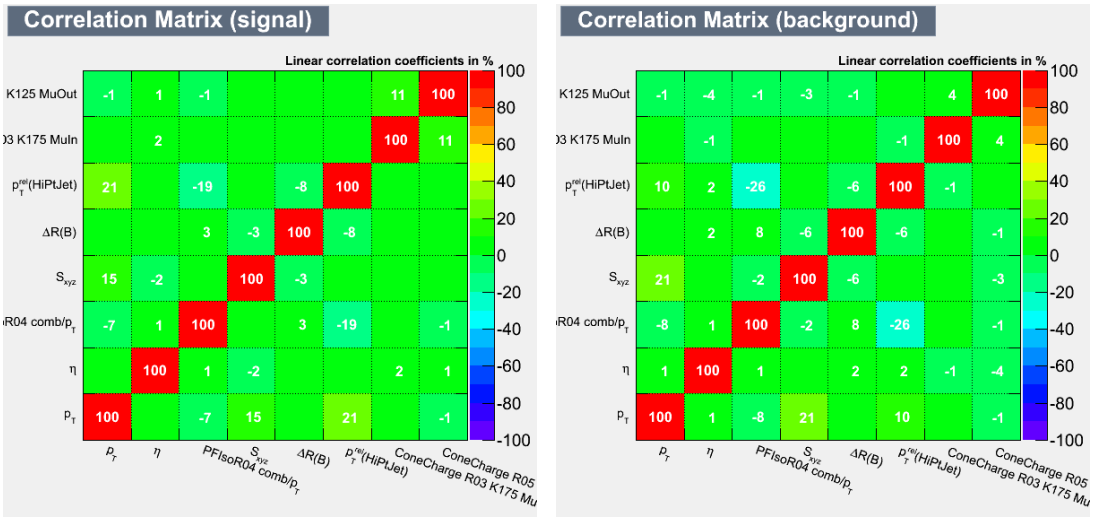
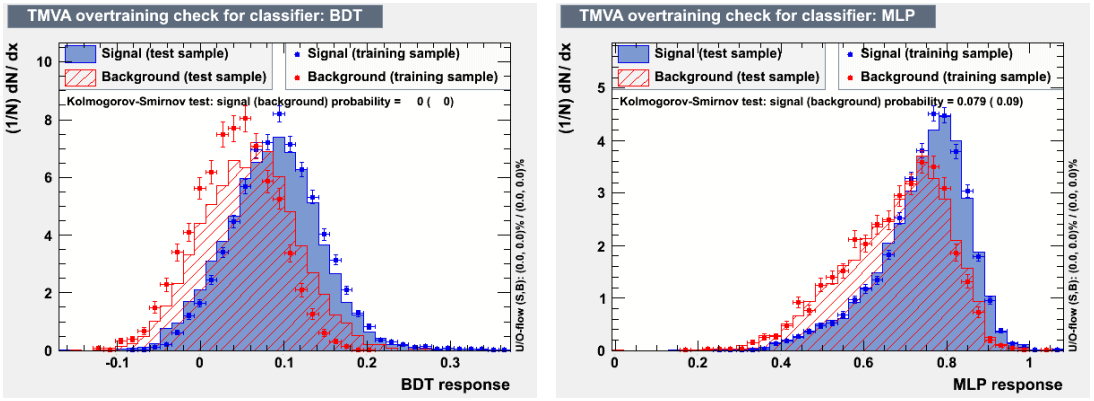


Figure 3.7: Relative Isolation. In red the same sign class, in blue the opposite sign class.



(a) Opposite sign class *(b)* Same sign class

Figure 3.8: Linear correlations matrices for the two classes with variables up to Relative Isolation



(a) BDT method *(b)* MLP method

Figure 3.9: Estimator after training and testing with variables up to Relative Isolation. In red the same sign class, in blue the opposite sign class.

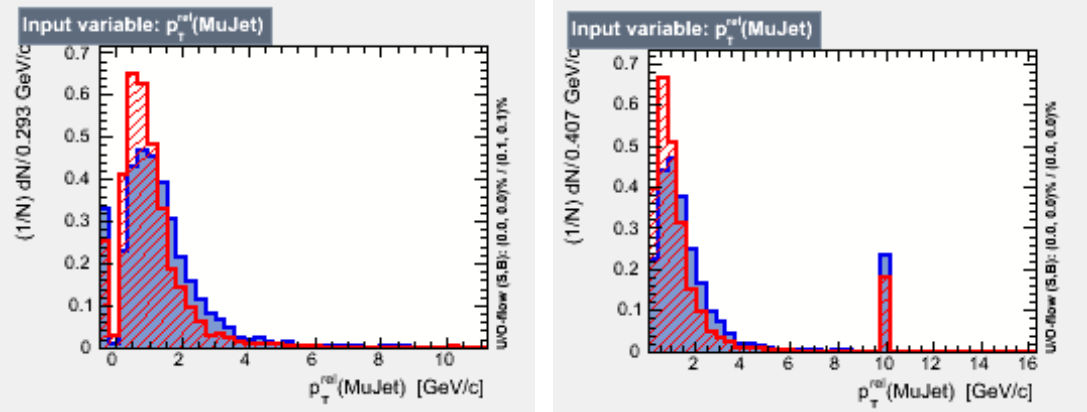
The muon's p_T^{rel}

One of the interesting variables with some discriminating power is the transverse momentum perpendicular to the direction of the jet of the muon. This variable is important for discriminating direct decays ($b \to \mu X$) from sequential decays ($b \to c \to \mu X'$) because direct muon has higher p_T^{rel} since it decays from a particle with higher mass and therefore has an higher kinematic energy limit.

Unfortunately loose cuts on the muon properties were requested (in order to obtain a bigger sample), so the muon is not always reconstructed in a jet. This means that in about 10% of the cases one of the variables is not defined. A temporary solution was to assign this events an arbitrary value outside the variable range, supposing that since in that peak the two classes where equally distributed, the methods would only use the rest of the range.

Two options were considered: placing this events before or after the range, as can be seen in 3.10(a) and 3.10(b). This assumption was not appropriate, and later on a more legitimate approach was implemented, as will be discussed in section 3.5.

Even just with this temporary solution the power of discrimination was slightly increased, as can be seen in figure 3.11 and 3.12.



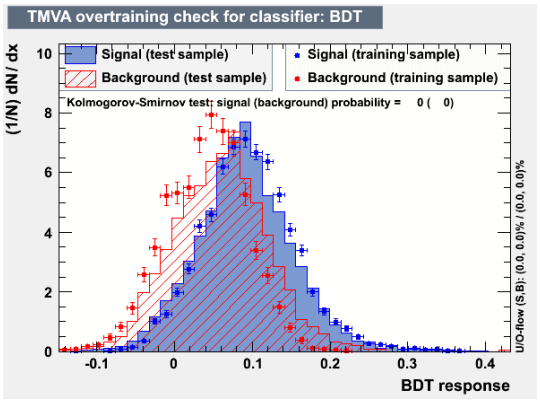
(a) p_T^{rel} with peak at -0.5 GeV

(b) p_T^{rel} with peak at 10 GeV

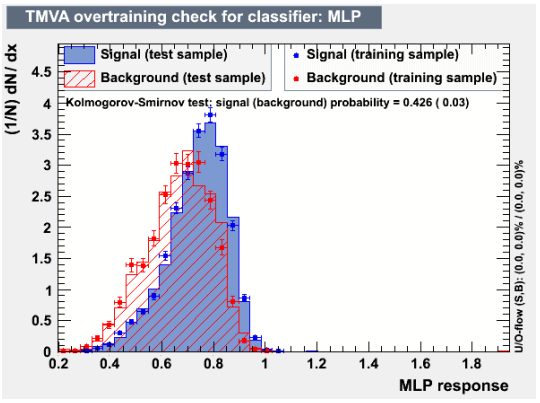
Figure 3.10: The two possibilities for p_T^{rel}. In red the same sign class, in blue the opposite sign class.

The Charge Cone

As we anticipated in paragraph 3.4.1, two charge cones will be used as training variables, one with the muon included and one with the muon excluded. There are two parameters that need to be chosen for each of them, k and ΔR from equation 3.1. While there were many possible combinations instead of checking all of them, the behaviour of the ROC integral under reasonable changes was studied. It is relevant that if there aren't any traces with ΔR less than the stated value, the charge cone can't be created. This is not

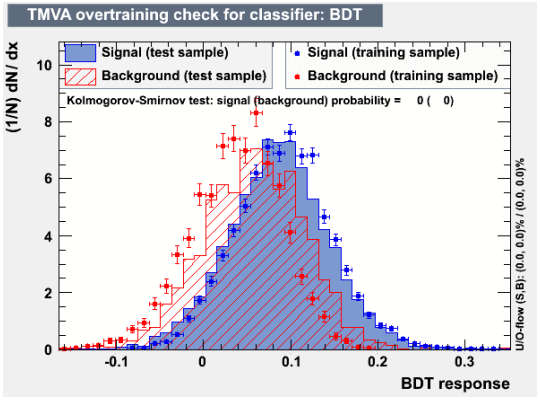


(a) BDT method

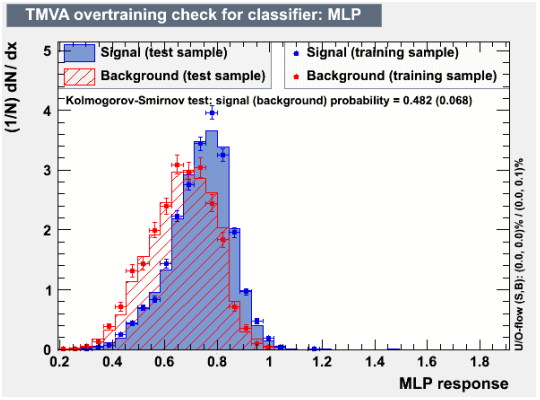


(b) MLP method

Figure 3.11: Estimator after training and testing with p_T^{rel} with peak at -0.5 GeV. In red the same sign class, in blue the opposite sign class.



(a) BDT method



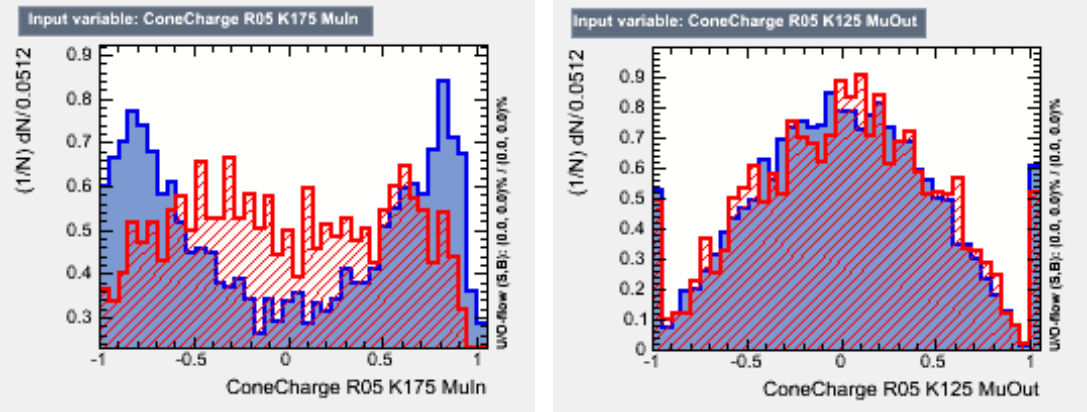
(b) MLP method

Figure 3.12: Estimator after training and testing with p_T^{rel} with peak at 10 GeV. In red the same sign class, in blue the opposite sign class.

a variable that is always presents but varying ΔR and k the efficiencies go from 89% to 100%.

The difference was $\sim 0.8\%$ of the ROC integral, so further studies weren't pursued and the variables with better results from the tests already done were chosen.

The final variables have $\Delta R < 0.5$, the charge cone with the muon has $k = 1.75$, while the other has $k = 1.25$, and can be seen in figure 3.13.



(a) Charge cone with muon

(b) Charge cone without muon

Figure 3.13: Charge Cones choosed after the optimization. In red the same sign class, in blue the opposite sign class.

3.4.2 Results

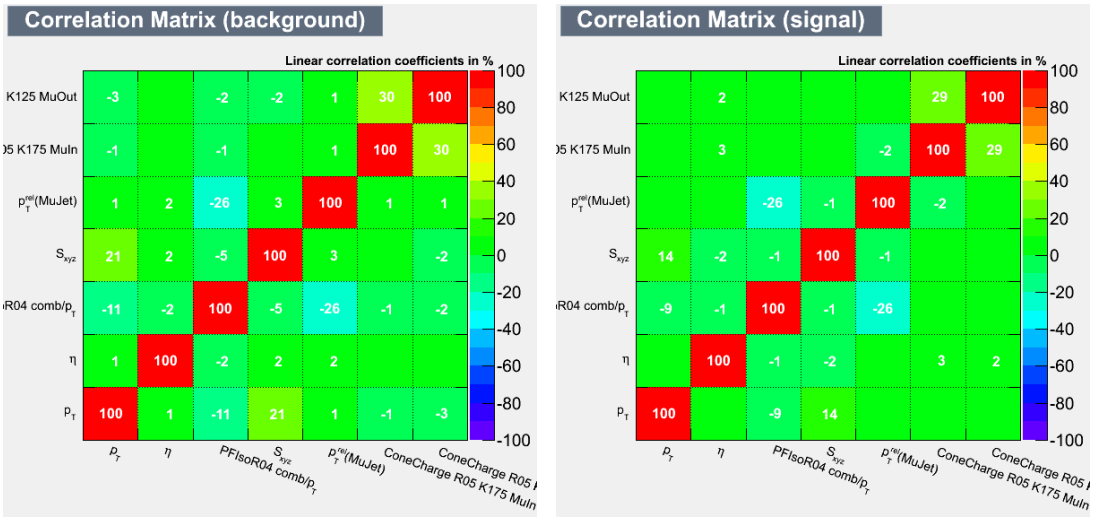
The chosen variables are:

- p_T of the muon
- η of the muon
- Relative Isolation
- S_{xyz}
- p_T^{rel} with the muon jet
- Charge Cone without muon $\Delta R = 05$ and $k = 1.25$
- Charge Cone with muon $\Delta R = 03$ and $k = 1.75$

The correlation between the variables can be seen in figure 3.14, and the result is presented in figure 3.15.

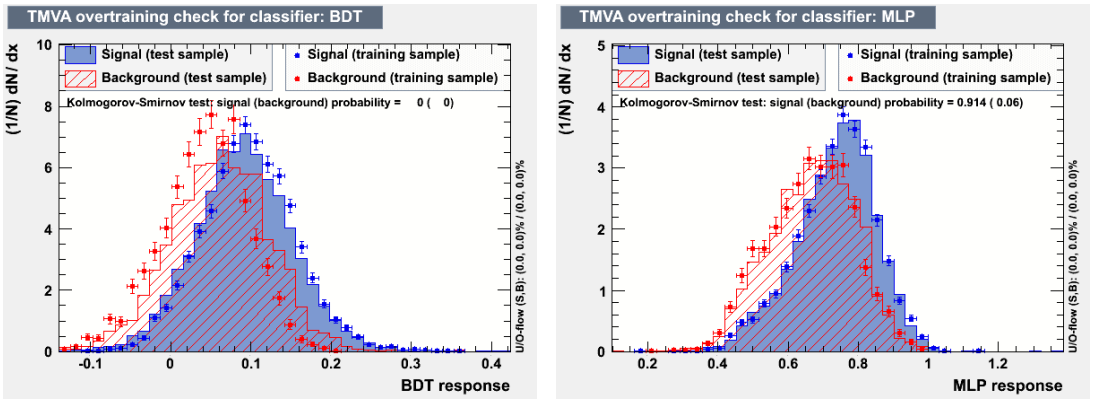
In the end, there isn't a huge difference in the results using BDT and MLP, the main reason why the MLP over the BDT was chosen is based on the agreement between the training and the test sample. This is given by the Kolmogorov-Smirnov test. As it can be seen from fig 3.15(a) the result of test for the BDT is zero. This means that the agreement between the two samples is very low. The MLP, on the other side, has a good agreement as can be seen in figure 3.15(b). This could be due to the small number of events in the sample.

The MLP method seemd the better option and from this point it will be the used method.



(a) Opposite sign class *(b)* Same sign class

Figure 3.14: Linear correlation coefficients between the final set of variables.



(a) BDT method *(b)* MLP method

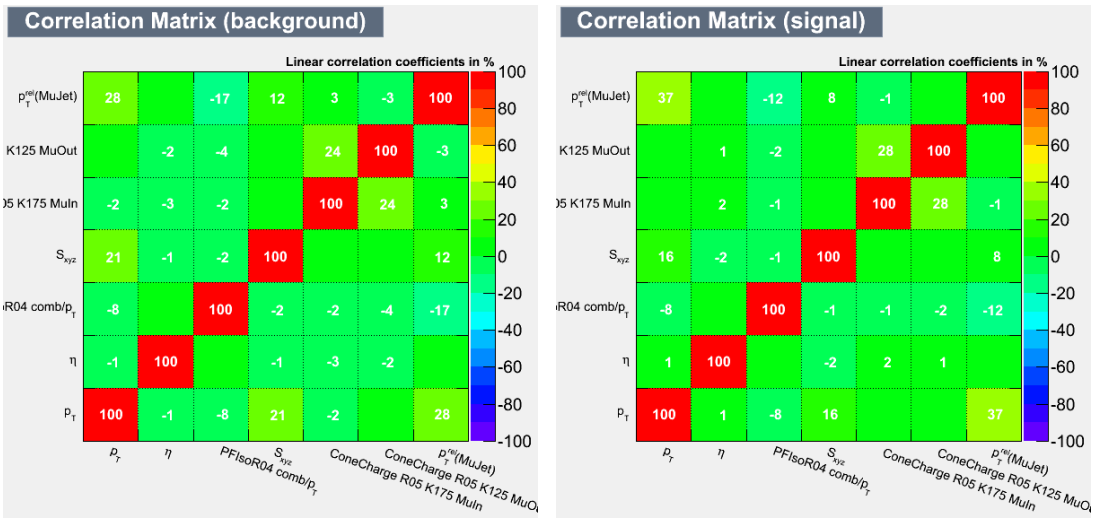
Figure 3.15: Estimator after training and testing with the final set of variables. In red the same sign class, in blue the opposite sign class.

3.5 Classification using categories

As we already anticipated there is another way to do the classification when we have a variable that is not defined for all the events. There are various way to combine different MVA methods, one of this is the category method that consists of dividing the events in disjoint subsamples with different properties. In each region an independent training is performed with the appropriate method and variables but with the advantage of an unique output function. In our analysis we considered two subsamples, one for the events

with the reconstructed muon's jet, the other for the events without a muon jet. In both cases we used the MLP method for its better performances, as described previously.

The power of discrimination with this new method increases a little as we can see in figure 3.16 and 3.17.



(a) Opposite sign class

(b) Same sign class

Figure 3.16: Linear correlation matrices with final variables for cathegory method

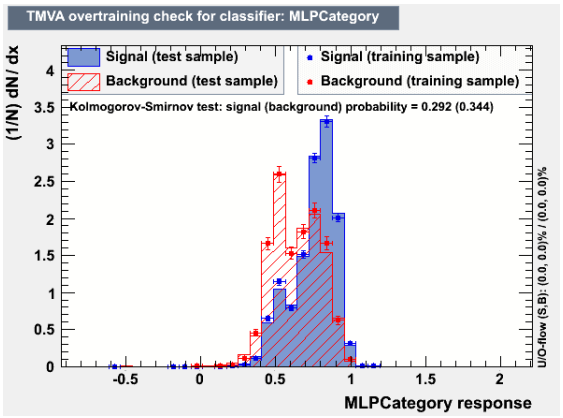


Figure 3.17: Estimator after training and testing with cathegory method. In red the same sign class, in blue the opposite sign class.

The MLP method has a lot of options that were studied to optimize our discriminator; a detailed account of the process can be found in appendix A.

3.6 Development of the OS lepton tagger

After optimizing the muon tagger, the weights were Applied to the same $B_S^0 \to J/\psi\phi$ sample in order to evaluate the tagging power as exposed in section 1.3. First of all a cut in the MLP discriminator was optimized to obtain a subsample enriched in correctly tagging events to optimize the tagging power

$$\mathcal{P}_{tag} = \varepsilon_{tag}\mathcal{D}^2 = \varepsilon_{tag}(1-2\omega)^2$$

that depends both on the efficiency ε_{tag} and of the mistag fraction ω. For the muon the cut is MLP$>$ 0.64 that corresponds to an efficiency of $\sim$ 75% with respect to the OS-muon efficiency. The result can be seen in table 3.10.

Table 3.10: Muon tagging performances evaluated on the $B_S^0 \to J/\psi\phi$, $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^*$ simulated samples, and on the $B^+ \to J/\psi K^+$ channel of the 2012 Data. A cut on the MLP discriminator MLP $>$ 0.64 is applied.

[%]	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC	$B^+ \to J/\psi K^+$ Data
MLP		> 0.64		
ε_{tag}	2.81 ± 0.02	2.86 ± 0.02	2.92 ± 0.03	3.34 ± 0.02
ω	23.1 ± 0.3	23.5 ± 0.4	23.5 ± 0.4	25.9 ± 0.4
$\mathcal{P}_{tag}$	0.82 ± 0.02	0.80 ± 0.02	0.82 ± 0.02	0.78 ± 0.03

An improved tagging performance can be obtained without reducing the efficiency by segmenting the samples in four bins of the MLP variable and calculating the tagging power and the wrong tag fraction for each of them independently. The overall tagging power is the sum of the tagging powers for the single categories. The single results and a global tagging power are presented in table 3.11

Table 3.11: Muon tagging performances evaluated on the $B_S^0 \to J/\psi\phi$ simulated sample. Four categories of MLP discriminator are used.

[%]	$B_S^0 \to J/\psi\phi$ 2012			
MLP	$[-0.05, 0.6405)$	[0.6405, 0.7405)	[0.7405, 0.815)	[0.815, 1.35)
ε_{tag}	0.932 ± 0.012	0.938 ± 0.013	0.936 ± 0.013	0.929 ± 0.012
ω	47.243 ± 0.672	30.725 ± 0.623	22.801 ± 0.569	15.515 ± 0.496
$\mathcal{P}_{tag}$	0.003 ± 0.001	0.139 ± 0.009	0.277 ± 0.012	0.442 ± 0.014
ε_{tag}^{tot}		3.74 ± 0.03		
$\mathcal{P}_{tag}^{tot}$		0.86 ± 0.02		

Later the per-event mistag was calculated measuring the mistag in 20 bins of the MLP discriminator as can be seen in figure 3.18 and interpolating the results with a curve. Various shapes were tried, using functions of tanh, arctan and of the error function (Erf). In the end the function that return the best agreement with the events was

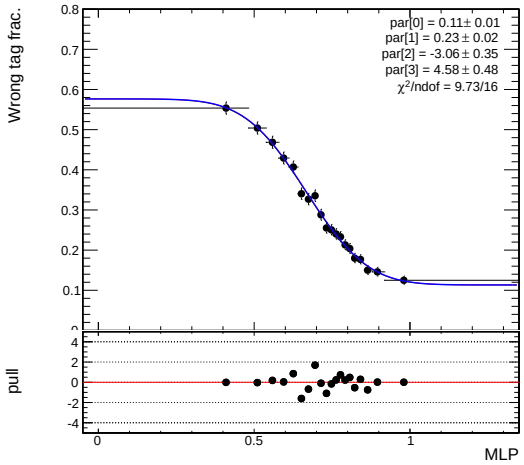


Figure 3.18: 20 bins of muon MLP fitted to obtain the per-event mistag in $B_S^0 \to J/\psi\phi$

$$f = p_0 + p_1(1 - \mathrm{Erf}(p_2 + p_3 \cdot MLP)) \tag{3.3}$$

the parameters obtained from the fit can be found in table 3.12

Table 3.12: Parameters resulting from the fit to the ω distribution in the $B_S^0 \to J/\psi\phi$ MC

parameters	$B_S^0 \to J/\psi\phi$ MC
p_0	0.11 ± 0.01
p_1	0.23 ± 0.02
p_2	-3.1 ± 0.4
p_3	4.6 ± 0.5

A similar process was followed for the electron tagger choosing the variables that work optimally for this particular multivariate discriminator:

- p_T
- η
- PFIsolation
- d_{xyz}
- PFmva
- Electron Charge Cone with $\Delta R = 0.3$, $k = 1.75$ and electron included

where PFmva is an electron identification criterion variable, optimized to discriminate electrons from background sources, such as pions [11] and the Electron Charge Cone has basically the same definition of the Muon Charge Cone.

The MLP variable obtained for the electron tagger can be found in figure 3.19.

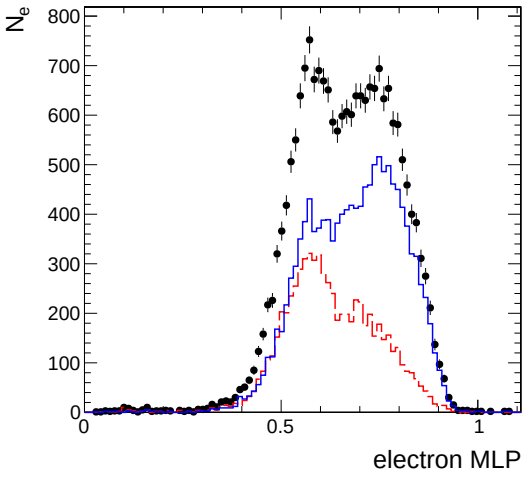


Figure 3.19: Electron MLP distribution for $B_s^0 \to J/\psi\phi$

As for the muons first of all a cut is defined to optimize the tagging power by ignoring the events with the wrong charge-flavor relationship. The cut is electron MLP > 0.64 that has an efficiency of $\sim 57\%$. The efficiency of the tagger becomes then $\sim 1.9\%$; the $\mathcal{P}_{tag}$ can be seen in table 3.13

Table 3.13: Electron tagging performances evaluated on the $B_S^0 \to J/\psi\phi$; a cut on the MLP discriminator MLP > 0.64 is applied.

[%]	$B_S^0 \to J/\psi\phi$ MC
ε_{tag}	1.90 ± 0.02
ω	25.5 ± 0.4
$\mathcal{P}_{tag}$	0.45 ± 0.02

The tagging power is then increased by dividing the whole sample in four categories of the electron MLP discriminator and calculating the wrong tag and tagging power of the single categories independently to obtain the total tagging power, as can bee seen in table 3.14.

A per-event mistag was then calculated dividing the electron MLP range in 20 bins and interpolating them with the error function see in equation 3.3 as can be seen in figure 3.20. The parameters can be found in table 3.15

Table 3.14: Electron tagging performances evaluated on the $B_S^0 \to J/\psi\phi$ simulated sample. Four categories of MLP discriminator are used.

[%]	$B_S^0 \to J/\psi\phi$ MC			
MLP	$[0, 0.573)$	$[0.573, 0.6675)$	$[0.6675, 0.761)$	$[0.761, 1.1)$
ε_{tag}	0.838 ± 0.012	0.839 ± 0.012	0.835 ± 0.012	0.832 ± 0.012
ω	47.687 ± 0.709	40.172 ± 0.698	29.434 ± 0.653	19.617 ± 0.574
$\mathcal{P}_{tag}$	0.002 ± 0.001	0.032 ± 0.005	0.141 ± 0.009	0.307 ± 0.012
ε_{tag}^{tot}	3.34 ± 0.02			
$\mathcal{P}_{tag}^{tot}$	0.48 ± 0.02			

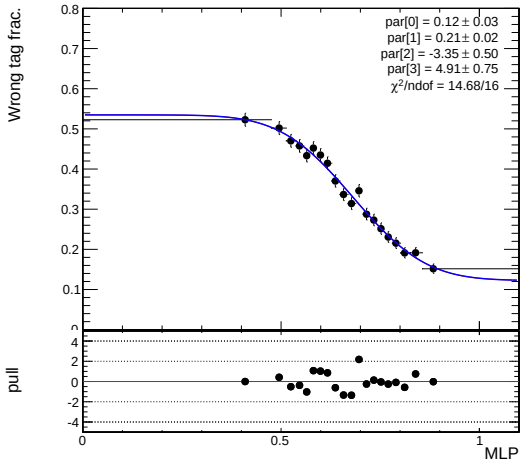


Figure 3.20: 20 bins of electron MLP for per-event mistag in $B_S^0 \to J/\psi\phi$

Table 3.15: Parameters resulting from the fit to the ω distribution in the $B_S^0 \to J/\psi\phi$ MC; a function of the form $p_0 + p_1 \cdot [1 - Erf(p_2 + p_3 \cdot MLP)]$ is used.

parameters	$B_S^0 \to J/\psi\phi$ MC
p_0	0.12 ± 0.03
p_1	0.21 ± 0.02
p_2	-3.4 ± 0.5
p_3	4.9 ± 0.8

3.7 OS lepton tagger test in B^+ and B^0 samples

The same weights obtained after training the muon and the electron tagger in the $B_S^0 \to J/\psi\phi$ sample were applied to the $B^+ \to J/\psi K^+$ Data and MC samples and the $B^0 \to J/\psi K^*$ MC sample.

The same procedure discussed in the previous section was employed to determine the tagging power of the taggers on these samples and compare the results on the MC samples.

Muon tagger The cut on the muon MLP variable was compared first, the results can be seen in table 3.16 along with the B_S^0 already presented, the agreement on the MC samples is very good.

Table 3.16: Muon tagging performances evaluated on the $B_S^0 \to J/\psi\phi$, $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^*$ simulated samples, and on the $B^+ \to J/\psi K^+$ channel of the 2012 Data. A cut on the MLP discriminator MLP > 0.64 is applied.

[%]	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC	$B^+ \to J/\psi K^+$ Data
MLP		> 0.64		
ε_{tag}	2.81 ± 0.02	2.286 ± 0.02	2.92 ± 0.03	3.34 ± 0.02
ω	23.1 ± 0.3	23.5 ± 0.4	23.5 ± 0.4	25.9 ± 0.4
$\mathcal{P}_{tag}$	0.82 ± 0.02	0.80 ± 0.02	0.82 ± 0.02	0.78 ± 0.03

The muon MLP variable was then divided in four bins each with his own efficiency and mistag in order to calculate a global $\mathcal{P}_{tag}$, the results can be seen in table 3.17 3.18 3.19 as well as 3.16 for the B_S^0 already presented. Once again the results are compatible between the MC samples.

The variable is then divided in 20 bins to be interpolated and to create the per-event mistag, as can be seen in figure 3.21. The results of the fit can be found in table 3.20.

The mistag obtained from the fit can be referred to as *calculated mistag* ω_{cal}, for a self tagging sample, the B^+ sample, we can also have a *measured mistag* ω_{mis} in bins of

Table 3.17: Muon tagging performances evaluated on the reconstructed $B^+ \to J/\psi K^+$ events 2012 data. Four categories of MLP discriminator are used.

[%]	$B^+ \to J/\psi K^+$ 2012 Data			
MLP	$[-0.05, 0.6405)$	$[0.6405, 0.7405)$	$[0.7405, 0.815)$	$[0.815, 1.35)$
ε_{tag}	1.219 ± 0.014	1.177 ± 0.014	1.130 ± 0.013	1.027 ± 0.013
ω	44.905 ± 0.918	33.267 ± 0.774	24.316 ± 0.660	19.313 ± 0.608
$\mathcal{P}_{tag}$	0.013 ± 0.005	0.132 ± 0.013	0.298 ± 0.017	0.387 ± 0.018
ε_{tag}^{tot}	4.55 ± 0.03			
$\mathcal{P}_{tag}^{tot}$	0.83 ± 0.03			

Table 3.18: Muon tagging performances evaluated on the $B^+ \to J/\psi K^+$ simulated sample. Four categories of MLP discriminator are used.

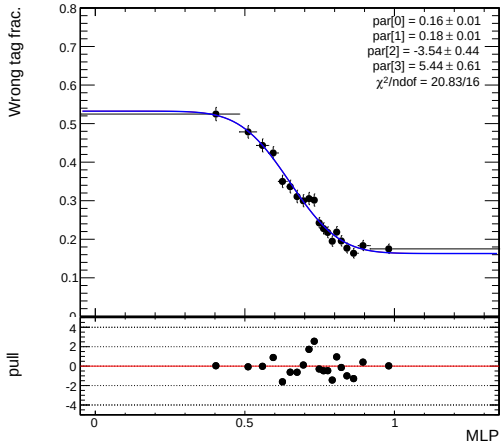
[%]	$B^+ \to J/\psi K^+$ 2012			
MLP	$[-0.05, 0.6405)$	$[0.6405, 0.7405)$	$[0.7405, 0.815)$	$[0.815, 1.35)$
ε_{tag}	0.991 ± 0.014	0.994 ± 0.014	0.952 ± 0.014	0.951 ± 0.014
ω	42.413 ± 0.698	30.668 ± 0.671	22.866 ± 0.661	16.861 ± 0.548
$\mathcal{P}_{tag}$	0.023 ± 0.004	0.141 ± 0.010	0.280 ± 0.013	0.418 ± 0.015
ε_{tag}^{tot}	3.84 ± 0.03			
$\mathcal{P}_{tag}^{tot}$	0.86 ± 0.02			

Table 3.19: Muon tagging performances evaluated on the $B^0 \to J/\psi K^*$ simulated sample. Four categories of MLP discriminator are used.

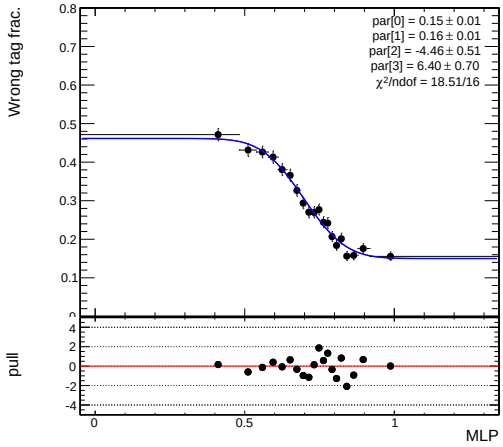
[%]	$B^0 \to J/\psi K^*$ 2012			
MLP	$[-0.05, 0.6405)$	$[0.6405, 0.7405)$	$[0.7405, 0.815)$	$[0.815, 1.35)$
ε_{tag}	1.004 ± 0.015	0.936 ± 0.014	0.987 ± 0.015	0.994 ± 0.015
ω	44.216 ± 0.715	31.139 ± 0.715	$22.030 \pm 0.0.626$	17.882 ± 0.579
$\mathcal{P}_{tag}$	0.013 ± 0.003	0.133 ± 0.010	0.309 ± 0.015	0.410 ± 0.016
ε_{tag}^{tot}	3.92 ± 0.02			
$\mathcal{P}_{tag}^{tot}$	0.87 ± 0.02			

Table 3.20: Parameters resulting from the fit to the ω distribution in the $B_S^0 \to J/\psi\phi$ MC, $B^+ \to J/\psi K^+$ MC, $B^0 \to J/\psi K^*$ MC and in 2012 $B^+ \to J/\psi K^+$ data. A function of the form $p_0 + p_1 \cdot [1 - Erf(p_2 + p_3 \cdot MLP)]$ is used.

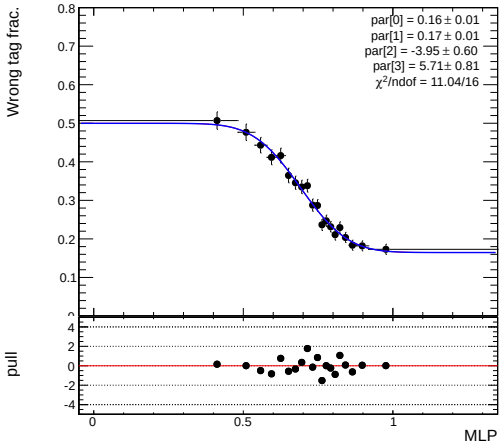
par	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC	$B^+ \to J/\psi K^+$ Data
p_0	0.11 ± 0.01	0.15 ± 0.01	0.16 ± 0.01	0.16 ± 0.01
p_1	0.23 ± 0.02	0.16 ± 0.01	0.18 ± 0.01	0.17 ± 0.1
p_2	-3.1 ± 0.4	-4.5 ± 0.5	-3.5 ± 0.4	-4.0 ± 0.6
p_3	4.6 ± 0.5	6.4 ± 0.7	5.4 ± 0.6	5.7 ± 0.8



(a) $B^0 \to J/\psi K^*$ MC



(b) $B^+ \to J/\psi K^+$ MC



(c) $B_S^0 \to J/\psi\phi$ Data

Figure 3.21: 20 bins of muon MLP for per-event mistag.

ω_{cal}. The relation should be linear in absence of biases in the definition of the mistag function and can be written as

$$\omega_{mis}(\omega_{cal}) = p_0 + p_1(\omega_{cal} - \omega')$$

where $\omega' = 0.35$ is a fixed parameter introduced to reduce the correlation between p_0 and p_1. In absence of bias we expect p_1 compatible with 1 and p_0 compatible with ω', and in this case the ω_{cal} should be corrected according to the resulting ω measured on data.

In order to extract p_0 and p_1, the measured mistag fraction is evaluated from fit to the B^+ mass in twenty bins of ω_{cal} and the resulting graph can be seen in figure 3.22. The parameter of the fit can be found in table 3.21

Table 3.21: Calibration parameters of the OS-μ tagger. The uncertainties are statistical only.

p_0	0.350 ± 0.004
p_1	1.00 ± 0.04
$corr(p_0, p_1)$	-0.642
$p_0 - p_1 \times \omega'$	-0.001 ± 0.011

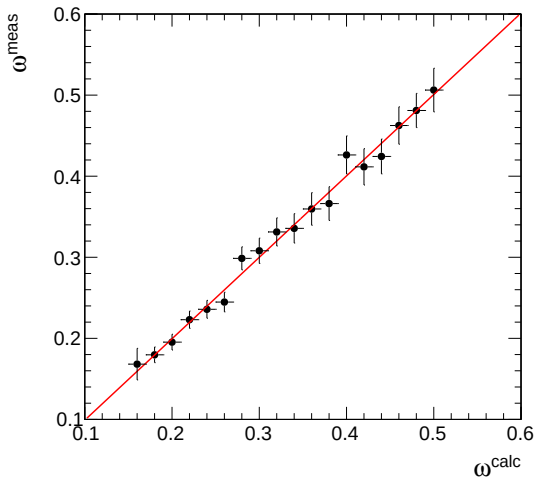


Figure 3.22: Calibration curve for the muon for the B^+ Data sample

Electron tagger For the simple cut MLP > 0.64 the tagging powers can be found in table 3.22 and the agreement is good ($< 3\sigma$).

For the classification in four bin of electron MLP, each with its independent efficiency and wrong tag, summed to calculate the total tagging power, the results can be found 3.23 3.24 3.25. The agreement between the MC samples is good.

Table 3.22: Electron tagging performances evaluated on the $B_S^0 \to J/\psi\phi$, $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^*$ simulated samples, and on the $B^+ \to J/\psi K^+$ channel of the 2012 Data. A cut on the MLP discriminator MLP > 0.64 is applied.

[%]	$B_S^0 \to J/\psi\phi$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC	$B^+ \to J/\psi K^+$ Data
MLP		> 0.64		
ε_{tag}	1.90 ± 0.02	1.91 ± 0.02	1.96 ± 0.02	2.11 ± 0.02
ω	25.5 ± 0.4	26.4 ± 0.5	25.5 ± 0.5	27.5 ± 0.5
$\mathcal{P}_{tag}$	0.45 ± 0.02	0.42 ± 0.02	0.47 ± 0.02	0.43 ± 0.02

Table 3.23: Electron tagging performances evaluated on the reconstructed $B^+ \to J/\psi K^+$ events 2012 data. Four categories of MLP discriminator are used.

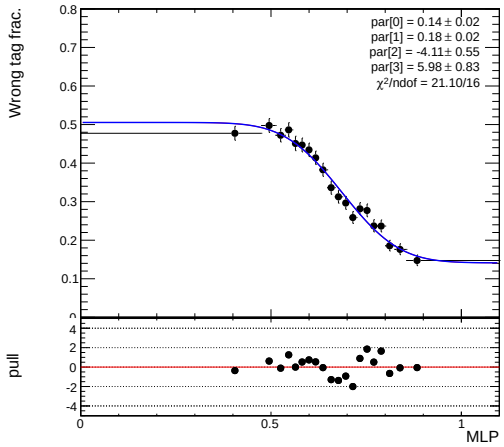
[%]	$B^+ \to J/\psi K^+$ Run2012 Data			
MLP	$[0, 0.573)$	$[0.573, 0.6675)$	$[0.6675, 0.716)$	$[0.716, 1.1)$
ε_{tag}	1.003 ± 0.013	1.018 ± 0.013	0.978 ± 0.012	0.871 ± 0.012
ω	47.693 ± 1.054	40.395 ± 0.94	30.162 ± 0.803	21.733 ± 0.704
$\mathcal{P}_{tag}$	0.002 ± 0.002	0.038 ± 0.008	0.154 ± 0.014	0.279 ± 0.016
ε_{tag}^{tot}	3.87 ± 0.03			
$\mathcal{P}_{tag}^{tot}$	0.47 ± 0.02			

Table 3.24: Electron tagging performances evaluated on the $B^+ \to J/\psi K^+$ simulated sample. Four categories of MLP discriminator are used.

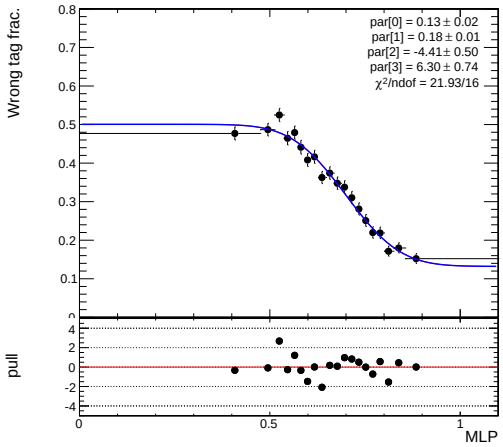
[%]	$B^+ \to J/\psi K^+$ MC			
MLP	$[0, 0.573)$	$[0.573, 0.6675)$	$[0.6675, 0.716)$	$[0.716, 1.1)$
ε_{tag}	0.876 ± 0.013	0.871 ± 0.013	0.831 ± 0.013	0.827 ± 0.013
ω	48.666 ± 0.75	39.89 ± 0.74	30.613 ± 0.716	18.791 ± 0.614
$\mathcal{P}_{tag}$	0.001 ± 0.001	0.036 ± 0.005	0.125 ± 0.009	0.322 ± 0.014
ε_{tag}^{tot}	3.41 ± 0.03			
$\mathcal{P}_{tag}^{tot}$	0.48 ± 0.02			

Table 3.25: Electron tagging performances evaluated on the $B^0 \to J/\psi K^*$ simulated sample. Four categories of MLP discriminator are used.

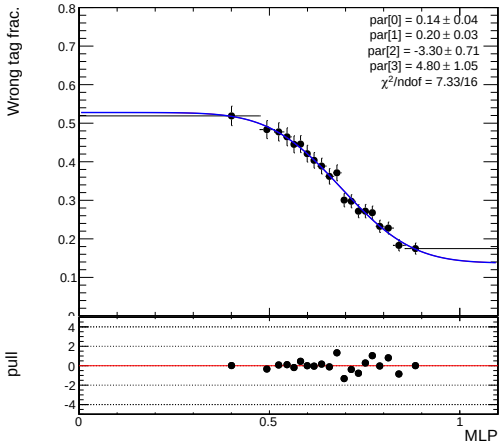
[%]	$B^0 \to J/\psi K^0$ MC			
MLP	$[0, 0.573)$	$[0.573, 0.6675)$	$[0.6675, 0.716)$	$[0.716, 1.1)$
ε_{tag}	0.856 ± 0.014	0.892 ± 0.014	0.879 ± 0.014	0.842 ± 0.014
ω	47.709 ± 0.801	40.325 ± 0.772	28.499 ± 0.72	19.717 ± 0.653
$\mathcal{P}_{tag}$	0.002 ± 0.001	0.033 ± 0.005	0.163 ± 0.011	0.309 ± 0.014
ε_{tag}^{tot}	3.47 ± 0.03			
$\mathcal{P}_{tag}^{tot}$	0.51 ± 0.02			



(a) $B^0 \to J/\psi K^*$ MC



(b) $B^+ \to J/\psi K^+$ MC



(c) $B^+ \to J/\psi K^+$ Data

Figure 3.23: 20 bins of electron MLP for per-event mistag

The MLP variable is then divided in 20 bins to be interpolated and to create the per-event mistag, as can be seen in figure 3.23. The results of the fit can be found in table 3.26.

Table 3.26: Parameters resulting from the fit to the ω distribution in the $B_S^0 \to J/\psi\phi$ MC, $B^+ \to J/\psi K^+$ MC, $B^0 \to J/\psi K^*$ MC, and in 2012 $B^+ \to J/\psi K^+$ data. A function of the form $p_0 + p_1 \cdot [1 - Erf(p_2 + p_3 \cdot MLP)]$ is used.

par	$B^0 \to J/\psi K^0$ MC	$B^+ \to J/\psi K^+$ MC	$B^0 \to J/\psi K^*$ MC	$B^+ \to J/\psi K^+$ Data
p_0	0.12 ± 0.03	0.13 ± 0.02	0.14 ± 0.02	0.14 ± 0.02
p_1	0.21 ± 0.02	0.18 ± 0.01	0.18 ± 0.01	0.20 ± 0.03
p_2	-3.4 ± 0.5	-4.4 ± 0.5	-4.1 ± 0.6	-3.3 ± 0.7
p_3	4.9 ± 0.8	6.3 ± 0.7	6.0 ± 0.8	4.8 ± 1.1

Once again the interpolating curve can be interpreted as *calculated mistag*, ω_{cal}, and for the B^+ sample a calibration curve can be derived from a graph of ω_{cal} in function of the ω_{mis}. The measured mistag fraction is evaluated from fit to the B^+ mass in twenty bins of ω_{cal} and the resulting graph can be seen in figure 3.24. The parameter of the fit, p_0 and p_1, can be found in table 3.27

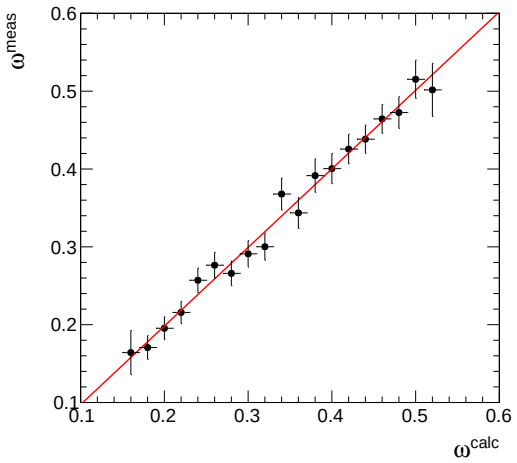


Figure 3.24: Calibration curve for the electron for the B^+ Data sample

Table 3.27: Calibration parameters of the OS-e tagger. The uncertainties are statistical only.

p_0	0.350 ± 0.004
p_1	1.01 ± 0.04
ρ	0.317
$p_0 - p_1 \times \omega'$	-0.004 ± 0.013

Chapter 4

$B^0 \to J/\psi K^*$ oscillation

4.1 Using the oscillation to measure w_{tag}

A measurement of B^0 can be used to cross check the performance of the OS lepton tagger, using a Same Side tagger to determine the flavor of the B^0 at the decaying time.

As was calculated in section 1.1.3, the time integrated mixing probability for the B^0, χ_d is defined as

$$\chi_d = \frac{\int_0^{+\infty} f^M(t)}{\int_0^{+\infty} f^U(t) + \int_0^{+\infty} f^M(t)}$$

where $f^M(t) = \Gamma e^{-t\Gamma}[1 - \cos(\Delta m\, t)]$ and $f^U(t) = \Gamma e^{-t\Gamma}[1 + \cos(\Delta m\, t)]$. If there are no tagging issues, χ_d can be measured experimentally as

$$\chi_d = \frac{N_M}{N_M + N_U}$$

where N_M and N_U are the number of mixed and unmixed events.

In the examined case, if we consider only the mistag induced by the lepton tagger, we have that the charge of the lepton is the same as the flavor of the B^0 if

- the $B^0\overline{B^0}$ mixing happened, and the event was correctly tagged, with a frequency of $f_{SCm} = \chi_d(1 - \omega_{tag})$
- there wasn't mixing but the event wasn't correctly tagged, with a frequency of $f_{SCu} = (1 - \chi_d)\omega_{tag}$

which means the fraction of events with the same charge is

$$f_{SC} = \chi_d(1 - \omega_{tag}) + (1 - \chi_d)\omega_{tag} = \chi_d + \omega_{tag} - 2\chi_d\omega_{tag} \tag{4.1}$$

The opposite sign-flavor correlation happens when

- the $B^0\overline{B^0}$ mixing happened and the event is wrongly tagged, with a frequency of $f_{OCm} = \chi_d\omega_{tag}$

- there wasn't mixing and the flavor was tagged correctly, with a frequency of $f_{OCu} = (1-\chi_d)(1-\omega_{tag})$

which means the fraction of events with the opposite charge is

$$f_{OC} = \chi_d\omega_{tag} + (1-\chi_d)(1-\omega_{tag}) = 1 + 2\chi_d\omega_{tag} - \chi_d - \omega_{tag}$$

Where $f_{SC} + f_{OC} = 1$.
Using equation 4.1 we can calculate ω_{tag} from f_{SC} and χ_d

$$\omega_{tag} = \frac{f_{SC} - \chi_d}{1 - 2\chi_d} \tag{4.2}$$

Note that we are only considering mixing on the reconstructed side, not on the opposite side. This is because the ω_{tag} already accounts for the mixing of the eventual neutral B produced on the opposite side.

4.1.1 Correction for the Same Side tagger mistag

The Same Side tagging will be discussed in section 4.2 but it will have his own mistag fraction (that will be refered as ω_{ss}, so the calculation of ω_{tag} will need to be corrected accounting that mistag.

The OS-tagger and SS-tagger will have the same charge if

- there was mixing and both taggers identified the flavour correctly, with frequency $f_{SC1} = \chi_d(1-\omega_{tag})(1-\omega_{ss})$
- there wasn't mixing and the OS-tagger misidentified the flavor, with frequency $f_{SC2} = (1-\chi_d)\omega_{tag}(1-\omega_{ss})$
- there wasn't mixing and the SS-tagger misidentified the flavor, with frequency $f_{SC3} = (1-\chi_d)(1-\omega_{tag})\omega_{ss}$
- there was mixing and both the taggers misidentified the flavor, with frequency $f_{SC4} = \chi_d\omega_{tag}\omega_{ss}$

The fraction of events with the same charge is the sum of the above,

$$f_{SC} = \chi_d + \omega_{ss}(1-2\chi_d) + \omega_{tag}(1-2\chi_d)(1-2\omega_{ss})$$

.

The OS-tagger and SS-tagger will have the opposite charge if

- there wasn't mixing and both taggers identified the flavor correctly, with frequency $f_{OC1} = (1-\chi_d)(1-\omega_{tag})(1-\omega_{ss})$
- there was mixing and the OS-tagger misidentified the flavor, with frequency $f_{OC2} = \chi_d\omega_{tag}(1-\omega_{ss})$

- there was mixing and the SS-tagger misidentified the flavor, with frequency $f_{OC3} = \chi_d(1-\omega_{tag})\omega_{ss}$
- there wasn't mixing and both the tagger misidentified the flavor, with frequency $f_{OC4} = (1-\chi_d)\omega_{tag}\omega_{ss}$

The fraction of events with the same charge is the sum of the above,

$$f_{OC} = 1 - \chi_d - \omega_{ss}(1-2\chi_d) - \omega_{tag}(1-2\chi_d)(1-2\omega_{ss})$$

.

Where $f_{SC} + f_{OC} = 1$.
The mistag of the opposite side tagger therefore becomes

$$\omega_{tag} = \frac{f_{SC} - \chi_d - \omega_{ss}(1-2\chi_d)}{1-2\chi_d-2\omega_{ss}(1-2\chi_d)} \tag{4.3}$$

It can be noticed that the correction introduced by the Same Side tagger is equivalent to changing the mixing probability χ_d to a new $\chi' = \chi_d + \omega_{ss}(1-2\chi_d)$

4.2 Same Side tagging using the K^*

Events are selected corresponding to the decay $B^0 \to J/\psi K^*$, where K^* is the $K^*(892)^0$, with a mass of $M_{K^*} = 895.81 \pm 0.19$ MeV, and whose main decay mode is $K^* \to K^+\pi^-$ ($\overline{(}K^*) \to K^-\pi^+$) with a lifetime so small that there is a negligible mixing probability. The charge of the $K^\pm$ gives the flavor of the K^* that gives the flavor of the B^0.

The issue is that CMS does not have a particle identification detector (like a Cherenkov detector) so there is no information about which of the two charged tracks that are produced in the K^* decay is the K and which is the π.

To solve the problem two hypothesis are created

1. the first particle is a K and the second is a π
2. the first particle is a π and the second is a K

and for each of them the event is reconstructed, leading to two mass hypothesis for the K^* and for the B^0, as well as two hypothesis for all their kinematic variables. Two main approaches where implemented to choose which particle is the K, cuts on the K^* invariant mass and a multivariate analysis.

4.2.1 Invariant Mass Cuts

The idea is that the combination with invariant mass closer to the K^* should be the right one. The same cuts already used in other analysis [12] were implemented. At least one of the two hypothesis must be within 80 MeV of the mass of the K^* and the closest to the value is chosen. If both the hypothesis are within 50 MeV of the mass of the K^* the

Table 4.1: Efficiency and mistag fraction with SS tagger with cuts on the invariant mass of the K^*.

	Efficiency	ω_{ss}	$\mathcal{P}_{ss}$
Invariant mass cuts	0.8347 ± 0.0019	0.0790 ± 0.0004	0.5916 ± 0.0135
One hyp within 80 MeV	0.7032 ± 0.0018	0.0544 ± 0.0004	0.5584 ± 0.0008

event is ignored since its width is $\Gamma_{K^*} = 47.4 \pm 0.6$ MeV and we can't choose one over the other. Efficiencies of this method, as well as the wrong tag fraction and the tagging power are presented in the first row of table 4.1.

An attempt to reduce the wrong tag fraction was performed selecting the events with only one hypothesis within 80 MeV in order to increase the number of the correctly tagged events although it means losing efficiency. The results are presented in the second row of table 4.1. The wrong tag fraction can be reduced to the 5.4%, but to furtherly reduce it a more sophisticated analysis is pursued.

4.2.2 Multivariate Analysis

To furtherly reduce the number of mistagged events a multivariate approach was tried. The two categories were defined as

1. the first track is the K and the second the π
2. the first track is the π and the second is the K

The used method was, once again, the recommended Artificial Neural Network of the TMVA package, the multilayer perceptron, MLP. The variables that were deemed interesting for this study where the two hypothesis of the K^* mass, p_T, η and ϕ of the two tracks. Composite variables were also tried, for example the difference between the two mass hypothesis and p_T. In the end the best combination of variables, in terms of minimizing ω_{ss} where

- Mass of the K^*, first hypothesis
- Mass of the K^*, second hypothesis
- transverse momentum p_T of the first track
- transverse momentum p_T of the second track
- η of the first track
- ϕ of the first track

There is a clear asymmetry in the chosen variables, the reason is that the correlation between ϕ of the first and of the second track is 85% and the correlation between η of the two tracks is 98%. Such correlated variables don't introduce new informations and

the best course of action is removing them. The input variables, the correlations and the output of the MLP method, with training and testing sample superimposed are presented in figure 4.1, 4.2 and 4.3 respectively.

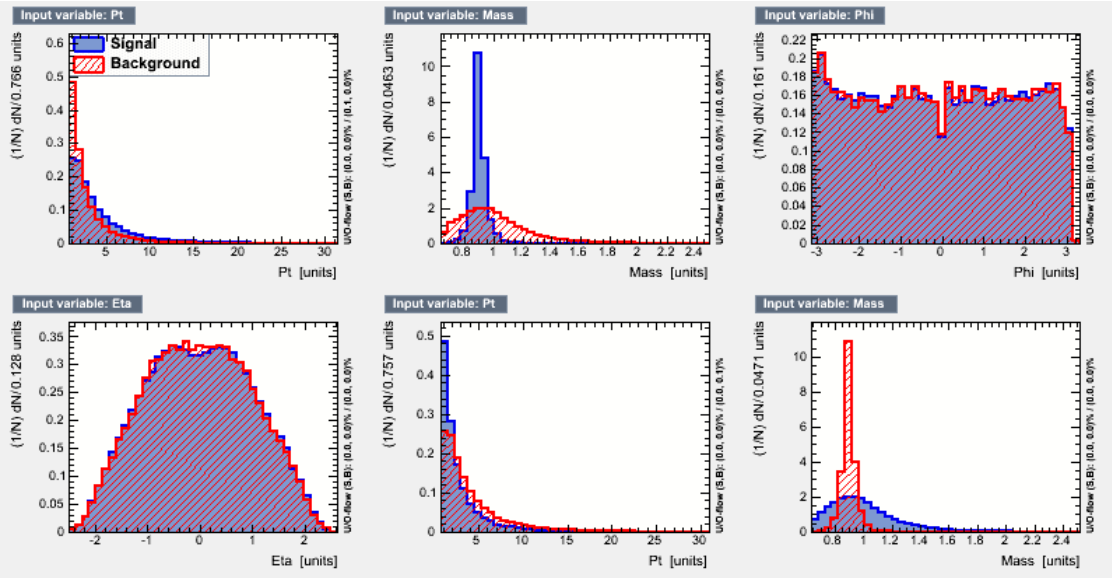
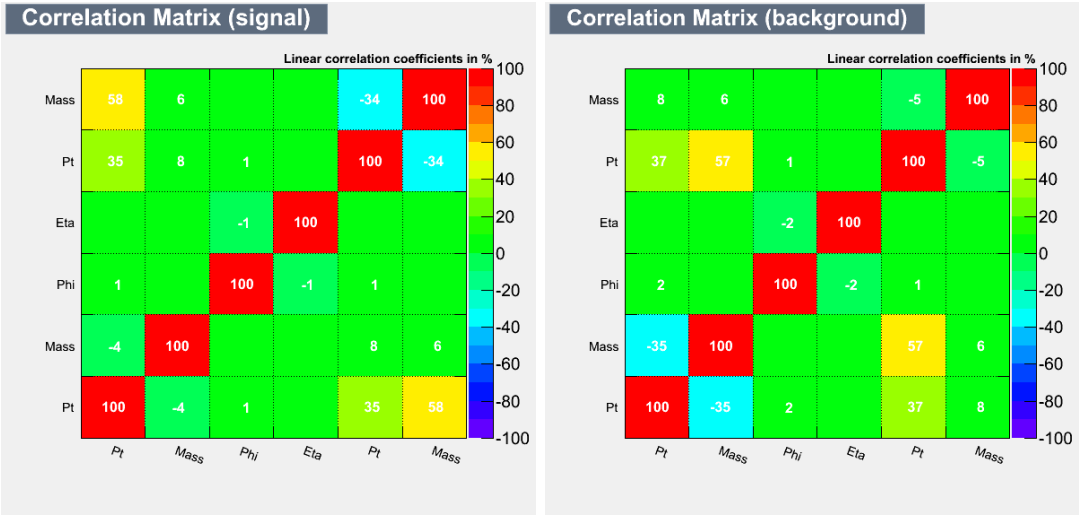


Figure 4.1: Input variables of the MLP method. In blu the first category, in red the second one.



(a) First category *(b)* Second category

Figure 4.2: Linear correlation coefficients for the variables used in the MLP method.

As can be seen in figure 4.3 there is a good separation between the categories, except for the central region, where the probability of one category being right and the other

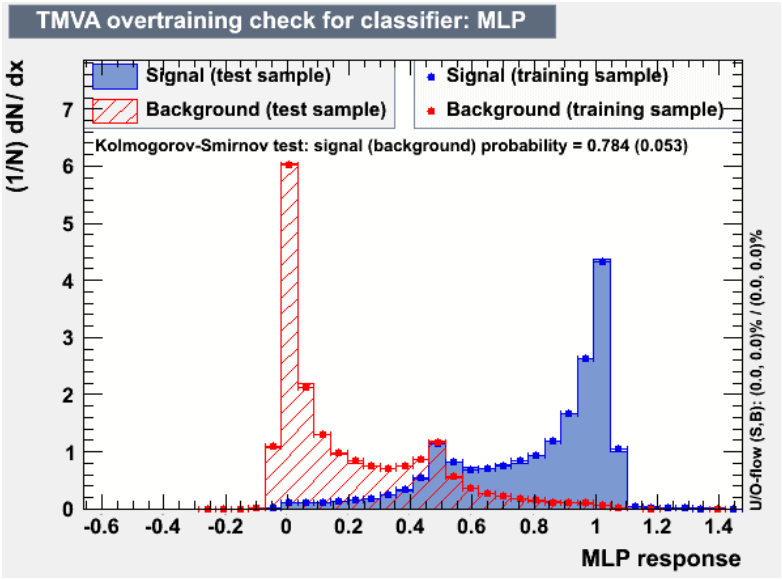


Figure 4.3: Output of the MLP method, in blu the first category, in red the second one.

wrong is $\sim 50\&$. The tagger is therefore defined as

- if MLP > 0.5 the first hypothesis is assumed correct
- if MLP < 0.5 the second hypothesis is assumed correct

To increase the rightly tagged event fraction a cut in the central region was implemented. The cut itself was defined as

$$|\text{MLP} - 0.5| > \text{Cut}$$

and different cuts were tried. Efficiency, wrong tag fraction, and tagging power are reported in table 4.2, as well as the results for the whole range, without cuts.

In the end the cut at 0.35 was chosen because with a 11% efficiency loss, with respect to the results obtained with the mass cuts method reported in table 4.1, there is a reduction of the wrong tag fraction of 40%.

4.2.3 Data Sample verification

Before applying the cuts to the Data sample the distribution of the MLP variable in the Data sample was checked. In order to do so, the Data sample was divided in 24 bins in the MLP range and for each of them a fit in the B^0 invariant mass, chosen with the MLP method, was performed. The background was described as a decreasing exponential, while the signal was described as the sum of two gaussians with the same mean. Two graphs are reported as examples in figure 4.4.

The two invariant mass distributions have different width because in the central region of the MLP there is a high mistag fraction, so the assumed hypothesis is wrong and the chosen value of the B^0 mass is wrong, and this is reflected in the gaussians width. Since

Table 4.2: Efficiency and wrong tagging study for different cuts for the MLP variable, with $|MLP-0.5| >$ Cut

Cut	Efficiency	ω_{ss}	$\mathcal{P}_{ss}$
-	1.000 ± 0.002	0.1287 ± 0.0005	0.5514 ± 0.0008
0.05	0.907 ± 0.002	0.0949 ± 0.0005	0.5952 ± 0.0009
0.10	0.851 ± 0.002	0.0774 ± 0.0004	0.6074 ± 0.0009
0.15	0.805 ± 0.002	0.0655 ± 0.0004	0.6077 ± 0.0009
0.20	0.763 ± 0.002	0.0558 ± 0.0004	0.6022 ± 0.0009
0.25	0.720 ± 0.002	0.0474 ± 0.0004	0.5902 ± 0.0009
0.30	0.674 ± 0.002	0.0397 ± 0.0004	0.5713 ± 0.0009
0.35	0.622 ± 0.002	0.0326 ± 0.0003	0.5432 ± 0.0008

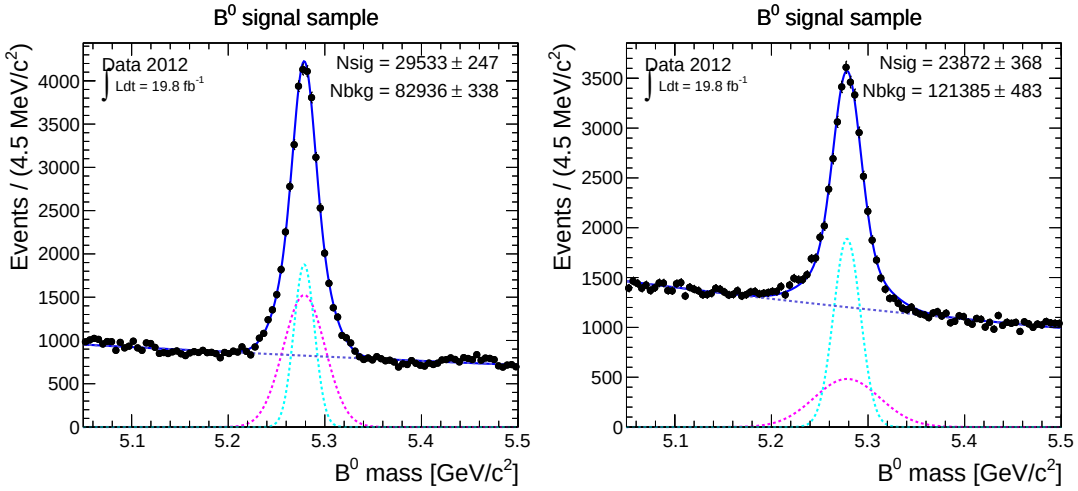
Table 4.3: Efficiency and wrong tagging study for different cuts for the MLP variable, with $|MLP-0.5| >$ Cut after the reweighting.

Cut	Efficiency	ω_{ss}	$\mathcal{P}_{ss}$
-	1.000 ± 0.002	0.1399 ± 0.0005	0.519 ± 0.001
0.05	0.899 ± 0.002	0.1043 ± 0.0005	0.563 ± 0.001
0.10	0.835 ± 0.002	0.0848 ± 0.0005	0.576 ± 0.001
0.15	0.785 ± 0.002	0.0720 ± 0.0004	0.575 ± 0.001
0.20	0.736 ± 0.002	0.0606 ± 0.0004	0.568 ± 0.001
0.25	0.687 ± 0.002	0.0507 ± 0.0004	0.554 ± 0.001
0.30	0.636 ± 0.002	0.0421 ± 0.0004	0.534 ± 0.001
0.35	0.579 ± 0.001	0.0339 ± 0.0004	0.503 ± 0.001

what we want to obtain is the number of events, this is not concerning, as long as the fit describes properly the distribution.

The MLP distribution in the Data sample and in the MC sample superimposed can be seen in figure 4.5.

The two distributions are tangibly different, so the MC sample need to be reweighted to account for this difference. The efficiency and wrong tag fraction were recalculated to account for that, the results can be found in table 4.3. The same cut was kept, even if the efficiency is decreased and ω_{SS} is bigger.



(a) MLP $\in (-0.05, 0.00)$ *(b)* MLP $\in (0.45, 0.50)$

Figure 4.4: Two examples of fit in the B^0 invariant mass in different ranges of MLP.

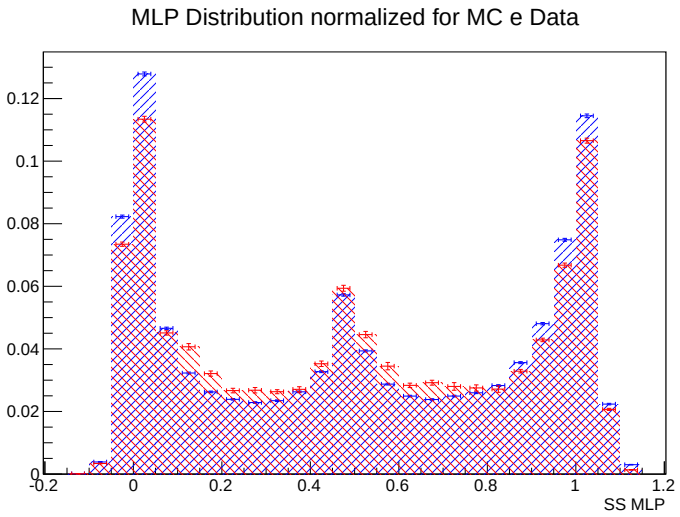


Figure 4.5: MLP distribution for the Data sample in red and for the MC sample in blue.

4.3 χ_d calculation on the MC

The measured value of χ_d is $\chi_d = 0.1873 \pm 0.0024$ but the value is expected to be different due to the trigger cuts. The value is expected to be higher since the `HLT4_Jpsi_Displaced_v9-v12` trigger requires that the J/ψ candidate has $L_{xy}/\sigma(L_{xy}) > 3$ and this translates in a suppression of the events with short lifetime that are mainly unmixed.

This study was performed on the MC sample, using the MC truth to determine if an event had mixed or not the mixed, unmixed and total proper time distribution are shown in figure 4.6. The resulting χ_d values for different proper time ranges are reported in table 4.4. In the end χ_d for the range $[0, \infty]$ was used to avoid further efficiency loss.

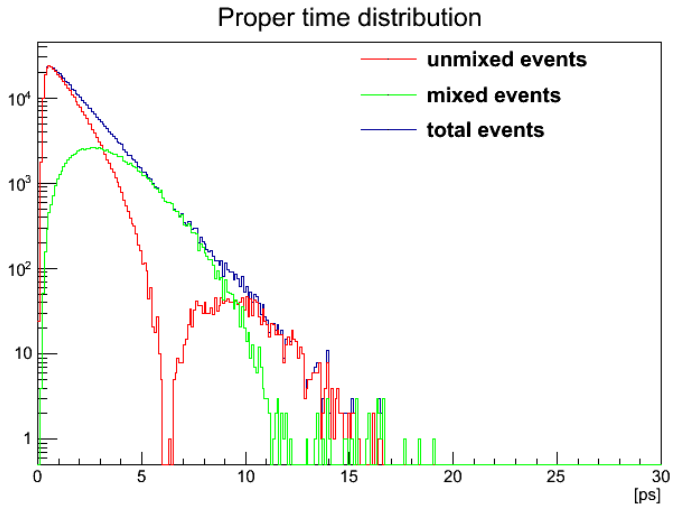


Figure 4.6: Proper time distribution for all the events, and the mixed and und unmixed subsamples.

Table 4.4: χ_d evaluated on the MC sample in various time ranges

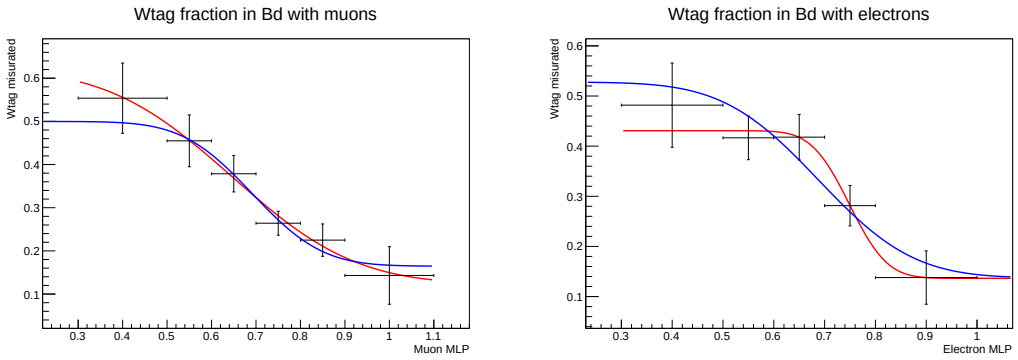
Time [ps]	χ_d
$[0; 1.5]$	0.0488 ± 0.005
$[0; 2.25]$	0.0867 ± 0.005
$[0; 3.0]$	0.1255 ± 0.006
$[0; 1.5]$	0.2294 ± 0.008

4.4 w_{tag} measurement on the Data sample

In order to measure the mistag fraction the Data sample was divided in 6 bins in the muon tagger MLP and 5 bins in the electron tagger MLP and the invariant mass of the B^0 was fitted to obtain the number of tagged events and the number of events that were

tagged with the same charge. In this way it is possible to evaluate the wtag fraction as explained in equation 4.3.

The functions were fitted with $f = p_0 + p_1(1 - \mathrm{Erf}(p_2 + p_3 \cdot MLP))$ even if the parameters have big errors due to the small number of points. The fit, as well as the function evaluated from the B^+ data sample, are presented in figure 4.7, the fit results can be found in table 4.5.



(a) Muons *(b)* Electrons

Figure 4.7: Wrong tag distribution as a function of the lepton MLP for the B^0 Data sample

Table 4.5: Parameter of the calibration fit

	muon	electrons
p_0	0.1 ± 0.2	0.14 ± 0.06
p_1	0.2 ± 0.2	0.15 ± 0.04
p_2	-2 ± 3	-9 ± 1
p_3	3 ± 4	1 ± 2

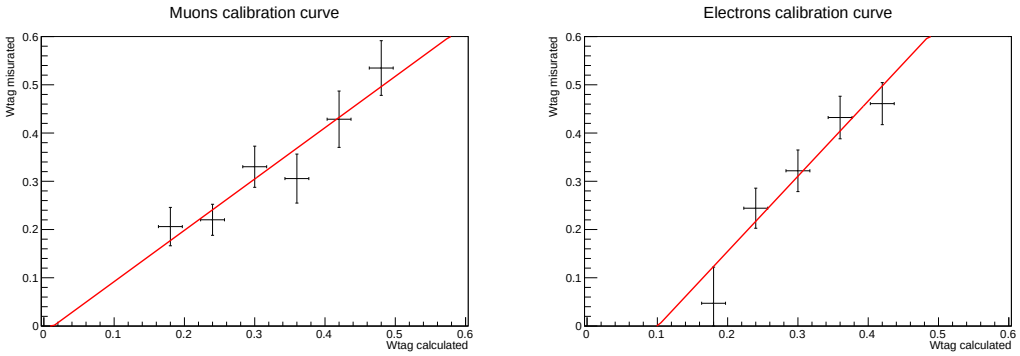
In order to do another check about the compatibility with the tagger in the B^+ data sample a study of the mistag fraction in bins of the calculated mistag fraction for the B^+ data sample was performed. Once again 6 bins were used for the muon and five for the electron. The function was the fitted with the function

$$\omega_{mis} = p_0 + p_1(\omega_{calc,B^+} - 0.35)$$

as shown in figure 4.8. The fit results are reported in table 4.6;

Table 4.6: Parameter of the calibration fit

	muon	electrons
p_0	0.36 ± 0.02	0.39 ± 0.03
p_1	1.1 ± 0.2	1.6 ± 0.3



(a) Muons

(b) Electrons

Figure 4.8: Two examples of fit in the B^0 invariant mass in different ranges of MLP.

The two measures are the same if $p_1 = 1$ and $p_0 = p_1 \cdot 0.35$. It may seem that the electron MLP tag is not compatible between B^0 and B^+ but due to the large errors in the measure it is in fact compatible.

Chapter 5

Conclusion

In this thesis a multivariate analysis was developed to tag the flavor of B neutral mesons at production time by exploiting the pair production of the $b\bar{b}$ quark to link the charge of an opposite side lepton (muon or eletron) to the flavour of the meson creating an opposite side lepton tagger. The independence of the tagger from the reconstructed B decay was checked by applying it to three simulated decay channels, $B_S^0 \to J/\psi\phi$, $B^+ \to J/\psi K^+$ and $B^0 \to J/\psi K^*$ which proved to be compatible. To measure its performances the tagger was applied to one self tagging data sample, $B^+ \to J/\psi K^+$.

To further check the independence of the tagger from the decay channel a multivariate same side tagger was developed for the decay $B^0 \to J/\psi K^*(K^+\pi^-)$ to inquire the flavor at decay time using the decay properties. Once the multivariate algorithm had identified the K and the π the tagger used the $K^* \to K^+\pi^-$ decay to connect the charge of the K^+ to the flavor of the B^0.

The performance of the OS lepton tagger on the $B^0 \to J/\psi K^*$ data sample were compatible with that of the B^+ sample.

The developed algorithms can be used in all the analysis that require b-flavour tagging and was used for CP violation measurements, in particular for the measurement of the CP violating phase ϕ_S in the channel $B_S^0 \to J/\psi\phi$

Appendix A

Optimization

The MLP method has a lot of options, some of which were deemed reasonable to improve the outcome, and will be discussed here. After deciding the set of variables it became necessary to choose the best combination of values for these options. Firstly, all of them were checked separately to see which ones induced some changes in the discrimination power. After this we collected the options of interest and tested them at the same time.

Training methods

During the training, for each training event the neural network output is computed and compared with the correct output. An error function that measures the agreement of the network response is computed and the set of weights that minimize the error function are updated to be used as input weight.

The two training methods differ in how the weights are updated. The first is called Back Propagation (BP) and uses first derivatives of the Error Function; the second is called the Broyden-Fletcher-Goldfarb-Shannon Algorithm (BFGS) and uses second derivatives. BFGS requires a smaller ammount of iterations but requires a time for one iteration proportional to the square number of the synapses, so it requires a longer time than BP. The ROC Integrals can be seen in Tab A.1

The BFGS Algorithm returned a better ROC integral so it was used in subsequent optimizations. It has two parameters that also need optimization, Tau and ResetStep. Tau is the learning rate, and is a parameter that defines how much the weights are changed in one epoch; it's default value is Tau = 3. The intermediate steps in the BFGS process become less accurate with the increasing number of iterations and are reset every ResetStep; it's default value is 50.

Table A.1: Training methods. The default value was BP.

Training	BP	BFGS
ROC integral	0.669	0.674

Table A.2: Variations in the number of cycles. The default value was 600.

Number of cycles	300	600	900
ROC integral	0.668	0.669	0.669

Table A.3: Variations in the hidden layers structure where N is the number of input variables. The defaul value is N+5 that means 1 hidden layer with N+5 neurons. Options with two hidden layers were tried.

Layers	N+5	N,N-1	N+2,N-2
ROC integral	0.669	0.664	0.665

This two options were only tested in subsequent optimizations.

Number of cycles

It is the number of cycles (*epochs*) executed while training the MLP. The defaul value used in previous analysis was 600. The values tried in the optimization can be seen in Tab A.2.

Hidden Layers

Both the number of hidden layers and the number of neuron in each layer are configurable. Increasing the number of neurons increases the precision of the MLP, but it also increases the training time. Increasing the number of layers can produce a similar performance with a smaller training time, a more robust network and less neurons. In the optimization both the number of layers and of neurons are changed as can be seen in Tab A.3.

Since both trials with two hidden layers resulted in a lower ROC integral, in the next optimazation different numbers of neurons for one layer were tried.

Neuron response function

The neuron response function ρ maps the neuron input (a vector of n entries) onto the neuron output. It can be separated in a $\Re^n \to \Re$ *synapse function* α and a $\Re \to \Re$ *neuron activation function* κ, so that $\rho = \alpha \circ \kappa$. Different forms for both α and κ can be chosen from preselected ones as can be seen in Tab A.4 and A.5.

The synapse function was found faulty in the training process if the sqsum or abssum forms were assigned, so it was decided to discard them and leave the default form.

Table A.4: Variations in the synapse function. The default value was sum.

Neuron input types	sum	sqsum	abssum
ROC integral	0.669	0.000	0.000

Table A.5: Variations in the neuron activation function. The default value was tanh.

Neuron types	tanh	sigmoid
ROC integral	0.669	0.667

Table A.6: Variations in the estimator type. The default value was the Mean Square Estimator (MSE).

Estimator types	MSE	sigmoid	tanh
ROC integral	0.669	0.669	0.669

Estimator Type

Estimator can have different forms, the optimization can be seen in Tab A.6.

The estimator type didn't change the ROC integral so it was left at its default value.

A.1 First concomitant testing

As an intermediate step all the options already discussed that needed further optimization were tried with a small number of values each (no more than 3) as can be seen in Tab A.7. The goal was to understand which of these options were really meaningful in order to increase the number of values only for those, and also understand the meaningful range of values for options such as the number of cycles and the layer structure. This time the values were not tested separately: for each value of an options all the value of the other options were tried. This led to 108 different trainings.

After looking at all the ROC values the only variable we fixed was Tau. Its default value Tau = 3 was kept.

A.2 Final concomitant testing

For the final optimization the values found in Tab A.8 were tried for a total of 250 trainings.

Table A.7: .

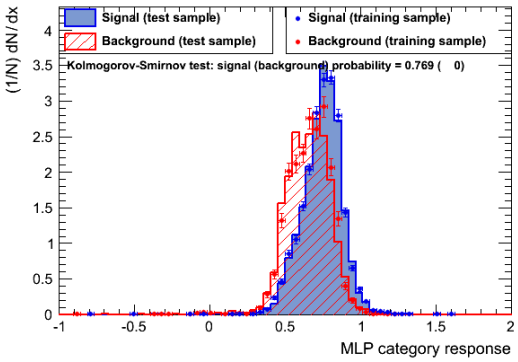
Option			
NCycle	400	600	800
HiddenLayer	N+3	N+5	N+7
NeuronType	tanh	sigmoid	
Tau	3	5	
ResetStep	25	50	100

Table A.8: .

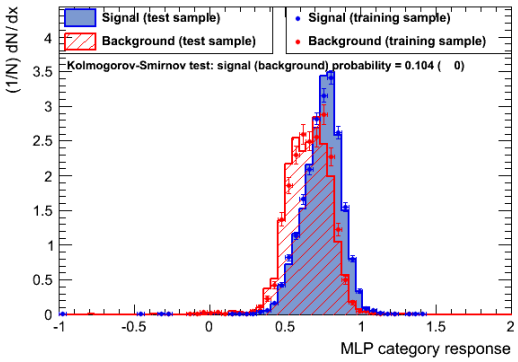
Option					
NCycle	450	550	650	750	850
HiddenLayer	N+2	N+4	N+5	N+6	N+8
NeuronType	tanh	sigmoid			
ResetStep	10	25	50	75	100

The best ROC Integral value was 0.682 that is 1.2% better than the value before the optimization (0.669). However, the optimization that lead to this value doesn't have a good agreement between the training and testing sample, as can be seen in figure A.1(a) and A.1(b). The first one is the result for 750 cycles with N+6 neurons in the hidden layer, tanh as the neuron type and ResetStep = 100; the second one is for 650 cycles with N+6 neurons, tanh as the neuron type and ResetStep = 75 and has a ROC integral of 0.681.

In the end, the optimization had a ROC integral value of 0.679 and 750 cycles, N+8 neurons, neuron type tanh and ResetStep = 100. The overtrain graph can be seen in Fig A.2



(a) ROC int = 0.682



(b) ROC int = 0.681

Figure A.1: Two examples of estimators with a good value of ROC integral but with poor agreement between train and test sample, particularly for the same sign class (in red)

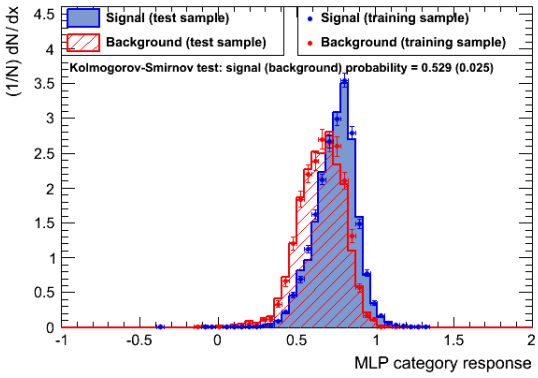


Figure A.2: The muon estimator with the final set of option values. In red the same sign class, in blue the opposite sign class

Appendix B

Bibliography

[1] Georges Aad, E Abat, J Abdallah, AA Abdelalim, A Abdesselam, O Abdinov, BA Abi, M Abolins, H Abramowicz, E Acerbi, et al. The ATLAS experiment at the CERN large hadron collider. *Journal of Instrumentation*, 3(08):S08003, 2008.

[2] R Aaij, C Abellan Beteta, B Adeva, M Adinolfi, C Adrover, A Affolder, Z Ajaltouni, J Albrecht, F Alessio, M Alexander, et al. Measurement of the cp-violating phase ϕs in decays. *Physics Letters B*, 713(4):378–386, 2012.

[3] T Aaltonen, B Álvarez González, S Amerio, D Amidei, A Anastassov, A Annovi, J Antos, G Apollinari, JA Appel, T Arisawa, et al. Measurement of the bottom-strange meson mixing phase in the full cdf data set. *Physical review letters*, 109(17):171802, 2012.

[4] Kenneth Aamodt, A Abrahantes Quintana, R Achenbach, S Acounis, D Adamová, C Adler, M Aggarwal, F Agnese, G Aglieri Rinella, Z Ahammed, et al. The alice experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08002, 2008.

[5] VM Abazov, B Abbott, BS Acharya, M Adams, T Adams, GD Alexeev, G Alkhazov, A Alton, G Alverson, GA Alves, et al. Measurement of the c p-violating phase ϕ s j/ψ ϕ using the flavor-tagged decay $b_s^0 \to j/\psi\ \phi$ in 8 fb-1 of p p^- collisions. *Physical Review D*, 85(3):032006, 2012.

[6] S Agostinelli, John Allison, K al Amako, J Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[7] A Augusto Alves Jr, LM Andrade Filho, AF Barbosa, I Bediaga, G Cernicchiaro, G Guerrer, HP Lima Jr, AA Machado, J Magnin, F Marujo, et al. The LHCb detector at the LHC. *Journal of Instrumentation*, 3(08):S08005, 2008.

[8] Elisabetta Barberio, Bob van Eijk, and Zbigniew Was. Photos—a universal monte carlo for qed radiative corrections in decays. *Computer Physics Communications*, 66(1):115–128, 1991.

[9] Elisabetta Barberio and Zbigniew Was. Photos-a universal monte carlo for qed radiative corrections: version 2.0. *Computer Physics Communications*, 79(2):291–308, 1994.

[10] GL Bayatyan, N Grigorian, VG Khachatrian, AT Margarian, AM Sirunyan, W Adam, M Brügger, J Erö, M Fierro, M Friedl, et al. Cms: the tridas project technical design report; v. 1, the trigger systems. 2000.

[11] J.P. Beaudette, D. Benedetti, and P.M. Electron reconstruction within the particle flow algorithm. *CMS Note*, 2010.

[12] Serguei Chatrchyan, A Apresyan, A Bornheim, J Bunn, Y Chen, E Di Marco, J Duarte, D Kcira, Y Ma, A Mott, et al. Angular analysis and branching fraction measurement of the decay $b^0 \to k_0^* \mu^+ \mu^-$. *Physics Letters B*, 727(1-3):77–100, 2013.

[13] Sergio Cittolin, Attila Rcz, and Paris Sphicas. Cms the tridas project: Technical design report, volume 2: Data acquisition and high-level trigger. *CMS trigger and dataacquisition project, Technical Design Report CMS, CERN, Geneva*, 2002.

[14] CMS Collaboration, S Chatrchyan, et al. The CMS experiment at the CERN LHC. *Jinst*, 3(08):S08004, 2008.

[15] CMs Collaboration, CMS Collaboration, et al. The muon project technical design report. *CERN/LHCC*, 32(404):1997, 1997.

[16] CMS collaboration, CMS Collaboration, et al. Particle-flow event reconstruction in cms and performance for jets, taus, and met. *CMS Physics Analysis Summary CMS-PAS-PFT-09-001*, 2009.

[17] Cms Collaboration et al. The cms electromagnetic calorimeter technical design report, 1997.

[18] CMS Collaboration et al. The cms hadronic calorimeter project: Technical design report. *CMS TDR, CERN, Geneva, CERN-LHCC-97-31*, 1997.

[19] CMS collaboration et al. Commissioning of the particle-flow event reconstruction with the first lhc collisions recorded in the cms detector. *CMS Physics Analysis Summary CMS-PAS-PFT-10-001*, 30, 2010.

[20] ATLAS Collaboration: G. Aad et al. Time-dependent angular analysis of the decay $b_s^0 \to j/\psi\phi$ and extraction of $\delta\gamma_s$ and the *cp*-violationg weak phase ϕ_s by atlas. *Journal of High Energy Physics*, 12(072), 2012.

[21] Marc Grabalosa Gándara, LHCb Collaboration, et al. Flavour tagging performance in lhcb. In *Journal of Physics: Conference Series*, volume 171, page 012103. IOP Publishing, 2009.

[22] Particle Data Group. Particle data group. http://pdg.lbl.gov/index.html. Accessed: 11/28/2014.

[23] Andreas Hoecker, Peter Speckmayer, Joerg Stelzer, Jan Therhaag, Eckhard von Toerne, Helge Voss, M Backes, T Carli, O Cohen, A Christov, et al. TMVA-toolkit for multivariate data analysis. *arXiv preprint physics/0703039*, 2007.

[24] Vishwajeet Jha. Measurement of b s lifetime difference. In *Proceedings of the twentieth DAE-BRNS high energy physics symposium: programme abstracts*, 2013.

[25] V Karimäki. The cms tracker system project. Technical report, CMS-TDR-005, 1997.

[26] David J Lange. The evtgen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 462(1):152–155, 2001.

[27] J Pazzini, F Simonetto, S Vigolo, T Jarvinen, and G Fedi. Flavour tagging studies. in press.

[28] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. Pythia 6.4 physics and manual. *Journal of High Energy Physics*, 2006(05):026, 2006.

[29] Thomas Speer, W Adam, R Frühwirth, A Strandlie, T Todorov, and M Winkler. Track reconstruction in the cms tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 559(1):143–147, 2006.

[30] Thomas Speer, Pascal Vanlaer, R Frühwirth, Wolfgang Waltenberger, and Kirill Prokofiev. Vertex fitting in the cms tracker. Technical report, CERN-CMS-NOTE-2006-032, 2006.